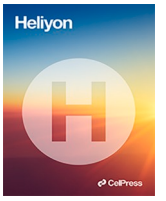
Heliyon 10 (2024) e39028

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Quantification of cancer biomarkers in urine using volatilomic approach

Elina Gashimova [a,*], Azamat Temerdashev [a], Dmitry Perunov [b], Vladimir Porkhanov [b], Igor Polyakov [b], Alexey Podzhivotov [a], Ekaterina Dmitrieva [c]

[a] Kuban State University, Stavropol'skaya St. 149, Krasnodar, 350040, Russia
[b] Research Institute – Regional Clinical Hospital, No 1 n.a. Prof. S.V. Ochapovsky, 1 May St. 167, Krasnodar, 350086, Russia
[c] Shenzhen MSU-BIT University, International University, Park Road, 1, Shenzhen, Longgang District, Guangdong Province, 518172, PR China

ARTICLE INFO

Keywords:
Urine analysis
Volatile organic compounds
GC-MS
Cancer
Diagnostics

ABSTRACT

Urine analysis is an attractive approach for non-invasive cancer diagnostics. In this study, a procedure for the determination of volatile organic compounds (VOCs) in human urine (acetone, acetonitrile, dimethylsulfide, dimethyl disulfide, dimethyl trisulfide, hexane, benzene, toluene, 2-butanone, 2-pentanone, pentanal) has been described including sample preparation using preconcentration of analytes in sorbent tubes followed by gas chromatography with mass spectrometry (GC-MS). Fractional factorial design and constrained surfaces design were used to optimize preconcentration of VOCs in sorbent tubes. The procedure was validated by analysis of synthetic urine containing VOC standards in the concentration range of 1–5000 ng/mL. Optimized procedure was applied to analyze urine samples of 89 healthy volunteers and 85 patients with cancer of various localizations: 42 patients with lung cancer, 25 – colon, 3 – stomach, 2 – prostate, 2 – esophageal, 2 – pancreas, 2 – kidney, 1 – ovarian, 1 – cervical, 1 – skin, 1 – liver. Concentrations of 2-butanone, 2-pentanone, acetonitrile, and benzene were found different in urine of patients with cancer and healthy individuals. Influence of cancer localization and tumor, nodule, metastasis stage on urine VOC profile was considered. The approach of using ratios of VOCs to the main ones instead of concentrations was considered. A diagnostic model based on significantly different VOC ratios was created to classify healthy individuals and patients with cancer using artificial neural network (ANN). The model was validated during construction by means of 3-fold cross-validation. Average area under receiver operating characteristic (ROC) curve on test dataset was 0.886. Average sensitivity and specificity of the created model were 91 % and 82 %.

1. Introduction

Diagnostics of various diseases using biomarkers from different matrices is one of the most interesting trends in modern medicine [1,2]. Implementation of diagnostic procedures with non-invasive sampling is especially interesting. It allows extending method throughput as well as enables patients to undergo diagnostics without discomfort. Searching for biomarkers among volatile organic compounds (VOCs) is a quickly evolving area which covers a few various matrices such as exhaled breath, urine, saliva, feces, sweat [3,

* Corresponding author.
E-mail address: elina.gashimova@yandex.ru (E. Gashimova).

https://doi.org/10.1016/j.heliyon.2024.e39028
Received 19 June 2024; Received in revised form 23 August 2024; Accepted 4 October 2024
Available online 5 October 2024

4]. VOCs can be directly obtained from the gas phase, which makes VOC profile attractive due to simple sample treatment and sample storage compared to other kinds of metabolites, which must be extracted from blood or body fluids prior to analysis [5]. VOCs can be easily transferred to the gaseous phase by heating the sample. Integrity of the sample during long-term storage is achieved by freezing [6]. Exhaled breath, saliva, urine, feces are actively explored as sources of volatile biomarkers to diagnose different diseases non-invasively [3]. The potential to develop approaches for diagnosing digestive neoplasia [3], infectious diseases [4], chronic liver disorders [7], gastrointestinal diseases [8,9] using VOC profile of various matrices is actively discussed in literature.

Despite the latest progress in the development of cancer medicines, cancer mortality rate is still high, which can be explained by the lack of ability to diagnose the disease in the early stages. Early diagnostics significantly improves course of the disease and reduces cancer mortality rate. Therefore, developing novel non-invasive approaches for diagnosing cancer is especially actual. Urine is increasingly recognized as a valuable subject of analysis for biomarkers discovery. It can be easily collected in a large amount. A sample of urine is steadier than exhaled breath, since exhaled breath is more sensitive to environmental exposures and immediate factors like smoking status and diet [10]. Urine reflects longer-term metabolic changes and dietary influences but may not respond quickly to acute environmental factors, which makes urine more stable [11]. Non-invasiveness, ready availability, and simple sample collection make urine one of the best subjects for analysis. Determination of volatile organic compounds (VOCs) in urine can be performed using several analytical methods. The simplest and the most applicable approach to routine analysis is using electronic noses with various analytical properties: metal oxide sensors [12–14], quartz crystal microbalance sensors [15], Cyranose 320 [16]. However, sensor systems are not capable to identify the biomarkers. They react on a full VOC profile of the sample which contains many confounding macro compounds. For this reason, the list of biomarkers should be determined to develop a selective sensor array.

Mass spectrometry is a high-powered analytical tool applied to analyze urine VOC profile. Utilization of proton transfer reaction mass spectrometry (PTR-MS) for urine analysis without sample treatment is described in paper [17]. Authors [18] demonstrate application of gas chromatography-ion mobility spectrometry (GC-IMS) do diagnose prostate cancer. High accuracy of diagnosing colorectal cancer and adenoma (95 %) was achieved applying urine analysis by field asymmetric ion mobility spectrometry (FAIMS) [19].

Despite the diversity of analytical methods as well as development of novel types of analytical tools, GC-MS remains the most powerful tool which can perform identification and quantification of many VOCs in the sample at the same time. Many works have been devoted to the GC-MS analysis of urine for the cancer diagnostics. One of them compared the effectiveness of field asymmetric ion mobility spectrometry (FAIMS), selected ion flow tube mass spectrometry (SIFT-MS), and gas chromatography with mass spectrometry (GC-MS) for colorectal cancer diagnostics [20]. GC-MS showed the highest clinical utility to diagnose colorectal cancer. Various kinds of sample preconcentration as well as analysis conditions were used. Solid phase microextraction (SPME) is the most popular pre-concentration technique. A majority of researchers provide the results of using several SPME fibers with subsequent choosing the best one. The number of peaks, peak area, and reproducibility are often used as criteria of SPME fiber priority [21–24]. Another approach to preconcentrate the analytes is using sorbent tubes or needle traps [25,26]. Chromatographic columns of various polarity are applied to separate the analytes [27–29]. To increase the efficiency of transferring analytes into gaseous phase, acid medium [24,30] or salting out agents [25,27] can be applied. Diverse sample preparation techniques as well as conditions of concentrating VOCs from urine sample are proposed, however the lack of a unique scheme of sample preparation requires optimization of this step of analysis. A lot of VOCs including 2-pentanone [12,21], dimethylsulfide [25], dimethyl disulfide [23], hexane, 2-butanone [25], acetone [12,25], etc. were shown as urinary biomarkers of cancer. Considering variations in putative cancer biomarkers in different studies [12,21–23,25, 26,30], the approach must identify a wide list of urine volatile metabolites to be studied as potential cancer biomarkers.

The aim of the present study is to optimize conditions of urine treatment and VOCs concentration from human urine including pH, salting agent, sample volume, temperature, temperature stabilization time and sorption time with the help of fractional factorial design and constrained surfaces design. The optimized procedure was validated using synthetic urine. The developed approach was applied for urine analysis of healthy subjects and patients with cancer of various localizations with subsequent statistical analysis and modelling using artificial neural networks.

2. Material and methods

2.1. Reagents

Acetonitrile, n-hexane, benzene, toluene, deuterated acetone (acetone-D6) (>99 %) were acquired from Sigma-Aldrich (St. Louis, USA), dimethylsulfide, dimethyl disulfide, dimethyl trisulfide, 2-butanone, 2-pentanone, pentanal were obtained from Mucklin (Shanghai, China). Acetone (99.9 %) and methanol were purchased from Ecos (Moscow, Russia). Acetic acid (98 %), uric acid, creatinine, urea, sodium sulfate, sodium hydroxide, sodium chloride, sodium citrate dihydrate, sodium tetraborate, sodium hydrogen phosphate dihydrate, sodium dihydrogen phosphate dihydrate, disodium hydrogen phosphate dihydrate, ammonium chloride, potassium oxalate monohydrate, potassium chloride, calcium chloride, magnesium sulfate heptahydrate, dipotassium phosphate (99 %) were purchased from Vectron (Moscow, Russia). A Milli-Q simplicity system (Milli-Q, Millipore, Molsheim, France) was used to obtain 18.2 MΩ cm water.

2.2. Preparation of standard solutions

Synthetic urine was prepared as described by N. Sarigul and co-authors [31]. A 1 mg/mL standard solution of acetone, 2-butanone, 2-pentanone, pentanal, acetonitrile was prepared in synthetic urine. A 1 mg/mL standard solution of dimethylsulfide, dimethyl

disulfide, dimethyl trisulfide, n-hexane, benzene, toluene was prepared in methanol. The former standard solution was diluted in synthetic urine to obtain calibration solutions with the following concentrations: 0.5, 1, 3, 5, 10, 50, 100, 300, 500, 750, 1000 ng/mL. The latter standard solution was first diluted in methanol to prepare intermediate solutions which were then diluted in synthetic urine. Dimethyl disulfide, dimethyl trisulfide were studied separately at the following concentrations: 10, 50, 100, 150, 200, 500, 750, 1000, 1500, 2500, 5000 ng/mL. The standard solutions were stored at 4 °C. The rest solutions were used one time. Acetone-D6 (internal standard) standard solution with the concentration of 1 mg/mL was prepared in synthetic urine; its final concentration was 50 ng/mL.

2.3. Instruments

Porapak N (50/80 mesh), Chromosorb 106 (60/80 mesh), Tenax TA (35/60 mesh), and multibed sorbent (Tenax GR (80/100 mesh), Carbopack B (60/80 mesh), Carbosieve SIII (60/80 mesh)) (Chromatec, Yoshkar-Ola, Russia) were used for the VOCs pre-concentration. A PV-2 aspirator (Chromatec, Yoshkar-Ola, Russia) was used to transfer VOC from urine. A gas chromatograph (Chromatec crystal 5000.2, Yoshkar-Ola, Russia) combined with electron ionization single quadrupole mass spectrometer (Chromatec MSD, Yoshkar-Ola, Russia) equipped with a TD2 thermal desorber (Chromatec, Yoshkar-Ola, Russia) was deployed for GC-MS analysis of the samples. A Supelco Supel-Q PLOT (30 m × 0.32 mm) column was used to separate the analytes. The acquisition of chromatographic data was executed applying the Chromatec Analytic (Chromatec, Yoshkar-Ola, Russia) software and NIST 2017 mass spectrum library, Version 2.3 (Gatesburg, USA). Conditions of thermal desorption, chromatographic separation, and mass spectrometric analysis were optimized earlier [32] (Table 1).

2.4. Human subjects and sample collection

First-morning urine samples were collected in standard universal sterile specimen containers from 89 healthy volunteers and 85 patients with cancer of various localizations: 42 patients with lung cancer, 25 – colon, 3 – stomach, 2 – prostate, 2 – esophageal, 2 – pancreas, 2 – kidney, 1 – ovarian, 1 – cervical, 1 – skin, 1 – liver cancer. The diagnosis of cancer patients was confirmed by histology, which was conducted according to the requirements for each cancer localization. Patients with comorbidities were excluded from the study. Information on participants is presented in Table 2. All the samples of cancer patients were obtained before treatment. Healthy status of the rest volunteers was confirmed by annual medical examination report. Informed consent was obtained from each subject at the time of enrolment. Local ethics committee of State budgetary healthcare institution Research Institute—Regional Clinical Hospital № 1 named after Professor S.V. Ochapovsky approved the study. Urine samples were frozen within an hour after collection and stored at −20 °C with subsequent analysis within one week after collection.

Table 1
Conditions of TD-GC-MS analysis.

Equipm-ent	Parameter			Value	
TD	Carrier gas			Nitrogen	
	Valve temperature, °C			150	
	Transfer line temperature, °C			180	
	Desorption temperature, °C		Tenax TA	250	
			Multi - bed	250	
			Chromosorb 106	220	
			Porapak N	150	
	Desorption time, min			5	
	Initial trap temperature, °C			−10	
	Final trap temperature, °C		Tenax TA	250	
			Multi - bed	250	
			Chromosorb 106	220	
			Porapak N	150	
	Trap heating time, min			2	
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 range, amu			36–220	
	Electron impact ionization, eV			70	
	Temperature program				
	Column	Heating rate, °C/min	Temperature, °C	Time, min	Carrier-gas flow rate, mL/min
	Supelco Supel-Q PLOT	0	50	0	1.30
		10	150	0	
		6	220	7	
		4	250	0	

Table 2
Characteristics of participants.

Group	Feature	Value
Healthy volunteer	Number (m/f)	89 (35/54)
	Age, range (median)	55 (39–67)
Cancer patient	Number (m/f)	85 (58/27)
	Age, range (median)	63 (45–73)
	Type of cancer	
	Lung	42
	Colon	25
	Stomach	3
	Bladder	3
	Prostate	2
	Esophageal	2
	Pancreas	2
	Kidney	2
	Ovarian	1
	Cervical	1
	Skin	1
	Liver	1
	Tumor nodule metastasis (TNM) stage	
	100	3
	101	1
	110	1
	120	2
	200	5
	201	5
	210	3
	211	2
	220	4
	231	2
	300	8
	301	4
	310	6
	311	2
	320	5
	321	6
	330	1
	400	5
	401	5
	410	4
	420	5
	421	4
	430	1
	431	1

2.5. Statistical analysis

Optimization of sample treatment procedure was fulfilled using a StatSoft STATISTICA software (version 10). Fractional factorial design was used to select factors affecting the sample treatment process. Optimization of temperature and time of sorption was conducted using a constrained surfaces design. Peak areas of main analytes: acetone, dimethyl sulfide, dimethyl disulfide, dimethyl trisulfide acetonitrile, benzene, toluene, 2-pentanone, 2-butanone, pentanal, hexane were considered to optimize sample preparation for the best performance conditions. Effects of experimental conditions including pH, salting agent, sample volume, temperature, temperature stabilization time, and sorption time at different levels on VOC peak areas were estimated for each compound separately.

Statistical analysis was conducted for VOCs occurring in more than 50 % of samples at least in group of cancer patients or healthy volunteers. The peak area ratio of VOC to the internal standard was used for statistical analysis. Additionally, ratios of VOCs to the main ones were used. For this, VOCs with 100 % frequency of occurrence were applied as a denominator and VOCs with frequency of occurrence exceeding 50 % were used as a numerator. The approach was described earlier in detail on the example of exhaled breath VOC profile [33]. The normality of distribution was determined by means of the Shapiro–Wilk test. The distribution of data was not normal. Therefore, non-parametric tests were used to determine statistically significant differences. The influence of common factors for both healthy individuals and patients with cancer, age, and gender on urine composition was evaluated in each group separately. For this, Mann–Whitney *U* test ($p < 0.05$) and Spearman rank correlation ($p < 0.05$) were used. Case-control variations were accessed using Mann–Whitney *U* test ($p < 0.05$). Differences between urine VOC profiles of patients with lung and colorectal cancer, TNM stage were assessed applying Mann–Whitney *U* test ($p < 0.05$) and Spearman rank correlation ($p < 0.05$).

A classification model was created to distinguish healthy subjects from cancer patients using artificial neural networks (ANN). The dataset was divided into three datasets: training dataset was used to teach algorithms how to discriminate healthy subjects and cancer patients. Training dataset includes 60 % of samples. Control dataset (10 % of samples) was used for the current assessment of the

quality of training, which makes it possible to prevent overfitting of the neural network and allows saving the recognition efficiency on test dataset [34]. Test dataset contained 30 % of samples. It was used for evaluating the performance of the trained model. The Broyden – Fletcher – Goldfarb – Shanno algorithm was used to train ANN [35]. Recurrent feedforward multilayer perceptron with fully connected one hidden layer was created. Various topologies of neural network were tested. The best one has the following topology: input layer consists of 17 VOC ratios; hidden layer contains 12 neurons; output layer includes 2 neurons, which evaluated whether the sample belonged to the healthy or cancer group. The activation function enables the model to learn complex patterns and relationships within the data. Activation function aimed to connect input and hidden layers was hyperbolic tangent, which provides benefits such as non-linearity, symmetry, and improved gradient flow; SoftMax was used in case of hidden and output layers, which outputs into a meaningful probability distribution [36]. Three-fold cross-validation was used to provide reliable results. Receiver operating characteristic (ROC) curves were constructed for each test datasets. Sensitivity and specificity of the models on training and test datasets were calculated.

3. Results and discussion

3.1. Optimization of conditions for VOCs preconcentration in sorbent tubes

Many analytical methods are applied to analyze urine VOC profile [15,20,26,27]. A diagnostic method applicable for clinical practice can be developed only after identification of potential biomarkers. For this, GC-MS remains the most suitable due to the ability to identify and quantify a huge number of analytes simultaneously. Sorbent tubes is one of the most reliable preconcentration approaches, which allows preconcentrating the analytes quantitatively. Considering variations in urine VOC profile and lack of consistency in results of various researchers in terms of putative biomarkers, conditions of urine analysis should be optimized, which was conducted in the study using experimental design. Optimization stage was conducted on the samples of healthy volunteers. For this, an averaged sample of 20 healthy persons was used.

The first issue was to determine type of sorbent for the preconcentration of VOCs. For this Porapak N, Chromosorb 106, Tenax TA, and multibed sorbent (Tenax GR, Carbopack B, and Carbosieve SIII) were studied. A 1.5-ml urine sample was heated to 60°C. After that, passive sampling was conducted during 60 min. Then, the tube was analyzed using GC-MS (Table 1). Three parallel measurements were conducted for each sorbent. The greatest number of VOCs was found to be preconcentrated using Tenax TA; also, the abundances of most VOCs were higher in case of Tenax TA, which fits the results of other researchers [37]. Therefore, Tenax TA was used for the preconcentration of urine VOCs.

The VOCs can be preconcentrated in static or dynamic regime, which were compared in the present study. Dynamic regime is expected to be more effective since the predefined volume of the sample is passed through the sorbent tube. However, static regime is simpler to fulfill because additional equipment (aspirator) is not required. Static and dynamic regimes of preconcentration were compared. The same conditions as previously were used: 2 urine samples with value of 1,5 ml were heated to 60°C and thermostated for 60 min. After that, sorbent tube was connected to the flask of each sample by silicone hose with a cap at the end for sorption during 60 min. One of the samples after static sorption was connected to the aspirator to pump the air above the sample through the tube during 1 min at a flow rate of 20 mL/min. The results were identical. Therefore, static approach was chosen.

Many factors such as salting agent addition or pH regulation are applied optionally in various studies [12,22,24,27,28]. It requires studying the efficiency of these stages and their necessity for analysis. Other factors, i.e., sample volume, temperature, temperature stabilization time and sorption time, also require optimization. The best way to optimize several parameters at the same time is using experimental design algorithms. The aim of the first step of experimental design was to determine the number of significant factors. For this, a 2^{6-1} fractional factorial design was used. The design allows us to study the influence of a much larger number of factors as well as their interaction with a reduced number of experiments [38]. It is achieved by omitting some combinations of factors. The levels of each factor were chosen for covering a wide range of studied conditions which allows estimating effect of each factor (Table 3). The experiment was repeated twice to provide reliable results. The Pareto charts were constructed for each VOC (Figs. 1 and 2). Two factors, namely, salting agent and temperature stabilization time as well as combination of factors, were found not to affect the preconcentration efficiency for almost all analytes. In other studies, salt was added, but its effect as well as rationality is not explained [27]. Another work demonstrates an increase in abundance of about 15 % after salt addition using univariate optimization strategy [25]. A temperature stabilization time value of 10 min and a lack of salting-out agent addition were selected in further experiments.

Considering a significant effect of sample volume on some VOCs without a significant influence in case of combination of the factor with other ones, sample volume was optimized separately from other factors. The sample volume was varied from 0.5 to 3.5 ml with a

Table 3
The minimum and maximum ranges for each factor for optimizing procedure of concentrating urine VOCs.

Factor	Minimum	Maximum
Sample volume, ml	0,5	1,5
Temperature, °C	40	60
Temperature stabilization time, min	10	60
Sorption time, min	10	60
pH	4	8
Salting agent (NaCl), mg	0	20 %

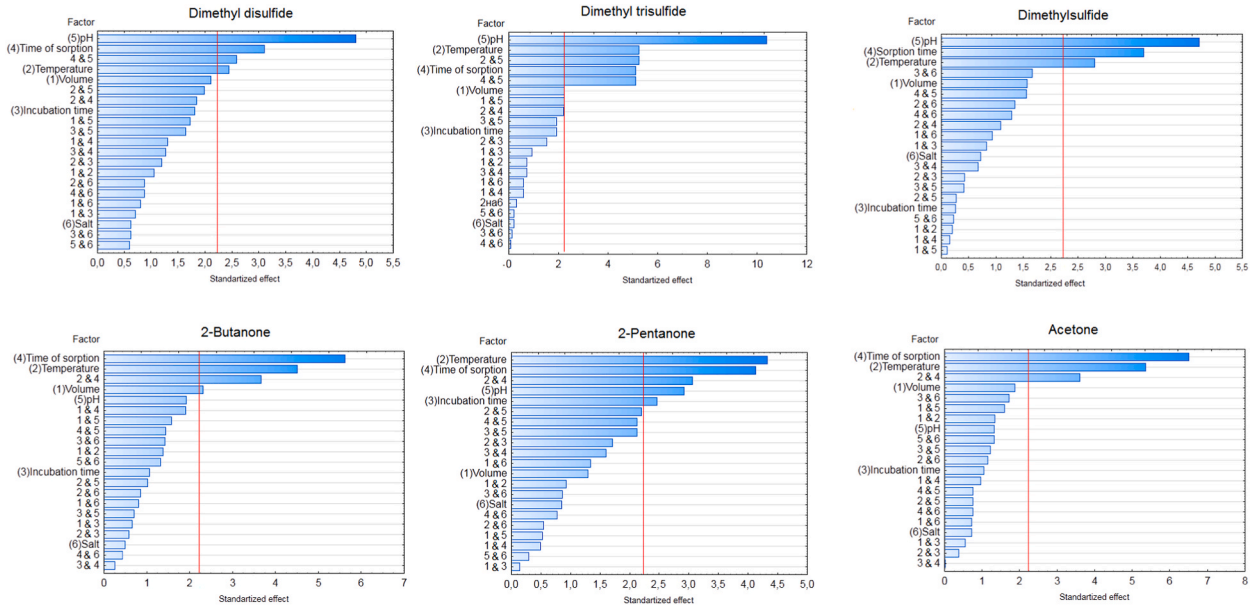


Fig. 1. Pareto charts for dimethyl sulfide, dimethyl disulfide, dimethyl trisulfide, 2-butanone, 2-pentanone, acetone.

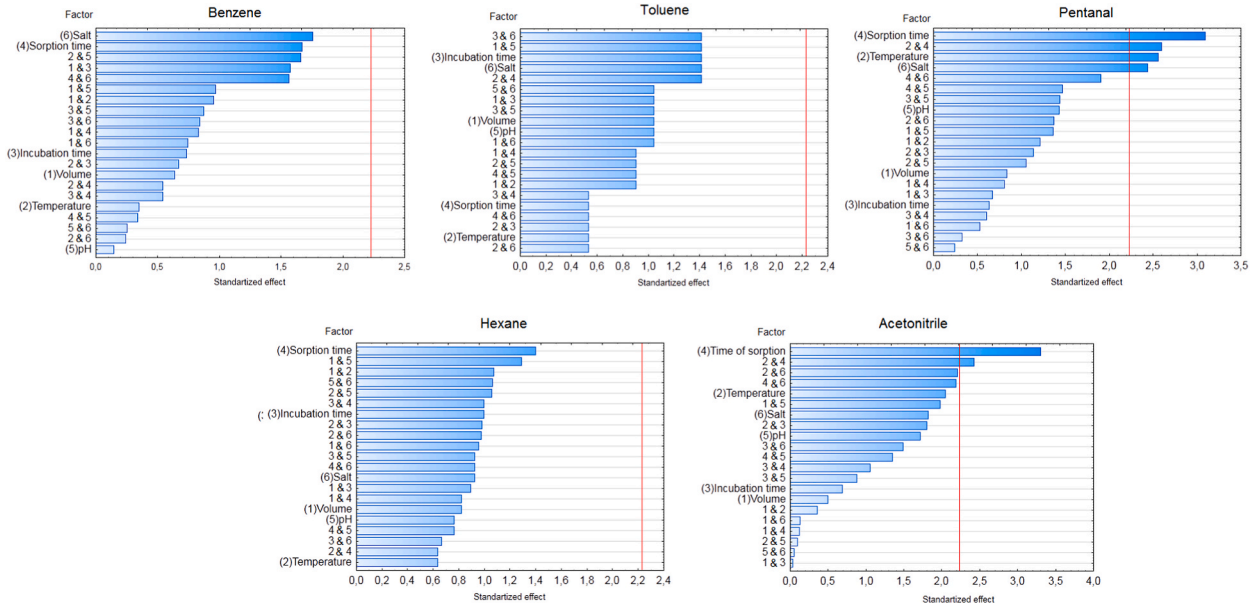


Fig. 2. Pareto charts for benzene, toluene, pentanal, hexane, acetonitrile.

step of 0.5 ml while other factors were fixed. A volume of 2.5 ml provided the highest peak areas for all studied VOCs. A decrease in peak areas in case of upper values can be explained by uneven distribution of the analytes between gas and aqueous phases. Optimization of pH was also conducted in a separate regime because of mixed impact on various analytes (Figs. 1 and 2). VOC peak areas achieved their highest values at pH 3, which was chosen for further analysis.

The effect of temperature and sorption time in combination was significant for many VOCs (Figs. 1 and 2). Therefore, these factors were optimized in combination. For this, 3-level constrained surfaces design was used. Temperature and time levels were 40, 50, 60°C and 30, 45, 60 min, respectively. The highest point of temperature 60°C was chosen to avoid destruction of high-molecular weight compounds of a sample. Time was restricted by 60 min to escape limitations on concentrating macro compounds in high levels with subsequent overload of electron multiplier.

Response surface plots were constructed for each VOC to determine the optimal values of temperature and time of preconcentration. Figs. 3 and 4 illustrate the response surface plots for each VOC. Based on the results obtained, the highest values of both

parameters provide the highest peak areas. The final sample preparation procedure was as follows: 2.5 ml of urine was adjusted by acetic acid to pH 3 with following incubation of the sample during 10 min at 60°C following preconcentration of VOCs during 60 min on a Tenax TA sorbent tube.

The method was evaluated for limits of detection (LOD), limits of quantification (LOQ), linearity, and intraday precision for the main VOCs present in urine. Peak area of the most abundant *m/z* was used as a quantitative analytical parameter, which was normalized relative to the peak area of the internal standard. LOD was assessed as the signal-to-noise ratio of 3, LOQ was calculated as the concentration of analyte determined with 15 % error. Calibration curves were constructed six times. Linearity was evaluated by the regression curves of standard solutions and expressed by the coefficient of determination R^2. The obtained parameters are presented in Table 4. Optimized conditions were found to have a linear response (R^2 in a range of 0.973–0.999) for all analytes. Intraday precision (expressed as RSD) was below 15 % for all analytes, which was found as satisfactory.

3.2. GC-MS analysis

Urine samples of 174 participants were analyzed using the optimized procedure. The results revealed applicability of proposed approach for the urine samples from both healthy individuals and cancer patients. All the VOCs were found in samples of both groups (Table 5). Concentrations of VOCs were calculated using calibration curves. As a rule, quantitative analysis is not conducted in studies aimed to identify biomarkers among VOCs in urine [25,27,30] Drabinskaya et al. [39] conducted quantitative assessment of VOCs in urine, however, several VOCs, e.g., dimethyl trisulfide, were not quantified in real samples.

Statistical analysis was conducted in relation to not only VOCs, but their ratios to the main ones occurring in 100 % of samples. This approach allows smoothing variations of the samples concerning individual physiological factors as well as highlight the differences, which are common for inter group differences [33]. Search of biomarkers should be conducted only in relation to VOCs, which are not affected by other factors. Age and gender were marked as factors affecting urine VOC profile earlier [40,41]. To avoid false results, influence of age and gender was evaluated in the present study. Systematic influence of age and gender was not observed for urinary VOCs. It can be explained by narrow age ranges of participants as well as the uniform distribution in case of gender. However, the cohort of participants was relatively small. Cohort expansion can modify the situation, which requires a study of influence of the factors in further research.

Concentrations of benzene (p = 0.033), acetonitrile (p = 0.000), 2-butanone (p = 0.001), and 2-pentanone (p = 0.002) were significantly higher in urine samples of cancer patients then in case of healthy individuals. 2-Pentanone was indicated as a biomarker of cancer by many scientists [12,21], which proves our findings. Acetone [12,25], dimethylsulfide [25], dimethyl disulfide [23] were also indicated as biomarkers, which was not observed in the present research. It is worth noting that cohorts of volunteers in the study was significantly larger, than in the studies [12,23,25], which highlights reliability of results obtained in the present study. Lung and colorectal cancer are the most prevalent and responsible for the largest number of cancer deaths [12,21]. Therefore, lung and colorectal cancer are the main cancer localizations in the study. The number of lung and colorectal cancer were 42 and 25 patients

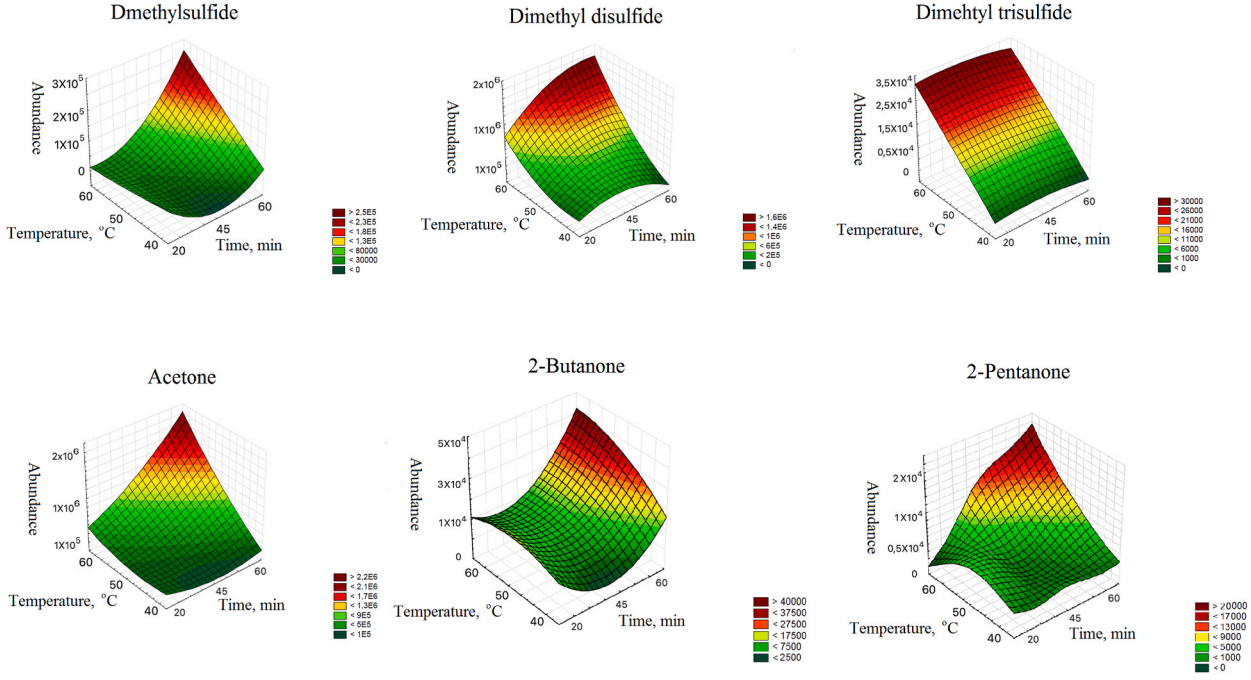


Fig. 3. Response surface plots for dimethyl sulfide, dimethyl disulfide, dimethyl trisulfide, 2-butanone, 2-pentanone, acetone.

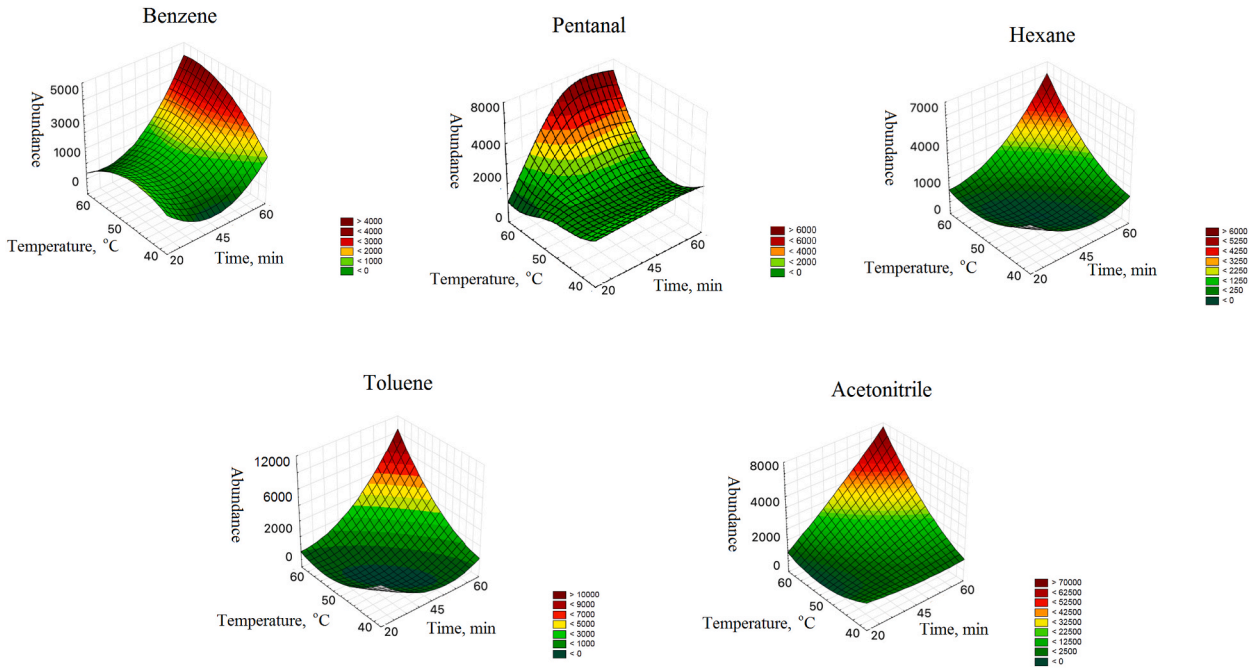


Fig. 4. Response surface plots for benzene, toluene, pentanal, hexane, acetonitrile.

Table 4
Analytical characteristics obtained from the standard solutions.

VOC	CAS number	*m/z*	LOD (ng/mL)[a]	LOQ (ng/mL)[b]	Concentration range (ng/mL)	R2	Interday precision (RSD. %)
Acetone	67-64-1	58	3	10	10–500	0.984	7.4
Benzene	71-43-2	78	3	10	10–500	0.987	12.4
Toluene	108-88-3	91	3	10	10–500	0.971	11.5
Acetonitrile	75-05-8	41	1	5	5–500	0.981	12.1
Hexane	110-54-3	57	3	10	10–500	0.984	13.4
2-Butanone	78-93-3	57	1	5	5–500	0.999	11.8
2-Pentanone	107-87-9	71	1	5	5–500	0.983	12.1
Pentanal	110-62-3	58	1	5	5–300	0.973	13.4
Dimethylsulfide	75-18-3	62	1	5	5–300	0.996	14.0
Dimethyl disulfide	624-92-0	94	50	200	200–2500	0.991	13.4
Dimethyl trisulfide	3658-80-8	126	150	500	500–5000	0.984	14.5

[a] Limit of detection.
[b] Limit of quantification.

accordingly. VOC profiles of them were compared: benzene ($p = 0.036$), acetonitrile ($p = 0.043$), and dimethylsulfide ($p = 0.012$) and some ratios were different between the groups, which shows potential of the VOCs to separate groups of patients with various cancer localizations. The groups as well as groups of various cancer localizations are needed to be expanded to prove the results. TNM influence on urine VOC profile was also assessed: some ratios were significantly correlated with TNM stage. Considering various cancer localizations, the findings are required to be proved by expanding cohorts of volunteers.

The ratios, which were different between healthy and cancer patients, were used as input variables for the creation of a classification model using ANN (Table 6). Average sensitivity, specificity, and overall accuracy of the model based on 3-fold cross validation were 90 %, 87 %, and 88 % on training dataset; 85 %, 75 %, and 80 % on control dataset; and 91 %, 82 %, and 86 % on test dataset. ROC curves of test datasets for each model are presented in Fig. 5. The average area under the ROC curve on test dataset was 0.886. High accuracy of the model on test dataset highlights potential of urine analysis to diagnose cancer. Application of GC-MS analysis and ANN allows classifying urine samples of patients with cancer and healthy individuals with high accuracy, which has great promise in the future.

3.3. Limitations

The group of cancer patients does not include all cancer localizations and the number of patients with some of cancer localizations it low, which significantly limits applying the proposed approach in clinical practice. Additionally, distribution of patients with

Table 5
Frequency of occurrence and concentration range the most common VOCs in the urine samples of patients with cancer and healthy volunteers.

VOC	Healthy volunteers			Patients with cancer		
	Frequency of occurring, %	Concentration range, ng/mL	Median, ng/mL	Frequency of occurring, %	Concentration range, ng/mL	Median, ng/mL
Acetone	100	17–500	111	100	26–500	133
Dimethylsulfide	100	2–37	4	100	1–45	5
Dimethyl disulfide	100	210–2500	1266	100	235–2500	1222
Dimethyl trisulfide	100	520–5000	2233	100	560–5000	2146
Acetonitrile	66	BQLL - 175	12	83	BQL - 200	41
Benzene	50	BQL - 35	11	30	BQL - 85	15
Toluene	49	BQL - 104	BQL	51	BQL - 466	15
Hexane	39	BQL - 20	BQL	53	BQL - 88	11
2-Pentanone	28	BQL - 39	BQL	55	BQL - 15	10
Pentanal	51	BQL - 28	7	45	BQL - 10	BQL
2-Butanone	80	BQL - 95	8	62	BQL - 73	6

*NQ – not quantified, BQLL – below limit of quantification.

Table 6
Ratios involved in creation of diagnostic model.

Ratio	p-level
Benzene/acetone	0.019
Acetonitrile/acetone	0.004
2-Butanone/acetone	0.000
Acetonitrile/dimethylsulfide	0.000
Toluene/dimethylsulfide	0.013
Acetone/dimethylsulfide	0.005
Benzene/dimethyl disulfide	0.024
Acetonitrile/dimethyl disulfide	0.001
Dimethylsulfide/dimethyl disulfide	0.018
2-Butanone/dimethyl disulfide	0.004
2-Pentanone/dimethyl disulfide	0.001
Butanal/dimethyl disulfide	0.009
Acetonitrile/dimethyl trisulfide	0.000
2-Pentanone/dimethyl trisulfide	0.000

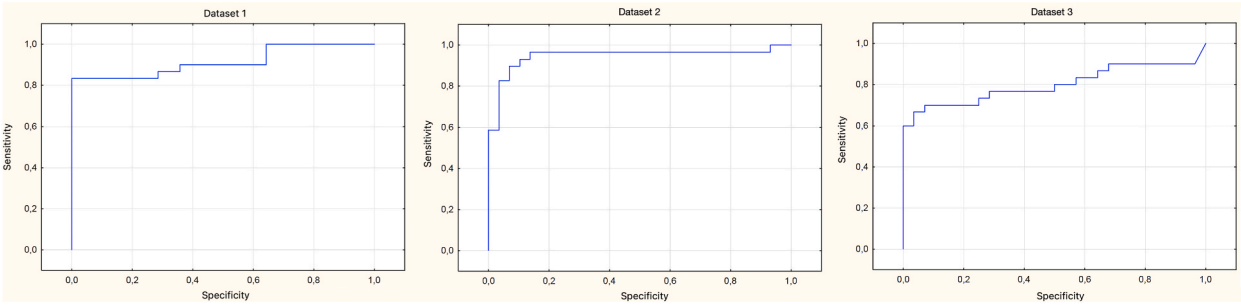


Fig. 5. ROC curves of 3 test datasets (3-fold cross-validation).

different cancer localizations varied, which does not allow finding out features specific for each cancer localization. Cohorts of patients with each cancer localization must be expanded to achieve reliable results. A lot of patients with various types of cancer localizations, stages, and histology should be involved in the study to create a test for diagnosing cancer or specific types of cancer from urine samples.

4. Conclusions

The paper presents the scheme of preconcentration of VOCs from urine with their subsequent determination using GC-MS. Concentration range for various analytes was 1–5000 ng/mL, which allows determining VOCs in urine samples of healthy volunteers and cancer patients. Concentrations of 2-butanone, 2-pentanone, acetonitrile, and benzene were found different in urine of patients with

cancer and healthy individuals, which can form the basis of the diagnostic method. The prospectiveness of using VOC ratios is demonstrated in the study, which allows creating a diagnostic model capable to distinguish patients with cancer and healthy subjects with high accuracy using artificial neural networks. The findings show utility of urine analysis for cancer diagnostics.

CRediT authorship contribution statement

Elina Gashimova: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation. **Azamat Temerdashev:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis. **Dmitry Perunov:** Writing – review & editing, Visualization, Validation, Investigation. **Vladimir Porkhanov:** Writing – review & editing, Conceptualization. **Igor Polyakov:** Writing – review & editing, Project administration, Conceptualization. **Alexey Podzhivotov:** Writing – review & editing, Validation, Investigation, Data curation. **Ekaterina Dmitrieva:** Writing – review & editing, Visualization, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Russian Science Foundation and Kuban Science Foundation (project no. 22-13-20018) and carried out using the scientific equipment of the Center for Environmental Analysis at the Kuban State University (unique identifier RFMEFI59317X0008).

References

[1] H. Sohrabi, et al., State-of-the-art cancer biomarker detection by portable (Bio) sensing technology: a critical review, Microchem. J. 177 (2022) 107248, https://doi.org/10.1016/j.microc.2022.107248.
[2] C.V. Berenguer, F. Pereira, J.A.M. Pereira, J.S. Camara, Volatilomics: an emerging and promising avenue for the detection of potential prostate cancer biomarkers, Cancers 14 (2022) 3982, https://doi.org/10.3390/cancers14163982.
[3] S. Sethi, R. Nanda, T. Chakraborty, Clinical application of volatile organic compound analysis for detecting infectious diseases, Clin. Microbiol. Rev. 26 (2013) 462–475, https://doi.org/10.1128/CMR.00020-13.
[4] H. P, M. Rangarajan, H.J. Pandya, Breath VOC analysis and machine learning approaches for disease screening: a review, J. Breath Res. 17 (2023) 024001, https://doi.org/10.1088/1752-7163/acb283.
[5] G. Riccio, S. Baroni, A. Urbani, V. Greco, Mapping of urinary volatile organic compounds by a rapid analytical method using gas chromatography coupled to ion mobility spectrometry (GC–IMS), Metabolites 12 (2022) 1072, https://doi.org/10.3390/metabo12111072.
[6] V.K. Pal, K. Kannan, Stability of volatile organic compound metabolites in urine at various storage temperatures and freeze-thaw cycles for 8 months, Environ Pollut 345 (2024) 123493, https://doi.org/10.1016/j.envpol.2024.123493.
[7] A.C. Dima, D.V. Balaban, A. Dima, Diagnostic application of volatile organic compounds as potential biomarkers for detecting digestive neoplasia: a systematic review, Diagnostics 11 (2021) 2317, https://doi.org/10.3390/diagnostics11122317.
[8] B.D. Hosfield, A.R. Pecoraro, N.T. Baxter, T.B. Hawkins, T.A. Markel, The assessment of fecal volatile organic compounds in healthy infants: electronic nose device predicts patient demographics and microbial enterotype, J. Surg. Res. 254 (2020) 340–347, https://doi.org/10.1016/j.jss.2020.05.010.
[9] S. el Manouni el Hassani, D. Berkhout, S. Bosch, M. Benninga, N. de Boer, T. de Meij, Application of fecal volatile organic compound analysis in clinical practice: current state and future perspectives, Chemosensors 6 (2018) 29, https://doi.org/10.3390/chemosensors6030029.
[10] W. Filipiak, et al., Dependence of exhaled breath composition on exogenous factors, smoking habits and exposure to air pollutants, J. Breath Res. 6 (3) (2012) 036008, https://doi.org/10.1088/1752-7155/6/3/036008.
[11] N. Drabińska, E. Jarocka-Cyrta, N.M. Ratcliffe, U. Krupa-Kozak, The profile of urinary headspace volatile organic compounds after 12-week intake of oligofructose-enriched inulin by children and adolescents with celiac disease on a gluten-free diet: results of a pilot, randomized, placebo-controlled clinical trial, Molecules 24 (2019) 1341, https://doi.org/10.3390/molecules24071341.
[12] H. Tyagi, E. Daulton, A.S. Bannaga, R.P. Arasaradnam, J.A. Covington, Non-invasive detection and staging of colorectal cancer using a portable electronic nose, Sensors 21 (2021) 5440, https://doi.org/10.3390/s21165440.
[13] J. Giro Benet, M. Seo, M. Khine, J. Guma Padro, A. Pardo Martnez, F. Kurdahi, Breast cancer detection by analyzing the volatile organic compound (VOC) signature in human urine, Sci. Rep. 12 (2022) 14873, https://doi.org/10.1038/s41598-022-17795-8.
[14] L. Capelli, C. Bax, F. Grizzi, G. Taverna, Optimization of training and measurement protocol for eNose analysis of urine headspace aimed at prostate cancer diagnosis, Sci. Rep. 11 (2021) 20898, https://doi.org/10.1038/s41598-021-00033-y.
[15] A.D. Asimakopoulos, et al., Prostate cancer diagnosis through electronic nose in the urine headspace setting: a pilot study, Prostate Cancer Prostatic Dis. 17 (2014) 206–211, https://doi.org/10.1038/pcan.2014.11.
[16] A. Filianoti, et al., Volatilome analysis in prostate cancer by electronic nose: a pilot monocentric study, Cancers 14 (2022) 2927, https://doi.org/10.3390/cancers14122927.
[17] X. Zou, et al., Detection of volatile organic compounds in a drop of urine by ultrasonic nebulization extraction proton transfer reaction mass spectrometry, Anal. Chem. 90 (2018) 2210–2215, https://doi.org/10.1021/acs.analchem.7b04563.
[18] Q. Liu, et al., Volatile organic compounds for early detection of prostate cancer from urine, Heliyon 9 (2023) e16686, https://doi.org/10.1016/j.heliyon.2023.e16686.
[19] E. Mozdiak, A.N. Wicaksono, J.A. Covington, R.P. Arasaradnam, Colorectal cancer and adenoma screening using urinary volatile organic compound (VOC) detection: early results from a single-centre bowel screening population (UK BCSP), Tech. Coloproctol. 23 (2019) 343–351, https://doi.org/10.1007/s10151-019-01963-6.
[20] C.E. Boulind, et al., Urinary volatile organic compound testing in fast-track patients with suspected colorectal cancer, Cancers 14 (2022) 2127, https://doi.org/10.3390/cancers14092127.
[21] Y. Hanai, et al., Urinary volatile compounds as biomarkers for lung cancer, Biosci. Biotechnol. Biochem. 76 (2012) 679–684, https://doi.org/10.1271/bbb.110760.

[22] T. Ligor, P. Adamczyk, T. Kowalkowski, I.A. Ratiu, A. Wenda-Piesik, B. Buszewski, Analysis of VOCs in urine samples directed towards of bladder cancer detection, Molecules 27 (2022) 5023, https://doi.org/10.3390/molecules27155023.
[23] C.L. Silva, M. Passos, J.S. Cmara, Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry, Br. J. Cancer 105 (2011) 1894–1904, https://doi.org/10.1038/bjc.2011.437.
[24] T. Zivković Semren, I. Brcic Karaconji, T. Safner, N. Brajenovic, B. Tariba Lovakovic, A. Pizent, Gas chromatographic-mass spectrometric analysis of urinary volatile organic metabolites: optimization of the HS-SPME procedure and sample storage conditions, Talanta 176 (2018) 537–543, https://doi.org/10.1016/j.talanta.2017.08.064.
[25] P. Porto-Figueira, J. Pereira, W. Miekisch, J.S. Câmara, Exploring the potential of NTME/GC-MS, in the establishment of urinary volatomic profiles. Lung cancer patients as case study, Sci. Rep. 8 (2018) 13113, https://doi.org/10.1038/s41598-018-31380-y.
[26] H. Tyagi, E. Daulton, A.S. Bannaga, R.P. Arasaradnam, J.A. Covington, Urinary volatiles and chemical characterisation for the non-invasive detection of prostate and bladder cancers, Biosensors 11 (2021) 437, https://doi.org/10.3390/bios11110437.
[27] T. Ligor, et al., Searching for potential markers of glomerulopathy in urine by HS-SPME-GC×GC TOFMS, Molecules 26 (2021) 1817, https://doi.org/10.3390/molecules26071817.
[28] P. Aggarwal, J. Baker, M.T. Boyd, S. Coyle, C. Probert, E.A. Chapman, Optimisation of urine sample preparation for headspace-solid phase microextraction gas chromatography-mass spectrometry: altering sample pH, sulphuric acid concentration and phase ratio, Metabolites 10 (2020) 1–17, https://doi.org/10.3390/metabo10120482.
[29] V. Longo, et al., Blood, urine and semen Volatile Organic Compound (VOC) pattern analysis for assessing health environmental impact in highly polluted areas in Italy, Environmental Pollution 286 (2021) 117410, https://doi.org/10.1016/j.envpol.2021.117410.
[30] K. Taunk, et al., Urinary volatomic expression pattern: paving the way for identification of potential candidate biosignatures for lung cancer, Metabolites 12 (2022) 36, https://doi.org/10.3390/metabo12010036.
[31] N. Sarigul, F. Korkmaz, İ. Kurultak, A new artificial urine protocol to better imitate human urine, Sci. Rep. 9 (2019) 20159, https://doi.org/10.1038/s41598-019-56693-4.
[32] E.M. Gashimova, et al., Evaluation of the possibility of volatile organic compounds determination in exhaled air by gas chromatography for the noninvasive diagnostics of lung cancer, J. Anal. Chem. 74 (2019) 472–479, https://doi.org/10.1134/S1061934819050034.
[33] E. Gashimova, A. Temerdashev, V. Porkhanov, I. Polyakov, D. Perunov, E. Dmitrieva, Non-invasive exhaled breath and skin analysis to diagnose lung cancer: study of age effect on diagnostic accuracy, ACS Omega 7 (2022) 42613–42628, https://doi.org/10.1021/acsomega.2c06132.
[34] R.B. Roy, et al., A comparative performance analysis of ANN algorithms for MPPT energy harvesting in solar PV system, IEEE Access 9 (2021) 102137–102152, https://doi.org/10.1109/ACCESS.2021.3096864.
[35] M.B. Reed, L-Broyden methods: a generalization of the L-BFGS method to the limited-memory Broyden family, Int. J. Comput. Math. 86 (2006) 606–615, https://doi.org/10.1080/00207160701656749.
[36] J. Bourquin, H. Schmidli, P. Hoogevest, H. Leuenberger, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research, Pharm. Dev. Technol 2 (1997) 95–109, https://doi.org/10.3109/10837459709022615.
[37] F.A. Franchina, D. Zanella, T. Dejong, J.-F. Focant, Impact of the adsorbent material on volatile metabolites during in vitro and in vivo bio-sampling, Talanta 222 (2021) 121569, https://doi.org/10.1016/j.talanta.2020.121569.
[38] G.E.P. Box, J.S. Hunter, The 2^{k-p} fractional factorial designs Part I, Technometrics 42 (2000) 28–47, https://doi.org/10.1080/00401706.2000.10485977.
[39] N. Drabińska, H.A. Azeem, U. Krupa-Kozak, A targeted metabolomic protocol for quantitative analysis of volatile organic compounds in urine of children with celiac disease, RSC Adv. 8 (2018) 36534–36541, https://doi.org/10.1039/c8ra07342b.
[40] R.B. Jain, Distributions of selected urinary metabolites of volatile organic compounds by age, gender, race/ethnicity, and smoking status in a representative sample of U.S. adults, Environ. Toxicol. Pharmacol. 40 (2015) 471–479, https://doi.org/10.1016/j.etap.2015.07.018.
[41] M. Monteiro, et al., GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients, J. Cell Mol. Med. 21 (2017) 2092–2105, https://doi.org/10.1111/jcmm.13132.