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