

Non-invasive Exhaled Breath and Skin Analysis to Diagnose Lung Cancer: Study of Age Effect on Diagnostic Accuracy

Elina Gashimova,* Azamat Temerdashev, Vladimir Porkhanov, Igor Polyakov, Dmitry Perunov, and Ekaterina Dmitrieva



Cite This: *ACS Omega* 2022, 7, 42613–42628



Read Online

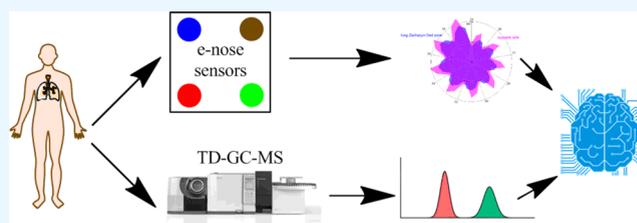
ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Development of simple, fast, and non-invasive tests for lung cancer diagnostics is essential for clinical practice. In this paper, exhaled breath and skin were studied as potential objects to diagnose lung cancer. The influence of age on the performance of diagnostic models was studied. Gas chromatography in combination with mass spectrometry (MS) was used to analyze the exhaled breath of 110 lung cancer patients and 212 healthy individuals of various ages. Peak area ratios of volatile organic compounds (VOCs) were used for data analysis instead of VOC peak areas.

Various machine learning algorithms were applied to create diagnostic models, and their performance was compared. The best results on the test data set were achieved using artificial neural networks (ANNs): classification of patients with lung cancer and young healthy volunteers: $88 \pm 4\%$ sensitivity and $83 \pm 3\%$ specificity; classification of patients with lung cancer and old healthy individuals: $81 \pm 3\%$ sensitivity and $85 \pm 1\%$ specificity. The difference between performance of models based on young and old healthy groups was minor. The results obtained have shown that metabolic dysregulation driven by the disease biology is too high, which significantly overlaps the age effect. The influence of tumor localization and histological type on exhaled breath samples of lung cancer patients was studied. Statistically significant differences between some parameters in these samples were observed. A possibility of assessing the disease status by skin analysis in the Zakharyin-Ged zones using an electronic nose based on the quartz crystal microbalance sensor system was evaluated. Diagnostic models created using ANNs allow us to classify the skin composition of patients with lung cancer and healthy subjects of different ages with a sensitivity of $69 \pm 2\%$ and a specificity of $68 \pm 8\%$ for the young healthy group and a sensitivity of $74 \pm 7\%$ and a specificity of $66 \pm 6\%$ for the old healthy group. Primary results of skin analysis in the Zakharyin-Ged zones for the lung cancer diagnosis have shown its utility, but further investigation is required to confirm the results obtained.



INTRODUCTION

Death rates of lung cancer are the highest among all types of cancer¹ mainly due to its difficult diagnostics in the early stage. Biopsy and computed tomography are the most useful methods for lung cancer diagnostics. The former procedure is invasive, while the latter involves radiation exposure which is potentially harmful for patients. Therefore, the development of a cheap, non-invasive, simple, and fast alternative to these tools is an urgent task in modern medicine. Different biological matrices are vigorously investigated for the development of new diagnostic methods.^{2,3} Among them, exhaled breath and skin are especially attractive for clinical diagnostics because these samples can be easily obtained and analyzed.⁴

A plethora of approaches have been applied for the analysis of exhaled breath to identify lung cancer biomarkers and create artificial olfactory systems.^{5,6} Among analytical methods, gas chromatography hyphenated with mass spectrometry (GC–MS) is highly informative for quantitative and qualitative analysis of exhaled breath. Results of exhaled breath analysis

applying GC–MS with preconcentration of volatile organic compounds (VOCs) on sorbent tubes^{7,8} or solid phase microextraction fibers (SPME)^{9–11} to identify lung cancer biomarkers were presented by different research groups. Implementation of multidimensional GC is especially attractive for solving such complex tasks as biomarker identification.⁷ Despite the informativeness of GC–MS, the analysis procedure is time-consuming and requires highly qualified staff. Ion mobility spectrometry (IMS),¹² proton transfer reaction MS,¹³ and selected ion flow tube (SIFT) MS¹⁴ are the other prospective approaches which provide rapid exhaled breath analysis without the preconcentration step.

Received: September 22, 2022

Accepted: November 1, 2022

Published: November 11, 2022



Table 1. Results Obtained by Different Research Groups in the Field of Exhaled Breath Analysis for the Diagnostics of Lung Cancer

biomarker	method	number of participants (age, median)		statistical data analysis (performance of diagnostic model)	reference
		lung cancer patient	healthy volunteer		
acetic acid; phenyl ether; furfural; heptadecane; 2-methyl-1-methyl-2-methylenecyclohexane; 7-oxabicyclo[4.1.0]heptane; 2-methylene-;4-undecen,5-methyl-, (E)-, anethole nonanoic acid	TD (Tenax GR, Carbopack B)-GCXGC-TOFMS(Rxi-5Sil + BPX-S0)	15 (62.0)	14 (58.0)	Fisher ratio, random forest, PCA	7
<i>n</i> -pentane; <i>n</i> -hexane; <i>n</i> -heptane; <i>n</i> -octane; <i>n</i> -dodecane; 3-methylpentane; cyclohexane; benzene; ethylbenzene; <i>n</i> -propylbenzene; propanal; <i>n</i> -butanal; <i>n</i> -hexanal; <i>n</i> -octanal; <i>n</i> -nonanal; <i>n</i> -decanal;1-butanol;2-butanone;2-pentanone	TD (carbonized molecular Sieve + Tenax TA + graphitized carbon black +) GC-MS (DB1)	81 (68.0)	89 (49.0)	Mann-Whitney <i>U</i> -test, logistic regression (sensitivity—32.0, specificity—88.0)	8
dimethyl sulfide; 1,4-pentadiene; ethyl acetate; methylcyclopentane; 2-propanol; isobutane; 2,4-dimethylpentane	SFME (CAR/PDMS)GC-MS(VF-624ms)	37 (66.0)	23 (61.0)	Mann-Whitney <i>U</i> -test, discriminant analysis (sensitivity—84.0; specificity—78.0)	9
dodecane; 2-hexanol/2-methylbutyl acetate; ethylbenzene; <i>n</i> -nonal/cyclohexanone; hexanal; heptanal; 3-methyl-1-butanol; 3-methyl-1-butanol; cyclohexanone; isopropylamine	SFME (CAR/PDMS)GC-MS(P-PoraBOND-Q)	108 (38–87)	121 (20–73)	decision tree, ANN (sensitivity—74.0, specificity—73.0)	22
VOC profile	IMS	50 (68.0)	39 (32.0)	Wilcoxon rank sum test, decision tree (sensitivity—100.0, specificity—76.0)	12
VOC profile	combined sensor system: (sensors: 12 metal oxide, 5 electrochemical, 1 solid electrolyte, 1 hot-wire)	46 (58.8)	41 (34.5)	support vector machine (sensitivity—97.8, specificity—90.2), <i>k</i> -nearest-neighbors (sensitivity—100.0, specificity—75.0), LogitBoost (sensitivity—87.0, specificity—82.9)	23
VOC profile	Aeonose	60 (65.0)	107 (63.0)	ANN (sensitivity—88.0, specificity—86.0)	15
VOC profile	quartz microbalances, functionalized with a different metalloporphyrin	70 (67.0)	76 (61.0)	PLS DA (sensitivity—81.0, specificity—91.0)	17
VOC profile	24 separate colorimetric indicators: metallochromic, acid-base and solvatochromic	92 (68.9)	137 (58.9)	C-statistics (performance—81.1)	18

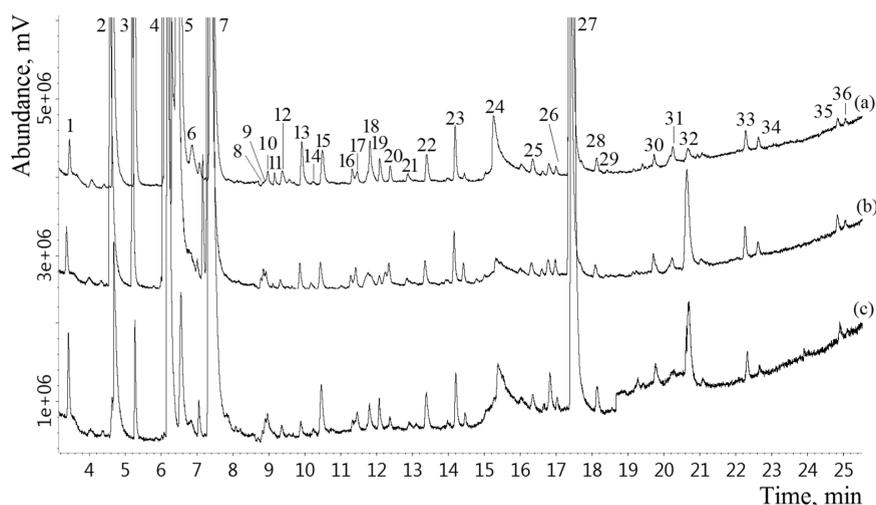


Figure 1. Total ion current chromatograms of a lung cancer patient (a) an old healthy individual (b) and a young healthy volunteer (c): 1—acetaldehyde, 2—ethanol, 3—acetonitrile, 4—acetone, 5—2-propanol, 6—dimethylsulfide, 7—*isoprene*, 8—*butanal*, 9—2,3-butandione, 10—2-butanone, 11—dimethyl carbonate, 12—ethylacetate, 13—hexane, 14—1-butanol, 15—benzene, 16—2-pentanone, 17—pentanal, 18—1-methylthiopropene, 19—dimethyl disulfide, 20—heptane, 21—1-pentanol, 22—toluene, 23—hexanal, 24—*N,N*-dimethylacetamide, 25—*m*-xylene + *p*-xylene, 26—heptanal, 27—phenol, 28—nonane, 29—dimethyl trisulfide, 30—octanal, 31—decane, 32—limonene, 33—nonanal, 34—undecane, 35—decanal, and 36—dodecane.

An alternative approach is to obtain a response not from an individual compound, but a combination of VOCs in a sample. Various kinds of electronic nose systems have already been proposed for this purpose. Cyranose 320¹⁵ and Aeonose¹⁶ are commercially available electronic noses which were applied for the analysis of exhaled breath to diagnose lung cancer. Other electronic olfaction systems based on a number of gas-sensing technologies such as nanomaterials,¹¹ quartz crystal microbalance systems,¹⁷ colorimetric sensors,¹⁸ and metal oxide gas sensors¹⁹ have also been applied to analyze exhaled breath.

To date, no compounds, presence or absence of which in exhaled breath can indicate lung cancer, have been found. In most cases, lung cancer patients and healthy volunteers can be classified using statistical data analysis. Various algorithms of machine learning, that is, logistic regression,¹⁰ support vector machine,¹⁹ decision trees,²⁰ gradient boost decision trees,²¹ random forest,⁷ and neural networks,^{15,19} have been used for the creation of classification models. Statistical data analysis allows us to achieve high predictive power even if differences in VOC concentrations are not obvious.

The findings of different scientists in the field of exhaled breath analysis to diagnose lung cancer using different analytical methods are presented in Table 1. As can be seen, there is no conformity in the list of biomarkers, statistical data analysis methods, performance of predictive models, and number of participants involved in the research. Also, in some papers, significant difference can be observed in the age of healthy individual and patient with lung cancer groups,^{12,22,23} which can cause false-positive results. Therefore, it seems important to compare different approaches for VOC determination, statistical data analysis methods, and groups with different ages and find out the optimal strategy.

Skin is another perspective source of potential biomarkers. VOCs are produced by skin through homolytic β -scission due to lipid oxidation activity of bacteria or UV radiation. Among the skin VOCs, alcohols, aldehydes, and ketones are the most prevalent compounds. Additionally, skin emanations contain acids, alkanes, aromatics, and esters.^{24,25} Skin analysis can be carried out by different techniques, namely, IMS, SPME²⁴ or

thermal desorption,²⁵ GC–MS, thermal desorption secondary electrospray ionization time of flight MS (TD–SESI–TOFMS),²⁶ and nanomaterial-based sensors array.²⁷ However, all these methods require sophisticated sample preparation equipment and are hard to implement.²⁴

The diagnostic potential of skin analysis has already been shown on the example of tuberculosis.²⁷ However, skin can also be a source of other diseases biomarkers, which remains to be explored. On the skin, there exist zones which can reflect the diseases of internal organs through reflected pain and hyperesthesia, which was first described by Zakharyin and Head.²⁸ Afterward, these zones were termed as Zakharyin-Ged zones. Accordingly, there is a possibility that the skin can reflect alterations occurring in lungs in lung cancer. However, this hypothesis has never been investigated before.

In this work, exhaled breath and skin of patients with lung cancer and healthy individuals were investigated. To study the influence of age, the healthy volunteer group was divided into two subgroups: young and old healthy individuals. The disease group was compared with two healthy volunteer subgroups separately. TD–GC–MS was applied for the analysis of exhaled breath samples. Performance of different machine learning algorithms was studied to create a diagnostic model. Also, the analysis of skin in Zakharyin-Ged zones relevant to the heart and lungs of patients with lung cancer and healthy subjects was conducted using electronic nose. Differences between skin chemical composition of investigated groups were estimated by chemometric methods.

RESULTS AND DISCUSSION

Exhaled Breath Analysis Using GC–MS. Numerous studies have demonstrated the ability of lung cancer diagnostics using analysis of exhaled breath.^{7,8,10,19} However, the findings obtained by various researchers were inconsistent (Table 1). Various analysis conditions, cohorts of participants, lists of putative biomarkers applied to develop diagnostic models, different learning algorithms, and test data set percentage were used, and different performances of created

Table 2. Frequency of VOCs Which Were Observed in Exhaled Breath of Healthy Subjects (Young and Old Groups) and Patients with Lung Cancer (%)

no	VOC	lung cancer patients	healthy individuals (matched age with LC group)	healthy individuals (young group)
1	acetone	100.0	100.0	100.0
2	isoprene	100.0	100.0	100.0
3	dimethylsulfide	100.0	100.0	100.0
4	1-methylthiopropene	100.0	96.0	92.0
5	2-pentanone	97.3	95.0	100.0
6	1-methylthiopropene	98.2	97.0	94.6
7	allyl methyl sulfide	98.2	93.0	96.4
8	dimethyl disulfide	97.3	93.0	94.6
9	2,3-butandione	85.7	83.0	92.9
10	acetonitrile	86.6	88.0	88.4
11	2-butanone	75.9	76.0	86.6
12	heptane	52.7	47.0	73.2
13	pentanal	50.0	51.0	71.4
14	dimethyl trisulfide	71.4	64.0	55.4
15	hexane	50.9	60.0	62.5
16	1-pentanol	56.2	50.0	58.9
17	benzaldehyde	59.8	62.0	59.8
18	nonanal	52.7	38.0	57.1
19	2-heptanone	57.1	40.0	43.8
20	3-heptanone	50.9	48.0	50.0
21	octanal	51.2	50.0	54.5
22	octane	50.9	54.0	47.3
23	toluene	50.9	42.0	29.3
24	1-butanol	24.1	39.0	47.3
25	hexanal	49.1	45.0	48.2
26	decanal	42.0	44.0	46.4
27	undecane	44.6	49.0	48.2
28	dodecane	45.6	47.0	45.5
29	decane	49.1	43.0	44.6
30	butyl acetate	21.4	24.0	47.3
31	butanal	36.7	39.0	49.1
32	heptanal	39.2	42.0	49.1
33	nonane	36.6	24.0	43.8
34	propylbenzene	44.6	14.0	17.8
35	benzene	31.2	32.0	25.0
36	1,3-pentadiene	25.0	11.0	15.1
37	ethylbenzene	25.0	24.0	13.3
38	1,4-pentadiene	22.3	9.0	14.2
39	<i>o</i> -xylene	20.5	22.0	11.6
40	<i>m</i> + <i>p</i> -xylene	19.6	23.0	13.4

models were observed. Previously, we have optimized analysis conditions and proposed a new data analysis approach applying VOC ratios instead of VOC peak area values. The efficiency of the approach was demonstrated for different analytical methods (GC–FID and GC–MS) and cohorts of participants.^{29,30} In this work, exhaled breath samples of three cohorts of participants were analyzed including 110 patients with lung cancer, 112 young healthy subjects, and 100 healthy individuals with the same age as lung cancer patients (old healthy individuals) by GC–MS. The GC–MS chromatograms obtained by analysis of the samples from each group are presented in Figure 1.

In total, 205 VOCs were discovered in the study. The compounds occurring in more than 50% of cases at least in one group, that is, lung cancer patients or healthy individuals (young and old groups), were considered for statistical analysis (Table 2). The frequency of VOCs occurring in the samples of

patients with lung cancer or healthy subject groups was calculated as follows

$$\text{frequency of VOC occurrence in the group} = \frac{\text{the number of samples containing VOC in the group}}{\text{the total number of samples in the group}}$$

A list of exhaled breath VOCs and frequency of their occurring are presented in Table 2.

Statistical analysis was conducted using peak area ratios of VOCs instead of VOC peak areas. Considering this approach, it seems reasonable to apply VOC with a frequency of 100% to avoid division by zero. The frequency of 100% was observed only for acetone, isoprene, and dimethylsulfide (Table 2). To consider a wider list of ratios, it was reasonable to apply the VOCs occurring the most frequently in the samples of both groups as a denominator, which was observed for the first 10 VOCs in Table 2. Among them, the lowest frequency was

Table 3. Correlation Coefficients between the Disease Status and all Considered Ratios of Patients with Lung Cancer and Young Healthy Subjects

VOC	acetone	allyl methyl sulfide	1-methylthiopropene	dimethyl disulfide	acetonitrile
hexane	-0.265	-0.254	-0.203	-0.242	-0.313
toluene	0.252	0.231	0.245	0.254	0.242
1-pentanol	0.137	0.012	0.070	0.070	-0.047
pentanal	0.192	-0.294	-0.256	-0.221	-0.293
dimethyl trisulfide	0.278	0.208	0.199	0.273	0.130
nonanal	0.008	-0.049	-0.051	-0.069	-0.143
heptane	-0.066	-0.147	-0.106	-0.121	-0.092
2-butanone	-0.025	-0.142	0.072	-0.036	-0.104
isoprene	0.245	0.970	0.103	0.068	-0.038
1-methylthiopropene	0.224	0.065	0.047	0.008	0.108
dimethylsulfide	0.076	0.104	0.095	0.049	-0.076
2,3-butandione	0.154	-0.020	0.137	0.021	0.087
2-pentanone	0.430	0.179	-0.339	-0.314	0.143
benzaldehyde	0.103	0.039	0.063	0.041	0.002
octanal	0.032	0.890	0.106	0.073	0.093
octane	0.106	0.050	0.069	0.071	0.031
3-heptanone	0.008	-0.031	-0.033	-0.041	-0.069
2-heptanone	0.126	0.086	0.070	0.108	-0.019

Table 4. Correlation Coefficients between the Disease Status and All Considered Ratios of Patients with Lung Cancer and Old Healthy Subjects

VOC	acetone	allyl methyl sulfide	1-methylthiopropene	dimethyl disulfide	acetonitrile
hexane	-0.132	-0.098	-0.108	-0.039	-0.260
toluene	-0.062	-0.053	-0.070	-0.014	-0.172
1-pentanol	0.255	0.225	0.233	0.261	0.101
pentanal	-0.019	-0.070	-0.045	-0.041	-0.159
dimethyl trisulfide	0.150	0.128	0.145	0.356	-0.070
nonanal	0.115	0.111	0.104	0.157	-0.043
heptane	0.115	0.114	0.140	0.164	0.029
2-butanone	0.065	-0.136	0.067	-0.093	-0.081
isoprene	0.138	0.125	0.081	-0.090	0.278
1-methylthiopropene	0.136	0.037	0.101	0.070	0.062
dimethylsulfide	0.081	0.075	0.132	0.096	-0.101
2,3-butandione	0.153	0.116	0.090	0.115	0.225
2-pentanone	0.289	0.087	-0.144	0.079	0.096
benzaldehyde	0.086	0.040	0.057	0.143	-0.175
octanal	0.108	0.115	0.039	0.161	0.048
octane	-0.039	-0.019	-0.006	0.051	-0.161
3-heptanone	0.449	0.431	0.447	0.451	0.392
2-heptanone	0.48	0.476	0.489	0.509	0.389

observed in the case of acetonitrile (86.6%). The frequency of occurring for rest compounds was lower and was different in the investigated groups. These VOCs were applied only as a numerator.

The aim of the initial step of the study was to determine the ratios with the highest difference between the patients with lung cancer and healthy individuals. For this purpose, we used Spearman's rang correlation test. To study the effect of age, the correlation coefficients between the group of patients with lung cancer and each group of healthy individuals were calculated separately. Correlation coefficients between the ratios and disease status of patients with lung cancer and young healthy subjects are reflected in Table 3, patients with lung cancer and old healthy subjects—in Table 4. The first column represents the peak areas of VOCs used as a numerator; the first line—as a denominator; and the row-column-intersection represents

the correlation coefficient between the ratio (numerator to denominator) and the disease status.

The ratios with the highest correlation coefficients, excluding duplicative ones, were selected to create diagnostic models. To eliminate the age effect, only the ratios correlated with the disease status in both groups, namely, young and old healthy volunteers, were used for the development of diagnostic models. Table 5 represents the correlation coefficients of ratios selected for the development of diagnostic models in both groups.

As presented in Table 5, correlation coefficients were higher for most of selected ratios, when the groups of lung cancer patients and young volunteers were compared. Two kinds of the models were created: based on patients with lung cancer and young healthy subjects and patients with lung cancer and old healthy subjects; their performance was compared.

Table 5. Ratios Selected for the Creation of Diagnostic Models

ratio	correlation coefficient	
	lung cancer patients/old healthy volunteers	lung cancer patients/young healthy volunteers
1-pentanol/acetone	0.224	0.156
1-methylthiopropene/acetone	0.136	0.224
dimethyl trisulfide/dimethyl disulfide	0.356	0.273
isoprene/acetone	0.138	0.245
pentanal/acetonitrile	-0.159	-0.293
hexane/acetonitrile	-0.260	-0.313
2-butanone/allyl methyl sulfide	-0.136	-0.142
2,3-butandione/acetone	0.153	0.154
2-pentanone/acetone	0.289	0.430

Several machine learning algorithms were applied to select the one with the highest performance. Among them, the main application field of logistic regression and support vector machine is binary classification which makes these algorithms attractive to solve the task. Random forest algorithm is more flexible because it can identify a broader scope of possible relationships between the model predictors (VOC peak area ratios) and the disease status. Also, one of the most widely used algorithms in medical diagnostics, neural networks, was used. The input values of each model were the same set of nine ratios (Table 5). The models were built using cross-validation. Performance of models created using three data sets for the groups of young and old healthy volunteers is shown in Tables 6 and 7.

Performance of diagnostic models varied significantly depending on the algorithm of machine learning in both types of the models. However, the best results on test data were achieved using neural networks in both types of the models with $87 \pm 2\%$ sensitivity and $83 \pm 3\%$ specificity in the case of young healthy subjects and $81 \pm 3\%$ sensitivity and $85 \pm 1\%$ specificity for old healthy volunteers. As can be seen from Tables 6 and 7, independently of algorithm of machine learning, sensitivity on test data was higher in the case of the

young healthy group, but specificity was almost the same for both groups. The models were created on the same set of ratios, statistically significantly different for both healthy subject groups, but the performance of diagnostic models was slightly higher in the case of the young group. Some researchers published the classification models, allowing us to discriminate patients with lung cancer and healthy subjects, but the age of participants was significantly different.^{12,23} Our findings prove the assumption that the best way to eliminate the age effect is to compare groups of the same age. In this study, the number of smokers was nearly the same, which allows us to eliminate the smoking impact. We cannot exclude smokers from the study because most lung cancer patients are active smokers, but the smoking factor significantly changes the VOC profiles of exhaled breath;³¹ therefore, to avoid obtaining false-positive results, the number of smokers should be equal in both groups of participants. The model for the lung cancer diagnostics using GC-FID analysis of exhaled breath was created by us previously.²⁹ It is worth noting that hexane/acetonitrile ratio values from the previous model also significantly correlated with the disease status in the present study for both young and old healthy groups (correlation coefficients of -0.260 and -0.313), which proves their reliance in terms of disease status indicators. The ratios included in the present models should also be tested on the larger cohort of people.

Concentrations of 3-hydroxy-2-butanone and 1-butanol were found to be significantly lower in the case of squamous cell carcinoma than in adenocarcinoma.³² However, other authors noted the lack of any difference in exhaled breath composition depending on histology.³¹ Influence of histology on the exhaled breath profile was studied in the current work. For this, exhaled breath of patients with small cell carcinoma, squamous cell carcinoma, and adenocarcinoma was analyzed. Kruskal-Wallis test was applied to find whether there exists the difference in VOC profiles of patients with different histologies. Statistically significant differences between VOC profiles of patients with different histologies were observed for dimethyl trisulfide/1-methylthiopropene ($p = 0,029$), 1-methylthiopropene/acetonitrile ($p = 0,035$), 3-heptanone/

Table 6. Performance of Diagnostic Models Using Different Machine Learning Algorithms for Patients with Lung Cancer and Young Healthy Volunteers

machine learning algorithm	data set	training data set		test data set	
		sensitivity, %	specificity, %	sensitivity, %	specificity, %
logistic regression	1	79	76	75	84
	2	81	83	71	72
	3	89	76	80	88
	mean \pm SD	83 ± 5	78 ± 4	75 ± 5	81 ± 8
random forest	1	87	71	77	64
	2	94	82	77	62
	3	92	80	77	76
	mean \pm SD	91 ± 4	78 ± 6	77 ± 0	67 ± 8
support vector machine	1	73	82	75	84
	2	76	86	71	76
	3	85	80	80	87
	mean \pm SD	78 ± 6	83 ± 3	75 ± 5	82 ± 6
neural networks	1	88	77	93	85
	2	85	82	85	79
	3	89	77	87	85
	mean \pm SD	87 ± 2	79 ± 3	88 ± 4	83 ± 3

Table 7. Performance of Diagnostic Models Using Different Machine Learning Algorithms for Patients with Lung Cancer and Old Healthy Volunteers

machine learning algorithm	data set	training data set		test data set	
		sensitivity, %	specificity, %	sensitivity, %	specificity, %
logistic regression	1	74	79	75	66
	2	78	83	69	67
	3	77	74	69	66
	mean \pm SD	76 \pm 2	79 \pm 4	71 \pm 3	66 \pm 1
random forest	1	82	90	75	76
	2	93	89	71	78
	3	90	87	83	75
	mean \pm SD	88 \pm 6	89 \pm 2	76 \pm 6	76 \pm 2
support vector machine	1	68	80	75	79
	2	75	84	66	69
	3	74	79	71	81
	mean \pm SD	72 \pm 4	81 \pm 3	71 \pm 5	76 \pm 6
neural networks	1	86	86	83	86
	2	93	89	77	86
	3	87	78	83	84
	mean \pm SD	89 \pm 4	84 \pm 6	81 \pm 3	85 \pm 1

allyl methyl sulfide ($p = 0,039$), and 3-heptanone/1-methylthiopropene ($p = 0,030$) ratios.

If the tumor is in the central part of a lung, it is closer to the airways than the peripheral tumor; therefore, VOC profiles of central and peripheral tumor can be different. Statistically significant differences in VOC peak areas and VOC peak area ratios in groups of patients with different tumor localizations were estimated by the Mann–Whitney test. Several parameters were significantly different in investigated groups: 1-pentanol ($p = 0,020$), 1-pentanol/2,3-butandione ($p = 0,006$), 1-pentanol/isoprene ($p = 0,023$), 1-pentanol/acetone ($p = 0,039$), dimethyl disulfide/acetonitrile ($p = 0,032$), and 2-butanone/isoprene ($p = 0,040$). Other research groups have never considered differences in VOC profiles depending on tumor localization; therefore, further research is required to prove the results obtained. It should be noted that among VOC ratios included in the creation of diagnostic models statistically significant difference in groups with central and peripheral tumor localization was observed only in the case of the 1-pentanol/acetone ($p = 0,039$) ratio.

Skin Analysis Using Electronic Nose. Owing to the ease of use, rapidity and simplicity, electronic nose has a great potential in a clinical context. Different research groups have demonstrated the ability of an electronic nose based on a quartz crystal microbalance sensor system coated with different metalloporphyrins to classify patients with lung cancer and healthy subjects by exhaled breath analysis.^{33,34}

Exhaled breath^{5,7,8} and skin^{24,25} contain a wide list of analytes, for example, water, alcohols, ketones, carbon acids, amines, ethers, esters, aromatic compounds, saturated hydrocarbons, ammonia, and acetonitrile. Analytical properties of electronic nose regarding different compounds were evaluated earlier. “Visual print” areas of a full array of sensors and three most sensitive sensors were applied to construct calibration curves.³⁰ The lowest sensitivity was observed for ethyl ester, hexane, ethyl acetate, and acetone with the quantification limits up to 50 g/m³. The limitations of electronic nose working parameters were evaluated. Ambient air of medical institutions contains different disinfectants (1-propanol and 2-propanol) at high concentrations, which affects electronic nose measurements. Also, it should be noted that significant alterations of

temperature (6 °C) and humidity (30%) greatly influence the signal; therefore, experimental conditions should be controlled, for example, by measurement of ethanol before the analysis and calculation of correction coefficients for samples. Analysis of exhaled breath applying the electronic nose was performed earlier. The electronic nose was unable to differentiate the samples of exhaled breath in relation of the disease status because the electrodes are sensitive to various VOCs, but larger quantities of water in the samples block the response of sensors toward other volatiles.³⁰

An original approach to monitor health status by the analysis of volatiles emitted from Zakharyin-Ged zones of skin using electronic nose has been proposed earlier.³⁵ In this study, an electronic nose based on the quartz crystal microbalance sensor system was used to analyze the skin in two Zakharyin-Ged zones corresponding to heart and lungs. Skin analysis using the electronic nose was performed by contacting the device with the skin. The time of contact was the same as in the case of analysis of VOC standards (80 s).³⁰ Desorption was performed with an opened cell during 120 s. Equilibration of electronic nose was achieved using ambient air.

Skin analysis in Zakharyin-Ged zones of 40 patients with lung cancer and 80 healthy individuals (40 young healthy volunteers and 40 old healthy individuals) was conducted. Recovery curves of skin measurements in the lung Zakharyin-Ged zone of patients with lung cancer and healthy individuals (young and old groups) obtained under identical conditions are presented in Figure 2. The difference in recovery curves of patients with lung cancer and healthy subjects of different age can be observed.

Alterations in the temperature and humidity significantly affect the results. To eliminate the influence, the skin was measured in Zakharyin-Ged and its opposite side zones; then, the ratios of parameters were used for statistical data analysis. Additionally, this approach eliminates the impact of individual skin metabolome and exogenous factors influencing the whole body but varying from one subject to another. Statistical analysis was conducted with regard to ΔF_{\max} , area of a “visual print” of all sensors ΔF_{\max} (Sv.p., Hz). Moreover, we considered the dynamics of sorption and desorption parameters by constructing “dynamic visual prints” using the

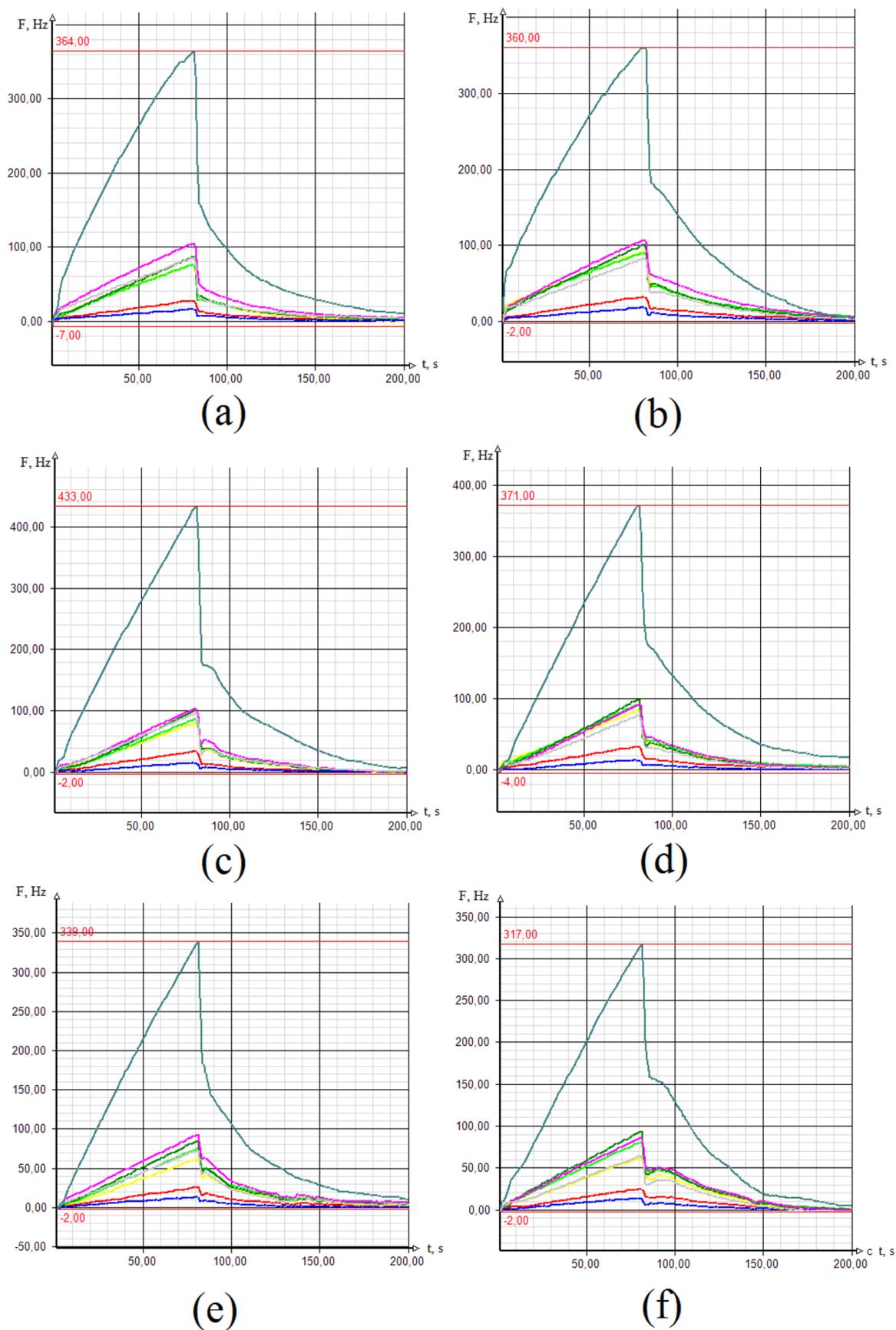


Figure 2. Typical recovery curves of skin in lungs Zakharyin-Ged zone and its opposite side of a patient with lung cancer (a,b), a young healthy volunteer (c,d), and an old healthy volunteer (e,f).

value of ΔF in the time of signal registration. We examined all samples manually: all sensors were considered separately during all time of analysis with the aim to find the time with maximal difference of the signal between the Zakharyin-Ged zone and its opposite side of healthy and patients with lung cancer groups. In accordance with the results, several algorithms of “dynamic visual prints” construction were created (Table 8).

Table 8. Algorithms of Constructing “Dynamic Visual Prints”

algorithm	sensor used	time of sensor responses recording, s
initial sorption	Zr1, DCH-18C6, GA1, PEGS	24, 26, 27, 28, 30, 32, 34, 36, 38
median sorption	Zr2, PEGS	10, 20, 30, 60
final sorption	GA2, Zr2	60, 63, 65, 67, 69, 70, 71, 73
sorption	Zr1, DCH-18C6, GA1, PEGS	20, 30, 40, 50, 60, 70
common dynamic visual print	MWCNT1, Zr1, DCH-18C6, GA1, Zr2, PEGS, MWCNT2	20, 30, 60, 100, 120, 140
desorption	GA1, Zr2, MWCNT2	110, 120, 130, 140, 150

At the initial step of statistical analysis, the data set was normalized by the transformation $\log(x + 10^6)$. Parameters with significantly different values in investigated groups were selected using one-way ANOVA. Statistically significant differences in parameter values were found only in the case of the lung Zakharyin-Ged zone (Table 9).

Table 9. Parameters with Statistically Significant Values in Groups of Patients with Lung Cancer and Healthy Subjects in the Lung Zakharyin-Ged Zone

parameter	patients with lung cancer and young healthy individuals (<i>p</i> -value)	patients with lung cancer and old healthy individuals (<i>p</i> -value)
ΔF_{\max} (Zr1)	0.005	0.015
ΔF_{\max} (DCH-18C6)	0.004	0.007
ΔF_{\max} (GA1)	0.012	0.032
ΔF_{\max} (GA2)	0.024	0.034
ΔF_{\max} (MWCNT2)	0.032	0.041
area of “visual print”	0.036	0.005
area of “dynamic visual print”: initial sorption	0.004	0.012
area of “dynamic visual print”: sorption	0.004	0.009

Typical “visual prints” and “dynamic visual prints” of a patient with lung cancer and young and old healthy subject are presented in Figure 3.

Classification of the groups was performed using PCA and DA based on one-way ANOVA analysis results. PCA is a useful method for the reduction of multidimensional data to its main components and simplifying data analysis procedure. It allows us to evaluate the significance of PC and their contribution to a discrimination by means of scree plots. As shown in Figure 4a,b, the main contribution to the discrimination is made by

the first factor which explains 80.85% of the variation in the case of young healthy individuals and for old healthy subjects—59.86%. As illustrated in Figure 4c,d, PCA based on these eight parameters does not provide clear clustering of the groups by the first two PCs in both young and old healthy volunteers, which demonstrates the importance of model creation for the assessment of the sensitivity and specificity of the proposed approach.

As illustrated in Figure 4a,b, it is hard to reduce the quantity of variables by subtracting them to PC. Therefore, diagnostic models using DA were created using the results of one-way ANOVA. Sensitivity and specificity of created models were 65 and 65% for the group of young healthy subjects and for the old healthy group—60 and 65%, respectively. Also, diagnostic models were created using one of the most effective machine learning algorithms, that is, neural networks. The input values of each model represented results of one-way ANOVA (Table 9). The models were built using cross-validation. Performance of models created using five data sets is shown in Table 10.

In the case of test data set, $69 \pm 2\%$ sensitivity and $68 \pm 8\%$ specificity were observed for the young healthy group; 74 ± 7 and $66 \pm 6\%$ for the old healthy group, which is higher than that in the case of DA. The sensitivity of the old healthy group was higher than the young healthy group, but specificity was about the same for both groups. The models were created on the same set of parameters, different for both healthy subject groups, but the performance of diagnostic models was too low for diagnostic purposes. To evidence the potency of the proposed approach, the skin analysis of a larger group of people is required. The study sheds a light on a new source of biomarker searching. It can be supposed that biochemical processes occurring as the result of tumor activity are reflected in lung Zakharyin-Ged zones. This assumption can be proved, for example, by monitoring a lung cancer patient before and after tumor resection to determine whether a tumor affects alterations in skin excretion profiles.

To the best of our knowledge, alterations in the skin VOC profile of Zakharyin-Ged zones occurring as a result of tumor have never been investigated before. However, the skin of healthy volunteers was analyzed using different analytical methods. The authors²⁴ determined 3-methyl-2-butenal, 6-methylhept-5-en-2-one, *sec*-butyl acetate, benzaldehyde, octanal, 2-ethylhexanol, nonanal, and decanal in skin headspace using IMS and GC–MS. Nonanal, decanal, and 6-methylhept-5-en-2-one were also found in skin emanations in another work.²⁵ Aldehydes can be present in the skin VOC profile as a result of metabolic processes in the skin. Alterations in skin VOCs occurring due to lung cancer have never been investigated before. In the current paper, an electronic nose was used to analyze skin. The parameters of the electronic nose cannot identify a single compound; they respond only to a group of compounds, which hinders the interpretation of results in terms of the qualitative composition of a sample. It would be an interesting issue for further research to explore qualitative and quantitative composition of Zakharyin-Ged zones in detail.

CONCLUSIONS

Classification of exhaled breath samples of patients with lung cancer and healthy subjects of different ages by using the VOC peak area ratios and applying various machine learning methods can be performed using GC–MS. The difference in performance of diagnostic models created using healthy

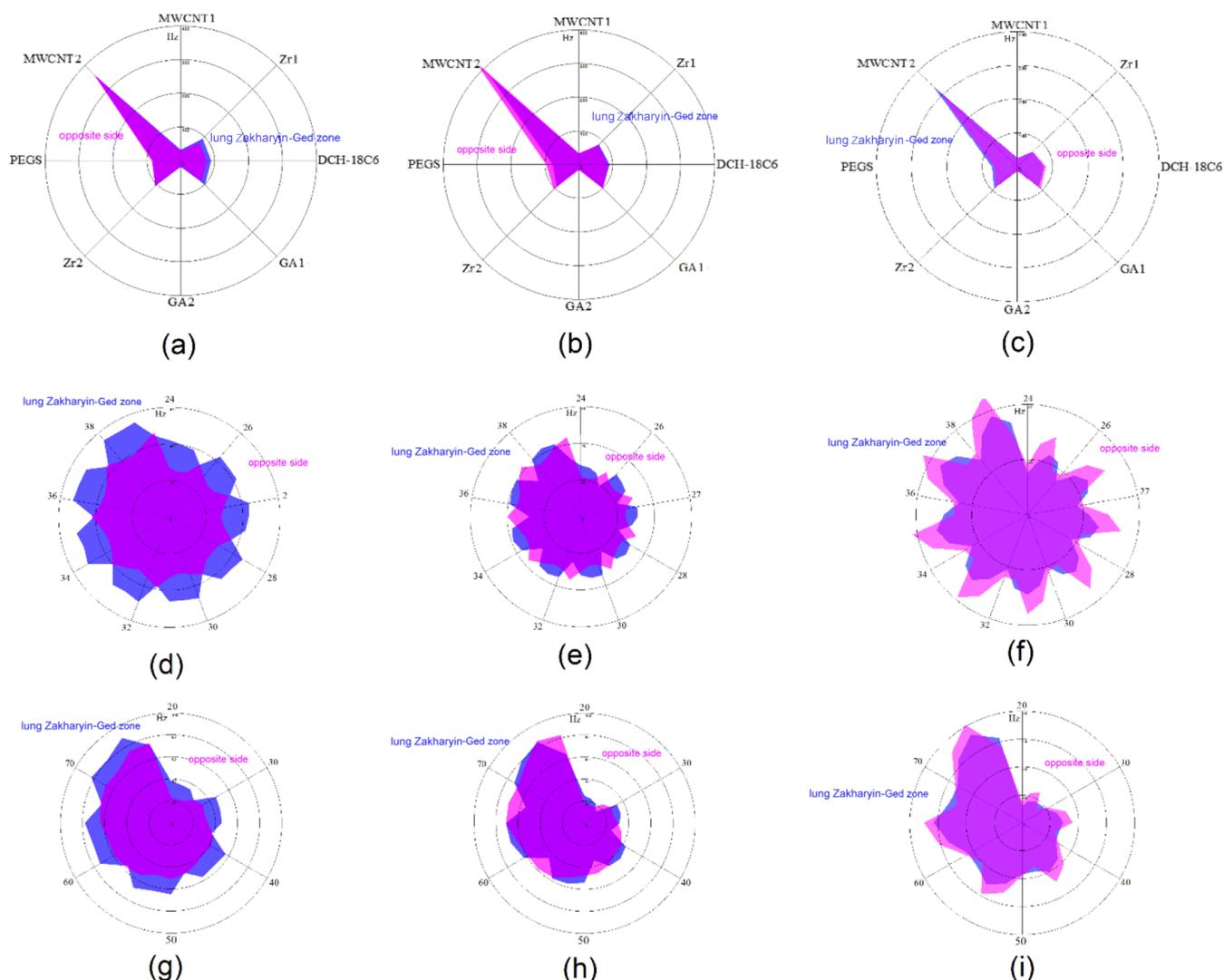


Figure 3. Typical “visual print” of the skin of a patient with lung cancer (a), a young healthy individual (b), and an old healthy volunteer (c); typical “dynamic visual prints” of the skin of a patient with lung cancer and young and old healthy volunteers: initial sorption (d,e,f) and sorption (g,h,i).

subjects with different ages was about the same, but it is important to select the parameters which are correlated with disease status independently of the age. Localization of tumor and histology can affect the exhaled breath VOC profile. A novel approach for lung cancer diagnostics by analyzing skin in Zakharyin-Ged zones using QMB electronic nose has a great potential, but further investigation is required to provide the reliance of the proposed approach.

EXPERIMENTAL SECTION

Materials. Ethyl ether, benzylamine, ethylenediamine, *t*-butylamine, methylamine (>95%), and formic acid (98%) were purchased from Acros Organics (USA). Methanol, ethanol, *n*-hexane, acetonitrile, toluene, and benzene (>95%) were obtained from Sigma-Aldrich (USA). Acetone, *n*-butanol, 2-butanol, and isoamyl alcohol (99.9%) were purchased from Ecos-1 (Russia). Ammonia, acetic acid, and 2-propanol were obtained from Vecton (Russia). Ethyl acetate and butyl acetate (99%) were purchased from Component-reaktiv (Russia). A Milli-Q simplicity system (Milli-Q, Millipore, France) was used to obtain 18.2 M Ω -cm water.

Human Subjects. The study included three groups of participants: patients with lung cancer, healthy individuals with young age (median—21 years old), and healthy individuals with age comparable with lung cancer patients (median—60 years old). Healthy status of participants was confirmed by a report of an annual medical examination. Criteria for including were the lack of inflammation processes and pathologies in lungs, which was confirmed by fluorography. Diagnosis of patients with lung cancer was verified by biopsy. The samples were collected before the beginning of any treatment or during the treatment with different regimes (Table 11). Demographic data were collected including sex, age, smoking status, and time since last smoking. Information on the participants is shown in Table 11. Each subject provided a declaration of agreement to take a part in the study. The study was conducted in conformity with guidelines and regulations of the local ethics committee of state budgetary healthcare institution “Research Institute—Regional Clinical Hospital No 1 named after Professor S.V. Ochapovsky”.

Collection of Exhaled Breath Samples. Tedlar (Supelco, Bellefonte, PA, USA) or Mylar (EKAN, Russia) sampling bags with the volume of 5 L were used to collect the samples of

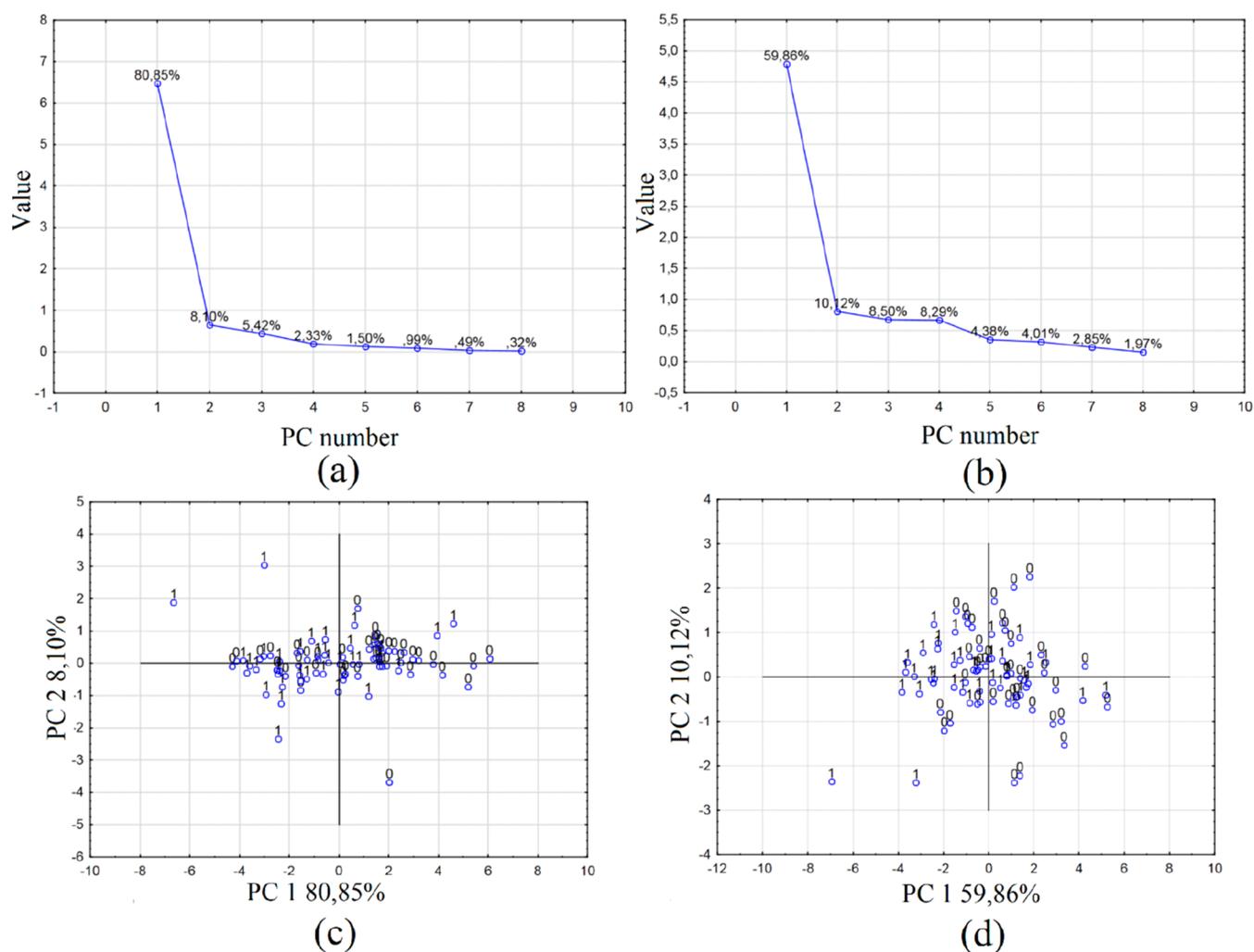


Figure 4. Scree plot and PCA score plot with parameters selected by one-way ANOVA to classify patients with lung cancer (a,c) and young healthy volunteers and old healthy volunteers (b,d).

Table 10. Performance of Diagnostic Models

data set	patients with lung cancer and young healthy individuals				patients with lung cancer and old healthy individuals			
	training data set		test data set		training data set		test data set	
	sensitivity, %	specificity, %	sensitivity, %	specificity, %	sensitivity, %	specificity, %	sensitivity, %	specificity, %
1	65	74	67	78	75	84	75	75
2	79	66	71	73	77	75	78	63
3	81	59	67	63	81	77	78	60
4	70	61	71	57	84	68	78	67
5	70	59	70	67	69	82	63	67
mean \pm SD	73 \pm 7	64 \pm 6	69 \pm 2	68 \pm 8	77 \pm 6	77 \pm 6	74 \pm 7	66 \pm 6

exhaled breath. Nitrogen was applied to clean the bags. We have identified the compounds which polluted the sample from the sampling bag material previously.²⁹ *N,N*-Dimethylacetamide and phenol are the pollutants from sampling bags. This finding allows us to use both sampling bags, but we excluded phenol and *N,N*-dimethylacetamide from a list of considered VOCs. Exhaled breath of patients with lung cancer and some healthy volunteers was sampled in the hospital. The rest samples of healthy volunteers were collected in a solvent-free room. Room air samples were collected each day of the sampling to consider the contribution of exogenous compounds. Sampling was conducted after the participants were

fasted overnight. Active smokers did not smoke for at least 2.5 h before breath sampling. After the participants had rested for 10 min in the room, subjects were asked to breathe deeply, hold the breath for 10 s, and breathe out calmly into the sampling bag until the bag was filled. The samples were processed within 6 h after collection.

GC–MS Analysis of Exhaled Breath. Exhaled breath samples of patients with lung cancer and healthy individuals were analyzed by GC–MS. A gas chromatograph (Chromatec crystal 5000.2, Russia) fitted with a quadrupole mass spectrometer with a source of electron ionization (Chromatec MSD, Russia) coupled to a thermal desorber TD2

Table 11. Clinical Characteristics of Subjects

group	parameter	total	male	female
healthy control (young group)	number	112	34	78
	age, range (median)	21–35 (21)	21–35 (21)	21–33 (23)
	number of smokers	17	10	6
healthy control (old group)	number	100	45	55
	age, range (median)	36–78 (61)	39–78 (63)	36–69 (61)
	number of smokers	9	5	4
lung cancer patient	number	110	87	23
	age, range	37–77	37–77	38–73
	age, median	63	60	65
	number of smokers	22	21	1
	type of lung cancer			
	small cell carcinoma	12	12	0
	adenocarcinoma	50	29	21
	squamous cell carcinoma	38	33	5
	non-small cell lung cancer	3	3	0
	thymoma	2	2	0
	neuroendocrine carcinoma	2	2	0
	non differentiated	5	4	1
	tumor localization			
	central	58	49	9
	peripheral	52	36	16
	mediastinum	2	2	0
	type of biopsy			
	endobronchial biopsy	96	75	21
	videothoracoscopy	14	11	3
	transbronchial biopsy	1	1	0
	transtorachal biopsy	1	1	0
	TNM (tumor, nodules, metastasis) stage			
	T1N0M1	1	0	1
	T2N0M0	8	5	3
	T2N0M1	5	5	0
	T2N1M0	8	6	2
	T2N1M1	1	1	0
	T2N2M0	2	2	0
	T2N2M1	6	5	1
	T2N3M0	1	0	1
	T2N3M1	2	1	1
	T3N0M0	6	5	1
	T3N0M1	2	1	1
T3N1M0	7	6	1	
T3N1M1	1	1	0	
T3N2M0	15	10	5	
T3N2M1	4	3	1	
T3N3M0	1	0	1	
T4N0M0	5	3	2	
T4N0M1	5	4	1	
T4N1M0	8	6	2	
T4N1M1	1	0	1	
T4N2M0	10	9	1	
T4N2M1	10	9	1	
T4N3M0	2	2	0	
T4N3M1	1	1	0	
	Chemotherapy Regimen			
	carboplatin + paclitaxel	25	21	4
	carboplatin + pemetrexed	8	7	1
	docetaxel	6	5	1
	hycamtin + carboplatin	6	3	3
	carboplatin + etoposide	5	5	0
	docetaxel + carboplatin	3	3	0
	cisplatin + vincristine	3	2	1
	etoposide + cisplatin	3	3	0

Table 11. continued

group	parameter	total	male	female
	Chemotherapy Regimen			
	cyclophosphamide + docetaxel	3	3	0
	cisplatin + pemetrexed	3	2	1
	pemetrexed + cisplatin + bevacizumab	3	2	1
	carboplatin + paclitaxel + doxorubicin	2	1	1
	docetaxel + cyclophosphan	2	2	0
	irinotecan	2	2	0
	docetaxel + carboplatin + resorba	1	1	0
	docetaxel + cisplatin	1	1	0
	docetaxel + cisplatin + bevacizumab	1	0	1
	carboplatin + mitotax	1	1	0
	paclitaxel + carboplatin + lomustine	1	1	0
	paclitaxel + vinorelbium	1	0	1
	gemcitabine + docetaxel	1	1	0
	gemcitabine + carboplatin	1	1	0
	doxorubicin + cisplatin + vincristine + cyclophosphamidum	1	1	0
	paclitaxel + maverex	1	1	0
	paclitaxel + vinorelbine	1	0	1
	carboplatin	1	1	0
	pemetrexed + cisplatin + bevacizumab + resorba	1	1	0
	docetaxel + cisplatin + bevacizumab	1	0	1
	immunotherapy regimen			
	nivolumab	3	2	1
	pembrolizumab,	3	3	0
	atezolizumab	2	1	1
	target therapy regimen			
	gefitinib	1	0	1

(Chromatec, Russia) was applied. Data were acquired and processed with mass spectral library NIST 2017, Version 2.3 (Gatesburg, USA) and Chromatec Analytic (Chromatec, Russia) software. VOC preconcentration and detection conditions were optimized previously. VOCs were separated using a Supelco Supel-Q PLOT (30 m × 0.32 mm × 15 μm) column because it provides separation of the greatest number of exhaled breath VOCs in comparison to cyanopropyl-phenyl methyl-polysiloxane, diphenyl-dimethyl-polysiloxane, and polyethylene glycol-TPA modified columns. Sorbent properties of the tubes with Porapak N (50/80 mesh), Chromosorb 106 (60/80 mesh), Tenax TA (35/60 mesh), and multibed sorbent [Tenax GR (35/60 mesh), Carbopack B (60/80 mesh), and Carbosieve SIII (60/80 mesh)] (Chromatec, Russia) were studied for the preconcentration of VOCs for GC–MS analysis. Tenax TA sorbent was the best to preconcentrate the analytes because the results were the most reliable and stable using the sorbent.²⁹ The sample VOCs were preconcentrated using a Tenax TA sorbent tube by passing a 0.5 L sample through it at a rate of 200 mL/min applying a PV-2 aspirator (Chromatec, Russia). Thermal desorption and GC–MS analysis conditions are given in Table 12. VOCs were identified using analytical standards. If the VOC standard was not available, the identification was conducted using the mass spectral library. The VOCs with match factor ≥85% were treated.

Skin Analysis Using Electronic Nose. Electronic nose “MCWbioG—8” (Multichannel biogas nanobalances, Voronezh, Russia) represents a chamber with the volume of 127 mL with the pipe to inject a gaseous sample. The laptop is used to control the device. It provides the ability to conduct exhaled breath and skin analysis independently from its location. The

Table 12. Thermal Desorber and GC–MS Operation Modes

	parameter	value	
thermal desorber	carrier gas	helium	
	valve temperature, °C	150	
	transition line temperature, °C	180	
	desorption temperature, °C	250	
	initial trap temperature, °C	−10	
	final trap temperature, °C	250	
	desorption time, min	5	
GC–MS	carrier gas	helium	
	injector temperature, °C	250	
	split ratio	1:10	
	ion source temperature, °C	200	
	transfer line temperature, °C	250	
	scan mode	full scan	
	scan range, amu	29–250	
	electron impact ionization, eV	70	
	temperature program		
	heating rate, °C/min	temperature, °C	time, min
0	50	0	
10	150	10	
6	220	11.7	
4	250	7.5	
	carrier-gas flow rate, mL/min	1.30	

bottom of the chamber can be hinged, which allows us to equilibrate the device by ambient air and take into account its influence. The chamber was opened during the time between analysis for the baseline equilibration. The work regimes of the device are presented in Figure 5. The analytical signal is obtained from sensors coated with various films of sorbent:

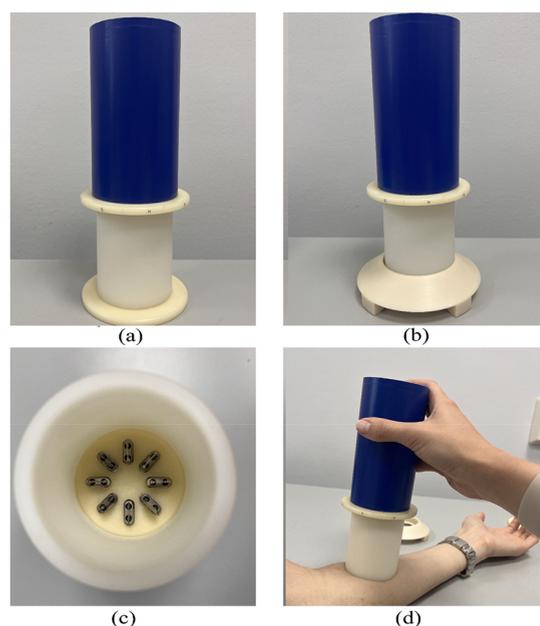


Figure 5. Regimes of electronic nose working (a—the standby mode of the device: the closed chamber, b—equilibrating the device by ambient air, c—sensors of the device, and d—skin sampling).

multiwalled carbon nanotubes 1, MWCNT1 (4.0 μg); zirconium oxide nitrate 1, Zr1 (2.0 μg); dicyclohexane-18-crown-6, DCH-18C6 (19.0 μg); biohydroxyapatite 1, GA1 (3.0 μg); biohydroxyapatite 2, GA2 (4.0 μg); zirconium oxide nitrate 2, Zr2 (2.3 μg); polyethylene glycol succinate, PEGS (12.0 μg); and multiwalled carbon nanotubes 2, MWCNT2 (3.0 μg). The resonators coated with MWCNT1, Zr1, DCH-18C6, GA1, and PEGS have reference frequency of 10.0 MHz and the rest ones have a reference frequency of 14 MHz.

Analysis of skin using an electronic nose was performed for 40 patients with lung cancer and 80 healthy subjects (40 young and 40 old healthy volunteers). Volunteers involved in skin measurements additionally were asked to refrain from using cosmetics on the day of sampling and to shower not later than 12 h before sampling. According to the diagnostic significance of the Zakharyin-Ged zones, the left forearm area of the skin reflects the status of the heart and the right side of III–IV neck segments area of skin—lungs and bronchus. Analysis of these heart and lungs Zakharyin-Ged zones and their opposite sides was performed using an electronic nose by contacting the

device with the skin. The location of Zakharyin-Ged zones, which have been measured in the study, is shown in Figure 6.

Several analytical parameters were considered for data treatment:³⁰ analytical signal ΔF_{max} (Hz) was calculated as the difference between the initial (i.e., before sample exposure) and maximal, during analysis, oscillation frequency F_0 and F_{max} ; ΔF_{max} of all sensors was applied to create a “visual print” and calculate its area (S v. p., Hz).

Statistical Analysis. The chromatograms were obtained in the full scan mode. The exhaled breath VOC peak areas were used as a quantitative parameter in GC–MS. They were calculated using the extracted ion chromatogram mode. The ambient air peak area values were subtracted from the exhaled breath sample ones. Negative results of subtraction were set to zero. Only the peak areas with values at least 20% higher than in room air were used for the statistical analysis for enhancing the reliability of the results. The rest peaks were set to zero too. The VOCs occurring in more than a half of samples were considered for statistical analysis. The ratios of the VOC peak areas to the main ones occurring in more than 86% of the samples and ratios of the main VOCs were used for statistical analysis.

The chemometric calculations were performed by StatSoft STATISTICA (version 10). The distribution normality was evaluated by applying the Kolmogorov–Smirnov test. The distribution was not normal. Spearman’s rank correlation test ($p = 0.05$) was used for the identification of the relationship between the parameters and disease status.

The ratios with the highest statistically significant correlation coefficients, excluding duplicative ones, were entered in the creation of diagnostic models. Two types of models were developed: to classify patients with lung cancer and young healthy volunteers and patients with lung cancer and old healthy subjects. The data set was randomly split: training (70%) and test (30%). *K*-Fold cross-validation method was applied to increase the reliability of the results. The data were split into three parts, one part was used as test data, and the remaining two parts were applied as training data. Different algorithms of machine learning were investigated (random forest, logistic regression, support vector machine, and artificial neural networks) to create a diagnostic model with the highest sensitivity and specificity. Logistic regression classifier with Rosenbross and quasi Newton optimization method was used to train the first model. Support vector machine algorithm can be used with linear and nonlinear kernel function. Sigmoid, linear, polynomial, and radial basis functions were investigated.

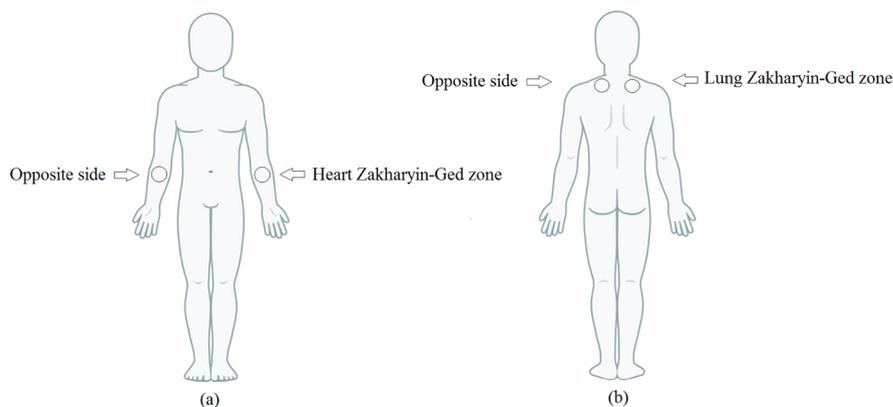


Figure 6. Location of heart (a) and lung (b) Zakharyin-Ged zones and their opposite sides.

The highest performance was observed using the radial basis function. Parameter γ affects the training accuracy: its large values lead to overfitting but in the case of small values, the kernel gradually reduces to a constant function. The range of γ values from 0.008 to 1.8 was investigated. The value of 0.08 was optimal because at lower values, the model accuracy was decreased, but the use of higher values led to the increase in model performance on the training data set without improvements of accuracy on test data set. In the case of random forest algorithm, the influence of different numbers of predictors on model performance was investigated. The highest accuracy on both test and training sets was achieved using three predictors. Multilayer perceptron with one hidden layer neural networks were applied for the diagnostic model creation. Different topologies of the neural network (1000) were tested; the best neural network was chosen. The hidden layer contained five neurons and the output layer included two neurons for the determination of the disease status. The Broyden–Fletcher–Goldfarb–Shanno algorithm was used for the training of the neural networks. Exponential activation function was used for the hidden layer; identity—for the output layer.

The performances for both training and test data were calculated for each model. The diagnostic accuracy of obtained models was compared.

Mann–Whitney test ($p = 0.05$) was used to identify the differences in ratio values between central and peripheral tumor localization groups. Statistically significant differences in three main histological tumor type groups (small cell carcinoma, squamous cell carcinoma, and adenocarcinoma) were evaluated by means of the Kruskal–Wallis test ($p = 0.05$).

Results obtained by skin measurements were analyzed in terms of the Zakharyin–Ged zone and its opposite side ratio. All parameters were normalized by the transformation $\log(x + 10^6)$. Normalization has led to changing the distribution of all parameters to normal, which was evidenced by the Kolmogorov–Smirnov test. One-way analysis of variance (ANOVA) was used for the selection of parameters with statistically significant differences in the investigated groups of participants (young and old healthy and lung cancer). Discriminant analysis and neural networks were applied to develop diagnostic models. Their performance was compared. In the case of neural networks, various topologies of the neural network were tested; the best neural network was chosen. The hidden layer contained nine neurons and the output layer included two neurons for the disease status determination. The Broyden–Fletcher–Goldfarb–Shanno algorithm was used to train the neural networks. Hyperbolic tangent was the hidden layer activation function; softmax–output layer. *K*-Fold cross-validation was applied to validate the neural network model. Considering low number of samples, the data set was split into five parts, one part was used as test data, and the remaining four parts were applied as a training data set.

AUTHOR INFORMATION

Corresponding Author

Elina Gashimova – Department of Analytical Chemistry,
Kuban State University, Krasnodar 350040, Russia;
orcid.org/0000-0002-9618-7580;
Email: elina.gashimova@yandex.ru

Authors

Azamat Temerdashev – Department of Analytical Chemistry,
Kuban State University, Krasnodar 350040, Russia

Vladimir Porkhanov – Research Institute—Regional Clinical
Hospital No 1 n.a. Prof. S.V. Ochapovsky, Krasnodar
350086, Russia

Igor Polyakov – Research Institute—Regional Clinical
Hospital No 1 n.a. Prof. S.V. Ochapovsky, Krasnodar
350086, Russia

Dmitry Perunov – Research Institute—Regional Clinical
Hospital No 1 n.a. Prof. S.V. Ochapovsky, Krasnodar
350086, Russia

Ekaterina Dmitrieva – Department of Analytical Chemistry,
Kuban State University, Krasnodar 350040, Russia

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c06132>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was conducted with the financial support of the Russian Science Foundation and Krasnodar Region Administration (project no. 22-13-20018) applying the scientific equipment of the Center for Environmental Analysis at the Kuban State University.

REFERENCES

- (1) Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33.
- (2) Gouzerh, F.; Bessière, J. M.; Ujvari, B.; Thomas, F.; Dujon, A. M.; Dormont, L. Odors and cancer: Current status and future directions. *Biochim. Biophys. Acta, Rev. Cancer* **2022**, *1877*, 188644.
- (3) Hua, Q.; Zhu, Y.; Liu, H. Detection of volatile organic compounds in exhaled breath to screen lung cancer: a systematic review. *Future Oncol.* **2018**, *14*, 1647–1662.
- (4) Casas-Ferreira, A. M.; Nogal-Sánchez, M. D.; Pérez-Pavón, J. L.; Moreno-Cordero, B. Non-separative mass spectrometry methods for non-invasive medical diagnostics based on volatile organic compounds: A review. *Anal. Chim. Acta.* **2019**, *1045*, 10.
- (5) Marzorati, D.; Mainardi, L.; Sedda, G.; Gasparri, R.; Spaggiari, L.; Cerveri, P. A review of exhaled breath key role in lung cancer diagnosis. *J. Breath Res.* **2019**, *13*, 034001.
- (6) Kim, C.; Raja, I. S.; LeeRaja, J.-M.; Lee, J. H.; Kang, M. S.; Lee, S. H.; Oh, J.-W.; Han, D.-W. Recent trends in exhaled breath diagnosis using an artificial olfactory system. *Biosensors* **2021**, *11*, 337.
- (7) Pesesse, R.; Stefanuto, P.-H.; Schleich, F.; Louis, R.; Focant, J.-F. Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC × GC-TOFMS. *J. Chromatogr. B* **2019**, *1114–1115*, 146–153.
- (8) Callol-Sanchez, L.; Munoz-Lucas, M. A.; Gomez-Martin, O.; Maldonado-Sanz, J. A.; Civera-Tejuca, C.; Gutierrez-Ortega, C.; Rodriguez-Trigo, G.; Jareno-Esteban, J. Observation of nonanoic acid and aldehydes in exhaled breath of patients with lung cancer. *J. Breath Res.* **2017**, *11*, 026004.
- (9) Schallschmidt, K.; Becker, R.; Jung, C.; Bremser, W.; Walles, T.; Neudecker, J.; Leschber, G.; Frese, S.; Nehls, I. Comparison of Volatile Organic Compounds from Lung Cancer Patients and Healthy Controls—Challenges and Limitations of an Observational Study. *J. Breath Res.* **2016**, *10*, 046007.
- (10) Corradi, M.; Poli, D.; Banda, I.; Bonini, S.; Mozzoni, P.; Pinelli, S.; Alinovi, R.; Andreoli, R.; Ampollini, L.; Casalini, A.; Carbonegani, P.; Goldoni, M.; Mutti, A. Exhaled breath analysis in suspected cases of non-small-cell lung cancer: a cross-sectional study. *J. Breath Res.* **2015**, *9*, 027101.
- (11) Broza, Y. Y.; Kremer, R.; Tisch, U.; Gevorkyan, A.; Shiban, A.; Best, L.; Haick, H. Nanomaterial-based breath test for short-term follow-up after lung tumor resection. *Nanomed. Nanotechnol. Biol. Med.* **2013**, *9*, 15–21.

- (12) Handa, H.; Usuba, A.; Maddula, S.; Baumbach, J. I.; Mineshita, M.; Miyazawa, T. Exhaled breath analysis for lung cancer detection using ion mobility spectrometry. *PLoS One* **2014**, *9*, 0114555.
- (13) Pleil, J. D.; Hansel, A.; Beauchamp, J. Advances in proton transfer reaction mass spectrometry (PTR-MS): applications in exhaled breath analysis, food science, and atmospheric chemistry. *J. Breath Res.* **2019**, *13*, 039002.
- (14) Meng, S.; Li, Q.; Zhou, Z.; Li, H.; Liu, X.; Pan, S.; LiPan, M.; Wang, L.; Guo, Y.; Qiu, M.; Wang, J. Assessment of an exhaled breath test using high-pressure photon ionization time-of-flight mass spectrometry to detect lung cancer. *JAMA Netw. Open* **2021**, *4*, No. e213486.
- (15) Tirzite, M.; Bukovskis, M.; Strazda, G.; Jurka, N.; Taivans, I. Detection of lung cancer in exhaled breath with an electronic nose using support vector machine analysis. *J. Breath Res.* **2017**, *11*, 036009.
- (16) Kort, S.; Brusse-Keizer, M.; Gerritsen, J. W.; Schouwink, H.; Citgez, E.; de Jongh, F.; van der Maten, J.; Samii, S.; van den Bogart, M.; van der Palen, J. Improving lung cancer diagnosis by combining exhaled-breath data and clinical parameters. *ERJ Open Res* **2020**, *6*, 00221.
- (17) Gasparri, R.; Santonico, M.; Valentini, C.; Borri, A.; Petrella, F.; Maisonneuve, P.; Pennazza, G.; D'Amico, A.; Di Natale, C. D.; Paolesse, R.; Spaggiari, L. Volatile signature for the early diagnosis of lung cancer. *J. Breath Res.* **2016**, *10*, 016007.
- (18) Mazzone, P. J.; Wang, X.-F.; Xu, Y.; Mekhail, T.; Beukemann, M. C.; Na, J.; Kemling, K. S.; Suslick, M.; Sasidhar, M. Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer. *J. Thorac. Oncol.* **2012**, *7*, 137–142.
- (19) Chang, J. E.; Lee, D. S.; Ban, S. W.; Oh, J.; Jung, S. H.; Kim, S. J.; Park, K.; Persaud, S.; Jheon, S. Analysis of volatile organic compounds in exhaled breath for lung cancer diagnosis using a sensor system. *Sens. Actuators, B* **2018**, *255*, 800–807.
- (20) Handa, H.; Usuba, A.; Maddula, S.; Baumbach, J. I.; Mineshita, M.; Miyazawa, T. Exhaled breath analysis for lung cancer detection using ion mobility spectrometry. *PLoS One* **2014**, *9*, 0114555.
- (21) Zou, Y.; Wang, Y.; Jiang, Z.; Zhou, Y.; Chen, Y.; Hu, Y.; Jiang, G.; Xie, D. Breath profile as composite biomarkers for lung cancer diagnosis. *Lung Cancer* **2021**, *154*, 206–213.
- (22) Rudnicka, J.; Walczak, M.; Kowalkowski, T.; Jezierski, T.; Buszewski, B. Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography-mass spectrometry versus trained dogs. *Sens. Actuators, B* **2014**, *202*, 615–621.
- (23) Liu, B.; Yu, H.; Zeng, X.; Zhang, D.; Gong, J.; Tian, L.; Qian, J.; Zhao, L.; Zhang, S.; Liu, R. Lung cancer detection via breath by electronic nose enhanced with a sparse group feature selection approach. *Sens. Actuators, B* **2021**, *339*, 129896.
- (24) Ruzsanyi, V.; Mochalski, P.; Schmid, A.; Wiesenhofer, H.; Klieber, M.; Hinterhuber, H.; Amann, A. Ion mobility spectrometry for detection of skin volatiles. *J. Chromatogr. B* **2012**, *911*, 84–92.
- (25) Zou, Z.; He, J.; Yang, H. An experimental method for measuring VOC emissions from individual human whole-body skin under controlled conditions. *Build. Environ.* **2020**, *181*, 107137.
- (26) Martin, H. J.; Reynolds, J. C.; Riazanskaia, S.; Thomas, P. High throughput volatile fatty acid skin metabolite profiling by thermal desorption secondary electrospray ionisation mass spectrometry. *Analyst* **2014**, *139*, 4279–4286.
- (27) Vishinkin, R.; Busool, R.; Mansour, E.; Fish, F.; Esmail, A.; Kumar, P.; Gharaa, A.; Cancilla, J. C.; Torrecilla, J. S.; Skenders, G.; Leja, M.; Dheda, K.; Singh, S.; Haick, H. Profiles of Volatile Biomarkers Detect Tuberculosis from Skin. *Adv. Sci.* **2021**, *8*, 2100235.
- (28) Muzhikov, V.; Vershinina, E.; Belenky, V.; Muzhikov, R. Assessing the Links Between Anthropometrics Data and Akabane Test Results. *J. Acupunct. Meridian. Stud.* **2018**, *11*, 31–38.
- (29) Gashimova, E.; Temerdashev, A.; Porkhanov, V.; Polyakov, I.; Perunov, D.; Azaryan, A.; Dmitrieva, E. Investigation of different approaches for exhaled breath and tumor tissue analyses to identify lung cancer biomarkers. *Heliyon* **2020**, *6*, No. e04224.
- (30) Gashimova, E.; Osipova, A.; Temerdashev, A.; Porkhanov, V.; Polyakov, I.; Perunov, D.; Dmitrieva, E. Exhaled breath analysis using GC-MS and an electronic nose for lung cancer diagnostics. *Anal. Methods* **2021**, *13*, 4793–4804.
- (31) Ligor, T.; Pater, L.; BuszewskiPater, B. Application of an artificial neural network model for selection of potential lung cancer biomarkers. *J. Breath Res.* **2015**, *9*, 027106.
- (32) Song, G.; Qin, T.; Liu, H.; Xu, G.-B.; Pan, Y.-Y.; XiongPan, F.-X.; Gu, K.-S.; Sun, G.-P.; Chen, Z.-D. Quantitative breath analysis of volatile organic compounds of lung cancer patients. *Lung Cancer* **2010**, *67*, 227–231.
- (33) Gasparri, R.; Santonico, M.; Valentini, C.; SeddaValentini, G.; Borri, A.; Petrella, F.; Maisonneuve, P.; Pennazza, G.; D'Amico, A.; Di Natale, C. D.; Paolesse, R.; Spaggiari, L. Volatile signature for the early diagnosis of lung cancer. *J. Breath Res.* **2016**, *10*, 016007.
- (34) D'Amico, A.; Pennazza, G.; Santonico, M.; Martinelli, E.; Roscioni, C.; Galluccio, G.; Paolesse, R.; Di Natale, C. D. An investigation on electronic nose diagnosis of lung cancer. *Lung Cancer* **2010**, *68*, 170.
- (35) Shuba, A.; Kuchmenko, T.; Umarchanov, R.; Chernitskiy, A. Portable E-nose for Diagnostic of Inflammation and Diverse Variation in Health Status of Humans and Animals. *Proceedings of the Fifth International Conference on Advances in Sensors, Actuators, Metering and Sensing (ALLSENSORS 2020)*; International Academy, Research, and Industry Association (IARIA), 2020; pp 56–62.