

An Optimized E-nose for Efficient Volatile Sensing and Discrimination

Gonçalo Santos^{1,*}, Cláudia Alves¹, Ana Carolina Pádua¹, Susana Palma¹,
Hugo Gamboa² and Ana Cecília Roque¹

¹UCIBIO, Departamento de Química, Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa,
2829-516 Caparica, Portugal

²Laboratório de Instrumentação Engenharia Biomédica e Física da Radiação (LIBPhys-UNL), Departamento de Física,
Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa, Monte da Caparica, 2829-516 Caparica, Portugal

Keywords: Electronic Nose, Volatile Organic Compounds, Machine Learning, Biomaterials.

Abstract: Electronic noses (E-noses), are usually composed by an array of sensors with different selectivities towards classes of VOCs (Volatile Organic Compounds). These devices have been applied to a variety of fields, including environmental protection, public safety, food and beverage industries, cosmetics, and clinical diagnostics. This work demonstrates that it is possible to classify eleven VOCs from different chemical classes using a single gas sensing biomaterial that changes its optical properties in the presence of VOCs. To accomplish this, an in-house built E-nose, tailor-made for the novel class of gas sensing biomaterials, was improved and combined with powerful machine learning techniques. The device comprises a delivery system, a detection system and a data acquisition and control system. It was designed to be stable, miniaturized and easy-to-handle. The data collected was pre-processed and features and curve fitting parameters were extracted from the original response. A recursive feature selection method was applied to select the best features, and then a Support Vector Machine classifier was implemented to distinguish the eleven distinct VOCs. The results show that the followed methodology allowed the classification of all the VOCs tested with 94.6% ($\pm 0.9\%$) accuracy.

1 INTRODUCTION

The human olfactory system is sophisticated and is able to detect spoiled food, to identify dangerous environments or pleasant experiences. Highly trained humans can even develop this sense beyond the usual limits. Hippocrates (430 B.C.), considered the father of medicine, could diagnose diseases by smelling patient's biological samples (Röck et al., 2008). Odours consist of sets of volatile organic compounds (VOCs). When VOCs flow through the human's nasal cavity, they reach the olfactory epithelium, where olfactory sensory neurons define action potentials up their own axons, until the olfactory bulb, which then transmits the signal to the brain, where recognition occurs (Barbosa et al., 2018).

The intriguing olfactory system has inspired researchers to develop technologies mimicking natural olfaction (Wasilewski et al., 2017; He et al., 2012; Pádua et al., 2018b).

The best example of artificial olfaction is the concept of electronic nose (E-nose). E-noses comprise of an array of semi-selective sensors (e.g. metal oxide semiconductor or MOS-based sensors, conduct-

ing polymers, piezoelectric quartz crystals) which change properties in the presence of VOCs (Arshak et al., 2004). Physico-chemical changes that occur in the sensors during VOCs exposure are typically converted to electrical signals by the transduction system. The acquired data is further analysed and processed and VOC recognition is performed.

A set of features is firstly extracted and selected from the signals obtained. (Yan et al., 2015) reviewed many feature extraction methods, including features extraction from the original response curves, curve fitting parameters and transform domains. As conclusion, the authors stated that the best method depends on the situation. Nevertheless, some advice was given: for steady-state features, normalization models are the best option; transient features, such as derivatives and integrals of original response curves, have more information than steady-state features (Yan et al., 2015).

After features selection, advanced computational techniques can be applied. Examples of linear classification methods are principal component analysis (PCA), linear discriminant analysis (LDA), functional discriminant analysis (FDA), partial least square dis-

criminant analysis (PLS-DA) and cluster analysis (CA); and examples of non-linear models include fuzzy logic and artificial neural networks (ANN) (Scott et al., 2007). Some articles reported the use of Support Vector Machine (SVM) for data acquired by E-noses, namely in classification problems e.g. to evaluate the quality of olive oil (Ordukaya and Karlik, 2017), to recognize fruit odors (Uçar and Özalp, 2017), and to discriminate alcoholic and non-alcoholic beers (Ghasemi-Varnamkhasti et al., 2015). SVM was also applied to address regression problems, such as the prediction of the concentration of formaldehyde (Xu et al., 2016).

Our research group has developed a new class of biomaterials for gas sensing. The sensing gels are composed of liquid crystal (LC) droplets self-assembled in the presence of ionic liquid, dispersed inside a biopolymeric matrix. These materials possess enormous versatility and have unique stimuli-responsive properties, changing their optical and electrical properties, when interacting with VOCs (Hussain et al., 2017). In order to explore the application of these materials in gas sensing, different E-nose versions were designed and assembled by our research group (Pádua et al., 2018b). After processing and analysing signals obtained from the interaction of VOCs with the sensing materials VOCs detection and identification can be made. The final goal is to achieve an accurate, miniaturized and user-friendly device that could run sample analysis in an automated way.

In the past, it was shown that 3 different compositions of biomaterials could accurately distinguish 11 VOCs, as seen by PCA (Hussain et al., 2017). In this work an optimized E-nose device combined with sophisticated signal processing methods, aimed to show that a single biomaterial sensor could automatically classify 11 different VOCs.

2 THE OPTIMIZED E-NOSE DEVICE

The presented E-nose explores the optical properties of the sensing films. Therefore, a light source and a light detector are required, separated by two crossed polarizers and a sensing film. Without sensing film, light does not reach the light detectors due the presence of crossed polarizers. When the sensing film is present, its optical properties rotate polarized light allowing it to reach the photodetectors. This is the detection principle of the E-nose V3. Previous E-nose versions used handmade solder boards with through-hole components, which would usually led to align-

ment problems between the light emitters and the light detectors. Those versions used LDRs (Light Dependent Resistors) as light detectors. Present version uses PCBs (Printed Circuit Boards) with SMD (Surface Mount Devices) components combined with tailor-made 3D printed structures to guarantee proper alignment. Also, the LDRs were replaced by photodiodes which are reported to be more sensitive. These changes allowed to assemble an improved version with increased viability.

The E-nose and its operation method are shown in Figure 1(a) and 1(b), respectively. The apparatus is composed of a delivery system, a detection system, and an acquisition and control system. All presented CAD images were designed using *Fusion 360* by *Autodesk* and printed with PLA (polylactid acid) using an *Ultimaker 3*.

2.1 Delivery System

The delivery system is composed by two air pumps, the sample chamber and the relays (see Figure 1(a)).

The sample chamber is inline with one of the pumps, being called the exposure pump, before connecting to the detection chamber. This allows the air containing VOCs in the sample chamber to be carried into the detection chamber. The other pump, the recovery pump, connects directly to the detection chamber as its purpose is to restore the initial conditions in this last. Both pumps use ambient air.

In a hierarchical order, the control system informs the relays when and which pump should they power on, then the pump either carry the VOCs to the detection chamber or renew the air inside it, depending on which pump was powered. By switching the activation of the pumps, the VOC exposure/recovery cycles are generated.

2.2 Detection System

The detection chamber is composed of an external chamber with a lid, and an internal structure. Inside this last, an optical detection system serves as a transducer to the sensors that change their optical properties on Polarised light in contact with VOCs.

2.2.1 External Chamber

The 3D drawings of the external chamber with a lid are shown in Figure 2(a).

The main goal of the external detection chamber (see Figure 2(b)) is to apply pressure on the glass chamber (part of the internal chamber), compressing it between the EPDM (ethylene propylene diene

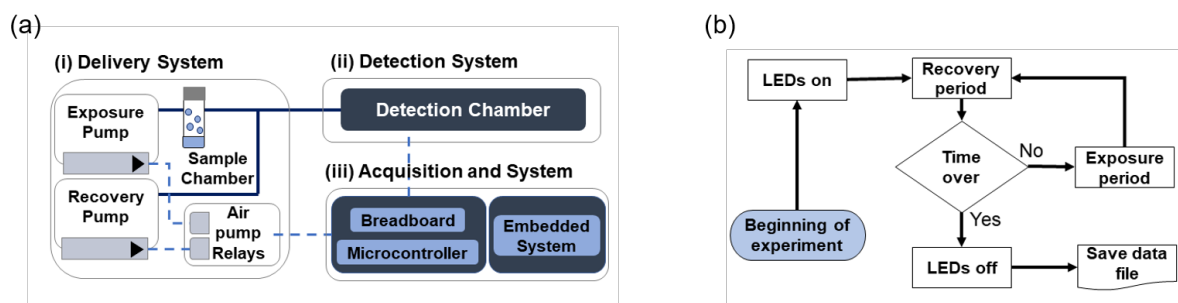


Figure 1: (a) Schematic of E-nose V3, composed by (i) Delivery system, (ii) Detection system, and (iii) Acquisition and control system. (b) Flow of instructions given by the embedded system.

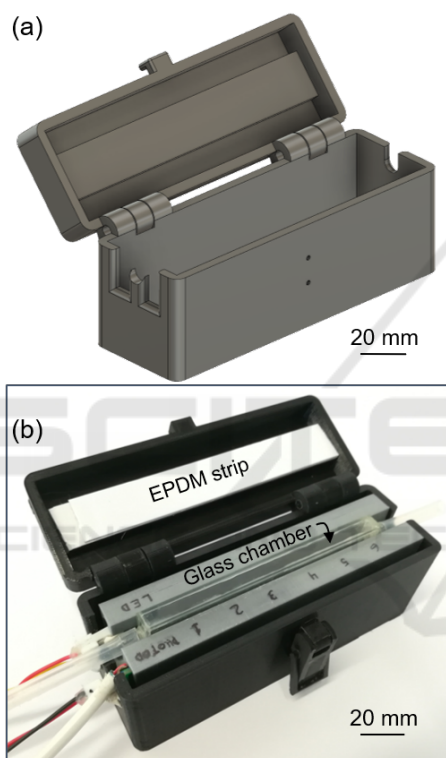


Figure 2: (a) 3D drawing of the external detection chamber. (b) Printed external detection chamber, enclosing the printed internal structure and respective connections. Lid is open for illustration purpose.

monomer) rubber strip of the lid and the films support. This compression is enough to close the chamber hermetically, for the pressures being applied in the system. Besides, it also increases the sensor robustness and isolates the light sensors from ambient light.

2.2.2 Internal Structure

The internal structure (see 3D drawings in Figure 3) serve as support to the PCBs, the polarizing filters and the sensing films support. It has two slots to insert two PCBs: one that has an emission cir-

cuit with LEDs, and other that has a detection circuit based on photodiodes. These PCBs are symmetrically aligned in front of each other. As such, each LED is placed exactly in front of its corresponding photodiode. Crossed polarizers are placed in between the PCBs, filtrating the light that is not polarized by the sensing probes. Moreover, the light guiders prevent interference of the light from other neighbor LEDs on the results of each individual LED/photodiode pair. Figure 3(e) is a picture of the fully assembled internal structure.

Emission Circuit

The emission circuit is composed by six LEDs, wired in parallel with the voltage source. It is a basic LED circuit where the current is, theoretically, evenly distributed through all the LEDs (i.e. all LEDs emit light with equal intensity). Also, a transistor in the common ground line acts as a switch to control LEDs.

The LEDs circuit schematic can be found in Figure 4.

Detection Circuit

Photodiodes are one of the most popular sensor types for light-based measurements. In the present case, the analyte detection relies on photodiodes for precise light intensity measurement. The photodiodes circuit schematics can be found in Figure 5.

Photodiodes generate a current proportional to the photons that strike their active area. This current needs to be converted into voltage, since the microcontroller (Arduino Due) analog input pins read voltage variations, and not current variations. Therefore, a circuit using a current-to-voltage converter (in this case a transimpedance amplifier) was implemented. Here, there is a need to point out the importance of V_{DIV} (Figure 5) voltage. The value of V_{DIV} is established as a compromise between the output voltage window and the reverse polarization of the photodiodes. The maximum output voltage is defined directly by V_{DIV} through the opamp (Operational Amplifier) positive input, and the minimum is 0V, being

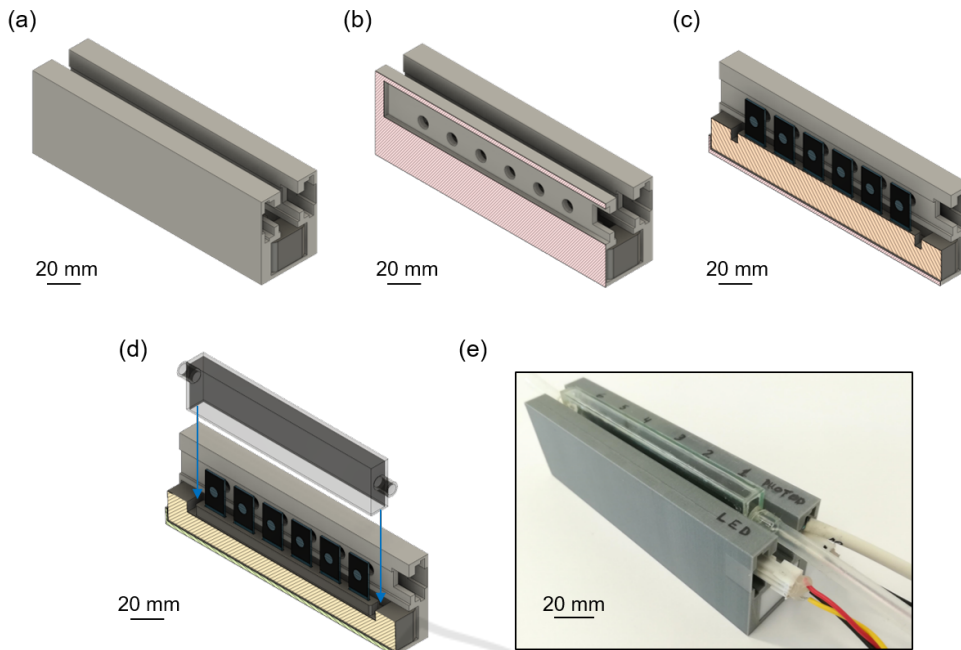


Figure 3: Main internal structure (light grey) and sensing films support (dark grey). The sensing films support is removable to ease the sensing films placement. **(a)** 3D drawing of the symmetrical internal structure. **(b)** Cross-section revealing one of the PCBs slot and respective light-guiders for the six LED/photodiode pairs. **(c)** Middle cross-section revealing the sensing films. **(d)** Middle cross-section revealing how the glass chamber fits in the sensing films support. **(e)** Assembled internal structure with respective components and connections.

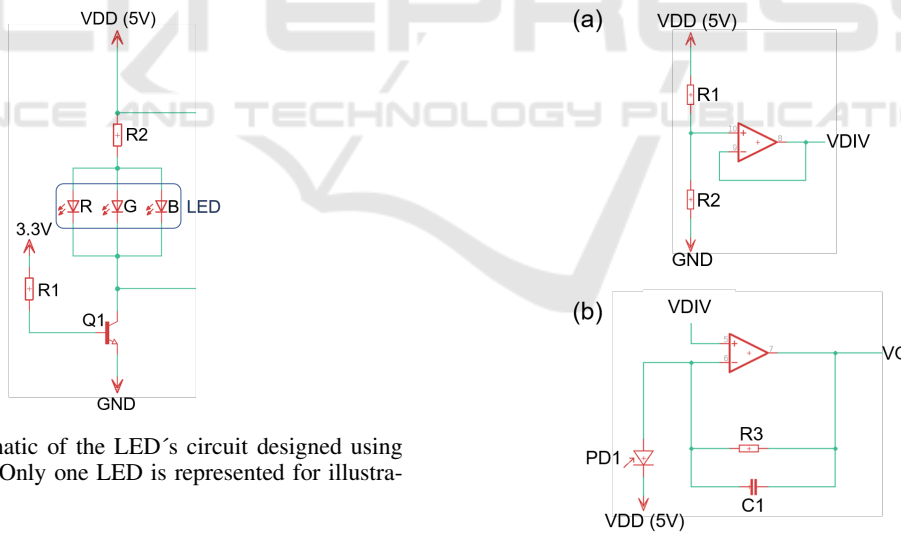


Figure 4: Schematic of the LED's circuit designed using Eagle software. Only one LED is represented for illustration purposes.

Figure 5: Schematic of photodiode's circuit designed using the software Eagle. **(a)** Generic voltage divider to generate V_{DIV} . **(b)** Schematic of photodiode's reverse polarizing circuit. One replica per photodiode.

the lowest value read by the arduino analog inputs (3.3 - 0V range). On the other hand, photodiode's reverse polarization voltage is the result of V_{DD} minus the opamp negative input voltage (or V_{DIV} , since both inputs share the same voltage). Increasing V_{DIV} will increase the output window but negatively impact the performance of the photodiodes, as they benefit from higher polarization voltages (become more sensitive to light). Opposite behaviour occurs if V_{DIV} is decreased.

A viable way to also control the system sensitivity to light is by resizing the resistor in the negative feedback loop, according to Equation 1 (Ohm Law). V_{DROP} is the voltage drop at the output, I_{PD} the current generated by the photodiode (PD) and R the re-

sistor value. The resizing of this resistor can be seen as an adjusting scale for the output signal, avoiding signal saturation.

$$V_{DROP} = R.I_{PD} \quad (1)$$

To confirm that the detection circuit is properly working without signal loss, either through signal saturation or insufficient sensitivity of the photodiodes to the lowest amounts of photons, one more test was performed.

A new setup (Figure 6) was assembled, including polarizing filters, and where a single LED/photodiode pair had the ability to rotate over the other. The experiment aimed to confirm that the photodiodes were sensitive enough to detect small light variations, specially close to the darkest point (i.e. when polarizers are crossed at 90°) corresponding to the maximum output voltage.

2.3 Acquisition and Control System

The instructions to control the device are given by an embedded system (Raspberry Pi 3 Model B) and executed by a microcontroller (Arduino Due). When the experiment begins, the LEDs are turned on, and the pumps work alternately while the photodiodes receive the signal and the microcontroller acquires them at a 90Hz sampling rate. At the end of the experiment, the LEDs are turned off and data is stored in a text file for later analysis (see Figure 1(b)). In terms of software, the latest version of the open access language Python (Python 3.6) was used with *Pymata* library, that allows the communication between the embedded system and the microcontroller.

3 CLASSIFICATION TOOLS

Together with the optimization of the device, analysis tools based on *Python* libraries (*SciPy*, *sklearn* and *novainstrumentation*) were also developed. A machine learning algorithm was trained with a set of selected features in order to distinguish 11 VOCs. In these tests, a single sensing film composition was used.

Before extracting the features of the signal, the signal was filtered using the median filter from *SciPy* library, kernel size equal to 11, to remove inductive noise caused mainly by the activation of the pumps. Then, the signal was filtered with *smooth* function from the *novainstrumentation* library. The E-nose records information that indicates when the recovery or the exposure pumps are working. This information was used to split the signal and obtain the individual

cycles (a cycle corresponds to exposure to gas sample followed by a recovery period, in which air restores the radial configuration of the LC-IL droplets, see Figure 7. Full phenomenon is described by (Hussain et al., 2017)). Every time there is an indication that the exposure pump is on, a cycle begins and a cut in the signal is made (Figure 8(a)).

After splitting the signal in individual cycles, those were normalised by Equation 2.

$$signal_{normalized} = \sum_n \frac{value_n - min(signal)}{max(signal) - min(signal)} \quad (2)$$

A time domain analysis was performed, combining two types of features: morphological features and parameters of curve fitting models.

A typical output signal, regarding an individual cycle, starts to rise when the sensors are being exposed to VOCs, until a maximum is reached which corresponds to the darkest state. On the other hand, the output signal decreases as the VOCs leave the chamber, restoring the radial configuration of the droplets and corresponding to the brightest state. This leads to a typical shape in the derivatives. Figure 8(b) represents a single cycle and its respective derivatives. The first derivative has two prominent peaks, with their position corresponding to the inflection points in the rise (positive peak) and the decrease (negative peak) of the signal. The second derivative has four prominent peaks; two positives when the signal is concave up and two negatives where it is concave down. The relative position of these peaks in relation to the beginning of the cycle, as well as its amplitudes, correspond to the first thirteen extracted features.

A suggestion made by R. Dutta and colleagues (Dutta and Dutta, 2006) was followed. Skewness and kurtosis which are measures of the asymmetry of the probability distribution of a random variable around its mean were extracted. Although the signal is not a probability density function, extracting these features helped increase the accuracy of their electronic nose classifier. Hence, they were included to see if could also be fitted for the current problem.

For the fitting model, the function considered was proposed by Holmberg, M. *et al.* (Equation 3) where θ_1 corresponds to signal amplitude, θ_2 to the rising slope, θ_3 to the position where the signal reaches half the rise, θ_4 to the descent slope and θ_5 to position where the signal reaches half of the descend (Holmberg et al., 1998). This model is based on the logistic function (Equation 4). The logistic function describes something that initially grows and that eventually starts to saturate and converge to a maximum value (in this case, total darkness, that occur when the droplets' configuration becomes isotropic). In Equat-

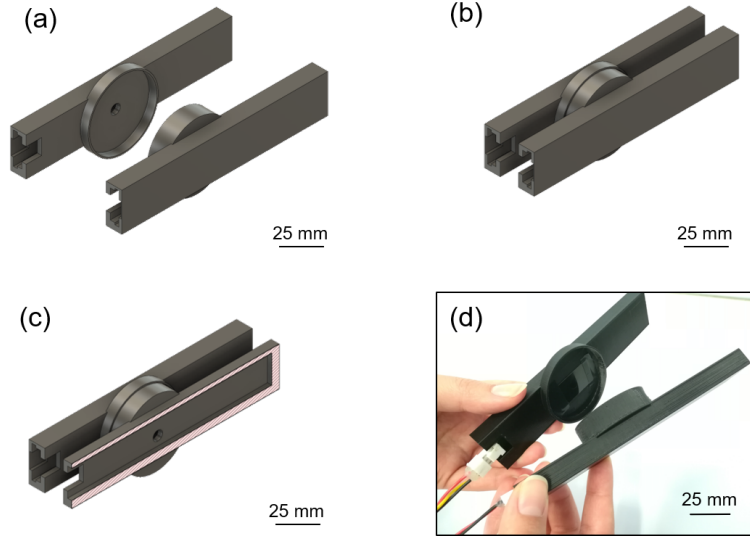


Figure 6: 3D printed structures designed to verify the detection circuit. (a) Separated pieces revealing the interior of the round chamber. (b) When together they create an isolated dark chamber, where the center point is the light-guider of the respective LED/photodiode pair. (c) Cross-section revealing the PCB slot. (d) Printed and assembled structure as used in the test.

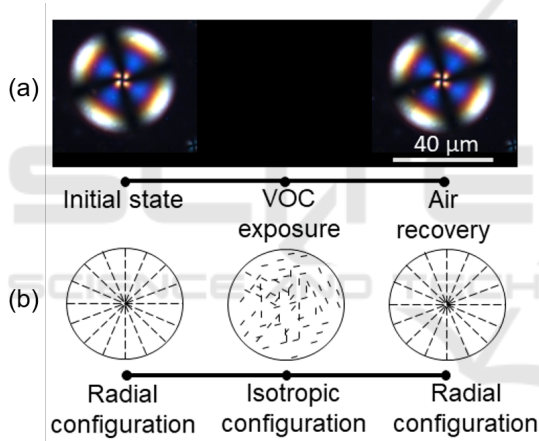


Figure 7: (a) Polarized Optical Microscopy image of a droplet when it interacts with a VOC. (b) Illustration of configuration changes when the VOC interacts with a droplet. Adapted from (Hussain et al., 2017).

tion 4, parameter a corresponds to that maximum value, b is the steepness of the curve and c indicates when in time the curve reaches its mid point.

$$f(x) = \theta_1 \cdot \frac{1}{1 + e^{\theta_2 \cdot (\theta_3 - x)}} \cdot \left(1 - \frac{1}{1 + e^{\theta_4 \cdot (\theta_5 - x)}} \right) \quad (3)$$

$$f(x) = \frac{a}{1 + e^{b \cdot (c - x)}} \quad (4)$$

The wrapper method Recursive Feature Elimination with Cross Validation (RFECV) was used for feature selection. This algorithm recursively eliminates features and in each iteration trains a classifier to ascertain the effect of the removed feature. It stops

when a good classification score is obtained. It then ranks feature performance as 1 if the remaining subset of features gave a good classification result. A. Jovic *et al.* stated that linear Support Vector Machines were feasible to use with wrapper methods (Jovic et al., 2015). Therefore, it was the learning model used. Stratified 10-fold cross-validation was used to obtain a classification score. In which the dataset is randomly divided in ten subsets, avoiding unbalanced set by maintaining class percentage, and using one tenth of the set as validation set and the rest as training set. Then the process is repeated ten times and the result is the average of the accuracies obtained.

Accuracy, given by Equation 5, was chosen as the scoring method, since it provides a measure of global performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

It gives the number of correctly classified examples: the sum of True Positives (TP) and True Negatives (TN) divided by the total number of examples (TP, TN, False Positives (FP) and False Negatives (FN)).

Finally, a Support Vector Machine classifier was chosen as the classification model. The standard SVM, illustrated in Figure 9, is a binary classifier. It finds a hyperplane that best separates the two classes and creates a decision boundary. The examples of the classes that are closer to the hyperplane are called the support vectors. The hyperplane can be described by the set of points that satisfy $\mathbf{w}^T \mathbf{x} = 0$, \mathbf{w} being the vector perpendicular to the hyperplane. The points

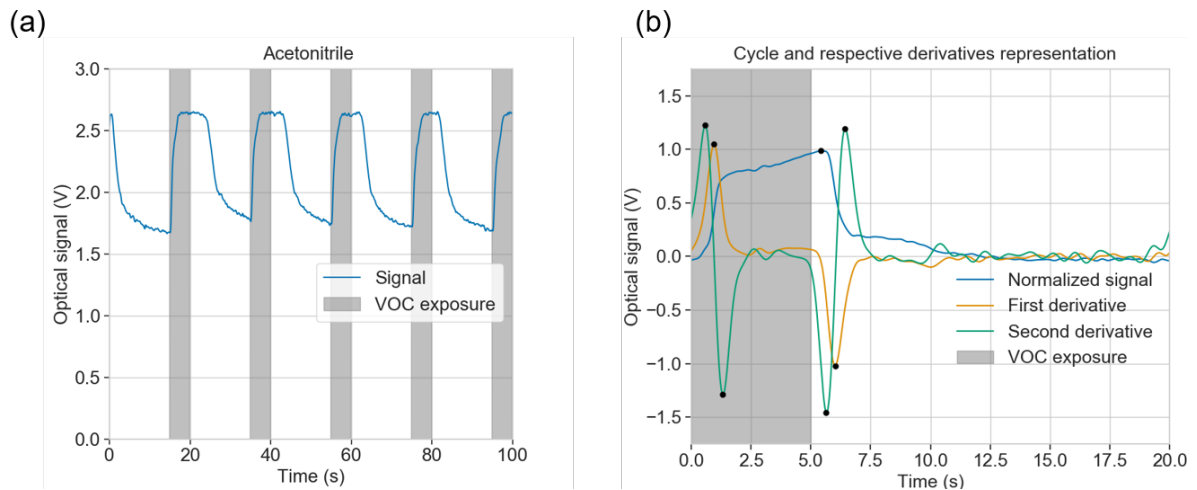


Figure 8: (a) Representation of some cycles with their exposure and recovery times. (b) Example of an individual cycle and its first and second derivatives. Black dots mark the peaks of interest and the darkened part of the plot correspond to the exposure time.

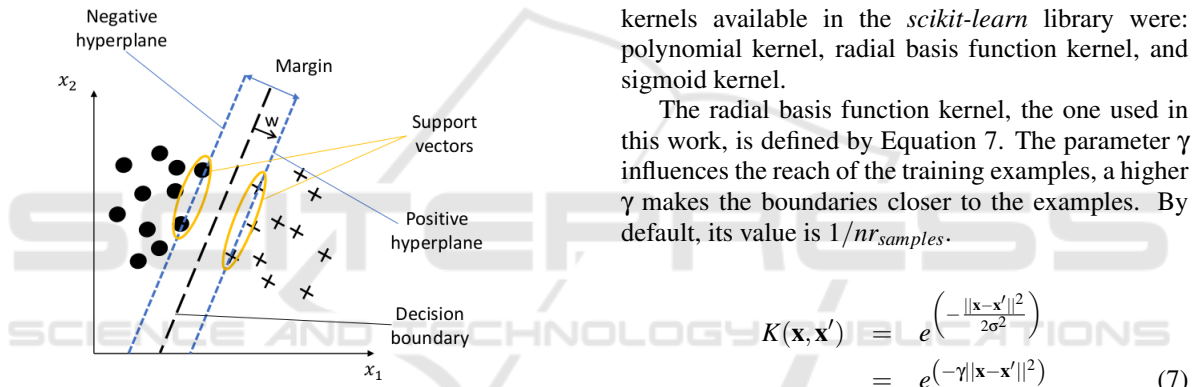


Figure 9: Schematic representation of Support Vector Machines. Adapted from (Sebastian, 2015).

that satisfy $w_o + \mathbf{w}^T \mathbf{x}_{pos} = 1$ form the *positive* hyperplane, and the ones that satisfy $w_o + \mathbf{w}^T \mathbf{x}_{neg} = -1$ form the *negative* hyperplane. By subtracting those expressions and normalizing by the length of the vector \mathbf{w} , Equation 6 is obtained.

$$\frac{\mathbf{w}^T (\mathbf{x}_{pos} - \mathbf{x}_{neg})}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (6)$$

The left term of Equation 6 is the distance between hyperplanes, the objective is to maximize it.

The performance of the SVM classifier depends on the kernel, the kernel parameter γ (if kernel is non-linear) and the parameter C .

The variable C influences the number of support vectors used by the model, a larger C means a higher cost for misclassification (Sebastian, 2015).

The kernel is used to transform the original features and project them onto a higher dimensional space via a mapping function (Sebastian, 2015). The

kernels available in the *scikit-learn* library were: polynomial kernel, radial basis function kernel, and sigmoid kernel.

The radial basis function kernel, the one used in this work, is defined by Equation 7. The parameter γ influences the reach of the training examples, a higher γ makes the boundaries closer to the examples. By default, its value is $1/nr_{samples}$.

$$K(\mathbf{x}, \mathbf{x}') = e^{\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)} = e^{(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)} \quad (7)$$

Parameters C and γ should be tuned together. For that, is possible to use *GridSearchCV* from *scikit-learn* library that performs a grid search with cross validation to choose the best parameters. This function was used with: values 0.1, 1, 10 and 100 for C ; values 0.01, 0.8, 1 and 10 for γ ; and the kernels available, already mentioned. Other parameters from the *SVC* *scikit-learn* library used the default values given.

4 RESULTS

4.1 Features of E-nose V3

E-nose V3 is an optimized version of E-nose V0, the first prototype. The system was improved, having a faster and cheaper data acquisition system composed by an Arduino microcontroller and a Raspberry Pi (under 100 euros) as opposed to the previous system composed by a computer and a NI myDAQ (above 600 euros). Now it possess a single interface making

Table 1: Comparative table of different e-nose versions.

E-nose Version		First version (Pádua et al., 2018b)	Latest version (present work)
Transduction system	Sampling rate (Hz)	10	90
	Data acquisition elements	NI myDAQ + Computer	Arduino Due + Raspberry Pi3
Detection chamber	Detection sensors	Photoresistors	Photodiodes
	Array capacity (nr sensors)	4	6
	Stability and scaling (scale 1-5)*	1	4
	Chamber volume	2L	28mL
Delivery system	Analysis frequency	0.37 cycles/min	3 cycles/min

* Stability and scaling scale:

- 1: Open and close the chamber was not practical. Chamber with flow losses.
- 2: Robust and more easy-to-use chamber. Difficult to change sensing films between experiments. Flow losses verified.
- 3: Easy to change sensing films. Hermetic. PCBs with through-hole technology. components.
- 4: Easy to change sensing films. Hermetic.. PCBs with surface-mount technology components. Has slots to insert the PCBs and support for sensing films. Has light guiders for LEDs.

it more user-friendly (Pádua et al., 2018a). Regarding the detection chamber, the number of sensors was increased and the chamber became more robust, stable and hermetic. Moreover, improvements in the delivery system allowed a 10 times faster analysis, comparing to the initial prototype.

Regarding the delivery system, pumps flow rate is 3 L/min, the flow rate at chamber entrance is 1.8 L/min for the exposure circuit, and 2.5 L/min for the recovery circuit. Table 1 summarizes the main differences between E-nose versions.

Considering the implementation of a new detector system, the results for the validation of the photodiodes are represented in Figure 10. It demonstrates that the photodiodes are suitable for the application and that no signal is lost at any given light intensity.

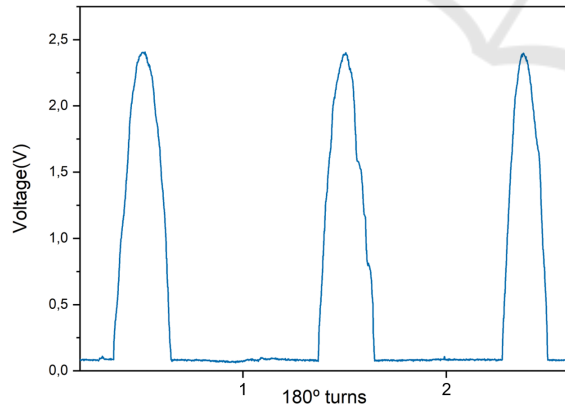


Figure 10: Proof of concept of the photodiodes sensitivity. The experiment consisted in the signal acquisition during three 180 degree turns. The absence of a saturation line at the maximum values approves the detection system.

Signals were acquired (Figure 11) to test the new version of the E-nose. A preliminary experiment was made to measure signal to noise ratio. Five sensing films were sequentially exposed to two different VOCs (acetone and toluene), the remaining position

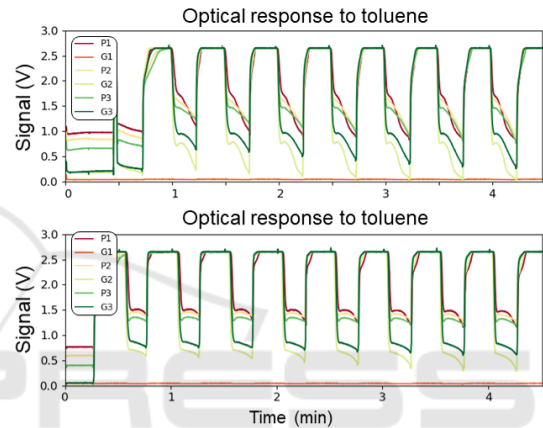


Figure 11: Signals acquired to test the new E-nose version. Position *G1* corresponds to the control measurement.

of the array was used as a control measurement.

The signal to noise ratio (SNR) was calculated by Equation 8. The result was approximately 50 dB.

$$\begin{aligned}
 SNR_{dB} &= 10 \log \left(\frac{P_{signal}}{P_{noise}} \right) \\
 &= 20 \log \left(\left(\frac{A_{signal}}{A_{noise}} \right)^2 \right) \quad (8)
 \end{aligned}$$

4.2 Classification Results

Three sensing gels, with the exact same composition, were exposed to a sequence of 11 VOCs (hexane, heptane, toluene, chloroform, dichloromethane, acetonitrile, acetone, diethyl ether, ethyl acetate, ethanol and methanol) in cycles of 5 seconds of exposure and 15 seconds of recovery. 15 mL of VOC was placed in the sample chamber and heated at 37°C during 15 minutes before beginning the experiment to ensure headspace saturation. Each sample interacted with the sensors in a total of 15 minutes. The remaining positions of the electronic nose were used as control

measurements. After the sequence was completed, sensing gels were stored. Two months later, the same procedure was repeated with the same sensing gels. The first cycles of each volatile (the first 5 for the first time the gels were exposed and the first 6 for the second time) were not considered due to the presence of noise, caused by the opening of the valve that allows VOC intake. This resulted in a total of approximately 2600 cycles for analysis (236 cycles per VOC).

The features were extracted, however, the curve fitting algorithm was not able to converge into a solution for the parameters of the approach of the article (Equation 3). In order to fit a model to the signal, it was necessary to simplify the approach. The solution was to divide the signal in two parts (from cycle beginning until its maximum and from the maximum until its end) and apply the logistic function (Equation 4) to each part of the signal (Figure 12).

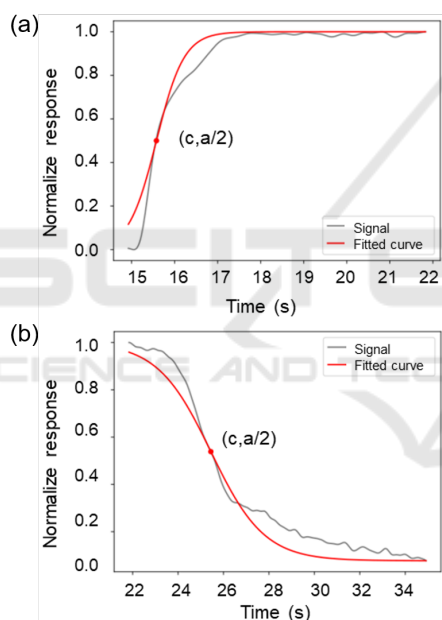


Figure 12: Estimated curves for (a), the first part of the cycle and for (b), the second part. The red point marks the coordinates where the signal bisects with the value of half the rise/descend.

After obtaining a set of features, RFECV algorithm used a quarter of the data and the rest remained untouched for validation purposes. The rank of the features is shown in Table 2. Features ranked as 1 are the ones with best performance, and therefore were the ones selected.

When tuning the classifier parameters, the best performance was obtained with the radial basis function kernel and parameters C: 100 and γ : 0.1.

Finally, to validate the classifier with unseen data, the validation set was used. The classifier managed

to obtain a global accuracy of 94.6% with a standard deviation of 0.9%. Figure 13 shows the correspondent normalized confusion matrix.

		Predicted										
		Acetone	Acetonitrile	Chloroform	Dichloromethane	Diethylether	Ethanol	Ethyl acetate	Heptane	Hexane	Methanol	Toluene
Actual	Acetone	0.98	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Acetonitrile	0.02	0.97	0.00	0.00	0.1	0.00	0.00	0.00	0.00	0.00	0.00
	Chloroform	0.01	0.00	0.94	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00
	Dichloromethane	0.00	0.00	0.00	0.95	0.02	0.00	0.01	0.00	0.01	0.00	0.00
	Diethylether	0.00	0.00	0.00	0.05	0.94	0.00	0.01	0.00	0.00	0.00	0.00
	Ethanol	0.00	0.02	0.00	0.00	0.00	0.90	0.00	0.00	0.00	0.08	0.00
	Ethyl acetate	0.01	0.00	0.03	0.02	0.00	0.00	0.92	0.00	0.00	0.00	0.00
	Heptane	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.02	0.00	0.00
	Hexane	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.02	0.95	0.00	0.00
	Methanol	0.00	0.01	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.91	0.00
	Toluene	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.96

Figure 13: Confusion matrix obtained for the classification of the 11 VOCs. Accuracy of 94.6% ($\pm 0.9\%$).

5 CONCLUSION AND FUTURE WORK

E-nose V3 is a tailor-made device using as gas sensors an innovative class of biomaterials which optical properties vary in the presence of VOCs (Husain et al., 2017). This new optimized device, together with the appropriate use of signal processing tools, allowed the accurate classification of 11 VOCs from different chemical classes using sensing films with the exact same composition. The results show that our sensor possess advantages over other types of sensors such as MOS-sensors or conducting polymer sensors, as these normally have limited selectivity towards VOC classes. In addition, our materials and sensing device operate at ambient conditions as opposed to MOS (300°C)(Wilson, 2012). In addition, used sensing materials are biodegradable and reusable with the potential to recycle its individual components, having a lower carbon footprint.

The detection chamber is hand-size and hermetic for the pressures being applied. Its low volume allows the use of short cycle periods, and consequently a fast analysis. The advantages of using photodiodes is that they are more sensible and have a quicker response time than photoresistors.

The transduction system based on Arduino and Raspberry Pi constitute a low-cost and independent

Table 2: Ranking of the selected features by RFECV with 10-fold cross validation.

Number	Description	10-fold validation score
1	Time to reach the maximum value of the cycle	7
2	Time to reach the maximum of the first derivative of the signal	4
3	Time to reach the minimum of the first derivative of the signal	1
4	Maximum value of the first derivative of the signal	3
5	Minimum value of the first derivative of the signal	2
6	Time to reach the maximum of the first half of the second derivative of the signal	1
7	Time to reach the minimum of the first half of the second derivative of the signal	10
8	Maximum value of the first half of the second derivative of the signal	1
9	Minimum value of the first half of the second derivative of the signal	1
10	Time to reach the maximum of the second half of the second derivative of the signal	11
11	Time to reach the minimum of the second half of the second derivative of the signal	9
12	Maximum value of the second half of the second derivative of the signal	1
13	Minimum value of the second half of the second derivative of the signal	1
14	Area under the signal	1
15	Skewness	1
16	Kurtosis	1
17	Parameter a from the first half of the signal	13
18	Parameter b from the first half of the signal	1
19	Parameter c from the first half of the signal	1
20	Error associated with the fit from the first half of the signal	5
21	Parameter a from the second half of the signal	12
22	Parameter b from the second half of the signal	1
23	Parameter c from the second half of the signal	6
24	Error associated with the fit from the second half of the signal	8

unit. The sampling rate of 90 Hz was enough to guarantee informative signal acquisition.

After data collection, two different feature extraction methods were used: features extraction from the original response curves and curve fitting parameters. The most informative features extracted from the signals were identified and reported in bold in Table 2. Applying those features, the accuracy rate for SVM was $94.6\% \pm 0.9\%$. Future work should include comparison of these results with other machine learning classification techniques.

The assembled system was able to distinguish the 11 tested VOCs. Although the identification of the VOCs was not made in real time, a step towards making this E-nose a device that is able to make automatic classification was taken.

The E-nose V3 device is robust and user-friendly. We are aware that there is still room for improvement. For example, in the future the detection system could benefit from the use of variable resistors, or multiple resistors per photodiode, enabling auto scaling or simply by easing the actual scaling where multiple PCBs assembled with different resistor values, are chosen to best suit the sensing probes in use based on the output voltage. Also, a longer array of LED/photodiodes pairs would allow for a wider combination of different sensing probes, or a larger number of replicas, which in both cases have a positive impact in the VOCs prediction algorithm.

Using the optimized e-nose version, the necessary

studies to evaluate more complex scenarios are possible, such as testing concentration limits, the effects of room conditions in the material and classification of complex samples where many different VOCs are mixed together.

ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the EU Horizon 2020 research and innovation programme (grant agreement No. **SCENT-ERC-2014-STG-639123**). This work was supported by the **Applied Molecular Biosciences Unit-UCIBIO** which is financed by national funds from FCT/MCTES (**UID/Multi/04378/2013**) and co-financed by ERDF under the PT2020 Partnership Agreement (**POCI-01-0145-FEDER-007728**). The authors thank FCT/MEC for the research fellowship PD/BD/105752/2014 for A.P.

REFERENCES

- Arshak, K., Moore, E., Lyons, G., Harris, J., and Clifford, S. (2004). A review of gas sensors employed in electronic nose applications. *Sensor Review*, 24(2):181–198.
- Barbosa, A. J., Oliveira, A. R., and Roque, A. C. A. (2018).

- Protein- and peptide-based biosensors in artificial olfaction. *Trends in Biotechnology*, 36:1244–1258.
- Dutta, R. and Dutta, R. (2006). Intelligent Bayes Classifier (IBC) for ENT infection classification in hospital environment. *BioMedical Engineering Online*, 5.
- Ghasemi-Varnamkhasi, M., Mohtasebi, S. S., Siadat, M., Ahmadi, H., and Razavi, S. H. (2015). Research paper: From simple classification methods to machine learning for the binary discrimination of beers using electronic nose data. *Engineering in Agriculture, Environment and Food*, 8:44 – 51.
- He, Q., Yan, J., Shen, Y., Bi, Y., Ye, G., Tian, F., and Wang, Z. (2012). Classification of Electronic Nose Data in Wound Infection Detection Based on PSO-SVM Combined with Wavelet Transform. *Intelligent Automation and Soft Computing*, 18(7):967–979.
- Holmberg, M., Gustafsson, F., Hörnsten, E. G., Winquist, F., Nilsson, L. E., Ljung, L., and Lundström, I. (1998). Bacteria classification based on feature extraction from sensor data. *Biotechnology Techniques*, 12(4):319–324.
- Hussain, A., Semeano, A. T. S., Palma, S. I. C. J., Pina, A. S., Almeida, J., Medrado, B. F., Pádua, A. C. S., Carvalho, A. L., Dionísio, M., Li, R. W. C., Gamboa, H., Ulijn, R. V., Gruber, J., and Roque, A. C. A. (2017). Tunable gas sensing gels by cooperative assembly. *Advanced Functional Materials*, 1700803:1–9. <https://doi.org/10.1002/afm.201700803>.
- Jovic, A., Brkic, K., and Bogunovic, N. (2015). A review of feature selection methods with applications. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1200–1205.
- Ordukaya, E. and Karlik, B. (2017). Quality control of olive oils using machine learning and electronic nose. *Journal of Food Quality*, Vol 2017 (2017).
- Pádua, A. C., Osório, D., Rodrigues, J., Santos, G., Porteira, A., Palma, S., Roque, A., and Gamboa, H. (2018a). Scalable and easy-to-use system architecture for electronic noses. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 179–186.
- Pádua, A. C., Palma, S., Gruber, J., Gamboa, H., and Roque, A. C. A. (2018b). Design and evolution of an opto-electronic device for voes detection. *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 48–55.
- Röck, F., Barsan, N., and Weimar, U. (2008). Electronic nose: Current status and future trends. *Chemical Reviews*, 108(2):705 – 725.
- Scott, S. M., James, D., and Ali, Z. (2007). Data analysis for electronic nose systems. *Microchimica Acta*, 156(3/4):183 – 207.
- Sebastian, R. (2015). *Python Machine Learning*. Packt Publishing Ltd.
- Uçar, A. and Özalp, R. (2017). Efficient android electronic nose design for recognition and perception of fruit odors using kernel extreme learning machines. *Chemometrics and Intelligent Laboratory Systems*, 166:69 – 80.
- Wasilewski, T., Gębicki, J., and Kamysz, W. (2017). Bioelectronic nose: Current status and perspectives. *Biosensors and Bioelectronics*, 87:480–494.
- Wilson, A. D. (2012). Review of electronic-nose technologies and algorithms to detect hazardous chemicals in the environment. *Procedia Technology*, 1(First World Conference on Innovation and Computer Sciences (INSODE 2011)):453 – 463.
- Xu, L., He, J., Duan, S., Wu, X., and Wang, Q. (2016). Comparison of machine learning algorithms for concentration detection and prediction of formaldehyde based on electronic nose. *Sensor Review*, 36(2):207.
- Yan, J., Xiuzhen, G., Shukai, D., Pengfei, J., Lidan, W., Chao, P., and Songlin, Z. (2015). Electronic nose feature extraction methods: A review. *Sensors*, Vol 15, Iss 11, Pp 27804-27831 (2015), (11):27804.