Volatile organic compounds of biofluids for detecting lung cancer by an electronic nose based on artificial neural network

Ehab I. Mohamed 1 *, Marwa A. Mohamed 2, Samir M. Abdel-Mageed 3, Taher S. Abdel-Mohdy 4, Mohamed I. Badawi 5, Samy H. Darwish 6

1 Alexandria University, Medical Research Institute, Department of Medical Biophysics, Alexandria, Egypt
2 Alexandria University, Medical Research Institute, Department of Chemical Pathology, Alexandria, Egypt
3 Alexandria University, Faculty of Science, Physics Department, Alexandria, Egypt
4 October 6 University, Faculty of Applied Medical Sciences, Department of Biomedical Equipment, Cairo, Egypt
5 Pharos University, Faculty of Allied Medical Sciences, Department of Medical Equipment, Alexandria, Egypt
6 Pharos University, Faculty of Engineering, Department of Electrical Engineering, Alexandria, Egypt

Abstract
Lung cancer (LC) incidence represents 11.5% of all new cancers, resulting in 1.72 million deaths worldwide in 2015. With the aim to investigate the capability of the electronic nose (e-nose) technology for detecting and differentiating complex mixtures of volatile organic compounds in biofluids ex-vivo, we enrolled 50 patients with suspected LC and 50 matching controls. Tissue biopsy was taken from suspicious lung mass for histopathological evaluation and blood, exhaled breath, and urine samples were collected from all participants and qualitatively processed using e-nose. Odor-print patterns were further analysed using the principal component analysis (PCA) and artificial neural network (ANN) analysis. Adenocarcinoma, non-small cell LC and squamous cell carcinoma were the predominant pathological types among LC patients. PCA cluster-plots showed a clear distinction between LC patients and controls for all biological samples; where the overall success ratios of classification for principal components #1 and #2 were: 95.46, 82.01, and 91.66% for blood, breath and urine samples, respectively. Moreover, ANN showed a better discrimination between LC patients and controls with success ratios of 95.74, 91.67 and 100% for blood, breath and urine samples, respectively. The e-nose is an easy noninvasive tool, capable of identifying LC patients from controls with great precision.

Keywords: Artificial Neural Network; Biological Fluids; Electronic Nose; Lung Cancer; Principal Component Analysis; Urine

Abbreviations: ANN, artificial neural network; AUC, area under curve; BMI, body mass index; e-nose, electronic nose; CT, computed tomography; CXR, chest x-ray; LC, lung cancer; NSCLC, non-small cell lung cancer; PCA, principal component analysis; ROC, receiver operating characteristic; SPME-GC/MS, solid-phase microextraction and gas chromatography-mass spectrometry; VOCs, volatile organic compounds

Introduction
Lung cancer (LC) is a major public health problem worldwide and its incidence represents 11.5% of all new cancers, that is 2.02 million people, which resulted in 1.72 million deaths in 2015 (Fitzmaurice et al., 2017). The average incidence rate of tracheal, bronchus, and LC in Egypt has been estimated to be 5.69% among males and 2.70% among females (4.22% for both) in 2008–2011. In fact, tracheal, bronchus, and LC cases are estimated to rise from 5233 in 2015 to 7293 in 2025, based on a 121.6% increase estimations in the Egyptian population (Ibrahim et al., 2014). The prognosis of LC patients depend directly on tumor size and its spread at the time of diagnosis; with survival time being inversely proportional to disease progression, many efforts have been extended to detect the disease at an early phase (Nardi-Agmon and Peled, 2017; Schmidt and Podmore, 2015).

We have recently shown the potential use of an electronic nose (e-nose) based on artificial neural network (ANN) for detecting volatile organic compounds (VOCs) in biofluids of cancer and tuberculosis patients (Mohamed et al., 2014; 2017a, b). VOCs, which are carbon-based chemicals classified on basis of their retention time and boiling point (WHO, 1989), emitted in the headspace over collected body fluids ex-vivo (i.e. blood, exhaled breath, urine, feces, and sweat) are believed to reflect endogenous metabolic processes at the tissue level, such as inflammation and oxidative stress (Nardi-Agmon and Peled, 2017; Schmidt and Podmore, 2015).
2017; Schmidt and Podmore, 2015). These VOCs are typically highly reactive substances capable of damaging DNA and proteins, a process that over time might promote developing cancer in various tissues. It was hypothesized that these pathological processes can also generate new VOCs, which are not produced during normal physiological processes, and/or alter the concentrations of existing ones, may therefore serve as biomarkers for detecting cancers (Schmidt and Podmore, 2015).

Patients often avoid regular preventive examinations and early screening procedures such as tissue biopsies, colonoscopies, X-ray mammographies, uterine dilation, and curettage because many of them are invasive and relatively expensive. Adding to their present hazards of significant negative side effects, they are often sufficiently painful to discourage patients from participating in preemptive, preventive screening procedures. We have employed e-nose devices, which are gas-detection technologies, for use as medical tools for various clinical applications. These rapid, efficient and relatively cheap devices have proven useful for the non-invasive diagnosis and early detection of chronic diseases and cancers, including LC. The benefits of cancers early detection are obvious, since prognoses are greatly improved due to early treatments.

The objective of the present study was to investigate the potential use of the e-nose technology for monitoring the headspace VOCs over blood, breath, urine and tissue biopsy samples collected from histopathologically proven LC patients and controls.

**Materials and methods**

This is a prospective case-control study, which enrolled 50 LC patients with either central or peripheral lung mass, who were admitted to the Alexandria University Main Hospital in Alexandria, Egypt. Other 50 patients without lung mass, who underwent bronchoscopy for clinical indications, mostly hemoptysis and pulmonary infection, were enrolled as controls. Patients with extra-thoracic malignancy, active tuberculosis and histopathological diagnosis of small cell LC were excluded from the study protocol. Patients with metabolic comorbidities (e.g. diabetes mellitus, obesity, and dyslipidemia), or renal failure, which could interfere with VOCs spectrum, were also excluded from the study protocol. All participants provided signed informed consent prior to their inclusion in the study. The study protocol was approved by the Ethical Committee of the Medical Research Institute, Alexandria University, Alexandria, Egypt.

**Physical examination and bronchoscopy**

History taking was carried out for all participants including sex, age, weight, height, body mass index (BMI) and smoking history. All participants also underwent clinical and radiological examinations (i.e. chest X-ray and Computed Tomography) to assess participants’ general health and find other suggestive factors due to LC (e.g. blood in the sputum, significant weight loss, and radiological changes in chest X-ray). Bronchoscopic and thoracoscopic investigations were carried out using a fiberoptic bronchoscope (Pentax-V18, Tokyo, Japan) under local anaesthesia with conscious sedation using midazolam 3–7 mg i.v., with or without propofol (Shure et al., 1985). Oxygen was supplied using nasal cannula in order to maintain saturation above 90% during the procedure. Moreover, oxygen saturation, blood pressure and electrocardiogram were continuously monitored throughout the procedure.

**Biological sample collection**

In case of presence of endobronchial mass or mucosal infiltration, forceps biopsies were taken. In case of peripheral lung mass or the visibility of pleural involvement, rigid thoracoscopy was performed and tissue pleural biopsy/lung biopsy was taken for evaluation. Furthermore, bronchoscopic forceps biopsies were taken from the healthy endobronchial mucosa of controls. Tissue biopsy samples were divided into two parts; one for routine histopathological evaluation and the other for e-nose measurement.

We also collected 20 ml middle-stream early morning urine samples in 200 ml sterile screw-capped containers and 1 ml blood samples in 4 ml heparinized vacutainer tubes from all participants. Mouth exhaled breath samples were collected in inert disposable sterile plastic sacks with a tight closing seal, using the system we detailed earlier (Mohamed et al., 2003; 2017a). The system consists of an elastic rubber tube with a one-way metal valve at its end, which was introduced into an inert disposable plastic sack. We instructed participants to sit comfortably, inhale deeply and exhale fully for five min, prior to using exhaling deeply from the mouth in the other end of the tube to fill sacks, for excluding air from the upper airways and eliminating environmental influences such as smoke, vapors and odors.

**Electronic nose measurements**

We used an e-nose (PEN3, Airsense Analytics GmbH, Schwerin, Germany) to detect odors of VOCs in the headspace over blood, urine and in breath samples at room temperature 25 °C, at the Medical Biophysics Department, Medical Research Institute, Alexandria University, Alexandria, Egypt. Dry air brings blood, urine and in breath samples in 4 ml heparinized vacutainer tubes from all patients and controls. Tissue biopsy samples were divided into two parts; forceps biopsies were taken. In case of peripheral lung mass or the visibility of pleural involvement, rigid thoracoscopy was performed and tissue pleural biopsy/lung biopsy was taken for evaluation. Furthermore, bronchoscopic forceps biopsies were taken from the healthy endobronchial mucosa of controls. Tissue biopsy samples were divided into two parts; one for routine histopathological evaluation and the other for e-nose measurement.

We also collected 20 ml middle-stream early morning urine samples in 200 ml sterile screw-capped containers and 1 ml blood samples in 4 ml heparinized vacutainer tubes from all participants. Mouth exhaled breath samples were collected in inert disposable sterile plastic sacks with a tight closing seal, using the system we detailed earlier (Mohamed et al., 2003; 2017a). The system consists of an elastic rubber tube with a one-way metal valve at its end, which was introduced into an inert disposable plastic sack. We instructed participants to sit comfortably, inhale deeply and exhale fully for five min, prior to using exhaling deeply from the mouth in the other end of the tube to fill sacks, for excluding air from the upper airways and eliminating environmental influences such as smoke, vapors and odors.

**Principal components analysis**

We used PCA technique to find the principal components that maximize the variance in e-nose sensor responses of each dataset (i.e. blood, breath, urine and tissue biopsy) between LC patients and controls. That is, PCA reduced the 10-dimensional correlated sensor responses to only 2-dimensional orthogonal components, while retaining the distance between points in a dataset. It starts with a correlation matrix between sensor responses, which is based on angles between vectors, to derive linear combinations of variables that reflect basic constructs in the sensor responses. PCA is the optimum transform in least square terms, where the greatest variance by any projection of the sensor responses comes to lie on the first coordinate (i.e. principal component #1) and the second greatest variance
Data were presented as mean ± standard deviation (SD). Statistical analysis performed in training and testing modes at neurons. During the training process, a set of examples is presented to the ANN to guess their output. Training is an adaptive process during which the weights associated to all the interconnected neurons change in order to provide the best possible response to all the observed stimuli. Estimation of LC diagnosis was carried out using the neural pattern recognition application of MATLAB 9.2 (MathWorks® Inc., Natick, MA, USA). It combines a modular, icon-based network design interface with an implementation of advanced artificial intelligence and learning algorithms using intuitive wizards together with an Excel™ interface. This provides a user-friendly intuitive interface to easily setup a simulation that automatically builds, trains and tests multiple neural network topologies, and generates a report of the simulation that automatically builds, trains and tests multiple neural network topologies, and generates a report of the results including the best performing model (Abdel-Mageed et al. 2015; Mohamed et al., 2017a).

The variables sex, age, weight, height, and BMI in addition to 10-sensor responses for blood, breath, urine and tissue biopsy samples were prepared in an Excel database as the input parameters. The Excel database was loaded within the MATLAB and used for classification according to reference pathological evaluation of biopsy. We used 60% of data for training, 25% for testing and 15% for validation. That is, of the 100 records examined by the ANN for being LC patients or controls, there were 60 records set for training, 25 for testing and 15 for validation (Mohamed et al., 2017a). We used a two-layer feedforward network, with a sigmoid transfer function in the hidden layer, and a softmax transfer function in the output layer. The default number of hidden neurons was set to two, which is equal to the number of elements in the target vector (i.e. either LC or control).

**Statistical analysis**

Data were presented as mean ± standard deviation (SD). Two-tailed t-test of significance was used to compare between various continuous variables of both groups. Significant P value was considered if P < 0.05. SPSS statistical package (Version 16; Chicago, IL, USA) was used for the analysis.

Confusion matrix, mean square error (MSE), Receiver Operating Characteristic (ROC) curve and area under curve (AUC) analysis; as well as predictive accuracy, sensitivity, and specificity for evaluating the ANN performance to identify/classify LC patients from controls were given. ROC is a plot of the true positive rate (sensitivity) versus the false positive rate (1 – specificity) as the threshold is varied. A perfect test would show points in the upper-left corner, with 100% sensitivity and 100% specificity (Mohamed et al., 2018).

**Results**

The demographical and clinical characteristics for LC patients and controls are shown in Table 1. The majority of the LC patients were in the age range of 54–68 year, with a male to female ratio of 3 : 1, while controls were in the age range of 51–62 year, with a male to female ratio of 2 : 1. There was no significant difference between both groups regarding sex, age or smoking status, yet the smoking index was significantly higher for LC patients than that for controls. That is, major number of LC patients were heavy smokers, since 52% reported to have been smoking more than two packets of cigarettes per day, 20% quit smoking 2–4 years ago, while 28% were non-smokers. On the other hand, 56% of the controls were mild smokers, who reported to have been smoking more than a packet of cigarettes per day, 16% were former smokers and the remaining 28% were non-smokers.

Furthermore, 64% of the LC patients were coughing up blood (hemoptysis), 72% suffering dyspnea, 50% were complaining of chest pain, and 24% showed significant weight loss due to the loss of appetite. This was evident by significantly (P < 0.001) lower mean weight and BMI for LC patients as compared to controls (Table 1). Thoracoscopic tissue biopsy was obtained from 16% of LC patients and 24% of controls while, the 84% of LC patients and 76% of controls who had central lesion were candidates to bronchoscopic forceps tissue biopsy. The majority of LC patients (72%) were in the advanced stages III and IV at time of admission and the remaining 28% were in early stages I and II that were pathologically proven to be carcinoid tumor (Table 1). Finally, histopathological analysis showed that adenocarcinoma was the most common diagnosis in 45.5% of LC cases, followed by undifferentiated non-small cell lung cancer (NSCLC) in 27.3% of them, then squamous cell carcinoma in 10%.

<table>
<thead>
<tr>
<th>Biopsy type, n (%)</th>
<th>Controls Lung cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoracoscopy bronchoscopy</td>
<td>12 (24%)</td>
</tr>
<tr>
<td>LC stage, n (%)</td>
<td>38 (76%)</td>
</tr>
<tr>
<td>Stage I</td>
<td>–</td>
</tr>
<tr>
<td>Stage II</td>
<td>–</td>
</tr>
<tr>
<td>Stage III</td>
<td>–</td>
</tr>
<tr>
<td>Stage IV</td>
<td>–</td>
</tr>
</tbody>
</table>

1 P < 0.001 as compared to controls; LC, lung cancer.
Cluster plots by PCA for blood, breath, and urine samples for LC patients and controls are shown in Fig. 1. A clear distinction between both groups is evident, where both principal components #1 and #2 explained more than 95.46, 82.01, and 91.66%, respectively of the variance in signals. This means that e-nose was capable of identifying samples from each group with no false-positive (controls) or false-negative (LC patients) results. This strengthens the fact that the VOCs in the head-space over blood, breath, and urine samples that overlap with VOCs for tissue biopsy samples of LC patients are significantly different from those for controls.

The confusion matrix and performance indices calculated for evaluating the ANN based on training, validation and testing phases of blood, breath and urine samples for LC patients and controls is shown in Table 2. The accuracy of the ANN for identifying blood, breath, and urine samples were 95.74, 91.67, and 100, respectively, as compared to pathological evaluation of biopsies. Moreover, the overall sensitivity and specificity of the ANN for classifying LC patients and controls was 100% for urine samples, while those for blood samples were 96.15% and 95.24% and for breath samples were 92.86 and 90% (Table 2). The performance plot of a well-trained ANN for the mean squared error versus epochs is shown in Fig. 2. The best validation and testing performance of blood, breath, and urine samples were achieved at epochs 7, 6 and 6 with the lowest mean squared error of 0.13, 0.20, and 0.20, respectively, for LC patients and controls. From these results, we can appreciate the highly successful discrimination of the ANN between LC patients and controls for all measured samples.

ROC curve of sensitivity (i.e. true positive rate) versus 1 – specificity (i.e. false positive rate) for all blood, breath and urine samples from LC patients and controls, is shown in Fig. 3. The best ANN classification is shown for urine samples of the ROC space (Fig. 3C), which has an AUC of “1” representing 100% sensitivity and specificity. The second best classification is that for blood (Fig. 3A), followed by breath samples (Fig. 3B).

Fig. 1. Cluster plots of principal component #1 against principal component #2 for an array of 10 metal-oxide sensors based e-nose measurements when applied to the headspace of A) blood, B) breath and C) urine samples from LC patients and controls. LC, lung cancer, CTR, controls. (For colour resolution see the on-line edition.)

Fig. 2. Performance of ANN for all A) blood, B) breath and C) urine samples from LC patients and controls. ANN, Artificial Neural Network; LC, lung cancer. (For colour resolution see the on-line edition.)
Table 2. Confusion Matrix and performance indices calculated for evaluating the ANN based on training, validation and testing phases for blood, breath, and urine samples from lung cancer (LC) patients and controls

<table>
<thead>
<tr>
<th>Target output</th>
<th>Blood (n = 47)</th>
<th>Breath (n = 48)</th>
<th>Urine (n = 51)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Lung cancer</td>
<td>Controls</td>
</tr>
<tr>
<td>Controls</td>
<td>20</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>LC</td>
<td>1</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>95.74</td>
<td>91.67</td>
<td>100</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>96.15 (78.42–99.80)</td>
<td>92.86 (75.04–98.75)</td>
<td>100 (82.83–100)</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>95.24 (74.13–99.75)</td>
<td>90 (66.87–98.25)</td>
<td>100 (84.50–100)</td>
</tr>
<tr>
<td>Prevalence (%)</td>
<td>55.32 (40.24–69.54)</td>
<td>58.33 (43.28–72.07)</td>
<td>47.06 (33.16–61.40)</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>96.15 (78.42–99.80)</td>
<td>92.86 (75.04–98.75)</td>
<td>100 (82.83–100)</td>
</tr>
<tr>
<td>False positive (%)</td>
<td>3.85 (0.20–21.58)</td>
<td>7.14 (1.25–24.96)</td>
<td>0 (0–17.17)</td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>95.24 (74.13–99.75)</td>
<td>90 (66.87–98.25)</td>
<td>100 (84.5–100)</td>
</tr>
<tr>
<td>False negative (%)</td>
<td>4.76 (0.25–25.87)</td>
<td>10 (1.75–33.13)</td>
<td>0 (0–15.50)</td>
</tr>
</tbody>
</table>

1 Numbers in parentheses are the “exact” Clopper-Pearson 95% confidence intervals (CI); ANN, artificial neural network.

Fig. 3. ROC curve analysis of sensitivity (true positive rate) against 1 – specificity (false positive rate) for all A) blood, B) breath and C) urine samples from LC patients and controls. ROC, receiver operating characteristic; LC, lung cancer.

Discussion

LC is the most common malignant tumor and the leading cause of cancer death among men and women worldwide (AICR, 2007; Mohamed et al., 2018). LC mortality is frequently associated with late diagnosis, therefore early diagnosis and immediate initiation of treatment are essential for significantly improving overall survival in high risk populations of asymptomatic patients (Nardi-Agmon and Peled, 2017; Schmidt and Podmore, 2015). Novel screening tests have been developed for the early diagnosis of LC, which has been remodeled on basis of its own genetic, biological and metabolic identity, such as exhaled breath analysis, serum biomarkers and urine analysis (Gasparri et al., 2018; Mohamed et al., 2018).

The majority of LC patients in this study suffered from hymoptysis and loss of appetite, which were responsible for significant weight and BMI loss. LC is responsible for both symptoms, yet heavy smoking given by the significantly higher smoking index for LC patients as compared to controls (i.e., 65.83 ± 36.43 vs. 25.22 ± 19.30, Table 1), is also responsible for the loss of appetite and loss of weight. Smoking is both a powerful risk factor for LC (Alberg et al., 2005), and is inversely associated with body weight (Wannamethee et al., 2001). Slimness is associated with increased LC risk, though the link generally has been attributed to smoking and preclinical disease. El-Zein et al. (2013), examined the relationship between BMI and LC, using an alternative approach to represent participants’ smoking history, which took into account several dimensions of smoking behavior. They showed that the association between BMI and LC is unlikely due to residual confounding by smoking. Yang et al. (2009), showed that lower BMI is associated with increased risk of LC incidence and mortality, albeit higher BMI has been associated with an increased risk of several other cancers (e.g., esophageal, pancreatic, colorectal, endometrial, postmenopausal breast, and prostate cancers) (AICR, 2007).

While early detection of small cancerous lesions carries the benefit of wider treatment options and better prognosis of LC patients, the process of obtaining a tissue biopsy to confirm a cancerous tissue is not free of complications and bears incon-
veniences and stress to the patient. Despite advances in technology and intensive research efforts, no molecular biomarker capable of identifying LC in the early stages has been found to be suitable for clinical use (Nolen et al., 2015; Zhang et al., 2015). Research efforts have been focused on identifying sensitive and specific blood-based biomarkers for early detection of LC, such as: identification of proteins, protein panels or antibodies to tumor-associated antigens; analysis of epigenetic changes such as methylation; microRNA profiling; and gene expression profiling (Tsay et al., 2013). That is, biomarkers derived from tumor tissue have been the main focus for establishing prognostic and predictive markers in LC (Xu-Weller and Carbone, 2017). In this context, we have recently investigated the prognostic potential of the tissue biomarkers miR-155 and miR-486-5p for diagnosing LC (Mohamed et al., 2018). We found that, miR-155 was highly overexpressed, yet didn’t correlate with LC stages, while miR-486-5p was extremely underexpressed and significantly correlated with LC stages. Thus, their detection represents an excellent diagnostic/prognostic tool to support more established techniques linked to LC spread locally and systemically.

We and others have shown that analysis of VOCs in biofluids by e-nose provided valuable information on metabolic processes altered by underlying diseases in the body (Machado et al., 2005; Mazzone et al., 2007; Mohamed et al., 2014; 2017a,b), which can be particularly promising for the non-invasive LC screening and diagnosis (Dragonieri et al., 2009; 2017). The present study showed that PCA explained more than 95.46, 82.01, and 91.66% of the variance found in e-nose signals for blood, breath and urine samples, respectively (Fig. 1). This means that e-nose was capable of identifying samples from each group with no false-positive (controls) or false-negative (LC patients) results.

Although VOCs detected in blood and urine, it is not known which volatile compounds are produced/consumed by tumor cells, since they may also be produced/consumed by surrounding noncancerous tissue cells or other regions of the body (Poulin and Krishnan, 1996; Silva et al., 2011), immune-competent cells (Aksenov et al., 2012), and infectious pathogens (Mohamed et al., 2017a). Ideally, one needs to compare e-nose patterns of VOCs in the headspace of cancer cells or tissues of a particular patient with those found in blood, breath, and/or urine of the same patient (Schmidt and Podmore, 2015). Thus, the overlap of blood, breath, and urine samples clusters with that of tissue biopsy for LC patients (as shown in Fig. 1), means that they have common VOCs characteristic for LC yet, significantly different from controls. In line with these findings, Deng et al. (2004) used solid-phase microextraction and gas chromatography-mass spectrometry (SPME-GC/MS) for investigating LC volatile biomarkers. They showed that 23 VOCs found in blood were also present in the exhaled breath of LC patients, meaning that there are characteristic compounds for identifying the presence of LC. Among those 23 VOCs, hexanal and heptanal were detected only in blood and breath of LC patients and were not found in controls. Guadagni et al. (2011), have shown later on that the endogenous urinary aldehydes: hexanal and heptanal can be used as biomarkers for identifying LC patients.

Furthermore, Buszewski et al. (2011) compared quantitative measurements of VOCs using SPME-GC/MS technique in the headspace of tissue and breath samples of LC patients and healthy controls. They detected 27 VOCs in the headspace above LC tissues, of these only 22 (mainly alcohols, aldehydes, ketones, and aromatic and aliphatic hydrocarbons) were found in both LC tissue and breath samples. Higher levels of ethanol, acetone, acetonitrile, 1-propanol, 2-propanol, carbon disulfide, dimethyl sulfide, 2-butanone, and 2-pentanone were detected in tissue and breath samples from LC patients in comparison with healthy non-smoking volunteers.

The accuracy of ANN classification of blood, breath, and urine samples for LC patients and controls were 95.74, 91.67, and 100%, respectively; with an AUC of “1” representing 100% sensitivity (i.e. no false negatives) and 100% specificity (i.e. no false positives) for urine samples (see Table 2 and Fig. 3). As we recently pointed out (Mohamed et al., 2017a), these results therefore support the hypothesis of diagnosing LC on basis of VOCs signals in urine, compared with pathological evaluation of tissue biopsies, and could be used in the clinical setting and for large-scale, non-invasive, low-cost mass screening of LC patients with great precision.

Recently, urine has attracted increased attention as a biospecimen that may be useful for the identification of cancer biomarkers. Urine is protein-rich and contains approximately 3.000 detectable protein species. Moreover, urine proteins are stable because the secreted proteins are mature, and urine is easy to collect in large volumes in a non-invasive manner (Zhang et al., 2015). In addition to its non-invasive nature and availability in large volumes, VOCs analysis of urine has an advantage over other biofluids (i.e. blood and exhaled breath), since analytes are concentrated by the kidneys before being excreted from the body. Furthermore, when compared to blood, the use of urine usually results in better detection limits as matrix effects may interfere with the release of the VOCs into the headspace over blood samples (Mills and Walker, 2000).

Conclusions

The e-nose was capable of identifying/classifying all studied biofluid samples with no false-positive (controls) or false-negative (LC patients) results. The accuracy of the ANN for identifying blood, breath, and urine samples were 95.74, 91.67, and 100, respectively, as compared to pathological evaluation of biopsies. These results therefore support the hypothesis of diagnosing LC on basis of VOCs signals in urine and could be used in the clinical setting and for large-scale, non-invasive, low-cost mass screening of LC patients with great precision.

Conflict of interests

The authors have no conflict of interests to declare.

Acknowledgements

The authors would like to express their deepest gratitude to the medical staff of Department of Chest Diseases at the Alexandria University Main Hospital, Alexandria, Egypt for the dedication and technical assistance provided with the bronchoscopy carried out for the study population.

The preliminary results of this study were presented in abstract form at the 2nd International Lung Cancer Symposium, which took place in Budapest; Hungary at the period from 16–17 April, 2016.

References


