Computers in Biology and Medicine 96 (2018) 227–232

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Early non-invasive detection of breast cancer using exhaled breath and urine analysis

Or Herman-Saffar [a], Zvi Boger [a,b], Shai Libson [c], David Lieberman [d], Raphael Gonen [a], Yehuda Zeiri [a,*]

[a] Biomedical Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
[b] OPTIMAL – Industrial Neural Systems, Be'er Sheva 84243, Israel
[c] Breast Health Center Soroka Medical Center, Ben-Gurion University, Beer Sheva, Israel
[d] Pulmonary Unit, Soroka University Medical Center and the Faculty of Health Sciences, Ben- Gurion University of the Negev, Israel

ARTICLE INFO

Keywords:
Exhaled breath
Urine
Breast cancer diagnosis
Artificial neural networks

ABSTRACT

The main focus of this pilot study is to develop a statistical approach that is suitable to model data obtained by different detection methods. The methods used in this study examine the possibility to detect early breast cancer (BC) by exhaled breath and urine samples analysis.

Exhaled breath samples were collected from 48 breast cancer patients and 45 healthy women that served as a control group. Urine samples were collected from 37 patients who were diagnosed with breast cancer based on physical or mammography tests prior to any surgery, and from 36 healthy women. Two commercial electronic noses (ENs) were used for the exhaled breath analysis. Urine samples were analyzed using Gas-Chromatography Mass-Spectrometry (GC-MS).

Statistical analysis of results is based on an artificial neural network (ANN) obtained following feature extraction and feature selection processes. The model obtained allows classification of breast cancer patients with an accuracy of 95.2% ± 7.7% using data of one EN, and an accuracy of 85% for the other EN and for urine samples.

The developed statistical analysis method enables accurate classification of patients as healthy or with BC based on simple non-invasive exhaled breath and a urine sample analysis. This study demonstrates that available commercial ENs can be used, provided that the data analysis is carried out using an appropriate scheme.

1. Introduction

Breast cancer (BC) is the most commonly diagnosed malignancy among females and the leading cause of death around the world. In 2016, breast cancer constituted 29% of all the identified new cases of cancer in the US, and 14% of the deaths caused by cancer [1]. The mortality of cancer in general, and BC in particular, is strongly connected with the sensitivity of tumor detection methods used [2]. Consequently, the development of new early tumor detection methods has been a highly active area of research for several decades. Development of new tumor detection schemes requires improved accuracy that can lead to detection of smaller tumors. However, the new scheme has also to be simple and inexpensive for implementation. Screening mammography is currently the main approach for early detection and has been proven to reduce breast cancer mortality. However, mammography has limitations that are associated with its ability to detect small tumors in dense breast tissue. The overall sensitivity of mammography is 75%–85%, which can decrease to 30%–50% in dense breast tissue [3]. Thus, new methods that can overcome these limitations are needed to identify tumor development at earlier stages of the cancer. One such method is the Dual-energy digital mammography [4,5]. This approach consists of high- and low-energy digital mammograms following administration of an iodine based contrast agent. In this method, the breast is exposed to the low- and high-energy X-ray beams during a single breast compression in mediolateral-oblique (MLO) projection. The breast is then released from compression, and the contrast agent is injected. Following a 3 min delay, the breast is compressed again, and another low- and high-energy exposures is performed to create pre- and post-contrast dual-energy images. These images allow to evaluate the contrast agent kinetics of uptake and washout. Subtraction of the images allows canceling the soft-tissue

* Corresponding author.
E-mail address: yehuda@bgu.ac.il (Y. Zeiri).

https://doi.org/10.1016/j.compbiomed.2018.04.002
Received 21 February 2018; Received in revised form 30 March 2018; Accepted 3 April 2018

contrast common to both images and to isolate the iodine signal in the region of angiogenesis. Dual-energy acquisitions are chosen to maximize and minimize, respectively, the ratio of the attenuation of the breast tissue to that of the iodine. Dual-energy enhanced mammography is an inexpensive technique useful in identification of lesions in dense breasts and capable of demonstrating cancers that are not visible at standard mammography. However, the improved resolution of this method is achieved by the exposure of the breast to an increased dose of X-ray irradiation.

An additional method for breast cancer detection is based on magnetic resonance imaging (MRI) imaging [6]. MRI imaging became increasingly important in the detection and delineation of breast cancer in daily practice. The main diagnostic value of MRI relies on specific situations such as detecting cancer in dense breast tissue and recognition of an occult primary breast cancer in patients presenting with cancer metastasis in axillary lymph nodes, among others. Nevertheless, the development of new MRI technologies such as diffusion-weighted imaging, proton spectroscopy and higher field strength 7.0 T imaging offer a new perspective in providing additional information in breast abnormalities. However, a major drawback of the MRI imaging technique is its high cost.

After tumor detection, in most cases a detailed analysis of the tumor tissue is performed following biopsy [7,8]. This procedure is invasive, and requires a high level of expertise and expensive equipment. Moreover, this approach can be used only to confirm BC after the tumor was identified. Another possibility is to use serum for the identification of BC biomarkers [9–15]. These methods are invasive, require very high degree of expertise, and can be implemented only in specialized laboratories.

Recently there have been attempts to detect various cancers including BC using analysis of exhaled breath and urine samples [16–24]. This type of diagnostic methods has important advantages. They are non-invasive, usually easy to implement and in many cases inexpensive. The analysis of body fluids can be performed using different techniques. One possibility is to examine the chemical composition and to identify biomarkers of the illness studied. This can be achieved using either gas or liquid chromatography coupled to mass spectrometry [16,23–25]. Other possibilities are to use electrochemical sensors [26] or different gas sensors (electronic noses, ENs) [16,20–22,27,28]. In this approach, the measurement of the exhaled breath sample yields a set of signals, the output of the sensors on the EN used, without details related to the chemical composition of the sample. The association between the outcome of all measurement types and the medical state of the individual examined is achieved by performing statistical analysis of the data collected. A wide variety of statistical methods can be utilized in the data analysis, including multivariate regression [29], principle component analysis [20,30,31], artificial neural networks (ANN) [32–35], fuzzy logic [19,35] and other methods.

The present article describes a scheme for data analysis based on artificial neural networks (ANN) that can be used to develop a reliable predictive model. The method is applied to results obtained in a pilot study in which samples of urine and exhaled breath were analyzed from women with initial stages of BC and from a control group of healthy individuals. In this pilot study, the breath samples were analyzed using two different commercial electronic noses. The urine samples were analyzed using gas chromatography with mass spectroscopy (GC-MS) and detected the volatile compounds in the urine. The main goal of the study is to demonstrate that analysis of the raw data leads to very poor models while application of feature extraction and feature selection to the measured data leads to highly accurate models that allow to detect early stages of BC.

2. Experimental and computational methods

2.1. Electronic noses

The exhaled breath analysis was performed using two different commercial ENs. Both ENs contain sensors whose electrical conductivity change when they are exposed to different gas mixtures. The first EN used was the MK4 model (E-Nose Pty Ltd) that contains 12 solid state oxide sensors that have different sensitivities to various gases. The second EN used was the Cyranose 320 (by Sensigent Intelligent Sensing Solutions) that has 32 polymer-based sensors each with a different sensitivity to various gases. The two ENs were attached to a mask through which exhaled breath of the patient was introduced into the ENs. For the MK4 EN, the patient breathes through the mask for a duration of 20 s while for Cyranose 320 the duration was 40 s. The two ENs differ in their sampling rates, the MK4 sampling rate is 0.25 Hz and EN output signal was collected during 416 s. The sampling rate of Cyranose 320 is approximately 2 Hz and the duration of EN output signal collection was 330 s. It should be noted that the shape of the sensor signals for patients and controls are similar with differenced only in their magnitude, rise and fall rate and similar characteristic parameters.

2.2. Urine sample analysis

Urine samples were collected from 37 sick and 36 healthy women. However, both urine and exhaled breath samples were obtained only from 35 sick and 31 healthy women. Gas Chromatography – Mass Spectrometry (GC-MS) analysis was performed using an Agilent 6890 series GC system (Agilent, USA) connected to Agilent 5973 network mass selective detector (Agilent). Further details of the urine sample GC-MS analysis are described in the Supplementary materials section.

2.3. Subjects

The sick women's samples were taken from patients who were diagnosed having breast with cancer based on physical or mammography tests prior to any surgery, irradiation or chemotherapy. All samples taken from sick patients were collected in the Breast Health Center in Soroka Medical Center. All the sick women were identified as having breast cancer by biopsy test after samples were collected. The control group consisted of healthy women who did not present any kind of cancer, pregnancy or acute inflammation when samples were collected. All women were asked to complete a questionnaire that contains the following questions: age, smoking or not, did they have cancer in the past, and if yes - when, known medical problems and medications used.

2.4. Data analysis method

Artificial Neural Network (ANN) is a non-linear modeling algorithm [32–35] that was used to analysis the data. The ANN architecture used is comprised of three layers-input, hidden and output feed forward, fully inter-connected with appropriate weights. The input data, also called features, are fed into the input layer. Additional details of the ANN used are presented in the Supplementary section.

The calculations reported here were carried out using the TURBONEURON [36] code in which the initial connection weights are assigned by the assumption of a linear relationship between the inputs and the outputs. To avoid being trapped in local minima during the training, the code incorporates a scheme of escape from such situations [33]. Additional details are presented in the supplementary materials section.

3. Results and discussion

3.1. Exhaled breath data

The signals obtained from the sensors in the two ENs used have quite different shapes. Typical signals are shown in Figs. SM-1 and SM-2 in the Supplementary materials section. The signal preprocessing started with subtraction of the sensor's initial conductivity from the entire signal. This leads, in most cases, to a zero baseline for all sensors in the ENs. Elimination of the baseline also allows to connect the signals obtained by the

different sensors in an EN to obtain a string of signals for each patient examined. Next, signal smoothing was performed using the *smooth* Matlab function (local quadratic regression) to remove noise. An example of a typical signal before and after the smoothing process is shown in Fig. 1. Clearly, this procedure of signal smoothening is not good enough.

As the next step, a feature extraction process was applied with the aim to completely eliminate the noise and to reduce the number of features (input values) describing the signals. The feature extraction method used is based on a simple fitting of the sensor response to the exhaled breath using an analytic model that describes the time dependence of the sensor response to its exposure to a gas mixture. The analytic model was adapted, with some modifications, from a simple description of the sensor electrical conductivity changes when it is exposed to a gas mixture [37]. Per this model, the sensor response is given by the expression,

$$R_i(t) = \begin{cases} 0, \ t < t_i \\ \beta_i \tau_i \tan^{-1}\left(\dfrac{t-t_i}{\tau_i}\right), \ t_i < t < t_i + T \\ \beta_i \tau_i \left[\tan^{-1}\left(\dfrac{t-t_i}{\tau_i}\right) - \tan^{-1}\left(\dfrac{t-t_i-T}{\tau_i}\right)\right], \ t > t_i + T \end{cases} \quad (1)$$

where: t_i is the time when the signal starts to rise; T is the time interval between beginning of the signal (t_i) until it reaches the maximum value and τ_i is the decay time constant of the signal. The quantity β_i is calculated by

$$\beta_i = \frac{\psi_i^{max}}{\tau_i \tan^{-1}\left(\frac{T}{\tau_i}\right)} \quad (2)$$

And it is related to the amplitude of the signal. The symbol ψ_i^{max} represents the value of the signal at the peak and is calculated by:

$$\psi_i^{max} = \beta_i \tau_i \tan^{-1}\left(\frac{T}{\tau_i}\right) \quad (3)$$

Since this model is not differentiable everywhere, the Matlab function *fminsearch* (function minimum search based on a simplex search method) was used to obtain the best fit of the required parameters. Since all the parameters required in the model described above have physical meaning it was possible to supply the fitting function with good initial guesses of the various parameters. Figs. 2–3 compare the fitting outcome for typical signals.

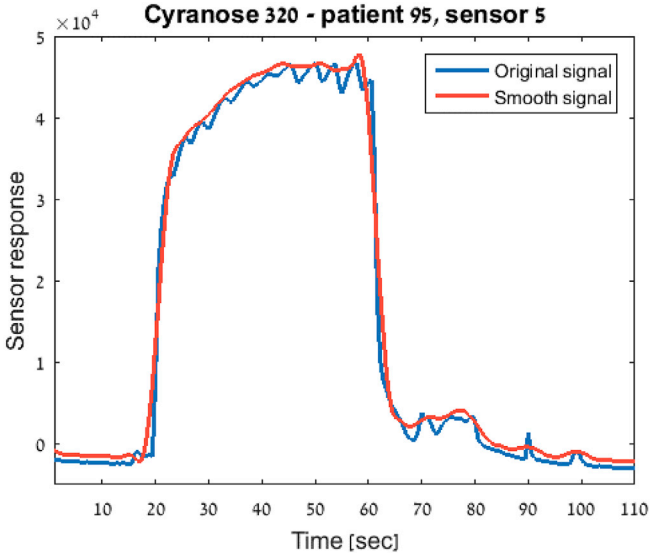

Fig. 1. Typical Cyranose 320 sensor response (blue) and the smoothened signal (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

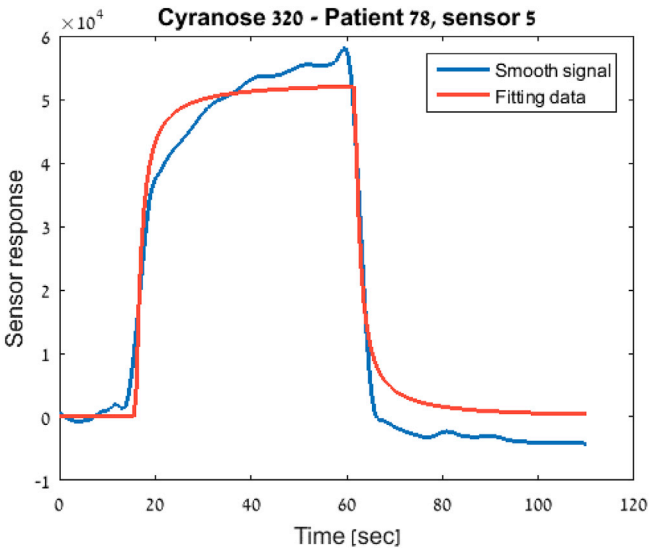

Fig. 2. Fitted analytic model (red) for typical measured signal. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

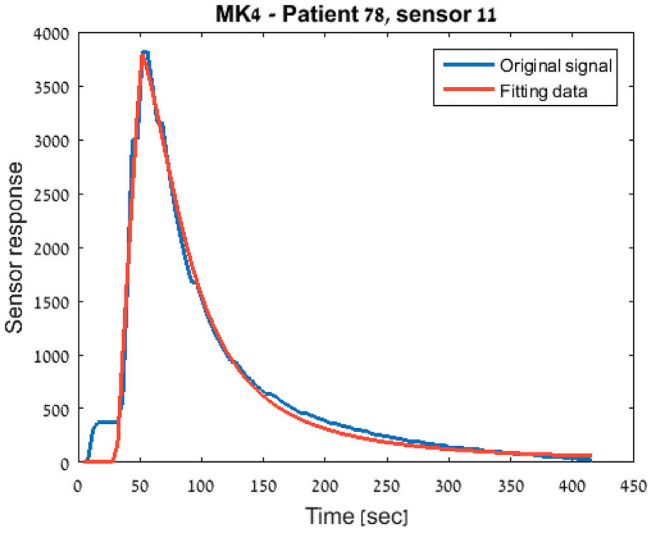

Fig. 3. Fitted analytic model (red) for typical measured signal. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

The fitted model clearly describes well the sensor signal decay, but inaccurately the initial rising part. If the fitting procedure is applied again only to the rising part of the measured signal one can obtain a very good representation of the whole sensor signal. The result of this double fitting process is presented in Figs. 4 and 5 for the signals shown in Figs. 2 and 3.

The application of this fitting process separately to the rising and decaying parts of the signal yields a very good noiseless description of the measured signals. Hence, the feature extraction process allows us to represent very accurately smooth sensor signals using only eight parameters. Three additional parameters were included:

R_{max} – the difference between the signal's peak and baseline (maximum height); A_{left} the area under the curve left of its maximum value (area under the signal rising part); and A_{signal} the area under the entire signal.

Besides these features, two additional patient related features were included – the number of years since the first time the patient was diagnosed having breast cancer (zero if she did not have breast cancer

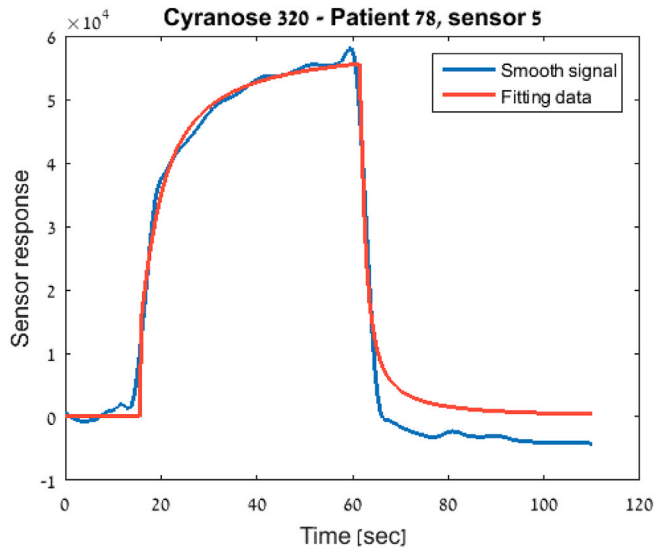

Fig. 4. Final fitted analytic model (red) for typical measured signal (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

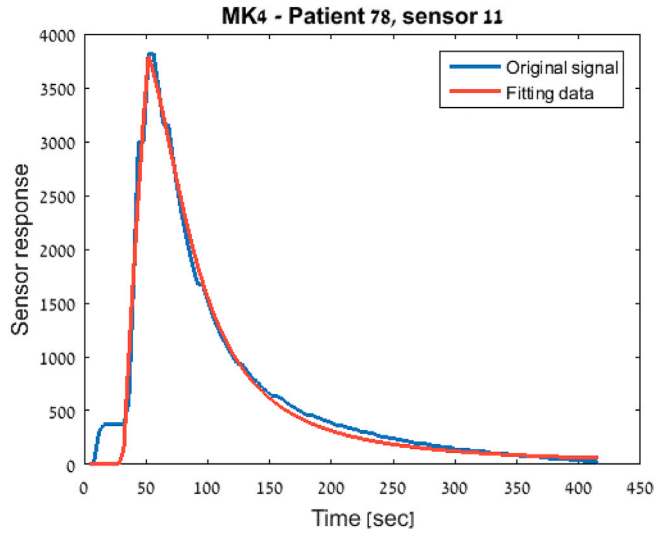

Fig. 5. Final fitted analytic model (red) for typical measured signal. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

before, most of the cases) and how many cigarettes the patient smokes every day (zero if she is a non-smoker). Thus, each sensor signal was described by 14 features, giving a total of 134 features for the MK4 EN and 354 features for the response of the Cyranose 320 EN. These values should be compared with 1248 features for the MK4 EN and 21,120 features of the Cyranose 320 EN used when the raw ENs output was employed to represent the EN signals.

The data obtained for exhaled breath analysis using the two ENs were used to form ANN models. Table 1 below summarizes the accuracy of the ANN models obtained when the preprocessed raw data is used (marked as "raw") for both ENs and the same data after using the analytical expression for feature extraction (FE). Moreover, we also examined the influence of "bad" sensor responses (signal shape highly distorted and containing multiple narrow peaks) on the accuracy of the models obtained. To do so we considered all the exhaled breath samples collected, a total of 85 (46 sick and 39 healthy women) samples examined by both ENs as "group 1". When all the samples that contain "bad" sensor signals

Table 1
Validation results of the ANN models obtained for the two sets of data "Group 1" and "Group 2" using both the raw data and the one following feature extraction (FE).

Type of EN	"Group"	Data type	Accuracy [%]	Sensitivity [%]	Specificity [%]
MK4	1	raw	67.5 ± 21.6	67.5 ± 18.5	67.5 ± 33.4
MK4	1	FE	70.4 ± 19.1	73.5 ± 22.6	67.5 ± 23.7
MK4	2	raw	73.8 ± 24.6	68.3 ± 28.8	80 ± 28.1
MK4	2	FE	76.2 ± 18.5	74.17 ± 25	78.3 ± 15.3
Cyranose 320	1	raw	60.9 ± 22.8	56 ± 29.1	66.7 ± 23.6
Cyranose 320	1	FE	57.4 ± 22	67 ± 29.7	46.7 ± 26.7
Cyranose 320	2	raw	51.7 ± 22.1	51.7 ± 21.4	52.5 ± 36.0
Cyranose 320	2	FE	55 ± 22.2	48.3 ± 18.8	61.7 ± 31.2

are eliminated we remain for both ENs with a total of 65 samples for both ENs (33 sick and 32 healthy women), this set will be termed as "group 2". A detailed description of the method to obtain: accuracy, sensitivity and specificity is given in section 2 of the Supplementary material.

Inspection of the results in Table 1 shows that there is a marked difference between the two ENsa. In the case of MK4, performing the feature extraction leads to some improvement in model performance for both raw data and the FE data. The elimination of samples containing "bad" signals results in a further improvement in model performance. However, models corresponding to data collected using the Cyranose 320 EN exhibited markedly lower performance in all cases. In this case, both feature extraction and removal of "bad" signals did not lead to any noticeable improvement of the model performance.

The performance of all the models presented in Table 1 is better than what one would get by random assignment of samples as "sick" or "healthy" (for a large enough statistical ensemble). However, for such models to be useful, one would like to improve these results and obtain much better performance. A possible reason for the low performance of the models might be due to existence of strong similarities between features that hinders the ANN training from reaching a good model. To overcome such a problem one can perform a feature selection process. Feature selection is aimed to produce a smaller set of features that accurately represents the original set but with reduced similarity among features. The feature selection method used is based on measuring the similarity between features among different samples, reducing the redundancy, allowing development of a simpler and more accurate classification model. The feature selection method employed is described in detail elsewhere [38]. In this method, a new similarity measure, called maximum information compression index between features X and Y, $\lambda_2(X,Y)$, was used together with the Pearson correlation measure $(\rho_{x,y})$ defined by:

$$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} \tag{4}$$

where cov(X,Y) is the covariance of X and Y, σ_X and σ_Y are the standard deviation of X and Y sets while μ_X and μ_Y are the corresponding mean values. The symbol E denotes expectation value. The sets X and Y are vectors containing the values of two different features in all the samples (all the patients in the present case). Thus, the size of the X and Y vectors is given by the number of patients. The maximum information compression index is calculated using the expression:

$$\lambda_2(X,Y) = \frac{1}{2}\left[var(X) + var(Y) - \sqrt{(var(X)+var(Y))^2 - 4var(X)var(Y)\left(1-\rho_{x,y}^2\right)} \right] \tag{5}$$

where var(X) is the variance of feature X. The maximum information compression index is a scalar whose magnitude is zero if X and Y are linearly dependent. Increasing values of $\lambda_2(X, Y)$ are obtained when the two features become linearly independent.

The feature selection process requires two steps, partitioning of the original feature set into a number of subsets followed by selection of representative features from each subset. The partitioning step is based on the Pearson correlation measure described above. To carry out this part, one computes the j nearest features to each feature. The feature having the most compact subset is selected and its j neighboring features are removed. The magnitude of the parameter j has to be defined to perform the selection process. For larger j values an increased number of features are removed and smaller number of features selected. Here, we used the algorithm described in details by Mitra et. Al. [39]. In the k-fold cross validation process the feature selection step is performed before each one of the k-fold ANN training processes. In each case the j value yielding the best model performance is chosen. The results of the models obtained after feature selection for the data collected by the two ENs are presented in Table 2.

Inspection of the results presented in Table 2 shows that the performance of all models improves dramatically when feature selection is included prior to model training. The performance of the MK4 EN is still noticeably better than that of the Cyranose 320 electronic nose. The results clearly demonstrate that both feature extraction and removal of "bad" signals lead to a markedly improved model performance of both ENs. The accuracy achieved in the models reported in Table 2 suggests that commercial ENs can be used for early breast cancer diagnosis if the statistical analysis is properly performed.

3.2. Urine data

The chemical composition of the urine samples was examined using GC-MS analysis. A typical chromatogram is shown in the Supplementary materials section, Fig. SM-1.

The data that were used in the statistical analysis is a feature vector that contains a list of all the peaks identified by the system software in the chromatogram. Each peak was characterized by its retention time (RT) and normalized peak area values. The peak area normalization and RT alignment were described in the Supplementary material section. The features vector contained 625 time intervals (boxes), each representing 2.4 s. The normalized peak area values were stored as features in the appropriate time intervals. Time intervals that did not contain any peaks were assigned the value zero. No feature extraction was performed in this case. The validation results of models obtained using the raw data and the model obtained after feature selection (FS) using the procedure described above are presented in Table 3. In both cases 73 (37 sick and 36 healthy women) urine samples were considered.

Table 2
Validation results of the ANN models obtained for the two sets of data "Group 1" and "Group 2" using both the raw data and the one following feature extraction (FE). In all cases a feature selection step was performed prior to the ANN training stage.

Type of EN	"Group"	Data type	Accuracy [%]	Sensitivity [%]	Specificity [%]
MK4	1	Raw	89.6 ± 8.7	91 ± 11.7	87.5 ± 13.2
MK4	1	FE	93.9 ± 12.1	93 ± 16.4	95 ± 10.5
MK4	2	Raw	93.6 ± 11.5	97.5 ± 7.9	90 ± 22.5
MK4	2	FE	94.3 ± 7.4	89.2 ± 14.2	100 ± 0
Cyranose 320	1	Raw	81.0 ± 11.2	87.5 ± 10.9	73.3 ± 23.2
Cyranose 320	1	FE	86.1 ± 12.8	87.5 ± 19.6	85 ± 21.1
Cyranose 320	2	Raw	82.6 ± 12.5	86.7 ± 23.3	77.5 ± 15.7
Cyranose 320	2	FE	84.5 ± 13.7	91.7 ± 13.6	78.3 ± 21.9

Table 3
Validation results of the ANN models obtained for the urine analysis by GC-MS. Here, "Raw" correspond the raw output obtained by the GC-MS while "Following FS" correspond the model obtained after feature selection.

Data type	Accuracy [%]	Sensitivity [%]	Specificity [%]
Raw	45.2 ± 24.5	48.3 ± 28	44.2 ± 36.2
Following FS	**85.2 ± 9.7**	**86.7 ± 14.3**	**85 ± 21.1**

The results presented in Table 3 show that the analysis of the raw data leads to very low performance model with very large standard deviations. These results are worse than pure random selection of the target groups (sick or healthy). When feature selection is applied, the performance of the model obtained improves dramatically and reaches that of the electronic noses described above.

4. Summary and conclusions

This study examined the possible detection of breast cancer at its initial stage by analysis of both exhaled breath and urine. Exhaled breath was analyzed using two different commercial ENs while urine samples were analyzed by GC-MS. For all types of measurements, very poor ANN models were obtained when pre-processed raw data were used. To improve data modeling, feature extraction and feature selection processes were applied to the exhaled breath measured data while only feature selection was used in the case of the urine features. The ANN models obtained using the processed data were very good to a level that they can be used as a non-invasive, simple, safe, painless and inexpensive screening method. It should be noted that although feature extraction alone, in the case of breath analysis, did not sufficiently improve the ANN models, it was necessary to obtain the good models reach following feature selection. The present study clearly demonstrates that appropriate statistical data processing should be performed to obtain reliable classification models. This study clearly demonstrates that inexpensive commercial ENs can be used successfully for exhaled breath analysis, provided that an appropriate data analysis procedure is used. It should be noted that in this study, feature extraction alone of data obtained in exhaled breath analysis was shown to improve the model performance, however it was not enough. The feature selection process seems to be the crucial step in improvement of the model performance. This was found to be true for both sets of exhaled breath measurements and urine analysis.

It should be noted that other types of cancer and their influence on the analysis of exhaled breath and urine samples have yet to be examined. It is possible that the two sample types examined here can also be used to identify other types of cancer. Indeed, it has been shown that exhaled breath analysis can be used to screen different types of cancer [20]. The accuracy of ANN based classification models described here for breast cancer detection, using both exhaled breath and urine samples, is comparable to those obtained in other studies of exhaled breath using different statistical analysis approaches [20,21,26,27,31]. It is clear that the use of commercial ENs could be sufficient to establish a good reliable screening method for cancer classification patients.

The promising classification models obtained in different studies, using a variety of measurement methods, raises a possibility that should be checked. Namely, is it possible to develop a coupling scheme between models conceived using data obtained by methods based on different physical or chemical principles (i.e. analysis of exhale breath and urine samples or ENs based on nano-particles embedded in polymers and solid state sensors etc.) such that the classification accuracy of the coupled models will be better than that of its constituent models. This possibility is being examined at present and will be examined in a future study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflicts of interest

The authors declare that they have no conflict of interest.

Declarations of interest

None.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.compbiomed.2018.04.002.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, CA A Cancer J. Clin. 2016 (66) (2016) 7–30. http://onlinelibrary.wiley.com/doi/10.3322/caac.21332/epdf.

[2] M.G. Marmot, D.G. Altman, D.A. Cameron, J.A. Dewar, S.G. Thompson, The benefits and harms of breast cancer screening: an independent review, Lancet 380 (2012) 1778–1786. https://doi.org/10.1016/S0140-6736(12)61611-0.

[3] M.S. Bae, W.K. Moon, J.M. Chang, H.R. Koo, W.H. Kim, N. Cho, A. Yi, B.L. Yun, S.H. Lee, M.Y. Kim, E.B. Ryu, M. Seo, Breast cancer detected with screening US: reasons for non-detection at mammography, Radiology 270 (2014) 369–377 https://doi.org/10.1148/radiol.13130724.

[4] C. Dromain, F. Thibaul, F. Diekmann, E.M. Fallenberg, R.A. Jong, M. Koomen, R.E. Hendrick, A. Tardivon, A. Toledano, Dual-energy contrast-enhanced digital mammography: initial clinical results of a multireader, multicase study, Breast Canc. Res. 14 (2012). R94, https://doi.org/10.1186/bcr3210.

[5] T. Knogler, P. Homolka, M. Hoernig, R. Leithner, G. Langs, M. Waitzbauer, K. Pinker, S. Leitner, T.H. Helbich, Application of BI-RADS descriptors in contrast-enhanced dual-energy mammography: comparison with MRI, Breast Care 12 (2017) 212–216, https://doi.org/10.1159/000478899.

[6] G.L.G. Menezes, F.M. Knuttel, B.L. Stehouwer, R.M. Pijnappel, M.A.A.J. van den Bosch, Magnetic resonance imaging in breast cancer: a literature review and future perspectives, World J. Clin. Oncol. 5 (2) (2014) 61–70, https://doi.org/10.5306/wjco.v5.i2.61.

[7] N. Cabioglu, M.S. Yazici, B. Arun, K.R. Broglio, G.N. Hortobagyi, J.E. Price, A. Sahin, CCR7 and CXCR4 as novel biomarkers predicting axillary lymph node metastasis in T1 breast cancer, Clin. Canc. Res. 11 (2005) 5686–5694, https://doi.org/10.1158/1078-0432.CCR-05-0014.

[8] M.C.U. Cheang, D. Voduc, C. Bajdik, S. Leung, S. McKinney, S.K. Chia, C.M. Perou, T.O. Nielsen, Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype, Clin. Canc. Res. 14 (2008) 1368–1377, https://doi.org/10.1158/1078-0432.CCR-07-1658.

[9] H. Zhao, J. Shen, L. Medico, D. Wang, C.B. Ambrosone, S. Liu, A pilot study of circulating mirnas as potential biomarkers of early stage breast cancer, PLoS One 5 (2010) e13735, https://doi.org/10.1371/journal.pone.0013735.

[10] N. Kosaka, H. Iguchi, T. Ochiya, Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis, Canc. Sci. 101 (2010) 2087–2092, https://doi.org/10.1111/j.1349-7006.2010.01650.x.

[11] J. Li, Z. Zhang, J. Rosenzweig, Y.Y. Wang, D.W. Chan, Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer, Clin. Chem. 48 (2002) 1296–1304. http://clinchem.aaccjnls.org/content/48/8/1296.

[12] Y. Sun, M. Wang, G. Lin, S. Sun, X. Li, J. Qi, J. Li, Serum microRNA-155 as a potential biomarker to track disease in breast cancer, PLoS One 7 (2012), https://doi.org/10.1371/journal.pone.0047003 e47003.

[13] N.P. Schaub, K.J. Jones, J.O. Nyalwidhe, L.H. Cazares, I.D. Karbassi, O.J. Semmes, E.C. Feliberti, R.R. Perry, R.R. Drake, Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients, J. Am. Coll. Surg. (2009) 208970–208978. https://doi.org/10.1016/j.jamcollsurg.2008.12.024.

[14] W. Chen, F. Cai, B. Zhang, Z. Barekati, X.Y. Zhong, The level of circulating miRNA-10b and miRNA-373 in detecting lymph node metastasis of breast cancer: potential biomarkers, Tumor Biol. 34 (2013) 455–462, https://doi.org/10.1007/s13277-012-0570-5.

[15] J. Li, R. Orlandi, C.N. White, J. Rosenzweig, J. Zhao, E. Seregni, D. Morell, Y. Yu, X.-Y. Meng, Z. Zhang, N.E. Davidson, E.T. Fung, D.W. Chan, Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry, Clin. Chem. 51 (2005) 2229–2235, https://doi.org/10.1373/clinchem.2005.052878.

[16] A. Krilaviciute, J.A. Heiss, M. Leja, J. Kupcinskas, H. Haick, H. Brenner, Detection of cancer through exhaled breath: a systematic review, Oncotarget 6 (2015) 38643–38657, https://doi.org/10.18632/oncotarget.5938.

[17] M. Phillips, R.N. Cataneo, B.A. Ditkoff, P. Fisher, J. Greenberg, R. Gunawardena, C.S. Kwon, F. Rahbari-Oskoui, C. Wong, Volatile markers of breast cancer in the breath, Breast J. 9 (2003) 184–191, https://doi.org/10.1046/j.1524-4741.2003.09309.x.

[18] M. Phillips, R.N. Cataneo, B.A. Ditkoff, P. Fisher, J. Greenberg, R. Gunawardena, C.S. Kwon, O. Tietje, C. Wong, Prediction of breast cancer using volatile biomarkers in the breath, Breast Canc. Res. Treat. 99 (2006) 19–21, https://doi.org/10.1007/s10549-006-9176-1.

[19] M. Phillips, R.N. Cataneo, C. Saunders, P. Hope, P. Schmitt, J. Wai, Volatile biomarkers in the breath of women with breast cancer, J. Breath Res. 4 (2010), https://doi.org/10.1088/1752-7155/4/2/026003, 026003.

[20] G. Peng, M. Hakim, Y.Y. Broza, S. Billan, R. Abdah-Bortnyak, A. Kuten, U. Tisch, H. Haick, Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nano-sensors, Br. J. Canc. 103 (2010) 542–551, https://doi.org/10.1038/sj.bjc.6605810.

[21] G. Shuster, Z. Gallimidi, A. Heyman Reiss, E. Dovgolevsky, S. Billan, R. Abdah-Bortnyak, A. Kuten, A. Engel, A. Shiban, U. Tisch, H. Haick, Classification of breast cancer precursors through exhaled breath, Breast Canc. Res. Treat. 126 (2011) 791–796.

[22] L. Kou, D. Zhang, D. Liu, A novel medical E-Nose signal analysis system, Sensors 17 (2017) 402–417, https://doi.org/10.1007/s10549-010-1317-x.

[23] P. Elia, S. Raizelman, E. Katorza, Y. Matana, O. Zeiri, Z. Boger, M. Cervelli, P. Mariotini, C. Vallone, F. Signore, Y. Zeiri, Biomarkers for the detection of pre-cancerous stage of cervical dysplasia, J. Mol. Biomarkers Diagn. 6 (2015), https://doi.org/10.4172/2155-9929.1000255, 1000255.

[24] C. Jurado, M.P. Giménez, T. Soriano, M. Menéndez, M. Repetto, Rapid analysis of amphetamine, methamphetamine, MDA, and MDMA in urine using solid-phase microextraction, direct on-fiber derivatization, and analysis by GC-MS, J. Anal. Toxicol. 24 (2000) 11–16, https://doi.org/10.1093/jat/24.1.11.

[25] B. Buszewski, M. Kesy, T. Ligor, A. Amann, Human exhaled air analytics: biomarkers of diseases, Biomed. Chromatogr. 21 (2007) 553–566, https://doi.org/10.1002/bmc.835.

[26] A. Chen, S. Chatterjee, Nanomaterials based electrochemical sensors for biomedical applications, Chem. Soc. Rev. 42 (2013) 5425–5438, https://doi.org/10.1039/C3CS35518G.

[27] G. Peng, U. Tisch, O. Adams, M. Hakim, N. Shehada, Y.Y. Broza, S. Billan, R. Abdah-Bortnyak, A. Kuten, H. Haick, Diagnosing lung cancer in exhaled breath using gold nanoparticles, Nat. Nanotechnol. 4 (2009) 669–673, https://doi.org/10.1038/nnano.2009.235.

[28] O. Kuzmych, B.L. Allen, A. Star, Carbon nanotube sensors for exhaled breath components, Nanotechnology 18 (2007) 375502, https://doi.org/10.1088/0957-4484/18/37/375502.

[29] W.D. Wayne, C. Lee Cross, Biostatistics: a Foundation for Analysis in the Health Sciences, tenth ed., 1995. ISBN-13: 978–1118302798.

[30] W. Svante, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1987) 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.

[31] 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.

[32] A. Ajith, Artificial Neural Networks Handbook of Measuring System Design, 2005, https://doi.org/10.1002/0471497398.mm421.

[33] Z. Boger, H. Guterman, Knowledge extraction from artificial neural networks models, in: Systems, Man and Cybernetics. Conf. On Computational Cybernetics and Simulation, IEEE, 1997, https://doi.org/10.1109/ICSMC.1997.633051.

[34] K. Gurney, An Introduction to Neural Networks (Routledge, Taylor & Francis), 1997. ISBN 0-203-45151-1.

[35] R. Blatt, A. Bonarini, E. Calabro, M.D. Torre, M. Matteucci, U. Pastorino, Lung cancer identification by an electronic nose based on an array of MOS sensors, in: Proceedings of International Joint Conference on Neural Networks Orlando, August 12-17, 2007, https://doi.org/10.1109/IJCNN.2007.4371167. Florida, USA.

[36] Turbo Neuron software, OPTIMAL - Industrial Neural Systems, Be'er Sheva, ISRAEL. optimal@peeron.com.

[37] L. Carmel, S. Levy, D. Lancet, D. Harel, A feature extraction method for chemical sensors in electronic noses, Sensor. Actuator. B Chem. 93 (2003) 67–76. https://doi.org/10.1016/S0925-4005(03)00247-8.

[38] P. Mitra, C.A. Murthy, S.K. Pal, Unsupervised feature selection using feature similarity, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 301–312, https://doi.org/10.1109/34.990133.