



HAL
open science

An electronic nose using conductometric gas sensors based on P3HT doped with triflates for gas detection using computational techniques (PCA, LDA, and kNN)

Aicha Boujnah, Aimen Boubaker, Sébastien Pecqueur, Kamal Lmimouni,
Adel Kalboussi

► To cite this version:

Aicha Boujnah, Aimen Boubaker, Sébastien Pecqueur, Kamal Lmimouni, Adel Kalboussi. An electronic nose using conductometric gas sensors based on P3HT doped with triflates for gas detection using computational techniques (PCA, LDA, and kNN). *Journal of Materials Science: Materials in Electronics*, 2022, 33, pp.27132-27146. 10.1007/s10854-022-09376-2 . hal-03851261

HAL Id: hal-03851261

<https://hal.science/hal-03851261v1>

Submitted on 14 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 An electronic nose using conductometric gas sensors based on 2 P3HT doped with triflates for gas detection using computational 3 techniques (PCA, LDA and kNN)

4 Aicha Boujnah¹, Aimen Boubaker¹, Sébastien Pecqueur², Kamal Lmimouni² & Adel Kalboussi¹

5 ¹Department of Physics, Faculty of sciences, University of Monastir, Monastir 5050, Tunisia.

6 ²Univ. Lille, CNRS, Centrale Lille, University Polytechnique Hauts-de-France, UMR 8520 - IEMN, F-
7 59000 Lille, France.

8 *Email:* aicha.boujnah.1992@gmail.com

9 **ABSTRACT**

10

11 This study presents the development of an electronic nose comprising eight homemade sensors with pure P3HT
12 and doped with different materials. The objective is to electronically identify the gases exposed on these sensors and
13 evaluate the accuracy of target-gas classification. The resistance variation for each sensor is measured over time and
14 the collected data were processed by three different identification techniques as following; principal component
15 analysis (PCA), linear discriminate analysis (LDA) and nearest neighbor analysis (kNN). The merit factor for the
16 analysis is the relative modulation of the resistance is very important and computationally gives different results. In
17 addition, the fact that we have sensors made with innovative materials where the reproducibility of the response for
18 the same material can be a constraint in the recognition. In contrast, we have shown that despite the lack of
19 reproducibility for the same material on two different sensors, and despite the instability during the ten last seconds,
20 we have good recognition rates and we can even say which algorithm is better. It is noted that the LDA is the most
21 reliable and efficient method for gas classification with a prediction accuracy equal to 100% whereas it reach 93.52%
22 and 73.14 % for PCA and kNN, respectively for other techniques for 40% of training dataset and 60% of testing
23 dataset.

24

25 **Keywords** Electronic nose, Conductometric gas sensor, Pattern recognition, Confusion matrix, LDA, PCA, kNN.

26

27 **1 Introduction**

28 Over the past decades, researchers have studied and developed electronic nose (e-nose) technology [1–3] that
29 has been widely used in many fields such as medicine [4], industry, military, and environment [5]. By analogy with
30 the human olfactory system, an e-nose is an instrument that combines different elements sensitive to gases on an
31 input array [3]. Information is then projected to a classifier that analyses interactions of gases with sensitive materials
32 by pattern recognition, so information is classified to identify the chemical identity of molecularly elementary or
33 constitutionally complex odors [2,6,7] . The performance of gas sensors can be evaluated by dint of several criteria

34 such as high sensitivity, selectivity/specificity, fast-response/recovery-time, robustness, low cost and simple
35 fabrication, good repeatability and long-term stability [8]. Based on the sensing materials and working principles,
36 sensing transduction can be classified into several types, such as electrochemical, resistive, capacitive, optical and
37 acoustical [9,10].

38 On conductometric gas sensors, resistance changes for the sensitive material when it interacts with the molecules
39 [11], caused by either-or-both change in charge carrier density and mobility [4]. Resistance in conductometric gas
40 sensors depends on changes in the partial pressure for different gases exposed, but also on the geometry of the pair
41 electrodes. Therefore, interdigitated electrodes structures with a sensing layer are widely used for several types of
42 sensors due to large width-to-length ratio, maximizing the sensitivity of the material's resistance upon small
43 environmental changes [12].

44 Conductive polymers have been widely studied in various applications such as in electronics [13,14],
45 optoelectronics[15,16], electromechanical devices on the one hand, and in chemical gas sensors on the other hand
46 [17]. It is widely recognized that researchers have used conductive polymers as a sensing layer for chemoresistors
47 gas sensors [8] because these materials have good mechanical properties such as robustness and flexibility [18]. In
48 addition, they not only offer high gas sensitivity; short response/recovery times but they also require low power
49 consumption and low design cost [19]. Several conducting polymers were used as a sensing layer, such as
50 poly(pyrrole), poly(aniline) and poly(3,4-ethylene dioxythiophene) [8,18]. Among polymers, the poly (3-
51 hexylthiophene) (P3HT) was studied in the previous work because it has high hole mobility, good stability under
52 ambient conditions [20,21]. In 1990, A. Assadi and coworkers [22] proposed a P3HT based field-effect transistor
53 and found that the interactions between alumina gas and P3HT results in a change in the electrical properties of
54 transistor. In addition, Bertoni C et al. opened the way to the use of P3HT nanofibers based sensing elements for the
55 realization of portable, real-time electronic noses for exhaled breath analysis [23]. Kuo et al. showed good
56 performances, such that the relative resistance modulation (dR/R) of the P3HT thin-film for different concentrations
57 of ammonia gas NH_3 was around 6% [21]. Doping materials such as P3HT is essential to improve the electrical
58 performances that in return enhances both charge-carrier injection and transport in semiconductor devices to
59 decrease its power consumption. In addition, doping can control the surface morphologies and physical/chemical
60 properties of the conductive polymer [24], affecting the surface energy and crystallinity [25]. Different dopants have
61 different doping yields on conducting polymers, but also different effects on the selectivity and sensitivity of target
62 gases. Joseph N. Barisci et al. [26] use a set of sensors based on doped-polypyrrole polymers for the detection of
63 BTEX compounds (benzene, toluene, ethylbenzene and xylene). In this study, they used different dopants to analyze
64 the target gases with sufficient sensitivity and reproducibility.

65 Cumulative effects of many gas sensing materials was conceptualized by the term electronic nose, first used in
66 the late 1980s by Gardner and coworkers, although the concept had been formalized somewhat earlier by Persaud
67 and Dodd [27]. Pattern recognition and classification techniques are the building blocks of the e-nose to identify
68 different types of gases exposed on the gas sensor [28,29]. Hence, several methods have been investigated for e-
69 nose analysis, such as principal component analysis (PCA) [30], support vector machines (SVM) [31,32], artificial
70 neural networks (ANN) [33,34] or k-nearest neighbor (kNN) [35].

71 In the present study, we report the results of an electronic nose based on conductometric gas sensors, where
72 P3HT presents the sensitive layer deposited onto interdigitated gold electrodes with seven different dopants, to
73 recognize and identify three different gases as following; are water, acetone and ethanol. The adsorption and

74 desorption datasets obtained from each sensor were used for the identification and prediction of those odors using
 75 three different machine-learning methods, i.e., PCA, LDA, and kNN and we discuss the results.

76 2 Data Acquisition and sensor response

77 In order to study the effect of the gas on the response of the sensor, all measurements of electrical device
 78 characterization were performed on an Agilent 4155 parameter analyzer in air. This air passes through vials
 79 containing the different solvents in the liquid phase, and the flow pushes the gas phase.

80 **Figure 1** describes the E-nose setup used for measuring the resistance of the gas sensor. Three valves that contain
 81 a check valve to prevent solvent contamination of the glass vials containing solvent manually controlled exposure
 82 to the various gases. Each solvent-containing vial (Water, Acetone and Ethanol) is manually exchanged in the setup
 83 during the 3-minute purge (bypass valve ON, analyte vial IN/OUT OFF), in order to ensure a continuous and steady
 84 flow of pressurized gas over the entire output current record for each device measured with the different P3HT
 85 materials. The exposed gas flow rate is always set to 1mL/s. During the experiment, no solvent condensation was
 86 observed either in the Teflon tube, at the gas capillary exhaust, or on the substrate of the sensitive material [36].

87 The saturating vapor concentration (in ppm) is the concentration of vapor in equilibrium with the pure substance
 88 at a given temperature (20°C) and atmospheric pressure. It is obtained directly from the saturation vapor pressure
 89 P_{sat} (in Pa) which is the partial pressure under the same conditions. According to the perfect gas law, we have:

$$90 P \times V = n \times R \times T \quad (1)$$

91 Where P is the vapor pressure (in Pa), V is the volume of the gas (in m^3), n is the quantity of gas (in mol), R is the
 92 universal constant of perfect gases ($8.314472 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$) and T is the temperature ($273.15 + X^\circ\text{C}$ in K).

93 The pressure P is proportional to the quantity of gas n . Therefore, we can find the gas concentration in ppm by
 94 this ratio:

$$95 C = \frac{n_{\text{sat}}(\text{gas})}{n(\text{total})} \times 10^6 = \frac{P_{\text{sat}}(\text{gaz})}{P_{\text{atm}}} \times 10^6 \quad (2)$$

96 With:

97 At $T=20^\circ\text{C}$, $P_{\text{sat}}(\text{Water}) = 2300 \text{ Pa}$, $P_{\text{sat}}(\text{Acetone}) = 22800 \text{ Pa}$ and $P_{\text{sat}}(\text{Ethanol}) = 5800 \text{ Pa}$.

98 $P_{\text{atm}} = 1013.25 \text{ hPa} = 101325 \text{ Pa}$.

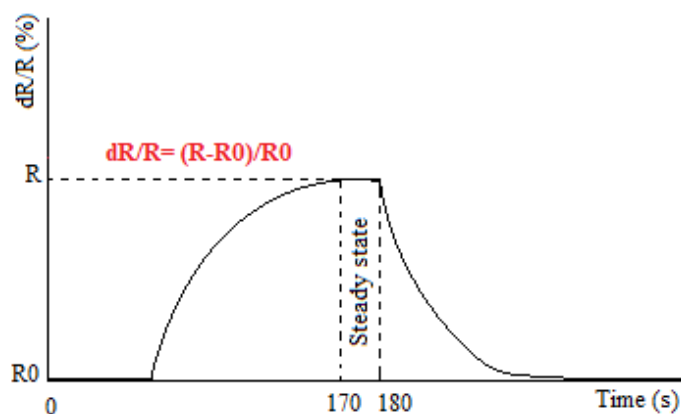
99 This is only an approximation because we are not necessarily in thermodynamic equilibrium; it would have to
 100 be verified experimentally to do metrology. However it not relevant within the framework of this study. In addition,
 101 we did not study the effect of gas concentrations on the response of the sensors and no metrology was carried out.



102
 103 **Fig. 1** E-nose setup used for measuring the resistance of gas sensor

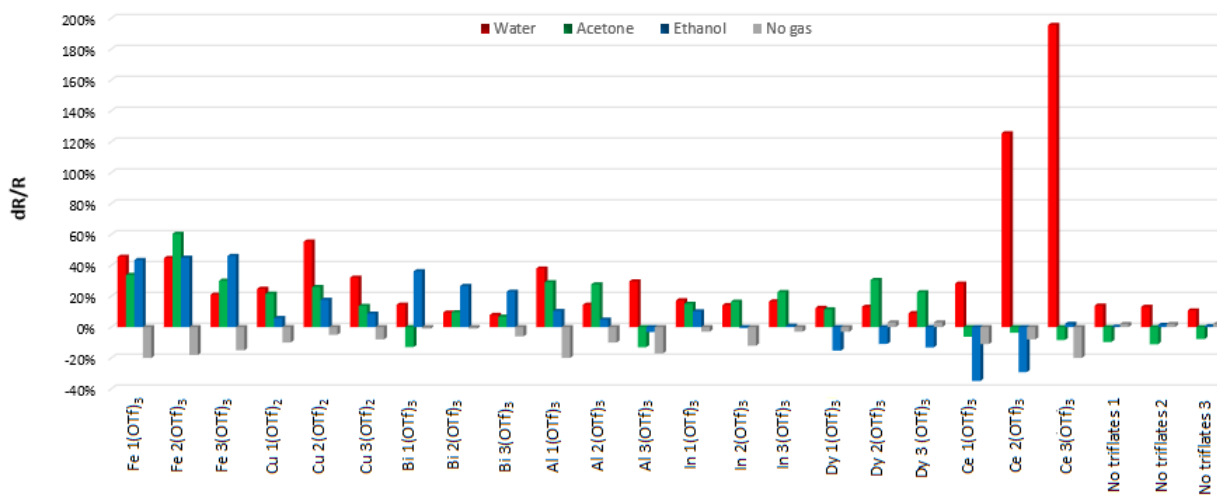
104 We collected all the data from the resistance versus time reading of these eight different sensors based on eight
 105 different materials. The parameter extraction step is important to reduce the dimension of the input parameters of
 106 the classification algorithm. The principle is to reduce the amount of information while keeping most of their
 107 variance contained in the sensor matrix. Therefore, after collecting all the measurements, we used the characteristic

108 parameters of the each sensor response to improve the classification. These parameters are the relative variation of
 109 the resistance dR/R , which defines the relative difference between the device resistance after 3 minutes of gas
 110 exposure and the resistance before opening the gas vial that can give information between the sensor and the odors
 111 (**Figure 2**). In addition, to filter the amount of data to be analyzed, we chose to use the data set that contains the
 112 dR/R at the last ten seconds of exposure. We chose to use the “steady state” dataset, which shows the best separation
 113 of data by nature of the gas [36].



114
115 **Fig. 2** Diagram of resistance modulation dR/R

116 **Figure 3** shows the mean value of the relative resistance variation (in %) of the different 24 devices (three tests
 117 were evaluated for each material) recorded over the last ten seconds per exposure to the three different stimulus
 118 water / acetone / ethanol. We observed a variability between the resistance modulations for all the devices doped
 119 with the same materials. This is due to the droplet deposition technique. However, we notice the reproducibility of
 120 the resistance modulation in particular in the cases of $Fe(OTf)_3$, $Cu(OTf)_2$, $In(OTf)_3$ and P3HT without dopant [36].



121
122 **Fig. 3** Representation of the mean value of the relative resistance modulation (in %) of the 24 sensors to the three different
 123 stimulus Water / Acetone / Ethanol

124 The exposure of these three different gases is done in a sequential way with all the possible combinations of
 125 gases in order to study the sensitivity and selectivity of sensors. **Figure 4** shows the sensor matrix that was used in
 126 this study. It presents the six different sequences of gas exposure. Each time, three gases are exposed: 1 is Water, 2
 127 is Acetone and 3 is Ethanol. Each 3-min gas exposure is initiated by a 3-min air purge. The measurement sequences
 128 that have been carried out for each device during two hours are:

129 \sum 1/ 3/ 2/ 1/ 2/ 3/ 2/ 1/ 3/ 2/ 3/ 1/ 3/ 1/ 2/ 3/ 2/ 1/ 3

130 With:

131 \sum : without compressed air flow and without probed gas

132 1 : with air purge

133 1: with compressed air flow and the probed gas is water

134 2: with compressed air flow and the probed gas is acetone

135 3: with compressed air flow and the probed gas is ethanol

136 For each gas, we took the 10 values of the relative resistance modulation dR/R during the last ten seconds (steady
137 state) as different rows in the matrix (180 rows). In addition, the 24 columns represent the number of extracted
138 features we can obtain at a specific time for the gas exposure sequences; dR/R for eight sensors with three tests per
139 sensor.

Possible sequences of gas exposure	Classes	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3
		Sensor 1	Sensor 1	Sensor 1	Sensor 2	Sensor 4	Sensor 8
Sequence 1	1 10 last values of steady state	$\left(\frac{dR}{R}\right)_{1,1}^1$	$\left(\frac{dR}{R}\right)_{1,2}^1$	$\left(\frac{dR}{R}\right)_{1,3}^1$	$\left(\frac{dR}{R}\right)_{2,1}^1$	$\left(\frac{dR}{R}\right)_{4,2}^1$	$\left(\frac{dR}{R}\right)_{8,3}^1$
		$\left(\frac{dR}{R}\right)_{1,1}^1$	$\left(\frac{dR}{R}\right)_{1,2}^1$	$\left(\frac{dR}{R}\right)_{1,3}^1$	$\left(\frac{dR}{R}\right)_{2,1}^1$	$\left(\frac{dR}{R}\right)_{4,2}^1$	$\left(\frac{dR}{R}\right)_{8,3}^1$
	3 10 last values of steady state	$\left(\frac{dR}{R}\right)_{1,1}^3$	$\left(\frac{dR}{R}\right)_{1,2}^3$	$\left(\frac{dR}{R}\right)_{1,3}^3$	$\left(\frac{dR}{R}\right)_{2,1}^3$	$\left(\frac{dR}{R}\right)_{4,2}^3$	$\left(\frac{dR}{R}\right)_{8,3}^3$
Sequence 2	2 10 last values of steady state	$\left(\frac{dR}{R}\right)_{1,1}^2$	$\left(\frac{dR}{R}\right)_{1,2}^2$	$\left(\frac{dR}{R}\right)_{1,3}^2$	$\left(\frac{dR}{R}\right)_{2,1}^2$	$\left(\frac{dR}{R}\right)_{4,2}^2$	$\left(\frac{dR}{R}\right)_{8,3}^2$
		$\left(\frac{dR}{R}\right)_{1,1}^2$	$\left(\frac{dR}{R}\right)_{1,2}^2$	$\left(\frac{dR}{R}\right)_{1,3}^2$	$\left(\frac{dR}{R}\right)_{2,1}^2$	$\left(\frac{dR}{R}\right)_{4,2}^2$	$\left(\frac{dR}{R}\right)_{8,3}^2$
	3 10 last values of steady state	$\left(\frac{dR}{R}\right)_{1,1}^3$	$\left(\frac{dR}{R}\right)_{1,2}^3$	$\left(\frac{dR}{R}\right)_{1,3}^3$	$\left(\frac{dR}{R}\right)_{2,1}^3$	$\left(\frac{dR}{R}\right)_{4,2}^3$	$\left(\frac{dR}{R}\right)_{8,3}^3$
Sequence 6	2 10 last values of steady state	$\left(\frac{dR}{R}\right)_{1,1}^1$	$\left(\frac{dR}{R}\right)_{1,2}^1$	$\left(\frac{dR}{R}\right)_{1,3}^1$	$\left(\frac{dR}{R}\right)_{2,1}^1$	$\left(\frac{dR}{R}\right)_{4,2}^1$	$\left(\frac{dR}{R}\right)_{8,3}^1$
		$\left(\frac{dR}{R}\right)_{1,1}^1$	$\left(\frac{dR}{R}\right)_{1,2}^1$	$\left(\frac{dR}{R}\right)_{1,3}^1$	$\left(\frac{dR}{R}\right)_{2,1}^1$	$\left(\frac{dR}{R}\right)_{4,2}^1$	$\left(\frac{dR}{R}\right)_{8,3}^1$
	3

Fig. 4 Sensor matrix with characteristic parameters

With:

$\left(\frac{dR}{R}\right)_{i,j}^c$: Represents the response of relative resistance modulation in the steady state for each sensor number i for test

j and for class c .

i : 1, 2, 3, 4, 5, 6, 7 and 8, is the number of the sensor to be tested.

146 j : 1, 2 and 3, is the number of measure tests for each sensor to be tested.

147 c : 1, 2 and 3, is the number of the class.

148 **3 Algorithms and pattern recognition**

149 This phase covers mainly the classification and representation of both pattern and results regarding their
150 importance as part of any system [37]. The main objective of a classification algorithm is to find the rules that model
151 the behavior of the sensor and thus discriminate between compounds or mixture of compounds. There are two types
152 of classification algorithms: supervised or unsupervised. Several methods can be found in the literature each with
153 its own specificities and variants [38–40]. It is recommended to test several classification algorithms in order to find
154 which algorithm will give the best classification performance. In this study case, three different classification
155 algorithms were tested: PCA, LDA, and k-NN. The Scikit-Learn library, developed in Python language was used to
156 test the different methods of recognition [41]. We used 180 individual observations. We chose randomly 40 % of
157 the dataset that was used for the training set (72 Observations) and the rest for the identification and the testing (108
158 Observations).

159 **Table 1** Algorithm accuracy of several random datasets

	PCA Accuracy (%) (X test=60%, X train=40%)	LDA Accuracy (%) (X test=60%, X train=40%)	KNN Accuracy (%) (X test=60%, X train=40%)
Random data set 1	86.11	100	63.6
Random data set 2	89.81	100	26.8
Random data set 3	93.52	100	73.1
Random data set 4	89.81	100	64.8

160
161 We took several data partitions and each time, we applied the three different algorithms to compare the calculated
162 accuracies. **Table 1** shows that with each random change in the dataset, the accuracy changes for both the PCA and
163 KNN classification methods. However, the LDA method gives an efficient classification with 100% accuracy.

164 **3.1 Principle Component Analysis (PCA)**

165 PCA is a statistical and unsupervised approach used for feature extraction and data compression [30,42]. Indeed,
166 the principle of this algorithm is to find a combination of input parameters that contains the most variance between
167 all data. The other interest is to remove the redundancy of information between these parameters [43]. The
168 disadvantage of this method is that it only allows decorrelating linearly correlated variables, but a non-correlation
169 does not mean a statistical independence.

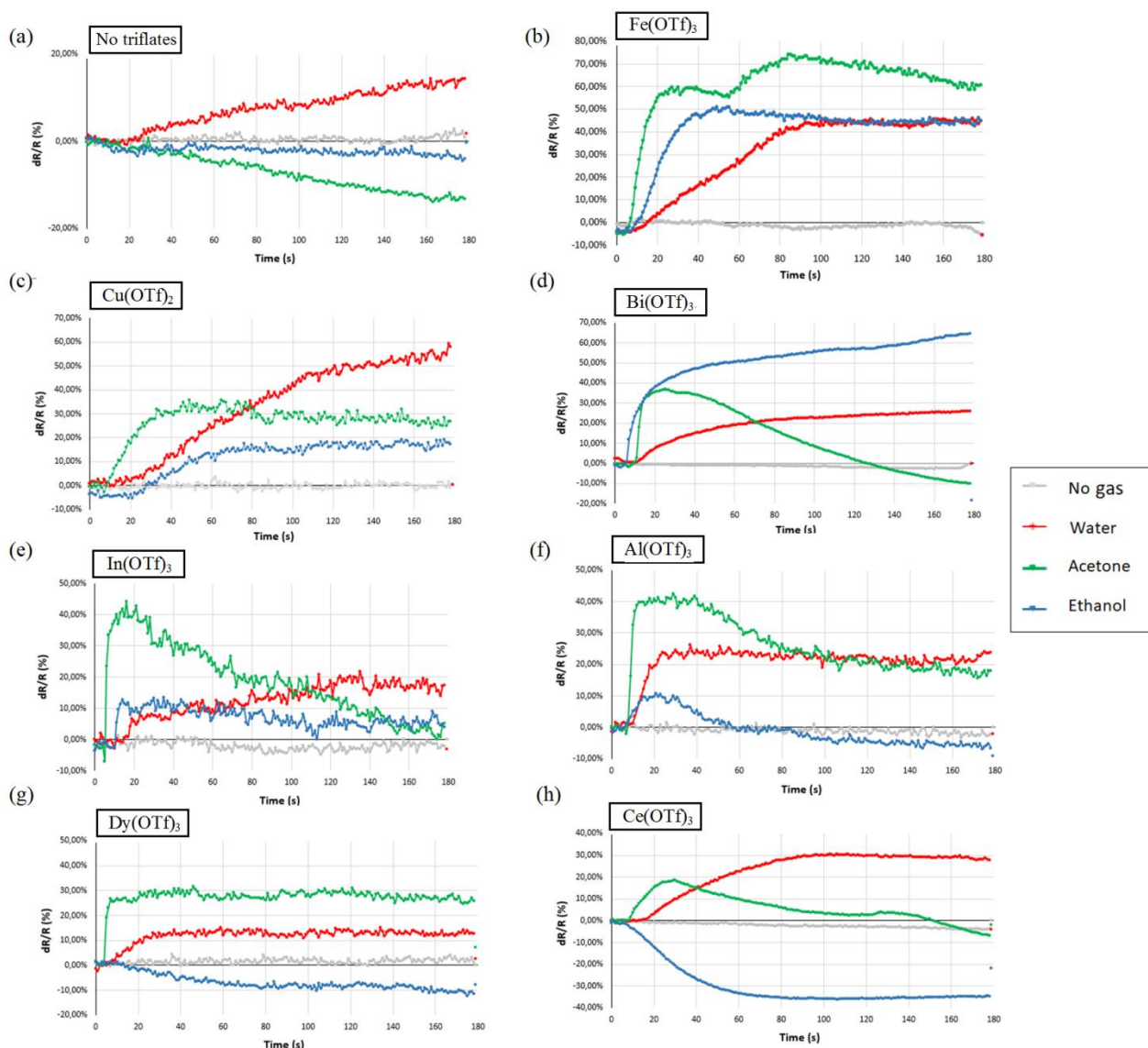
170 **3.2 Linear Discrimination Analysis (LDA)**

171 LDA is a supervised approach, which is considered as the most commonly used classification algorithm since it
172 gives good results and allows easy interpretation [44]. The basic function of this algorithm is to maximize variance
173 difference between data classes while minimize the variance difference inside each individual class [7]. Therefore,
174 if there is any groupings in data, the LDA is a powerful tool to recognize them. However, the disadvantage of this
175 classification algorithm is that it can only discriminate classes that are linearly separable.

176 **3.3 K- Nearest Neighbors (kNN)**

177 The K-nearest neighbors (kNN) algorithm is a Machine Learning algorithm that belongs to the class of simple
178 and easy-to-implement supervised learning algorithms that can be used to solve classification and regression
179 problems. In fact, the kNN is a special type of algorithm that does not use a statistical model [45]. It is "non-

180 parametric" based only on training data. The kNN algorithm is based on the selection of the k closest data points to
 181 the point under study in order to predict its value. In this work, we choose k between 1 and 50 as the best choice for
 182 the application since the highest classification rate was achieved with this value. The algorithm becomes much
 183 slower as the number of observations and independent variables increase.



184
 185 **Fig. 5** The resistance variation dR/R (%) versus time of the sensor array with 7 different dopants (triflates) during exposure to
 186 the different gases **a** no triflate, **b** $\text{Fe}(\text{OTf})_3$, **c** $\text{Cu}(\text{OTf})_2$, **d** $\text{Bi}(\text{OTf})_3$, **e** $\text{In}(\text{OTf})_3$, **f** $\text{Al}(\text{OTf})_3$, **g** $\text{Dy}(\text{OTf})_3$, and **h** $\text{Ce}(\text{OTf})_3$

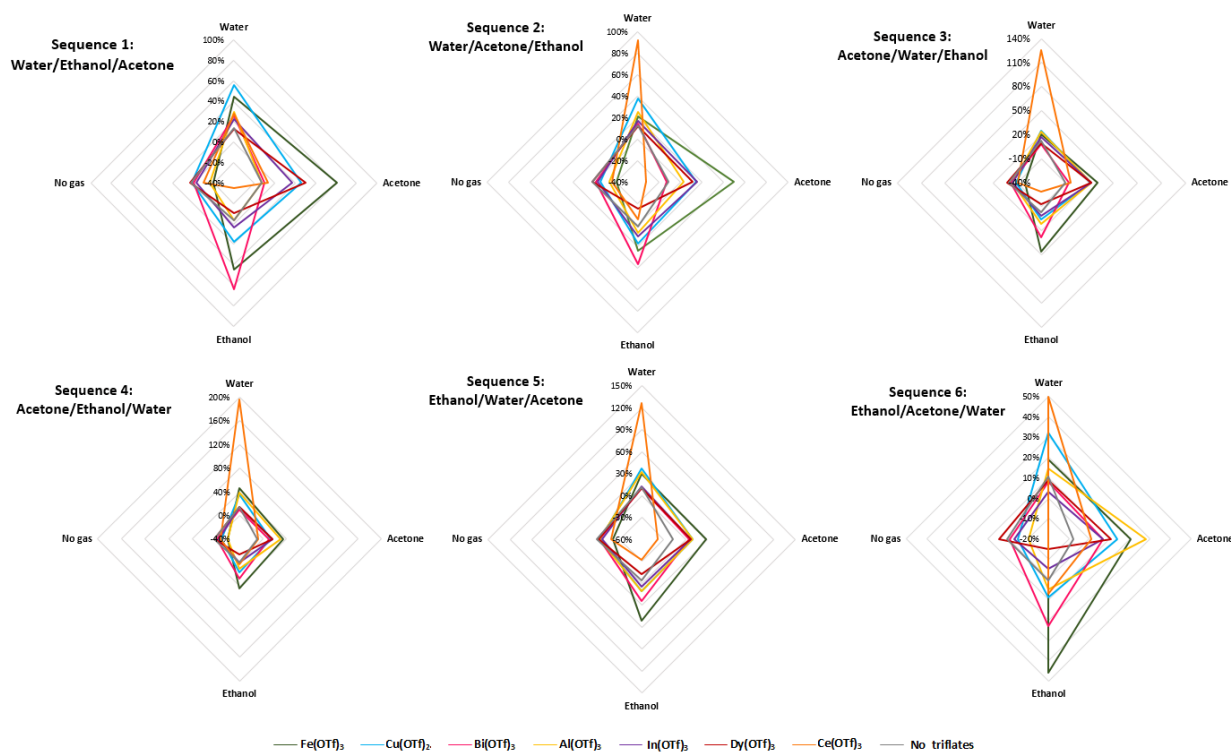
187 4 Results and Discussion

188 4.1 Sensor Response

189 We recently reported on the doping of P3HT with different triflates, which showed gas-specific sensitivity [36].
 190 We selected one of three devices for each gas sensor during a single sequence of exposure of the three gases taken
 191 at random to visualize the change in resistance over three minutes. The sensor responses show a small base resistance
 192 drift over time for some dopants (but not in the resistance modulation, which is used as a factor of merit for the
 193 analysis). This indicates the temporal stability of the triflate sensitized sensing elements over two hours of data

194 acquisition (The raw data are available as supplementary information S3 and S4 in the article ‘Mildly-doped
 195 polythiophene with triflates for molecular recognition’ [36]).

196 **Figure 5** shows the dR/R response of eight different sensor materials: P3HT without triflates and P3HT doped
 197 with Fe(OTf)₃, Cu(OTf)₂, Bi(OTf)₃, In(OTf)₃, Al(OTf)₃, Dy(OTf)₃ and Ce(OTf)₃. As we can see, each sensor showed
 198 different signature to each transported gas (water, acetone and ethanol). This would explain that each gas has an
 199 impact on the device conductance.



200
 201 **Fig. 6** Radar plot of six sequences of exposure of gases for eight different devices

202 The values of the resistance variation are specific from gas to another and from dopant to another that favor
 203 either increases or decreases in resistance. Moreover, **Fig. 6** represents the radar graph of each sequence of exposure
 204 of three gases for the eight devices. As we observe, in the absence of gas exposure for all eight devices, the dR/R
 205 variation of the response of each sensor is very low about 0%. This explains that the intervention of probed gas
 206 molecules is necessary to vary the devices resistance [24,36]. Moreover, it is noticed that pure P3HT without triflates
 207 gives relatively low resistance variation responses compared to other sensors with different dopants as shown in **Fig.**
 208 **5**. It interacts with water and acetone with a resistance variation about 14% and -13% respectively. While the reaction
 209 with ethanol is very limited.

210 Sensitivity and selectivity can be improved and optimized by introducing dopants that cause the modification of
 211 the energy band structure and morphology. Thus, they have a significant impact on the materials' conductivity
 212 [24,25]. Furthermore, in **Fig. 3**, it was found that devices doped with Fe(OTf)₃ is the most sensitive one to ethanol
 213 than the others. It presents a resistance variation of about 46%. Similarly, the devices doped with this same material
 214 produce the highest response with 35.3% as resistance variation when exposed to acetone while devices doped with
 215 Bi(OTf)₃ present the lowest response among all sensors with only 9.42% variations. While, devices doped with
 216 Ce(OTf)₃ are the most sensitive to water as it gives a response of 125.5% of resistance variation. In
 217 addition, the order of gas exposure has a strong effect on the resistance modulation changes. As we observed in **Fig.**

218 **6**, in the sequence 3, in the case of $\text{Ce}(\text{OTf})_3$ for example, when exposing water after acetone, gives a resistance
219 modulation about 125.5% which is different than when exposing water after ethanol in the sequence 4 (195.7%).

220 **4.2 Result of PCA classification**

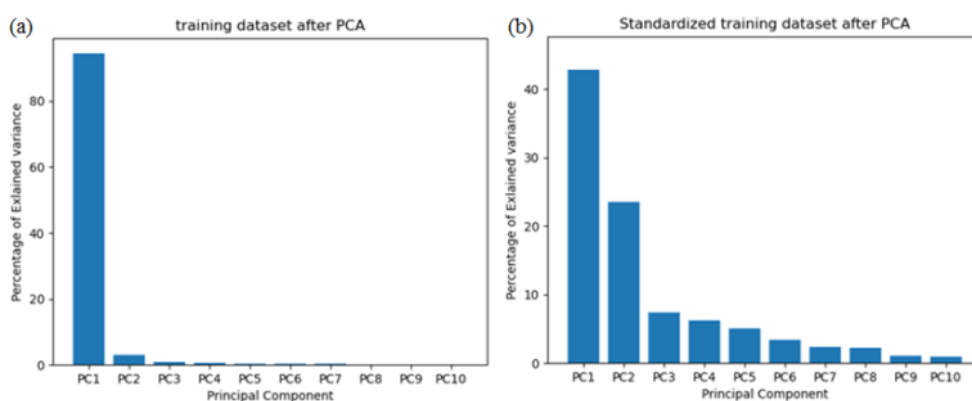
221 **4.2.1 Importance of feature Scaling**

222 The principal components are chosen to contain the maximum variance in the data and to be orthogonal [42,46].
223 While simplifying the interpretation of the data by the first two or three principal components (PC1, PC2, and PC3)
224 in two or three dimensions and preserving most of the variance in the data. The cumulative percentage explained is
225 obtained by adding the successive proportions of variances explained. For example for 94.4 % and 3.1 % is equal to
226 97.5 %, and so on. Therefore, about 97.5 % and 66% of the total variance is explained by the first two eigenvalues
227 using data without and with Standard scaler applied respectively (**Figure 7**).

228 Feature scaling through standardization would be an important preprocessing step for many machine-learning
229 algorithms, which consists of transforming numeric columns to a common scale. There are multiple normalization
230 techniques in statistics. The z-score method (often called standardization) transforms the data into a distribution with
231 a mean of 0 and a standard deviation of 1. Each standardized value is computed by subtracting the mean of the
232 corresponding feature and then dividing by the standard deviation.

233 Alternatively, we can use the StandardScaler class available in the Scikit-learn library to perform the z-score.
234 First, we create a standard_scaler object. Then, we calculate the parameters of the transformation (in this case the
235 mean and the standard deviation) using the .fit() method. Next, we call the .transform() method to apply the
236 standardization to the data frame. The .transform() method uses the parameters generated from the .fit() method to
237 perform the z-score.

238 While many algorithms (such as SVM, kNN, and logistic regression) require features to be normalized, Principle
239 Component Analysis (PCA) is a prime example of such an importance for it. In PCA, we are interested in the
240 components that maximize the variance. To illustrate this, PCA is performed to compare the use of data with
241 StandardScaler applied, to unscaled data. The results are visualized in the meantime a clear difference is noticed
242 (**Figure 8**).



243 **Fig. 7** The score plot of the data variance by the principal components **a** for training dataset after PCA, and **b** for standardized
244 training dataset after PCA
245

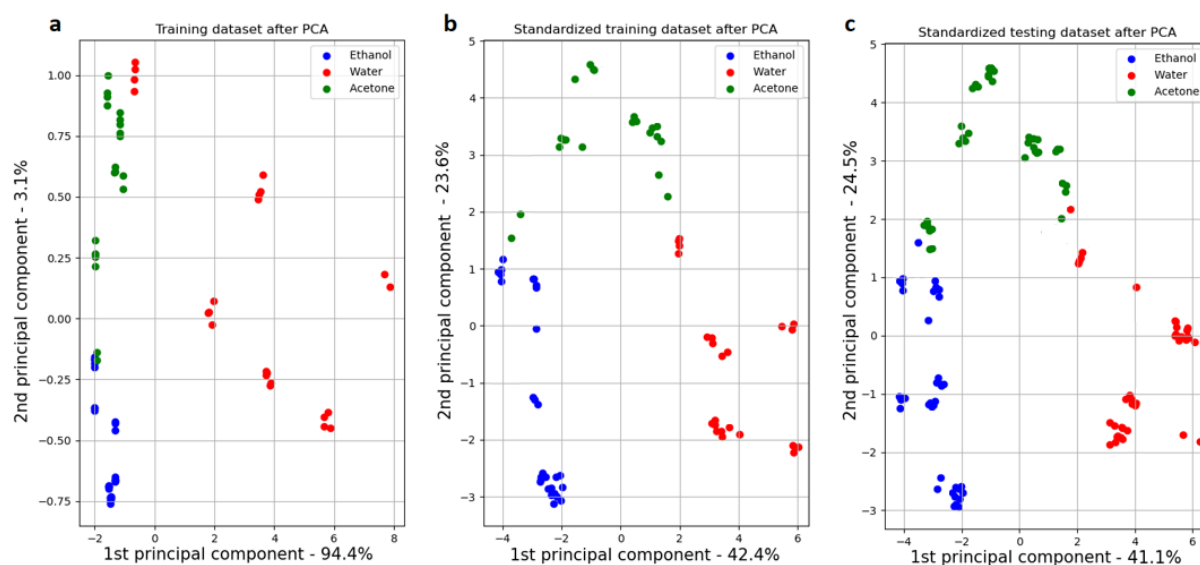
246 **4.2.2 Classification**

247 Here, we use the value of the resistances instead of the resistance modulation, and we apply the PCA method.
248 As we can see in **Fig. S1** (which is provided as supplementary information), the classification of the three classes is
249 not good. We observe an overlap between clusters with a low rate of accuracy (57.41%) because of the drift of

250 sensor responses. Therefore, we are interested only in studying the data of the resistance modulation dR/R in gas
251 classification.

252 **Figure 8** shows a 2D plot for the coefficients of the Principal Component Analysis for the three different gases
253 that group into three different clusters. Each gas is shown in different colors for visual identification. While these
254 clusters separated from each other, the separation between water, ethanol and acetone is defined. It is more
255 interesting to split data on train and test datasets to visually analyze the classification results. The data elements of
256 random data set 3 were split randomly on 40% for training and the rest for the testing. By the PCA before
257 standardizing training dataset (**Fig. 8a**), we observe that the clusters of acetone and ethanol are grouped on the left
258 of the graph, and the cluster of water is located on the right and in the middle of the graph. In addition, we notice
259 that there is a partial overlap between the water and acetone cluster and between the ethanol and acetone cluster.
260 The first two components, PC1 and PC2, contain 97.5% of the variance of the data. The first principal component,
261 PC1, explains 94.4% of the total variation, while 3.1% of the total variance is explained by PC2. This shows high
262 identification accuracy close to 82.51% for un-scaled data as compared to the rest of the variance components. The
263 results obtained by PCA after scaling provided a perfect classification. In **Fig. 8b**, about 66% of the total variance
264 of the data is displayed. PC1 and PC2 explain 42.4% and 23.6% of the variance respectively. The classification
265 accuracy obtained by the PCA method with standardization of training dataset was 93.52% and we observed that
266 the three gas groups are separated from each other with a little overlap.

267 After performing the PCA on the training set, the PCA was tested using test set, to generate overviews of the
268 observations. The score plots of the testing set using the first two components shows an overlap between
269 acetone/water and acetone/ethanol (**Fig. 8c**).

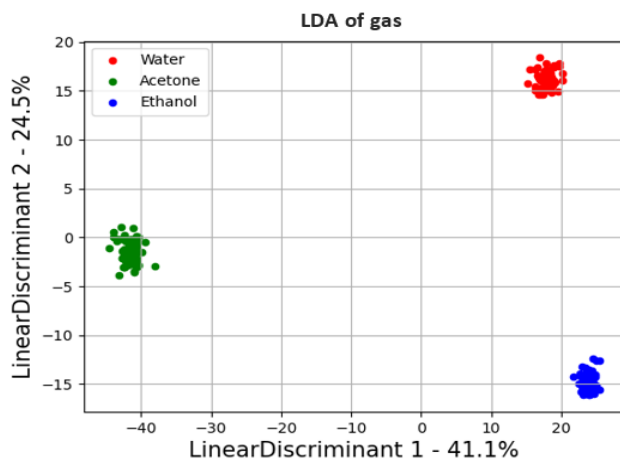


270
271 **Fig. 8** PCA classification with **a** training dataset after PCA, **b** Standardized training dataset after PCA, and **c** Standardized
272 testing dataset after PCA

273 4.3 Result of LDA classification

274 Similar to PCA, LDA also evaluates the distribution and distances, within and between classes [46]. The first
275 two components are 0.41 and 0.24. This technique can improve the resolution of existing classes. To detail the
276 classification performance, the different classification parameters such as sensitivity (or recall), accuracy and
277 specificity (or F1 score) were calculated (**Table 2**). The results of the LDA analysis show a clear identification as
278 shown in **Fig. 9**. The three gas groups are completely separated from each other. Prediction accuracy for the normal

279 test dataset with LDA is 100%. The results indicate that the e-nose can be used successfully as a rapid method for
 280 gas identification.

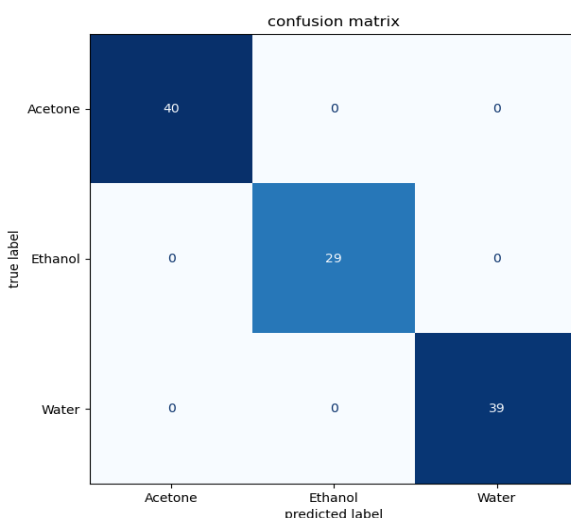


281
 282

Fig. 9 LDA classification

283 The use of a confusion matrix helps to evaluate the performance of a Machine Learning model and to check how
 284 often its predictions are correct compared to reality in classification problems. The rows of this matrix represent the
 285 true classes while the columns represent the predicted classes. The diagonal elements represent the number of points
 286 for which the observations are correctly classified, while the off-diagonal elements are those that are mislabeled by
 287 the classifier. The higher the diagonal values of the confusion matrix, the better indicating many correct predictions.
 288 The advantage of these matrices is that they are very simple to read and understand. They allow visualizing very
 289 quickly data and statistics in order to analyze the performance of a model and to identify trends that can help to
 290 modify the parameters.

291 As showed in **Fig. 10**, the non-diagonal cells are all zero, confirm that, the discrimination of classes is due
 292 without errors. In addition, we can see that the diagonal cells indicate the number of correct reclassifications by the
 293 trained observation data sets. For example, in 40 cases, acetone was correctly classified, corresponding to 100%
 294 accuracy of all 108 observational data sets. Because there was not a single case of incorrect classification, the
 295 discrimination value was 100% correct.



296
 297
 298

Fig. 10 Confusion matrix of LDA method

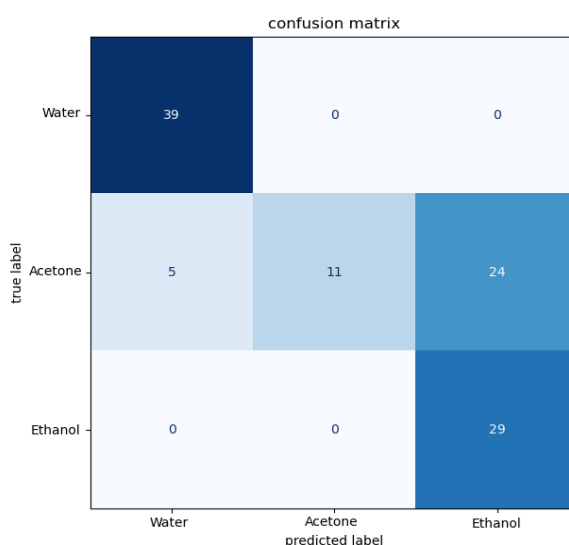
Table 2 Results of LDA and kNN algorithms

Algorithm	Precision %		Recall % (Sensitivity)		F1-score % (Specificity)		Support	
	kNN	LDA	kNN	LDA	kNN	LDA	kNN	LDA
Acetone	100	100	28	100	43	100	40	40
Water	89	100	100	100	94	100	39	39
Ethanol	55	100	100	100	71	100	29	29
Accuracy of algorithm	KNN 73.14 %			LDA 100 %			108	108

300

301 **4.4 Result of kNN classification**

302 For this last algorithm, we choose k between 1 and 50 to determine the accuracy of the prediction for unknown
 303 observations. KNN requires features to be normalized and we applied StandardScaler to unscaled data. This model
 304 shows an accuracy of 73.14 % only. We notice that the accuracy for acetone and ethanol is perfect at 100%. On the
 305 other hand, water shows an accuracy of 80 % (**Figure 11**). Hence, the misclassification is worse than the one by the
 306 LDA prediction algorithm.



307

308

Fig. 11 Confusion matrix of kNN method

309 In the next step, we tested to what extent the accuracy of the kNN algorithm depends on the number of neighbors
 310 (**Figure 12**). However, while increasing k from 50 up to 100, we did not see any further increase in accuracy. In
 311 contrast, the accuracy dropped slowly. Thus, the distances between the classes sufficiently separate when using 50
 312 neighbors. We may also be interested in a way to choose the k for which the classification would be the best. One
 313 way to find this is to plot the k value and the corresponding error rate for the dataset. Thus, we can detect the best
 314 prediction rate is obtained for a k between 1 and 50.

315 A brief comparison of our initial output results to those of other similar tools such as k-nearest neighbors (kNN)
 316 is shown in **Table 2**. As shown above, the performance of LDA model was perfect. While the kNN model showed
 317 a 73.14 % prediction performance, it is failed to successfully predict all of the labels for the test subset.

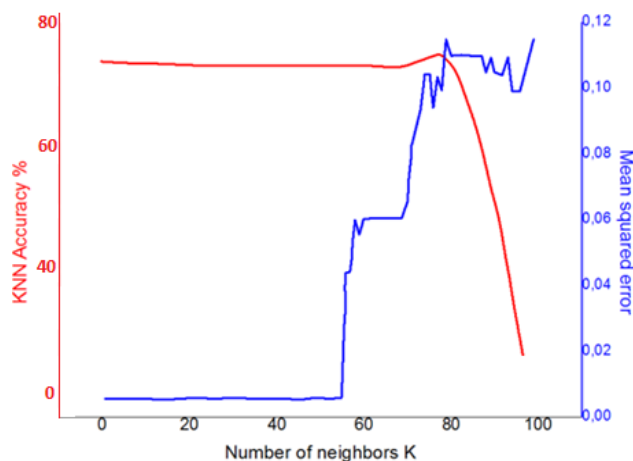


Fig. 12 kNN Accuracy depend on number of neighbors k

4.5 Effect of training datasets on accuracy algorithms

In this part, we were interested in testing several data sets in order to know the efficiency of each of these three algorithms when we chose a smaller training data set and a larger test data set. The objective is to have a device that we are able to calibrate the least number of times possible to be able to test it the most number of times possible. As we can see in Fig. 13, Both PCA and kNN classification methods give a high accuracy of about 100% for a high training datasets (80%). Each time we decrease the training datasets, the matching becomes weak and subsequently the accuracy of the classification decreases. For example, at a small training dataset, which is equal to 20%, the accuracy is 88.9% and 32% for PCA and kNN algorithm, respectively. On the other hand, the LDA technique shows a perfect accuracy for large testing datasets: 100% accuracy for a training dataset equal to 7%.

In this case, we can say that this method is efficient for the classification of different gases. In addition, LDA is widely used for various technologies of E-noses and materials other than conductometric sensor. Okur, S. et al, studied the identification of mint scents using a Quartz Crystal Microbalance (QCM) sensors Based E-Nose and verified that LDA method gives good results of classification with 100% accuracy [45,47]. Moreover, Akbar, M. A. et al, studied the classification of gases using 4x4 Array of tin-oxide based Gas Sensor and they compare results of PCA and LDA methods. The obtained results reveal that LDA reaches 100% classification accuracy with only 50% of the collected sample data. They verified that for the given sample data LDA with fewer components provide better classification than PCA [48].

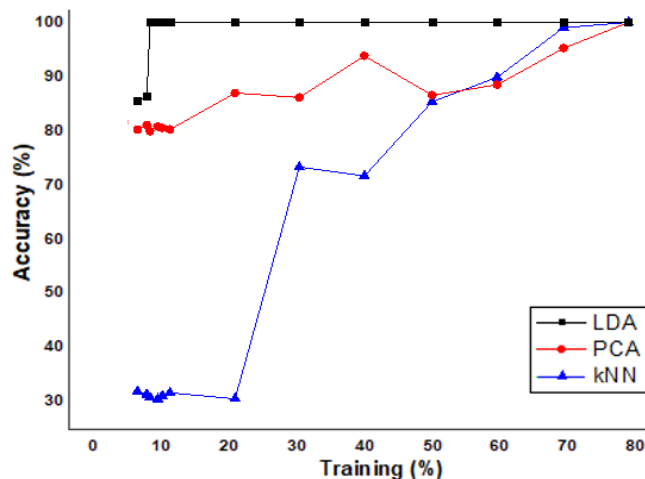


Fig. 13 Algorithm accuracies for several training datasets

339 **5 Conclusion**

340 In this work, we have developed the results of analysis and identification of the three studied gases: water,
341 acetone and ethanol. The raw data of the temporal responses of the gas sensors represent too much data. We have
342 therefore defined representative parameters of these curves, allowing a fast and reliable exploitation of the data using
343 supervised and unsupervised analysis algorithms.

344 The extracted parameters are the variation of the relative resistance dR/R recorded for the eight different
345 materials, three devices per material, with the output currents recorded overtime when exposed to the three different
346 gases, exposed six sequences of three minutes per gas [36]. We were interested in the steady-state data that contains
347 the output currents at the last 10 seconds of exposure. This state shows a clearer classification of the gas classes
348 compared to the other states [36]. We compared the following analysis methods: PCA, LDA and kNN for 40% of
349 training dataset. They present good identification of the gases with high accuracy. For the PCA algorithm, the
350 accuracy reaches 93.52%, 100% for the LDA and 73.14% for the kNN. In addition, we tested these techniques using
351 smaller training dataset and larger testing dataset, and we concluded that LDA is the efficient method for gas
352 discrimination.

353 In order to improve the classification performance of the e-nose system, future work will be doing the feature
354 engineering and using some feature extraction and feature reduction techniques. Among these feature parameters,
355 we noted the derivative and the integral as being transient parameters and that could even result in better
356 classification performance.

357 Thanks to this study that emphasize on the algorithm's choice, we will challenge our system with larger number
358 of classes with more gases (providing a protocol for exposing each gas for shorter time will be optimized with a
359 different setup, in preparation...)

360 **Supporting Information**

363 Supplementary information is available in the online version of the paper.

365 **Acknowledgements**

366 This work was in collaboration between the laboratory of microelectronics and instrumentation of the faculty of
367 sciences of Monastir, Tunisia and the central laboratory of IEMN in Lille, France. In addition, the gas sensor devices
368 used in this work are already fabricated and characterized and published in our paper entitled "Mildly-doped
369 polythiophene with triflates for molecular recognition [36]".

372 **Compliance with Ethical Standards**

374 This article does not contain any studies involving human participants performed by any of the authors.

376 **Declaration of Competing Interest**

377 The authors declare that they have no known competing financial interests or personal relationships that could have
378 appeared to influence the work reported in this paper.

380
381

382 **Research Data Policy and Data Availability Statements**

383

384 The authors agree with the availability of data transparency as per journal guidelines.

385

386 **Authors' contribution statement**

387

388 **Aicha Boujnah:** Methodology, Formal analysis, Investigation, Writing – original draft, Writing - review & editing,

389 Visualization. **Aimen Boubaker:** Writing – review & editing, Project administration, Funding acquisition.

390 **Sebastien Pecqueur:** Validation, Formal analysis, Investigation, Writing– review & editing, Supervision. **Kamal**

391 **Lmimouni:** Writing – review & editing, Project administration. **Adel Kalboussi:** Writing – review & editing,

392 Project administration, Funding acquisition.

393

394 **References**

395

396 1. J. Tan and J. Xu, *Artif. Intell. Agric.* **4**, 104 (2020) <https://doi.org/10.1016/j.aiia.2020.06.003>

397 2. F. Rabeb, B. Souhir, K. Abdermaceur, and S. Mounir, in 14th Int. Conf. Sci. Tech. Autom. Control Comput.
398 Eng. - STA'2013 (IEEE, 2013), pp. 174–178 <https://doi.org/10.1109/STA.2013.6783126>

399 3. J. A. Covington, S. Marco, K. C. Persaud, S. S. Schiffman, and H. T. Nagle, *IEEE Sens. J.* **21**, 12969 (2021)
400 <https://doi.org/10.1109/JSEN.2021.3076412>

401 4. L. Tiggemann, S. C. Ballen, C. M. Bocalon, A. M. Graboski, A. Manzoli, J. Steffens, E. Valduga, and C.
402 Steffens, *Innov. Food Sci. Emerg. Technol.* **43**, 112 (2017) <https://doi.org/10.1016/j.ifset.2017.08.003>

403 5. S. Bedoui, R. Faleh, H. Samet, and A. Kachouri, in 10th Int. Multi-Conferences Syst. Signals Devices 2013
404 (IEEE, 2013), pp. 1–6 <https://doi.org/10.1109/SSD.2013.6564152>

405 6. W. Hu, L. Wan, Y. Jian, C. Ren, K. Jin, X. Su, X. Bai, H. Haick, M. Yao, and W. Wu, *Adv. Mater. Technol.*
406 **1800488** (2018) <https://doi.org/10.1002/admt.201800488>

407 7. R. Gutierrez-Osuna, *IEEE Sens. J.* **2**, 189 (2002) <https://doi.org/10.1109/JSEN.2002.800688>

408 8. Y. Wang, A. Liu, Y. Han, and T. Li, *Polym. Int.* **69**, 7 (2020) <https://doi.org/10.1002/pi.5907>

409 9. A. D. Wilson, *Procedia Technol.* **1**, 453 (2012) <https://doi.org/10.1016/j.protcy.2012.02.101>

410 10. A. Nasri, M. Pétrissans, V. Fierro, and A. Celzard, *Mater. Sci. Semicond. Process.* **128**, 105744 (2021)
411 <https://doi.org/10.1016/j.mssp.2021.105744>

412 11. G. Korotcenkov, V. Brinzari, and B. K. Cho, *Microchim. Acta* **183**, 1033 (2016)
413 <https://doi.org/10.1007/s00604-015-1741-z>

414 12. F. A. Sébastien Pecqueur, Maurizio Mastropasqua Talamo, David Guérin, Philippe Blanchard, Jean Roncali,
415 Dominique Vuillaume, *Adv. Electron. Mater.* **4**, (2018) <https://doi.org/10.1002/aelm.201800166>

416 13. M. Angelopoulos, *IBM J. Res. Dev.* **45**, 57 (2001) <https://doi.org/10.1147/rd.451.0057>

417 14. M. Yasin, T. Tauqeer, K. S. Karimov, S. E. San, A. Kösemen, Y. Yerli, and A. V. Tunc, *Microelectron. Eng.*
418 **130**, 13 (2014) <https://doi.org/10.1016/j.mee.2014.08.010>

419 15. H. S. Nalwa, *Handbook of Advanced Electronic and Photonic Materials and Devices* (2001).

420 16. H. Seon, B. Kim, and J. Kang, *IEEE Trans. Nucl. Sci.* **64**, 1739 (2017)
421 <https://doi.org/10.1109/TNS.2016.2645228>

422 17. S. Nambiar and J. T. W. Yeow, *Biosens. Bioelectron.* **26**, 1825 (2011)
423 <https://doi.org/10.1016/j.bios.2010.09.046>

- 424 18. R. Megha, F. A. Ali, Y. T. Ravikiran, C. H. V. V. Ramana, A. B. V. Kiran Kumar, D. K. Mishra, S. C.
425 Vijayakumari, and D. Kim, *Inorg. Chem. Commun.* **98**, 11 (2018) [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.inoche.2018.09.040)
426 [inoche.2018.09.040](https://doi.org/10.1016/j.inoche.2018.09.040)
- 427 19. D. K. A. and S. K. Gupta, *Science and Technology of Chemiresistive Gas Sensors* (2007).
- 428 20. K. Ferchichi, R. Bourguiga, K. Lmimouni, and S. Pecqueur, *Synth. Met.* **262**, 116352 (2020)
429 <https://doi.org/10.1016/j.synthmet.2020.116352>
- 430 21. C.-G. Kuo, J.-H. Chen, Y.-C. Chao, and P.-L. Chen, *Sensors* **18**, 37 (2017) <https://doi.org/10.3390/s18010037>
- 431 22. A. Assadi, G. Gustafsson, M. Willander, C. Svensson, and O. Inganäs, *Synth. Met.* **37**, 123 (1990)
432 [https://doi.org/10.1016/0379-6779\(90\)90135-8](https://doi.org/10.1016/0379-6779(90)90135-8)
- 433 23. C. Bertoni, P. Naclerio, E. Viviani, S. Dal Zilio, S. Carrato, and A. Fraleoni-Morgera, *Sensors* **19**, 1296 (2019)
434 <https://doi.org/10.3390/s19061296>
- 435 24. Z. Ma, W. Shi, K. Yan, L. Pan, and G. Yu, *Chem. Sci.* **10**, 6232 (2019) <https://doi.org/10.1039/C9SC02033K>
- 436 25. J.-S. Do and S.-H. Wang, *Sensors Actuators B Chem.* **185**, 39 (2013)
437 <http://dx.doi.org/10.1016/j.snb.2013.04.080>
- 438 26. J. N. Barisci, G. G. Wallace, M. K. Andrews, A. C. Partridge, and P. D. Harris, *Sensors Actuators B Chem.*
439 **84**, 252 (2002) [https://doi.org/10.1016/S0925-4005\(02\)00033-3](https://doi.org/10.1016/S0925-4005(02)00033-3)
- 440 27. J. W. Gardner and P. N. Bartlett, *Sensors Actuators B Chem.* **18**, 210 (1994) [https://doi.org/10.1016/0925-](https://doi.org/10.1016/0925-4005(94)87085-3)
441 [4005\(94\)87085-3](https://doi.org/10.1016/0925-4005(94)87085-3)
- 442 28. R. Faleh, M. Othman, S. Gomri, K. Aguir, and A. Kachouri, *IEEE Sens. J.* **16**, 3123 (2016)
443 <https://doi.org/10.1109/JSEN.2016.2521578>
- 444 29. S. Fuentes, V. Summerson, C. Gonzalez Viejo, E. Tongson, N. Lipovetzky, K. L. Wilkinson, C. Szeto, and R.
445 R. Unnithan, *Sensors* **20**, 5108 (2020) <https://doi.org/10.3390/s20185108>
- 446 30. V. Schroeder, E. D. Evans, Y.-C. M. Wu, C.-C. A. Voll, B. R. McDonald, S. Savagatrup, and T. M. Swager,
447 *ACS Sensors* **4**, 2101 (2019) <https://doi.org/10.1021/acssensors.9b00825>
- 448 31. S. Güney and A. Atasoy, *Sensors Actuators B Chem.* **166–167**, 721 (2012)
449 <https://doi.org/10.1016/j.snb.2012.03.047>
- 450 32. O. Djedidi, M. A. Djeziri, N. Morati, J.-L. Seguin, M. Bendahan, and T. Contaret, *Sensors Actuators B Chem.*
451 **339**, 129817 (2021) <https://doi.org/10.1016/j.snb.2021.129817>
- 452 33. R. Ionescu, E. Llobet, X. Vilanova, J. Brezmes, J. E. Sueiras, J. Calderer, and X. Correig, *Analyst* **127**, 1237
453 (2002) <https://doi.org/10.1039/B205009A>
- 454 34. C. Gonzalez Viejo, S. Fuentes, A. Godbole, B. Widdicombe, and R. R. Unnithan, *Sensors Actuators B Chem.*
455 **308**, 127688 (2020) <https://doi.org/10.1016/j.snb.2020.127688>
- 456 35. Z. Haddi, A. Amari, H. Alami, N. El Bari, E. Llobet, and B. Bouchikhi, *Sensors Actuators B Chem.* **155**, 456
457 (2011) <https://doi.org/10.1016/j.snb.2010.12.047>
- 458 36. A. Boujnah, A. Boubaker, A. Kalboussi, K. Lmimouni, and S. Pecqueur, *Synth. Met.* **280**, 116890 (2021)
459 <https://doi.org/10.1016/j.synthmet.2021.116890>
- 460 37. H. MEN, K. NING, and D. CHEN, *Sensors and Transducers* **157**, 57 (2013).
- 461 38. M. Abdelkhalek, S. Alfayad, F. Benouezdou, M. B. Fayek, and L. Chassagne, *IEEE Access* **7**, 98267 (2019)
462 <https://doi.org/10.1109/ACCESS.2019.2928875>
- 463 39. A. Berna, *Sensors* **10**, 3882 (2010) <https://doi.org/10.3390/s100403882>

- 464 40. P. Giungato, M. Renna, R. Rana, S. Licen, and P. Barbieri, *Food Res. Int.* **115**, 65 (2019)
465 <https://doi.org/10.1016/j.foodres.2018.07.067>
- 466 41. G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, *GetMobile Mob. Comput.*
467 *Commun.* **19**, 29 (2015)
- 468 42. M. Bougrini, K. Tahri, Z. Haddi, T. Saidi, N. El Bari, and B. Bouchikhi, *J. Sensors* **2014**, 1 (2014)
469 <https://doi.org/10.1155/2014/245831>
- 470 43. M. A. H. Khan, B. Thomson, R. Debnath, A. Motayed, and M. V. Rao, *IEEE Sens. J.* **20**, 6020 (2020)
471 <https://doi.org/10.1109/JSEN.2020.2972542>
- 472 44. D. H. Y. Jan Hendrik Leopold^{1, 2}, Lieuwe D J Bos^{1, 3}, Peter J Sterk³, Marcus J Schultz¹, Niki Fens³, Ildiko
473 Horvath⁴, Andras Bikov⁴, Paolo Montuschi⁵, Corrado Di Natale⁶, *J. Breath Res.* **9**, (2015)
- 474 45. S. Okur, M. Sarheed, R. Huber, Z. Zhang, L. Heinke, A. Kanbar, C. Wöll, P. Nick, and U. Lemmer,
475 *Chemosensors* **9**, 31 (2021) <https://doi.org/10.3390/chemosensors9020031>
- 476 46. S. A. Fatemi Heydarabad, M. H. Raoufat, S. Kamgar, and A. Karami, *J. Food Meas. Charact.* **13**, 107 (2019)
477 <https://doi.org/10.1007/s11694-018-9924-z>
- 478 47. X. Su, D. Chen, N. Li, A. C. Stevenson, G. Li, and R. Hu, *Sensors Actuators A Phys.* **305**, 111938 (2020)
479 <https://doi.org/10.1016/j.sna.2020.111938>
- 480 48. M. A. Akbar, A. Ait Si Ali, A. Amira, F. Bensaali, M. Benammar, M. Hassan, and A. Bermak, *IEEE Sens. J.*
481 **16**, 5734 (2016) <https://doi.org/10.1109/JSEN.2016.2565721>
- 482