

## Применение машинного обучения для диагностики некоторых социально значимых заболеваний по выдыхаемому человеком воздуху методом инфракрасной лазерной спектроскопии

© Иг.С. Голяк<sup>1</sup>, П.В. Бережанский<sup>2</sup>, А.Ю. Седова<sup>2</sup>, Т.А. Гутырчик<sup>2</sup>, О.А. Небритова<sup>1</sup>, А.Н. Морозов<sup>1</sup>, Д.Р. Анфимов<sup>1</sup>, И.Б. Винтайкин<sup>1</sup>, А.А. Коноплева<sup>1</sup>, П.П. Дёмкин<sup>1</sup>, И.Л. Фуфурин<sup>1</sup>

<sup>1</sup>Московский государственный технический университет им. Н.Э. Баумана, 105005 Москва, Россия

<sup>2</sup>ГБУЗ Морозовская детская городская клиническая больница ДЗМ, 119049 Москва, Россия

e-mail: igorgolyak@yandex.ru

Поступила в редакцию 01.12.2022 г.

В окончательной редакции 25.01.2023 г.

Принята к публикации 30.01.2023 г.

Исследованы инфракрасные спектры воздуха, выдыхаемого несколькими группами волонтеров: страдающих диабетом первого типа, бронхиальной астмой и пневмонией. Для регистрации инфракрасных спектров применен перестраиваемый квантово-каскадный лазер, излучающий в диапазоне длин волн от 5.3 до 12.8  $\mu\text{m}$  в импульсном режиме с шириной импульса 50 ns, мощностью до 150 mW и шагом перестройки 1  $\text{cm}^{-1}$ . Лазер оптически сопряжен с астигматической газовой кюветой типа Эрриота с длиной оптического пути 76 m. Обнаружено отличие интенсивности селективных линий молекул-биомаркеров в спектрах выдыхаемого воздуха здоровых волонтеров от аналогичных показателей волонтеров, страдающих определенным заболеванием. На примере таких методов, как метод опорных векторов (SVM), метод  $k$ -ближайших соседей ( $k$ -NN) и алгоритм случайного леса (RandomForest), показана возможность классификации волонтеров по инфракрасным спектрам их выдыхаемого воздуха. Применение методов понижения размерности (PCA и  $t$ -SNE) позволило повысить точность классификации болезней до 98% по метрике accuracy.

**Ключевые слова:** инфракрасная спектроскопия, квантово-каскадный лазер, диагностика, выдыхаемый воздух, диабет первого типа, пневмония, хроническое заболевание, машинное обучение.

DOI: 10.21883/OS.2023.06.55917.109-23

### 1. Введение

Социально значимые заболевания [1–3], такие как сахарный диабет, бронхиальная астма, пневмония требуют ранней диагностики, что особенно важно для своевременной терапии. Неинвазивные методы исследования предпочтительны из-за малого воздействия на организм пациента, что особенно важно в тех случаях, когда диагностические тесты необходимо делать на протяжении всей жизни пациента. Современные методы диагностики выдыхаемого воздуха выявляют большое количество летучих органических соединений (ЛОС), выделяемых телом человека [4]. Изменения их концентраций часто сопряжены с некоторыми заболеваниями [5] или нарушениями метаболизма в целом. Перспективным является определение ЛОС в качестве маркеров, имеющих прогностическое значение для обнаружения развития метаболических нарушений, среди которых особое место занимает сахарный диабет. Применение такого рода предикторов при скрининге больших групп населения и формирование на этой основе профилактических мероприятий является важной как социальной, так и медико-биологической проблемой, особенно когда это касается здоровья детей. Одним из перспективных летучих соеди-

нений, сопряженных с особенностями метаболизма, является ацетон. Вариации его содержания в выдыхаемом воздухе или моче достаточно точно отражают изменения липидного обмена, в частности бета-окисление липидов.

Однако есть необходимость более четкой идентификации соответствующих газо-метаболических профилей у больных сахарным диабетом, отражающих системные перестройки метаболизма в условиях нормы и патологии, поскольку легкие углеводороды являются промежуточными или побочными продуктами многих метаболических циклов [6]. Так, например, образование ацетона обусловлено вовлечением жирных кислот в энергетический метаболизм [7] при сахарном диабете [8]. Также образование ацетона в выдыхаемом воздухе возможно при голодании [9], длительной интенсивной физической работе [10], изменении состояния энтеральной среды [11].

В последние годы аллергические заболевания респираторного тракта приобретают большую значимость из-за высокого роста распространенности среди населения, особенно в детской популяции [12]. По данным отчета Глобальной сети по борьбе с бронхиальной астмой (БА) (The Global Asthma Network, GAN), около 339.4 млн человек страдают данным заболеванием, приблизительно

14% из них — дети. По результатам центра по контролю и профилактике заболеваний (Centers for Disease Control and Prevention, CDC) в 2015 г. в США имелось 24.6 млн (7.8%) человек с установленным диагнозом бронхиальная астма. Из них 6.2 млн (8.4%) — дети: 4.7% пришлось на долю пациентов в возрасте до 4 лет, 9.8% — от 5 до 14 лет, 9.8% — от 15 до 17 лет. Известно, что больше половины случаев начала бронхиальной астмы приходится на ранний и дошкольный возраст, однако именно у данных пациентов диагностика представляет особую сложность [13]. Современная стратегия лечения бронхиальной астмы направлена на предупреждение развития патологии и купирование обострений. Разработка новых методов ранней диагностики и изучение механизмов развития заболевания позволит повысить качество жизни детей и их родителей. Известны методы ранней диагностики бронхиальной астмы у детей раннего возраста путем определения монооксида углерода (СО) в выдыхаемом воздухе. У детей раннего возраста, страдающих астмой, пороговое значение СО составляет более 2 ppm, у здоровых детей значение СО не превышает 1 ppm. Данный метод может быть использован как дополнительный в качестве неинвазивного метода контроля астмы и диагностики астмы у детей раннего возраста [14]. Динамический мониторинг FeNO является эффективным индикатором воспаления дыхательных путей и, таким образом, представляет собой важный клинический инструмент для оценки течения астмы [15], концентрация которого снижается в ответ на терапию глюкокортикостероидами [16]. Показана статистическая значимость различия концентраций пероксида водорода в конденсате выдыхаемого воздуха между пациентами с астмой и здоровыми пациентами (соответственно 480 и 780 nmol) [17].

Ежегодно в мире регистрируется около 150 миллионов случаев пневмонии у детей дошкольного возраста. Тяжелое течение пневмонии наблюдается в 7–13% случаев и обуславливает до 11–20 миллионов госпитализаций ежегодно. Показано [18], что основными веществами, обнаруживаемыми в выдыхаемом воздухе у больных пневмонией, были пентан и изопрен. Изопрен является основным углеводородом, присутствующим в альвеолярном газе, и образуется из изопентенилпирофосфата и диметилаллилпирофосфата — продуктов превращения мевалоната в холестерин. Производство мевалоната является этапом в биосинтезе тканевого холестерина. Изопрен является побочным продуктом синтеза стеролов, и любое изменение скорости синтеза холестерина вызовет сопоставимые изменения количества выделяемого изопрена на выдохе.

Таким образом, с точки зрения физиологии можно определить некоторые зависимости между профилем дыхания и состоянием здоровья человека. Однако, часто невозможно выделить одно конкретное химическое соединение и определить референсные значения, что обуславливает необходимость анализа состава выдыхаемого воздуха как единого целого. Инфракрасная

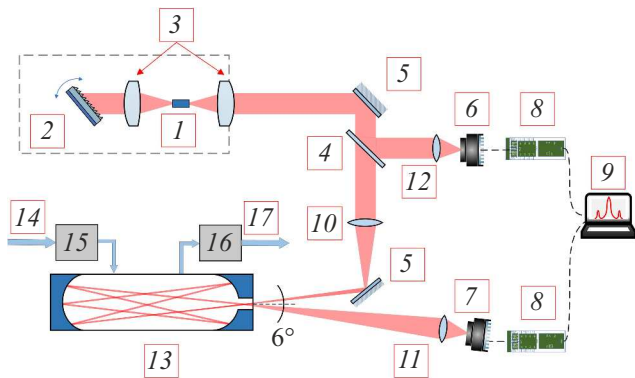
спектроскопия в среднем инфракрасном диапазоне (диапазон „отпечатков пальцев“) позволяет одновременно регистрировать характерные спектральные линии ряда химических соединений, что позволяет разработать метод диагностики ряда заболеваний по анализу выдыхаемого воздуха. Так, например, в работе [19] показаны возможности применения квантово-каскадного лазера, перестраиваемого в диапазоне  $1150\text{--}1250\text{ cm}^{-1}$  для анализа ацетона в выдыхаемом воздухе. В работе [20] показана возможность детектирования порядка 60 ppt NO на длине волны  $5.6\text{ }\mu\text{m}$  с помощью перестраиваемого инфракрасного лазера. Однако существенное перекрытие ряда селективных линий для компонентов выдыхаемого воздуха делает решение обратной задачи спектроскопии очень сложной.

Методы машинного обучения получили широкое применение в различных областях науки и техники, а также в медицине [21], что позволяет решать проблемы диагностики в различных областях, таких как визуализация медицинских данных, диагностика рака, системы на чипе и др. [22]. Машинное обучение применяется для изучения важных клинических параметров, таких как извлечение медицинской информации, прогнозирование заболеваний и стадий развития. Таким образом, это помогает в планировании и поддержании статуса пациента. Кроме того, он обеспечивает эффективный мониторинг состояния здоровья, помогая в анализе данных и отправке своевременных предупреждений [23]. Чтобы система машинного обучения была полезной при решении задач медицинской диагностики, необходимы следующие характеристики: высокая скорость выполнения, способность надлежащим образом обрабатывать недостающие данные и зашумленные данные, прозрачность диагностических знаний, способность объяснять решения и способность алгоритма сокращать количество тестов, необходимых для получения надежного диагноза. Алгоритмы машинного обучения активно применяются в приложениях по автоматизированной диагностике. Такие алгоритмы помогают медицинским работникам в прогнозировании заболеваний, а также для повышения точности диагностики заболеваний [24,25].

В работе [26] предложен метод диагностики острого инфаркта миокарда по выдыхаемому воздуху с помощью методов машинного обучения. Контрольная группа включала 30 человек и здоровая — 42. Кластеризация данных была проведена методом PCA на основе шести биомаркеров острого инфаркта миокарда ( $\text{C}_5\text{H}_{12}$ ,  $\text{N}_2\text{O}$ ,  $\text{NO}_2$ ,  $\text{C}_2\text{H}_4$ ,  $\text{CO}$ ,  $\text{CO}_2$ ) с дальнейшей классификацией методом опорных векторов. Достигнутая чувствительность и специфичность составили 0.82 и 0.93. В работе [27] проведено исследование о возможности применения искусственного интеллекта и методов машинного обучения при диагностике респираторных заболеваний, таких как астма и хроническая обструктивная болезнь легких. В работе [28] показано применение этих методов для анализа состояния пациентов, страдающих раком легких и хронической обструктивной болезнью легких,

**Таблица 1.** Поло-возрастное распределение (человек) здоровых волонтеров и волонтеров с установленным заболеванием

	Женский пол, младше 14 лет	Женский пол, старше 14 лет	Мужской пол, младше 14 лет	Мужской пол, старше 14 лет	Сумма
Диабет 1-го типа	10	12	18	20	60
Бронхиальная астма	6	2	13	11	32
Внебольничная пневмония	4	0	1	0	5
Здоровые	14	9	15	22	60



**Рис. 1.** Принципиальная схема лабораторной установки.

по данным инфракрасной фотоакустической спектроскопии. В работах [29,30] методы машинного обучения применялись для анализа многокомпонентных газовых смесей. В работах [31,32] показаны возможности применения машинного и глубокого обучения для анализа выдыхаемого человеком воздуха.

### Экспериментальная установка

Принципиальная схема лабораторной установки, которая применялась в рамках данного исследования, приведена на рис. 1.

Установка состоит из квантово-каскадного лазера (1), излучающего в диапазоне длин волн 5.8–12.3 мкм в импульсном режиме с длиной импульса 50 ns и мощностью до 150 mW. Лазер построен по схеме Литтрова: (2) — дифракционная решетка, (3) — асферические линзы. Инфракрасное излучение через систему зеркал (5) попадает на светоделитель (4) и часть пучка через согласующую оптику (12) попадает на опорный фотоприемник (6). Далее излучение направляется на линзу (10) с фокусным расстоянием  $f \sim 350$  mm, расположенную так, чтобы сфокусировать луч внутри объема газовой кюветы на расстоянии  $1/3-1/2$  ее длины. Газовая кювета (13) типа Эрриота позволяет получить 238 отражений лазерного луча и оптический

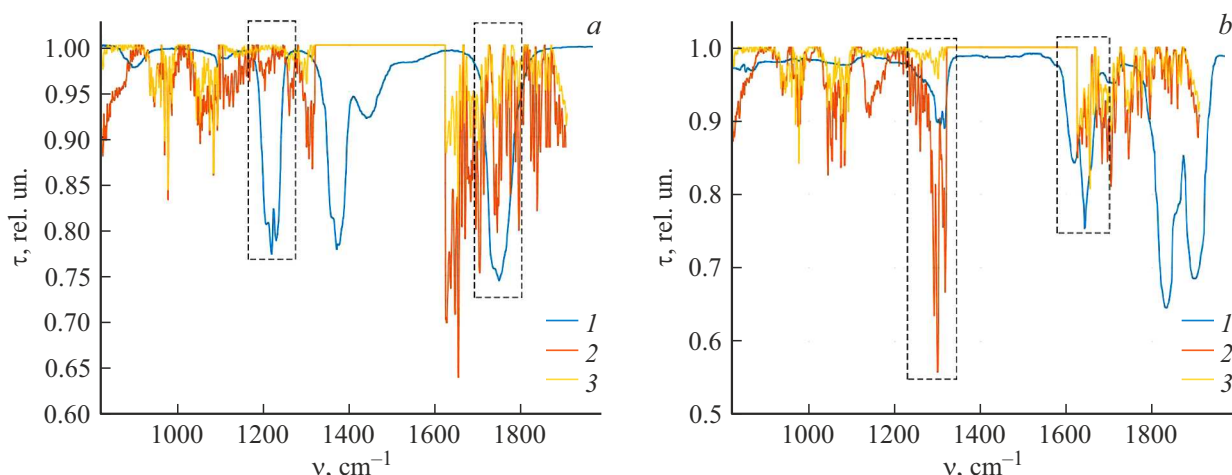
путь 76 m. Выходящий луч фокусируется с помощью согласующей оптики (11) на сигнальном фотоприемнике (7). В качестве фотоприемников (6) и (7) использовались кадмий–ртуть–теллурические фотоприемники, охлаждаемые каскадом ячеек Пельтье, обнаружительная способность  $D^*$  которых составляла значение  $\sim (6-8) \cdot 10^9$  cm $\sqrt{Hz/W}$ . Сигнал с обоих фотоприемников считывается 18-разрядным аналого-цифровым преобразователем (8), после чего референтный и сигнальный спектры передаются на компьютер (9). Анализируемая проба (14) через осушитель (15) попадает в многоходовую кювету (13). Концентрация водяного пара после осушки составляет значение порядка 1 g/m $^3$ . На выходе кюветы установлен вакуумный насос (16), который поддерживает заданное давление и откачивает пробу (17).

Далее на рис. 2 представлены характерные спектры выдыхаемого воздуха здорового волонтера, пациента с установленным диагнозом диабет 1 типа, и пациента с диагнозом бронхиальная астма, а также спектры ацетона и монооксид азота как характерных биомаркеров диабета и астмы соответственно.

На рис. 2 выделены области спектра, для которых наиболее выражены различия, соответствующие поглощению молекул-биомаркеров. Однако необходимо отметить, что спектр выдыхаемого воздуха содержит несколько сотен компонент, спектры ряда которых существенно перекрываются, поэтому невозможно утверждать о влиянии на характерную линию поглощения только одного вещества. Таким образом, для анализа используется достаточно широкая область инфракрасного спектра.

### Исследуемые группы

Для проведения клинических исследований были отобраны четыре группы испытуемых, что в общей сумме составляет 157 человек. В первую группу входят здоровые волонтеры в количестве 60 человек (табл. 1). Во вторую группу включены пациенты, страдающие внебольничной пневмонией (5 человек). Третья группа — пациенты с подтвержденным диагнозом диабета первого



**Рис. 2.** (а) Спектр ацетона [33] (1), спектры выдыхаемого воздуха пациента, страдающего диабетом первого типа (2), и здорового волонтера (3); (б) спектр NO [33] (1), спектры выдыхаемого воздуха пациента, страдающего бронхиальной астмой (2), и здорового волонтера (3).

типа (60 человек) и пациенты с бронхиальной астмой (32 человека) четвертая.

Исследование проводилось на базе МГТУ им. Н.Э. Баумана и ГБУЗ Морозовская детская городская клиническая больница ДЗМ. Протокол исследования одобрен этическим комитетом Государственного бюджетного учреждения здравоохранения „Детская клиническая больница им. Морозова“, Департамента здравоохранения города Москвы (Москва, Россия), № 174 от 18 января 2022 года. Все процедуры, выполненные в исследовании с участием людей, соответствуют этическими стандартами институционального и/или национального комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики. Каждый участник исследования предоставил согласие на обработку персональных данных и заполнил анкету для сбора анамнеза.

## Результаты и обсуждения

В настоящей работе основную задачу представляла классификация волонтеров по инфракрасным спектрам их выдыхаемого воздуха на 4 группы: волонтеры без диагностированных отклонений (здоровые), пациенты с диабетом первого типа, пациенты с диагностированной астмой и пациенты с внебольничной пневмонией. Во время эксперимента от каждого пациента было получено 100 спектров, которые впоследствии были усреднены. Перед классификацией выполнена предобработка данных. Для полученных спектров пропускания проводилась коррекция базовой линии, после чего спектры были нормализованы таким образом, чтобы каждый спектр имел среднее значение равное 0 и стандартное отклонение, равное 1, что позволяет привести данные к единому масштабу.

В ходе исследований зарегистрировано 204 спектра (для некоторых пациентов пробы повторялись), из которых 71 — здоровые волонтеры, 77 — диабет первого типа, 32 — бронхиальная астма, 24 — внебольничная пневмония. Тестовая выборка составила 20% от представленного набора данных. Валидация проводилась по метрике точность (accuracy). Для визуализации данных были использованы метод главных компонент (PCA) и метод стохастического вложения соседей с  $t$ -распределением ( $t$ -SNE), результаты применения которых представлены на рис. 3.

На основе данных, представленных на рис. 3, показана принципиальная возможность классификации экспериментальных данных. На рис. 3 для метода PCA показано 2 размерности (для наглядности), однако объясненная дисперсия составляет значение 0.45, хотя визуально видно, что группы могут быть разделимы. Данными с десятью признаками покрывается 90% объяснённой дисперсии [34], что является достаточным значением для надежной классификации.

Классификация спектров пропускания была проведена с помощью трех методов машинного обучения: метод опорных векторов (SVM), метод  $k$ -ближайших соседей ( $k$ -NN), случайный лес (Random Forest). В табл. 2 приведены результаты классификации полноразмерных данных (без понижения размерности). В качестве метрик применены *precision*, *recall* и  $F1$ , которые могут быть описаны следующими формулами:

$$Precision = \frac{TP}{TP + FP},$$

$$Recall = \frac{TP}{TP + FN},$$

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall},$$

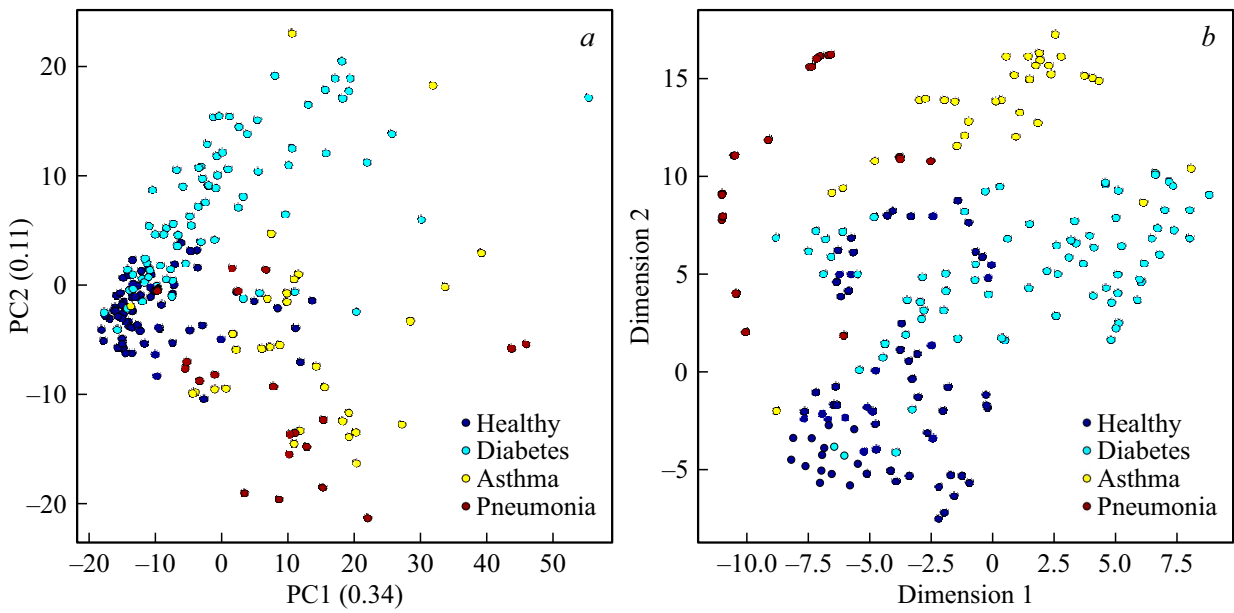


Рис. 3. Визуализация классификации данных с помощью методов (а) PCA и (b) t-SNE.

Таблица 2. Оценка эффективности классификации для полноразмерных данных

Метод	Группы	Precision	Recall	F1
SVM	Здоровые	0.87	0.93	0.90
	Диабет	0.93	0.88	0.90
	Астма	1.00	1.00	1.00
	Пневмония	1.00	1.00	1.00
k-NN	Здоровые	0.72	0.93	0.81
	Диабет	0.88	0.88	0.88
	Астма	1.00	1.00	1.00
	Пневмония	1.00	0.20	0.33
Random Forest	Здоровые	0.77	0.91	0.83
	Диабет	0.94	0.94	0.94
	Астма	1.00	0.88	0.93
	Пневмония	1.00	0.50	0.67

где  $TP$  — истинно положительный результат,  $FP$  — ложно положительный результат,  $TN$  — истинно отрицательный результат,  $FN$  — ложно отрицательный результат.

В табл. 3 приведены результаты классификации экспериментальной выборки на основе данных пониженной размерности (10 признаков).

Для определения оптимального метода для классификации, проведена оценка алгоритмов на тестовой выборке по метрике ассурасу. Результат приведен в табл. 4.

Таблица 3. Результаты классификации для данных пониженной размерности

Метод	Группы	Precision	Recall	F1
SVM	Здоровые	0.93	1.00	0.97
	Диабет первого типа	1.00	0.94	0.97
	Астма	1.00	1.00	1.00
	Пневмония	1.00	1.00	1.00
k-NN	Здоровые	0.93	0.93	0.93
	Диабет первого типа	0.94	0.94	0.94
	Астма	0.86	1.00	0.92
	Пневмония	1.00	0.80	0.89
Random Forest	Здоровые	0.92	0.86	0.89
	Диабет первого типа	0.88	0.94	0.91
	Астма	1.00	1.00	1.00
	Пневмония	1.00	1.00	1.00

Полученные результаты показывают, что модели машинного обучения для заданной выборки волонтеров и методики регистрации спектров позволяют классифицировать группы людей по их инфракрасным спектрам дыхания со значением метрик  $Precision$  и  $Recall$  не менее 0.8. Метод опорных векторов показал наилучший результат для классификации описанных в работе за-

**Таблица 4.** Результаты расчета точности для полученных классификаторов

Данные	SVM	k-NN	Random Forest
Полноразмерные данные	0.93	0.83	0.93
Данные пониженной размерности	0.98	0.93	0.93

болеваний по инфракрасным спектрам воздуха, выдыхаемого человеком. Понижение размерности повышает точность классификации за счет выделения наиболее значимых признаков, а также позволяет существенно повысить скорость вычислений. Количество значимых признаков было выбрано из условия достижения 90% объяснённой дисперсии.

## Заключение

Проведен анализ инфракрасных спектров выдыхаемого воздуха для четырех групп волонтеров: здоровых, страдающих диабетом первого типа, бронхиальной астмой и пневмонией. Инфракрасные спектры зарегистрированы при помощи перестраиваемого квантово-каскадного лазера, излучающего в диапазоне длин волн от 5.3 до 12.8  $\mu\text{m}$ . Излучение лазера осуществлялось в импульсном режиме, с шириной импульса 50 ns и мощностью до 150 mW, с шагом перестройки 1  $\text{cm}^{-1}$ . Лазер был оптически сопряжен с астигматической газовой кюветой типа Эрриота с длиной оптического пути 76 m. Для здоровых волонтеров и волонтеров, страдающих диабетом или астмой обнаружено различие в интенсивности спектров пропускания на длинах волн, соответствующих веществам-биомаркерам (ацетон и монооксид азота).

Для диагностики заболевания по инфракрасному спектру применялись следующие методы классификация: SVM, k-NN и Random Forest. При помощи методов PCA и t-SNE показано, что снижение размерности до 10 наиболее значимых признаков покрывает 90% объясненной дисперсии. Понижение размерности инфракрасных спектров позволяет повысить точность классификации. Метод опорных векторов позволил получить значение *accuracy* не менее 98% при классификации описанной выборки волонтеров по 4 классам состояния здоровья с учетом понижения размерности экспериментальных спектров выдыхаемого воздуха до 10 признаков. Описанный в работе метод анализа инфракрасных спектров выдыхаемого воздуха позволяет разработать приборы экспресс-диагностики состояния здоровья человека для дальнейшего применения в медицинской диагностике.

## Конфликт интересов

Авторы заявляют, что у них нет конфликта интересов.

## Финансирование работы

Работа выполнена в рамках реализации программы стратегического академического лидерства „Приоритет-2030“, утвержденной постановлением Правительства Российской Федерации от 13 мая 2021 г. № 729.

## Список литературы

- [1] T. Grant, E. Croce, E.C. Matsui. *Annals of Allergy, Asthma & Immunology*, **128** (1), 5(2022). DOI: 10.1016/j.ana.2021.10.002
- [2] E. Uphoff, B. Cabieses, M. Pinart, M. Vald?s, J.M. Antó, J. Wright. *Eur. Respir. J.*, **46** (2), 364(2015). DOI: 10.1183/09031936.00114514
- [3] И.И. Дедов, М.В. Шестакова, А.Ю. Майоров, О.К. Викулова, Г.Р. Галстян, Т.Л. Кураева, Е.А. Шестакова. Алгоритмы специализированной медицинской помощи больным сахарным диабетом. Под ред. И.И. Дедова, М.В. Шестаковой, А.Ю. Майорова. 9-й выпуск. Сахарный диабет, **22** (1S1), 1-144 (2019). DOI: 10.14341/DM221S1
- [4] B. de Lacy Costello, A. Amann, H. Al-Kateb, C. Flynn, W. Filipiak, T. Khalid, D. Osborne, N. M. Ratcliffe. *J. Breath Res.*, **8** (1), 014001 (2014). DOI: 10.1088/1752-7155/8/1/014001
- [5] A. Bajtarevic, C. Ager, M. Pienz, M. Klieber, K. Schwarz, M. Ligor, A. Amann. *BMC cancer.*, **9**(1), 1(2009).
- [6] R.H. Eckel, S.M. Grundy, P.Z. Zimmet. *The Lancet*, **365** (9468), 1415 (2005). DOI: 10.1016/S0140-6736(05)66378-7
- [7] P.R. Galassetti, B. Novak, D. Nemet, C. Rose-Gottron, D.M. Cooper, S. Meinardi, D.R. Blake. *Diabetes. Technol. Ther.*, **7**(1), 115(2005).
- [8] I. Ueta, Y. Saito, M. Hosoe, M. Okamoto, H. Ohkita, S. Shirai, H. Tamura, K. Jinno. *Chromatogr. B*, **877** (24), 2551 (2009). DOI: 10.1016/j.jchromb.2009.06.039
- [9] M. P. Kalapos. *Med. Hypotheses*, **53** (3), 236(1999).
- [10] A.K. Mork, G. Johanson. *Toxicol. Lett.*, **164** (1), 6(2006).
- [11] S.K. Kundu, J.A. Bruzek, R. Nair, A.M. Judilla. *Clinic. Chem.*, **39** (1), 87 (1993).
- [12] В.М. Ганузин, Н.Л. Черная, Г.С. Маскова. *Доктор. Ру*, **19** (3), 57 (2020). DOI:10.31550/1727-2378-2020-19-3-57-60
- [13] P.V. Berezanskij, T.A. Gutyrychik, Y.V. Vekshina, N.A. Gutyrychik, N.A. Shapiey, T.I. Ushina. *Med. Pharm. J. „Pulse“*, **24**(11), 101(2022). DOI: 10.26787/nydha-2686-6838-2022-24-11-101-107
- [14] Y. Ohara, T. Ohara, K. Hashimoto, M. Hosoya. *J. Med. Sci.*, **66**(2), 78 (2020). DOI: 10.5387/fms.2019-02
- [15] S.A. Kharitonov, D. Yates, R.A. Robbins, P.J. Barnes, R. Logan-Sinclair, E.A. Shinebourne. *The Lancet*, **343** (8890), 133(1994). DOI: 10.1016/S0140-6736(94)90931-8
- [16] P.E. Silkoff, P. McClean, M. Spino, L.A. Erlich, A.S. Slutsky, N. Zamel. *Chest*, **119** (5), 1322 (2001). DOI: 10.1378/chest.119.5.1322
- [17] S. Svensson, A. Olin, M. Lärstad, G. Ljungkvist, K. Torén. *J. Chromat. B*, **809**(2), 199 (2004). DOI: 10.1016/j.jchromb.2004.06.027

- [18] J.K. Schubert, W.P.E. Müller, A. Benzing, K. Geiger. *Intensive Care Med.*, **24**(5), 415 (1998).  
DOI: 10.1007/s001340050589
- [19] A. Reyes-Reyes, R.C. Horsten, H.P. Urbach, N. Bhattacharya. *Analytical Chem.*, **87**(1), 507 (2015).  
DOI: 10.1021/ac504235e
- [20] L. Richard, D. Romanini, I. Ventrillard. *Sensors*, **18**(7), 1997 (2018). DOI: 10.3390/s18071997
- [21] I. Kononenko. *Art. Intell. Med.*, **23**(1), 89 (2001).  
DOI: 10.1016/S0933-3657(01)00077-X
- [22] M. Shehab, L. Abualigah, Q. Shambour, M.A. Abu-Hashem, M.K.Y. Shambour, A.I. Alsalibi, A.H. Gandomi. *Comp. Biol. Med.*, **145**, 105458 (2022).  
DOI: 10.1016/j.combiomed.2022.105458
- [23] M. Gharaibeh, D. Alzu'bi, M. Abdullah, I. Hmeidi, M.R. Al Nasar, L. Abualigah, A.H. Gandomi. *Big Data Cogn. Comp.*, **6**(1), 29 (2022). DOI: 10.3390/bdcc6010029
- [24] M. Li, Z.H. Zhou. *IEEE Transactions on Systems, Man, and Cybernetics Part A*, **37** (6), 1088 (2007).  
DOI: 10.1109/TSMCA.2007.904745
- [25] A. Choudhury, D. Gupta. In *Recent developments in machine learning and data analytics* (Springer, Singapore, 2019). V. 67. DOI: 10.1007/978-981-13-1280-9\_6
- [26] A.V. Borisov, A.G. Syrkina, D.A. Kuzmin, V.V. Ryabov, A.A. Boyko, O. Zaharova, V.S. Zasedatel, Y.V. Kistenev, *J. Breath Res.*, **15**, 027104 (2021).  
DOI:10.1088/1752-7163/abebd4
- [27] A. Kaplan, H. Cao, J. M. FitzGerald, N. Iannotti, E. Yang, J.W.H. Kocks, K. Kostikas, D. Price, H.K. Reddel, I. Tsiligianni, C.F. Vogelmeier, P. Pfister, P. Mastoridis/ *J. Allergy Clinic. Immunol. Pract.*, **9**, 2255 (2021).  
DOI: 10.1016/j.jaip.2021.02.014
- [28] Y.V. Kistenev, A.V. Borisov, D.A. Kuzmin, O.V. Penkova, N. Kostyukova, A.A. Karapuzikov. *J. Biomed. Opt.*, **22** (1), 017002 (2017). DOI: 10.1117/1.jbo.22.1.017002
- [29] I.L. Fufurin, D.R. Anfimov, E.R. Kareva, A.V. Scherbakova, P.P. Demkin, A.N. Morozov, I.S. Golyak. *Opt. Engin.*, **60** (8), 082016 (2021). DOI: 10.1117/1.OE.60.8.082016
- [30] I.S. Golyak, E.R. Kareva, I.L. Fufurin, D. R. Anfimov, A.V. Scherbakova, A.O. Nebritova, P.P. Demkin, A.N. Morozov. *Comp. Opt.*, **46** (4), 650 (2022).  
DOI: 10.18287/2412-6179-CO-1058
- [31] I.L. Fufurin, P.V. Berezhanskiy, I.S. Golyak, D.R. Anfimov, E.R. Kareva, A.V. Scherbakova, P.P. Demkin, O.A. Nebritova, A. Morozov. *Materials*, **15** (9), 2984(2022).  
DOI: 10.3390/ma15092984
- [32] I.S. Golyak, I.L. Fufurin, E.R. Kareva, D.R. Anfimov, A.V. Scherbakova, A.N. Morozov, P.P. Demkin. In *Saratov Fall Meeting 2020: Optical and Nanotechnologies for Biology and Medicine*. Proc. SPIE. **11845**, 169 (2021)  
DOI: <https://doi.org/10.1117/12.2590835>
- [33] NIST Chemistry WebBook [Электронный ресурс].  
URL: <https://webbook.nist.gov/chemistry/>
- [34] S.E. Maxwell, H.D. Delaney, K. Kelley. *Designing Experiments and Analyzing Data. A Model Comparison Perspective* (Routledge, New York, 2017).  
DOI: 10.2307/2532173