Repeatability Study on a Classifier for Gastric Cancer Detection from Breath Sensor Data

Turppa, Emmi; Polaka, Inese; Vasiljevs, Edgars; Kortelainen, Juha M.; Shani, Gidi; Leja, Marcis; Haick, Hossam

Published in:
2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)

DOI:
10.1109/BIBE.2019.00087

Published: 01/10/2019

Document Version
Peer reviewed version

Link to publication

Please cite the original version:
Turppa, E., Polaka, I., Vasiljevs, E., Kortelainen, J. M., Shani, G., Leja, M., & Haick, H. (2019). Repeatability Study on a Classifier for Gastric Cancer Detection from Breath Sensor Data. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)* (pp. 450-453). Article 8941778 IEEE Institute of Electrical and Electronic Engineers. https://doi.org/10.1109/BIBE.2019.00087

Repeatability Study on a Classifier for Gastric Cancer Detection from Breath Sensor Data

Emmi Turppa[1], Inese Polaka[2], Edgars Vasiljevs[2], Juha M. Kortelainen[1], Gidi Shani[3], Marcis Leja[2], and Hossam Haick[3]

[1] VTT Technical Research Centre of Finland Ltd, Tampere, Finland
[2] Institute of Clinical and Preventive Medicine, University of Latvia, Riga, Latvia
[3] Department of Chemical Engineering and the Russel Berrie Nanotechnology Institute, Technion, Haifa, Israel

e-mail: emmi.turppa@vtt.fi (E. Turppa), inese.polaka@gmail.com (I. Polaka), edvasiljevs@gmail.com (E. Vasiljevs), juha.m.kortelainen@vtt.fi (J. M. Kortelainen), gidishani@gmail.com (G. Shani), cei@latnet.lv (M. Leja), hhossam@technion.ac.il (H. Haick)

This project has received funding from the *European Union's Horizon 2020 research and innovation programme (Euratom research and training programme 2014-2018)* under grant agreement No 644031.

DOI: 10.1109/BIBE.2019.00087
Originally published in 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)
Date of conference: 28-30 Oct 2019.
Date added to IEEE Xplore: 26 Dec 2019.

Abstract

The SNIFFPHONE device is a portable multichannel gas sensor, aiming to detect gastric cancer (GC) from breath samples. It employs gold nanoparticle (GNP) sensors reacting to volatile organic compounds (VOCs) in the exhaled breath, a non-invasive technique to support early diagnosis. This study evaluates the repeatability of the SNIFFPHONE classification result for measurements conducted on healthy subjects over a short period of time of less than 10 minutes. Due to the portable nature of the device, repeatability is studied with respect to varying measurement location. We find the classification results repeatable with a statistically significant 81 % Pearson correlation coefficient, even though the raw sensor responses are not concluded repeatable.

Keywords

breath sensor, cancer detection, decision support for health, gastric cancer, volatile organic compounds

Introduction

Several types of cancer produce their own recognizable set of volatile organic compounds (VOCs) that escape the body through, e.g., exhalation [1]. Therefore, it can be possible to identify a cancer non-invasively through a breath sample. Previous studies have already displayed promising results for detecting gastric cancer (GC) from breath samples [2–6]. One novel device relying on this concept is the SNIFFPHONE device, still under development at the time of writing [7].

The SNIFFPHONE is a non-invasive portable device, providing user-friendly and cost-effective decision support that can interact with a user via their mobile phone. It employs gold nanoparticle (GNP) sensors, where a particular mixture of VOCs adhere to ligand-coated GNPs, and alter the resistance of the system [2]. This technique promotes earlier diagnosis as compared to the conventional diagnostic methods that depend on observing a visible malformation [8]. The SNIFFPHONE device is expected to improve gastric cancer survival rates and reduce the number of unnecessary, invasive clinical tests. As a diagnostic device, the reliability of the measurements crucially affects its potential for being adopted into use by healthcare professionals. Consistent reproducibility of the classification results for diagnostic support is an essential prerequisite for successful uptake of the device in medical use.

In this paper, we study repeatability considering the possibility of using the SNIFFPHONE device at different locations. Essentially, a measurement given by a specific individual should be repeatable to a reasonable extent within a short time frame, i.e., within such a period that no considerable changes occur in the actual VOC content of the study subject's breath. In this study, the measurements are repeated within less than 10 minutes. The VOC content of the ambient air usually varies at different locations and is partly conveyed into the breath analyser with the exhaled air. Hence, changes in the sensor responses can be expected, yet they should not affect the result of the diagnostic test.

The main object of this study is to evaluate the repeatability of the SNIFFPHONE classifier result for each pair of samples via correlation analysis. The SNIFFPHONE classifier employs linear discriminant analysis to identify cancer cases. It transforms the sensor responses into a single measure describing the probability of the user having gastric cancer. The SNIFFPHONE classifier has been trained on a data set obtained in clinical trials on 196 subjects, including both control cases and verified cancer cases. In this work, we study data obtained from a separate group of twenty subjects.

Material and Methods

The SNIFFPHONE Measurement Protocol

As the sensors' gold nanoparticles coated with chemoresistive nanomaterial react to the breath sample, the resistance alterations of eight distinct sensor signals are collected. The coatings are chemically diverse, reacting to different VOC mixtures. Importantly, the breath sample contains a residue of the ambient air VOCs, which do not contribute to identifying a cancer yet affect the sensor responses. Therefore, the SNIFFPHONE sampling procedure includes two parts: the baseline measurement and the breath measurement. The baseline measurement collects a sample of the ambient air. It is purely to find the VOCs of the measurement environment, so that their effect can be distinguished from that of the breath sample VOCs.

In each measurement, a pump integrated within the SNIFFPHONE device transfers the gas sample into a concealed gas chamber, where the GNP sensors are located. The chamber is closed as the pumping finishes, and the baseline or breath sample is measured for 15 or 30 seconds, respectively. The sampling interval is 839 milliseconds. After completing the measurement, the chamber is flushed with the ambient air to remove residual breath sample VOCs.

Study Data

For the study data, twenty healthy subjects of ages 25 to 55 were recruited. The data set contains two samples per subject collected successively at two distinct locations, the transition taking 2 to 6 minutes. The samples were collected indoors, in Latvia. The subjects were not given any specific instructions regarding food consumption before the test. In most cases, the subject had eaten shortly before the measurement.

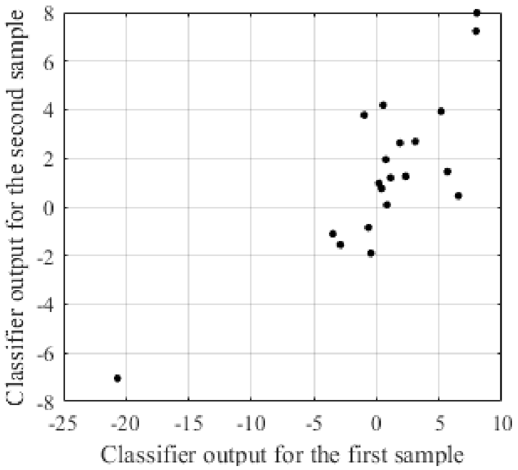

Fig. 1. Correlation between the linear outputs of the SNIFFPHONE classifier for the first and second samples of a subject, when the two samples were given at different locations. The Spearman correlation for this data is 66.6 % and Pearson correlation 81.0 %.

The study data were collected in the presence of an instructor, overseeing the collection of each measurement. All participants signed a form of consent.

Analysis Methods

To assess repeatability, we evaluate the similarity between two samples collected from each of the study subjects. First, we analyse the similarity of the SNIFFPHONE classifier output for the paired samples. We use Pearson's correlation coefficient to assess the linear correlation between paired samples, and Spearman's rank correlation coefficient to measure monotonic correlation. To assess the correlation p-values, a significance level α of 0.05 is selected considering the small sample sizes.

The SNIFFPHONE classifier employs linear discriminant analysis to identify cancer cases, applying a logistic function to transform the linear output to values describing the probability of gastric cancer. We use the classifier's linear output to study repeatability in this paper. The SNIFFPHONE classifier has been trained on a separate data set of 196 subjects, 23 % of which were verified to carry gastric cancer. The training data were obtained in clinical trials in Latvia.

Additionally, the similarity of the raw data in the paired samples is evaluated. For this purpose, we summarize the raw response signals by their minimum, mean, and maximum values. For the *i*th measurement and sensor channel *s*, we define, respectively,

$$f_{min}(i,s) = (\min(x_{i,s}) - \overline{y_{i,s}}) / \overline{y_{i,s}} , \qquad (1)$$

$$\overline{f}(i,s) = (\overline{x_{i,s}} - \overline{y_{i,s}}) / \overline{y_{i,s}} , \qquad (2)$$

$$f_{max}(i,s) = (\max(x_{i,s}) - \overline{y_{i,s}}) / \overline{y_{i,s}} , \qquad (3)$$

where $x_{i,s}$ contains the sensor response to the *i*th breath sample on the sensor channel *s*, and $y_{i,s}$ the corresponding baseline measurement. These features represent the relative change of the breath measurement with respect to the average value over the baseline measurement.

The repeatability of the summarized raw data is evaluated using Spearman's rank correlation coefficient. We assess repeatability individually for each of the eight signals, originating from eight distinct sensor channels. These eight signals provide the input to the SNIFFPHONE classifier.

Results

Fig. 1 presents the correlation between the output values given by the SNIFFPHONE classifier to the paired samples. It should be noted that the shown classifier output values are before applying a non-linear logistic function, which would finally transform the results into probability values. Strikingly, the Spearman correlation for the presented data is 66.6 % with a p-value of 1.8×10^{-3}, implying good and statistically significant monotonic correlation. Furthermore, the Pearson correlation is a strong 81.0 % with a p-value of 1.5×10^{-5} for significant linear correlation. Altogether, the SNIFFPHONE classifier assigned 17 of 20 subjects into the same class on both instances. Essentially, this implies location-independent repeatability for the diagnostic support provided by the SNIFFPHONE device.

As for the raw data, the measurement temperature between the paired samples varied slightly, with a mean absolute difference of 0.50 ± 0.69 degrees Celsius. The first principal component of the raw data, explaining 67.5 % of the data,

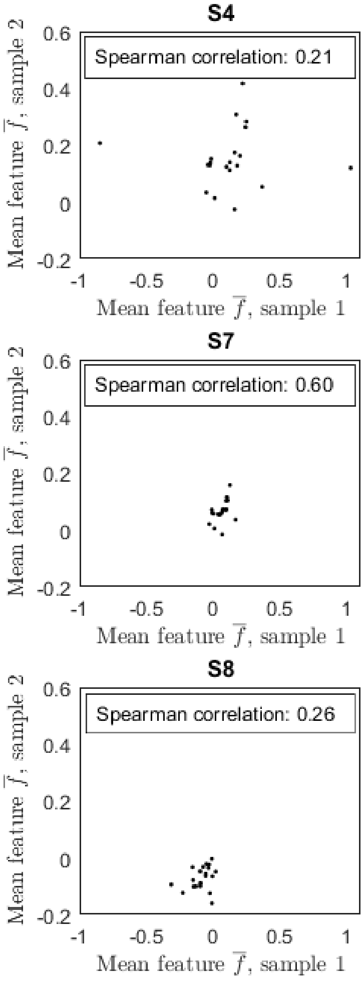

Fig. 2. Visualized correlation between the paired samples for selected sensor channels four (top, S4), seven (middle, S7), and eight (bottom, S8). Here we use the mean features to summarize raw signals.

has a Spearman correlation coefficient of 0.053 when compared with respect to the temperature differences, implying that the data are not significantly affected by temperature variations, although using the second principal component does reveal a 26.9 % correlation for 28.4 % of the data.

The Spearman correlation coefficients between the raw features of the paired samples vary between 0.09 and 0.60 for the different sensor channels, with the lowest coefficient corresponding to sensor channel six. The most promising repeatability result is given by sensor channel seven, while other correlations are below 50 %. Fig. 2 visualizes the paired sample features for channels four, seven, and eight, together with the Spearman correlation coefficients. The correlation p-values are presented in Table I. The p-values confirm that only sensor channel seven exhibits statistically significant correlation. The presented results have been computed with the mean features, which gave the most critical results.

TABLE I Spearman correlation of paired samples in each sensor channel

	Sensor channel							
Spearman's correlation	**S1**	**S2**	**S3**	**S4**	**S4**	**S6**	**S7**	**S8**
Spearman's rho	0.38	0.44	0.33	0.21	0.34	0.09	**0.60**	0.26
p-value	0.103	0.054	0.152	0.371	0.145	0.691	**0.007**	0.275

Overall, the results do not support channel-wise measurement repeatability of raw data. However, the measurement results given by the individual sensor channels are not used as such for diagnostic support, and our results do support the repeatability of the classification result despite varying location.

Conclusion

The paired samples in this study were collected at two separate locations within a few minutes of each other. We were therefore able to presume that no significant changes in the real VOC content of the breath would occur within the studied time span. In contrast, changes in the raw sensor responses might have been anticipated due to the changing VOC content of the ambient air. Relevantly, we did not find significant correlation between raw data features of the paired samples, except for one sensor channel.

Importantly, Spearman correlation of 66.6 % and Pearson correlation of 81.0 % was obtained with the SNIFFPHONE classifier, which implies repeatability of the SNIFFPHONE classification result regardless of measurement location. Significantly, this also denotes location-independent repeatability of the measure used for diagnostic support.

The SNIFFPHONE sensor responses are based on resistivity and are therefore affected by the ambient temperature. A heating unit will be integrated into the following SNIFFPHONE prototype, under development at the time of writing, to maintain a constant temperature in all measurements. This will allow a broader repeatability study with respect to additional variables as it removes the bias introduced by temperature variations between repeated samples. This bias would not be straightforward to fix in already acquired data without additional measurements specifically designed for the purpose. This is why we leave broader repeatability studies to be carried out with the next prototype. In future studies, it would be advisable to use larger subject groups.

Acknowledgment

The investigators acknowledge the biobank infrastructure of the Institute of Clinical and Preventive Medicine, University of Latvia and Riga East University hospital as well as the infrastructure of Riga East University hospital, Digestive Diseases Centre GASTRO along with Latvia University that allowed the recruitment of study subjects.

The authors thank Prof. M. van Gils of VTT Technical Research Centre of Finland Ltd for the guidance in preparing this paper.

References

[1] M. K. Nakhleh et al., “Diagnosis and classification of 17 diseases from 1404 subjects via pattern analysis of exhaled molecules,” *ACS Nano*, vol. 11, pp. 112–125, Dec 2016.

[2] H. Amal etl al., “Detection of precancerous gastric lesions and gastric cancer through exhaled breath,” *Gut*, vol. 65, pp. 400–407, Apr 2015.

[3] Z-q Xu et al., “A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions,” *British Journal of Cancer*, vol.108, pp. 941–950, Mar 2013.

[4] C. M. Durán-Acevedo et al., “Exhaled breath analysis for gastric cancer diagnosis in Colombian patients,” *Oncotarget*, vol. 9, pp. 28805–28817, June 2018.

[5] S. Kumar et al., “Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esphago-gastric cancer,” *Analytical Chemistry*, vol. 85, pp. 6121–6128, May 2013.

[6] Y. Chen et al., “Breath analysis based on surface-enhanced Raman scattering sensors distinguishes early and advanced gastric cancer patients from healthy persons,” *ACS Nano*, vol. 10, pp. 8169–8179, July 2016.

[7] “SNIFFPHONE, Smart phone for disease detection from exhaled breath”, *sniffphone.eu*, 2019. [Online]. Available: https://www.sniffphone.eu. [Accessed: Sep. 4, 2019].

[8] W. K. Leung et al., “Screening for gastric cancer in Asia: current evidence and practice,” *The Lancet Ontology*, vol. 9, pp. 279–287, Mar 2008.