

PAPER • OPEN ACCESS

Detection of lung cancer with electronic nose and logistic regression analysis

To cite this article: Madara Tirzite *et al* 2019 *J. Breath Res.* **13** 016006

View the [article online](#) for updates and enhancements.

Recent citations

- [Recognition of breathprints of lung cancer and chronic obstructive pulmonary disease using the Aeonose[®] electronic nose](#)
Ekaterina Krauss *et al*
- [A review of exhaled breath: a key role in lung cancer diagnosis](#)
Davide Marzorati *et al*
- [Electronic nose: a non-invasive technology for breath analysis of diabetes and lung cancer patients](#)
Bhagaban Behera *et al*



BREATH BIOPSY

Breath Biopsy Panel for Focused Biomarker Discovery in Respiratory Disease Research

Providing high confidence identification of non-invasive breath biomarkers to distinguish, monitor and assess therapeutic responses across a range of chronic inflammatory airway diseases

[WATCH OUR INTRODUCTORY WEBINAR](#)





PAPER

Detection of lung cancer with electronic nose and logistic regression analysis

OPEN ACCESS

RECEIVED

17 January 2018

REVISED

9 September 2018

ACCEPTED FOR PUBLICATION

17 September 2018

PUBLISHED

20 November 2018

Madara Tirzīte , Māris Bukovskis, Gunta Strazda, Normunds Jurka and Immanuels Taivans

Faculty of Medicine, University of Latvia, Raina Blv. 19, Riga, Latvia

E-mail: madara.tirzite@aslimnica.lv**Keywords:** electronic nose, e-nose, electronic smell sensor, artificial olfactory sensor, logistic regression analysis, lung cancer, volatile organic compounds

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Abstract**

Lung cancer is a very common malignancy with a low five-year survival rate. Artificial olfactory sensor (electronic nose) is a tool that recently has been studied as a probable optimal screening tool for early detection of lung cancer, but still no statistical method has been put forward as the preferable one. The aim of the study was to explore the use of logistic regression analysis (LRA) to analyse patients' exhaled breath samples with electronic nose in order to differentiate lung cancer patients (regardless of the stage of the cancer) from patients with other lung diseases and healthy individuals. Patients with histologically or cytologically verified, untreated lung cancer, patients with other lung diseases such as benign lung tumors, chronic obstructive pulmonary disease, asthma, pneumonia, etc, and healthy volunteers were enrolled in the study, in total 252 cancer patients and 223 patients without cancer. Breath sample collection and analysis were performed with Cyranose 320 sensor device and data further analysed using LRA. The LRA correctly differentiated lung cancer patients from no-cancer patients. The overall sensitivity in detecting patients having cancer was 95.8% for smokers and 96.2% for non-smokers and the overall specificity was 90.6% for non-smokers and 92.3% for smokers. Exhaled breath analysis by electronic nose using LRA is able to discriminate lung cancer patients from patients with other lung diseases and from healthy individuals.

Introduction

Lung cancer is one of the leading malignant diseases in the world. According to the World Cancer Report 2014, the data from 2012 reveal that there were 1.8 million lung cancer cases detected worldwide (13% of all cases of cancer) [1]. Unfortunately, the mortality from lung cancer is high and the overall five-year survival rate is low, only 17% [1, 2]. These numbers arise from the fact that in its early stages lung cancer usually evolves with no significant symptoms as well as with no suspected radiological changes. Thus lung cancer is mostly detected in already advanced stages and consequently linked with higher lethality. The survival rate for lung cancer increases if the cancer is diagnosed at an early stage. Only ~15% of lung cancer patients are diagnosed at an early stage and more than half of the lung cancer patients have already died within a year from the detection of lung cancer [2].

None of the available diagnostic methods—nor computed tomography/positron emission tomography, nor fibrobronchoscopy with biopsy or sputum cytology—have been accredited as a reliable screening method, since each has its disadvantages. There is a lack of simple, cheap and widely available tool for early diagnostics of lung cancer, and it is essential to search for it. One such potential tool is the electronic smell sensor or electronic nose (e-nose). It has been tested in various specialities and divisions of medicine, not only the respiratory medicine, for more than two decades. Several data analysis methods have been applied to study the data gained with electronic nose. The respiratory medicine e-nose, combined with various statistical analysis methods, has been mostly used in studies exploring differentiation between chronic obstructive pulmonary disease (COPD), asthma, pneumonia and lung cancer. The results obtained have been promising. Still no statistical data analysis method has been found to be superior over

others, yet the results may vary significantly depending on the method chosen. In a retrospective study that reviewed 73 studies on the analysis of volatile organic compounds (VOC) numerous different data statistical analysis methods were identified [3], while in the recently published European Respiratory Society Task Force publication on biomarkers in lung disease no current recommendations are given regarding the statistical analysis of VOC [4]. Regarding the minimum reporting standards for data analysis in metabolomics, there is no preferred algorithm and the preferred method in each case should depend on the problem [5]. The ERS Task Force recommends to use principal components analysis as a clustering method for exploratory data analysis [4]. Due to this ambiguity, we chose for our study to use the logistic regression analysis (LRA) and to explore the possibility of using LRA as a statistical tool for differentiation of the lung cancer patients from patients with other lung diseases and from healthy individuals, through analysing the data obtained from patients' exhaled breath with the artificial olfactory sensor (electronic nose).

Methods

Study patients

The patients and healthy volunteers were enrolled in a study at the Department of Lung Diseases and Centre of Thoracic Surgery at Pauls Stradins Clinical University Hospital in Riga, Latvia, from April 2011 to September 2013. Approval from the Ethical committee of scientific research at the University of Latvia, Institute of Experimental and clinical medicine, was received. All patients signed an informed consent.

The patients were divided in 'cancer' and 'no-cancer' groups and then further in 'smokers' and 'non-smokers' groups (see below). In the 'cancer' group we included patients with verified lung cancer. In the 'no-cancer' group we included patients with other lung diseases, as well as healthy volunteers.

For the 'cancer' group patients the clinical diagnosis of lung cancer was specified [6]. Patients with complications of lung cancer, like post-obstructive atelectasis, pneumonia, carcinomatous lymphangitis, destruction of the tumor mass, etc were also included [7]. All lung cancer patients were newly diagnosed and had not received any prior specific anti-cancer therapy before the breath sampling.

In the 'no-cancer' group we included both healthy volunteers and:

- patients with known COPD, diagnosed according to the Global Initiative of Chronic Obstructive Lung Disease (GOLD) Report, updated 2011,
- patients with bronchial asthma, diagnosed according to the Global Initiative for Asthma (GINA) Report, updated 2011,

- patients with pneumonia, diagnosed according to the British Thoracic Society Guidelines for the Management of Community Acquired Pneumonia in Adults, updated 2009,
- patients with pulmonary embolism (PE), diagnosed according to the Guidelines on the diagnosis and management of acute PE of the European Society of Cardiology, updated 2008,
- patients with bronchiectasis, diagnosed according to the British Thoracic Society Guidelines for non-CF bronchiectasis, updated 2010,
- patients with tuberculosis, diagnosed according to the WHO Guidelines for diagnosis of tuberculosis, updated 2011,
- patients with histologically or cytologically verified benign lung tumors [7].

Exclusion criteria (for all patient groups):

- patients who were unable to perform the manoeuvres necessary to gain the breath sample;
- patients with uncertain anamnesis regarding possible lung diseases;
- no clear diagnosis regardless of thorough investigations [7].

475 individuals were included in the study in total, 336 males and 139 females. We then divided them in two further groups (after classifying them as 'cancer' or 'no-cancer'). The first group ('non-smokers') included non-smokers and ex-smokers, 265 individuals in total. Ex-smokers had to have ceased smoking at least a year ago. In the second group ('smokers') we included only smokers, 210 individuals in total. The first, 'non-smokers' group, consisted of 133 patients with verified lung cancer (from the 'cancer' group) and 132 patients from the 'no-cancer' group. The second group, 'smokers', consisted of 119 patients from the 'cancer' group and 91 individuals from the 'no-cancer' group.

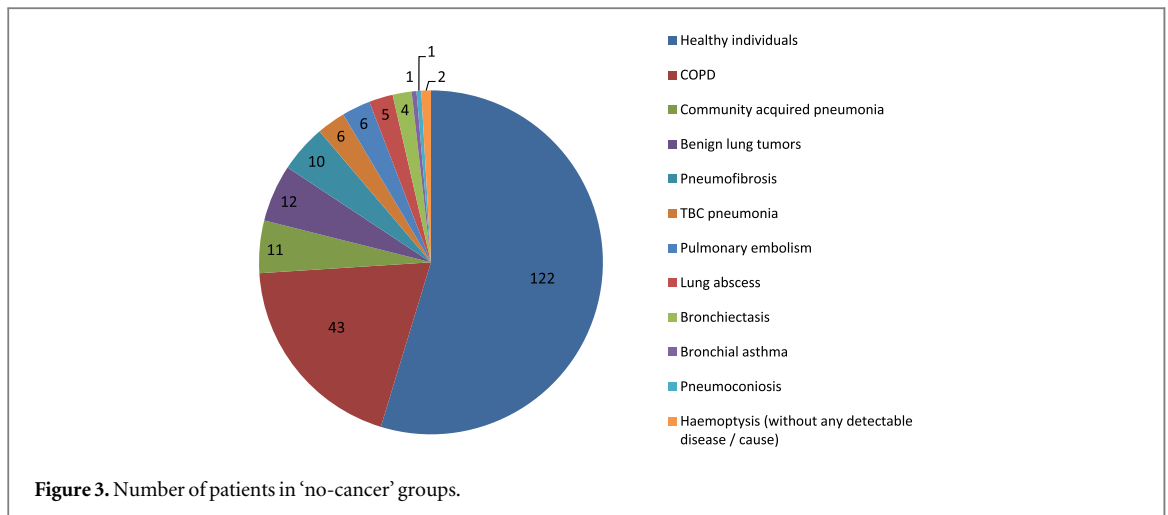
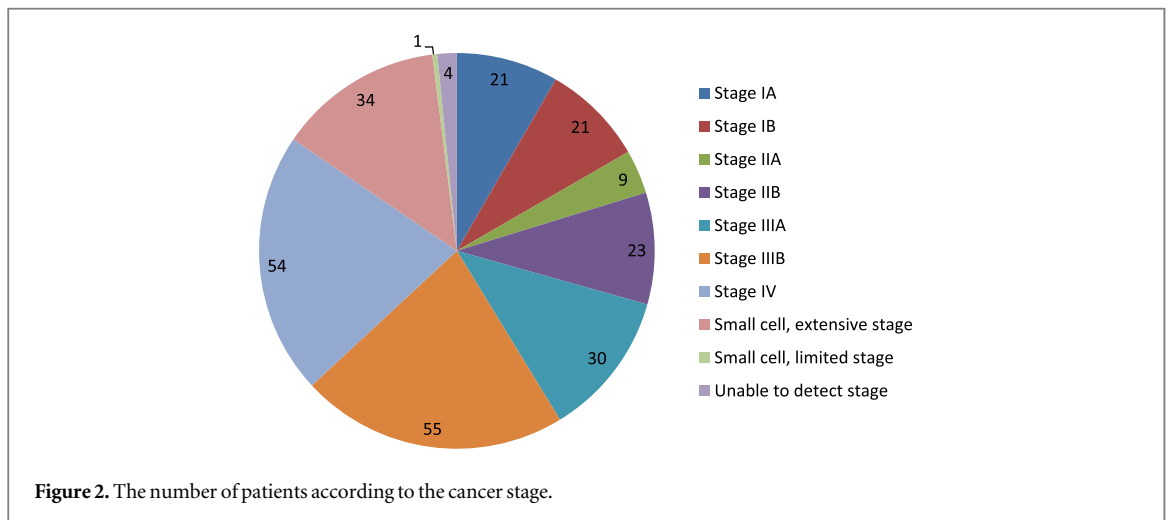
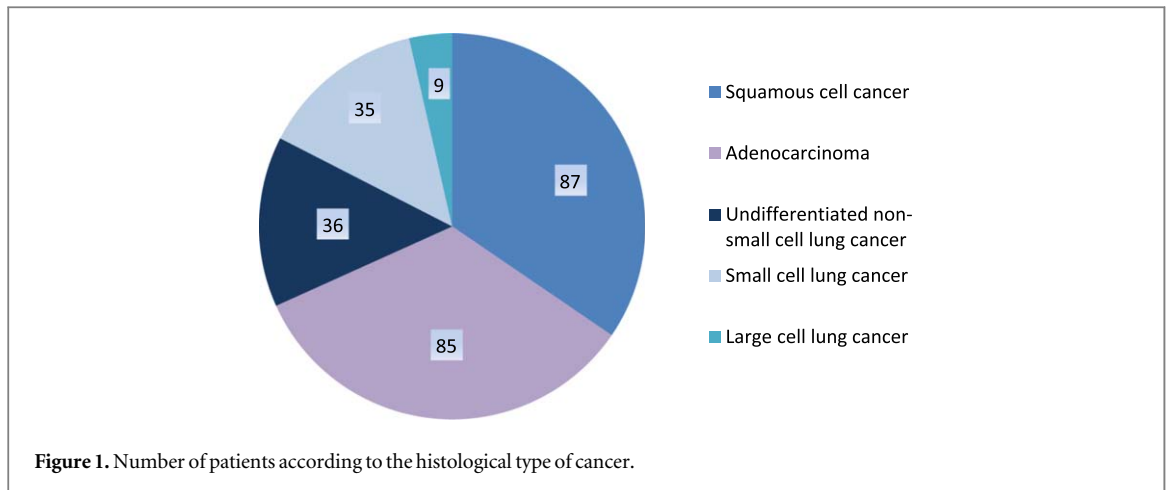
The division of lung cancer patients according to the histological forms of the cancer type and according to the cancer stage is depicted below in figures 1 and 2.

The division of no-cancer individuals into disease groups is represented in figure 3.

Study subjects had to fill in a questionnaire, which contained questions about demographics, concomitant diseases and smoking history. After filling in the questionnaire, a breath sample was collected [7]. The study design is depicted in figure 4.

Exhaled breath sampling and analysis

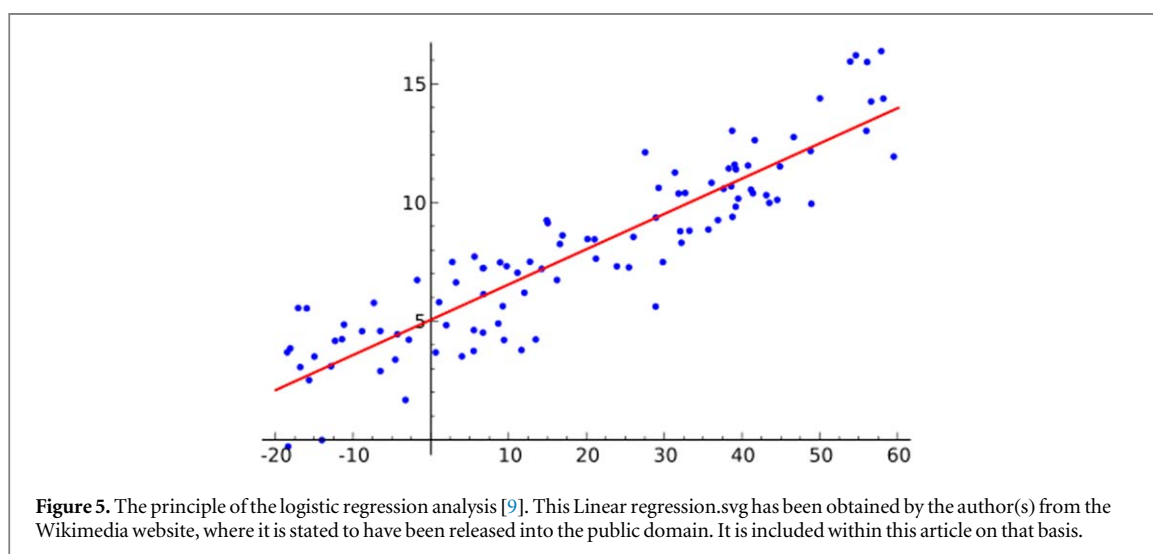
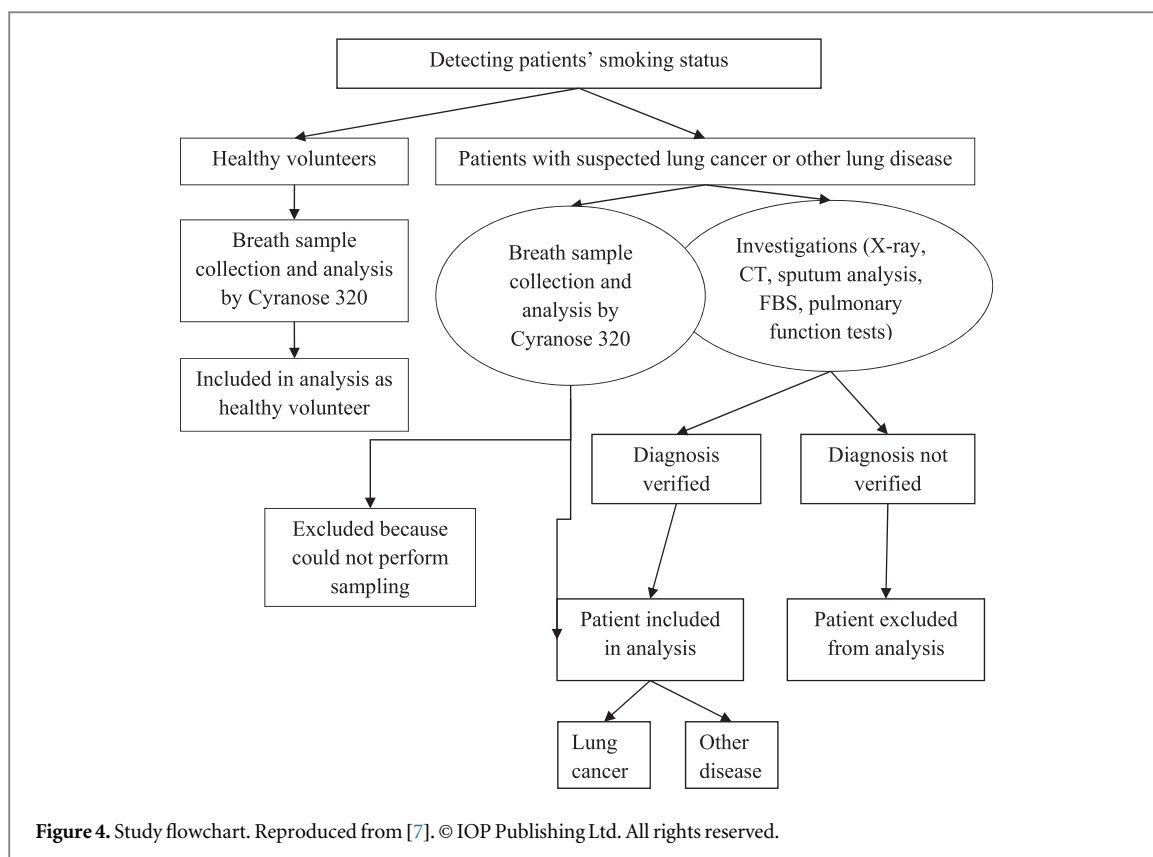
Sampling and analysis of exhaled breath were done according to a standardized method published by Dragonieri, with some modifications [8].



Initially, patients breathed tidally activated carbon filtered (*Nordic Safety, Norway*) air with clipped nose, through a T-shaped two-way non-rebreathing valve (*Hans Rudolph Inc., USA*) for 5 min to clean the exhaled breath from ambient air pollution. The second step was inhalation to total lung capacity and full exhalation into polyethylene terephthalate bag. This was followed by immediate analysis with the artificial olfactory sensor device [7]. The approximate flow rate

of the expiratory flow was 250–500 ml sec⁻¹, we did not separate the dead space sample. Later the analysis model in LRA was designed specifically for such sample acquisition model.

Exhaled breath analysis was done within 5 min after its collection with the artificial olfactory sensor device *Cyranose 320 (Smith's Detection, USA)*. The cycle of exhaled breath sample analysis consisted of a 20 s long period of baseline ambient air registration, a



60 s long period of the analysis of the sample, a 5 s long interim period, when air sample was disconnected from the olfactory sensor, and a 180 s long rinsing cycle. During the analysis 32 sensor curves of electrical resistance were registered [7].

Statistical analysis of the data

The data derived were statistically analysed using LRA (*Statistica 7.0*). As continuous predictors, we chose the relative maximum (R_{max}), area under the curve (AUC_{0-60}) and $tg\alpha_{0-60}$ for each curve of 32 sensors. Age, smoking status (smoker, ex-smoker, non-smoker),

smoking history and ambient temperature (t °C) at the moment of taking the air sample were considered additional predictor factors.

The LRA calculates the probability of each outcome falling into one or other group, and tries to maximize the likelihood and utility of the decision. It takes into account the constant and changing variables (figure 5) and the key is that the changing variables can only be expressed in one of the two modes (e.g., black/white) and the model predicts the probability of the data falling into their respective groups [9].

LRA is about maximizing the probability of the data to be a part of each group. The method can be

combined with additional methods, such as Kernel or Bayesian models [10].

When analysing the results, we calculated: sensitivity (number of true positives/(number of true positives + number of false negatives)), specificity (number of true negatives/(number of true negatives + number of false positives)), PPV—positive predicted value (number of true positives/(number of true positives + number of false positives)), NPV—negative predicted value (number of true negatives/(number of true negatives + number of false negatives)) [7, 11–14]

Data models for LRA

We developed two data models. The **first model** included and analysed only ‘non-smokers’—‘cancer’ individuals versus ‘no-cancer’ individuals. The **second model** included and analysed only smokers—‘cancer’ individuals versus ‘no-cancer’ individuals. The selected characteristic values of the detectors for the two models of lung cancer identification by LRA are schematically depicted in figures 6 and 7.

Results

The total number of ‘cancer’ patients was 252 and the total number of ‘no-cancer’ patients was 223.

In the first model, the mean age of ‘cancer’ patients was 68.73 ± 2.51 years, 46 (34.6%) were female, 87 (65.4%) were male, the mean age of ‘no-cancer’ patients was 53.84 ± 2.50 years, 65 (49.2%) were female, 67 (50.8%) were male.

In the second model, the mean age of ‘cancer’ patients was 62.07 ± 2.63 years, 7 (5.9%) were female, 112 (94.1%) were male. The mean age of ‘no-cancer’ patients was 49.71 ± 3.02 years, 15 (16.5%) were female, 76 (83.5%) were male.

First model

128 of 133 cancer patients were diagnosed correctly, sensitivity 96.2% (table 1).

Second model

114 out of 119 cancer patients were diagnosed correctly, sensitivity 95.8% (table 2).

In the test sample of the next 100 investigated patients (after those whose results we analysed and showed above), 82 cases were predicted correctly.

For descriptive statistics of comparison between ‘cancer’ and ‘no-cancer’ patients regarding smoking exposure in pack-years, see tables 3 and 4.

Mann–Whitney U test showed a significant difference between the ‘cancer’ and ‘no-cancer’ groups regarding their smoking exposure in pack-years ($p < 0.0001$) (table 5), while t -test found no significant difference in patients’ height and weight (table 5).

Discussion

We would like to assume that we have gained a very good sensitivity (95.8% in ‘smoker’ and 96.2% in ‘non-smoker’ groups) and specificity (92.3% in ‘smoker’ and 90.6% in ‘non-smoker’ group) markers. Both the overall sensitivity and specificity of the LRA were slightly impacted by the patients being ‘non-smokers’ (non-smokers and ex-smokers) or smokers. Both markers in both groups as well as PPV and NPV were all above 90%. It means that by using LRA, we could correctly classify patients as having or not having lung cancer with more than 90% precision.

Regarding the studies in respiratory medicine where LRA was used to analyse data obtained with electronic nose, we have to admit that there are only a few studies where LRA was used at all, not even limiting our search to studies devoted to the detection of lung cancer with e-nose and LRA. Thaler with colleagues used electronic nose and analysed patients’ samples trying to distinguish biofilm-producing from non-biofilm-producing *Pseudomonas aeruginosa* and *Staphylococcus aureus* strains (two species each). Binary classification of the bacteria (biofilm producing versus non-biofilm producing) by logistic regression was performed with various sets of data (e.g., taking into account the days of illness). The testing accuracy varied between 80.6%–100% for both *Pseudomonas aeruginosa* species and 72.2%–91.7% for both *Staphylococcus* species [15].

The use of LRA in probability of possible airway bacterial colonization of patients with COPD was studied by Sibila and colleagues. The air over cultures of specimens of clinically stable COPD patients and healthy controls was sampled using electronic nose and afterwards analysed with LRA. The accuracy was 88% for colonized and 83% for non-colonized patients [16].

The use of electronic nose together with logistic regression method analysis in detecting head and neck squamous cell carcinoma was studied by Leunis and colleagues. The conducted study was rather small—23 patients. The sensitivity gained with the chosen method was 90%, specificity—80% [17].

A different type of electronic nose sensor technology was used in another study by Thaler and colleagues. They sought to find a reliable difference between groups of patients with or without chronic bacterial sinusitis. Just as we had in our study, usually the metal oxide sensor e-nose is used in medical studies, but in this study the colorimetric sensor arrays were chosen. Nevertheless, the data obtained were later analysed with logistic regression method and the accurate classification rate found was 90% [18].

Schnabel *et al* used LRA to try to detect ventilator-associated pneumonia and analysed samples from 72 patients. The sensitivity was 88% with a specificity of 66% when it was used to distinguish VAP patients from the control group, but the values fell to a

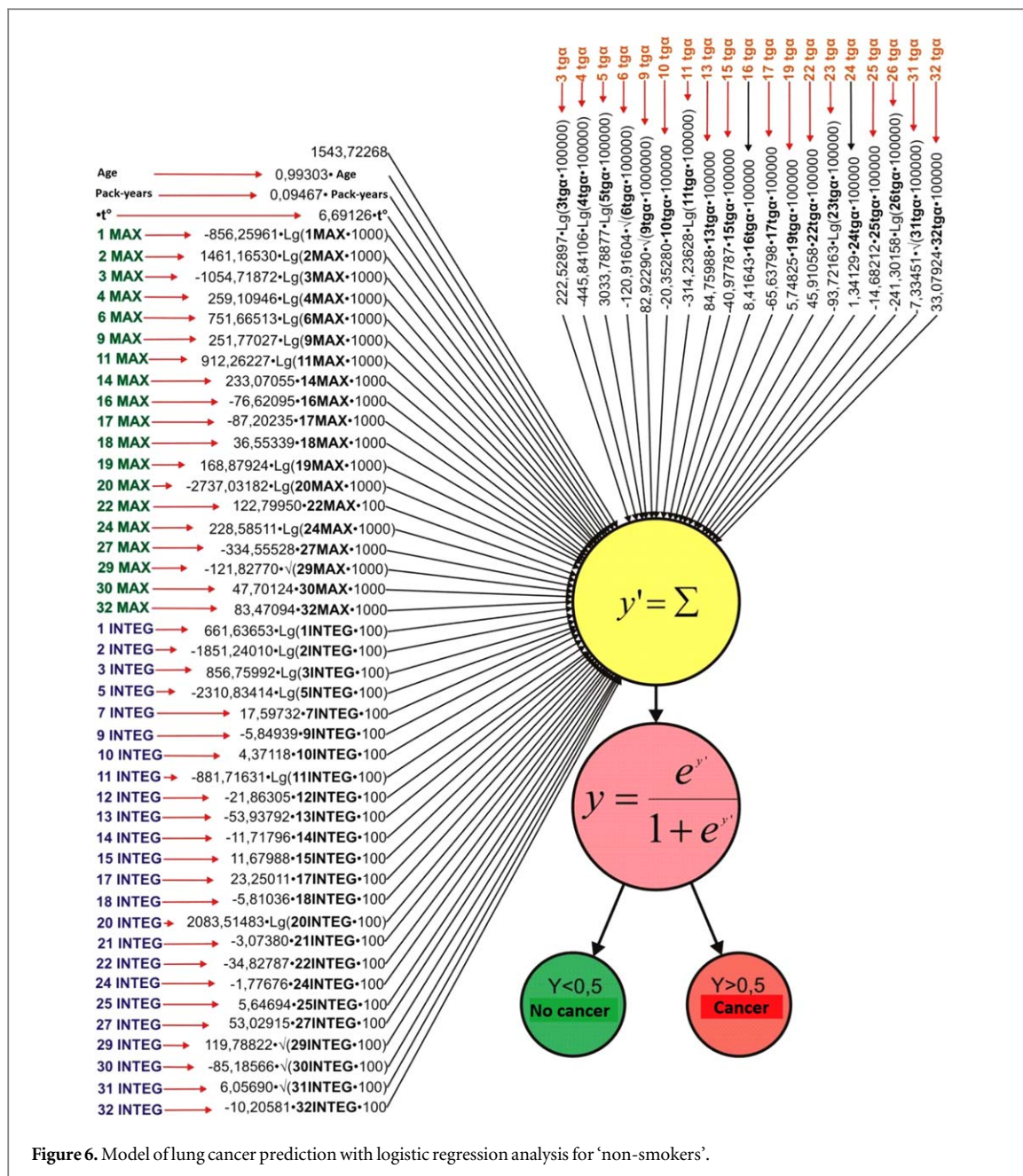


Figure 6. Model of lung cancer prediction with logistic regression analysis for 'non-smokers'.

sensitivity of 76% with a specificity of 56% when it was used to distinguish between patients with clinical signs of VAP. The final diagnosis adjusted even further after receiving the results of the bronchoalveolar lavage [19].

We performed *PubMed* database search on 20th August, 2017, searching for studies where electronic nose could have been used together with LRA in the detection of lung cancer, but could not find any relevant study.

As we have already mentioned, e-nose can be and in some studies has been used together with various methods of statistical analysis, yet until now there is no superior recommended method to be used when analysing data obtained with it. Various statistical methods have been used in analysing data obtained via electronic nose and aimed at detection of lung cancer,

all of which show good results. One of the first publications regarding detection of lung cancer with e-nose was published in 2005 by Machado *et al.* They conducted a study where they used a support vector machine and obtained 91.9% specificity with 66.6% positive predictive value and 93.4% negative predictive value regarding the detection of lung cancer [20].

Another study was published in 2009 by Dragonieri *et al.*, who used canonical discriminant analysis. Their cross-validation results were 85% correct when distinguishing lung cancer from COPD, and cross-validation value 90% correct when distinguishing lung cancer from healthy controls [21].

When using LRA, there is an option to choose the variables and to explore their effect on results. We have included ambient temperature as a predictive

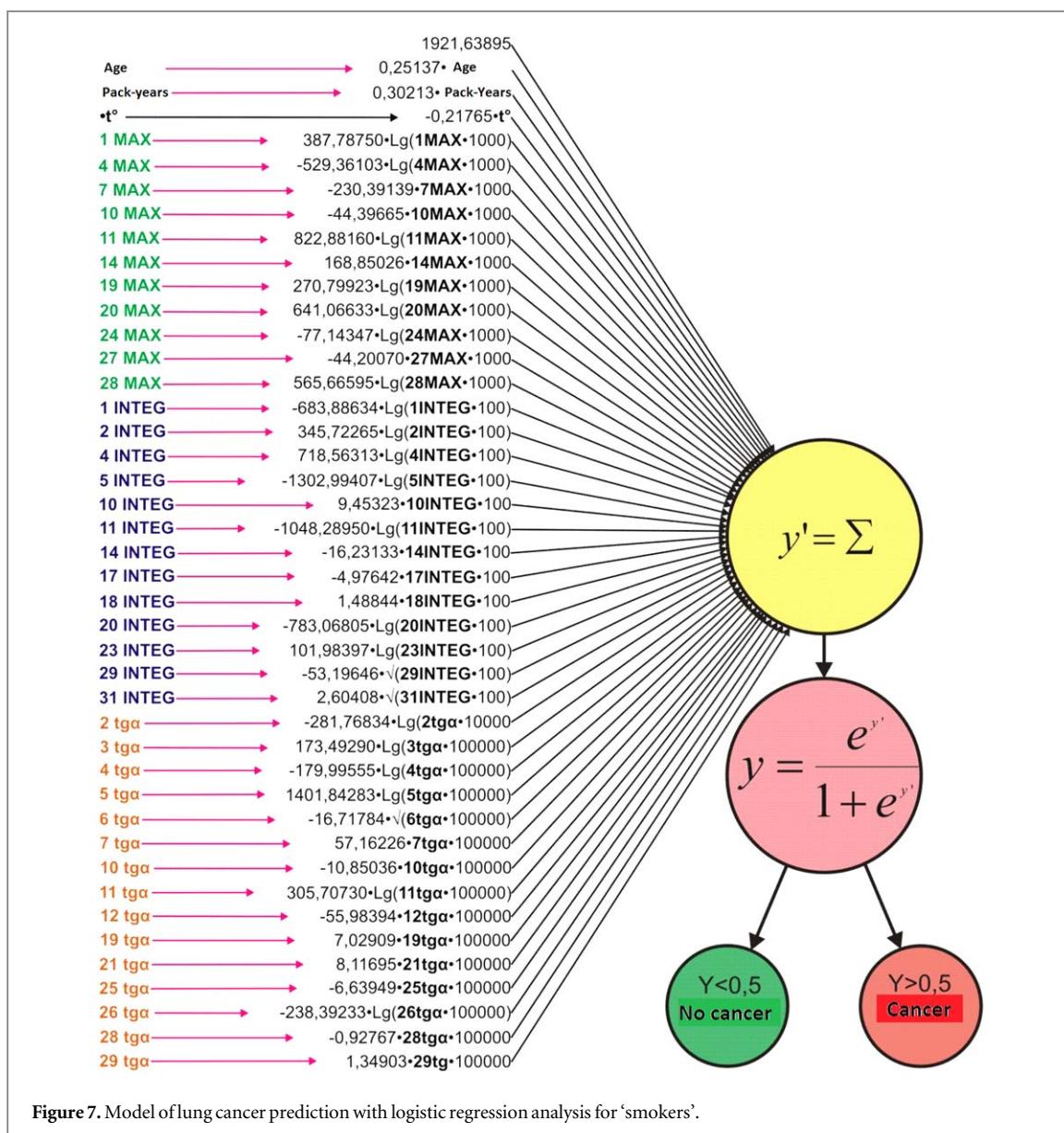


Figure 7. Model of lung cancer prediction with logistic regression analysis for 'smokers'.

Table 1. Logistic regression analysis in the first model.

Non-smokers (n = 265)	Lung cancer (verified)	No-cancer (verified)	
Lung cancer (detected with LRA)	128	12	PPV 91.4%
No-cancer (detected with LRA)	5	120	NPV 96.0%
	Sensitivity 96.2%	Specificity 90.6%	

Table 2. Logistic regression analysis in second model.

Smokers (n = 210)	Lung cancer (verified)	No-cancer (verified)	
Lung cancer (detected with LRA)	114	7	PPV 94.2%
No-cancer (detected with LRA)	5	84	NPV 94.4%
	Sensitivity 95.8%	Specificity 92.3%	

factor because in real life when we were gathering patients' breath samples during the summer, as the indoor temperature exceeded 36 °C, we noticed involuntary changes in data registration and those data could not be included in the study. We therefore believe that if the ambient temperature starts to

approach the e-nose's internal temperature, the e-nose and its sensors could fail.

We could say that our results have a high success rate in detecting lung cancer but we have to stress once again that there is practically no data regarding the use of electronic nose and LRA to compare our results

Table 3. 'Cancer' patients, descriptive statistics of the group.

	Valid N	Mean	Minimum	Maximum	Std. Dev.	SE
Age	252	65.8413	38.0000	84.0000	9.185 48	0.578 631
Height	252	172.1627	150.0000	195.0000	8.811 96	0.555 102
Weight	252	73.8889	45.0000	113.0000	13.656 70	0.860 291
Pack-years	250	31.8720	0.0000	120.0000	21.948 59	1.388 151

Table 4. 'No-cancer' patients, descriptive statistics of the group.

	Valid N	Mean	Minimum	Maximum	Std. Dev.	SE
Age	223	51.2197	19.0000	94.0000	18.942 57	1.268 488
Height	221	171.3801	150.0000	193.0000	9.172 11	0.616 983
Weight	222	74.3018	46.0000	130.0000	17.384 76	1.166 788
Pack-years	220	14.3814	0.0000	87.0000	18.467 32	1.245 067

Table 5. Coherence of patient data regarding pack-years, height and weight.

	'Cancer' ($n = 252$) \pm SD	'No-cancer' ($n = 223$) \pm SD	P value
Gender, male n (%)	202 (80.2%)	134 (60.1%)	
Age, years	65.8 \pm 9.2	51.2 \pm 18.9	<0.0001
Height, cm	172.2 \pm 8.8	171.4 \pm 9.2	0.3449
Weight, kg	73.9 \pm 13.7	74.3 \pm 17.4	0.7726
Smoking history, pack-years	31.9 \pm 21.9	14.4 \pm 18.5	<0.0001

with others, because majority of authors have chosen other data analysis methods.

As we were focussing on the detection of lung cancer, we did not analyse patients' data regarding the link of other diseases with specific breathprint pattern. The number of the patients in each 'non-cancer' disease group was too small to purposefully analyse them and we did not do that, aiming at analysis of lung cancer specifically. Just as well we did not analyse differences between different cancer stages and different histological types. We assume that would be a worthwhile and valuable further investigation, even though in one of the studies where histological type was taken into account, no differences were observed [22].

We admit that we have not taken into account the possible influence of medications taken by the study subjects. As the medication groups used by the study subjects were various and most of them used not only one medication, but combination of medications covering different pharmaceutical groups, it would be incorrect to start analysing the possible influence of medications as the number of patients using each separate medication would be quite small and proper analysis could not be possible as each different pharmaceutical group could have a different impact on the data results and if studied, they should be parted in appropriate groups. There were no patients in the 'cancer' group that had end-stage disease and were severely ill. In healthy volunteers group the patients were not taking any medications. We did not ask the patients about their use of over the counter medications. For further studies in larger patient groups it

could be possible to analyse the possible influence of the used medications.

Just as well additional attention and further recommendations are needed for the optimal rate of exhalation flow. Optimal expiratory flow rate has been discussed since, but when we started our study and set the methodology there were no data available regarding that. As was established by Bikov *et al*, the expiratory flow rate influenced the results of healthy individuals, but not the results of lung cancer patients [22].

A thorough analysis of studies that have used e-nose as their diagnostic tool has been carried out recently to explore the impact and differences arising from different data validation and analysis systems [23], covering 46 prior studies. The authors found that no classification or analysis method has resulted in consistent results of the analysed training set, even after internal validation, and proposed to use external validation as well. This comparative explorative study was not carried out at the time when we were working with our patients and the number of the completed and published e-nose studies was much smaller than it is now. Thus, we have not performed such a thorough analysis with both wide internal and external validation groups ourselves, but we strongly agree that this should be encouraged in further studies.

Regarding LRA as a subgroup of principal component analysis, it is a possible tool to be used and an optimal one taking into account that in LRA one can choose to include or exclude the factors that essentially affect the results. Those might be either factors that depend on the sensor detector, chemical substances,

etc or a combination of two or more factors that can be later avoided and the analysis formula altered, taking those into account. Just as well there might be different approaches to the analysis of each and every factor, e.g., relative maximum, area under the curve, $tg\alpha$ or other. The impact might arise from exhalation rate as well. LRA is good in the way that it is kind of a self-learning tool that might be improved with time, use and problems encountered. With every following patient the model is learning, adapting and calculating the optimal variant to analyse the incoming data. Thus detector changes over time and other similar problems might be avoided with the use of a self-learning model. Though studies regarding the use of e-nose in the detection of lung cancer do not cease and are delivering more and more promising results, there still are some challenges [24, 25]. These are linked with both e-nose device and sensor properties, such as humidity, temperature, sensor stability etc, as discussed above, and the study design/patient selection properties, as well as the choice of statistical methods.

We are looking forward for more recommendations regarding optimal statistical analysis system to be used for data obtained with electronic nose from patients' exhaled breath samples in order to detect lung cancer.

Conclusions

LRA is a good tool to be used with high specificity and sensitivity in the detection of lung cancer. The results obtained with the means of LRA vary slightly depending on the chosen variables. The results of exhaled breath analysis depend on the chosen statistical method.

Acknowledgments/support statement

The project was implemented upon receipt of ERAF activity 2.1.1.1., No. of the agreement 2010/0303/2DP/2.1.1.1.0/10/APIA/VIAA/043.

Certain images in this publication have been obtained by the author(s) from the Wikipedia/Wikimedia website, where they were made available under a Creative Commons licence or stated to be in the public domain. Please see individual figure captions in this publication for details. To the extent that the law allows, IOP disclaim any liability that any person may suffer as a result of accessing, using or forwarding the image(s). Any reuse rights should be checked and permission should be sought if necessary from Wikipedia/Wikimedia and/or the copyright owner (as appropriate) before using or forwarding the image(s).

ORCID iDs

Madara Tirzite  <https://orcid.org/0000-0003-0557-7768>

References

- [1] Stewart B W and Wild C P 2014 *World Cancer Report 2014* International Agency for Research on Cancer
- [2] American Cancer Society 2015 *Cancer Facts & Figures 2015* (Atlanta, GA: American Cancer Society)
- [3] Krilaviciute A, Heiss J A, Leja M, Kupcinskis J, Haick H and Brenner H 2015 Detection of cancer through exhaled breath: a systematic review *Oncotarget* **6** 28643–38657
- [4] Horváth I et al 2017 A European respiratory society technical standard: exhaled biomarkers in lung disease *Eur. Respir. J.* **49** 1600965
- [5] Goodacre R et al 2007 Proposed minimum reporting standards for data analysis in metabolomics *Metabolomics* **3** 231–41
- [6] Sobin L H, Gospodarowicz M K and Wittekind C 2009 *The TNM Classification of Malignant Tumours* 7th edn (Hoboken, NJ: Wiley-Blackwell)
- [7] Tirzite M, Bukovskis M, Strazda G, Jurka N and Taivans I 2017 Detection of lung cancer in exhaled breath with an electronic nose using support vector machine analysis *J. Breath Res.* **11** 036009
- [8] Dragonieri S et al 2007 An electronic nose in the discrimination of patients with asthma and controls *J. Allergy Clin. Immunol.* **120** 856–62
- [9] Walker S H and Duncan D B 1967 Estimation of the Probability of an Event as a Function of Several Independent Variables *Biometrika* **54** 167–79
- [10] Swersky K Support Vector Machines versus Logistic Regression. Lecture materials. University of Toronto (Accessed: 23 June 2016)
- [11] Fawcett T 2006 An introduction to ROC analysis *Pattern Recognition Letters* **27** 861–74
- [12] Altman D G and Bland M J 1994 Statistics Notes: Diagnostic tests 2: predictive values *BMJ* **309** 102
- [13] The McGraw-Hill Companies, Inc. 2002 *Concise Dictionary of Modern Medicine* (New York: McGraw-Hill)
- [14] Farlex, Inc. 2012 *Segen's Medical Dictionary*
- [15] Thaler E R, Huang D, Giebeig L, Palmer J, Lee D, Hanson C W and Cohen N 2008 Use of an electronic nose for detection of biofilms *Am. J. Rhinol.* **22** 29–33
- [16] Sibila O et al 2014 Identification of airway bacterial colonization by and electronic nose in chronic obstructive pulmonary disease *Respir. Med.* **108** 1608–14
- [17] Leunis N, Boumans M L, Kremer B, Din S, Stobberingh E, Kessels A G and Kross K W 2014 Application of an electronic nose in the diagnosis of head and neck cancer *Laryngoscope* **124** 1377–81
- [18] Thaler E R, Lee D D and Hanson C W 2008 Diagnosis of rhinosinusitis with a colorimetric sensor array *J. Breath Res.* **2** 037016
- [19] Schnabel R M, Boumans M L L, Smolinska A, Stobberingh E E, Kaufmann R, Roekaerts P M H J and Bergmans D C J J 2015 Electronic nose analysis of exhaled breath to diagnose ventilator associated pneumonia *Respir. Med.* **109** 1454–9
- [20] Machado R F et al 2005 Detection of lung cancer by sensor array analysis of exhaled breath *Am. J. Respir. Crit. Care Med.* **171** 1286–91
- [21] Dragonieri S, Annema J T, Schot R, van der Schee M P C, Spanevello A, Carratu P, Resta O, Rabe K F and Sterk P J 2009 An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD *Lung Cancer* **64** 166–70
- [22] Bikov A et al 2014 Expiratory flow rate, breath hold and anatomic dead space influence electronic nose ability to detect lung cancer *BMC Pulmonary Med.* **14** 202
- [23] Leopold J H et al 2015 Comparison of classification methods in breath analysis by electronic nose *J. Breath Res.* **9** 046002
- [24] Wilson A D and Baietto M Applications and advances in electronic-nose technologies *Sensors* **9** 5099–148
- [25] Li W, Liu H-Y, Ji Z-R, Qiao P-P, Pi X-T, Chen J and Deng L-H 2014 Advances in the early detection of lung cancer using analysis of volatile organic compounds: from imaging to sensors *Asian Pacific Journal of Cancer Prevention* **15** 4377–84