

Breath Analysis for Detecting Diseases on Respiratory, Metabolic and Digestive System

Lu Kou¹, David Zhang^{1,2*}, Jane You¹, Yingbin Jiang²

¹Biometrics Research Center, Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China; ²Department of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China

Correspondence to: David Zhang, *csdzhang@comp.polyu.edu.hk

Keywords: E-Nose, Chemical Sensors, Breath Analysis

Received: November 26, 2018

Accepted: January 26, 2019

Published: January 29, 2019

Copyright © 2019 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

ABSTRACT

Recently, biological technology and computer science are of great importance in medical applications. Since one's breath biomarkers have been proved to be related with diseases, it is possible to detect diseases by analysis of breath samples captured by e-noses. In this paper, a novel medical e-nose system specific to disease diagnosis was used to collect a large-scale breath dataset. Methods for signal processing, feature extracting as well as feature & sensor selection were discussed for detecting diseases on respiratory, metabolic and digestive system. Sequential forward selection is used to select the best combination of sensors and features. The experimental results showed that the proposed system was able to well distinguish healthy samples and samples with different diseases. The results also showed the most significant sensors and features for different tasks, which meets the relationship between diseases and breath biomarkers. By selecting best combination of different sensors and features for different tasks, the e-nose system is shown to be helpful and effective for diseases diagnosis on respiratory, metabolic and digestive system.

1. INTRODUCTION

Traditional diagnosis methods include blood, urine tests and some other methods. Nowadays, biological technology and computer science are playing their roles in medical applications. Technic developments in bioimaging [1], biochips [2] and biosensors [3] have brought new aids in disease diagnosis.

Breath analysis is a way to detect some diseases by the examination of certain compounds in human breath, which is largely composed of oxygen, carbon dioxide, water vapor and nitric oxide, as well as less than 100 ppm (parts per million) of mixture with over 500 kinds of components. The components include carbon monoxide, methane, hydrogen, acetone and numerous volatile organize compounds (VOCs) [4, 5].

Table 1 summarizes the concentrations of typical compositions from human breath [6].

A breathing process includes three stages. The first stage is the exchange of gases between the outside air and the alveoli and between the alveoli and the blood in pulmonary capillary. The second is the exchange between oxygen and carbon dioxide in blood during gas transportation in the blood. The third is the exchange of gases between blood and tissue cells. During this process, endogenous molecules produced by metabolic processes are separated from blood and enter into the alveolar air via the alveolar pulmonary membrane [6-8] and thus into the exhaled breath. Variation in the concentration of these molecules can suggest various diseases or at least changes in metabolism [9]. For instance, nitric oxide in breath can be measured as an indicator of asthma or other conditions characterized by airway inflammation [10]. Increased pentane and carbon disulfide have been observed in the breath of patients with schizophrenia [11]. Breath concentration of volatile organic compounds (VOCs) such as cyclododecatriene, benzoic acid, and benzene is much higher in lung cancer patients than in control groups [12]. Acetone has been found to be more abundant in the breath of diabetics [13, 14], and breath ammonia is significantly higher in patients with renal diseases [15]. These molecules are considered as biomarkers of the presence of diseases and clinical conditions. Much can be learnt from them about the overall state of an individual's metabolism or physical condition.

Compared with these methods, breath analysis has many advantages [16]. Firstly, breath analysis is a non-invasive method, and it causes least harm to both the subjects and the personnel who collect the samples. Secondly, its result can be obtained immediately. Thirdly, the sample collection is quite easy for a subject, since the only requirement to collect a breath sample is that the subject must be breathing. Therefore, increasing interest has been expressed about the applications of breath analysis in medicine and clinical pathology [17, 18]. However, most of the existing trials [19] on breath diagnosis only focus on very limited kinds of diseases. The reason may be complicated. Lack of specific breath analysis system and method, as well as a large enough dataset will all block the process of research. Also, some of the diseases may not be so related to the breath system, which will not give a satisfactory result. Basing on the process of gas exchange in breath, we can see that diseases of respiratory system and metabolic system will strongly affect one's exhaled breath. On the other hand, since the respiratory passage is connected to digestive system, it may be possible to detect diseases of digestive system by breath analysis. We thus collected a breath analysis dataset with both healthy samples and samples with different kinds of diseases of respiratory, metabolic and digestive system by a specific e-nose system. Experiments were organized on the collected dataset to discover a proper method of breath analysis for disease diagnosis.

2. LITERATURE REVIEW

2.1. Breath Biomarker and Diseases

Nowadays, the concentration of some biomarkers in breath has been proven to be related with certain diseases. For example, lung diseases may alter volatile organic compounds (VOCs) in breath because both Mycobacteria and oxidative stress resulted from Mycobacterial infection generate distinctive VOCs. And digestive system diseases will reflect in hydrogen of ones' exhaled breath because sugar could not be fully digested and would be decomposed or fermented and produce more hydrogen. By selecting proper sensors that can respond to the components, it is possible to analyze a person's breath odor and thus his or her health state. A few examples will further prove these points. The level of nitric oxide can be used as a diagnostic for asthma [20]. Patients with renal disease have higher concentrations of ammonia [21]. The concentration of VOCs, such as cyclododecatriene, benzoic acid, and benzene is much higher in lung cancer patients [22]. **Table 2** lists the relationship between biomarkers and some typical diseases.

2.2. Breath Analysis with E-Nose

Currently, the measurement of exhaled breath is usually performed by two common gas analysis apparatuses, gas chromatography (GC) [26] or electronic nose (e-nose) [27]. GC can separate and identify molecules that are responsible for typical odors occurring in specific diseases, which is very accurate for

Table 1. Typical compositions of human breath.

Concentration	Molecules
percentage	oxygen, water, carbon dioxide
parts-per-million	acetone, carbon monoxide, methane, hydrogen, isoprene, benzenemethanol
parts-per-billion	formaldehyde, acetaldehyde, 1-pentane, ethane, ethylene, other hydrocarbons, nitric oxide, carbon disulfide, methanol, carbonyl sulfide, methanethiol, ammonia, methylamine, dimethyl sulfide, benzene, naphthalene, benzothiazole, ethane, acetic acid

Table 2. Breath biomarkers and related diseases.

Diseases	Breath Biomarkers
diabetes [6]	acetone
kidney disease [23]	ammonia
lung disease [12]	benzene, 1,1-oxybis-, 1,1-biphenyl, 2,2-diethyl, furan, 2,5-dimethyl-, etc.
Respiratory disease [24]	pentane, nitric oxide, carbon monoxide
digestive system disease [25]	hydrogen

disease identification. But this kind of apparatus is expensive and not portable. Its sampling and assaying processes are complicated and time consuming (about one hour for one sample), and its results require expert's interpretation [28]. Therefore, it is hard to use such apparatus as a domestic or clinical tool.

Electronic noses, or e-noses, are devices that "smell" or detect odor. An e-nose consists of a mechanism for chemical detection, such as an array of electronic sensors, and a mechanism for processing. Using e-nose is a less expensive and more portable way for breath analysis. Recently, e-nose has gradually been used in medicine for the diagnosis of renal disease [29], diabetes [30], lung cancer [31], and asthma [32]. Though all of these methods work satisfactorily in breath analysis, their results could possibly be improved. That is because, commercial e-noses, with their marketing concerns, have to provide versatility in applications, such as coffee, wine, and fragrances identification. The versatility, however, limits their performance in disease detection since their sensor selection has to match broad applications.

The idea of e-nose is inspired by the mechanisms of human olfaction. In general, basic elements of an e-nose system include an "odor" sensor array, a data preprocessor, and a pattern recognition engine [33]. Among them, the sensor array, like signal receptors, is the key part of e-nose. The application of sensor array on odor recognition was demonstrated firstly by Persaud and Dodd in 1982 [34]. Currently, e-nose has been much developed and used to fulfill a large number of industrial needs, such as food, chemistry, fragrances, security, and environment [35]. Recently, the feasibility of using e-noses for monitoring the health of human and diagnosing diseases in an early stage has been demonstrated [36].

As early as 1997, Wang *et al.* [37] designed an e-nose with one SnO₂ thin film sensor for diabetes diagnosis. The authors tested their device by using the breath samples collected from 18 patients and 14 healthy persons. The concentration of blood sugar of the subjects was used as reference. The results showed that the e-nose was able to diagnose diabetes with a sensitivity of 77.8% and a specificity of 35.7%.

In 2001, Lin *et al.* [36] reported a study about the application of e-nose with six quartz crystal sensors to detect renal diseases. Discriminant Analysis (DA) was carried out to analyze the sensor signals. The clinical test result showed that the e-nose could discriminate the breath samples from 30 normal subjects, 83 uremia patients, and 61 chronic renal disease patients with a total correct classification of 86.78%.

In 2009, Ogorodnik *et al.* [38] analyzed VOCs from a breath sample of a patient with different lung diseases by using an e-nose with ten MOSFET sensors and four SnO₂ sensors. In total, 66 individuals – 23 with asthma, 3 with chronic obstructive pulmonary disease (COPD), 12 with pneumonia, 13 with lung cancer, 4 in the past operation state (removed lung cancer), and 11 healthy volunteers were tested at two different times and ANN analysis was employed to classify the samples of cancer and other lung diseases. The results showed that the e-nose could identify lung cancer with 100% accuracy, healthy subjects with 100% accuracy, and asthma with 82.6% accuracy.

In 2010, Guo *et al.* [39] proposed a method of monitoring the blood glucose levels of diabetics via measuring the concentration of breath acetone. A SVM classifier was used to evaluate the accuracy of classifying the samples into the groups with different blood glucose levels. The results indicated that the system was not only able to distinguish between breath samples from patients with diabetes and healthy subjects, but also to represent the fluctuation of blood glucose of diabetics. In the same year, Guo *et al.* [40] improved accuracy of diabetes condition monitoring by using a SRC method. Coupled with SRC, the system was able to classify these levels with a much better accuracy than the accuracy reported in [41].

In 2017, Yan *et al.* [42] proposed a domain learning method using domain features and independence maximization. A logic regression classifier was used to evaluate the accuracy on diabetes, chronic kidney disease and cardiopathy. The results showed that the method can improve the classifying performance by reducing the instrumental variation and time-varying drift.

Table 3 concludes some existing trials on medical breath analysis with e-noses.

3. BREATH ANALYSIS

3.1. Breath Dataset

To evaluate the performance of our device, a large-scale breath dataset was collected. We cooperated with Guangzhou Hospital of Traditional Chinese Medicine and collected data from inpatient volunteers with a specific e-nose device [43].

In the device, there were 12 different odor sensors selected based on the relationship between breath biomarkers and diseases into consideration. The sensors in the device should be sensitive to the VOCs, carbon dioxide, humidity, and temperature. Thus, a sensor array with 11 sensors was optimized for the purpose of detecting one's breath. The sensor array included six ordinary metal oxide semiconductor (MOS) sensors, three temperature modulated MOS sensors, a carbon dioxide sensor, and a temperature-humidity sensor. Specifically, the temperature-humidity sensor had two input channels for temperature and humidity respectively. Therefore, there were, in total, 12 input channels. The model, manufacturer and function of the sensors are listed in **Table 3**. The suffix “-TM” indicates a temperature-modulated sensor.

For each sample, we first collected the patient's breath and recorded the signals. The diagnosis was then given by an authoritative doctor as the classification labels. Then some biochemical indicators, such as blood glucose, blood pressure and blood lipids were collected. Finally, in this dataset, there were in total over 10,000 samples of 47 classes, including 1491 healthy samples and samples of 46 different kinds of diseases. In this paper, a subset of healthy samples and samples for six kinds of diseases are selected for experiments, including breast disease, cardiopathy, diabetes, lung disease, kidney disease and gastritis. **Table 4** shows the number of each class used in selected subset.

All the samples were collected from hospitals in Guangzhou. However, since most of the healthy samples were provided by medically-examined young people while disease samples by elder patients, it is difficult to make age-matched subsets, which is a limitation of this dataset.

3.2. Preprocessing

Before analyzing data, original signals should be preprocessed so as to be transformed into standard samples. Four steps were taken: faulty signal removal, de-noising, baseline manipulation and normalization.

Table 3. Trials on medical breath analysis with e-noses.

Year	Researcher	Disease	Data size	Accuracy
1997	Wang <i>et al.</i>	Diabetes	32	77.8%
2001	Lin <i>et al.</i>	Kidney	174	86.78%
2009	Ogorodnik <i>et al.</i>	Lung cancer	66	100%
		Diabetes	225	87%
2011	Guo <i>et al.</i>	Kidney disease	218	86%
		Digestive disease	218	70%
		Diabetes	431	85.29%
2017	Yan <i>et al.</i>	Chronical kidney disease (CKD)	340	80.18%
		Cardiopathy	97	91.67%

Table 4. Number of samples in each class.

Type	Class	Data size
Healthy	Healthy	1291
	Lung	498
Respiratory system	Respiratory	130
	Diabetes	585
Metabolic system	Kidney	473
	Digestive system	310

A faulty signal is a common problem in devices with sensors. In our system, causes of faulty signals were complicated, including mis-operation, bad connection and device damage. In order to make the system more robust, these signals should be removed before analysis.

De-noising aims to remove the noise from the original signals by utilizing a low-pass filter since the signal is mainly interfered by high-frequency noise.

The purpose of baseline manipulation is to compensate baseline drift. The baseline value is the average response in the baseline stage of each sensor. The value is then subtracted from the whole response curve to eliminate the interference of background noise of the sensors [44]. Assume that for each sensor transient of each sample, there are k dimensions, where $k = 1, \dots, N_k$, and b dimensions in the baseline stage, where $b = 1, \dots, N_b$. The response at time t_k is denoted as $R(t_k)$. The baseline response is $B(t_b)$. Then baseline manipulation can be computed as:

$$R^B(t_k) = R(t_k) - \frac{1}{N_b} \sum_{t_b=1}^{N_b} B(t_b) \quad (1)$$

Normalization is used to compensate for sample-to-sample variations caused by analyte concentration. $R^B(t_k)$ is a sample after the baseline manipulation step, and the normalized response $R^{BN}(t_k)$ can be defined as:

$$R^{BN}(t_k) = \frac{R^B(t_k)}{\max(R^B(t_k))} \quad (2)$$

3.3. Feature Extraction

To reduce the dimension of the origin features, principal component analysis (PCA) can be used. PCA projects high-dimensional data into a low-dimensional subspace while keeping most of the data variance.

Some low-dimensional geometric features can also be extracted from the origin response curves. One of the traditional features of gas sensors is steady state response. When a gas sensor is used to sense a gas sample, its response will reach a steady state in a few minutes. The steady state response has a close relationship with the concentration of the measured gas. Therefore, the 9D feature vector contains most of the information needed for disease screening.

However, additional useful information is carried in the transient responses [45]. Transient responses are often related to the change of gas flow (injection/purge) or temperature (for TM sensors). The feature set includes magnitude, difference, derivative, second derivative, integral, slope and phase features, as well as features in frequency domain such as fast Fourier transformation (FFT) and wavelet. The extracted features in both space domain and frequency domain are described in Table 5.

4. EXPERIMENTS

4.1. Feature Selection for Different Diseases

For each class, the first 50 samples collected by the first device were chosen as the labeled training sets and the rest were test samples. Logistic regression method was adopted as the classifier. Sequential forward selection (SFS) method was used to optimize the features. SFS method is a greedy strategy. In each iteration, one feature was selected from all features that could achieve the best classification accuracy together with the features already selected. Figure 1 shows the results of forward selection in different disease diagnosis tasks. For each graph, the horizontal axis represents the number of features used and the vertical axis represents the classification accuracy. We will find the peak point of each curve and record which features are selected in the peak points.

In Table 6, we concluded the best combination of features and sensors selected for each task (the peak points in each curve of Figure 1). It can be found that wavelet features contributed most in all the tasks. Pca feature, MeanMag feature, phase feature and derivative feature differed in different tasks. Other features did not improve the performance of the system. The sensors that contributed most to meeting the relationship between diseases and breath biomarkers in each task are listed in Table 2 and Table 7. S7 (TGS2602) contributes most to metabolic diseases.

4.2. Multi-Disease Diagnosis

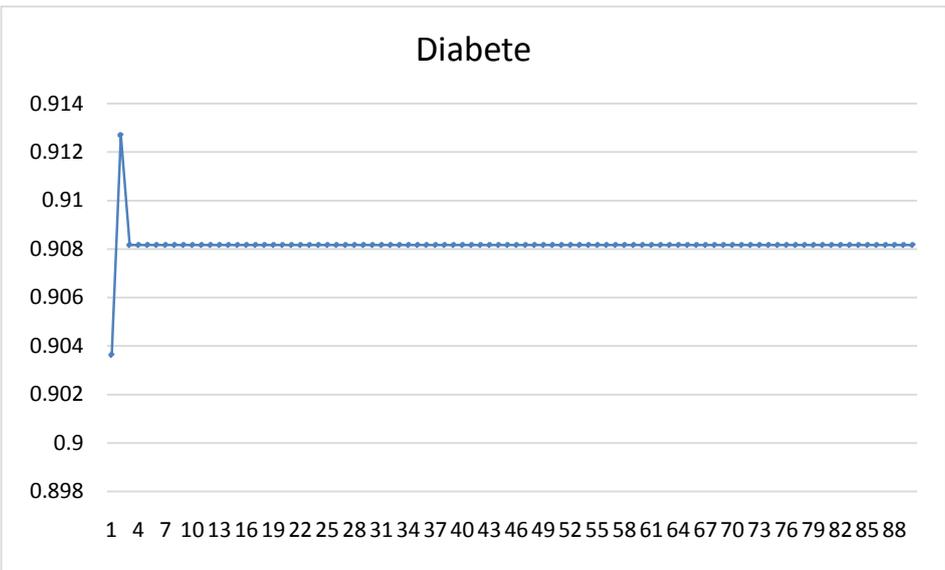
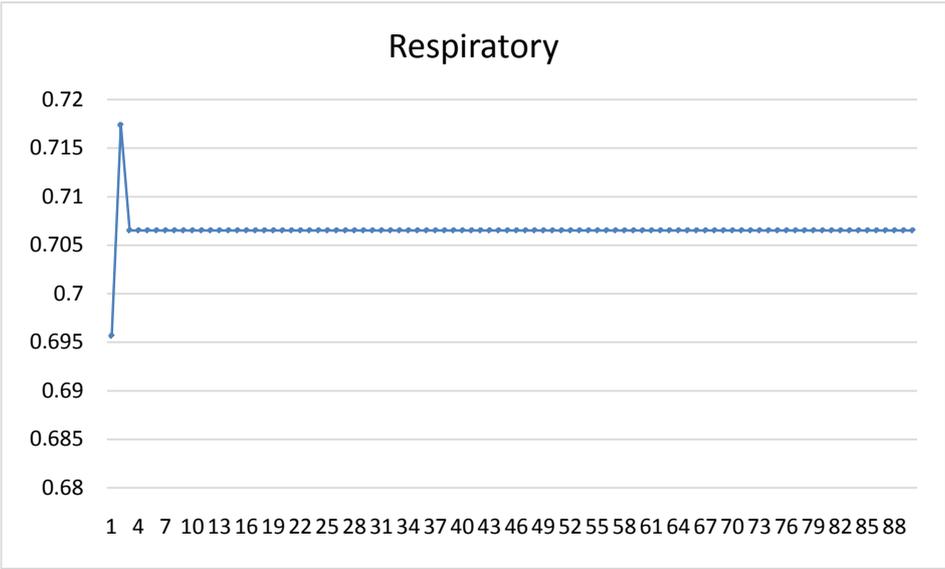
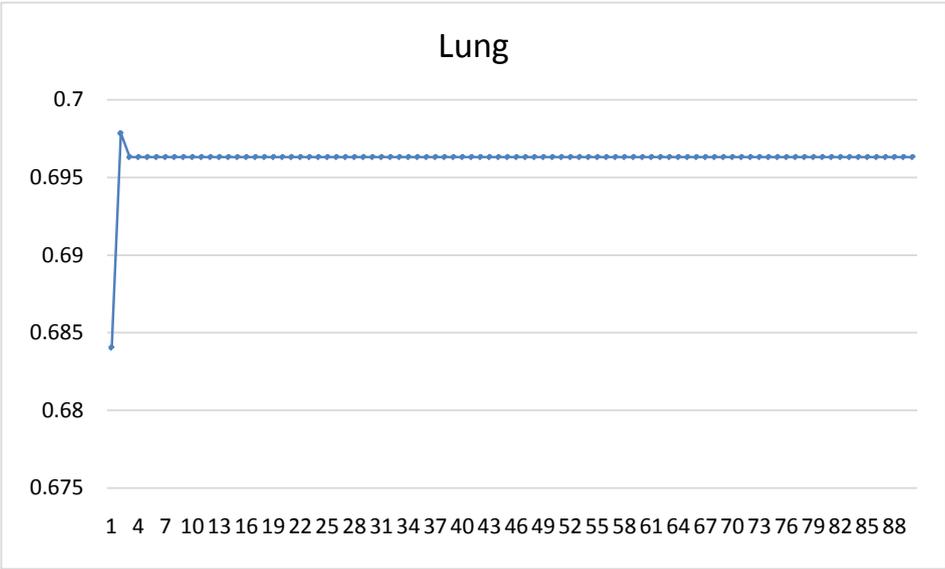
To further evaluate the performance of the system, we built a classification tree to deal with the multi-disease diagnosis problem. Synthetic Minority Over-sampling Technique (SMOTE) [46] was used in the experiment. SMOTE method is an over-sampling approach in which the minority class is over-sampled by creating “synthetic” examples rather than over-sampling with replacement.

Figure 2 shows the result of the classification tree. When an unknown sample comes, it will first be tested if it is diabetes. Then check if it is healthy or kidney disease in turn. Finally, if the sample belongs to none of these classes, it may be disease in lung, stomach or respiratory system. Accuracy of each level of the classification tree is 90%, 75%, 63% and 76%, respectively. It can be seen that metabolic diseases are easier to be separated out than diseases of respiratory system or digestive system.

However, this is only for one-label samples. For those patients with more than one disease, multi-labeled methods should be considered.

5. CONCLUSIONS

This paper discussed breath analysis method for diseases on respiratory, metabolic and digestive system.



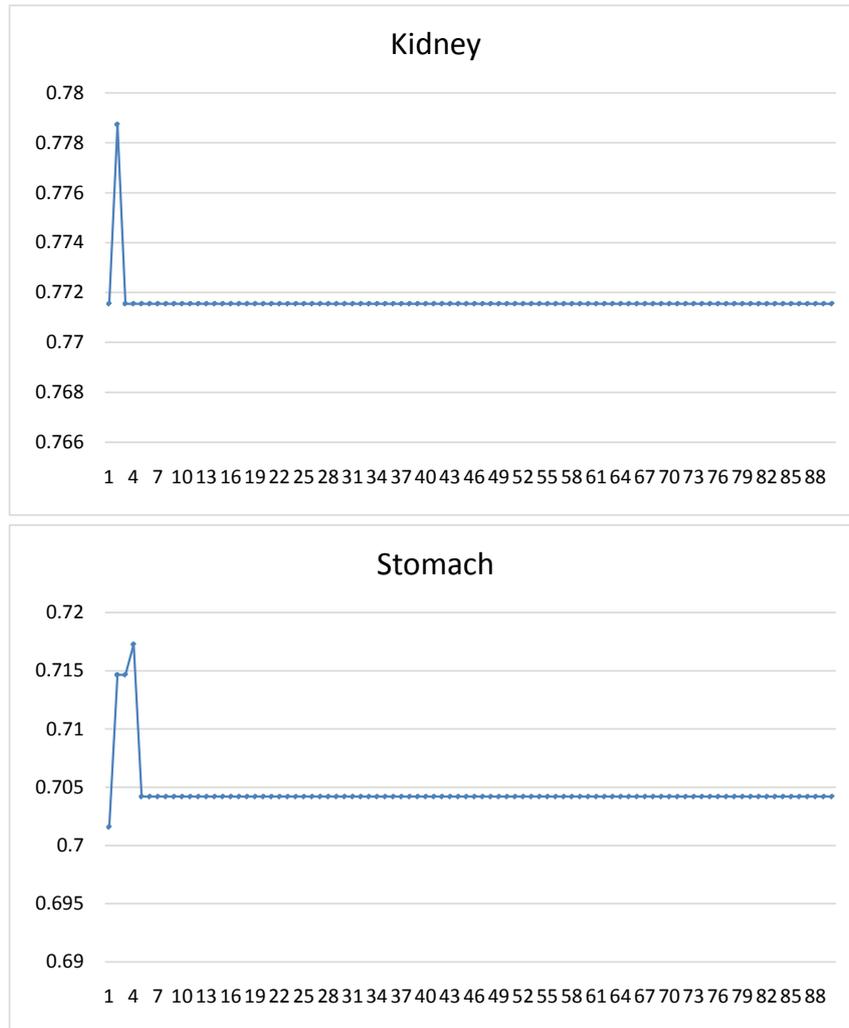


Figure 1. Forward selection result of six binary-classification tasks. For each graph, the horizontal axis represents the number of features used and the vertical axis represents the classification accuracy.

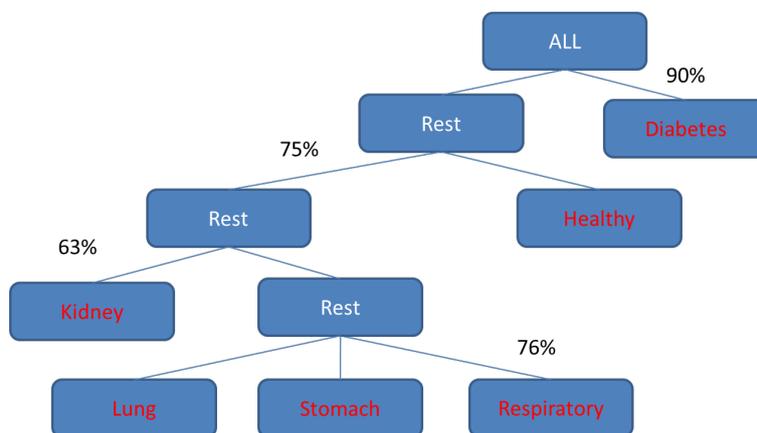


Figure 2. The optimal classification tree for multi-diseases diagnosis.

Table 5. Summary of the transient features. PCA: principal component analysis.

Feature	Characteristics
PCA	Reduced dimension of the origin features with PCA method.
Magnitude	Down-sampled values of the curve's magnitude M .
	The maximum magnitude.
	Down-sampled values of the normalized magnitude $M/\max(M)$. Mean values of the magnitude.
Derivative	Down-sampled values of the curve's derivative D .
	The maximum and minimum derivative.
Space	Second derivative
	The maximum and minimum second derivative in both the injection and purge stage.
	Integral
	The integral of the five intervals of the curve; the intervals are the same with the difference feature.
Slope	The slope of the five intervals of the curve; the intervals are the same with the difference feature.
Phase Feature	The phase feature is proposed in [44]. First, the response is transformed to the phase space, which is spanned by its magnitude and derivative. Then, the phase feature is defined as $\int_{M(t_i)}^{M(t_{i+1})} DdM$
Frequency	FFT
	Fast Fourier transformation
	Wavelet
	Wavelet transformation

Table 6. Experimental results on different disease diagnosis tasks.

Task (vs Healthy)	Accuracy	Features	
Respiratory system	Lung	0.698	Pca of S9, Wavelet of S4
	Rerspiratory	0.717	Derivative of S8, Wavelet of S6
Metabolic system	Diabetes	0.913	phase of S7, Wavelet of S4
	Kidney	0.779	Derivative of S10, Wavelet of S7
Digestive system	Stomatch	0.717	Maxmag of S4, Slope of S6 Wavelet of S6

Table 7. Summary of the sensor array VOCs: volatile organic compounds; ppm: parts per million; TM: temperature-modulated.

Channel	Model	Manufacturer	Function	Sensitivities (ppm)
1	TGS4161	Figaro Inc., Japan	CO ₂	350 - 10,000
2	TGS826	Figaro Inc., Japan	VOCs, NH ₃	30 - 5000
3	QS01	FIS Inc., Japan	VOCs, H ₂ , CO	1 - 1000
4	TGS2610D	Figaro Inc., Japan	H ₂ , VOCs	500 - 10,000
5	TGS822	Figaro Inc., Japan	VOCs, H ₂ , CO	50 - 5000
6	TGS2602-TM	Figaro Inc., Japan	VOCs, NH ₃ , H ₂ S	1 - 30
7	TGS2602	Figaro Inc., Japan	VOCs, NH ₃ , H ₂ S	1 - 30
8	TGS2600-TM	Figaro Inc., Japan	H ₂ , VOCs, CO	1 - 100
9	TGS2603	Figaro Inc., Japan	NH ₃ , H ₂ S	1 - 10
10	TGS2620-TM	Figaro Inc., Japan	VOCs, H ₂	50 - 5000
11			Temperature	
12	HTG3515CH	Humirel Inc., France	Humidity	

The sensors, features and signal preprocessing methods were introduced, and a specific medical dataset was collected. Experiments were taken on the new collected datasets to discover the best selection of sensors and features for different diseases with an SFS method. A classification tree is built on the collected data for the multi-class task.

The experimental results showed that better accuracy could be achieved by an optimal combination of features and sensors for different tasks. Wavelet feature is the most significant feature in majority of disease diagnosis tasks, while different sensors contribute differently. The most significant sensors for different diseases are just in accord with the relationship of diseases and biomarkers listed in [Table 2](#). The multi-class classification tests also produced a satisfactory output. Metabolic diseases are more likely to be diagnosed correctly than other diseases by means of breath analysis.

However, there is still more to work on how to find other diseases that may be detected from human breath, as well as multi-labeled diagnosis. Moreover, discovering the deeper relationship between one's breath and the body states such as blood glucose, blood fat, blood press and other biochemical indicator is also worthy of studying.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this paper.

REFERENCES

1. Alibolandi, M., Abnous, K., Ramezani, M., *et al.* (2014) Synthesis of AS1411-Aptamer-Conjugated CdTe Quantum Dots with High Fluorescence Strength for Probe Labeling Tumor Cells. *Journal of Fluorescence*, **24**, 1519-1529. <https://doi.org/10.1007/s10895-014-1437-5>
2. Han, H.C., Lo, H.C., Wu, C.Y., *et al.* (2015) Nano-Textured Fluidic Biochip as Biological Filter for Selective Survival of Neuronal Cells. *Journal of Biomedical Materials Research Part A*, **103**, No. 6. <https://doi.org/10.1002/jbm.a.35338>
3. Mehrotra, P. (2016) Biosensors and Their Applications—A Review. *Journal of Oral Biology and Craniofacial Research*, **6**, 153-159. <https://doi.org/10.1016/j.jobcr.2015.12.002>

4. Yan, K., Zhang, D., Wu, D., Wei, H. and Lu, G. (2014) Design of a Breath Analysis System for Diabetes Screening and Blood Glucose Level Prediction. *IEEE Transactions on Biomedical Engineering*, **61**, 2787-2795. <https://doi.org/10.1109/TBME.2014.2329753>
5. Cao, W. and Duan, Y. (2007) Current Status of Methods and Techniques for Breath Analysis. *Critical Reviews in Analytical Chemistry*, **37**, 3-13. <https://doi.org/10.1080/10408340600976499>
6. Risby, T.H. and Solga, S. (2006) Current Status of Clinical Breath Analysis. *Applied Physics B*, **85**, 421-426. <https://doi.org/10.1007/s00340-006-2280-4>
7. D'Amico, A., Di Natale, C., Paolesse, R., Macagnano, A., Martinelli, E., Pennazza, G., Santonico, M., Bernabei, M., Roscioni, C., Galluccio, G., *et al.* (2008) Olfactory Systems for Medical Applications. *Sensors and Actuators B: Chemical*, **130**, 458-465. <https://doi.org/10.1016/j.snb.2007.09.044>
8. Miekisch, W., Schubert, J.K. and Noeldge-Schomburg, G.F.E. (2004) Diagnostic Potential of Breath Analysis-Focus on Volatile Organic Compounds. *Clinica Chimica Acta*, **347**, 25-39. <https://doi.org/10.1016/j.cccn.2004.04.023>
9. Röck, F., Barsan, N. and Weimar, U. (2008) Electronic Nose: Current Status and Future Trends. *Chemical Reviews*, **108**, 705-725. <https://doi.org/10.1021/cr068121q>
10. Deykin, A., Massaro, A.F., Drazen, J.M. and Israel, E. (2002) Exhaled Nitric Oxide as a Diagnostic Test for Asthma: Online versus Offline Techniques and Effect of Flow Rate. *American Journal of Respiratory and Critical Care Medicine*, **165**, 1597-1601. <https://doi.org/10.1164/rccm.2201081>
11. Phillips, M., Sabas, M. and Greenberg, J. (1993) Increased Pentane and Carbon Disulfide in the Breath of Patients with Schizophrenia. *Journal of Clinical Pathology*, **46**, 861-864. <https://doi.org/10.1136/jcp.46.9.861>
12. Phillips, M., Altorki, N., Austin, J.H.M., Cameron, R.B., Cataneo, R.N., Greenberg, J., Kloss, R., Maxfield, R.A., Munawar, M.I., Pass, H.I., *et al.* (2007) Prediction of Lung Cancer Using Volatile Biomarkers in Breath. *Cancer Biomarkers*, **3**, 95-109. <https://doi.org/10.3233/CBM-2007-3204>
13. Deng, C., Zhang, J., Yu, X., Zhang, W. and Zhang, X. (2004) Determination of Acetone in Human Breath by Gas Chromatography-Mass Spectrometry and Solid-Phase Microextraction with On-Fiber Derivatization. *Journal of Chromatography B*, **810**, 269-275. [https://doi.org/10.1016/S1570-0232\(04\)00657-9](https://doi.org/10.1016/S1570-0232(04)00657-9)
14. Fleischer, M., Simon, E., Rumpel, E., Ulmer, H., Harbeck, M., Wandel, M., Fietzek, C., Weimar, U. and Meixner, H. (2002) Detection of Volatile Compounds Correlated to Human Diseases through Breath Analysis with Chemical Sensors. *Sensors and Actuators B: Chemical*, **83**, 245-249. [https://doi.org/10.1016/S0925-4005\(01\)01056-5](https://doi.org/10.1016/S0925-4005(01)01056-5)
15. Davies, S., Spanel, P. and Smith, D. (1997) Quantitative Analysis of Ammonia on the Breath of Patients in End-Stage Renal Failure. *Kidney International*, **52**, 223-228. <https://doi.org/10.1038/ki.1997.324>
16. Van Berkel, J., Dallinga, J.W., Moller, G.M., Godschalk, R.W.L., Moonen, E., Wouters, E.F.M. and Van Schooten, F.J. (2008) Development of Accurate Classification Method Based on the Analysis of Volatile Organic Compounds from Human Exhaled Air. *Journal of Chromatography B*, **861**, 101-107. <https://doi.org/10.1016/j.jchromb.2007.11.008>
17. Di Francesco, F., Fuoco, R., Trivella, M.G. and Ceccarini, A. (2005) Breath Analysis: Trends in Techniques and Clinical Applications. *Microchemical Journal*, **79**, 405-410. <https://doi.org/10.1016/j.microc.2004.10.008>
18. Dweik, R.A. and Amann, A. (2008) Exhaled Breath Analysis: The New Frontier in Medical Testing. *Journal of Breath Research*, **2**, Article ID: 030301. <https://doi.org/10.1088/1752-7163/2/3/030301>
19. Di Natale, C., Paolesse, R., Martinelli, E., *et al.* (2014) Solid-State Gas Sensors for Breath Analysis: A Review. *Analytica Chimica Acta*, **824**, 1-17. <https://doi.org/10.1016/j.aca.2014.03.014>
20. Pedrosa, M., Cancelliere, N., Barranco, P., *et al.* (2010) Usefulness of Exhaled Nitric Oxide for Diagnosing Asthma. *Journal of Asthma*, **47**, 817-821. <https://doi.org/10.3109/02770903.2010.491147>

21. Chen, W., Laiho, S., Vaittinen, O., *et al.* (2016) Biochemical Pathways of Breath Ammonia (NH₃) Generation in Patients with End-Stage Renal Disease Undergoing Hemodialysis. *Journal of Breath Research*, **10**, 036011. <https://doi.org/10.1088/1752-7155/10/3/036011>
22. Phillips, M., Altorki, N., Austin, J.H., Cameron, R.B., Cataneo, R.N., Greenberg, J., Kloss, R., Maxfield, R.A., Munawar, M.I. and Pass, H.I. (2007) Prediction of Lung Cancer Using Volatile Biomarkers in Breath. *Cancer Biomark*, **3**, 95-109. <https://doi.org/10.3233/CBM-2007-3204>
23. Davies, S., Spanel, P. and Smith, D. (1997) Quantitative Analysis of Ammonia on the Breath of Patients in End-Stage Renal Failure. *Kidney International*, **52**, 223-228. <https://doi.org/10.1038/ki.1997.324>
24. Olopade, C.O., Zakkar, M., Swedler, W.I. and Rubinstein, I. (1997) Exhaled Pentane Levels in Acute Asthma. *Chest*, **111**, 862-865. <https://doi.org/10.1378/chest.111.4.862>
25. Eisenmann, A., Amann, A., Said, M., Datta, B. and Ledochowski, M. (2008) Implementation and Interpretation of Hydrogen Breath Tests. *Journal of Breath Research*, **2**, Article ID: 046002. <https://doi.org/10.1088/1752-7155/2/4/046002>
26. Nakhleh, M.K., Amal, H., Jeries, R., Broza, Y.Y., Aboud, M., Gharra, A., Ivgi, H., Khatib, S., Badarneh, S., Har-Shai, L., Glass-Marmor, L., Lejbkowitz, I., Miller, A., Badarny, S., Winer, R., Finberg, J., Cohen-Kaminsky, S., Perros, F., Montani, D., Girerd, B., Garcia, G., Simonneau, G., Nakhoul, F., Baram, S., Salim, R., Hakim, M., Gruber, M., Ronen, O., Marshak, T., Doweck, I., Nativ, O., Bahouth, Z., Shi, D., Zhang, W., Hua, Q., Pan, Y., Tao, L., Liu, H., Karban, A., Koifman, E., Rainis, T., Skapars, R., Sivins, A., Ancans, G., Liepniece-Karele, I., Kikuste, I., Lasina, I., Tolmanis, I., Johnson, D., Millstone, S.Z., Fulton, J., Wells, J.W., Wilf, L.H., Humbert, M., Leja, M., Peled, N. and Haick, H. (2017) Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules. *ACS Nano*, **11**, 112-125. <https://doi.org/10.1021/acsnano.6b04930>
27. Phillips, M. (1997) Method for the Collection and Assay of Volatile Organic Compounds in Breath. *Analytical Biochemistry*, **247**, 272-278. <https://doi.org/10.1006/abio.1997.2069>
28. Amann, A., Poupert, G., Telsler, S., Ledochowski, M., Schmid, A. and Mechtcheriakov, S. (2004) Applications of Breath Gas Analysis in Medicine. *International Journal of Mass Spectrometry*, **239**, 227-233. <https://doi.org/10.1016/j.ijms.2004.08.010>
29. Li, J., Zhang, D., Li, Y., Wu, J. and Zhang, B. (2016) Joint Similar and Specific Learning for Diabetes Mellitus and Impaired Glucose Regulation Detection. *Information Sciences*, **384**, 191-204.
30. Yu, J., Byun, H., So, M. and Huh, J. (2005) Analysis of Diabetic Patient's Breath with Conducting Polymer Sensor Array. *Sensors and Actuators B: Chemical*, **108**, 305-308. <https://doi.org/10.1016/j.snb.2005.01.040>
31. Van Hooren, M.R., Leunis, N., Brandsma, D.S., Dingemans, A.M.C., Kremer, B. and Kross, K.W. (2016) Differentiating Head and Neck Carcinoma from Lung Carcinoma with an Electronic Nose: A Proof of Concept Study. *European Archives of Oto-Rhino-Laryngology*, **273**, 3897-3903. <https://doi.org/10.1007/s00405-016-4038-x>
32. Brekelmans, M.P., Fens, N., Brinkman, P., Bos, L.D., Sterk, P.J., Tak, P.P. and Gerlag, D.M. (2016) Smelling the Diagnosis: The Electronic Nose as Diagnostic Tool in Inflammatory Arthritis. A Case-Reference Study. *PLoS ONE*, **11**, e0151715. <https://doi.org/10.1371/journal.pone.0151715>
33. Craven, M.A., Gardner, J.W. and Bartlett, P.N. (1996) Electronic Noses-Development and Future Prospects. *Trends in Analytical Chemistry*, **15**, 486-493. [https://doi.org/10.1016/S0165-9936\(96\)00061-1](https://doi.org/10.1016/S0165-9936(96)00061-1)
34. Persaud, K. and Dodd, G. (1982) Analysis of Discrimination Mechanisms in the Mammalian Olfactory System Using a Model Nose. *Nature*, **299**, 352-355. <https://doi.org/10.1038/299352a0>
35. Rock, F., Barsan, N. and Weimar, U. (2008) Electronic Nose: Current Status and Future Trends. *Chemical Reviews*, **108**, 705-725. <https://doi.org/10.1021/cr068121q>

36. Lin, Y.J., Guo, H.R., Chang, Y.H., Kao, M.T., Wang, H.H. and Hong, R.I. (2001) Application of the Electronic Nose for Uremia Diagnosis. *Sensors and Actuators B: Chemical*, **76**, 177-180. [https://doi.org/10.1016/S0925-4005\(01\)00625-6](https://doi.org/10.1016/S0925-4005(01)00625-6)
37. Wang, P., Tan, Y., Xie, H. and Shen, F. (1997) A Novel Method for Diabetes Diagnosis Based on Electronic Nose. *Biosensors and Bioelectronics*, **12**, 1031-1036.
38. Ogorodnik, V., Kleperis, J., Taivans, I., Jurka, N. and Bukovskis, M. (2008) Electronic Nose for Identification of Lung Diseases. *Latvian Journal of Physics and Technical Sciences*, **45**, 60-67. <https://doi.org/10.2478/v10047-008-0026-2>
39. Guo, D., Zhang, D., Li, N., Zhang, L. and Yang, J. (2010) A Novel Breath Analysis System Based on Electronic Olfaction. *IEEE Transactions on Biomedical Engineering*, **57**, 2753-2763. <https://doi.org/10.1109/TBME.2010.2055864>
40. Guo, D., Zhang, D., Li, N., Zhang, L. and Yang, J. (2010) Diabetes Identification and Classification by Means of a Breath Analysis System. *International Conference on Medical Biometrics*, Hong Kong, 28-30 June 2010, 52-63. https://doi.org/10.1007/978-3-642-13923-9_6
41. Guo, D., Zhang, D. and Li, N. (2010) Monitor Blood Glucose Levels via Breath Analysis System and Sparse Representation Approach. *Sensors*, 1238-1241.
42. Yan, K., Kou, L. and Zhang, D. (2017) Learning Domain-Invariant Subspace Using Domain Features and Independence Maximization. *IEEE Transactions on Cybernetics*, **48**, 288-299.
43. Kou, L., Zhang, D. and Liu, D. (2017) A Novel Medical E-Nose Signal Analysis System. *Sensors*, **17**, 402. <https://doi.org/10.3390/s17040402>
44. Marco, S. and Gutierrez-Galvez, A. (2012) Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review. *IEEE Sensors Journal*, **12**, 3189-3214. <https://doi.org/10.1109/JSEN.2012.2192920>
45. Martinelli, E., Falconi, C., D'Amico, A. and Di Natale, C. (2003) Feature Extraction of Chemical Sensors in Phase Space. *Sensors and Actuators B: Chemical*, **95**, 132-139. [https://doi.org/10.1016/S0925-4005\(03\)00422-2](https://doi.org/10.1016/S0925-4005(03)00422-2)
46. Chawla, N.V., Bowyer, K.W., Hall, L.O., *et al.* (2002) SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, **16**, 321-357. <https://doi.org/10.1613/jair.953>