

DOI: 10.18721/JPM.14311
УДК 543.51+ 681.2–5

АЛГОРИТМ ОБРАБОТКИ ДАННЫХ МАСС-СПЕКТРОМЕТРИЧЕСКОГО АНАЛИЗА ДЛЯ ПЕРВИЧНОЙ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ПО ВЫДЫХАЕМЫМ ГАЗАМ

**В.В. Манойлов, Л.В. Новиков, А.Г. Кузьмин,
Ю.А. Титов, И.В. Заруцкий**

Институт аналитического приборостроения Российской академии наук,
Санкт-Петербург, Российская Федерация

В работе предложен алгоритм обработки масс-спектров газов, выдыхаемых пациентами. Масс-спектры регистрируются на квадрупольном масс-спектрометре MS7-200 с электронной ионизацией и прямым капиллярным вводом пробы. Алгоритм основан на преобразовании массива спектров (не менее десяти) в пространство главных компонент. Вероятность заболевания определяется по евклидову расстоянию координат пациента от центроида. Тестирование алгоритма проведено на данных масс-спектров газов, выдыхаемых больными с онкологическими заболеваниями. Предлагаемая процедура имеет ряд преимуществ перед традиционными лабораторными методами; алгоритм использует многомерную плотность вероятности распределения параметров выдыхаемых газов контрольных групп и тестируемого пациента; позволяет за короткое время составить общую картину вероятных заболеваний.

Ключевые слова: диагностика, метод главных компонент, многомерная плотность вероятности, обработка многомерных данных

Ссылка при цитировании: Манойлов В.В., Новиков Л.В., Кузьмин А.Г., Титов Ю.А., Заруцкий И.В. Алгоритм обработки данных масс-спектрометрического анализа для первичной диагностики заболеваний по выдыхаемым газам // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2021. Т. 14. № 3. С. 146–157. DOI: 10.18721/JPM.14311

Статья открытого доступа, распространяемая по лицензии CC BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>)

AN ALGORITHM FOR PROCESSING MASS SPECTROMETRIC ANALYSIS DATA FOR INITIAL DIAGNOSTICS OF DISEASES BY EXHALED GASES

**V.V. Manoilov, L.V. Novikov, A.G. Kuzmin,
Yu.A. Titov, I.V. Zarutskii**

Institute for Analytical Instrumentation of the Russian Academy of Sciences,
St. Petersburg, Russian Federation

An algorithm for processing mass spectra of gases exhaled by patients has been proposed in the paper. The mass spectra are recorded on the MS7-200 quadrupole mass spectrometer, with electronic ionization and with direct capillary injection of the sample. The algorithm is based on transforming an array of spectra (not less than 10) in the space of principal components. The probability of a disease is determined through the Euclidean distance of the patient's coordinates from the centroid. Testing of the algorithm was carried out on the data of mass spectra of gases exhaled by cancer patients. The proposed procedure has several advantages over traditional laboratory methods. The algorithm uses the multidimensional probability density of the distribution of the parameters of the exhaled gases of control groups and the patient being tested and allows to compile an overall picture of the patient's probable diseases in a short time.



Keywords: diagnostics, principal component analysis, multivariate probability density, multivariate data processing

Citation: Manoilov V.V., Novikov L.V., Kuzmin A.G., Titov Yu.A., Zarutskii I.V., An algorithm for processing mass spectrometric analysis data for initial diagnostics of diseases by exhaled gases, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 14 (3) (2021) 146–157. DOI: 10.18721/JPM.14311

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>)

Введение

В последнее время для диагностики заболеваний применяется анализ состава выдыхаемых газов [1, 2]. Он позволяет выявлять ряд серьезных недугов, среди которых болезнь Альцгеймера [3], эпилепсия [4], болезни почек [5], онкологические заболевания [6] и другие. Важной составляющей в развитии указанного способа диагностики является алгоритмическое обеспечение приборов, используемых для подобного анализа. Компьютерная программа, реализующая предлагаемый в работе алгоритм, выполняет обработку масс-спектров выдыхаемых газов; состав газов измеряют с помощью прибора с квадрупольным анализатором. Для проверки алгоритма были проведены измерения выдыхаемых газов у пациентов с онкологическими заболеваниями.

При классическом подходе к анализу масс-спектров, изначально производится так называемая первичная, или предварительная обработка масс-спектров; при этом выполняется обнаружение масс-спектрометрических пиков и оценка их параметров: позиция на шкале массовых чисел и амплитуда (иногда площадь). В ряде случаев первичной обработки масс-спектров выполняется математическое улучшение разрешающей способности прибора путем разделения наложившихся спектральных линий [7].

За первичной следует вторичная обработка данных. Описываемый ниже алгоритм обработки масс-спектров реализуется именно на этом этапе в офлайн-режиме, по информации, записанной в памяти компьютера в виде линейчатых спектров. В результате работы программы, реализующей данный алгоритм, выполняется предварительная оценка того, насколько вероятно принадлежность испытуемых лиц к классу больных с определенной патологией либо к классу практически здоровых людей.

Таким образом, данный алгоритм дополняет возможности классификации методами дискриминантного и кластерного анализов по многомерным статистическим данным. Классификация масс-спектров выдыхаемых газов методами дискриминантного и кластерного анализов описана в работе [8].

Главным отличием масс-спектров выдыхаемых газов у пациентов с какими-либо патологиями от таковых у практически здоровых людей является, прежде всего, присутствие дополнительных компонент, например, с массами 71, 64, 59 и 55 Да. Присутствие указанных линий в масс-спектре выдыхаемых газов у обследуемого лица указывает на высокую вероятность патологии в состоянии его организма. По аналогии с отпечатками пальцев, для каждого масс-спектра выдыхаемого воздуха характерен свой неповторимый профиль (набор пиков разной интенсивности), поэтому можно предположить, что различные профили «отпечатков» выдоха могут служить индикаторами определенной патологии. После набора достаточного количества таких данных появится возможность определять с весьма высокой эффективностью конкретный тип заболевания и либо назначать дополнительные обследования, либо сразу начинать лечение пациента.

Основные предпосылки

Массив данных масс-спектров под названием «здоровые люди» образует матрицу. По всем столбцам матрицы, кроме первого, записываются интенсивности пиков их масс-спектра на заданных значениях массы. Первый столбец матрицы содержит номер масс-спектра выдыхаемых газов определенного человека, а каждая ее строка – информацию об интенсивности отдельных линий масс-спектра.

Обозначим такую матрицу как $\mathbf{XS} = [x_{ij}]$, где i, j – номера объектов (людей) и признаков (массовых чисел) соответственно, $i = 1, 2, \dots, I, j = 1, 2, \dots, J$; x_{ij} – интенсивность j -й спектральной компоненты i -го здорового человека.

Аналогично образуем матрицу $I_0 \times J$ для масс-спектров выдыхаемых газов у больных. Каждая строка такой матрицы содержит интенсивности отдельных линий соответствующего масс-спектра. Обозначим эту матрицу $\mathbf{XI}_1 = [x_{ij}^0]$, где i, j, J имеют тот же смысл, но $i = 1, 2, \dots, I_0$; x_{ij}^0 – интенсивность j -й спектральной компоненты i -го больного.

Объединим матрицы здоровых и больных испытуемых и получим единую матрицу данных \mathbf{X} размера $(I_0 + I) \times J$:

$$\mathbf{X} = ([\mathbf{XS}; \mathbf{XI}_1]).$$

Преобразуем данную матрицу по алгоритму построения пространства главных компонент (ГК) и вычислим матрицу счетов \mathbf{T} .

На основе клинических обследований большого количества больных и здоровых пациентов полученные масс-спектры разделяются на два множества: контрольные и тестируемые группы.

Контрольные группы масс-спектров – это данные для пациентов с известным диагнозом их заболевания и данные для здоровых людей. В то же время в процессе обработки множества тестируемых масс-спектров – иная ситуация: неизвестно, принадлежат ли они здоровым людям или пациентам с какой-либо патологией.

Число масс-спектров выдыхаемых газов у пациентов в каждой из двух групп должно быть не менее тридцати. Каждый полученный масс-спектр выдоха пациента из контрольной группы представляется в многомерном пространстве в виде точки. Такие точки всех пациентов группы, собранные вместе, образуют «облако» исходных данных. Чем ближе к центру (центроиде) облака находится точка, определяющая масс-спектр выдоха тестируемого пациента, тем выше вероятность того, что пациент страдает заболеванием, соответствующим контрольной группе. Если пациент здоров, то вероятность патологии его состояния оценивается по контрольной группе здоровых пациентов; чем дальше от центра (центроиды) облака находится точка, определяющая масс-спектр выдоха тестируемого пациента, тем выше вероятность патологии его состояния.

С математической точки зрения, масс-спектры выдыхаемых газов пациентов одной группы образуют пространство данных с некоторым известным заболеванием. Вероятность принадлежности тестируемого пациента к одной из этих групп проверяется последовательно по формуле многомерной плотности вероятности.

Результатом описанного обследования является полная вероятностная картина заболевания тестируемого пациента. Это позволяет оперативно разработать пути дальнейшего, более тщательного обследования и последующего лечения.

Теоретическое описание работы алгоритма

Пусть x_{ij} – j -й параметр, а в конкретном случае – величина интенсивности спектральной линии для данных i -го пациента одной из групп, причем $i = 1, 2, \dots, I$ и $j = 1, 2, \dots, J$, и рас-



сматривается набор из I пациентов J регистрируемых спектральных линий выдыхаемого воздуха. Образует из этих данных матрицу размером $I \times J$. Назовем такую матрицу обучающей и обозначим ее как \mathbf{X} . Столбцы этой матрицы обозначим как X_j : $\mathbf{X} = [X_1, X_2, \dots, X_j, \dots, X_J]$ [9].

Вектор X_j является случайной величиной. Предположим, что эта случайная величина имеет нормальное распределение с математическим ожиданием \bar{X}_j и дисперсией σ_j^2 . Элементы матрицы \mathbf{X} формируют в J -мерном пространстве «облако». Обозначим такое облако через G , состоящее из $I \times J$ точек, с центроидой в точке $\bar{\mathbf{X}}$. Ее координаты выражаются как

$$\bar{\mathbf{X}} = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_J].$$

Для пациента, состояние которого исследуется, характерна патология (или он здоров), если измеренный вектор параметров выдыхаемого им воздуха имеет вид

$$X_d = [X_{d,1}, X_{d,2}, \dots, X_{d,j}].$$

$X_{i,j}$ является параметром в J -мерном пространстве G , т. е. $X_d \in G$. Данное условие будет выполняться, если вероятность P отклонения точки X_d от центроиды $\bar{\mathbf{X}}$ не превышает некоторый порог α . Для того, чтобы вычислить данную вероятность, построим многомерное распределение вероятности принадлежности события пространству G , предполагая, что это распределение подчиняется нормальному закону [10]:

$$P(\mathbf{X}_d) = W * \exp \left\{ -\frac{1}{2} (\mathbf{X}_d - \bar{\mathbf{X}})^T \mathbf{K}_X^{-1} (\mathbf{X}_d - \bar{\mathbf{X}}) \right\}, \mathbf{X}_d \in G, \quad (1)$$

где \mathbf{K}_X – ковариационная матрица, $\mathbf{K}_X = E \left[(\mathbf{X} - \bar{\mathbf{X}}) * (\mathbf{X} - \bar{\mathbf{X}})^T \right]$ (E – символ математического ожидания); W – нормирующий множитель; $(\dots)^T$ – символ транспонирования матрицы.

Очевидно, что вероятность P этого события равна единице, если для очередного проверяемого пациента окажется, что $\mathbf{X}_d \equiv \bar{\mathbf{X}}$. Это условие выполняется, если в формуле (1) положить $W = 1$. Тогда условие, при котором тестируемый пациент будет принадлежать к контрольной группе, имеет вид $P(\mathbf{X}_d) < \alpha$, где величина α выбирается методом экспертной оценки.

Стоит отметить, что расчет требуемой вероятности P по формуле (1) связан со значительными вычислительными трудностями, вызванными наличием большого числа параметров J и корреляционных связей между столбцами матрицы \mathbf{X} . С целью сжатия данных и сокращения размерности пространства G , используется ортогональное преобразование данных в пространство главных компонент – метод главных компонент (МГК) [11].

Чтобы перейти в пространство ГК, необходимо сформировать новую матрицу, состоящую из всех строк матрицы \mathbf{X} и строки \mathbf{X}_d . Обозначим эту матрицу как \mathbf{XI} . В новой системе координат получим:

$$\mathbf{XI} = \mathbf{TP}^T + \mathbf{e} = \sum_{j=1}^A t_j \mathbf{p}_j^T + \mathbf{e}, \quad (2)$$

где \mathbf{p}_j^T – собственные функции ковариационной матрицы \mathbf{K}_X ; \mathbf{T} называется матрицей счетов и выражается как $\mathbf{T} = [T_1, T_2, \dots, T_A]$, ее размерность – $I \times A$; \mathbf{P}^T называется матрицей нагрузок, ее размерность – тоже $I \times A$; \mathbf{e} – матрица остатков (шумов), ее размерность – $(I \times J)$; векторы-столбцы T_j ($j = 1, 2, \dots, A$) называются главными компонентами (A – число главных компонент).

Величина A намного меньше числа переменных J . Данное обстоятельство означает, что почти вся основная информация о состоянии тестируемого пациента сосредоточена в нескольких первых ГК. Последняя строка матрицы (2), вектор T_d , – координаты параметров пациента в пространстве ГК:

$$\mathbf{T}_d = [t_{d,1}, t_{d,2}, \dots, t_{d,A}].$$

Средние значения столбцов матрицы \mathbf{T} равны нулю, а дисперсия – вектор σ^2 с элементами – равна $\sigma_j^2 = \lambda_j$, т. е. собственным числам ковариационной матрицы. Свойство разложения по ГК таково, что дисперсия быстро уменьшается уже к четвертой ГК, а столбцы матрицы \mathbf{T} не коррелированы, т. е.

$$\mathbf{T}_m \mathbf{T}_n' = \begin{cases} 0 & \text{при } n \neq m, \\ \lambda_j & \text{при } n = m. \end{cases}$$

Если принять во внимание это обстоятельство, то в новой системе координат формула (1) принимает вид:

$$P(\mathbf{T}_d) = \exp \left\{ -\frac{1}{2} \mathbf{T}_d \sigma^{-2} \mathbf{T}_d' \right\} = \exp \left\{ -\frac{1}{2} \sum_{j=1}^A \frac{t_{d,j}^2}{\sigma_j^2} \right\} \quad (3)$$

и евклидово расстояние от пациента с индексом d до центроиды контрольной группы равно

$$D(d) = \sqrt{\sum_{j=1}^A t_{d,j}^2}. \quad (4)$$

Основные шаги алгоритма

Рассматриваемый алгоритм можно разделить на два этапа: обучение и диагностика; они выполняются последовательно во времени.

На первом этапе (обучение), по результатам клинического обследования большого количества пациентов, формируется K матриц \mathbf{X}^k ($k = 1, 2, \dots, K$), где K – число подгрупп с информацией об интенсивности линий спектра выдыхаемых газов для каждого вида патологии. Следует отметить, что число строк I^k каждой матрицы должно быть больше числа столбцов J^k . Этап обучения выполняется один раз, по результатам накопления данных по спектрам газов, выдыхаемых пациентами (их обследуют в клинике).

На втором этапе алгоритма (диагностика) выполняется измерение спектральных компонент газов, выдыхаемых тестируемыми пациентами, и в несколько шагов выполняется вычисление вероятности их принадлежности к одной из контрольных групп.

Шаг 1. На основе данных по интенсивностям спектральных компонент на заданных массах образуется вектор-строка $\mathbf{X}_d = \{x_{d,1}, x_{d,2}, \dots, x_{d,J}\}$.

Шаг 2. Вычисляется матрица $\mathbf{X}\mathbf{I}^k = [\mathbf{X}^k; \mathbf{X}_d]$ размером $(I^k + 1 \times J^k)$; последняя строка в этой матрице – вектор X_d .

Шаг 3. Вычисляются нормированные значения этой матрицы. Коэффициентом нормировки является максимальное значение ее элементов.

Шаг 4. Рассчитывается матрица счетов \mathbf{T}^k и вектор собственных чисел λ_j^k по алгоритму метода ГК [12].

Шаг 5. Находятся главные компоненты и определяется число A (количество главных компонент) матрицы \mathbf{T}^k таким образом, чтобы $\sigma_A^k = \sqrt{\lambda_A^k} < \varepsilon$. Без потери достоверности можно положить $\varepsilon = (0,001, \dots, 0,01)$. Последняя строка матрицы $\sigma_A^k = \sqrt{\lambda_A^k} < \varepsilon$ – главные компоненты параметров выдыхаемых газов обследуемого пациента.

Шаг 6. Вычисляется вероятность $P(\mathbf{T}_d^k)$, $k = 1, 2, \dots, K$, по формуле (3).

Шаг 7. Переход к шагу 2 ($k = k + 1$), для того чтобы определить вероятность принадлежности обследуемого пациента к другой группе возможных патологий.

Шаг 8. Анализ результатов вычисления вероятности $P(\mathbf{T}_d^k)$, $k = 1, 2, \dots, K$, для всех K патологий, с целью определения патологии с наибольшей вероятностью.

Тестирование разработанного алгоритма

Выдыхаемые газы обследуемых пациентов вводились в квадрупольный масс-спектрометр, который имеет диапазон детектирования 1 – 200 Да, разрешение по массовым числам – 0,5 Да; скорость регистрации одного масс-спектра – до 15 с.

Были проведены измерения выдыхаемых воздуха у 43 пациентов с диагностированными онкологическими заболеваниями¹. Эти пациенты были разделены на контрольную группу (36 чел.) и группу тестируемых пациентов (7 чел.). При этом интенсивность компонентов выдыхаемого воздуха первого тестируемого пациента, например, совпадала с центроидой контрольной группы. Также для проверки эффективности алгоритма были привлечены еще 10 тестируемых пациентов² с другими видами патологии. Пациентам обеих групп тестируемых были присвоены порядковые номера: соответственно, $d = (1, 2, \dots, 7)$ и $d = (1, 2, \dots, 10)$.

Масс-спектрометр регистрировал масс-спектр воздуха, выдыхаемого каждым пациентом, в диапазоне от 50 до 100 Да. В результате обследования большого числа пациентов и последующего анализа полученных данных было установлено, что в выдыхаемых газах содержатся компоненты спектра на массах 53, ..., 69 Да (всего 12), которые являются маркерами большинства заболеваний. Из исходных данных, содержащих информацию о контрольной группе, была сформирована матрица, в которой содержатся 12 столбцов; по ее строкам были расположены интенсивности компонент спектра – маркеров.

На рис. 1 показан пример масс-спектра выдыхаемого газа одним из семи тестируемых пациентов. По горизонтальной оси откладывается масса однозарядных ионов в дальтонах (русское обозначение – Да, международное – Da). Поочередно выбирались данные по выдыхаемым газам тестируемых пациентов и выполнялись шаги 1 – 6 алгоритма. При этом оказалось, что в пространстве ГК достаточно положить $A = 6$, так как $\sigma_A^k \cong 0$. По формулам (4) и (3) были вычислены евклидово расстояние до центроиды контрольной группы и вероятность диагностики онкологического заболевания у каждого тестируемого пациента. Для всех тестируемых пациентов эти результаты представлены в таблице.

На рис. 2,*a* изображено пространство первых двух главных компонент матрицы (PC1 и PC2), где нанесены координаты пациентов контрольной группы 1 (см. таблицу) (о) и семи тестируемых (+) пациентов. Из данных таблицы для этой группы, состоящей из 7 чел., видно, что по мере удаления координат масс-спектра выдыхаемых газов от центра облака (центроиды) вероятность наличия онкологического заболевания у этих пациентов уменьшается, но остается большой.

На рис. 2,*b* в пространстве первых двух PC (PC1 и PC2) изображены координаты пациентов контрольной группы 2 и десяти тестируемых пациентов. Оказалось, что все пациенты

¹ Данные о выдыхаемых газах для этой группы пациентов предоставлены Национальным медицинским исследовательским центром имени В.А. Алмазова, Санкт-Петербург.

² Данные о выдыхаемых газах для этой группы предоставлены Санкт-Петербургским клиническим научно-практическим центром специализированных видов медицинской помощи (онкологическим).

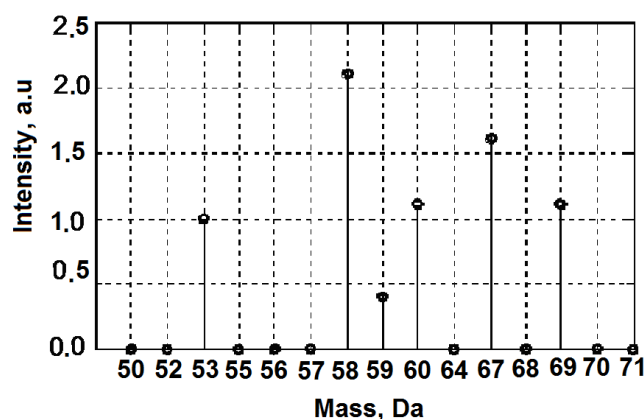


Рис. 1. Масс-спектр газов, выдыхаемых пациентом с патологией его состояния

Т а б л и ц а

**Результаты оценки вероятности онкологического заболевания
у пациентов двух контрольных групп**

Номер пациента	Евклидово расстояние	Вероятность заболевания
<i>Группа 1</i>		
1	0,0116	0,9854
2	0,1076	0,1888
3	0,1499	0,2055
4	0,2044	0,1164
5	0,2554	0,2864
6	0,2892	0,3458
7	0,3247	0,1197
<i>Группа 2</i>		
1	2,8791	0,0159
4	4,5739	0,0000
3	4,7337	0,0000
2	5,5927	0,0000
5	5,6321	0,0000
8	5,7129	0,0000
9	5,7322	0,0000
6	5,8520	0,0000
10	5,8635	0,0000
7	5,9168	0,0000

П р и м е ч а н и я. 1. Данные для пациентов *группы 1* и *2* были получены в разных медицинских учреждениях (см. соответственно сноски 1 и 2 на стр. 151).

2. Пациенты *группы 2* страдали онкологическим заболеванием, но другого типа, по сравнению с заболеваниями пациентов *группы 1*.

(кроме первого) находятся вне облака контрольной группы. Из данных таблицы для группы 2 видно, что у этих пациентов вероятность онкологического заболевания, но другого типа, по сравнению с онкологическим заболеванием пациентов группы 1, ничтожно мала.

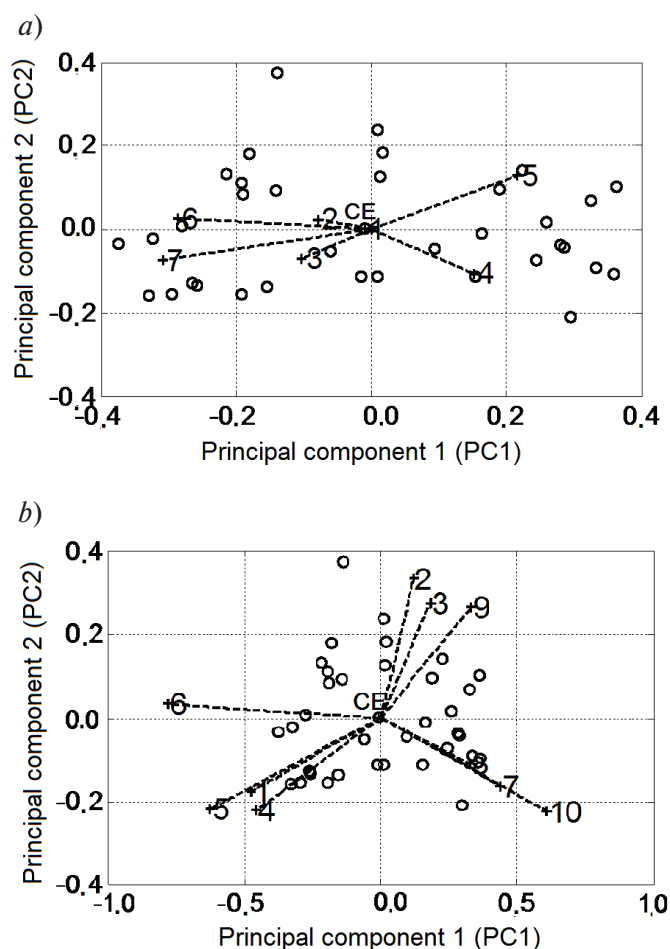


Рис. 2. Представление результатов обследований в пространстве главных компонент (PC1 и PC2) матрицы для контрольных групп 1 (a) и 2 (b) (см. таблицу).
 Центральные точки (CE) обозначают евклидов центр, остальные точки (o) – координаты (данные масс-спектров) пациентов соответствующих групп. Пунктирами показаны евклидовы расстояния до пациентов (+), которым присвоены номера

Заключение

Разработанный нами и представленный в настоящей статье алгоритм медицинской диагностики по многомерной плотности вероятности распределения параметров выдыхаемых газов контрольных групп и тестируемых пациентов, позволяет за короткое время выявить общую совокупность наиболее вероятных патологий пациента и при необходимости рекомендовать ему провести дальнейшие обследования. Предложенные алгоритм и метод могут быть также полезны для проведения скрининговых обследований пациентов, например, при проведении диспансеризации.

Эффективность успешного применения алгоритма зависит от объема и достоверности составления массива данных масс-спектров «здоровые» и массива данных «больные». В массив данных «здоровые» должны быть занесены данные зарегистрированных масс-спектров газов, выдыхаемых людьми, которые были предварительно обследованы различными медицинскими специалистами (с большой тщательностью). Эти специалисты должны подтвердить отсутствие патологий различного рода на момент проведения эксперимента по масс-спектрометрическому анализу выдыхаемых газов. Так например, при заполнении массива данных «здоровые» могут быть взяты масс-спектры выдыхаемых газов у лиц, находящихся на воен-

ной службе, или курсантов военных училищ и академий, которые были тщательно обследованы медицинскими специалистами и были признаны годными к службе или прохождению обучения. В последнее время авторами данной статьи были проведены эксперименты по масс-спектрометрическому анализу выдыхаемых газов у более чем 30 курсантов Военно-медицинской академии им. С.М. Кирова (г. Санкт-Петербург), которые были тщательно обследованы медицинскими специалистами и позволили сформировать массив данных «здоровые» и уточнить результаты тестирования алгоритма, представленные в таблице.

В массив данных «больные» должно быть занесено большое количество масс-спектров больных с точно установленным диагнозом, например онкологическим заболеванием определенного типа. Такой массив данных «больные» был составлен в процессе выполнения данной работы, но для повышения достоверности принятия решения о состоянии здоровья обследуемых пациентов такой массив данных следует увеличить хотя бы вдвое-втрое.

Для дальнейшего расширения возможностей описанного алгоритма желательно его применить для обработки данных масс-спектров выдыхаемых газов не только для приборов с квадрупольным анализатором, но и для приборов с другими типами анализаторов: статическим, времяпролетным и др. Для проведения исследований на таких приборах можно использовать систему ввода пробы в анализатор, аналогичную той, которая применялась в наших экспериментах.

Работа выполнена в рамках государственного задания 075-00780-19-02 (номер гос. регистрации АААА-А19-119053190069-2)

СПИСОК ЛИТЕРАТУРЫ

1. Li X., Huang D., Zeng J., Chan Ch.K., Zhou Z. Positive matrix factorization: A data preprocessing strategy for direct mass spectrometry-based breath analysis // *Talanta*. 2019. Vol. 192. 15 January. Pp. 32–39.
2. Li X., Huang L., Zhu H., Zhou Z. Direct human breath analysis by secondary nano electrospray ionization ultrahigh resolution mass spectrometry: Importance of high mass resolution and mass accuracy // *Rapid Communications in Mass Spectrometry*. 2017. Vol. 31. No. 3. Pp. 301–308.
3. Tiele A., Wicaksono A., Daulton E., Ifeachor E., Eyre V., Clarke S., Timings L., Pearson S., Covington J.A., Li X. Breath-based non-invasive diagnosis of Alzheimer's disease: A pilot study // *Journal of Breath Research*. 2020. Vol. 14. No. 2. P. 026003.
4. Van Dartel D., Schelhaas H.J., Colon A.J., Kho K.H., de Vos C.C. Breath analysis in detecting epilepsy // *Journal of Breath Research*. 2020. Vol. 14. No. 3. P. 031001.
5. Chen Ch.-Ch., Hsieh J.-Ch., Chao Ch.-H., Yang W.-Sh., Li X., Cheng H.-T., Chan Ch.-K., Lu Ch.-J., Meng H.-F., Zan H.-W. Correlation between breath ammonia and blood ureanitrogen levels in chronic kidney disease and dialysis patients // *Journal of Breath Research*. 2020. Vol. 14. No. 3. P. 036002.
6. Kononov A., Korotetsky B., Jahatspanian I., et al. Online breath analysis using metal oxide semiconductor sensors (electronic nose) for diagnosis of lung cancer // *Journal of Breath Research*. 2020. Vol. 14. No. 1. P. 016004.
7. Manoilov V.V., Novikov L.V., Belozertsev A.I., Zarutskii I.V., Titov Yu.A., Kuzmin A.G., El-Salim S.Z. Processing of mass spectra of exhaled gases based on correlation algorithms // *Journal of Analytical Chemistry*. 2020. Vol. 75. No. 13. Pp. 1678–1684.
8. Manoilov V.V., Novikov L.V., Zarutskii I.V., Kuzmin A.G., Titov Yu.A. Methods for processing mass spectrometry signals from exhaled gases for medical diagnosis // *Biomedical Engineering*. 2020. Vol. 53. No. 5. Pp. 355–359.



9. **Большаков А.А., Каримов Р.Н.** Методы обработки многомерных данных и временных рядов. М.: Горячая линия-Телеком, 2007. 522 с.
10. **Ширяев А.Н.** Вероятность. В 2 кн. 4-е изд., перераб. и доп. Т. 1. Вероятность—1. М.: МЦНМО, 2007. 552 с.
11. **Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.** Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
12. **Эсбенсен К.** Анализ многомерных данных: избранные главы. Пер. с англ. Под ред. О.Е. Родионовой. Черноголовка.: ИПХФ РАН, 2005. 161 с.

Статья поступила в редакцию 21.06.2021, принята к публикации 19.07.2021.

СВЕДЕНИЯ ОБ АВТОРАХ

МАНОЙЛОВ Владимир Владимирович — доктор технических наук, заведующий лабораторией автоматизации измерений и цифровой обработки сигналов Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26
manoilov_vv@mail.ru

НОВИКОВ Лев Васильевич — доктор физико-математических наук, ведущий научный сотрудник лаборатории оптики заряженных частиц и математического моделирования Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26
novik38@mail.ru

КУЗЬМИН Алексей Георгиевич — кандидат физико-математических наук, старший научный сотрудник лаборатории экологической масс-спектрометрии Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26
agqz55@rambler.ru

ТИТОВ Юрий Алексеевич — научный сотрудник лаборатории автоматизации измерений и цифровой обработки сигналов Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26
toplm@mail.ru

ЗАРУЦКИЙ Игорь Вячеславович — кандидат технических наук, старший научный сотрудник лаборатории автоматизации измерений и цифровой обработки сигналов Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26
igorzv@yandex.ru

REFERENCES

1. **Li X., Huang D., Zeng J., et al.**, Positive matrix factorization: A data preprocessing strategy for direct mass spectrometry-based breath analysis, Talanta. 192 (15 January) (2019) 32–39.

2. **Li X., Huang L., Zhu H., Zhou Z.**, Direct human breath analysis by secondary nano electrospray ionization ultrahigh resolution mass spectrometry: Importance of high mass resolution and mass accuracy, *Rapid Commun. Mass Spectrom.* 31 (3) (2017) 301–308.
3. **Tiele A., Wicaksono A., Daulton E., et al.**, Breath-based non-invasive diagnosis of Alzheimer's disease: A pilot study, *J. Breath Res.* 14 (2) (2020) 026003.
4. **Van Dartel D., Schelhaas H.J., Colon A.J., et al.**, Breath analysis in detecting epilepsy, *J. Breath Res.* 14 (3) (2020) 031001.
5. **Chen Ch.-Ch., Hsieh J.-Ch., Chao Ch.-H., et al.**, Correlation between breath ammonia and blood ureanitrogen levels in chronic kidney disease and dialysis patients, *J. Breath Res.* 14 (3) (2020) 036002.
6. **Kononov A., Korotetsky B., Jahatspanian I., et al.**, Online breath analysis using metal oxide semiconductor sensors (electronic nose) for diagnosis of lung cancer, *J. Breath Res.* 14 (1) (2020) 016004.
7. **Manoilov V.V., Novikov L.V., Belozertsev A.I., et al.**, Processing of mass spectra of exhaled gases based on correlation algorithms, *J. Analytical Chem.* 75 (13) (2020) 1678–1684.
8. **Manoilov V.V., Novikov L.V., Zarutskii I.V., et al.**, Methods for processing mass spectrometry signals from exhaled gases for medical diagnosis, *Biomed. Eng.* 53 (5) (2020) 355–359.
9. **Bolshakov A.A., Karimov R.N.**, *Metody obrabotki mnogomernykh dannykh i vremennykh ryadov* [Methods for processing multivariate data and time series], Goryachaya Liniya-Telekom Publishing, Moscow, 2007 (in Russian).
10. **Shiryayev A.N.**, *Probability*, Springer-Verlag, New-York, 1996.
11. **Ayvyazyan S.A., Bukhshtaber V.M., Enyukov I.S., Meshalkin L.D.**, *Prikladnaya statistika. Klassifikatsiya i snizheniye razmernosti* [Applied statistics. Classification and dimension reduction], Finance and Statistics Publishing, Moscow, 1989 (in Russian).
12. **Esbensen K.H.**, *Multivariate data analysis – in Practice*, CAMO, Oslo, 1994 – 2002.

Received 21.06.2021, accepted 19.07.2021.

THE AUTHORS

MANOILOV Vladimir V.

Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
manoilov_vv@mail.ru

NOVIKOV Lev V.

Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
novik38@mail.ru

KUZMIN Aleksey G.

Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
agqz55@rambler.ru

TITOV Yuriy A.

Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
topl@mail.ru

ZARUTSKII Igor V.

Institute for Analytical Instrumentation of the Russian Academy of Sciences

26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation

igorzv@yandex.ru