

## ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

## INFORMATICS AND PROGRAMMING

Научная статья

УДК 004.048

doi: 10.17223/19988605/73/14

**Нейронные сети в диагностике злокачественных новообразований  
по выдыхаемому воздуху****Артём Викторович Обходский<sup>1</sup>, Денис Евгеньевич Кульбакин<sup>2</sup>,  
Елена Владимировна Обходская<sup>3</sup>, Владислав Сергеевич Лаконкин<sup>4</sup>,  
Евгений Олегович Родионов<sup>5</sup>, Виктор Иванович Сачков<sup>6</sup>,  
Владимир Иванович Чернов<sup>7</sup>, Евгений Лхаматцыренович Чойнзон<sup>8</sup>**<sup>1, 4</sup> *Национальный исследовательский Томский политехнический университет, Томск, Россия*<sup>2, 4, 5, 7, 8</sup> *Научно-исследовательский институт онкологии Томского национального исследовательского  
медицинского центра Российской академии наук, Томск, Россия*<sup>3, 6</sup> *Национальный исследовательский Томский государственный университет, Томск, Россия*<sup>5</sup> *Сибирский государственный медицинский университет, Томск, Россия*<sup>1</sup> *art707@yandex.ru*<sup>2</sup> *kulbakin\_d@mail.ru*<sup>3</sup> *fil330a@yandex.ru*<sup>4</sup> *vsl13@tpu.ru*<sup>5</sup> *rodionov\_eo@oncology.tomsk.ru*<sup>6</sup> *vicsachkov@gmail.com*<sup>7</sup> *chernov@tnimc.ru*<sup>8</sup> *choynzonov@tnimc.ru*

**Аннотация.** Рассматриваются результаты исследования набора архитектур нейронных сетей с изменяемыми параметрами обучения для обработки данных с газоаналитических медицинских приборов, предназначенных для неинвазивной диагностики рака легких и верхних дыхательных путей. Алгоритм обеспечивает структурирование формата входных паттернов для нейронной сети с учетом обеспечения критерия максимума информации во входных данных. Диагностические данные с газоаналитических медицинских приборов представляют собой массивы целочисленных значений кодов с аналого-цифровых преобразователей. Алгоритм нейросетевой обработки данных реализован на языке программирования Python. В исследовании использовались оцифрованные пробы выдыхаемого воздуха от 154 человек. Для случаев отдельной дифференциации здоровых добровольцев, пациентов с раком легких и верхних дыхательных путей алгоритм нейросетевой обработки данных показал точность, в среднем превышающую 87%.

**Ключевые слова:** обработка данных; классификатор; искусственная нейронная сеть; архитектура нейронной сети; формат входных данных; оптимизация входных данных; параметры обучения; признак дифференцирования; эффективность классификатора.

**Благодарности:** Исследование выполнено при финансовой поддержке гранта РНФ в рамках научного проекта № 23-15-00177 «Теоретическое и клиническое обоснование молекулярной оценки состава выдыхаемого воздуха для диагностики онкологических заболеваний», <https://rscf.ru/project/23-15-00177/>

**Для цитирования:** Обходский А.В., Кульбакин Д.Е., Обходская Е.В., Лаконкин В.С., Родионов Е.О., Сачков В.И., Чернов В.И., Чойнзон Е.Л. Нейронные сети в диагностике злокачественных новообразований по выдыхаемому воздуху // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 122–133. doi: 10.17223/19988605/73/14

## Neural networks in the diagnosis of malignant neoplasms by exhaled air

Artem V. Obkhodskiy<sup>1</sup>, Denis E. Kulbakin<sup>2</sup>, Elena V. Obkhodskaya<sup>3</sup>, Vladislav S. Lakonkin<sup>4</sup>,  
Evgeniy O. Rodionov<sup>5</sup>, Victor I. Sachkov<sup>6</sup>, Vladimir I. Chernov<sup>7</sup>, Evgeny L. Choynzonov<sup>8</sup>

<sup>1,4</sup> National Research Tomsk Polytechnic University, Tomsk, Russian Federation

<sup>2,4,5,7,8</sup> Cancer Research Institute, Tomsk National Research Medical Center  
of the Russian Academy of Sciences, Tomsk, Russian Federation

<sup>3,6</sup> National Research Tomsk State University, Tomsk, Russian Federation

<sup>5</sup> Siberian State Medical University, Tomsk, Russian Federation

<sup>1</sup> art707@yandex.ru

<sup>2</sup> kulbakin\_d@mail.ru

<sup>3</sup> fil330a@yandex.ru

<sup>4</sup> vs113@tpu.ru

<sup>5</sup> rodionov\_eo@oncology.tomsk.ru

<sup>6</sup> vicsachkov@gmail.com

<sup>7</sup> chernov@tmimc.ru

<sup>8</sup> choynzonov@tmimc.ru

**Abstract.** The results of a study the set of neural network architectures with variable learning parameters for processing data from gas-analytical medical devices designed for noninvasive diagnosis of lung and upper respiratory tract cancer are considered. The algorithm provides structuring of the input pattern format for the neural network, taking into account the criterion of maximum information in the input data. Diagnostic data from gas-analytical medical devices are arrays of integer values of codes from analog-to-digital converters. The neural network data processing algorithm is implemented in the Python programming language. The study used digitized exhaled air samples from 154 people. For cases of separate differentiation of healthy volunteers, patients with lung and upper respiratory tract cancers, the neural network data processing algorithm showed an average accuracy exceeding 86%.

**Keywords:** data processing; classifier; artificial neural network; neural network architecture; input data format; input data optimization; learning parameters; differentiation feature; classifier efficiency.

**Acknowledgments:** The study was supported by the Russian Science Foundation grant № 23-15-00177 “Theoretical and clinical substantiation of the molecular assessment of the composition of exhaled air for the diagnosis of oncological diseases”, <https://rscf.ru/en/project/23-15-00177>

**For citation:** Obkhodskiy, A.V., Kulbakin, D.E., Obkhodskaya, E.V., Lakonkin, V.S., Rodionov, E.O., Sachkov, V.I., Chernov, V.I., Choynzonov, E.L. (2025) Neural networks in the diagnosis of malignant neoplasms by exhaled air. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 122–133. doi: 10.17223/19988605/73/14

### Введение

Интенсивное внедрение технологий искусственного интеллекта и нейронных сетей в медицинскую практику, в том числе в скрининг, позволяет обеспечить своевременный доступ людей к медицинским услугам в условиях значительного увеличения объема биомедицинских данных, необходимых для постановки диагноза, а также для назначения эффективного лечения с учетом индивидуальных особенностей пациентов. Автоматизация процессов диагностики и назначения терапии с применением программного обеспечения на алгоритмах искусственного интеллекта становится все более востребованной при отсутствии возможности личного посещения медицинских учреждений, пиковом увеличении количества обследуемых людей и интенсивном развитии технологий персонализированной медицины [1].

Повышение эффективности исследований и развитие инструментов анализа медицинских данных предполагает разные направления для применения технологий искусственного интеллекта. В работе [2] применялись нейросетевые алгоритмы многоклассовой сегментации изображений и метрики оценки качества mIoU и mDis для решения задачи различения нечетких локальных визуальных

признаков в биологических тканях во время сегментации изображений лапароскопической хирургии. Были достигнуты сравнительно высокие показатели эффективности нейросетевого алгоритма в среднем на уровне mIoU и mDice 78,19 и 83,84% соответственно. Данный подход позволяет автоматизировать процесс сегментации изображений на практике.

Алгоритмы искусственного интеллекта применяются при разработке лекарств, в том числе при выявлении и оптимизации противовоспалительных агентов для терапии рака. В работе [3] были выделены основные провоспалительные механизмы, участвующие в прогрессировании опухоли, такие как NF-κB, STAT3, COX-2 и ось JAK/IL-6, и продемонстрировано, как инструменты искусственного интеллекта, включая глубокое обучение и языковые модели, могут ускорить разработку и проверку терапевтических кандидатов. Появление индивидуализированной терапии рака, основанной на глубоком анализе микроокружения опухоли, тоже представляет собой революционную трансформацию в онкологии [4, 5].

В последние годы алгоритмы искусственного интеллекта все чаще входят в состав технологических инструментов цифровой ортопедии, в том числе в области виртуализированной дополненной реальности, 3D-печати и хирургических роботов [5, 6]. В медицинской практике становится частым применение дистанционной роботизированной хирургии, что позволяет решать проблему неравномерного распределения медицинских специалистов высшей категории на местах.

Визуализация в медицине является одним из основных инструментов для диагностики и лечения разных заболеваний. Интеграция технологий машинного зрения с искусственным интеллектом применяется в гепатологии [7], диагностике и лечении кожных заболеваний [8], онкологии [9, 10] и др.

Современные большие языковые модели (LLM) искусственного интеллекта в медицине обещают решать задачи от поддержки принятия клинических решений до обучения пациентов [11]. Достижения LLM в виде агентов еще больше расширяют их полезность, обеспечивая мультимодальную обработку и многозадачную обработку в сложных диагностических и клинических процессах [12]. Большие языковые модели, такие как GPT-4o, DeepSeek-R1, Gemini 2.0, Command-R, Claude 3, Qwen и Grok 3, демонстрируют различные частные характеристики, но в целом общую возможность в формировании сложной логики. Они быстро расширяются, ускоряя переход медицинского сектора от парадигмы Интернет+ к парадигме ИИ+ [13, 14].

Современные технологии искусственного интеллекта дают возможность для освоения и применения новых принципов работы медицинских изделий, которые ранее не могли быть оценены как эффективные в силу того, что обработка данных в этих медицинских приборах выполнялась оператором, а программная реализация алгоритмов обработки данных представлялась нерешаемой задачей [15]. Учитывая этот факт, технологии выявления летучих органических соединений в выдыхаемом воздухе все чаще рассматриваются в качестве перспективного инструмента для ранней диагностики рака. Это обусловлено неинвазивностью метода и возможностью проведения обследования человека за короткое время, что особенно важно при скрининге. Благодаря низкой растворимости в крови, летучие органические соединения легко проникают в альвеолярный воздух и выводятся из организма с выдыхаемым воздухом, что делает их доступными для анализа и диагностики заболеваний.

Общим результатом рассмотренных выше работ является повышение эффективности процессов за счет внедрения автоматизированных алгоритмов интеллектуальной обработки данных. В нашей работе представлены результаты создания базовой нейронной сети и алгоритма обработки данных, предназначенных для нейросетевой классификации данных с газоаналитических медицинских приборов диагностики рака легких и верхних дыхательных путей. Применение технологий искусственного интеллекта напрямую может влиять на увеличение средней продолжительности и улучшение качества жизни для большого числа людей, поскольку в случаях отдельных заболеваний раннее выявление практически определяет исход выживаемости человека в борьбе с заболеванием или способствует снижению уровня инвалидности.

## 1. Данные с газоаналитического медицинского прибора

В исследовании измерение концентрации компонентов выдыхаемого воздуха проводится с помощью газоаналитических медицинских приборов. Во всех приборах реализуется единый порядок взятия

проб выдыхаемого воздуха у всех участников исследования в контролируемых нормальных внешних условиях. Во время работы прибора полупроводниковые неселективные датчики функционируют в режиме термоциклирования, т.е. периодически нагреваются и охлаждаются. В процессе сбора данных регистрируется 10 полных циклов – от начальной точки нагрева до завершающей стадии остывания процесса термоциклирования [15, 16]. Всего в приборе применяется 24 датчика, а выходной сигнал с каждого датчика преобразуется в цифровой десятичный код в диапазоне от 0 до 1 023 с помощью аналого-цифрового преобразователя (АЦП). Пример формы сигнала с датчика 12 для групп пациентов с раком легких и здоровых добровольцев представлен на рис. 1.

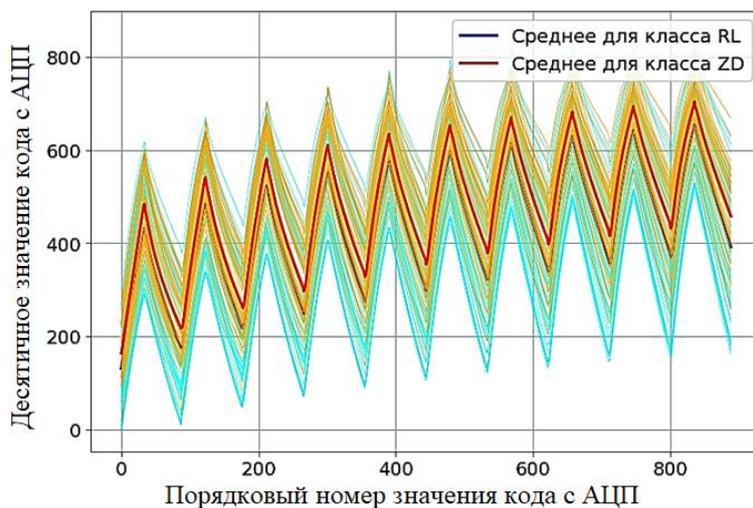


Рис. 1. Форма сигналов с датчика 12 для групп пациентов с раком легких (класс RL) и здоровых добровольцев (класс ZD)  
Fig. 1. Waveform from sensor 12 for groups of lung cancer patients (class RL) and healthy volunteers (class ZD)

В качестве критерия для возможной последующей качественной дифференциации двух классов с помощью нейронной сети на начальном этапе исследования бралось во внимание превышение более чем на 1% относительного отклонения средних значений этих классов. Такое отличие демонстрирует, например, датчик 12 на последнем импульсе термоциклирования. Средние значения сигналов для двух классов датчика 12 представлены на рис. 2. На последнем – десятом – импульсе сигналы с датчиков выходят на уставку, и возникает статическое различие между классами. Для различных патологий наборы датчиков, сигналы которых имеют большие отклонения средних значений, отличаются. В частности, для выборки здоровых пациентов и больных с раком головы и шеи разница средних значений для датчика № 12 меньше в 2–3 раза, а значит, извлекаемые паттерны будет сложнее различить нейронной сетью по этому датчику.

Выходные XML-файлы с газоаналитических медицинских приборов содержат избыточное количество данных в виде кодов АЦП, которые нецелесообразно в полном объеме применять в нейросетевой обработке. В связи с этим они преобразуются в один входной для нейронной сети txt-файл, в котором содержится информация о количестве объектов обучающей выборки, размеры входного и выходного слоев нейронной сети, массивы данных обучающей выборки, содержащие значения кодов АЦП последнего импульса термоциклирования. При чтении массива данных десятого импульса термоциклирования из XML-файлов применяется прореживание данных, что позволяет уменьшить размер входного слоя в пять раз без значительной потери качества нейросетевого классификатора. Каждый из 18 подряд выстроенных паттернов (признаков, значений) во входном массиве данных нейронной сети принадлежит 24 датчикам. Таким образом, размер входного слоя составляет 432, размер выходного слоя – 2. В выходном слое возможны два случая: [1, 0] – здоровый доброволец (состояние 1, класс ZD); [0, 1] – пациент с раком легких (состояние 2, класс RL). На рис. 3 представлена сигнатура по 24 датчикам для одного паттерна (признака, значения, точки) в пике десятого импульса термоциклирования. Данная сигнатура обладает такой специфичностью, что для человека задача классификации по ней представляется невозможной, но не для нейронной сети.

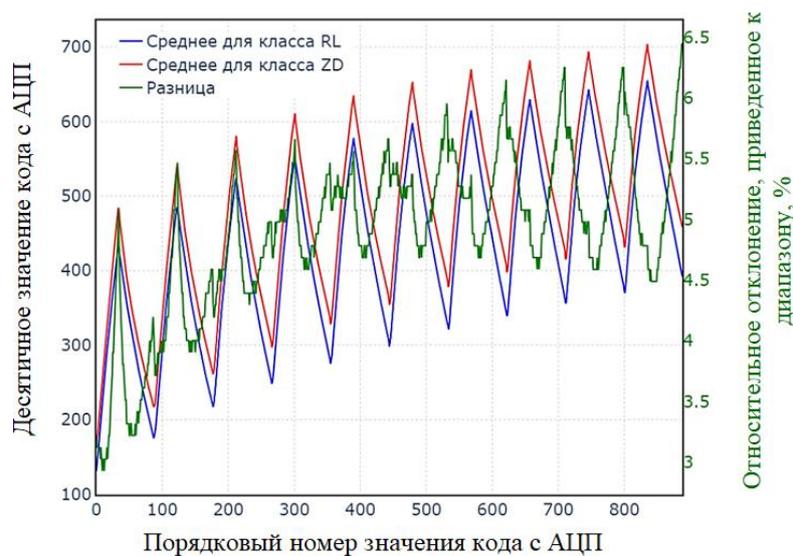


Рис. 2. Средние значения сигналов с датчика 12 для групп пациентов с раком легких (класс RL) и здоровых добровольцев (класс ZD) и относительное отклонение между ними  
 Fig. 2. The average values of the signals from sensor 12 for groups of patients with lung cancer (class RL) and healthy volunteers (class ZD) and the relative deviation between them

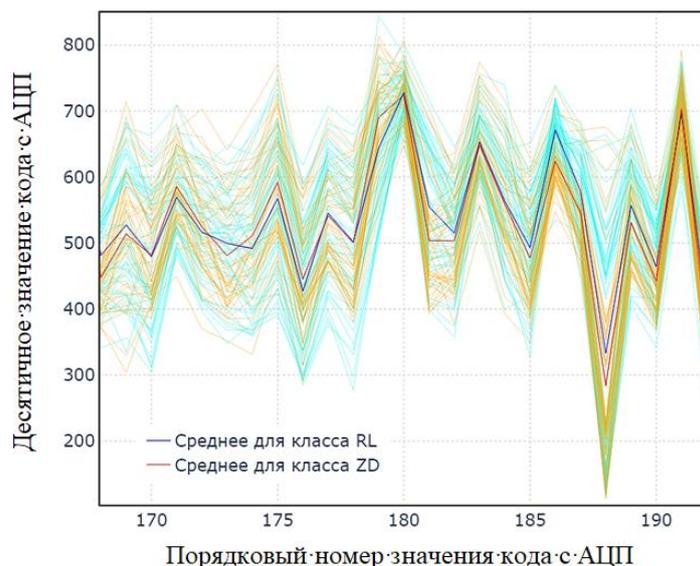


Рис. 3. Паттерн для групп пациентов с раком легких (класс RL) и здоровых добровольцев (класс ZD) по 24 датчикам для восьмой (пиковой) точки десятого импульса термоциклирования  
 Fig. 3. A pattern for groups of lung cancer patients (class RL) and healthy volunteers (class ZD) with 24 sensors for the eighth (peak) point of the tenth thermal cycling pulse

Не менее важным вопросом является изучение влияния сущности паттернов на точность нейросетевого классификатора. С учетом определенной далее в работе архитектуры и гиперпараметров нейронной сети исследования показали, что использование паттернов в виде разницы последней волны и первой или отношение последней волны на первую не вносят особого вклада как с точки зрения математики, так и с точки зрения физики газовых процессов. Основной вклад вносится последней волной, когда происходит выход на уставку.

## 2. Архитектура нейронной сети

Задача поиска оптимальной архитектуры нейронной сети для обработки данных является одной из ключевых в области машинного обучения и искусственного интеллекта. Это связано с тем, что

архитектура нейронной сети напрямую влияет на ее способность решать поставленные задачи: качество работы модели, ее производительность, вычислительную эффективность и возможность масштабирования [17].

Исходя из последующего применения в предварительной диагностике заболеваний, наиболее подходящими кандидатами для дифференциации здоровых добровольцев и пациентов с патологией являются нейронная сеть прямого распространения типа многослойный персептрон (полносвязная нейронная сеть) и сверточная нейронная сеть. В отличие от полносвязной нейронной сети сверточная выявляет не столько вклад каждого отдельного паттерна в результат, сколько вклад совокупностей этих паттернов. На рис. 4 приведена оптимальная архитектура нейронной сети для классификации имеющегося набора экспериментальных данных с газоаналитических медицинских приборов.

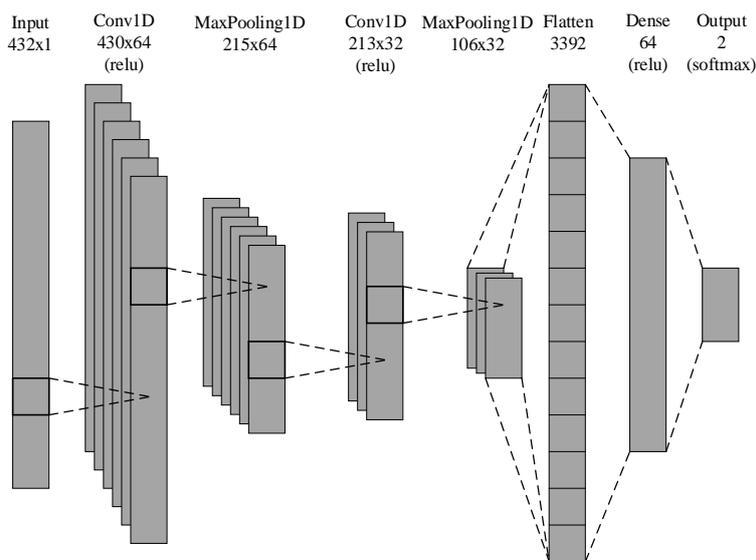


Рис. 4. Архитектура нейронной сети для диагностики рака легких и верхних дыхательных путей по выдыхаемому воздуху  
Fig. 4. Neural network architecture for lung and upper respiratory tract cancer diagnosis by exhaled air

Для поиска оптимальной архитектуры нейронной сети использовались такие стратегии, как случайный поиск [16] и байесовская оптимизация. Суть обеих методик заключается в автоматизации процесса подбора параметров нейронной сети. При случайном поиске параметры подбирались случайным образом, а при байесовской оптимизации использовалась вероятностная модель для представления неизвестной целевой функции (функции потерь). Приоритет в работе был отдан байесовской оптимизации.

При использовании этой стратегии определялось количество и содержание скрытых слоев, в каждом слое задавались диапазон изменения количества нейронов и шаг, с которым это количество менялось. Задавались варианты функций активации, например linear, tanh, sigmoid или relu. Для компиляции модели выбирался оптимизатор, например Adam или SGD, задавался диапазон изменения скорости обучения, выбиралась категория потерь. Для решаемой задачи это всегда categorical\_crossentropy. Определялась метрика, например accuracy, precision, recall или f1-score. В нашей работе оценка качества нейронной сети проводилась по метрике accuracy. Далее определялись количество комбинаций гиперпараметров для проверки, количество запусков для каждой комбинации, количество эпох и размер партии. В результате реализации байесовской оптимизации получена архитектура нейронной сети с параметрами, обеспечивающими лучшие показатели по выбранной метрике, точность на первой перекрестной группе для рака легких составила 90%. Во всех скрытых слоях нейронной сети использовалась функция активации relu. Для используемых наборов данных нейронные сети с тремя скрытыми слоями показали себя эффективными: обучение проходит относительно быстро, а качество получается сравнительно высоким. В дальнейшем в нейросетевом классификаторе предполагается использование трехслойной архитектуры сверточной нейронной сети.

Нейронная сеть в работе обучалась методом обратного распространения сигнала ошибки с учетом сверточной архитектуры. Завершение процесса обучения нейронной сети происходило по достижении

минимума ошибки, получаемого на верифицируемом множестве данных образцов выдыхаемого воздуха. Верифицирующее множество наборов данных выделялось из исходного набора данных проб выдыхаемого воздуха, из которого также отбирались пробы для обучения без пересечения обучающих и верифицирующих наборов.

Полная размерность входного слоя нейронной сети соответствует количеству значений с аналого-цифровых преобразователей по 24 датчикам прореженного десятого импульса термоциклирования и составляет 432 значения. Выходной слой – два нейрона, принимающие значения в диапазоне от 0 до 1. Исходный порог разделения проб составляет 0,5, но в процессе обучения истинное значение порога уточняется с учетом разного количества проб в классах обучающего набора нейронной сети. Значение на втором нейроне, близкое к 1, выше порога разделения означает вероятностную оценку принадлежности пробы выдыхаемого воздуха к классу пациентов с раком легких или верхних дыхательных путей, ниже порога до 0 – к классу здоровых добровольцев.

При определении оптимальных параметров обучения нейронной сети на наборах данных, полученных с газоаналитических медицинских приборов, проводился подбор гиперпараметров: количества эпох обучения, размера партии и скорости обучения. В результате оптимальное значение скорости обучения `learning_rate` составило 0,001. Исследование зависимости эффективности классификатора от количества скрытых слоев показало, что чем их больше, тем точнее классификатор, однако увеличение количества скрытых слоев более трех не способствует значительному увеличению точности нейронной сети, но способствует существенному увеличению времени ее обучения. Размер партии влияет на скорость обучения нейронной сети и ее эффективность, этот гиперпараметр всегда должен быть как можно больше, но он ограничен ресурсами вычислительной машины. В нашей работе размер партии `batch_size` составил 128. Количество эпох обучения, как правило, выбирается с учетом необходимости научить нейронную сеть обобщать, а не запоминать. Исследования показали, что для нашего набора данных этот процесс происходит за 20 эпох.

## 5. Экспериментальная часть

Программная реализация алгоритма нейросетевой обработки данных осуществлялась в среде Jupyter Notebook на языке программирования Python под управлением операционной системы Linux Ubuntu. Для этого использовались встраиваемые библиотеки TensorFlow и Keras.

При запуске алгоритма выполняются предварительная обработка исходных данных, обучение нейронной сети и перекрестная проверка. После обработки исходных данных осуществляются их анализ и визуализация. По результатам обучения нейронной сети формируются зависимости потерь и точности от эпох, визуализация t-SNE и выполняется решение обратной задачи нейронной сети. По результатам перекрестной проверки строятся зависимости ROC и PR [18], вычисляются их показатель AUC [19] и пороги разделения проб, формируется диаграмма распределения проб и матрицы неточностей.

В первом эксперименте по нейросетевой обработке данных с газоаналитических медицинских приборов использовался набор проб от 47 здоровых добровольцев и 53 пациентов с раком легких. После обучения нейронной сети были сформированы метрики оценки ее эффективности, в том числе ROC-характеристика, и определено значение параметра AUC. График ROC-характеристики представлен на рис. 5. Точность предварительной диагностики с классификатором на основе сверточной нейронной сети составила в среднем 90% при средних показателях чувствительности 86,79% и специфичности 93,62%.

Порог разделения положительных и отрицательных проб определен путем анализа ROC-характеристики. Его значение составило 0,534, и оно учитывается при проведении перекрестной проверки и для определения показателей специфичности и чувствительности нейросетевого классификатора. Величина AUC-ROC составила 0,950, что указывает на высокое качество классификации проб выдыхаемого воздуха с помощью сверточной нейронной сети и имеющегося обучающего набора экспериментальных данных.

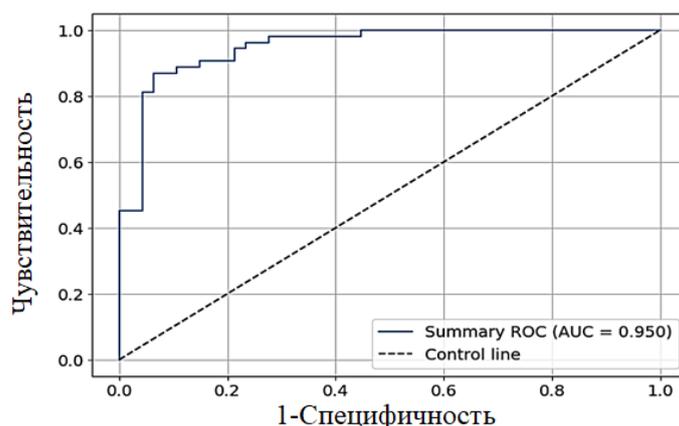


Рис. 5. ROC-характеристика качества обучения нейронной сети для классификации здоровых добровольцев и пациентов с раком легких

Fig. 5. ROC-characteristic of the neural network training quality for classifying healthy volunteers and patients with lung cancer

Во втором эксперименте были проведены аналогичные исследования с использованием набора данных, включающего 47 здоровых добровольцев и 53 пациента с раком верхних дыхательных путей, в том числе головы и шеи. Этот эксперимент позволил расширить исследование и оценить воспроизводимость полученных результатов в другой клинической группе. Анализ данных пациентов с раком верхних дыхательных путей проводился с использованием аналогичных методологических подходов, как и в первом эксперименте, что обеспечило сопоставимость результатов. Полученные результаты подтвердили статистическую значимость выявленных закономерностей и продемонстрировали потенциал применяемого подхода для диагностики онкологических заболеваний разных локализаций. Результаты формирования ROC-характеристики для второго эксперимента приведены на рис. 6.

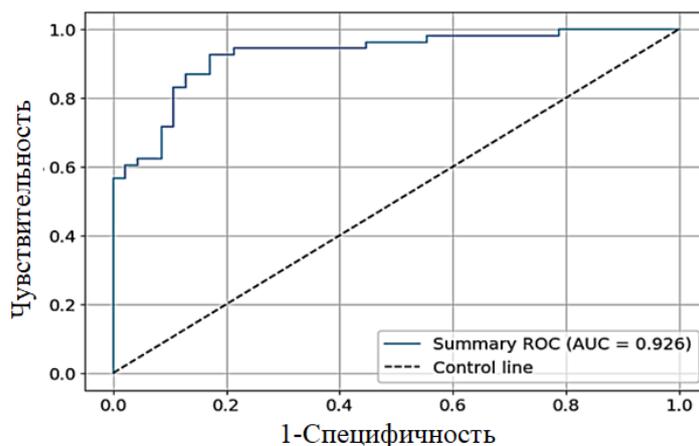


Рис. 6. ROC-характеристика качества обучения нейронной сети для классификации здоровых добровольцев и пациентов с раком верхних дыхательных путей

Fig. 6. ROC-characteristic of the neural network training quality for classifying healthy volunteers and patients with upper respiratory tract cancer

В результате второго эксперимента точность предварительной диагностики составила в среднем 87% при показателях чувствительности 92,45% и специфичности 82,98%. Порог разделения положительных и отрицательных проб выдыхаемого воздуха был найден с помощью анализа ROC-характеристики и составил 0,448. Величина AUC составила 0,926, что ниже, чем в первом эксперименте, но определяющее высокое качество работы классификатора.

Результаты исследований, представленные в работе, подтверждают наличие обобщенного признака дифференциации выдыхаемого воздуха от здоровых добровольцев, пациентов с раком легких и пациентов с раком верхних дыхательных путей. На имеющемся сравнительно небольшом наборе экспериментальных данных с учетом применяемых метрик качества обучения нейронной сети этот

признак дифференциации классов определяется с высокой достоверностью ( $P < 0,05$ ). Решение обратной задачи позволяет выявить датчики и, как следствие, группы летучих органических соединений, характерные для отдельной рассматриваемой патологии в сравнении с группой здоровых добровольцев.

На текущем этапе исследования изучались возможность и эффективность использования цифровых паттернов выдыхаемого воздуха с разным форматом и разными архитектурами нейронных сетей для дифференциации здоровых добровольцев и пациентов с патологией. С учетом полученных результатов можно также заключить, что большой перспективой обладает технология оцифровки не только выдыхаемого воздуха человека, но и всех возможных прямых и косвенных данных, характеризующих состояние здоровья человека, включая анамнестические показатели, биохимические и др., с формированием медицинского информационного цифрового двойника человека. Применение цифровых двойников человека совместно с алгоритмами нейросетевой обработки данных, учет дополнительных сведений, например таких, как возраст, имеющиеся хронические заболевания и др., позволит выявлять закономерности в изменении состояния здоровья и с высокой точностью определять рекомендации для его улучшения.

### Заключение

В работе предложена архитектура нейронной сети для проведения процедуры предварительной диагностики злокачественных новообразований легких и верхних дыхательных путей по анализу выдыхаемого воздуха. В исследовании применялись полносвязные и сверточные нейронные сети для классификации сигналов, поступающих в виде цифровых кодов с неселективных полупроводниковых газовых датчиков. Сравнительный анализ архитектур показал, что сверточная нейронная сеть (CNN) демонстрирует большую эффективность по сравнению с полносвязной (DNN) на данных с газоаналитических медицинских проборов. Это обусловлено способностью CNN выявлять общие сигнатуры в сигнале за счет использования ядер свертки, что особенно актуально при анализе временных рядов термоциклирования датчиков. Полносвязные сети, хотя и способны обучаться сложным зависимостям, страдают от избыточности параметров, что приводит к более быстрому переобучению, особенно при увеличении объема входных данных.

Исследование показало, что использование полного набора данных сигнала с одного цикла измерения не является необходимым для достижения высокой точности классификации. Наибольшую информативность несет последняя волна сигнала – финальный этап термоциклирования, когда датчики достигают максимальной чувствительности к отдельным летучим органическим соединениям малой концентрации в выдыхаемом воздухе. Прореживание данных в пределах последнего импульса термоциклирования не оказывает существенного влияния на качество классификатора, что позволяет сократить объем входных данных без потери точности. В то же время подача на вход всей последовательности сигнала увеличивает время обучения, требует больше вычислительных ресурсов и ускоряет процесс перехода в состояние переобучения, особенно в случае полносвязной архитектуры нейронной сети.

Сверточная архитектура оказалась более устойчивой к вариативности экспериментальных данных и обеспечила лучшую обобщающую способность. Характеристики точности CNN достигли уровня 86,79% по показателю чувствительности и 93,62% по специфичности для 47 проб выдыхаемого воздуха от здоровых добровольцев и 53 – от пациентов с раком легких; 92,45% по показателю чувствительности и 82,98% по специфичности для 47 проб здоровых добровольцев и 53 – пациентов с раком верхних дыхательных путей, что сопоставимо с эффективностью современных рентгенологических методов визуальной диагностики рассматриваемых патологий. Эти показатели были получены в условиях обеспечения баланса между сложностью модели и объемом входных данных, что обусловлено необходимостью как выбора оптимальной архитектуры нейронной сети, так и оптимизации формата входных данных и информационного пространства входных признаков дифференциации. Таким образом, сверточная нейронная сеть с использованием только последнего импульса сигнала термоциклирования представляет собой достаточное и рациональное решение с точки зрения производительности, точности и устойчивости к переобучению, что делает ее предпочтительной для интеграции в газо-

аналитические медицинские приборы скрининга злокачественных новообразований легких и верхних дыхательных путей.

#### Список источников

1. Kumar A. et al. Revolutionizing healthcare with 5G and AI: Integrating emerging technologies for personalized care and cancer management // *Intelligent Hospital*. 2025. V. 1 (1). Art. 100005. P. 1–10. doi: 10.1016/j.inhs.2025.100005
2. Peng K. et al. Image Segmentation Network for Laparoscopic Surgery // *Biomimetic Intelligence and Robotics*. 2025. V. 5. P. 1–12. doi: 10.1016/j.birob.2025.100236
3. Paithankar V., Devnani D., Nimburkar T.A. A Review Article on "AI-Guided Discovery of Novel Antiinflammatory Agents for Cancer Therapy: A New Era in Drug Development" // *Intelligent Hospital*. 2025. V. 1 (1). Art. 100007. doi: 10.1016/j.inhs.2025.100007
4. Feng S., Yin X., Shen Y. Artificial intelligence powered precision: Unveiling the tumor microenvironment for a new frontier in personalized cancer therapy // *Intelligent Medicine*. 2025. V. 5. P. 95–98. doi: 10.1016/j.imed.2025.02.001
5. Zhang L. et al. Osteosarcoma knowledge graph question answering system: deep learning-based knowledge graph and large language model fusion // *Intelligent Medicine*. 2025. V. 5. P. 99–110. doi: 10.1016/j.imed.2024.12.001
6. Asadi O., Hajhosseini M., Shirzadi S. et al. Improvement of classification accuracy of functional near-infrared spectroscopy signals for hand motion and motor imagery using a common spatial pattern algorithm // *Intelligent Medicine*. 2025. V. 5. P. 123–131. doi: 10.1016/j.imed.2024.05.004
7. Feng Y. et al. Application of artificial intelligence-based computer vision methods in liver diseases: a bibliometric analysis // *Intelligent Medicine*. 2025. V. 5. P. 111–122. doi: 10.1016/j.imed.2024.09.008
8. Li T. et al. Application of multimodal deep learning in the auxiliary diagnosis and treatment of dermatological diseases // *Intelligent Medicine*. 2025. V. 5. P. 132–140. doi: 10.1016/j.imed.2024.10.002
9. Ram U.S., Pogue J.A., Soike M. et al. Assessing quantitative performance and expert review of multiple deep learning-based frameworks for computed tomography-based abdominal organ auto-segmentation // *Intelligent Oncology*. 2025. V. 1. P. 160–171. doi: 10.1016/j.intonc.2025.03.003
10. Wang S., Guo X., Ma J. et al. AWCDL: Automatic weight calibration deep learning for detecting HER2 status in whole-slide breast cancer image // *Intelligent Oncology*. 2025. V. 1. P. 128–138. doi: 10.1016/j.intonc.2025.03.008
11. Adeoye J., Su Y. Deep learning with data transformation improves cancer risk prediction in oral precancerous conditions // *Intelligent Medicine*. 2025. V. 5. P. 141–150. doi: 10.1016/j.imed.2024.11.003
12. Chen X., Xiang J., Lu S. et al. Evaluating large language models and agents in healthcare: key challenges in clinical applications // *Intelligent Medicine*. 2025. V. 5. P. 151–163. doi: 10.1016/j.imed.2025.03.002
13. Wei M., Teh B.S., Xu B. The new paradigm for global oncologists in the era of generative AI // *Intelligent Oncology*. 2025. V. 1. P. 189–192. doi: 10.1016/j.intonc.2025.06.002
14. Cai W. DeepSeek AI: Transforming medical AI with cost-efficiency, transparency, and privacy preservation // *Intelligent Oncology*. 2025. V. 1. P. 172–175. doi: 10.1016/j.intonc.2025.03.006
15. Chernov V.I., Choynzonov E.L., Kulbakin D.E., Obkhodskaya E.V., Obkhodskiy A.V., Popov A.S., Sachkova A.S., Sachkov V.I. Cancer Diagnosis by Neural Network Analysis of Data from Semiconductor Sensors // *Diagnostics*. 2020. V. 10 (9). Art. 677. doi: 10.3390/diagnostics10090677
16. Обходский А.В., Обходская Е.В., Кульбакин Д.Е., Чойнзонов Е.Л., Родионов Е.О., Сачков В.И., Чернов В.И., Попов А.С. Программный комплекс нейросетевой классификации данных газоаналитического обследования дыхательной системы // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2024. № 69. С. 112–123. doi: 10.17223/19988605/69/12
17. Сорокоумова Д.А., Корелин О.Н., Сорокоумов А.В. Построение и обучение нейронной сети для решения задачи распознавания речи // *Труды Нижегородского государственного технического университета им. П. Е. Алексеева*. 2015. № 3 (110). С. 77–84.
18. Maaten van der L.J.P., Hinton G.E. Visualizing High-Dimensional Data Using t-SNE // *J. of Machine Learning Research*. 2008. V. 9. P. 2579–2605.
19. Sokolova M., Japkowicz N., Szpakowicz S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation // *Advances in Artificial Intelligence*. 2006. V. 4304. P. 1015–1021.

#### References

1. Kumar, A. et al. (2025) Revolutionizing healthcare with 5G and AI: Integrating emerging technologies for personalized care and cancer management. *Intelligent Hospital*. 1(1). Art. 100005. pp. 1–10. doi: 10.1016/j.inhs.2025.100005
2. Peng, K. et al. (2025) Image Segmentation Network for Laparoscopic Surgery. *Biomimetic Intelligence and Robotics*. 5. pp. 1–12. doi: 10.1016/j.birob.2025.100236
3. Paithankar, V., Devnani, D. & Nimburkar, T.A. (2025) A Review Article on "AI-Guided Discovery of Novel Antiinflammatory Agents for Cancer Therapy: A New Era in Drug Development." *Intelligent Hospital*. 1(1). Art. 100007. doi: 10.1016/j.inhs.2025.100007
4. Feng, S., Yin, X. & Shen, Y. (2025) Artificial intelligence powered precision: Unveiling the tumor microenvironment for a new frontier in personalized cancer therapy. *Intelligent Medicine*. 5. pp. 95–98. doi: 10.1016/j.imed.2025.02.001
5. Zhang, L. et al. (2025) Osteosarcoma knowledge graph question answering system: deep learning-based knowledge graph and large language model fusion. *Intelligent Medicine*. 5. pp. 99–110. doi: 10.1016/j.imed.2024.12.001

6. Asadi, O., Hajhosseini, M., Shirzadi, S. et al. (2025) Improvement of classification accuracy of functional near-infrared spectroscopy signals for hand motion and motor imagery using a common spatial pattern algorithm. *Intelligent Medicine*. 5. pp. 123–131. doi: 10.1016/j.imed.2024.05.004
7. Feng, Y. et al. (2025) Application of artificial intelligence-based computer vision methods in liver diseases: a bibliometric analysis. *Intelligent Medicine*. 5. pp. 111–122. doi: 10.1016/j.imed.2024.09.008.
8. Li, T. et al. (2025) Application of multimodal deep learning in the auxiliary diagnosis and treatment of dermatological diseases. *Intelligent Medicine*. 5. pp. 132–140. doi: 10.1016/j.imed.2024.10.002.
9. Ram, U.S., Pogue, J.A., Soike, M. et al. (2025) Assessing quantitative performance and expert review of multiple deep learning-based frameworks for computed tomography-based abdominal organ auto-segmentation. *Intelligent Oncology*. 1. pp. 160–171. doi: 10.1016/j.intonc.2025.03.003
10. Wang, S., Guo, X., Ma, J. et al. (2025) AWC DL: Automatic weight calibration deep learning for detecting HER2 status in whole-slide breast cancer image. *Intelligent Oncology*. 1. pp. 128–138. doi: 10.1016/j.intonc.2025.03.008
11. Adeoye, J. & Su, Y. (2025) Deep learning with data transformation improves cancer risk prediction in oral precancerous conditions. *Intelligent Medicine*. 5. pp. 141–150. doi: 10.1016/j.imed.2024.11.003
12. Chen, X., Xiang, J., Lu, S. et al. (2025) Evaluating large language models and agents in healthcare: key challenges in clinical applications. *Intelligent Medicine*. 5. pp. 151–163. doi: 10.1016/j.imed.2025.03.002
13. Wei, M., The, B.S. & Xu, B. (2025) The new paradigm for global oncologists in the era of generative AI. *Intelligent Oncology*. 1. pp. 189–192. doi: 10.1016/j.intonc.2025.06.002
14. Cai, W. (2025) DeepSeek AI: Transforming medical AI with cost-efficiency, transparency, and privacy preservation. *Intelligent Oncology*. 1. pp. 172–175. doi: 10.1016/j.intonc.2025.03.006
15. Chernov, V.I., Choynzonov, E.L., Kulbakin, D.E., Obkhodskaya, E.V., Obkhodskiy, A.V., Popov, A.S., Sachkova, A.S. & Sachkov, V.I. (2020) Cancer Diagnosis by Neural Network Analysis of Data from Semiconductor Sensors. *Diagnostics*. 10(9). Art. 677. doi: 10.3390/diagnostics10090677
16. Obkhodskiy, A.V., Kulbakin, D.E., Obkhodskaya, E.V., Popov, A.S., Rodionov, E.O., Sachkov, V.I., Chernov, V.I. & Choynzonov, E.L. (2024) Neural network classification software for the gas analytical survey data of respiratory system. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 69. pp. 112–123. doi: 10.17223/19988605/69/12
17. Sorokoumova, D.A., Korelin, O.N. & Sorokoumov, A.V. (2015) Postroeniye i obuchenie neyronnoy seti dlya resheniya zadachi raspoznavaniya rechi [Construction and Training of a Neural Network for Solving Speech Recognition Tasks]. *Trudy Nizhegorodskogo gosudarstvennogo tekhnicheskogo universiteta im. R. Ye. Alekseeva*. 3(110). pp. 77–84.
18. Maaten, van der L.J.P. & Hinton, G.E. (2008) Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*. 9. pp. 2579–2605.
19. Sokolova, M., Japkowicz, N. & Szpakowicz, S. (2006) Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. *Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science*. 4304. pp. 1015–1021.

#### **Информация об авторах:**

**Обходский Артём Викторович** – кандидат технических наук, доцент инженерной школы ядерных технологий Национального исследовательского Томского политехнического университета (Томск, Россия). E-mail: art707@yandex.ru

**Кульбакин Денис Евгеньевич** – доктор медицинских наук, заведующий отделением Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра РАН (Томск, Россия). E-mail: kulbakin\_d@mail.ru

**Обходская Елена Владимировна** – кандидат технических наук, старший научный сотрудник лаборатории химических технологий Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: fil330a@yandex.ru

**Лаконкин Владислав Сергеевич** – лаборант-исследователь Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра Российской академии наук; студент инженерной школы ядерных технологий Национального исследовательского Томского политехнического университета (Томск, Россия). E-mail: vsl13@tpu.ru

**Родионов Евгений Олегович** – кандидат медицинских наук, старший научный сотрудник Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра Российской академии наук; ассистент кафедры онкологии Сибирского государственного медицинского университета (Томск, Россия). E-mail: rodionov\_eo@oncology.tomsk.ru

**Сачков Виктор Иванович** – доктор химических наук, заведующий лабораторией химических технологий Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: vicsachkov@gmail.com

**Чернов Владимир Иванович** – профессор, доктор медицинских наук, член-корреспондент РАН, заместитель директора по научной и инновационной работе, заведующий отделением радионуклидной терапии и диагностики Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра РАН (Томск, Россия). E-mail: chernov@tnimc.ru

**Чойнзоннов Евгений Лхамцыренович** – профессор, доктор медицинских наук, академик РАН, директор Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра РАН (Томск, Россия). E-mail: choynzonov@tnimc.ru

**Вклад авторов:** все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

**Information about the authors:**

**Obkhodskiy Artem V.** (Candidate of Technical Sciences, National Research Tomsk Polytechnic University, Tomsk, Russian Federation), E-mail: art707@yandex.ru

**Kulbakin Denis E.** (Doctor of Medical Sciences, Head of Department of Head and Neck Tumors, Cancer Research Institute, Tomsk National Research Medical Center of the RAS, Tomsk, Russian Federation). E-mail: kulbakin\_d@mail.ru

**Obkhodskaya Elena V.** (Candidate of Technical Sciences, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: fil330a@yandex.ru

**Lakonkin Vladislav S.** (Laboratory Research Assistant, Tomsk National Research Medical Center of the Russian Academy of Sciences; Student, School of Nuclear Technology, National Research Tomsk Polytechnic University, Tomsk, Russian Federation). E-mail: vsl13@tpu.ru

**Rodionov Evgeniy O.** (Candidate of Medical Sciences, Cancer Research Institute, Tomsk National Research Medical Center of the RAS; Siberian State Medical University, Tomsk, Russian Federation). E-mail: rodionov\_eo@oncology.tomsk.ru

**Sachkov Victor I.** (Doctor of Chemistry Sciences, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: vicsachkov@gmail.com

**Chernov Vladimir I.** (Professor, Doctor of Medical Sciences, Head of Nuclear Medicine Department, Cancer Research Institute; Tomsk National Research Medical Center of the RAS, Tomsk, Russian Federation). E-mail: chernov@tnimc.ru

**Choynzonov Evgeny L.** (Professor, Doctor of Medical Sciences, Member of the RAS, Director of the Cancer Research Institute, Tomsk National Research Medical Center; Head of Oncology Department of Siberian State Medical University, Tomsk, Russian Federation). E-mail: choynzonov@tnimc.ru

**Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.**

*Поступила в редакцию 27.07.2025; принята к публикации 02.12.2025*

*Received 27.07.2025; accepted for publication 02.12.2025*