How Well Does Your E-nose Detects Cancer? Application of Artificial Breath Analysis for Performance Assessment

Justin D.M. Martin[a,\*](#_bookmark0) [[1]](#footnote-1), Falzone Claudia a, Anne-Claude Romain a

a *Department of Environmental sciences, Sensing of Atmospheres and Monitoring (SAM), SPHERES Research Unit, University of Liège, 6700 Arlon, Belgium*

*\* Corresponding author at: University of Liège, Arlon Campus, 6700 Arlon, Belgium E-mail address: jdm.martin@uliege.be (J. Martin).*

*Keywords:* Metal oxide, Detection, Lung cancer, Benchmarking, Electronic nose, Breath.

# Abstract

Comparing electronic nose performance is a challenging task because of a lack of standardised method. This paper proposes a method for defining and quantifying an indicator of the effectiveness of multi-sensor systems in detecting cancers by artificial breath analysis. To build this method, an evaluation of the performances of an array of metal oxide sensors built for use as a lung cancer screening tool was conducted. Breath from 20 healthy volunteers has been sampled in FEP sampling bags. These healthy samples were analysed with and without the addition of 9 VOC cancer biomarkers, chosen from literature. The concentration of the VOC added was done in increasing amounts. The more VOC were added, the better the discrimination between “healthy” samples (breath without additives) and “cancer” samples (breath with additives) was. By determining at which level of concentration the electronic nose fails to reliably discriminate between the two groups, we estimate its ability to well predict the presence of the disease or not in a realistic situation. In this work, a home-made electronic nose is put to the test. The results underline that the biomarkers need to be about 5.3 times higher in concentration than in real breath for the home-made nose to tell the difference between groups with a sufficient confidence.

# Introduction

Detecting subtle variations in breath, sometimes around parts per billion (ppb), is a difficult task. Nevertheless, dogs have shown their ability to tell the difference between people with a breath indicating lung cancer (LC), and people without. For example, Feil et al. used a dog to successfully detect 40 out of 41 cancer samples amongst 150 healthy controls, using a combination of urine and breath samples [1]. Amazingly, ants have shown some skill in doing so as well [2].

A lot of research has been done on cancer biomarkers in breath, and no consensus has been found on which volatile organic compound (VOC) is a marker of cancer [3]. Classifying breath might be more complex than isolating a few VOCs. Maybe a more appropriate tool would better account for the complexity of the breath, and allow to generalise this approach as cancer screening ?

Of course, LC screening already exists, we can find examples in the now infamous NELSON study [4], the MILD [5] and the NLST [6]. However, these studies show that on average, about 3% of cases in the population at risk come up positive for malignant LC, and it is recommended to do the screening again each year. Since this kind of screening uses low-dose computed tomography and chest X-rays, screening a country-wide population could quickly become a challenge for healthcare centres: multipurpose, expensive machines that need qualified personnel are required to make this possible. It is possible that making this kind of screening more common could become a logistical bottleneck. So far, the large-scale implementation of this kind of screening has not taken place. A simpler, inexpensive, non-invasive screening tool – that could be used at the general practitioner’s office for example – could therefore reduce the workload of healthcare centres by selecting the truly suspicious cases amongst the population at risk.

Several methods stood up to the challenge, such as variants of gas chromatography mass spectrometry (GCMS) [7], colorimetric sensors [8], and of course electronic noses. Each method is being refined by their respective communities and have their own issues to solve before being used at a large scale. E-noses try to prove their ability to go above and beyond by identifying the stage of the detected LC and distinguish it from the most common comorbidities [9]. However, there is one common problem amongst all LC screening studies.

The issue with emerging methods is that there are not enough samples to be analysed to reach strong statistical significance. The conclusion of most articles, regardless of the method, has some variation of “promising results” and “should increase the amount of data to validate the findings”. The reasons behind this are not clear, and maybe linked with the small amount of positive LC diagnostics that usually happen in a single healthcare centre over an usual study timeframe. To reach significant numbers, studies like NELSON had to work with a large amount of healthcare centres and have a timespan of several years. This requires a lot of resources, which most research groups likely have little access to. As reported before [10,11], 58% of all LC breath biomarkers GCMS studies between 1985 and 2019 fail to gather more than 100 participants, and 85% have no more than 200.

Of course, there are a few exceptions that begin to reach a believable level of statistical significance, for example a recent study included 2308 participants for a Time Of Flight Mass Spectrometry (TOF-MS) analysis of breath for LC detection [12]. The Pathacov Project [13], which is also the frame of this paper, reported having reached around 1200 participants with a goal for 300 more for a GCMS study on LC breath biomarkers.

The electronic nose, another promising screening tool, also struggles with reaching large number of participants to prove the worth of the method in the medical field. For example, one of the largest studies gathered 575 participants [14] (i.e. 13 times less than the NELSON study). Even big studies like Nakhleh et al. included a small number of cancer patients (45 out of 813) [15]. Conclusions drawn from such small number of samples can be overly optimistic and give little information about the final performance of the device once in use by patients in real life conditions.

When developing an electronic nose, a lot of very important conception aspects need to be addressed [3]. These lead to choices which influence the ability of the electronic nose to perform as a screening tool, but in often unpredictable ways. A consequence of this is the very diverse prototypes that are being developed, without any real way to decide which one has the best design, or if modifications would improve or not their performances.

This paper presents the metric of performance using a previously reported method [3] to evaluate the ability of an electronic nose to discriminate between healthy breath and artificial breath with cancer biomarkers. This type of approach has been used previously in a medical context [16], but this time increasing concentrations are used to find the limit of discrimination. The performance metric can be used to compare devices and allow the improvement of designs. Conclusions regarding the adequacy of electronic noses for early lung cancer detection can be drawn from the application of this metric on an experimental e-nose.

The number of active projects in this very specific field is surprisingly high, and the number of publications is significant as shown on table 1. The metric could therefore help compare which approach performs better.

Table 1 was obtained by following the method: Scopus and PubMed were searched using the following keywords: *(“e-nose” OR “array” OR “sensor” OR “breath”) AND (“lung cancer”) NOT (“colorimetric” OR “in-vivo” OR “condensate”)*. The references of the selected documents were examined to find other pertinent articles. 525 articles published from 2000 to 2023 in English language were reviewed. Meta-analysis and review articles were not read. 43 articles were relevant to the subject. 9 were excluded for the following reasons: for the lack of a healthy control group (4), for lack of information on the disease researched (1), for presenting a single sensor as an e-nose (1), for the lack of classification results (3). In the end, the 34 selected articles are displayed in Table 1.

**Table 1**

List of publications using a clinical trial with e-noses for lung cancer detection. Publications cover years from 2000 to the time of publication.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Reference and year** | **Method details** | **Sampled population** | **Visualisation and feature extraction** | **Classification method** | **Validation and results** |
| **(Se = Sensitivity, Sp = Speciticity, Acc= Accuracy, LC = Lung Cancer, H = Healthy (control), OD = Other Diseases )** | | | | | |
| 2003 Di Natale [17] | LibraNose | 42LC/18H  LC includes patients before/after surgery | Not mentioned | PLS-DA | Leave-one-out Cross validation  Se 87.5% Sp 95.5% Acc 90,3% |
| 2005 Chen [18] | Custom system with GC column, SPME fiber adsorption and 2 SAW sensors. | 20LC/15H  +7 OD | Not mentioned | ANN with back-propagation (BP) | Tested on 5LC/5H subgroup.  Se 80% Sp 80% |
| 2005 Machado [19] | Cyranose 320 (commercial e-nose with polymer sensors) + Mylar bag sampling | 28LC/107H | PCA + CDA | Support Vector Machine (SVM) | Separate validation population (63%).  Se 71,4% Sp 91,9% Acc 90,3% |
| 2007 Blatt [20] | Custom E-nose with 6 MOS sensors | 43LC/58H | PCA + LDA | Fuzzy k-NN | Cross-validation  Se 95.3% Sp 90.5% Acc 92.6% |
| 2009 Dragonieri [21] | Cyranose 320 and tedlar bag with filter | 10LC/10H +10OD (COPD) | PCA | LDA | Cross validation value + Mahalanobis distance  85% correct classification, with M-distance of 3.73 between LC and COPD. |
| 2010 D’Amico [22] | Libranose (from C. Di Natale team) | 28LC/36H +28OD | PLS-DA | PLS-DA | Leave-one-out cross-validation.  Se 90% Sp 100% Acc 93.7% |
| 2012 Chapman [23] | Cyranose 320 with 2L sampling bag (material unspecified) | 20 LC/42H +18OD (asbestos related study) | PCA | LDA | Cross validation  Se 90%, Sp 91%, Acc 95%. |
| 2012 Peled [24] | E-nose (from H. Haick team)  Mylar bag +  Tenax adsorption | 50LC (malignant) /19 benign nodules and afflictions | Not mentioned | Discriminant factor analysis (DFA) | Leave-one-out cross validation  Se 86%, Sp 88% Acc 88%? AUC 0.986 |
| 2012 Wang [25] | “HENS” E-nose with 9 commercial MOS + GC-SAW in parallel. Preconcentration with Tenax-TA | 47LC/42H | PCA | ANN | Leave-one-out cross-validation  Se 93.6% Sp 83.4% |
| 2013 Broza [26] | Custom E-nose with 6 nanoparticle sensors (from H. Haick team) and mylar bag (offline sampling) | 12LC/5H  Before/after surgery | Discriminant Factor Analysis (DFA) | DFA | Cross validation  Se 100% Sp 80% |
| 2014 Bikov [27] | Cyranose 320 | 27LC/37H  Tested influence of sampling conditions | PCA | LDA | Leave-one-out cross validation  Se 63% Sp 78% Acc 72% |
| 2015 De Vries [28] | SpiroNose (7 MOS sensors in a spirometry apparatus) | 31LC/45H  +68OD | PCA | LDA | Leave-one-out cross validation  Acc 88% |
| 2015 McWilliams [29] | Cyranose 320 with Mylar bag. | 25LC/166H (control group consists of active and former smokers. Comorbidity considered) | MWW (discriminant descriptor extraction),  Scatter plots. | CART (2-3 nodes) (Classification and Regression Trees) + DFA. | Leave-One-Out Cross Validation  Se 84% Sp 81,3% |
| 2016 Gasparri [30] | Libranose (8 QMB sensors) + tedlar bag (alveolar breath) | *70 LC/76H* | Not mentioned | Partial Least Square Discriminant Analysis (PLS-DA) | 30% of dataset for validation  *Se 81% Sp 91%* |
| 2016 Rocco [31] | Bionote (7 QCM sensors) Uses Tenax sorption | 23LC/77H | Not mentioned | PLS-DA | Leave-one-out cross validation  Se 86% Sp 95% |
| 2017 Cai [32] | zNose4200 (micro-GC with SAW sensor detector and preconcentration) | 57LC/72H +118 in external validation dataset | Chi-square + ANOVA | Logistic regression Analysis (LRA) | External validation dataset  Se 76% Sp 94% |
| 2017 Li [33] | Custom E-Nose with commercial sensors, offline sampling with Tedlar gas bags. | 24LC/23H +5 OD | Compared PCA, LDA, Laplacian Eigenmap, local linear embedding, t-Stochastic Neighbor Embedding | Fuzzy k-Nearest Neighbors (kNN) and SVM | 10-fold cross validation.  Best results with LDA-Fuzzy 5-NN  Se 91,6% Sp 91,7% Acc 91,6%  And LDA-SVM  Se 90,8% Sp 84,2% Acc 57,6% |
| 2017 Nakhleh [15] | Custom e-nose (from H. Haick team) | 45LC/591H +768OD (16 diseases) | Linear Regression Models | DFA | 23% of dataset for blind validation.  Average performances for LC:  Se 84% Sp 88% Acc 86% |
| 2017 Shlomi [34] | E-nose with 40 custom sensors (from H. Haick team) + Tenax® sorption | 89LC/30Benign  Different cancer mutations and smoking habits | Not mentioned | DFA | Leave-one-out cross validation  Se 75% Sp 93% ROC-AUC = 0.89 |
| 2017 Tizite [35] | Cyranose 320 + PET sampling bag | 165LC/79H  +91 with other diseases | Not mentioned | SVM | 25% of dataset for validation  LC vs OD Se 88,9% Sp 66,7%  LC vs H Se 97,8% Sp 68,8%  Cancer staging: 40% misclassification |
| 2018 Huang [36] | Cyranose 320, tedlar bag + endotracheal sampling. | *56LC/188H* | PCA | LDA, SVM | 20% of dataset for validation, double cross validation with two nested loops. External validation one year later.  Se 83,3% Sp 86,2% |
| 2018 Kort [37–39] | Aeonose | 144LC/146H (Non-Small Cell Lung Cancer (NSCLC)) | ANOVA, Kruskall Wallis or chi-squared test | ANN | 10% of dataset for validation  Se 94,4% Sp 32,9% |
| 2018 Van de Goor [40] | Aeonose (experimental e-nose with 3 MOS sensors and sorption step on Tenax®) | 60LC/107H | Sample t test  Fisher’s exact test | Artificial neural network (ANN) | 10% of dataset for blinded validation.  Se 83% Sp 84% |
| 2019 Kononov [41] | Custom E-nose (6 MOS) with direct breathing | 65LC/53H | PCA | Logistic regression | 30% of dataset for validation + Grid-search cross validation.  *Se 95% Sp 100% Acc 97.2%* |
| 2019 Lu [42] | Custom E-Nose, offline sampling with Tedlar gas bags (commercial sensors of mixed types) | 98LC/116H | Gated recurrent unit based autoencoder (GRU-AE), compared with PCA. | Compared 7 different models, including adaboost, several ensemble pruning methods and complete ensemble. | 50-fold cross validation.  GRU-AE + MSEP acchiev best results with ;  Se 94,22% Sp 92,80% Acc 93,55%  GRU-AE treated data demonstrated better results than data treated with PCA  Could differentiate stage III and IV with about 80% Acc, Se and Sp. |
| 2019 Marzorati [43] | Custom e-nose with 4 commercial MOS | 6LC/10H | Not mentioned | ANN | Leave-one-out cross-validation.  Se 85.7% Sp 100% Acc 93.8% |
| 2019 Tirzite [44] | Cyranose 320 | 252LC/223H +OD? | Not mentioned | LRA | No validation  Se 95.8% Sp 92.3% |
| 2020 Saidi [45] | Custom E-nose with 6 custom UV irradiated WO3 sensors (MOS). | 32LC/12H | PCA | DFA | Leave-one-out + Bootstrap validation  Acc 98.6% |
| 2021 Rodriguez-Aguilar [46] | Cyranose 320 | 30LC/50H  +100OD (Breast cancer and COPD) | PCA | Canonical analysis of principal coordinates (CAP) | leave-one-out cross-validation  91.4% Acc |
| 2021 Binson [47–49] | Custom e-nose with 5 commercial MOS sensors | 32LC/72H + 38OD (COPD) | PCA | Naïve Bayes, LDA, LRA, kNN and SVM | 3-fold, 5-fold, 10-fold cross-validation.  Se 91.3% Sp 84.4% Acc 94.4%(with kNN) |
| 2021 Chen [9] | Custom e-nose with 11 commercial sensors: 2 electrochemical sensors, 7 MOS, 1 hot wire and 1 catalytic combustion. | 101LC/134H  Staging classification (III/IV) | Kernel PCA | SVM + Extreme gradient boosting (XGBoost) | 10-fold cross-validation  Se 93.6% Sp 91.1% Acc 93.6% |
| 2023 De Vries [50] | SpiroNose | 211LC/682 COPD  5.45% of included COPD patients developed LC | PCA | LDA | Validation on 1/3 of data  Se 86%, Sp 89, Acc 87% |
| 2023 Kort [14] | Aeonose | 239LC/336H | Not mentioned | Average between ANN, Support Vector Machine (SVM), Random Forest (RF), XGBoost, logistic regression. | Validation on 1/3 of data.  Se 88%, Sp 52%  Included the number of pack-years which improved results. |
| 2023 Hao [51] | ILD3000 (3 electrochemical sensors, + 1 MOS sensor) | 103LC/61H | PCA and Genetic Algorithm (GA) | Improved Adaptative Boosting (ImAdaBoost) + others tested | k-fold cross-validation + external validation dataset  Se 92.5% Sp 92.2% Acc 92.4% |

Table 1 teaches us that the approaches to breath sensing for lung cancer detection are numerous, and are all promising in their accuracy to detect lung cancer. 54% undergo small size clinical trials (with less than 100 participants) to prove their approach adequate, which give some encouraging evidence of the power of the method, but small sample pools lack statistical power. There is too much difference between studies to reliably compare results solely on their accuracy, sensitivity and specificity. A metric of performance to compare devices would help in improving existing approaches.

# Material and methods

The experiment consisted of collecting healthy human breath, injecting this breath with biomarkers in different concentrations using mass flow controllers, and then analysing breath (with or without biomarkers) using TD-GC-MS and an electronic nose (SAMBre). The steps are detailed below.

* 1. *Breath sampling volunteers*

The collection of breath started with a call for volunteers for healthy breath sampling. It took place on the ULiège Environmental Science Campus in the city of Arlon, Belgium. Volunteers were not compensated for their participation. The goal of the recruitment was to take in as many different volunteers as available, as the participation rate was expected to be low. The number of smokers and ex-smokers is therefore rather low as they are a minority on site. Having subgroups equal in size (i.e., according to sex, smoking or sport habits) was not deemed attainable. The group is therefore not ideal in terms of age, smoking habits, or sex ratio, and could result in a perceptible bias on the resulting data.

To participate, volunteers had to be 18 years old at least and not have any breathing issues or lung illness. To increase participation, the volunteers could come on the day that was most fitting in their schedule. They had to confirm the days a week prior and participate at least 4 times over the course of the study.

Finally, 21 healthy volunteers participated to the campaign. They had to fill in a short questionnaire about their biometrics (sex, age, size, and weight) and habits (history and frequency of smoking, frequency of sport activities) before the first sampling, and report if any change occurred during the study.

The details of the volunteers are available in Table 2.

**Table 2**

Statistics for the 21 healthy volunteers

|  |  |  |  |
| --- | --- | --- | --- |
| **Male (%)** | 8 (38.0%) | **BMI – Average** | 23.5 |
| **Age – Average** | 37.5 | **BMI – Median** | 24.3 |
| **Age – Median** | 36.0 | **BMI – Range** | 18.4 – 29.4 |
| **Age – Range** | 21 - 62 | **Sport – Daily (%)** | 2 (9.5%) |
| **Smokers (%)** | 2 (9.5%) | **Sport – Weekly (%)** | 10 (47.6%) |
| **Ex-smokers (%)** | 3 (14.3%) | **Sport – Monthly (%)** | 5 (23.8%) |
| **Never-smokers (%)** | 16 (76.2%) | **Sport – Rarely (%)** | 4 (19.1%) |

For breath collection, it was requested of participants that they fast overnight and respect a “nothing by mouth” policy the morning of the sampling (i.e., no smoking, teeth brushing, chewing, or eating. Drinking water was allowed). Then, they were given water for mouth rinsing and filled a 10L Fluorinated Ethylene Propylene (FEP) sampling bag through a spirometry saliva filter in a Polytetrafluoroethylene (PTFE) holder.

The breath of the 21 healthy participants was collected on 26 different days across four months (August 2022 and January to March 2023), leading to a total of 126 unique breath samples being collected. An average of 5 breath samples was collected on each occasion; each participant gave about 6 samples over the course of the study on average (with a median of 3.5).

The details of the breath samples are available in Table 3.

**Table 3**

Statistics for the 126 healthy breath samples collected.

|  |  |  |  |
| --- | --- | --- | --- |
| **Male (%)** | 40 (31.7%) | **BMI – Average** | 24.0 |
| **Age – Average** | 36.6 | **BMI – Median** | 26.0 |
| **Age – Median** | 34.0 | **BMI – Range** | 18.4 – 29.4 |
| **Age – Range** | 21 - 62 | **Sport – Daily (%)** | 15 (11.9%) |
| **Smokers (%)** | 7 (5.6%) | **Sport – Weekly (%)** | 61 (48.4%) |
| **Ex-smokers (%)** | 18 (14.2%) | **Sport – Monthly (%)** | 32 (25.4%) |
| **Never-smokers (%)** | 101 (80.2%) | **Sport – Rarely (%)** | 1. 14.3%) |

* 1. *Artificial cancer breath creation*

Once the breath is collected, half of each sample is injected with VOC (Volatile Organic Compound) biomarkers using the following method.

The whole procedure is identical to one being used in previous publications [3,11]. Parts per million (ppm) level concentrations are reached using microinjections: a micro-syringe is used to insert between 0.1 and 2µL of pure compound in a pre-filled 8L FEP sampling bag (HedeTech®, Dalian, China). The injection of the compounds is followed by 30 minutes of heating in an oven at 60°C. Reaching sub-ppm level concentrations is done through a dilution of the injected bag using mass flow controllers. In this case, the dilution air is human breath from volunteers’ samples.

The injected compounds are those proposed in [3], and come from a citation frequency analysis of literature about GCMS studies of lung cancer biomarkers in breath. Injected volumes and compounds are as follow:

* Acetone 2,1 µL (Merck© KGaA, Darmstadt, Germany)
* Ethanol 1,8 µL (Thermo Scientific©, Pittsburgh, USA)
* 2-propanol 1,8 µL (Merck© KGaA, Darmstadt, Germany)
* 1-propanol 0,5 µL (Thermo Scientific©, Pittsburgh, USA)
* 2-pentanone 0.1 µL (Thermo Scientific©, Pittsburgh, USA)
* N-dodecane 0.1 µL (VWR international©, Radnor, USA)
* Hexanal 0.1 µL (Merck© KGaA, Darmstadt, Germany)
* Toluene 0.1 µL (Thermo Scientific©, Pittsburgh, USA)
* 2-butanone 0.1 µL (Merck© KGaA, Darmstadt, Germany)

The bag containing these vaporized compounds was put under 1 bar of pressure in a pressure chamber. The flow of the concentrated biomarkers was modulated in regard to the desired final concentration in artificial cancer breath. To reach literature concentrations, 3.93mL/min was used. Multiples of that value were used to reach the concentrations of other test groups: 7.86mL/min (2x), 15.72mL/min (4x), 30.0mL/min (8x). All were diluted against 1300mL/min of breath.

The method used in this paper has been previously presented [12]. The sample creation process is summarized in Figure 1. The following is the application of the methodological considerations.

A diagram of a sample

Description automatically generated

**Fig. 1:** Sample creation process detailed in [12], using real breath with and without biomarker additions for analysis by e-nose.

* 1. *Electronic nose analysis*

Every sample (both healthy and artificial cancer) was sucked in through an experimental electronic nose, SAMBre, for 5 minutes. Before and after each sample, reference air was sucked in for at least 5 minutes. The flow is always kept at 200mL/min by a downstream pump and flowmeter. The reference air is pure analytical-grade air (Air Liquide®, Paris, France) humidified at 40% HR at 20°C then stored in a 25L FEP bag. The switching between samples and reference air is done through a manual 3-way valve.

The composition of the SAMBre electronic nose has been detailed in another publication [11], and can be found in supplementary materials.

All data treatment has been done using the JMP® software (SAS Institute®, USA).

* 1. *TD-GC-MS Analysis*

78 of the 252 samples (31%), both with and without biomarkers addition, were adsorbed on Tenax® TA (Buchem™ B.V., Apeldoorn, The Netherlands) cartridges for Thermal Desorption Gas Chromatography Mass Spectrometry (TD-GC-MS) analysis. 1L of each sample was pumped at 120mL/min through the sorbent cartridge. Tenax has been chosen based on its frequent use in breath analysis in literature, as it does not retain water and works well with VOCs. The method used for this analysis and the materials has already been reported before [11]. The samples were chosen at random and processed right away. All the samples were not absorbed mainly for logistical reasons (limited number of cartridges, and the long time needed for the analysis and the cleaning process). There are several goals for this analysis:

* Check the reproducibility and accuracy of the dilution process and the composition of the samples.
* Observe any tendencies in the samples that could correlate to MOS gas sensor behaviour.

Before analysis, a calibration was done with all 9 compounds using three concentrations (equivalent to 0.005 ppmv, 0.1 ppmv, 0.3 ppmv in air). The concentrations were obtained by microinjections of pure compounds in vials of methanol with a septum cap. The resulting mixture is then spiked on cartridges using a calibration solution loading rig (Markes International®, Llantrisant, UK). The regression line obtained is then used to compute the concentration of the corresponding compounds based on the area under the chromatogram peaks.

* 1. *E-nose performance index*

In regard of a discrimination task, a better performing e-nose can classify samples into categories with less error – using an identical classification algorithm. Classification error rate should therefore be usable as a metric of e-nose performance in a discrimination task. In this regard, a key factor is the Classification Performance Criterion (CPC). To determine the best-performing devices, a severe CPC is essential. We use the P4 metric as a measure of classification quality, defined by the following equation [52]:

Equation 1:

In this equation, TP represents the number of true positives, TN is the number of true negatives, while FP and FN denote the number of false positives and false negatives, respectively. The P4 metric is selected for its ability to encompass precision, recall, specificity, and negative predictive value, thus providing a more comprehensive evaluation of classification quality than metrics like F1 or Accuracy [52].

To meet the CPC, the e-nose must achieve a P4 score of 0.9 or higher with a minimum of 35 samples from each group (healthy, cancer, and others if applicable), totalling in this case 70 samples. Future comparisons with results from real medical sampling conditions will inform whether the CPC needs to be adjusted to be more or less stringent.

Following the establishment of the CPC, the next step involves calculating the Discrimination Performance Score (DPS) of the e-nose. This involves exposing the e-nose to samples of healthy breath and artificial cancer breath, which is created by adding biomarkers to the analyzed healthy breath. The biomarkers are initially added at concentrations realistic to those reported in literature. The e-nose's performance in classifying these samples accurately is then evaluated. If the e-nose fails to meet the CPC at these initial concentrations, the concentration of biomarkers is doubled for a repeated test and so on until success is achieved. The point at which the success is achieved gives the DPS, as per Equation 2:

Equation 2:

Here, DPS is the Discrimination Performance Score, 'A' is the highest concentration multiplier at which the e-nose fails to meet the CPC, and 'B' is the lowest concentration multiplier at which the e-nose succeeds. A score around 1 indicates the e-nose's ability to detect realistic levels of biomarkers in artificial cancer breath, with higher scores suggesting better resolution (Figure 2). For additional precision in the DPS, testing intermediate concentrations between 'A' and 'B' is possible, particularly when finer accuracy is needed. However, this requires larger sample sizes to ensure statistical significance.

**1**

**<1**

**>1**

Performance is adequate for the task.

Underperforming

Samples are too subtle for the e-nose to work as intended.

Overperforming

The e-nose could perform with subtler samples.

The higher the better. Twice higher, twice better.

**Fig. 2.** Diagram illustrating the DPS interpretation for three cases.

The classification algorithm chosen for this evaluation is k-Nearest Neighbours (kNN). Its selection is due to its simplicity and effectiveness as a non-parametric method, suitable for non-normal and/or heteroscedastic datasets. In cases of small sample pools, like those in this study, assumptions of parametric methods are often unmet, making kNN a preferable choice over methods like LDA.

* 1. *Calculation of sample size*

According to the reference method [3], the minimum value for this kind of test has been chosen as 30 breath samples of each group (healthy/cancer). For this study, several concentrations of additives were planned, therefore each subgroup of concentration had at least 30 healthy and 30 artificial cancer breath samples.

* 1. *Ethics approval*

The study was approved by the ULiège-CHU ethics committee and volunteers gave their informed consent to the study. Data collection and handling is GDPR compliant. This study was done under the supervision of a medical practitioner specialised in medical research.

* 1. *Data analysis*

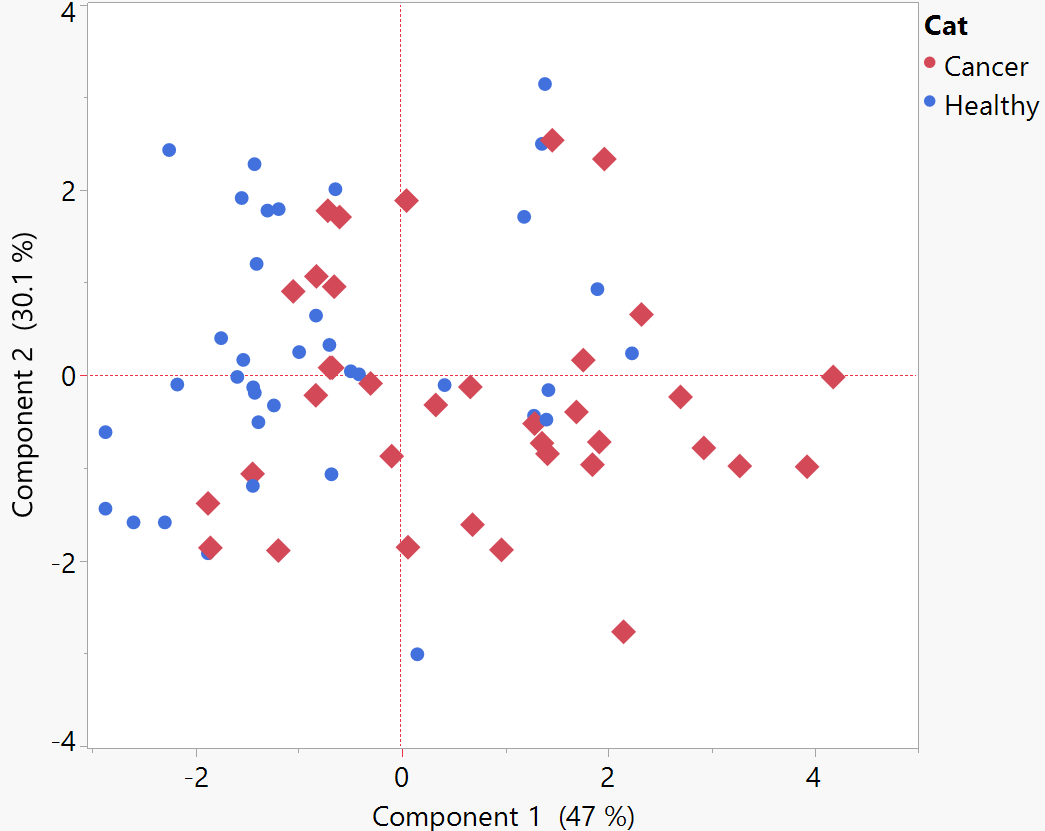
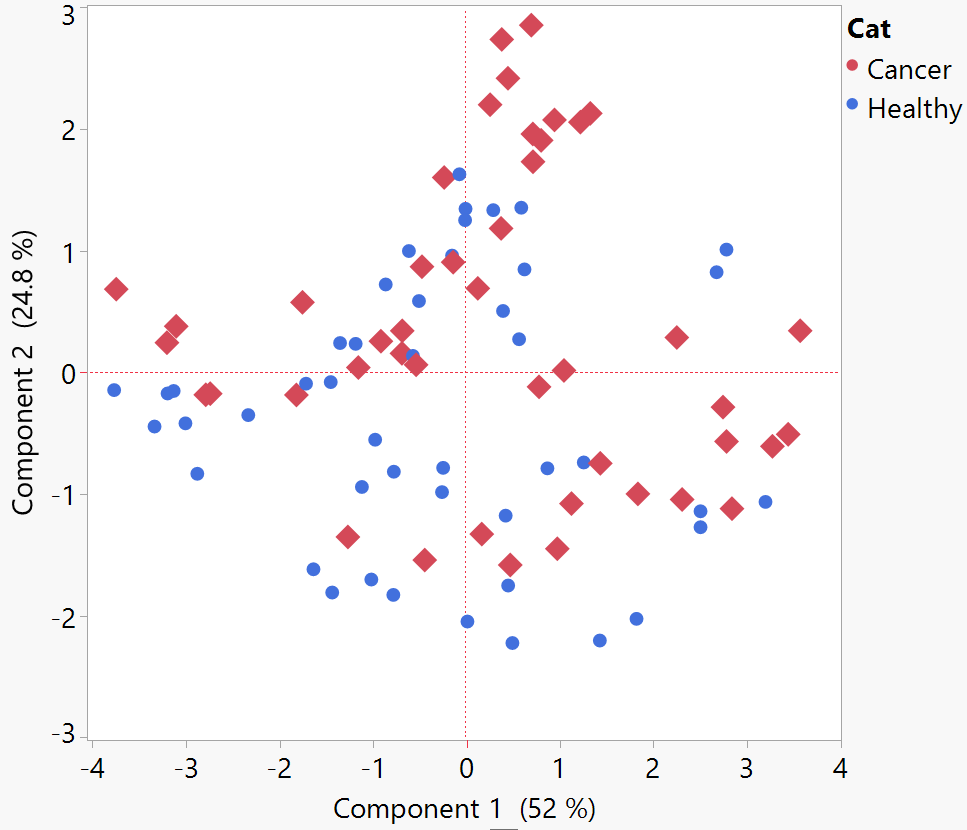
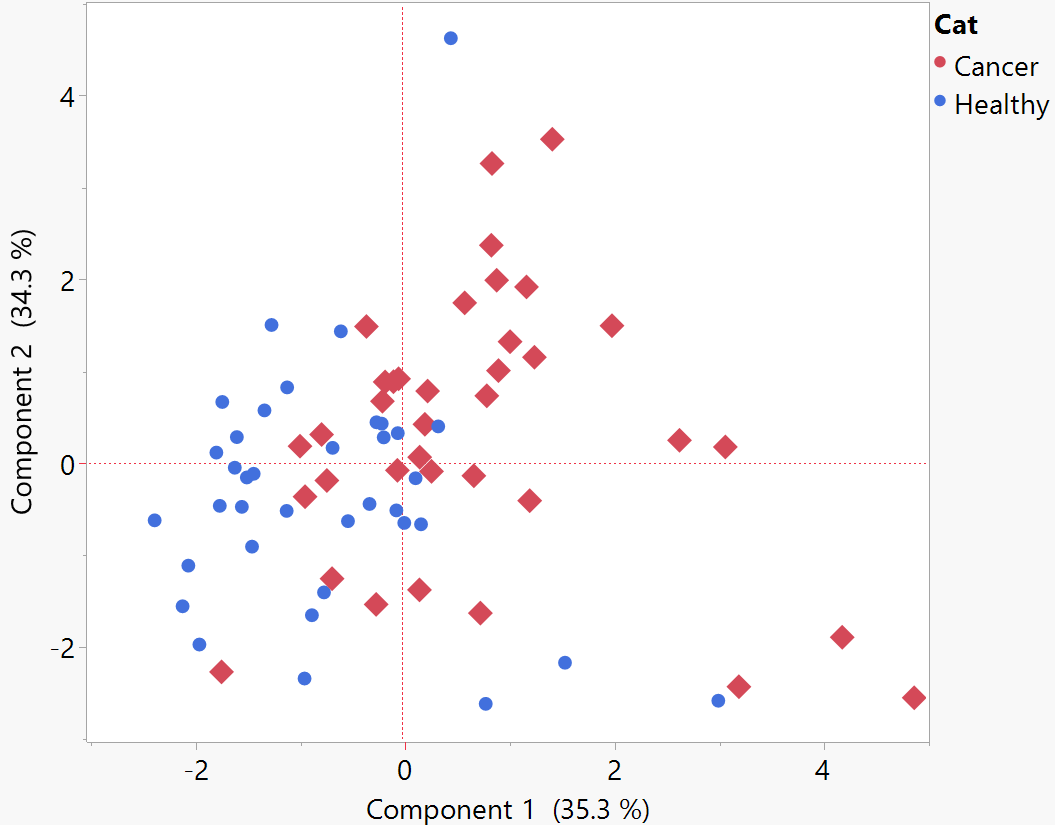
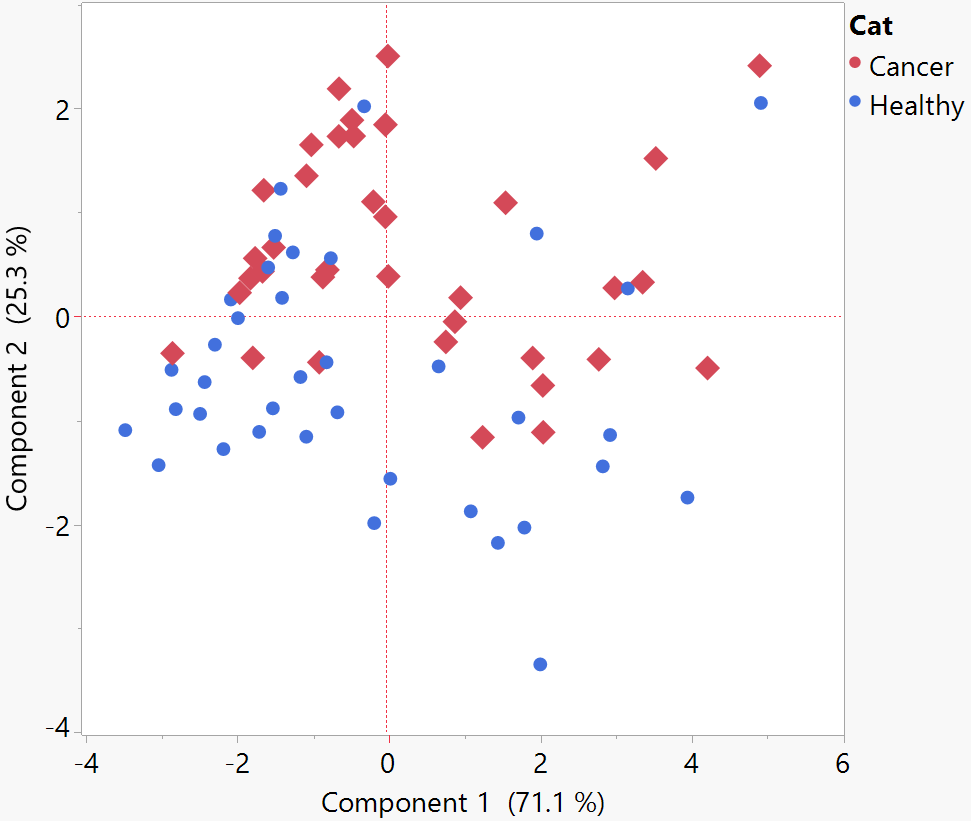
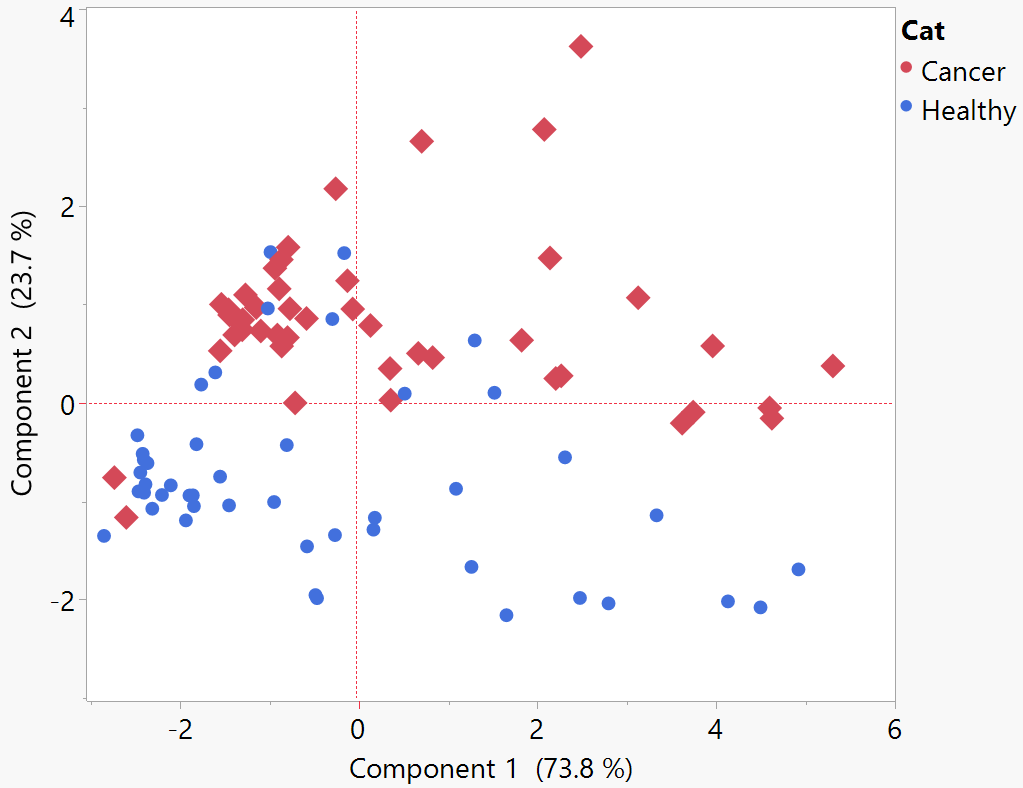
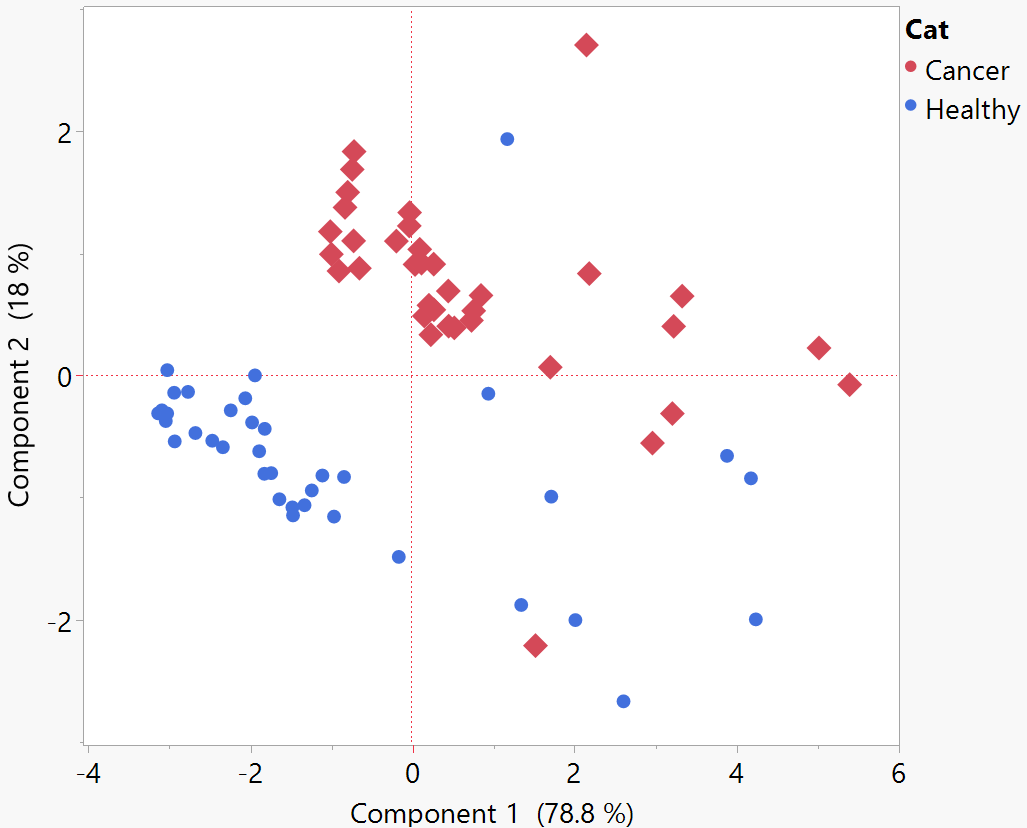
From the raw data, for each sample, the maximum conductance value of each sensor is extracted for analysis. The conductance value of the baseline (i.e., the stable conductance of the sensor in the reference air) is subtracted from this value. There is an apparent drift between data from 2022 and 2023. The drift can be compensated using calibration data. In this case, calibration is done before measurement at least once a month. Increasing concentrations of ethanol (0.1, 0.5, 1.0, 1.5, 2.0 ppm) are analysed by the electronic nose, which gives a calibration line that allows for drift compensation. More information on this can be found in the supplementary materials.

# Results and discussion

* 1. *Concentration effect removal*

As illustrated on Fig 3, the normalization of the data removes the concentration effect. This has the consequence of making artificial cancer breath a homogeneous group, which is less distinct from the healthy breath group.

**Fig. 3.** PCA score plot of the data without normalisation (a,b,c), and with normalisation (d,e,f) separated by concentration. Healthy samples are blue dots, while artificial cancer breath are red squares. 1.5x group is (a,d), 4x group is (b,e) and 8x group is (c,f). Concentration is expressed relatively to concentrations expected in real cancer breath according to literature. Despite normalization, boundaries between groups are clearer as concentration increases. The third principal component is available in supplementary materials.



**A**

**B**

**C**

**F**

**E**

**D**

This means that the electronic nose is sensitive to differences in concentration but lacks resolution in mixture identification. In this case, sensors are highly correlated to one another, with two main groups of sensors giving complementary data: MP901 and G8530T, G2530T are correlated. T2603, G3530T, G1430T are also highly correlated with one another. The main selection criterion for these sensors was their lower limit of detection, which in commercial sensors often indicate either SnO (tin oxide) or ZnO (zinc oxide) based sensors [53]. This probably explains the appearance of two groups of highly correlated sensors. This is however unfortunate as less correlated sensors lead to higher information quality, as it means each sensor evaluates an aspect of the sample other sensors do not [54].

It is however interesting to note that, without normalization, most artificial cancer breath of realistic concentration (Fig. 3A) is very close to healthy samples, but with a distinct centroid. This is also very visible on the third component (see supplementary materials). This indicates that despite the low resolution, the electronic nose is still able to label mixtures as artificial cancer breath regardless of the low amount of biomarkers added.

* 1. *Discrimination efficiency*

As stated before (see 2.1), the quality of an electronic nose in a discrimination task can be evaluated by the number of misclassifications. For each concentration group, the number of neighbours (k) taken in consideration was the one giving the smaller percentage of misclassification, with k between 1 and 10. The results for the k-NN analysis are detailed on Table 4 below.

**Table 4**

Results for a k-Nearest Neighbours training using normalized data. Different concentrations of biomarkers in breath, expressed relatively to mean literature cancer breath concentrations.

|  |  |  |  |
| --- | --- | --- | --- |
| **Concentration** | **1.5x** | **4x** | **8x** |
| **P4** | **0.709** | **0.879** | **0.932** |
| **Misclassified samples (%)** | **28.76%** | **12.08%** | **6.75%** |
| **Sensitivity** | **0.806** | **0.867** | **0.946** |
| **Accuracy** | **0.712** | **0.879** | **0.932** |
| **Precision** | **0.674** | **0.886** | **0.921** |
| **F1 Score** | **0.734** | **0.876** | **0.933** |
| **Healthy samples** | **37** | **46** | **36** |
| **Artificial cancer samples** | **36** | **44** | **37** |

Quality of the classification indeed rises with higher concentrations as expected. It is notable that, even with low biomarkers additions, the electronic nose still manages to perform relatively well, with a sensitivity of 0.81, which have been considered noteworthy in other publications [23,55–57]. This is however misleading, as Fig. 3D and 3A shows, most of the samples are very close to the boundaries of the healthy group. This level of accuracy would not likely survive a larger sample pool.

The required performances on such small sample pools should be very high to be considered noteworthy, as only high-quality performances devices are likely to perform as intended once in real sampling conditions with a larger number of samples. This should translate in a very high threshold needed to trigger the performance index (see 3.4).

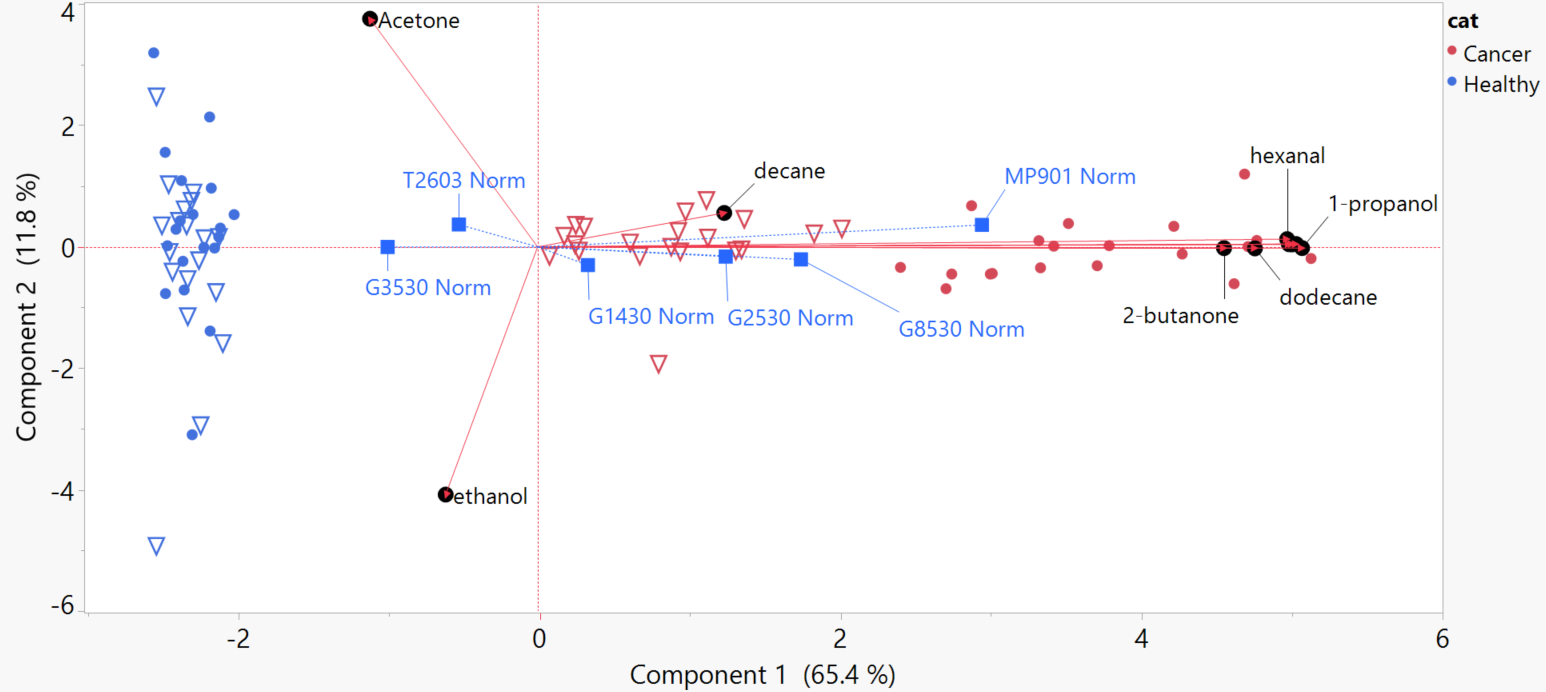
* 1. *Effect of confounders*

As said before, a questionnaire was submitted to the volunteers with questions about basics habits (smoking and sports) as well as biometrics. We have found that with or without normalisation, no cluster or trend could be seen for the recorded parameters. Active and former smokers were not apart from nonsmokers. This is likely thanks to the “nothing by mouth” policy which allows for limited contamination of samples. Persons from various body mass indexes, age, frequency of sport practices, are blended together in PCA on all components as well.

According to the reference method [3], it was recommended to include artificial Chronic Obstructive Pulmonary Disease (COPD, frequent comorbidity with lung cancer) breath and artificial smoker breath (provided healthy smokers are not available) in the trial along the artificial cancer breath. This is intended to evaluate the ability of the e-nose to not confuse COPD/smoking markers for cancer. Due to logistic and scaling limitations, it was chosen not to include theses group in the trial. The inclusion of other types of artificial breath will be considered for future trials.

* 1. *TD-GC-MS analysis*

Principal component analysis on the results of the TD-GC-MS analysis shows that artificial cancer breath samples are clearly distinct from healthy breath samples, as illustrated by Figure 4 and that additions of biomarkers worked as intended. Several observations arise from this:

* The spread of healthy breath sample is very limited on the first component, while the spread of artificial cancer breath is much wider, showing a greater variety in the composition of artificial cancer breath. This is likely linked to a poor reproducibility of the injection and dilution process used to create artificial cancer breath, as the first component is linked to biomarkers concentration except for acetone and ethanol. More precise dilution and evaporations systems, such as the use of a permeation oven, could improve this aspect. The use of microfluidics for precise mixture creation would also be a very interesting path of development. The variety in composition does not hinder the whole experiment as variation within breath is expected, and conclusions can be drawn from the result even with the observed level of variations in mixtures.
* The spread of healthy samples is very wide on the second component, which is mostly linked to ethanol and acetone concentration. Both these compounds are not well retained on Tenax, which might explain the variability. Ethanol being used as a disinfectant on lab material between volunteers, contamination was expected, and is likely the explanation behind healthy outliers. Better drying procedure are to be considered in future experiments, or the inclusion of non-emissive disinfection procedures such as autoclaving. Fasting has also been reported to induce higher acetone breath content [58], which could explain some of the spread in Fig. 4.
* Sensors 2530T, 8530T and MP901 have responses more strongly correlated to biomarkers than other sensors, which confirms their adequate ability to discriminate healthy and cancer breath. It is however impossible with this dataset to identify which biomarkers correlate the most with the sensors’ responses.

**Fig. 4.** PCA bi-plot of the TD-GC-MS data, with normalized sensor responses correlation added to the loadings plot (in blue squares). Red is for the cancer group; blue is for the healthy group. Triangles show the “4x” dataset, while dots show the “8x” dataset.

* 1. *Performance index of SAMBre*

As a reminder, the performance criterion is met if the P4 CPC reaches 0.9 or more with at least 35 samples of each group.

The criterion is met somewhere between the “4x” concentration group and the “8x” concentration group. According to the Equation 1 (see 2.1), the performance score is therefore:

(1/4+1/8)/2 = 0.1875

Which indicates that the biomarkers need to be about 5 times more concentrated than in real breath to be reliably discriminated by the device.

In its current state, SAMBre is therefore unsuited for medical screening of cancer. However, concentrating the biomarkers using sorption on Tenax should enable the concentration needed to increase the performance score. Swapping sensors for more sensitive and less correlated counterparts is another strategy to increase the performance score.

The quality of this score is however without real validation: comparison with performances during a large-scale screening campaign using real patients should be realised to confirm that a low/high score translates into low/high classification performances in the field. A clinical trial is scheduled to happen in the future to validate the benchmarking approach. Using the present technique on experimental devices that have already been tested in hospitals would also be an interesting endeavour.

# Conclusion

In this paper, we aimed to demonstrate how it is possible to test the discriminative power of an electronic nose using artificial cancer breath and give it a performance score that allows comparison with other devices. Our investigation involved several meticulous steps, from recruiting healthy volunteers for breath collection to the creation of artificial cancer breath samples through the addition of biomarkers at varying concentrations.

Over four month, 21 volunteers donated a total of 126 healthy breath samples. No significant effect from basic habits and biometrics were found on the results in the PCA.

E-noses have promising capabilities in distinguishing the breath of cancer patients from healthy patients or cancer from other conditions, as our literature review suggests. The e-nose developed for this paper was submitted to the benchmark and managed to obtain sufficient discrimination at higher concentration of biomarkers. While it managed to discriminate at realistic biomarkers concentrations, the resolution was not good enough for medical use, emphasizing the importance of sensitivity and diversity in sensors for achieving reliable detection. The complexity of the breath and the very low concentrations of cancer biomarkers reported in the literature, combined with the insufficient limit of detection of commercial MOS sensors resulted in the necessity to increase the concentrations of biomarkers 8 times before reaching the performance criterion. To enhance the e-nose's performance, we suggested potential strategies, such as concentrating biomarkers using sorption on Tenax and exploring more sensitive sensor options.

Additionally, we addressed the sensor drift over time using a calibration and linear correction method. Calibration and drift correction emerged as essential elements in ensuring the quality of our results.

Finally, we recognize the need for further investigations with larger sample sizes and real patient data. A comparison of the score obtained by a device with the performance of the same device in a real medical screening campaign context is desirable to confirm the relevance of this approach.

# Contributions to this paper

**Anne-Claude ROMAIN**: Supervision, Funding acquisition, Writing – review & editing, Resources, Project administration.

**Justin DM Martin:** Writing,Formal analysis, Conceptualization, Experimental operation, Investigation, Data curation.

**Claudia FALZONE:** Writing – review & editing

# Declaration of Competing Interest

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

# Data availability

Due to ethical considerations and GDPR requirements, supporting data is not available.

# Safety procedures

All researchers involved in breath sampling and analysis adhered to standard safety procedures, including the use of lab coats, disposable gloves, and safety goggles. Additionally, personnel were trained on the safe handling of chemicals and equipment. Sterile mouthpieces and filters were used to minimize the risk of contamination between participants. A single use spirometry filter was used for each sample, and the filter holder was cleaned and disinfected between each sampling. The collection and analysis of collected breath samples was conducted in a controlled laboratory environment with proper ventilation. All samples, once processed, were disposed of using a suction tube leading outside the building. Sampling bags were thoroughly cleaned between each sample by filling the bag with clean air then using an oven at 60°C for 30 minutes and repeating the process three times.

# Acknowledgements

This work was financially supported by Interreg France-Wallonie-Vlaanderen in the context of the Pathacov project. This work is also supported by the ULiège SPHERES research group.

We would like to thank Florence SCHLEICH for her assistance for the procedure with the Ethics Committee and general supervision of this study.

We would like to thank Simon-Pierre LIEGEOIS for his enthusiasm and help with data treatment.

We would also like to thank all participants to this study.

# References

[1] C. Feil, F. Staib, M.R. Berger, T. Stein, I. Schmidtmann, A. Forster, C.C. Schimanski, Sniffer dogs can identify lung cancer patients from breath and urine samples, BMC Cancer. 21 (2021) 917. https://doi.org/10.1186/s12885-021-08651-5.

[2] B. Piqueret, B. Bourachot, C. Leroy, P. Devienne, F. Mechta-Grigoriou, P. d’Ettorre, J.-C. Sandoz, Ants detect cancer cells through volatile organic compounds, iScience. 25 (2022). https://doi.org/10.1016/j.isci.2022.103959.

[3] J.D.M. Martin, A.-C. Romain, Building a Sensor Benchmark for E-Nose Based Lung Cancer Detection: Methodological Considerations, Chemosensors. 10 (2022) 444. https://doi.org/10.3390/chemosensors10110444.

[4] Y. Ru Zhao, X. Xie, H.J. de Koning, W.P. Mali, R. Vliegenthart, M. Oudkerk, NELSON lung cancer screening study, Cancer Imaging. 11 (2011) S79–S84. https://doi.org/10.1102/1470-7330.2011.9020.

[5] U. Pastorino, M. Silva, S. Sestini, F. Sabia, M. Boeri, A. Cantarutti, N. Sverzellati, G. Sozzi, G. Corrao, A. Marchianò, Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy, Ann Oncol. 30 (2019) 1162–1169. https://doi.org/10.1093/annonc/mdz117.

[6] N.P. Chudgar, P.R. Bucciarelli, E.M. Jeffries, N.P. Rizk, B.J. Park, P.S. Adusumilli, D.R. Jones, Results of the National Lung Cancer Screening Trial: Where Are We Now?, Thorac Surg Clin. 25 (2015) 145–153. https://doi.org/10.1016/j.thorsurg.2014.11.002.

[7] V.A. Binson, M. Subramoniam, Exhaled Breath Volatile Organic Compound Analysis for the Detection of Lung Cancer-A Systematic Review, Journal of Biomimetics, Biomaterials and Biomedical Engineering. 56 (2022) 17–35. https://doi.org/10.4028/p-dab04j.

[8] C. Hou, J. Lei, D. Huo, K. Song, J. Li, X. Luo, M. Yang, H. Fa, Discrimination of Lung Cancer Related Volatile Organic Compounds with a Colorimetric Sensor Array, Analytical Letters. 46 (2013) 2048–2059. https://doi.org/10.1080/00032719.2013.782550.

[9] K. Chen, L. Liu, B. Nie, B. Lu, L. Fu, Z. He, W. Li, X. Pi, H. Liu, Recognizing lung cancer and stages using a self-developed electronic nose system, Computers in Biology and Medicine. 131 (2021) 104294. https://doi.org/10.1016/j.compbiomed.2021.104294.

[10] J. Martin, A.-C. Romain, A Gas sensor array to Identify complex VOC mixtures in breath, in: 2021. https://orbi.uliege.be/handle/2268/256787 (accessed March 17, 2023).

[11] J.D.M. Martin, A.-C. Romain, Experimental Evaluation of Gas Sensors Array for the Identification of Complex Voc Mixtures in Human Breath, 1. 85 (2021) 199–204. https://doi.org/10.3303/CET2185034.

[12] J. Li, Y. Zhang, Q. Chen, Z. Pan, J. Chen, M. Sun, J. Wang, Y. Li, Q. Ye, Development and validation of a screening model for lung cancer using machine learning: A large-scale, multi-center study of biomarkers in breath, Frontiers in Oncology. 12 (2022). https://doi.org/10.3389/fonc.2022.975563.

[13] rbalia, Newsletter finale : Résultats du projet et retour sur l’évènement de clôture !, PATHACOV. (2023). https://pathacov-project.com/newsletter-finale-resultats-du-projet-et-retour-sur-levenement-de-cloture/ (accessed March 17, 2023).

[14] S. Kort, M. Brusse-Keizer, H. Schouwink, E. Citgez, F.H. de Jongh, J.W.G. van Putten, B. van den Borne, E.A. Kastelijn, D. Stolz, M. Schuurbiers, M.M. van den Heuvel, W.H. van Geffen, J. van der Palen, Diagnosing Non-Small Cell Lung Cancer by Exhaled Breath Profiling Using an Electronic Nose: A Multicenter Validation Study, Chest. 163 (2023) 697–706. https://doi.org/10.1016/j.chest.2022.09.042.

[15] M.K. Nakhleh, H. Amal, R. Jeries, Y.Y. Broza, M. Aboud, A. Gharra, H. Ivgi, S. Khatib, S. Badarneh, L. Har-Shai, L. Glass-Marmor, I. Lejbkowicz, A. Miller, S. Badarny, R. Winer, J. Finberg, S. Cohen-Kaminsky, F. Perros, D. Montani, B. Girerd, G. Garcia, G. Simonneau, F. Nakhoul, S. Baram, R. Salim, M. Hakim, M. Gruber, O. Ronen, T. Marshak, I. Doweck, O. Nativ, Z. Bahouth, D. Shi, W. Zhang, Q. Hua, Y. Pan, L. Tao, H. Liu, A. Karban, E. Koifman, T. Rainis, R. Skapars, A. Sivins, G. Ancans, I. Liepniece-Karele, I. Kikuste, I. Lasina, I. Tolmanis, D. Johnson, S.Z. Millstone, J. Fulton, J.W. Wells, L.H. Wilf, M. Humbert, M. Leja, N. Peled, H. Haick, Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules, ACS Nano. 11 (2017) 112–125. https://doi.org/10.1021/acsnano.6b04930.

[16] Q. Chen, Z. Chen, D. Liu, Z. He, J. Wu, Constructing an E-Nose Using Metal-Ion-Induced Assembly of Graphene Oxide for Diagnosis of Lung Cancer via Exhaled Breath, ACS Appl. Mater. Interfaces. 12 (2020) 17713–17724. https://doi.org/10.1021/acsami.0c00720.

[17] C. Di Natale, A. Macagnano, E. Martinelli, R. Paolesse, G. D’Arcangelo, C. Roscioni, A. Finazzi-Agrò, A. D’Amico, Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors, Biosensors and Bioelectronics. 18 (2003) 1209–1218. https://doi.org/10.1016/S0956-5663(03)00086-1.

[18] X. Chen, M. Cao, Y. Li, W. Hu, P. Wang, K. Ying, H. Pan, A study of an electronic nose for detection of lung cancer based on a virtual SAW gas sensors array and imaging recognition method, Meas. Sci. Technol. 16 (2005) 1535–1546. https://doi.org/10.1088/0957-0233/16/8/001.

[19] R.F. Machado, D. Laskowski, O. Deffenderfer, T. Burch, S. Zheng, P.J. Mazzone, T. Mekhail, C. Jennings, J.K. Stoller, J. Pyle, J. Duncan, R.A. Dweik, S.C. Erzurum, Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath, Am J Respir Crit Care Med. 171 (2005) 1286–1291. https://doi.org/10.1164/rccm.200409-1184OC.

[20] R. Blatt, A. Bonarini, E. Calabró, M. Della Torre, M. Matteucci, U. Pastorino, Fuzzy k-NN Lung Cancer Identification by an Electronic Nose, in: F. Masulli, S. Mitra, G. Pasi (Eds.), Applications of Fuzzy Sets Theory, Springer, Berlin, Heidelberg, 2007: pp. 261–268. https://doi.org/10.1007/978-3-540-73400-0\_32.

[21] S. Dragonieri, J.T. Annema, R. Schot, M.P.C. van der Schee, A. Spanevello, P. Carratú, O. Resta, K.F. Rabe, P.J. Sterk, An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD, Lung Cancer. 64 (2009) 166–170. https://doi.org/10.1016/j.lungcan.2008.08.008.

[22] A. D’Amico, G. Pennazza, M. Santonico, E. Martinelli, C. Roscioni, G. Galluccio, R. Paolesse, C. Di Natale, An investigation on electronic nose diagnosis of lung cancer, Lung Cancer. 68 (2010) 170–176. https://doi.org/10.1016/j.lungcan.2009.11.003.

[23] E.A. Chapman, P.S. Thomas, E. Stone, C. Lewis, D.H. Yates, A breath test for malignant mesothelioma using an electronic nose, European Respiratory Journal. 40 (2012) 448–454. https://doi.org/10.1183/09031936.00040911.

[24] N. Peled, M. Hakim, P.A. Bunn, Y.E. Miller, T.C. Kennedy, J. Mattei, J.D. Mitchell, F.R. Hirsch, H. Haick, Non-invasive Breath Analysis of Pulmonary Nodules, Journal of Thoracic Oncology. 7 (2012) 1528–1533. https://doi.org/10.1097/JTO.0b013e3182637d5f.

[25] D. Wang, K. Yu, Y. Wang, Y. Hu, C. Zhao, L. Wang, K. Ying, P. Wang, A hybrid electronic noses’ system based on mos-saw detection units intended for lung cancer diagnosis, J. Innov. Opt. Health Sci. 05 (2012) 1150006. https://doi.org/10.1142/S1793545811500064.

[26] Y.Y. Broza, H. Haick, Nanomaterial-based sensors for detection of disease by volatile organic compounds, Https://Doi.Org/10.2217/Nnm.13.64. (2013). https://doi.org/10.2217/nnm.13.64.

[27] A. Bikov, M. Hernadi, B.Z. Korosi, G. Zsamboki, Z. Sutto, A. Tarnoki, D. Tarnoki, G. Losonczy, I. Horvath, Expiratory Flow Rate, Breath Hold And Anatomic Dead Space Influence Electronic Nose Ability To Detect Lung Cancer, in: A102. A VIEW TO THE LUNG: LUNG CANCER SCREENING AND THE EVALUATION OF PULMONARY NODULES, American Thoracic Society, 2014: pp. A2262–A2262. https://doi.org/10.1164/ajrccm-conference.2014.189.1\_MeetingAbstracts.A2262.

[28] R. de Vries, P. Brinkman, M.P. van der Schee, N. Fens, E. Dijkers, S.K. Bootsma, F.H.C. de Jongh, P.J. Sterk, Integration of electronic nose technology with spirometry: validation of a new approach for exhaled breath analysis, J. Breath Res. 9 (2015) 046001. https://doi.org/10.1088/1752-7155/9/4/046001.

[29] A. McWilliams, P. Beigi, A. Srinidhi, S. Lam, C.E. MacAulay, Sex and Smoking Status Effects on the Early Detection of Early Lung Cancer in High-Risk Smokers Using an Electronic Nose, IEEE Transactions on Biomedical Engineering. 62 (2015) 2044–2054. https://doi.org/10.1109/TBME.2015.2409092.

[30] R. Gasparri, M. Santonico, C. Valentini, G. Sedda, A. Borri, F. Petrella, P. Maisonneuve, G. Pennazza, A. D’Amico, C. Di Natale, R. Paolesse, L. Spaggiari, Volatile signature for the early diagnosis of lung cancer, J. Breath Res. 10 (2016) 016007. https://doi.org/10.1088/1752-7155/10/1/016007.

[31] R. Rocco, R.A. Incalzi, G. Pennazza, M. Santonico, C. Pedone, I.R. Bartoli, C. Vernile, G. Mangiameli, A. La Rocca, G. De Luca, G. Rocco, P. Crucitti, BIONOTE e-nose technology may reduce false positives in lung cancer screening programmes, Eur J Cardiothorac Surg. 49 (2016) 1112–1117. https://doi.org/10.1093/ejcts/ezv328.

[32] X. Cai, L. Chen, T. Kang, Y. Tang, T. Lim, M. Xu, H. Hui, A Prediction Model with a Combination of Variables for Diagnosis of Lung Cancer, Med Sci Monit. 23 (2017) 5620–5629. https://doi.org/10.12659/MSM.904738.

[33] W. Li, H. Liu, D. Xie, Z. He, X. Pi, Lung Cancer Screening Based on Type-different Sensor Arrays, Sci Rep. 7 (2017) 1969. https://doi.org/10.1038/s41598-017-02154-9.

[34] D. Shlomi, M. Abud, O. Liran, J. Bar, N. Gai-Mor, M. Ilouze, A. Onn, A. Ben-Nun, H. Haick, N. Peled, Detection of Lung Cancer and EGFR Mutation by Electronic Nose System, J Thorac Oncol. 12 (2017) 1544–1551. https://doi.org/10.1016/j.jtho.2017.06.073.

[35] M. Tirzīte, M. Bukovskis, G. Strazda, N. Jurka, I. Taivans, Detection of lung cancer in exhaled breath with an electronic nose using support vector machine analysis, J Breath Res. 11 (2017) 036009. https://doi.org/10.1088/1752-7163/aa7799.

[36] C.-H. Huang, C. Zeng, Y.-C. Wang, H.-Y. Peng, C.-S. Lin, C.-J. Chang, H.-Y. Yang, A Study of Diagnostic Accuracy Using a Chemical Sensor Array and a Machine Learning Technique to Detect Lung Cancer, Sensors (Basel). 18 (2018). https://doi.org/10.3390/s18092845.

[37] S. Kort, M.M. Tiggeloven, M. Brusse-Keizer, J.W. Gerritsen, J.H. Schouwink, E. Citgez, F.H.C. de Jongh, S. Samii, J. van der Maten, M. van den Bogart, J. van der Palen, Multi-centre prospective study on diagnosing subtypes of lung cancer by exhaled-breath analysis, Lung Cancer. 125 (2018) 223–229. https://doi.org/10.1016/j.lungcan.2018.09.022.

[38] S. Kort, M. Brusse-Keizer, J.-W. Gerritsen, J. van der Palen, Data analysis of electronic nose technology in lung cancer: generating prediction models by means of Aethena, J. Breath Res. 11 (2017) 026006. https://doi.org/10.1088/1752-7163/aa6b08.

[39] S. Kort, M. Brusse-Keizer, H. Schouwink, E. Citgez, F.D. Jongh, J.W. Gerritsen, J.V.D. Palen, Combining exhaled-breath analysis data with clinical parameters to improve the diagnosis of lung cancer, European Respiratory Journal. 54 (2019). https://doi.org/10.1183/13993003.congress-2019.PA3030.

[40] R. van de Goor, M. van Hooren, A.-M. Dingemans, B. Kremer, K. Kross, Training and Validating a Portable Electronic Nose for Lung Cancer Screening, Journal of Thoracic Oncology. 13 (2018) 676–681. https://doi.org/10.1016/j.jtho.2018.01.024.

[41] A. Kononov, B. Korotetsky, I. Jahatspanian, A. Gubal, A. Vasiliev, A. Arseniev, A. Nefedov, A. Barchuk, I. Gorbunov, K. Kozyrev, A. Rassadina, E. Iakovleva, M. Sillanpää, Z. Safaei, N. Ivanenko, N. Stolyarova, V. Chuchina, A. Ganeev, Online breath analysis using metal oxide semiconductor sensors (Electronic Nose) for diagnosis of lung cancer, J. Breath Res. (2019). https://doi.org/10.1088/1752-7163/ab433d.

[42] B. Lu, L. Fu, B. Nie, Z. Peng, H. Liu, A Novel Framework with High Diagnostic Sensitivity for Lung Cancer Detection by Electronic Nose, Sensors (Basel). 19 (2019). https://doi.org/10.3390/s19235333.

[43] D. Marzorati, L. Mainardi, G. Sedda, R. Gasparri, L. Spaggiari, P. Cerveri, A Metal Oxide Gas Sensors Array for Lung Cancer Diagnosis Through Exhaled Breath Analysis, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019: pp. 1584–1587. https://doi.org/10.1109/EMBC.2019.8856750.

[44] M. Tirzïte, M. Bukovskis, G. Strazda, N. Jurka, I. Taivans, Detection of lung cancer with electronic nose and logistic regression analysis, J. Breath Res. 13 (2018) 016006. https://doi.org/10.1088/1752-7163/aae1b8.

[45] T. Saidi, M. Moufid, K. de Jesus Beleño-Saenz, T.G. Welearegay, N. El Bari, A. Lisset Jaimes-Mogollon, R. Ionescu, J.E. Bourkadi, J. Benamor, M. El Ftouh, B. Bouchikhi, Non-invasive prediction of lung cancer histological types through exhaled breath analysis by UV-irradiated electronic nose and GC/QTOF/MS, Sensors and Actuators B: Chemical. 311 (2020) 127932. https://doi.org/10.1016/j.snb.2020.127932.

[46] M. Rodríguez-Aguilar, L. Díaz de León-Martínez, P. Gorocica-Rosete, R. Pérez-Padilla, C.A. Domínguez-Reyes, J.A. Tenorio-Torres, O. Ornelas-Rebolledo, G. Mehta, B.N. Zamora-Mendoza, R. Flores-Ramírez, Application of chemoresistive gas sensors and chemometric analysis to differentiate the fingerprints of global volatile organic compounds from diseases. Preliminary results of COPD, lung cancer and breast cancer, Clinica Chimica Acta; International Journal of Clinical Chemistry. 518 (2021) 83–92. https://doi.org/10.1016/j.cca.2021.03.016.

[47] V.A. Binson, M. Subramoniam, L. Mathew, Discrimination of COPD and lung cancer from controls through breath analysis using a self-developed e-nose, J. Breath Res. (2021). https://doi.org/10.1088/1752-7163/ac1326.

[48] V.A. Binson, R. Akbar, N. Thankachan, S. Thomas, Design and construction of a portable e-nose system for human exhaled breath VOC analysis, Materials Today: Proceedings. 58 (2022) 422–427. https://doi.org/10.1016/j.matpr.2022.02.388.

[49] V.A. Binson, M. Subramoniam, Design and development of an e-nose system for the diagnosis of pulmonary diseases, Acta Bioeng Biomech. 23 (2021). https://doi.org/10.37190/ABB-01737-2020-03.

[50] R. de Vries, N. Farzan, T. Fabius, F.H.C. De Jongh, P.M.C. Jak, E.G. Haarman, E. Snoey, J.C.C.M. In ’T Veen, Y.W.F. Dagelet, A.-H. Maitland-Van Der Zee, A. Lucas, M.M. Van Den Heuvel, M. Wolf-Lansdorf, M. Muller, P. Baas, P.J. Sterk, Prospective Detection of Early Lung Cancer in Patients With COPD in Regular Care by Electronic Nose Analysis of Exhaled Breath, Chest. (2023). https://doi.org/10.1016/j.chest.2023.04.050.

[51] L. Hao, G. Huang, An improved AdaBoost algorithm for identification of lung cancer based on electronic nose, Heliyon. 9 (2023). https://doi.org/10.1016/j.heliyon.2023.e13633.

[52] M. Sitarz, Extending F1 metric, probabilistic approach, (2022). https://doi.org/10.48550/arXiv.2210.11997.

[53] A.K. Pathak, K. Swargiary, N. Kongsawang, P. Jitpratak, N. Ajchareeyasoontorn, J. Udomkittivorakul, C. Viphavakit, Recent Advances in Sensing Materials Targeting Clinical Volatile Organic Compound (VOC) Biomarkers: A Review, Biosensors. 13 (2023) 114. https://doi.org/10.3390/bios13010114.

[54] K.J. Johnson, S.L. Rose-Pehrsson, Sensor Array Design for Complex Sensing Tasks, Annual Rev. Anal. Chem. 8 (2015) 287–310. https://doi.org/10.1146/annurev-anchem-062011-143205.

[55] W. Li, H. Liu, D. Xie, Z. He, X. Pi, Lung Cancer Screening Based on Type-different Sensor Arrays, Sci Rep. 7 (2017) 1969. https://doi.org/10.1038/s41598-017-02154-9.

[56] A. McWilliams, P. Beigi, A. Srinidhi, S. Lam, C.E. MacAulay, Sex and Smoking Status Effects on the Early Detection of Early Lung Cancer in High-Risk Smokers Using an Electronic Nose, IEEE Transactions on Biomedical Engineering. 62 (2015) 2044–2054. https://doi.org/10.1109/TBME.2015.2409092.

[57] R.F. Machado, D. Laskowski, O. Deffenderfer, T. Burch, S. Zheng, P.J. Mazzone, T. Mekhail, C. Jennings, J.K. Stoller, J. Pyle, J. Duncan, R.A. Dweik, S.C. Erzurum, Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath, Am J Respir Crit Care Med. 171 (2005) 1286–1291. https://doi.org/10.1164/rccm.200409-1184OC.

[58] D. Smith, P. Spanel, S. Davies, Trace gases in breath of healthy volunteers when fasting and after a protein-calorie meal: a preliminary study, J Appl Physiol (1985). 87 (1999) 1584–1588. https://doi.org/10.1152/jappl.1999.87.5.1584.

1. \* Corresponding author at: University of Liège, Arlon Campus, 6700 Arlon, Belgium

   *E-mail address:* [jdm.martin@uliege.be](mailto:zhouyying@swu.edu.cn) (J. Martin).

   [DOI](https://doi.org/10.1016/j.snb.2023.133518)

   Received [date] ; Received in revised form [date]; Accepted [date]

   Available online [date]

   xxxx-xxxx/© [year] Elsevier B.V. All rights reserved. [↑](#footnote-ref-1)