

ВИШНЯКОВ В.А., ХЭ Т.

## СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ ДЛЯ ИТ-ДИАГНОСТИКИ ЛЕГКИХ

Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь

Предметом исследований является использование технологии обработки голоса пациента в ИТ-медицине. Цель статьи – разработать нейронную сеть для диагностики заболеваний легких с помощью звукового анализа голоса пациента. Исследование включает в себя обучение нейронной сети, разработку мобильной программы для сбора звука пациента, извлечение звуковых характеристик на стороне сервера, диагностику звуковых данных с использованием обученной нейронной сети и возврат результатов диагностики в мобильную программу приложения. Представлена блок-схема обработки голоса от исходного сигнала до извлечения аудиофайла, в качестве примера приведено извлечение функций MFCC и FBank. Приведена структура сверточной нейронной сети (CNN), которая была обучена на стандартном наборе данных респираторных заболеваний. Приведен упрощенный процесс классификации звуков дыхания, необходимых для прогнозирования заболеваний легких. Для практической реализации использована в среде программирования Python сеть VGGish, которая имеет сетевые параметры, обученные с помощью набора данных. Эксперименты проведены на платформе Android service framework, которая разделена на две части: Android front-end и серверную. Интерфейсная часть реализует интерактивную функцию пользователя и отвечает за ввод аудиоданных. После загрузки аудио сервер выполнит предварительную обработку аудио, и вызовет CNN для классификации аудио, результаты возвращаются во внешний модуль на смартфоне. Лучшая точность модели достигла 83,6 %.

**Ключевые слова:** ИТ-медицина, анализ голоса, модель нейронной сети, распознавание заболевания легких

### Введение

Метод диагностики заболеваний по голосу пациента зависит от субъективного восприятия врача и его опыта в медицинской практике, который, к сожалению, не может быть воспроизведен быстро и эффективно. Более того, одно и то же заболевание может проявляться по-разному у разных пациентов. Например, индивидуальные различия между людьми, возрастные различия между людьми и время болезни могут привести к различным диагнозам, и эти обоснованные диагнозы, которые ставятся с помощью человека-диагноста, являются субъективными.

Однако акустическая диагностика также имеет несравнимые преимущества, такие как отсутствие повреждений тела пациента, дискомфорта, вызванного инвазивными исследованиями, а также простые, удобные и быстрые методы обнаружения. Эти причины делают очень актуальным использование звука пациента для ИТ-диагностики заболеваний. С развитием компьютерных технологий в медицинской сфере уже стало возможным измерять звуковые сигналы по объективным цифровым критериям. Появление все большего числа характеристик звукового сигнала также позволило диагностировать заболевания пациента по звуку его голоса.

В работе были предложены возможные количественные показатели для обоснованной голосовой диагностики [1]. Все большее число патологических исследований демонстрируют корреляцию между голосом и болезнью. В 2007 году Макс А. Литтл произвел повторную выборку образцов здорового голоса из базы данных KaupENTAX Model 4337 (MEEI) с частотой 25 кГц и образцов голоса с частотой 50 кГц до 44,1 кГц и достиг точности классификации 91,8 %, используя только два нелинейных признака [2]. В работе [3] обсуждается не только корреляция между голосом и патологией рака легких, но также корреляция между раком легких и его голосовыми характеристиками.

Цель статьи – разработать нейронную сеть для диагностики заболеваний легких с помощью звукового анализа голоса пациента. Исследование включает в себя обучение нейронной сети, разработку мобильной программы для сбора звука пациента, извлечение звуковых характеристик на стороне сервера, диагностику звуковых данных с использованием обученной нейронной сети и возврат результатов диагностики в мобильную программу приложения.

### Выделение признаков

Поскольку аудиоданные не имеют функций, которые легко наблюдать, таких как изображения

или текстовые данные, вопрос о том, какие функции можно использовать для лучшего представления информации, содержащейся в аудио, всегда был в центре внимания исследователей. В современных основных методах извлечения звуковых объектов, прежде чем извлекать объекты из аудиосигнала, сигнал сначала должен быть подвергнут многоступенчатой операции предварительной обработки, включая предварительное выделение, кадрирование и дискретное преобразование Фурье. На рисунке 1 представлена блок-схема обработки голоса от исходного сигнала до извлечения аудиофайла, в качестве примера приведено извлечение функций MFCC и FBank.

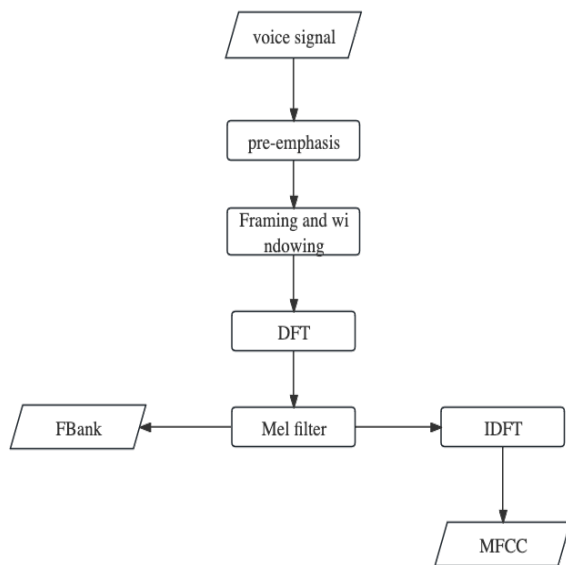


Рисунок 1. Процесс извлечения признаков из MFCC и FBank

## Структура сверточной нейронной сети

Сверточные нейронные сети (CNN) могут использовать ядра свертки на входном изображении для присвоения весов различным изучаемым параметрам путем вычисления и обучения различным частям изображения и извлечения их из них. Карта объектов использует описания объектов, чтобы позволить модели нейронной сети классифицировать изображения или идентифицировать различные объекты на изображении. По сравнению с другими моделями классификационных сетей сверточные нейронные сети требуют гораздо меньшего количества параметров, тем самым снижая нагрузку на компьютер. В сверточной нейронной сети ее нейронным узлам не нужно связывать значение каждого пикселя на изображении. Им часто требуется только связать информацию с частями изображения, а затем интегрировать информацию о каждой части изображения в общее изображение посредством перемещения ядра свертки. Информация. Характеристики подключений к локальной сети и совместное использование параметров ядра свертки в сверточной нейронной сети значительно уменьшают количество подключений между нейронными узлами в сверточной нейронной сети и значительно уменьшают проблему, вызванную существованием большого количества параметров в нейронной сети. После обучения сверточной нейронной сети пользователи анализируют и обнаруживают, что она также может извлекать эти признаки, указывая на то, что СНС позволяет компьютеру обучаться. научился правильно извлекать признаки для идентификации изображений.

Структура сети CNN показана на рисунке 2, она была обучена на стандартном наборе данных респираторных заболеваний [4].

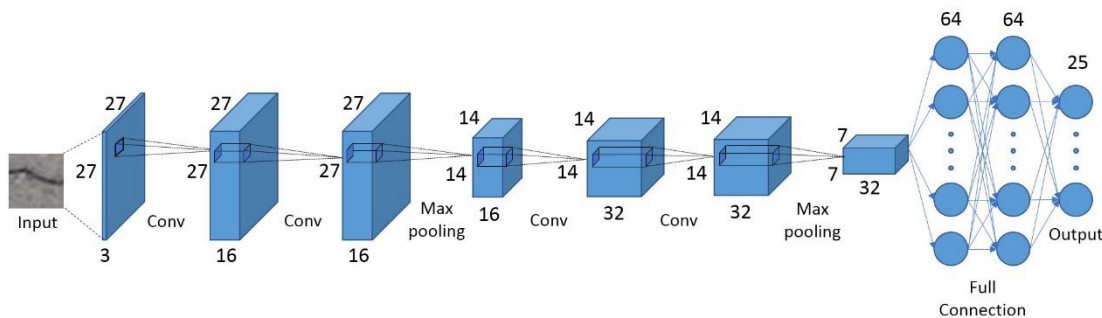


Рисунок 2. Полная схема сети CNN

С помощью описанного выше процесса модель может распознать особенности изображения. После свертки и объединения в пул конечный результат сглаживается и вводится в полностью подключенную нейронную сеть для классификации.

## Процесс диагностики заболеваний легких по звукам дыхания

После завершения обучения модели CNN [5] упрощенный процесс классификации звуков

дыхания [6], необходимых для прогнозирования заболеваний, включает шаги.

1. Выполнить предварительную обработку и процесс выделения признаков для звуков дыхания.
2. Выполнить оптимизацию признаков для извлеченных признаков.
3. Применить обученную CNN для классификации нового пациента с использованием модели VGGish (в среде Python).
4. Вывести результаты классификации.

Блок-схема прогнозирования заболевания легких по звуку дыхания показана на рисунке 3.

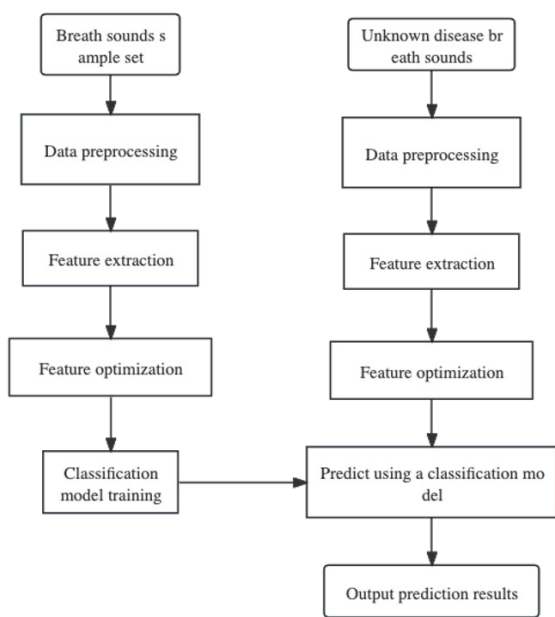


Рисунок 3. Блок-схема ИТ-диагностики заболеваний легких по звукам дыхания

### Сервис Android для идентификации заболеваний легких по звуку дыхания

Платформа Android service framework показана на рисунке 5. Система разделена на две части: Android front-end и серверную. Интерфейсная часть реализует интерактивную функцию пользователя и отвечает за ввод аудиоданных. После загрузки аудио сервер сгенерирует IP-адрес терминала пользователя. Папка используется для хранения аудио. После нажатия на интерфейс для диагностики сервер выполнит предварительную обработку аудио, такую как извлечение признаков, и вызовет сохраненную модель для классификации аудио. Наконец, результаты возвращаются во внешний модуль "получить результаты". Лучшая точность модели достигла 83,6 %.

Согласно структуре внутреннего уровня модельной сети, можно видеть, что VGGish – это форма функций logmel в сочетании с сетью, подобной VGG [7]. Ее преимущество заключается в том, что сеть VGGish имеет сетевые параметры, обученные с помощью большого объема данных, что улучшает способность модели к обобщению; если используется только оператор If, структура модели не загружает параметры предварительного обучения VGGish, улучшение производительности модели может быть неочевидным.

В качестве примера возьмем аудио спектрограмму размером  $1 \times 96 \times 64$ , ее преобразование размеров с помощью слоя свертки  $3 \times 3$  и слоя объединения  $2 \times 2$  Max модуля VGGish показано на рисунке 4.

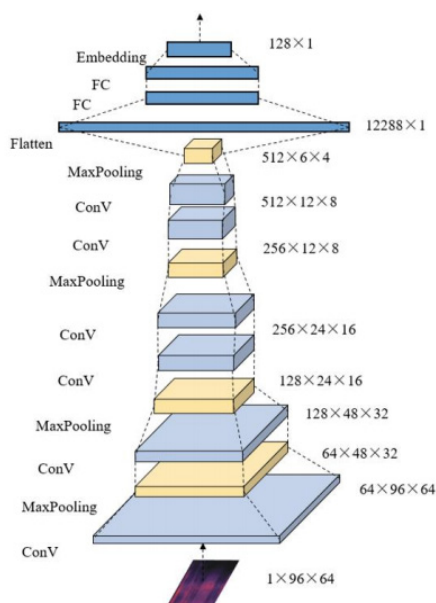


Рисунок 4. Внутренняя структура сети VG Dish

### Демонстрация использования приложения

1. Открыть приложение и войти в интерфейс домашней страницы, как показано на рисунке 6. Выбрать диагностику звуков дыхания, используя дыхание пациента или загружая аудиофайлы.

2. Нажать кнопку "использовать звуки", чтобы войти в интерфейс звуков дыхания, как показано на рисунке 7. Собирая звуки дыхания, необходимо убедиться, что вы находитесь в спокойной обстановке. Поднесите микрофон вашего мобильного телефона поближе к горлу, нажмите и удерживайте кнопку "Hold to speak", чтобы зафиксировать звуки дыхания. Время сбора по умолчанию составляет 20 секунд, а затем отпустите его. Нажмите и удерживайте кнопку "Говорить". В это время появится кнопка "Проанализировать". Нажмите кнопку "Проанализировать", чтобы загрузить аудиофайл на сервер и дождаться возврата результата.

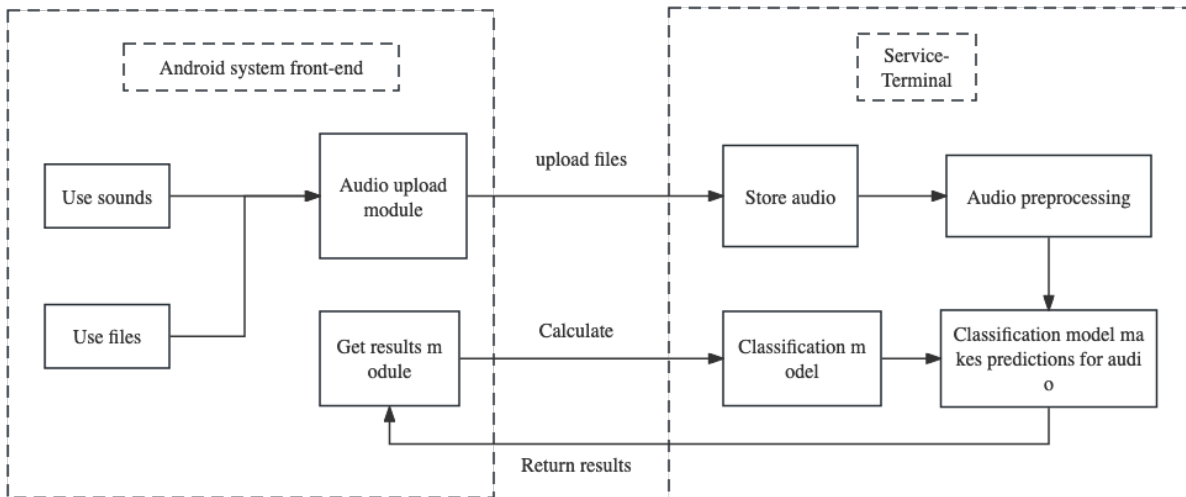


Рисунок 5. Структура сервисной системы Android



Рисунок 6. Домашняя страница приложения

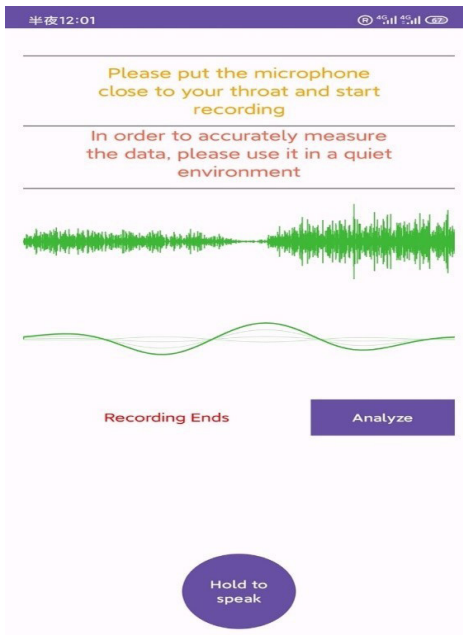


Рисунок 7. Использование интерфейса "Звук дыхания"

3. Возвращаемые результаты с использованием звуков дыхания показаны на рисунке 8. Темным шрифтом показан тип и достоверность прогнозируемого заболевания.

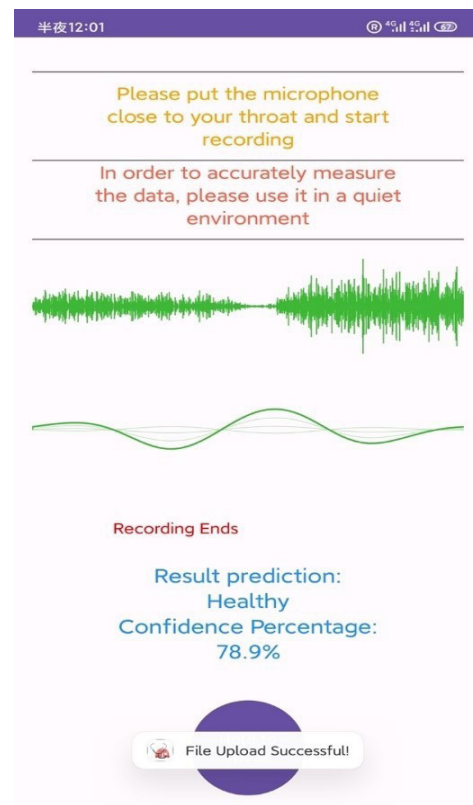


Рисунок 8. Результаты диагностики с использованием звуков дыхания.

## Заключение

1. Рассмотрены основные направления голосового анализа в ИТ-медицине. Разработана нейронная сеть для диагностики заболеваний легких

с помощью звукового анализа голоса пациента. Исследование включает в себя обучение нейронной сети, разработку мобильной программы для сбора звука пациента, извлечение звуковых характеристик на стороне сервера, диагностику звуковых данных с использованием обученной нейронной сети и возврат результатов диагностики в мобильную программу приложения.

2. Представлена блок-схема обработки голоса от исходного сигнала до извлечения аудиофайла, в качестве примера приведено извлечение функций MFCC и FBank. Приведена структура сверточной нейронной сети (CNN), которая была обучена на стандартном наборе данных респираторных заболеваний. Приведен упрощенный процесс

классификации звуков дыхания, необходимых для прогнозирования заболеваний легких.

3. Для практической реализации использована в среде программирования Python сеть VGGish, которая имеет сетевые параметры, обученные с помощью набора данных. Эксперименты проведены на платформе Android service framework, которая разделена на две части: Android front-end и серверную. Интерфейсная часть реализует интерактивную функцию пользователя и отвечает за ввод аудиоданных. Сервер выполнит предварительную обработку аудио, и вызывает CNN для классификации аудио, результаты возвращаются во внешний модуль на смартфоне. Лучшая точность модели достигла 83,6 %.

## REFERENCES

1. Hwang E.J., Park S., Jin K.-N. et al. Development and Validation of a Deep Learning based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. *Clinical Infectious Diseases*, 2019, Vol. 69, Issue 5, pp. 739-747.
2. Little M.A., McSharry, P.E., Roberts, S.J. et al. Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. *BioMed Eng OnLine* 6, 23 (2007). DOI: 10.1186/1475-925X-6-23
3. Das N., Topalovic M., Janssens W. Artificial intelligence in diagnosis of obstructive lung disease: current status and future potential. *Current Opinion in Pulmonary Medicine*, 2018, Volume 24, Issue 2, pp. 117–123.
4. Rocha B.M., Filos D., Mendes L. et al. A Respiratory Sound Database for the Development of Automated Classification. In: *Precision Medicine Powered by pHealth and Connected Health*, Singapore: Springer Singapore. 2018, pp. 33-37. (IFMBE Proceedings; vol. 66).
5. Amoh J., Odam K. Deep Neural Networks for Identifying Cough Sounds. *IEEE Transactions on Biomedical Circuits and Systems*. 2016, Vol. 10, Issue 5, pp. 1003-1011.
6. Aykanat M., Kılıç Ö., Kurt B., Saryal S. Classification of lung sounds using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, 2017, Vol. 2017, Issue 1, pp. 65.
7. The Robust Feature Extraction of Audio Signal by Using VGGish Model [Electronic resource]. – Access mode :<https://www.linkedin.com/pulse/robust-feature-extraction-audio-signal-using-vggish-ijcsis>. – Access date: 20.10.2023.

VISHNIAKOU U.A., HE T.

## CONVOLUTIONAL NEURAL NETWORK FOR IT LUNG DIAGNOSTICS

*Belarusian State University of Informatics and Radioelectronics  
Minsk, Republic of Belarus*

*The subject of research is the use of voice processing technology of the patient in IT medicine. The purpose of the article is to develop a neural network for the diagnosis of lung diseases using sound analysis of the patient's voice. The study includes training of a neural network, development of a mobile program for collecting patient sound, extraction of sound characteristics on the server side, diagnostics of sound data using a trained neural network and return of diagnostic results to the mobile application program. A block diagram of voice processing from the source signal to the extraction of an audio file is presented, as an example, the extraction of MFCC and FBank functions is given. The structure of a convolutional neural network (CNN), which was trained on a standard dataset of respiratory diseases, is given. A simplified process of classification of breathing sounds necessary for the prediction of lung diseases is given. For practical implementation, the VGGish network is used in the Python programming environment, which has network parameters trained using a data set. The experiments were carried out on the Android service framework platform, which is divided into two parts: Android front-end and server. The interface part implements the interactive user function and is responsible for entering audio data. After downloading the audio, the server will pre-process the audio, and call CNN to classify the audio, the results are returned to an external module on the smartphone. The total accuracy of the model reached 83.6 %.*

**Keywords:** IT medicine, voice analysis, neural network model, lung disease recognition



**Вишняков Владимир Анатольевич**, д.т.н., профессор, профессор БГУИР, кафедра ИКТ. Область научных интересов: информационное управление и безопасность, электронный бизнес, интеллектуальные системы управления, сети интернет вещей, блокчейн. Член 2-х докторских Советов по защите диссертаций. Автор более 500 научных работ, в том числе 8 монографий (2 на английском языке), 4-х учебных пособий с грифом Министерства образования, 8-и томного учебного комплекса «Информационный менеджмент», 195 научных статей.

**Vishniakov Uladzimir Anatolyevich**, Doctor of Technical Sciences, Professor, Professor of BSUIR, Department of ICT. Research interests: information management and security, electronic business, intelligent control systems, IoT network. Member of 2 doctoral Councils for the defense of dissertations. Author of more than 470 scientific papers, including 6 monographs (1 in English), 4 textbooks with the stamp of the Ministry of Education, 8-volume educational complex "Information Management", 175 scientific articles.

**E-mail:** vish2002@list.ru

**Хэ Тао**, магистрант факультета ИКТ Белорусского государственного университета информатики и радиоэлектроники. Научные интересы: интеллектуальные диагностические системы.

**He Tao**, master student of ICT department of Belarusian State University of Informatics and Radioelectronics. Research interests: intelligent diagnostics system.