

Применение методов машинного обучения для анализа инфракрасных спектров выдыхаемого человеком воздуха

Фуфурин Игорь Леонидович

МГТУ им. Н.Э. Баумана

igfil@mail.ru

Бережанский Павел Вячеславович

ГБУЗ «Морозовская ДГКБ ДЗМА»

p.berezhanskiy@mail.ru

Голяк Игорь Семенович

МГТУ им. Н.Э. Баумана

golyakis@bmstu.ru

Анфимов Дмитрий Романович

МГТУ им. Н.Э. Баумана

anfimovdr@student.bmstu.ru

Небритова Ольга Александровна

МГТУ им. Н.Э. Баумана

onebritova@bmstu.ru

Дёмкин Павел Павлович

АО «Центр прикладной физики МГТУ им. Н.Э. Баумана»

Demkin.Pavel1996@yandex.ru

Морозов Андрей Николаевич

МГТУ им. Н.Э. Баумана

amor@bmstu.ru

В настоящее время приблизительно 10,5 % населения земного шара в возрасте 20–79 лет в настоящее время живут с диабетом. Экспресс-диагностика ряда социально значимых заболеваний, в том числе сахарного диабета представляет в настоящее время актуальную задачу. В работе описан способ экспресс-диагностики ряда заболеваний по анализу выдыхаемого человеком воздуха. Метод основан на инфракрасной лазерной спектроскопии с применением квантового каскадного лазера, излучающего в импульсном режиме с пиковой мощностью до 150 МВт в спектральном диапазоне 5,3...12,8 мкм, и многоходовая газовая ячейка Эрриота с длиной оптического пути 76 м. По результатам анализа 1200 инфракрасных спектров выдыхаемого воздуха у 60 здоровых добровольцев (контрольная группа) и 60 добровольцев с подтвержденным диабетом I типа с применением одномерной сверточной нейронной сети показаны результаты классификации accuracy 99,7 %, recall 99,6 % и площадь под кривой AUC 99,9 %. Дополнительно исследованы 30 волонтеров с бронхиальной астмой. Применение метода опорных векторов дает значение accuracy порядка 95 % при классификации волонтеров на 3 группы (здоровые, диабет I типа, астма).

Ключевые слова: инфракрасная спектроскопия, ранняя диагностика, машинное обучение, сверточная нейронная сеть

Введение

Неинвазивная диагностика — одно из важнейших направлений развития современной медицины. Во всем мире насчитывается 537 миллионов взрослых в возрасте 20–79 лет (10,5 % из всех взрослых в этой возрастной группе) страдают диабетом в 2021 г. По оценкам международной федерации диабета (IDF), в 2021 году число детей

(0–19 лет) и подростков с диабетом I типа составит около 1,2 миллиона. По прогнозам, к 2030 г. это число возрастет до 643 миллионов, а к 2045 г. — до 783 миллионов [1]. Показана возможность анализа уровня глюкозы по выделениям кожи и в дыхании человека [2, 3]. В последнее время активно развивается область, связанная с анализом летучих органических соединений, выделяющихся из человеческого организма [4]. Показано, что подобные соединения выделяются из дыхания (872 соединения), слюны (359 соединений), крови (154 соединения), молока (256 соединений), кожных выделений (532 соединения), мочи (279 соединений) и кала (381 соединение). Ряд летучих органических соединений представляет собой биомаркеры конкретного заболевания человека. Так, средняя концентрация ацетона в здоровом дыхании колеблется от 293 до 870 ppb, а этанола — от 27 до 153 ppb [5]. У пациентов, страдающих сахарным диабетом средняя концентрация ацетона может превышать 1800 ppb [6]. В работе [7] показано, что измерение уровня выдыхаемого монооксида углерода (eCO) применимо для оценки состояния младенцев и детей ясельного возраста со стабильной астмой и во время острого приступа астмы. В работах [8–10] показано применение методов машинного обучения для анализа спектров многокомпонентных газовых смесей, в т.ч. выдыхаемого человеком воздуха. В работе [11] показаны возможности применения глубокого обучения для диагностики сахарного диабета I типа по инфракрасным спектрам выдыхаемого воздуха.

Экспериментальная установка

На рис. 1 представлены внешний вид и оптическая схема экспериментальной установки.

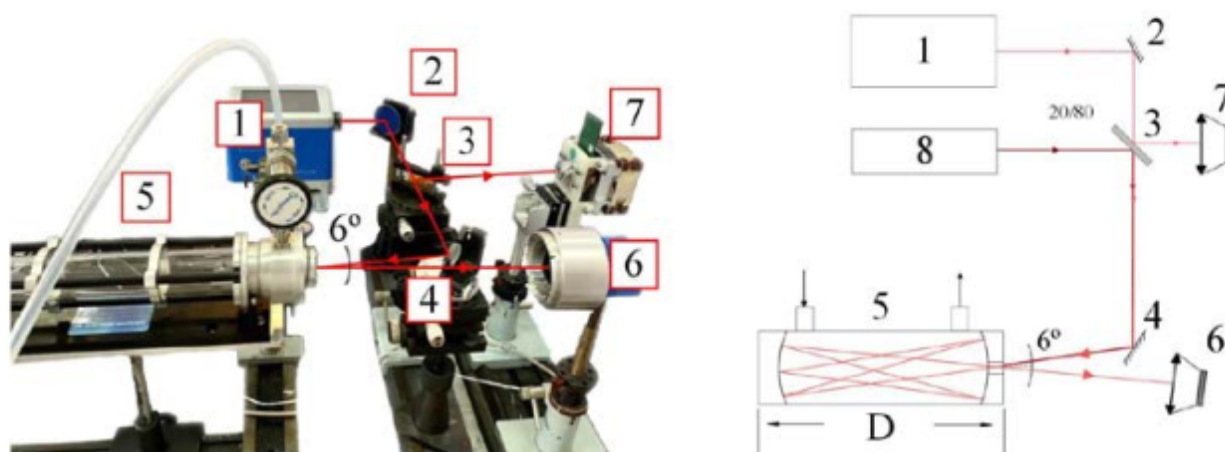


Рис. 1. Экспериментальная установка

Экспериментальная установка состоит из инфракрасного перестраиваемого квантово-каскадного лазера 1, излучающего в диапазоне длин волн 5,3...12,8 мкм в импульсном режиме с пиковой мощностью 150 мВт и длительностью импульса 50нс; многоходовой астигматической кюветы типа Эрриота с длиной оптического пути 76 м и базой 0,32 м. Принцип работы установки состоит в следующем: ИК-излучение от лазера 1 через зеркала 2, 4 и светоделитель 3 направляется на опорный фотоприемник 7 и газовую кювету 5. После заданного числа переотражений в кювете излучение регистрируется сигнальным фотоприемником 6. Воздушная проба волонтера предвари-

тельно осушается с помощью нафionoвого осушителя, что позволяет приблизительно на порядок снизить концентрации водяного пара в анализируемой пробе.

Исследуемые группы

Исследование проводилось с августа 2021 г. по сентябрь 2022 г. на базе МГТУ им. Н.Э. Баумана и ГБУЗ «Морозовская детская городская клиническая больница» ДЗМ. Протокол исследования был одобрен этическим комитетом Государственного бюджетного учреждения здравоохранения «Детская клиническая больница им. Морозова» Департамента здравоохранения города Москвы (Москва, Россия), № 174 от 18 января 2022 г.

На рис. 2 приведена половозрастная диаграмма здоровых волонтеров и волонтеров, страдающих диабетом I типа.

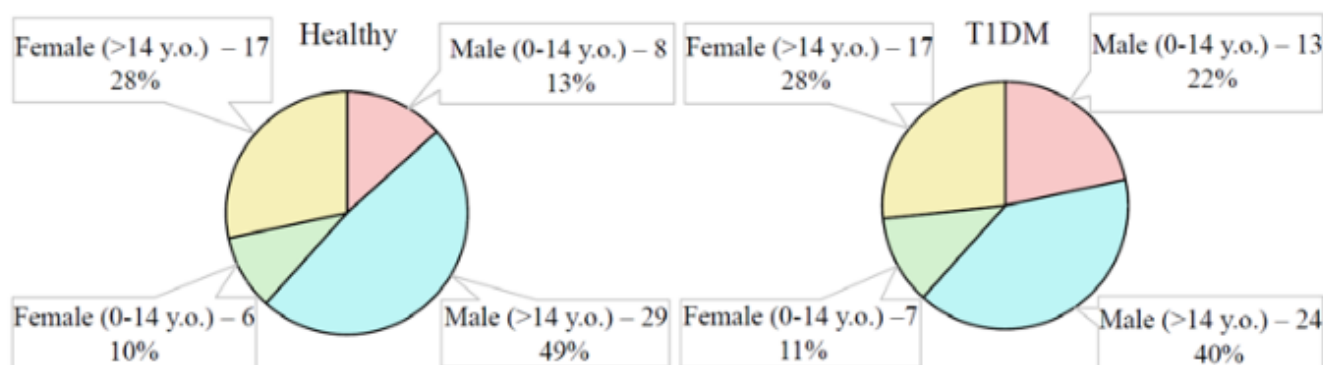


Рис. 2. Половозрастная диаграмма исследуемых групп

Дополнительно были исследованы 30 волонтеров, страдающих бронхиальной астмой.

Результаты и обсуждения

В нашей последней работе [11] показана возможность классификации здоровых волонтеров и волонтеров, страдающих сахарным диабетом. Для классификации применялась однослойная сверточная нейронная сеть, параметры которой и методы валидации описаны в [11]. Точность классификации при этом не менее 99 %.

Дополнительный интерес представляет возможность разделения пациентов одновременно по нескольким классам. Данный подход существенно расширит возможности метода и интерес потенциальных потребителей. Для этой цели дополнительно исследованы 30 пациентов, страдающих астмой. Таким образом, набор данных состоял из 60 здоровых волонтеров (600 спектров), 60 волонтеров, страдающих сахарным диабетом (600 спектров), и 30 волонтеров, страдающих бронхиальной астмой (300 спектров).

Для разделения волонтеров по ИК спектам выдыхаемого воздуха применялись методы машинного обучения. Так методы понижения размерности (метод главных компонент — PCA), позволяют визуально разделить исследуемые группы на классы (рис. 3). Видно, что существует принципиальная возможность классификации анализируемых групп (3 класса — здоровые, диабет, астма).

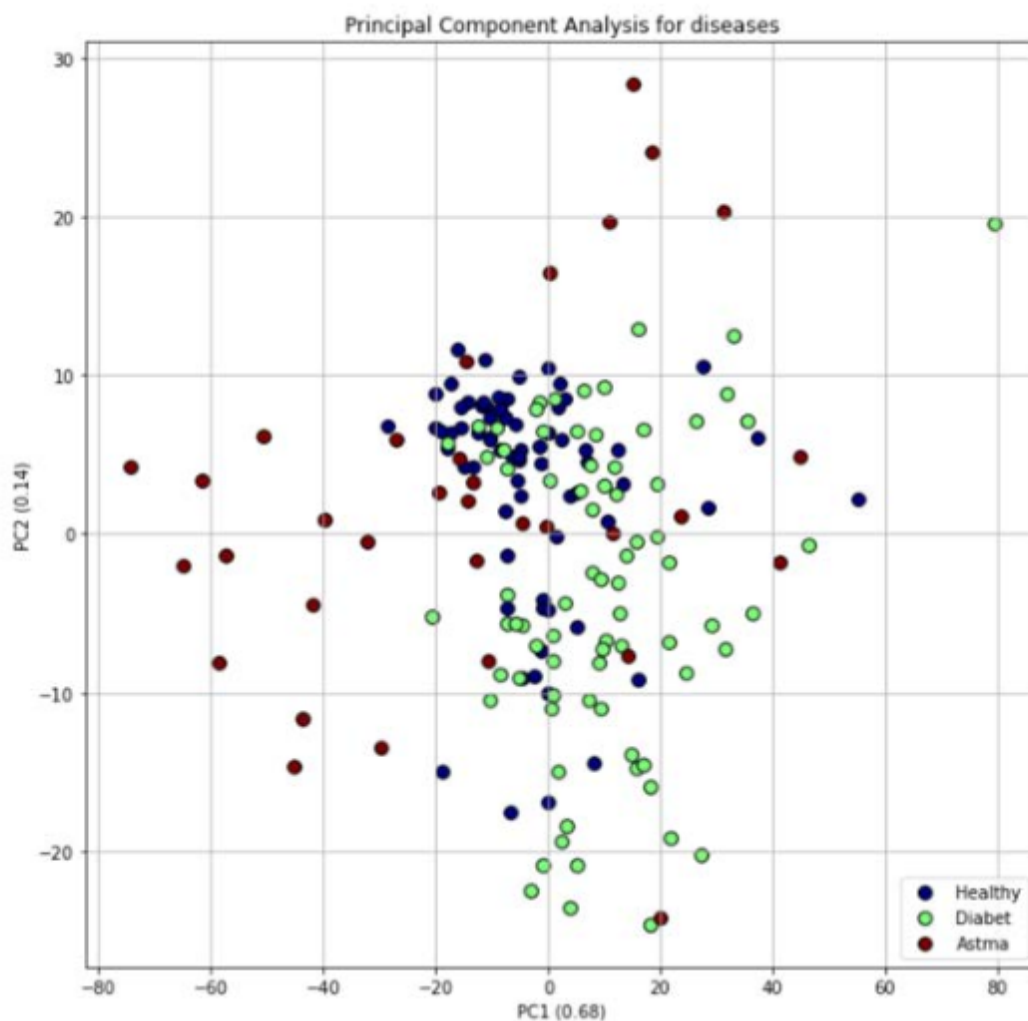


Рис. 3. Применение метода главных компонент для разделения пациентов на 3 класса — здоровые, диабет и астма

Результаты применения метода опорных векторов для анализируемых групп сведены в таблицу.

Результаты применения метода опорных векторов

Класс	Precision	Recall	F1-score
Здоровые	0,90	1,0	0,95
Диабет	1,0	0,88	0,93
Астма	1,0	1,0	1,0

Значение ассигасы в этом случае составляет 0,95.

Приведенные в работе результаты показывают принципиальную возможность применения описанного в работе метода для одновременной диагностики широкого ряда заболеваний, что, однако, требует многочисленных проверок на больших наборах данных и проведения клинических испытаний.

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

Литература

- [1] International Diabetes Federation. IDF Diabetes Atlas, 10th ed. International Diabetes Federation. Brussels, Belgium, 2021.
- [2] Bayrakli I. Breath analysis using external cavity diode lasers: a review. *Journal of Biomedical Optics*, 2017, vol. 22, no. 4, art. 040901. DOI: <https://doi.org/10.1117/1.JBO.22.4.040901>
- [3] Turner C. Potential of breath and skin analysis for monitoring blood glucose concentration in diabetes. *Expert Review of Molecular Diagnostics*, 2011, vol. 11, no. 5, pp. 497–503. DOI: <https://doi.org/10.1586/erm.11.31>
- [4] de Lacy Costello B. et al. A review of the volatiles from the healthy human body. *Journal of Breath Research*, 2014, vol. 8, no. 1, art. 014001. DOI: <https://doi.org/10.1088/1752-7155/8/1/014001>
- [5] Diskin A.M., Pan I P., Smith D. Time variation of ammonia, acetone, isoprene and ethanol in breath: a quantitative SIFT-MS study over 30 days. *Physiological Measurement*, 2003, vol. 24, no. 1, pp. 107–119. DOI: <https://doi.org/10.1088/0967-3334/24/1/308>
- [6] Deng C. et al. Determination of acetone in human breath by gas chromatography–mass spectrometry and solid-phase microextraction with on-fiber derivatization. *Journal of Chromatography*, 2004, vol. 810, no. 2, pp. 269–275. DOI: <https://doi.org/10.1016/j.jchromb.2004.08.013>
- [7] Ohara Y. et al. Exhaled carbon monoxide levels in infants and toddlers with episodic asthma. *Fukushima Journal of Medical Science*, 2020, vol. 66, no. 2, pp. 78–87. DOI: <https://doi.org/10.5387/fms.2019-02>
- [8] Fufurin I.L. et al. Numerical techniques for infrared spectra analysis of organic and inorganic volatile compounds for biomedical applications. *Optical Engineering*, 2021, vol. 60, no. 08. DOI: <https://doi.org/10.1117/1.OE.60.8.082016>
- [9] Golyak I.S. et al. Numerical methods of spectral analysis of multicomponent gas mixtures and human exhaled breath. *Computer Optics*, 2022, vol. 46, no. 4, pp. 650–658. DOI: <https://doi.org/10.18287/2412-6179-CO-1058>
- [10] Fufurin I.L. et al. Machine learning applications for spectral analysis of human exhaled breath for early diagnosis of diseases. *Optics in Health Care and Biomedical Optics X*. Ed. Luo Q. et al. SPIE, 2020. DOI: <https://doi.org/10.1117/12.2584043>
- [11] Fufurin I. et al. Deep Learning for type 1 diabetes mellitus diagnosis using infrared quantum cascade laser spectroscopy. *Materials*, 2022, vol. 15, no. 9, art. 2984. DOI: <https://doi.org/10.3390/ma15092984>