



Исследования и разработки в области новых
информационных технологий и их приложений

<https://doi.org/10.25559/SITITO.021.202503.455-466>

УДК 004.89

Сравнительный анализ архитектур ResNet и типов спектральных данных для классификации речи пациентов после гемиглоссэктомии

С. Д. Томила^{*}, Д. И. Новохрестова, П. Ю. Лаптев,
Е. Ю. Костюченко

Оригинальная статья

ФГАОУ ВО «Томский государственный университет систем управления и радиоэлектроники», г. Томск, Российская Федерация

Адрес: 634050, Российская Федерация, г. Томск, пр. Ленина, д. 40

^{*} tsd@fb.tusur.ru

Аннотация

Введение. Рак языка – серьезная онкологическая проблема, лечение которой часто сопровождается нарушениями речевых функций. Гемиглоссэктомия, как основной метод хирургического вмешательства, требует длительной реабилитации, а традиционные методы оценки восстановления речи основываются на субъективных наблюдениях. В данной работе предложен упрощенный метод оценки речи пациентов после гемиглоссэктомии с использованием нейронных сетей ResNet, что позволяет перейти к объективному анализу на основе спектральных данных и повысить точность диагностики.

Материалы и методы. Исследование основано на аудиозаписях 29 пациентов, перенёсших гемиглоссэктомию с реконструкцией и без неё. Записи были сделаны на трёх этапах: до операции, сразу после неё и после трёхмесячной реабилитации. Аудиофайлы преобразовывались в три типа спектральных данных: спектрограммы, мел-спектрограммы и MFCC (мел-частотные кепстральные коэффициенты). В ходе обучения моделей применялись различные архитектуры семейства ResNet – 18, 34, 50, 101 и 152 – с вариацией параметров dropout и размеров входных изображений. Данные были случайным образом разделены на обучающий (70%), проверочный (15%) и тестовый (15%) наборы.

Результаты исследования. Максимальная точность классификации составила 0,69 для полного набора данных (ResNet-152), 0,71 для пациентов без реконструкции и 0,77 для пациентов с реконструкцией (ResNet-34). Разделение данных по типу операции позволило улучшить точность до 0,7692 для пациентов с реконструкцией при использовании мел-спектрограмм и ResNet-34, что подтверждает их преимущество для анализа нелинейных изменений частотных характеристик.

Обсуждение и заключение. Предложенный метод позволяет упростить оценку речи, обеспечивая объективный и стандартизированный мониторинг динамики восстановления речевых функций. Это снижает нагрузку на медицинский персонал и повышает эффективность реабилитационных программ. Для дальнейшего улучшения точности планируется реализовать ансамблевый подход, объединяющий результаты анализа различных спектральных представлений, а также анализировать только значимые части спектрограмм, что позволит повысить надёжность классификации.

Ключевые слова: нейронная сеть, анализ речи, речевая реабилитация, архитектура ResNet, MFCC, спектрограммы, мел-спектрограммы

Финансирование: Данная работа выполнялась в рамках Программы развития ТУСУР на 2025-2036 годы Программы стратегического академического лидерства «Приоритет 2030».

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Томила С. Д., Новохрестова Д. И., Лаптев П. Ю., Костюченко Е. Ю. Сравнительный анализ архитектур ResNet и типов спектральных данных для классификации речи пациентов после гемиглоссэктомии // Современные информационные технологии и ИТ-образование. 2025. Т. 21, № 3. С. 455-466.
<https://doi.org/10.25559/SITITO.021.202503.455-466>

© Томила С. Д., Новохрестова Д. И., Лаптев П. Ю., Костюченко Е. Ю., 2025



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Research and Development in the Field of New IT
and Their Applications

Comparative Analysis of ResNet Architectures and Spectral Data Types for Speech Classification in Patients after Hemiglossectomy

S. D. Tomilina*, D. I. Novokhrestova, P. Yu. Laptev,
E. Yu. Kostyuchenko

Original article

Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation

Address: 40 Prospekt Lenina, Tomsk 634050, Russian Federation

* tsd@fb.tusur.ru

Abstract

Introduction. Tongue cancer is a serious oncological condition that often leads to speech impairments. Hemiglossectomy, the primary surgical intervention, requires prolonged rehabilitation, whereas traditional speech recovery assessments rely on subjective observations. This study introduces a simplified method for evaluating the speech of patients after hemiglossectomy using ResNet neural networks, thereby enabling objective spectral data-based analysis and improving diagnostic precision.

Materials and Methods. The study is based on audio recordings of 29 patients who underwent hemiglossectomy with and without reconstruction. Recordings were made at three stages: before surgery, immediately after surgery, and after a three-month rehabilitation period. Audio files were transformed into three types of spectral data: spectrograms, mel-spectrograms, and MFCCs (Mel-Frequency Cepstral Coefficients). Various ResNet architectures (18, 34, 50, 101, and 152) with different dropout parameters and image sizes were used for model training. The data were randomly divided into training (70%), validation (15%), and test (15%) sets.

Results. The maximum classification accuracy was 0.69 for the full dataset (ResNet-152), 0.71 for patients without reconstruction, and 0.77 for patients with reconstruction (ResNet-34). Data segmentation by surgery type improved accuracy to 0.7692 for patients with reconstruction using mel-spectrograms and ResNet-34, confirming their advantage in analyzing nonlinear changes in frequency characteristics.

Discussion and Conclusion. The proposed method simplifies speech assessment by providing objective and standardized monitoring of speech function recovery dynamics. This reduces the workload on medical staff and improves the effectiveness of rehabilitation programs. To further improve accuracy, an ensemble approach is planned that combines the results of analyzing different spectral representations and focuses on significant parts of spectrograms, which will increase classification reliability.

Keywords: neural network, speech analysis, speech rehabilitation, ResNet architecture, MFCC, spectrograms, mel-spectrograms

Funding: The study was carried out under the TUSUR Development Program for 2025-2036 as part of the Priority 2030 Strategic Academic Leadership Program.

Conflict of interests: The authors declares no conflict of interest.

For citation: Tomilina S.D., Novokhrestova D.I., Laptev P.Yu., Kostyuchenko E.Yu. Comparative Analysis of ResNet Architectures and Spectral Data Types for Speech Classification in Patients after Hemiglossectomy. *Modern Information Technologies and IT-Education*. 2025;21(3):455-466. <https://doi.org/10.25559/SITITO.021.202503.455-466>



Введение

Рак языка является серьезной медицинской проблемой, число случаев заболевания которой растет. По прогнозам, к 2030 году уровень заболеваемости в России может достичь 29 случаев на 100 тысяч человек [1]. Этот рост отражает глобальную тенденцию увеличения заболеваемости раком полости рта, охватывающую как Россию, так и другие страны [2, 3]. Влияние таких факторов риска, как курение, употребление алкоголя, вирус папилломы человека (ВПЧ) и культурные особенности, остается сложным и многогранным [2]. Онкологические заболевания челюстно-лицевой области занимают значительную долю онкологической патологии и оказывают заметное влияние на общую смертность [4].

Хирургическое вмешательство в виде гемиглоссэктомии, заключающееся в удалении части языка, является одним из ключевых методов лечения рака языка [3, 5]. Однако эта процедура часто вызывает значительные нарушения речевой, жевательной и глотательной функций, что связано с серьезными изменениями структуры и функционирования полости рта и речевого аппарата [6, 7]. Такие последствия подчеркивают необходимость разработки более эффективных методов диагностики, лечения и реабилитации, поскольку даже после успешного удаления опухоли и реабилитации пациенты могут испытывать трудности в повседневном общении.

Хотя использование ResNet для анализа спектральных данных и не является новым, применение этой технологии для оценки речи пациентов после хирургического вмешательства при раке языка представляет собой инновационный подход. Предложенный метод позволяет объективно и быстро оценивать изменения речевых функций на разных этапах реабилитации (до операции, после операции и в процессе восстановления). Это, в свою очередь, способствует разработке более эффективных и персонализированных программ реабилитации, так как специалисты получают количественные данные о прогрессе пациентов.

В клинической практике оценка динамики восстановления речи пациентов после гемиглоссэктомии зачастую основывается на субъективных методах. Логопеды и врачи анализируют изменения в речи пациентов, но в большинстве случаев это происходит на слух, без использования формализованных метрик или систематического сравнения с предыдущими записями. При этом:

- большое количество пациентов и длительный период реабилитации (от 3 до 12 месяцев) затрудняют объективную оценку изменений в речи;

- отсутствие автоматизированных инструментов приводит к тому, что специалисты вынуждены полагаться на память и субъективное восприятие, что может снижать точность оценки;

- необходимость сравнения записей при их наличии (например, до операции, после операции и в процессе реабилитации) требует значительных временных затрат, особенно при большом потоке пациентов.

В клинической практике это может быть использовано для разработки персонализированных программ реабилитации, учитывающих специфику речевых нарушений после различных типов хирургических вмешательств.

В настоящее время не существует стандартных методов оценки эффективности речевой реабилитации после хирургического лечения рака языка [3]. Ранее был предложен подход к оценке качества и разборчивости речи, основанный на сравнении огибающей речевого сигнала с использованием различных показателей сходства, а также использовании методов классификации с полученными значениями показателей для расчета окончательной оценки качества речи [8,9]. Использование спектрального анализа данных с использованием нейронных сетей может повысить точность оценки изменений в речи пациентов и оптимизации реабилитации.

Несмотря на широкое применение нейронных сетей, включая архитектуры ResNet, для анализа аудиосигналов и речи в медицине, в современной литературе отсутствуют исследования, посвященные систематическому сравнению эффективности различных типов спектральных данных (мел-спектрограммы, MFCC, спектрограммы) и архитектур ResNet для оценки речи пациентов после гемиглоссэктомии – операции, существенно влияющей на речевые функции. В существующих работах [8, 9] анализируют общие метрики разборчивости речи (сравнение огибающей речевого сигнала), но не учитывают специфику хирургических вмешательств (например, реконструкцию языка) и не анализируют, как выбор типа спектральных данных влияет на точность классификации. Это ограничение особенно актуально для пациентов после гемиглоссэктомии, где изменения в речи могут варьироваться в зависимости от типа операции (с реконструкцией или без неё). В данной работе проводится сравнительное исследование архитектур ResNet и типов спектральных данных для классификации речи пациентов после гемиглоссэктомии – задачи, ранее не рассматривавшейся в литературе.

Целью данного исследования является сравнение эффективности различных архитектур ResNet и типов спектральных данных (мел-спектрограммы, MFCC, спектрограммы) для классификации речи пациентов после гемиглоссэктомии с целью автоматизации и улучшения процесса отслеживания динамики речевой реабилитации. Это позволит повысить точность оценки восстановления речи, снизить нагрузку на медицинский персонал и улучшить индивидуальные программы реабилитации. В отличие от существующих работ, где анализируются общие метрики разборчивости речи, в данном исследовании:

1. Определяются оптимальные комбинации архитектур ResNet и спектральных данных для максимальной точности классификации по специфической медицинской задаче – классификации речи пациентов после гемиглоссэктомии.

2. Сравняется эффективность моделей отдельно для пациентов с реконструкцией языка и без неё, что ранее не анализировалось в литературе.

3. Оценивается влияние типа хирургического вмешательства на точность классификации, что позволяет разработать персонализированные подходы к диагностике и реабилитации.



Для достижения поставленной цели необходимо решить следующие задачи:

- разработать программное обеспечение для преобразования аудиозаписей в спектральные данные: спектрограммы, мел-спектрограммы и MFCC;
- обучить различные архитектуры нейронных сетей ResNet для анализа спектральных данных и определения состояния речи у пациентов;
- выполнить перекрестную проверку моделей и оценить результаты работы нейронных сетей с использованием различных форматов спектральных данных;
- сравнить точность классификаторов с использованием различных форматов данных.

1 Связанные исследования

Рак языка – одно из наиболее распространённых онкологических заболеваний полости рта, характеризующееся злокачественным поражением тканей языка. Согласно данным ВОЗ и современным исследованиям [1, 2], заболеваемость раком языка неуклонно растёт. Лечение рака языка часто требует хирургического вмешательства, одним из наиболее распространённых методов которого является гемиглоссектомия – операция по частичному или полному удалению половины языка. Гемиглоссектомия может проводиться как с реконструкцией удалённой части (например, с использованием лоскутов тканей из других областей тела), так и без неё. Выбор метода зависит от стадии заболевания, локализации опухоли и индивидуальных особенностей пациента.

После операции пациенты сталкиваются с серьёзными нарушениями речевых, жевательных и глотательных функций, что существенно снижает качество их жизни. Реабилитация после гемиглоссектомии может занимать от 3 до 12 месяцев, в течение которых пациенты проходят курсы логопедической терапии, направленной на восстановление артикуляции и разборчивости речи. Однако в клинической практике оценка эффективности реабилитации зачастую основывается на субъективных методах, что затрудняет объективный мониторинг прогресса.

На сегодняшний день отсутствуют исследования, посвящённые систематическому анализу речи пациентов после гемиглоссектомии с использованием спектральных данных (мел-спектрограммы, MFCC, спектрограммы) и нейронных сетей (в частности, архитектур ResNet). В существующих работах в основном изучают общие метрики разборчивости речи [8, 9] или анализ аудиосигналов в других медицинских контекстах, таких как диагностика сердечно-сосудистых заболеваний или распознавание эмоций. Однако специфика речи пациентов после гемиглоссектомии, особенно с учётом типа хирургического вмешательства (с реконструкцией или без неё), ранее не рассматривалась в литературе. Это создаёт пробел в исследованиях, который данная работа призвана восполнить.

Использование искусственных нейронных сетей в медицине существенно повлияло на процессы лечения и диагностики заболеваний. Внедрение машинного обучения и нейронных сетей позволило расширить и

улучшить возможности традиционных подходов, позволяя упростить процесс обработки больших объемов данных и выявить скрытые закономерности, что в дальнейшем будет способствовать разработке наиболее результативных методов лечения и диагностики. В аналитическом обзоре [10] демонстрируется разнообразие применений нейронных сетей в медицине, от автоматизации рутинных задач до сложных процессов анализа медицинских изображений и геномных данных, что, в свою очередь, повышает качество и доступность медицинской помощи.

Современные исследования демонстрируют высокую точность работы нейронных сетей в анализе медицинских изображений. Например, в [11] архитектура ResNet-18 показала точность до 91,93% при классификации формы языка в традиционной китайской медицине, а в исследовании [12] ResNet-101 достигла точности 96,32% в диагностике COVID-19 и пневмонии на рентгеновских снимках грудной клетки. В классификации ультразвуковых изображений опухолей яичников [13] модель ResNeXt-50 продемонстрировала точность 95,2% на сегментированных данных, где авторы подчеркнули важность предобработки изображений. Для диагностики глаукомы [14] ResNet-50 достигла точности 99,86% на наборе данных ACRIMA, а в исследовании [15] по диагностике остеоартрита коленного сустава ResNet-18 с трансферным обучением показала точность 90%, что подтверждает преимущества использования предобученных моделей. В работе [16] при анализе цифровых маммограмм на наборе данных EMBED ResNet-152 продемонстрировала точность 90,4%, что указывает на то, что глубокие архитектуры хорошо справляются с задачей диагностики рака молочной железы.

Далее, одним из ярких примеров является интеграция логических нейронных сетей (LAN) и байесовских сетей для учета неопределенностей в медицинских данных, представленная в исследовании [17]. Модель Ballon, реализованная в TensorFlow, показала точность 69%, что подчеркнуло потенциал модели в медицинском моделировании, несмотря на несколько более низкий результат по сравнению с традиционными байесовскими нейронными сетями. В работе [18] авторы исследуют потенциал квантовых вычислений в сочетании с искусственными нейронными сетями для создания моделей, способных имитировать неврологические расстройства. В [19] авторы изучают возможности искусственного интеллекта для повышения эффективности и скорости медицинских процессов, уделяя особое внимание таким технологиям, как сверточные нейронные сети и передача обучения. Основное внимание уделяется объяснению основных концепций машинного обучения, искусственных нейронных сетей и глубокого обучения, а также иллюстрации их практического применения в медицине. В исследовании [20] представлен метод ECRCNN (*Extraction of Classification Rules from Convolutional Neural Network*; метод, направленный на извлечение правил классификации из обученной сверточной нейронной сети), направленный на диагностику сердечных заболеваний путем извлечения понятных



правил из сверточных нейронных сетей. Этот алгоритм генерирует ценную информацию о факторах, способствующих развитию сердечно-сосудистых заболеваний, и демонстрирует высокую точность прогнозирования сердечно-сосудистых заболеваний, что делает его полезным инструментом для медицинских работников.

Результаты исследования [21] подтверждают значительный потенциал нейронных сетей в улучшении диагностики, прогнозировании заболеваний и разработке индивидуальных подходов к лечению. В [22] рассматривается возможность использования нейронных сетей для обработки аудиозаписей, зашумленных белым шумом. Результаты исследования показали, что спектрограммы заметно изменяются при воздействии белого шума, что позволяет эффективно обучить нейронную сеть распознавать уровни шума. В работе [23] авторы использовали метод «звуковой дактилоскопии», преобразуя аудиопоток в спектрограммы для последующего анализа с использованием сверточных нейронных сетей (CNN). Результаты показали, что CNN демонстрирует разную степень успеха в распознавании эмоций, что подчеркивает сложность задачи и необходимость дальнейших исследований. Завершая обзор текущих исследований, в работе [24] предложен метод распознавания кашля с использованием CNN и анализа спектрограмм. На базе данных из 2437 аудиофайлов с аугментацией изображений достигнута точность распознавания 70,86%, что подтверждает потенциал метода для телемедицинских приложений.

Рассмотренные работы демонстрируют широкие возможности использования нейронных сетей как в медицине, так и при анализе аудиоданных. В области медицины нейронные сети помогают автоматизировать рутинные задачи, повысить точность диагностики и разработать новые методы лечения. Отдельные исследования показывают, что использование квантовых вычислений и глубоких нейронных сетей может значительно повысить эффективность моделирования неврологических расстройств, а методы поддающегося объяснению искусственного интеллекта способствуют повышению прозрачности и надежности медицинских решений.

Таким образом, нейронные сети и методы глубокого обучения оказывают значительное влияние на различные области науки и техники, предоставляя новые инструменты и методы анализа данных. В медицине это приводит к улучшению диагностики и индивидуализации лечения, а в анализе аудиоданных – к разработке новых подходов к обработке и интерпретации аудиосигналов. Эти достижения подчеркивают важность дальнейших исследований и разработок в области применения нейронных сетей, которые обещают значительные прорывы в повышении качества жизни и безопасности людей.

2 Подготовка данных и описание нейронной сети

2.1 Обучающий набор данных: параметры и преобразование

Набор данных включает аудиозаписи пациентов с раком языка после гемиглоссэктомии с реконструкцией и

без реконструкции. Записи были сделаны до хирургического лечения, после хирургического лечения и в процессе речевой реабилитации. Данные были собраны онкологами НМИЦ онкологии им. Н.Н. Блохина. Одна запись представляет собой произношение пациентом одного слога из списка слогов, которые записываются в процессе речевой реабилитации для оценки динамики восстановления речи.

Первая часть набора данных включает данные о 14 пациентах, перенесших гемиглоссэктомию с реконструкцией. Записи были сделаны до операции, сразу после нее и в конце трехмесячного курса реабилитации. На каждом этапе у каждого пациента проводится как минимум один сеанс записи, состоящий из записи 30 слогов. Однако у некоторых пациентов было несколько сеансов записи в разные дни, что увеличило общее количество записей. Таким образом, общее количество записей составило 1410. После удаления пустых и посторонних аудиозаписей (24 записи) для анализа осталось 1386 записей.

Вторая часть набора данных содержит данные о 15 пациентах, которым была выполнена гемиглоссэктомию без реконструкции. Аналогично первой части, у некоторых пациентов было несколько сеансов записи, что также увеличило общее количество записей до 1410. После удаления пустых и посторонних аудиозаписей из записей (25 записей) для анализа осталось 1385 записей.

Для дальнейшего анализа обе части данных были объединены в следующие группы:

- «до операции» – 974 аудиозаписи;
- «после операции» – 846 аудиозаписей;
- «после реабилитации» – 951 аудиозапись.

Аудиофайлы были преобразованы в спектрограммы, мел-спектрограммы и мел-частотные кепстральные коэффициенты (MFCC) и сохранены в виде изображений с разрешением 1560×1500 пикселей для сохранения максимального количества деталей из спектрограмм, что важно для точного анализа и классификации аудио данных, так как мелкие детали могут содержать важную информацию о характеристиках речи. Однако из-за ограничений вычислительных ресурсов, в последующем, разрешение изображений было уменьшено для успешного обучения модели. Пример изображений в виде спектральных данных изображен на рисунке 1.

В отличие от предыдущих исследований [8, 9], где анализировались общие метрики разборчивости речи, в данной работе впервые учитывается тип хирургического вмешательства (с реконструкцией и без неё) как ключевой фактор, влияющий на точность классификации.

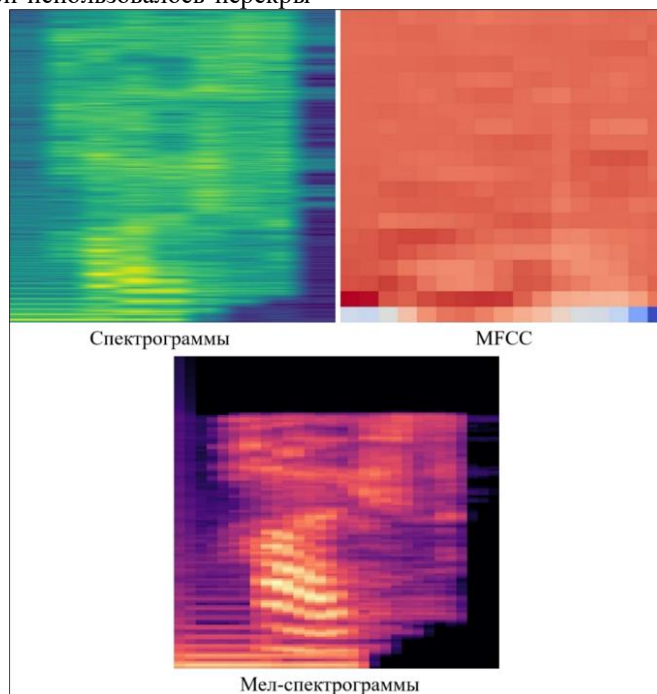
При анализе аудиосигналов для создания спектрограмм использовалось кратковременное преобразование Фурье (STFT, Short-Time Fourier Transform) с размером окна 1024 отсчёта. Этот метод позволяет разделить сигнал на временные сегменты и применить преобразование Фурье к каждому из них, обеспечивая баланс между временным и частотным разрешением.

Ось абсцисс спектрограммы соответствует временной шкале, а ось ординат — частотной шкале. Размер окна



(1024 отсчёта) определяет частотное разрешение, что позволяет выделять важные частотные характеристики сигнала, такие как форманты, критичные для анализа речи. Временное разрешение зависит от шага между окнами, который определяется перекрытием. Для улучшения временного разрешения и уменьшения артефактов на границах окон использовалось перекры-

тие в 50%, то есть каждое новое окно начинается с середины предыдущего. При размере окна 1024 отсчёта шаг между окнами составляет 512 отсчётов. Это позволяет получить более плавные и детализированные спектрограммы, что критично для точного анализа речевых сигналов.



Р и с. 1. Пример входных данных: а) спектрограмма; б) MFCC; в) мел-спектрограмма. Слог [к'ам] (по-русски КАМ)
F i g. 1. Example of input data: a) spectrogram; b) MFCC; c) mel-spectrogram. Syllable [k'am] (in Russian KAM)

Источник: здесь и далее в статье все таблицы и рисунки составлены авторами.
Source: Hereinafter in this article all tables and figures were made by the authors.

Для минимизации растекания спектра (спектрального размытия) использовалась оконная функция Хэннинга, которая снижает уровень боковых лепестков и искажения спектрального представления сигнала.

Разбиение данных на обучающий, проверочный и тестовый наборы выполнялось случайным образом. Это обеспечивало равномерное распределение данных, что важно для обобщающей способности модели. Разбиение проводилось независимо от конкретных пациентов и произносимых слогов, чтобы избежать смещения модели в сторону индивидуальных особенностей речи или звуковых паттернов.

2.2 Архитектура нейронной сети

Архитектура ResNet была выбрана для анализа речи и аудиосигналов на основе спектральных данных благодаря ее способности противодействовать градиентному затуханию, использованию пропускных слоев и успешному применению в анализе речи. Семейство ResNet включает в себя различные модели, которые отличаются по глубине и количеству параметров, включая ResNet-18, ResNet-34, ResNet-50, ResNet-101 и ResNet-152.

Аналогичный эксперимент с mel-спектрограммами также был проведен на архитектуре нейронной сети ResNet [25]. В результате однократного тестирования

модель показала точность 0,6875. При проведении 10-кратной перекрестной проверки средняя точность составила 0,6756. В рамках эксперимента планируется обучить нейронную сеть как на полном наборе данных, так и с разделением по видам хирургического вмешательства (гемиглоссэктомия с реконструкцией и гемиглоссэктомия без реконструкции).

3 Обучение и результаты

Обучение моделей на основе архитектур ResNet. Данные пациентов были разделены на три класса: «До операции», «После операции» и «После реабилитации». Изображения для обучения и тестирования модели были получены из аудиозаписей, соответствующих классам «До операции», «После операции» и «Во время реабилитации».

Для обучения модели данные были разделены на обучающий набор (70% от общего количества изображений), проверочный набор (15% изображений) и тестовый набор (15% изображений).

В процессе обучения использовались изображения MFCC размером 390×375 пикселей, и обучение проводилось в течение 30 периодов. Скорость обучения (lr) была установлена равной 0,0001. В ходе экспериментов изменялся параметр dropout, который



применялся на полносвязном слое перед финальным классификационным слоем. Dropout – это метод регуляризации, используемый для предотвращения переобучения нейронных сетей путём временного исключения (зануления) случайного подмножества нейронов на каждом этапе обучения с заданной вероятностью. Это помогает сделать модель более стабильной, так как сеть учится работать с различными комбинациями входных данных. Параметр dropout изменялся в пределах от 0.2 до 0.5 с шагом 0.05. Наиболее подходящим вариантом, обеспечивающим наилучшую точность классификации, оказался Dropout (0.25).

Было проведено сравнение различных архитектур ResNet (18, 34, 50, 101, 152) с использованием изображений разного размера: 260×250 , 390×375 , 520×500 . Наилучшие результаты представлены в Таблице 1: ResNet-18 с размером изображения 520×500 пикселей, ResNet-34 с размером изображения 390×375 пикселей, ResNet-152 с размерами изображения 390×375 и 520×500 пикселей.

Затем было проведено обучение на спектрограммах. Использовались изображения размером 390×375 пикселей, и обучение проводилось в течение 30 периодов. Скорость обучения (lr) была установлена равной 0,0001. Во время обучения параметр dropout был изменен, наиболее подходящими значениями были 0,3 и 0,5.

Поскольку архитектуры ResNet-50 и 101 показали худшие результаты на MFCC, чем другие, позже, на спектрограммах, было решено сосредоточиться на архитектурах ResNet-18, 34 и 152. Поэтому было проведено дальнейшее сравнение различных комбинаций архитектур (ResNet-18, 34, 152) и изображений разного размера (260×250 , 390×375 , 520×500) с наилучшими параметрами dropout (0,3, 0,5). Наилучшие результаты показаны в Таблице 2: ResNet-18 с dropout (0,3) на изображении 390×375 , ResNet-34 с dropout (0,3) на изображении 260×250 , ResNet-152 с dropout (0,5) на изображениях 390×375 и 520×500 .

Заключительным этапом было обучение на мел-спектрограммах со скоростью обучения 0,0001 с использованием изображений размером 390×375 пикселей в течение 30 периодов. Во время обучения параметр dropout изменялся, чтобы предотвратить переобучение. Наилучшие результаты были достигнуты при dropout 0,25, 0,45 и 0,5. На основе этих результатов были проведены сравнения между архитектурами ResNet (18, 34, 152). Наилучшими комбинациями были ResNet-18 с dropout (0,45), ResNet-34 с dropout (0,45) и ResNet-152 с dropout (0,5). Эти конфигурации были протестированы на изображениях разного размера. Наилучшие результаты показаны в Таблице 3: ResNet-34 с dropout (0,45) на изображениях с разрешением 260×250 и 390×375 пикселей и ResNet-152 с dropout (0,5) на изображениях с разрешением 390×375 и 520×500 пикселей.

Т а б л и ц а 1. Результаты для MFCC на полном наборе данных

Table 1. Results for MFCC for the full dataset

Epochs = 30, lr = 0.0001	Точность (Accuracy)
ResNet-152, dropout (0.25), изображение 390×375 пикселей	0.6538
ResNet-152, dropout (0.25), изображение 520×500 пикселей	0.6442
ResNet-18, dropout (0.25), изображение 520×500 пикселей	0.6322
ResNet-34, dropout (0.25), изображение 390×375 пикселей	0.6322

Т а б л и ц а 2. Результаты для спектрограмм на полном наборе данных

Table 2. Results for spectrograms for the full dataset

Epochs = 30, lr = 0.0001	Точность (Accuracy)
ResNet-152, dropout (0.5), изображение 520×500 пикселей	0.6418
ResNet-152, dropout (0.5), изображение 390×375 пикселей	0.6250
ResNet-18, dropout (0.3), изображение 390×375 пикселей	0.6226
ResNet-34, dropout (0.3), изображение 260×250 пикселей	0.6154

Т а б л и ц а 3. Результаты для мел-спектрограмм на полном наборе данных

Table 3. Results for mel-spectrograms for the full dataset

Epochs = 30, lr = 0.0001	Точность (Accuracy)
ResNet-152, dropout (0.5), изображение 390×375 пикселей	0.6875
ResNet-152, dropout (0.5), изображение 520×500 пикселей	0.6683
ResNet-34, dropout (0.45), изображение 260×250 пикселей	0.6659
ResNet-34, dropout (0.45), изображение 390×375 пикселей	0.6538



Затем отобранные ранее лучшие модели нейронной сети для каждого типа входных данных (мел-спектрограммы, MFCC и спектрограммы) были дополнительно обучены на разделённом наборе данных – отдельно для пациентов после гемиглоссектомии с реконструкцией и без неё. Такой подход обоснован следующим:

1. Предварительный отбор лучших конфигураций (архитектур ResNet, параметров dropout и размеров изображений) на полном наборе данных позволил исключить заведомо менее эффективные комбинации, что сократило вычислительные затраты;
2. Разделение данных по типу хирургического вмешательства (с реконструкцией и без неё) обусловлено медицинскими особенностями: речевые паттерны пациентов после реконструкции языка могут значительно отличаться от таковых у пациентов без реконструкции. Это подтверждается клиническими наблюдениями и требует отдельного анализа;
3. Возможность улучшения результатов: хотя некоторые комбинации могли показать себя лучше на отдельных классах, предварительный отбор на полном наборе данных позволил выделить наиболее перспек-

тивные архитектуры и параметры, которые затем были оптимизированы для каждого подмножества данных.

Результаты обучения на разделённых данных представлены в Таблицах 4–6. В столбцах Δ показана разница между точностью на разделённых наборах данных и точностью на полном наборе данных. Положительные значения указывают на улучшение точности, отрицательные – на ухудшение.

Были получены следующие результаты:

- 1) Для полного набора данных наилучший результат точности составляет 0,6875 для мел-спектрограмм в модели ResNet-152, параметр dropout (0,5) и размер изображения 390×375 пикселей;
- 2) Для разделённого набора данных, для части без реконструкции, наилучший результат точности составляет 0,7115 для мел-спектрограмм в модели ResNet-152, параметр dropout (0,5) и размер изображения 520×500 пикселей;
- 3) Для разделённого набора данных, для части с реконструкцией, наилучший результат точности составляет 0,7692 для мел-спектрограмм в модели ResNet-34, параметр dropout (0,45) и размер изображения 260×250 пикселей.

Т а б л и ц а 4. Результаты для MFCC на разделённом наборе данных

Table 4. Results for MFCC for the divided dataset

Epochs = 30, lr = 0.0001	Точность (Accuracy) для части с реконструкцией	Δ (с реконструкцией)	Точность (Accuracy) для части без реконструкции	Δ (без реконструкции)
ResNet-152, dropout (0.25), изображение 390×375 пикселей	0.7163	+0.0625	0.5913	-0.0625
ResNet-152, dropout (0.25), изображение 520×500 пикселей	0.7163	+0.0721	0.5577	-0.0865
ResNet-18, dropout (0.25), изображение 520×500 пикселей	0.7163	+0.0841	0.5817	-0.0505
ResNet-34, dropout (0.25), изображение 390×375 пикселей	0.6923	+0.0601	0.5577	-0.0745

Т а б л и ц а 5. Результаты для спектрограмм на разделённом наборе данных

Table 5. Results for spectrograms for the divided dataset

Epochs = 30, lr = 0.0001	Точность (Accuracy) для части с реконструкцией	Δ (с реконструкцией)	Точность (Accuracy) для части без реконструкции	Δ (без реконструкции)
ResNet-152, dropout (0.5), изображение 520×500 пикселей	0.7212	+0.0794	0.6106	-0.0312
ResNet-152, dropout (0.5), изображение 390×375 пикселей	0.7115	+0.0865	0.6058	-0.0192
ResNet-18, dropout (0.3), изображение 390×375 пикселей	0.6346	+0.0120	0.5337	-0.0889
ResNet-34, dropout (0.3), изображение 260×250 пикселей	0.6490	+0.0336	0.6154	+0.0000

Т а б л и ц а 6. Результаты для мел-спектрограмм на разделённом наборе данных

Table 6. Results for mel-spectrograms for the divided dataset

Epochs = 30, lr = 0.0001	Точность (Accuracy) для части с реконструкцией	Δ (с реконструкцией)	Точность (Accuracy) для части без реконструкции	Δ (без реконструкции)
ResNet-152, dropout (0.5), изображение 390×375 пикселей	0.6875	+0.0000	0.6979	+0.0104
ResNet-152, dropout (0.5), изображение 520×500 пикселей	0.7019	+0.0336	0.7115	+0.0432
ResNet-34, dropout (0.45), изображение 260×250 пикселей	0.7692	+0.1033	0.6346	-0.0313
ResNet-34, dropout (0.45), изображение 390×375 пикселей	0.6346	-0.0192	0.6587	+0.0049



Разделение набора данных на части с реконструкцией и без неё демонстрирует различное влияние на точность классификации в зависимости от типа спектральных данных. Для MFCC разделение данных позволяет значительно повысить точность на 6,97% для пациентов с реконструкцией, но снижает её на 6,85% для пациентов без реконструкции. При использовании спектрограмм точность увеличивается на 5,29% для пациентов с реконструкцией, но снижается на 3,48% для пациентов без реконструкции. Мел-спектрограммы показывают наименьшее изменение точности: увеличение на 2,94% для пациентов с реконструкцией и небольшое увеличение на 0,68% для пациентов без реконструкции. В среднем, разделение данных позволяет повысить точность на 5,07% для пациентов с реконструкцией, но снижает точность на 3,22% для пациентов без реконструкции.

4 Обсуждение

Полученные результаты демонстрируют, что мел-спектрограммы в сочетании с архитектурой ResNet-34 обеспечивают наивысшую точность классификации (до 0,7692 для пациентов после гемиглоссэктомии с реконструкцией), что может быть объяснено их способностью улавливать нелинейные изменения частотных характеристик, специфичных для речи после реконструкции языка. Этот вывод особенно важен, так как в существующих исследованиях [21, 22] ResNet применялся для анализа зашумленных аудиосигналов или эмоциональной окраски речи, но не для оценки речи пациентов после онкологических операций.

Предложенная трёхклассовая модель классификации («до операции», «после операции», «во время реабилитации») позволяет логопедам объективно и качественно оценивать динамику восстановления речи пациентов. Например:

- сравнение с базовым состоянием («до операции»). Логопед может загрузить аудиозапись пациента в систему и получить классификацию, которая покажет, насколько текущая речь пациента приближена к его речи до операции. Если система классифицирует запись как «во время реабилитации», но близкую по метрикам к классу «до операции», это свидетельствует о значительном прогрессе;
- выявление «застойных» состояний. Если в течение нескольких недель записи пациента продолжают классифицироваться как «после операции» без перехода в класс «во время реабилитации», это сигнализирует о необходимости коррекции реабилитационной программы (например, изменения упражнений или интенсивности занятий).

Разделение данных по типу хирургического вмешательства (с реконструкцией и без неё) значительно улучшило точность модели (например, точность для пациентов без реконструкции выросла с 0,5913 до 0,7115 при использовании мел-спектрограмм и ResNet-152). Это подтверждает гипотезу о том, что речевые паттерны у пациентов после гемиглоссэктомии существенно зависят от типа операции, и учет этого фактора позволяет оптимизировать процесс классификации.

Полученные результаты демонстрируют, что предложенный метод классификации речи пациентов позволяет объективно оценивать динамику восстановления речевых функций у пациентов после гемиглоссэктомии. Метод обеспечивает точность классификации до 0,77 (для пациентов с реконструкцией языка) при использовании мел-спектрограмм и архитектуры ResNet-34.

Благодаря данной классификации речи на три класса, логопеды получают инструмент для объективного, быстрого и стандартизированного мониторинга прогресса пациентов. Это позволяет:

1. Снизить субъективность оценки за счёт количественных метрик, основанных на спектральном анализе;
2. Своевременно корректировать программы реабилитации на основе данных системы, а не только субъективных наблюдений;
3. Повысить эффективность работы медицинского персонала, автоматизируя рутинный анализ записей и фокусируясь на пациентах с низкой динамикой.

Таким образом, предложенный метод не только улучшает точность диагностики речевых нарушений, но и способствует персонализации реабилитационных программ, что в конечном итоге повышает качество жизни пациентов после гемиглоссэктомии.

Сравнение с аналогичными исследованиями показывает, что в данной работе впервые предложен подход к выбору подходящих конфигураций ResNet для анализа речи пациентов после гемиглоссэктомии. Разработанный метод не только подтверждает эффективность использования ResNet для анализа речи, но и предлагает практический инструмент для автоматизации оценки прогресса пациентов после операции. Это открывает перспективы для упрощения процесса диагностики и мониторинга эффективности реабилитации в онкологии, способствуя снижению нагрузки на медицинский персонал и повышению эффективности реабилитационных программ.

Для дальнейшего улучшения результатов по точности анализа речевых и аудиосигналов планируется:

- реализовать ансамблевый подход, объединяющий результаты анализа различных спектральных представлений одного и того же аудиосигнала. Конкретно, для каждого аудиофрагмента будут генерироваться три типа спектральных данных: стандартная спектрограмма, мел-спектрограмма и MFCC. Каждый вид представления будет обрабатываться отдельной специализированной моделью нейронной сети, обученной на соответствующем типе спектрограмм. Полученные вероятности классификации от всех трёх моделей будут агрегироваться (например, с помощью усреднения или взвешенного голосования), что позволит снизить влияние случайных ошибок и повысить надёжность итогового результата;
- выбирать и анализировать только значимые части спектрограмм: обрезать спектрограммы, анализируя только значимые частоты (например, исключить частоты выше или ниже определенного порога), а также применять частотные фильтры и устранение шума. Такой подход может повысить качество вводимых данных и, следовательно, точность моделей.



Заключение

В рамках данного исследования проведено сравнение эффективности различных архитектур ResNet и типов спектральных данных (мел-спектрограммы, MFCC, спектрограммы) для классификации речи пациентов после гемиглоссэктомии. Полученные результаты позволяют сделать следующие ключевые выводы:

1. Мел-спектрограммы в сочетании с ResNet-34 обеспечивают наивысшую точность классификации (до 0,7692) для пациентов после реконструкции языка, что может быть связано с их способностью улавливать тонкие изменения частотных характеристик, специфичных для этой группы пациентов;
2. Разделение данных по типу операции (с реконструкцией и без неё) значительно улучшает точность модели, подчеркивая важность учета медицинских особенностей пациентов при разработке алгоритмов анализа речи;
3. Результаты в случае без реконструкции хуже, чем при ее проведении. Это может быть объяснено медицинскими причинами, по которым принято решение о проведении восстановительной процедуры. Поскольку операция с реконструкцией применяется в тех случаях, когда опухоль поражает значительную часть языка (что существенно влияет на речь), данные с реконструкцией могут содержать более разнообразные и сложные структуры. Также было высказано предположение, подкрепленное мнением логопедов, что реконструкция оказывает положительное влияние на речевую реабилитацию пациентов;
4. Оптимальные конфигурации ResNet (например, ResNet-152 с мел-спектрограммами для пациентов без реконструкции) могут быть использованы для создания клинических инструментов автоматизированной диагностики и мониторинга речевой реабилитации.

В данном исследовании разработан метод классификации речи пациентов после гемиглоссэктомии на три класса: «до операции», «после операции» и «во время реабилитации». Этот метод позволяет объективно оценивать изменения речи пациентов на основе спектрального анализа с использованием нейронных сетей ResNet. Благодаря классификации аудиозаписей на три группы удастся упростить процесс отслеживания динамики восстановления речи, сравнивая текущие записи с базовыми (до операции). Это не только позволяет объективно оценивать изменения речевых функций, но и снижает нагрузку на медицинский персонал за счёт автоматизации анализа аудиозаписей, выделяя случаи, требующие особого внимания. Таким образом, практическая значимость исследования заключается в возможности стандартизировать и упростить оценку динамики восстановления речи у пациентов после гемиглоссэктомии, повышая эффективность реабилитационных программ.

Внедрение предложенного метода в клиническую практику позволит автоматизировать оценку динамики восстановления речи у пациентов после гемиглоссэктомии. Это не заменяет диагностику онкологических заболеваний, а дополняет её, обеспечивая объективный и быстрый мониторинг речевых функций в процессе реабили-

тации. Таким образом, технология способствует повышению эффективности медицинского обслуживания за счёт своевременной коррекции реабилитационных программ и снижения нагрузки на специалистов.

Таким образом, в данном исследовании впервые продемонстрировано, что сочетание мел-спектрограмм с архитектурой ResNet-34 позволяет достичь точности 0,7692 в классификации речи пациентов после гемиглоссэктомии с реконструкцией языка. Это превосходит результаты, достигнутые при использовании MFCC и спектрограмм. Полученные данные подтверждают предположение о том, что мел-спектрограммы более эффективны для анализа нелинейных изменений частотных характеристик, характерных для речи пациентов после реконструкции языка.

Важным преимуществом разработанного метода является переход от субъективной оценки речи логопедом к объективному анализу на основе спектральных данных. Разделение данных по типу хирургического вмешательства позволяет значительно улучшить точность модели, что подчеркивает необходимость учета медицинских особенностей пациентов. Такой подход не заменяет субъективный опыт логопеда, а дополняет его количественными метриками, повышая надежность и воспроизводимость оценки динамики восстановления речи.

Результаты исследования открывают новые перспективы для клинического применения нейронных сетей в онкологии, включая автоматизацию диагностики речевых нарушений и персонализацию программ реабилитации. Дальнейшие исследования могут быть направлены на интеграцию предложенных методов в клиническую практику, а также на анализ других типов спектральных данных (например, комбинаций спектрограмм) для повышения точности классификации.



References

1. Ganina Ch., Kropotov M., Saprina O., Kosova E., Aketova T., Isaeva M., Gerasimov O. Quality of life of patients with tongue cancer after hemiglossectomy. *Opukholi golovy i shei* = Head and Neck Tumors. 2023;13(3):32-42. <https://doi.org/10.17650/2222-1468-2023-13-3-32-42>
2. Dil O.S., Kulbakin D.E., Choinzonov E.L., et al. Pregnancy and oral cancer. *Issues of oncology*. 2023;69(6):1014-1020. (In Russ., abstract in Eng.) <https://doi.org/10.37469/0507-3758-2023-69-6-1014-1020>
3. Sdvizhkov A.M., Kozhanov L.G., Shatskaia N.Kh., Belov E.N. Analysis of the causes of cancer negligence and low survival in the patients with malignant neoplasms of ENT and oral cavity in the city of Moscow. *Russian Bulletin of Otorhinolaryngology*. 2014;(1):20-23. (In Russ., abstract in Eng.)
4. Dasgupta S., Samuel S. Assessment of Speech in Patients Undergoing Hemiglossectomy with Pri-mary Closure and Radiotherapy – A Prospective Study. *Annals of Maxillofacial Surgery*. 2022;12:157. https://doi.org/10.4103/ams.ams_76_22
5. Monish V., Jaya V., Rani R. Effect of tongue reconstruction following hemiglossectomy on articulation and speech intelligibility. *IP Indian Journal of Anatomy and Surgery of Head, Neck and Brain*. 2022;7:103-106. <https://doi.org/10.18231/j.ijashnb.2021.027>
6. Girish K.S., Thejaswini M.R., Surabhi M.Y. Functional Outcomes in a Client with Hemiglossectomy and Segmental Mandibulectomy: A Case Report. *Research & Reviews Journal of Surgery*. 2022;11(2):1-6.
7. Dasgupta S., Samuel S. Assessment of Dysphagia after Hemiglossectomy and Radiotherapy: A Prospective Study. *Journal of Health and Allied Sciences NU*. 2022;13(2):257-261. <https://doi.org/10.1055/s-0042-1755451>
8. Novokhrestova D., Hodashinsky I., Kostyuchenko E., Sarin K., Bardamova M. Fuzzy Classifier for Speech Assessment in Speech Rehabilitation. In: Prasanna S.R.M., Karpov A., Samudravijaya K., Agrawal S.S. (eds.) Speech and Computer. SPECOM 2022. *Lecture Notes in Computer Science*. Vol. 13721. Cham: Springer; 2022. p. 522-532. https://doi.org/10.1007/978-3-031-20980-2_44
9. Novokhrestova D., Kostyuchenko E., Borovskoy A. Classifier-Based Combined Measure of Syllable Pronunciation Similarity in Speech Rehabilitation. In: Kovalev S., Kotenko I., Sukhanov A. (eds.) Proceedings of the Seventh International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'23). IITI 2023. *Lecture Notes in Networks and Systems*. Vol. 777. Cham: Springer; 2023. p. 127-136. https://doi.org/10.1007/978-3-031-43792-2_13
10. Lysova M.E., Kuznetsov M.E. Neural networks in medicine. Automation using artificial intelligence. In: Achievements of science and technology-DNiT-11-2023: Collection of scientific articles based on the materials of the II All-Russian Scientific Conference, Krasnoyarsk. Krasnoyarsk: Krasnoyarsk Regional House of Science and Technology of the Russian Union of Scientific and Engineering Public Associations; 2023. Vol. 7. p. 581-586. (In Russ., abstract in Eng.) <https://doi.org/10.47813/dnit-II.2023.7.581-586>
11. Yi Zh., Lai X., Sun A., Fang S. Tongue feature recognition to monitor rehabilitation: deep neural network with visual attention mechanism. *Frontiers in Bioengineering and Biotechnology*. 2024;12:1392513. <https://doi.org/10.3389/fbioe.2024.1392513>
12. Yenikaya M., Kerse G., Oktaysoy O. Artificial intelligence in the healthcare sector: comparison of deep learning networks using chest X-ray images. *Frontiers in Public Health*. 2024;12:1386110. <https://doi.org/10.3389/fpubh.2024.1386110>
13. Wu M., Cui G., Lv Sh., Chen L., Tian Z., Yang M., Bai W. Deep convolutional neural networks for multiple histologic types of ovarian tumors classification in ultrasound images. *Frontiers in Oncology*. 2023;13:1154200. <https://doi.org/10.3389/fonc.2023.1154200>
14. Vijaya K.V., Sharma L. Multi-stage glaucoma classification using pre-trained convolutional neural networks and voting-based classifier fusion. *Frontiers in Physiology*. 2023;14:1175881. <https://doi.org/10.3389/fphys.2023.1175881>
15. Yeoh P., Lai Kh. W., Goh S.L., Hasikin Kh., Wu X., Li P. Transfer learning-assisted 3D deep learning models for knee osteoarthritis detection: Data from the osteoarthritis initiative. *Frontiers in Bioengineering and Biotechnology*. 2023;11:1164655. <https://doi.org/10.3389/fbioe.2023.1164655>
16. Hwang I., Trivedi H., Brown-Mulry B., Zhang L., Nalla V., Gastounioti A., Gichoya J., Seyyed-Kalantari L., Banerjee I., Woo M. Impact of multi-source data augmentation on performance of convolutional neural networks for abnormality classification in mammography. *Frontiers in Radiology*. 2023;3:1181190. <https://doi.org/10.3389/fradi.2023.1181190>
17. Diaz Ochoa, Juan G., Maier, Lukas, Csiszár, Orsolya. Bayesian logical neural networks for human-centered applications in medicine. *Frontiers in Bioinformatics*. 2023;3:1082941. <https://doi.org/10.3389/fbinf.2023.1082941>
18. Youvan D. Neurological Disorder Simulations: Leveraging Quantum Computing and Neural Networks. 2024. <https://doi.org/10.13140/RG.2.2.10851.98086>
19. Kufel J., Bargiel-Łączek K., Kocot Sz., Koźlik M., Bartnikowska W., Janik M., Czogalik Ł., Dudek P., Magiera M., Lis A., Paszkiewicz I., Nawrat Z., Cebula M., Gruszczynska K. What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine. *Diagnostics*. 2023;13:2582. <https://doi.org/10.3390/diagnostics13152582>
20. Chakraborty M. Rule extraction from convolutional neural networks for heart disease prediction. *Biomedical Engineering Letters*. 2024;14:649-661. <https://doi.org/10.1007/s13534-024-00358-3>



21. Joshi Sr., Kumar V., Venkataramanan V., C S K. A Review on Neural Networks and its Applications. *Journal of Computer Technology & Applications*. 2023;14(2):60-71. <https://doi.org/10.37591/jocta.v14i2.1062>
22. Volkov N.A., Ivanov A.V., Karpova N.E., Chuvakov A.V. *Issledovanie zashumlennoi audiozapisi rechi dlya obrabotki v avtomatizirovannoi sisteme* [Investigation of noisy audio recordings of speech for processing in an automated system]. *Bezopasnost' tsifrovyykh tekhnologii* = Digital Technology Security. 2022;(2):9-20. (In Russ., abstract in Eng.) <https://doi.org/10.17212/2782-2230-2022-2-9-20>
23. Semeniyuk V.V., Skladchikov M.V. Development of an algorithm for recognizing human emotions using a convolutional neural network using Python tools. *Engineering Bulletin of the Don*. 2023;(12):231-245. (In Russ., abstract in Eng.) EDN: TBESZT
24. Dvoynikova A.A. Cough recognition using spectrogram analysis. In: Bobtsov A.A. (ed.) *The Almanac of scientific works of young scientists of ITMO University*. St. Petersburg: ITMO National Research University; 2022. Vol. 2. p. 230-234. (In Russ., abstract in Eng.) EDN: TGGMEA
25. Novokhrestova D., Tomilina S., Laptev P., Kostyuchenko E. Classifier Based on Neural Networks to Determine the Patient's Speech State. In: Kovalev S., Kotenko I., Sukhanov A., Li Y., Li Y. (eds.) *Proceedings of the Eighth International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'24)*, Vol. 2. IITI 2024. *Lecture Notes in Networks and Systems*. Vol. 1210. Cham: Springer; 2024. p. 236-244. https://doi.org/10.1007/978-3-031-77411-9_22

Поступила 14.06.2025; одобрена после рецензирования 21.09.2025; принята к публикации 06.10.2025.

Submitted 14.06.2025; approved after reviewing 21.09.2025; accepted for publication 06.10.2025.

Об авторах:

Томила Светлана Дмитриевна, преподаватель кафедры комплексной информационной безопасности электронно-вычислительных систем, ФГАОУ ВО «Томский государственный университет систем управления и радиоэлектроники» (634050, Российская Федерация, г. Томск, пр. Ленина, д. 40), **ORCID:** <https://orcid.org/0009-0002-2526-6377>, tsd@fb.tusur.ru

Новохрестова Дарья Игоревна, доцент кафедры комплексной информационной безопасности электронно-вычислительных систем, ФГАОУ ВО «Томский государственный университет систем управления и радиоэлектроники» (634050, Российская Федерация, г. Томск, пр. Ленина, д. 40), кандидат технических наук, **ORCID:** <https://orcid.org/0000-0002-4931-1681>, ndi@fb.tusur.ru

Лаптев Павел Юрьевич, преподаватель кафедры безопасности информационных систем, ФГАОУ ВО «Томский государственный университет систем управления и радиоэлектроники» (634050, Российская Федерация, г. Томск, пр. Ленина, д. 40), **ORCID:** <https://orcid.org/0000-0002-5305-3312>, lpu@fb.tusur.ru

Костюченко Евгений Юрьевич, и.о. заведующего кафедрой безопасности информационных систем, ФГАОУ ВО «Томский государственный университет систем управления и радиоэлектроники» (634050, Российская Федерация, г. Томск, пр. Ленина, д. 40), кандидат технических наук, доцент, **ORCID:** <https://orcid.org/0000-0001-8000-2716>, key@fb.tusur.ru

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Svetlana D. Tomilina, Lecturer of the Chair of Complex Information Security of Computer Systems, Tomsk State University of Control Systems and Radioelectronics (40 Prospekt Lenina, Tomsk 634050, Russian Federation), **ORCID:** <https://orcid.org/0009-0002-2526-6377>, tsd@fb.tusur.ru

Dariya I. Novokhrestova, Associate Professor of the Chair of Complex Information Security of Computer Systems, Tomsk State University of Control Systems and Radioelectronics (40 Prospekt Lenina, Tomsk 634050, Russian Federation), Cand. Sci. (Tech.), **ORCID:** <https://orcid.org/0000-0002-4931-1681>, ndi@fb.tusur.ru

Pavel Yu. Laptev, Lecturer of the Chair of Information System Security, Tomsk State University of Control Systems and Radioelectronics (40 Prospekt Lenina, Tomsk 634050, Russian Federation), **ORCID:** <https://orcid.org/0000-0002-5305-3312>, lpu@fb.tusur.ru

Evgeny Yu. Kostyuchenko, Acting Head of the Chair of Information System Security, Tomsk State University of Control Systems and Radioelectronics (40 Prospekt Lenina, Tomsk 634050, Russian Federation), Cand. Sci. (Tech.), Associate Professor, **ORCID:** <https://orcid.org/0000-0001-8000-2716>, key@fb.tusur.ru

All authors have read and approved the final manuscript.