



УДК 334.029.3

СЕТЬ «ИНТЕРНЕТА ВЕЩЕЙ» ДЛЯ ДИАГНОСТИКИ НА ОСНОВЕ АНАЛИЗА ГОЛОСА

ВИШНЯКОВ В. А.,
д. т. н., профессор, профессор кафедры ИКТ

ХЭ Т.,
магистрант кафедры ИКТ

*Белорусский государственный университет
информатики и радиоэлектроники*

Звуки дыхания человека играют большую роль в диагностике заболеваний легких. Различные характеристики заболеваний легких имеют разные характеристики звука дыхания. Разработана и обучена модель сверточной нейронной сети (СНС) с использованием набора данных о звуках дыхания пациентов – ИСВНІ. Эта нейронная сеть включена в состав системы «интернета вещей» (ИВ), включающей смартфон и сервер. В ходе эксперимента выполнен ввод звуков дыхания пациента (через смартфон), выделение признаков и их оптимизация. Для распознавания они поступают в сеть VGGish (VGG классификатор, расположенный на сервере). Результат выводился на смартфон. Система реализована на базе ОС Android. Результаты экспериментов показывают, что реализация предложенного алгоритма распознавания достигла точности классификации заболеваний, связанных со звуками дыхания 83%.

Ключевые слова: ИТ-диагностика легких, обработка звуков голоса, машинное обучение, нейронная сеть, сеть интернет вещей.

ВВЕДЕНИЕ

Одной из подсистем умного города является ИТ-медицина. В ней важную роль занимает диагностика пациентов, которая строится на технологии сетей «интернета вещей» [1]. Заболевания легких являются одной из наиболее распространенных причин смерти во всем мире. Ежегодно более 4 млн человек умирают преждевременно. Для диагностики пациентов используются рентгеновские снимки грудной клетки и компьютерная томография (КТ). Однако эти методы сканирования являются дорогостоящими и могут увеличить риск развития рака в результате облучения. Стетоскоп также используется для прослушивания звуков в легких пациента, но для постановки надежного диагноза требуется высоко профессиональный врач.

С развитием «интернета вещей» и технологий искусственного интеллекта в области ИТ-диагностики можно создавать системы удаленного обследования пациентов. Для этого в систему «интернета вещей» включается нейронная сеть, обученная на классификации

звуков дыхания пациентов с заболеваниями легких. В статье используется набор данных – ИСВНІ, приведенный в [2], содержащий звуки дыхания пациентов с заболеваниями легких. Эти данные используются для обучения сверточной нейронной сети для достижения эффекта точной диагностики заболеваний легких. Эта СНС встраивается в сеть ИВ для построения удаленной диагностики легких пациента без участия врача.

ОСНОВНАЯ ЧАСТЬ

Классификация звуковых сигналов. Типы сигналов, частотные диапазоны и методы анализа, используемые для внутренних и внешних сигналов легочных шумов, обобщены в таблице 1.

Звуки дыхания вызываются потоком воздуха через легкие и могут быть просто классифицированы на две категории: нормальные или ненормальные. Эти звуки обычно делятся на трахеобронхиальные и везикулярные. Первые происходят из трахеи и более крупных бронхиальных дыхательных путей [2],

Виды звуков	Частотный диапазон	Метод анализа
Звуки дыхания	10-120Hz, 60-2000Hz	Вейвлет-регрессия Мейера, БПФ*
Грудной резонанс	80-1000Hz	Коэффициент корреляции, БПФ
Трансоральный резонанс дыхательных путей	200-2000Hz, 50-2000Hz	БПФ, FRF*, анализ мощности, спектрограмма
Сонограмма дыхательных путей с катетером	20-1600Hz, 50-1600Hz	СПМ*, БПФ

* БПФ – быстрое преобразование Фурье.
 СПМ – спектральная плотность мощности.
 FRF – функция частотной характеристики.

Таблица 1. Типичные диапазоны звуков и методы их анализа

а вторые – из небольших ветвей трахеального дерева, удаленных от трахеи, или других участков в дистальных отделах паренхимы легкого. Отсутствие нормальных звуков дыхания или появление нечленораздельных звуков может быть признаком заболевания легких [3].

Модель и алгоритм извлечения звуков дыхания. 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 многоступенчатым операциям предварительной обработки, включая предварительное выделение, кадрирование и дискретное преобразование Фурье [4]. На рисунке 1 представлена блок-схема алгоритма обработки звука от исходного сигнала до извлечения аудиофайлов, в качестве примера приведено извлечение функций MFCC (частотный кепстральный коэффициент Mel) и FBank (FilterBank).

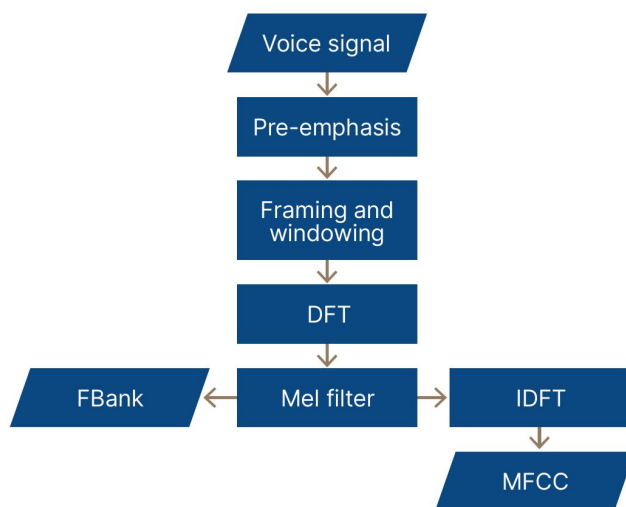


Рисунок 1. Процесс извлечения признаков из MFCC и FBank

Реакция человеческого уха на звуковой спектр нелинейна. Fbank – это алгоритм фронтальной обработки, который обрабатывает аудио способом, аналогичным человеческому уху, что может улучшить производительность распознавания речи [5]. Общими шагами для получения характеристик fbank речевых сигналов являются: предварительное выделение, кадрирование, управление окнами, кратковременное преобразование Фурье (STFT), фильтрация mel и т. д. Выполняется дискретное косинусное преобразование (DCT) в fbank, чтобы получить функции mfcc.

На частотный кепстральный коэффициент Mel (MFCC) не влияет такая информация, как высота звука. Кепстральный коэффициент является широко признанной и эффективной характеристикой в задачах распознавания звуков пациентов. Поэтому функции MFCC часто используются в задачах, связанных с анализом голоса. Кепстральный коэффициент частоты Mel (MFCC) основан на слуховом восприятии человека и выводится на основе его расчетов. Хотя функция MFCC обладает большим количеством преимуществ, она также имеет некоторые ограничения [6]. Например, при наличии фонового шума эффективность функции будет снижена. Алгоритм извлечения функции из аудиофайлов включает:

1. Ввод аудиоданных в формате WAV.
2. Повтор дискретизации звука на частоте 16 кГц.
3. Функция Window настроена на Hanning Window, длина окна равна 25 мс, а размер шага равен 10 мс. Выполнение быстрого преобразования Фурье звука, чтобы получить спектрограмму.
4. Использование набор фильтров Mel 64-го порядка для вычисления спектра Mel на спектрограмме, полученной на предыдущем шаге;
5. Вычисление логарифма (mel-спектр+0,01), чтобы получить стабильный спектр Mel;
6. Выполнение кадрирования 0,96 в секунду, при котором кадры не складываются в стопку. Каждый кадр содержит 64 полосы Mel, а каждая секунда содержит в общей сложности 96 кадров.

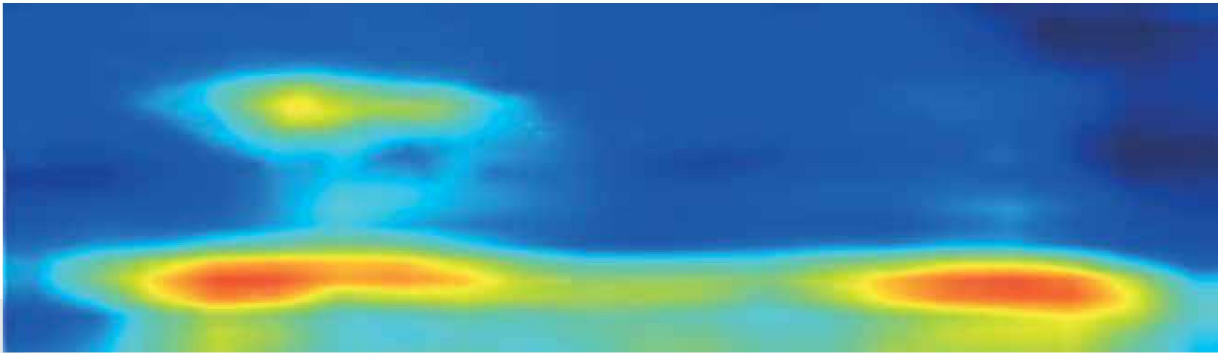


Рисунок 2. GradCam анализирует исходный образец

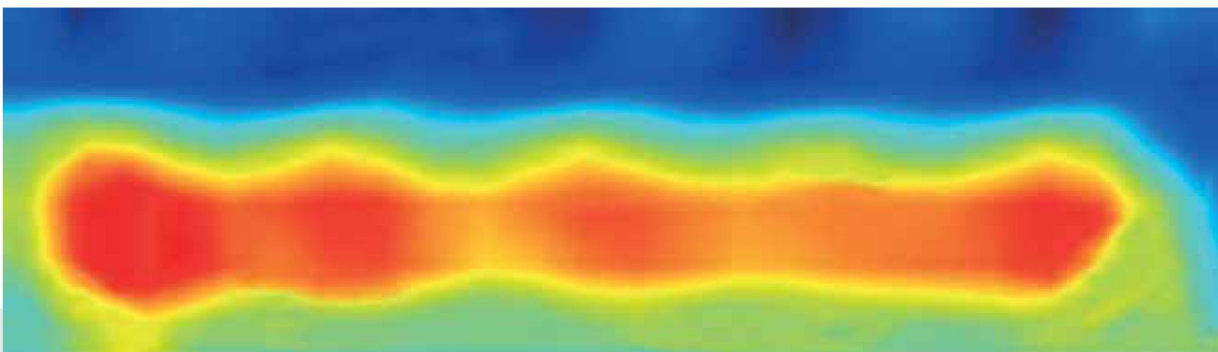


Рисунок 3. Оптимизированная выборка для анализа GradCam

Следовательно, последним входом в сеть является матрица размером $10 \times 96 \times 64$. 10 представляет 10 временных шкал, 96 представляет количество кадров в каждой временной шкале, а 64 представляет количество объектов Mel. После описанной выше обработки получается спектрограмма Mel, генерируемая звуками дыхания, которая является входными данными модели VGGish.

Оптимизация функций на основе GradCam. GradCam расшифровывается как карта активации взвешенного градиентного класса, которая может выводить области на входном изображении, важные для прогнозирования и классификации целей, в виде тепловых карт. Этот метод использует сверточную нейронную сеть, в которой модель извлекает более специфические свойства определенных категорий объектов из входного изображения на более низких уровнях.

Функции, используемые в базовой модели данной статьи, были выбраны в диапазоне частот 100-2000 Гц [7]. После предварительных экспериментов, когда GradCam использовался для анализа образцов, классифицированных по базовой модели, в высокочастотных областях их спектров были обнаружены значительные области с низким уровнем нагрева. Как показано на рисунке 2, области с высокой температурой на рисунке представляют части, которые более важны для модели классификации, а области с низкой температурой

представляют части, которые относительно неважны для прогнозирования цели и классификации.

Можно видеть, что фокус модели на входном изображении в основном сосредоточен в нижней половине входного изображения, что указывает на то, что некоторые области в диапазоне высоких частот звука не могут обеспечить хорошие характеристики для классификационного суждения модели. Дальнейший анализ показывает, что многие образцы имеют пустые области, особенно в диапазоне частот 1500-2000 Гц. Это может отрицательно сказаться на производительности сети, описанной в данной статье. Чтобы оптимизировать эффект модели, в этой статье выборочно вырезаются пустые линии из высокочастотных областей этих спектров. Цель этого состоит в том, чтобы гарантировать, что сеть фокусируется на интересующей области и уменьшает помехи в несущественных областях, тем самым повышая производительность модели. Выбрана область выше частоты звукового сигнала Mel, равной 1500 Гц. После отсека высокочастотных областей сеть начинает уделять больше внимания нижней половине спектрограммы, как показано на рисунке 3.

Структура модели СНС. После выделения признаков звуков дыхания получается спектрограмма Mel. Используем среду программирования Python. После нарезки вводим VGGish, чтобы получить 128-мерный вектор. Следующим шагом является

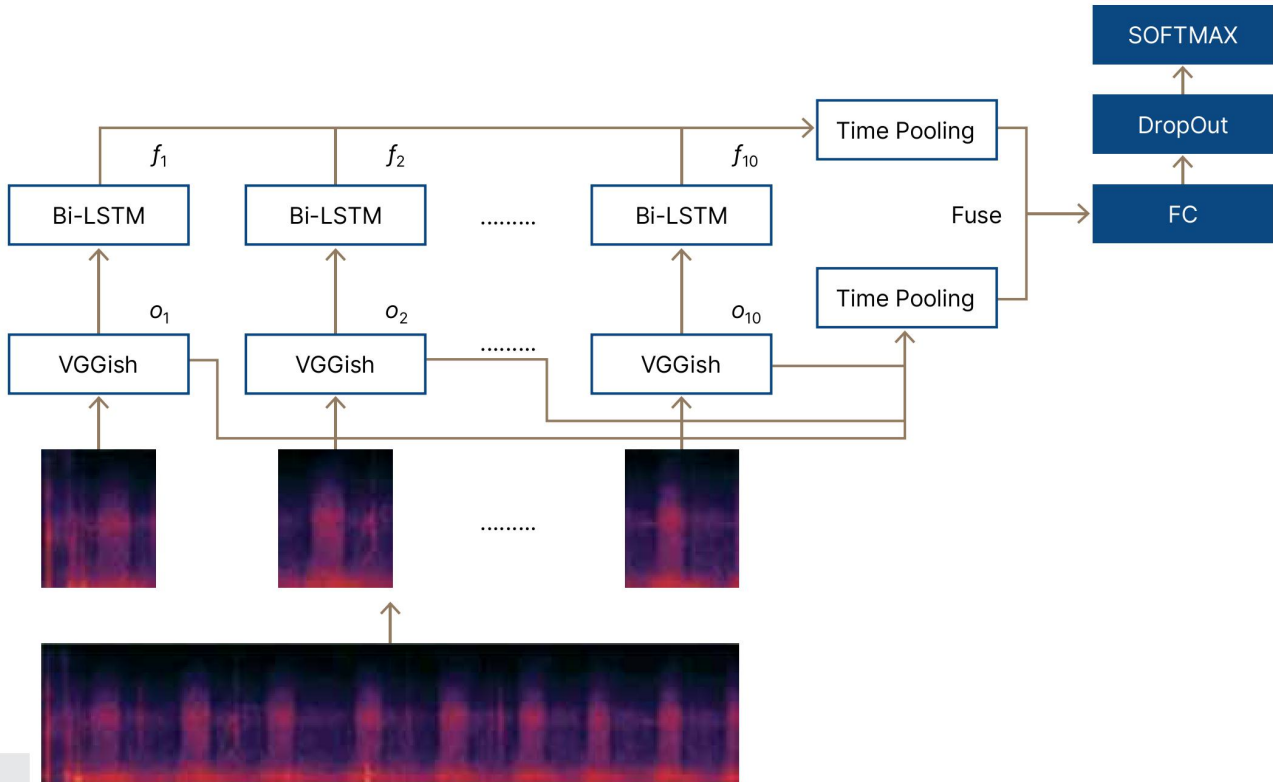


Рисунок 4. Структура модели CHC, основанная на VGGish

ввод вектора в сеть Blstm для обучения. В то же время, чтобы уменьшить вероятность потери информации частью нейронной сети (CHC) в процессе обучения, используется объединение времени для интеграции выходных данных VGGish и BiLSTM и обучение с использованием полностью связанных выходных векторов для улучшения способности модели к обобщению [8, 9]. Структура модели CHC показана на рисунке 4.

Алгоритм диагностики заболеваний легких по звукам дыхания. После завершения обучения модели CNN упрощенный процесс классификации звуков дыхания, необходимых для прогнозирования заболеваний, выглядит следующим образом:

1. Выполнение предварительной обработки выделения признаков для звуков дыхания пациента.

2. Выполнение оптимизации для извлеченных признаков.
3. Выполнение классификации с использованием модели VGGish.
4. Выведение результатов распознавания.

Анализ результатов эксперимента. Алгоритм протестирован на скомпилированном наборе данных для проверки эффективности модели [10]. В то же время, для проверки способности модели к обобщению в статье используется пятикратная перекрестная проверка для расчета экспериментальных результатов в эксперименте. Результаты экспериментов приведены в таблице 2.

Судя по результатам, эффект распознавания нормальных звуков дыхания и симптомов ХОБЛ является наилучшим. Точность и оценка F1 нормальных звуков дыхания являются самыми

	Upper respiratory tract infection (URTI)	Healthy	Хроническая обструктивная болезнь легких (COPD)	Бронхоэктазы (Bronchiectasis)	Общая
Точность	0,7831	0,8685	0,8506	0,7590	0,8360
Recall	0,6915	0,8810	0,880	0,7326	
F1-score	0,7345	0,8747	0,8689	0,7335	

Таблица 2. Результаты экспериментов с моделью VGGish-Lstms

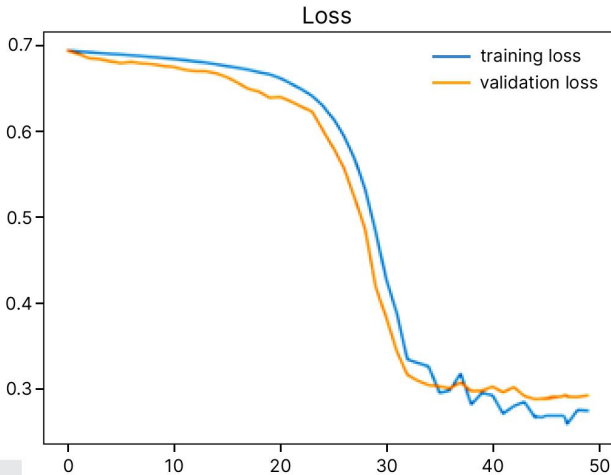


Рисунок 5. Диаграмма снижения потерь при обучении модели VGGish-Lstm

высокими, но частота запоминания ниже, чем при ХОБЛ. Инфекция верхних дыхательных путей и бронхоэктазы были менее точными, чем обычные звуки дыхания, при этом бронхоэктазы имели более высокий балл F1. Общая точность распознавания достигла 83,6%. Это указывает на то, что модель может лучше определять заболевание по звукам дыхания. Как следствие, получаем определенный эффект, особенно при определении наличия заболевания легких. На рисунке 6 показан процесс, при котором функция потерь модели продолжает уменьшаться до тех пор, пока

она не сойдется в процессе обучения. Из рисунка 5 видно, что модель достигает сходимости примерно на 45-й эпохе.

Структура сети ИВ для идентификации заболеваний легких по звукам дыхания. Сеть ИВ реализована на смартфоне и сервере, показана на рисунке 6. Она разделена на две части: интерфейсная Android front-end (смартфон) и серверная. 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 отправляются обратно в приложение «Получить результаты».

Сеть ИТ-диагностики пациентов по звуку дыхания, основанная на клиенте и сервере ОС Android. Клиент для Android разработан с использованием Java и предоставляет функции взаимодействия с пользователем, включая «запись с микрофона» и «загрузку аудиофайлов» для сбора звуков дыхания. Серверная часть построена с использованием фреймворка Flask и разработана с помощью Python. Сервер получает, хранит и обрабатывает аудиофайлы, включая предварительную обработку, извлечение признаков, а затем использует предварительно обученные модели для классификации и диагностики. Результаты диагностики возвращаются в клиент Android и отображаются пользователю.

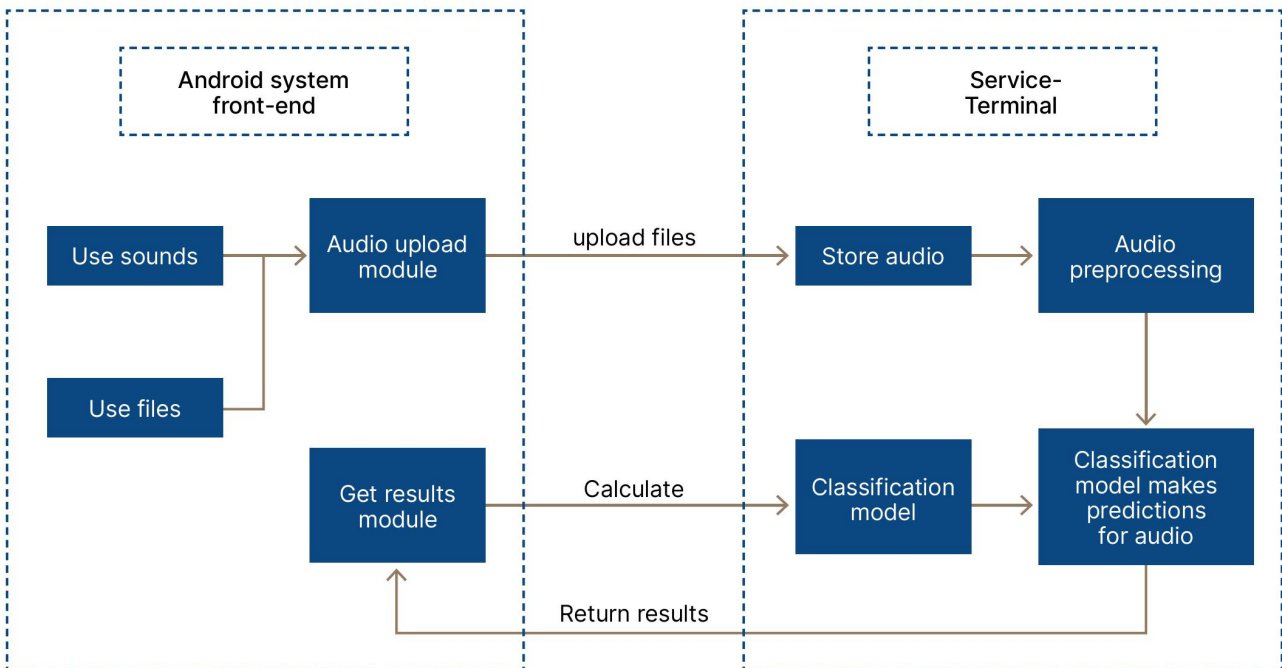


Рисунок 6. Структура сети ИВ на базе Android



ЗАКЛЮЧЕНИЕ

Разработана и обучена модель СНС с использованием набора данных о звуке дыхания ИСВНІ пациентов. Эта нейронная сеть включена в состав системы ИВ, включающей смартфон и сервер. В ходе эксперимента выполнены ввод дыхания пациента (через смартфон), выделение признаков и их оптимизация, которые были использованы

для распознавания в сети VGGish (VGG-подобный классификатор аудиосетей, расположенный на сервере). Результат выводился на смартфон. Система реализована на базе ОС Android. Результаты экспериментов показывают, что реализация предложенного алгоритма распознавания достигла точности классификации заболеваний, связанных со звуками дыхания до 83 %.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Вишняков, В. А. Специализированные IoT-сети: модели, структуры, алгоритмы, программно-аппаратные средства=Specialized IoT systems: Models, Structures, Algorithms, Hardware, Software Tools / В. А. Вишняков. – Минск: БГУИР, 2023. – 184 с.
2. Rocha, В.М., 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.
3. Göğüş, F. Z., Karlık B., Harman G. Classification of asthmatic breath sounds by using wavelet transforms and neural networks. International Journal of Signal Processing Systems, 2015, 3(2), pp. 106–111.
4. Shi, Y., i Y., Cai M., et al. A Lung Sound Category Recognition Method Based on Wavelet Decomposition and BP Neural Network[J]. International journal of biological sciences, 2019, 15(1), pp. 195–207.
5. Liu, Y., Zhang C., Zhao Y. The feature extraction and classification of lung sounds based on wavelet packet multiscale analysis. Chinese Journal of Computers, 2006, 29(5), pp. 769–777.
6. Li, S., Liu Y. Feature extraction of lung sounds based on bispectrum analysis. Third International Symposium on Information Processing, Qingdao, 2010, pp. 393–397.
7. Chen, H., Yuan X., Pei X. et al. Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. IEEE Access, 2019, 7, pp. 32845–32852.
8. Acharya, J., Basu A. Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. IEEE transactions on biomedical circuits and systems, 2020, 14(3), pp. 535–544.
9. Gemmeke, J. F., Ellis D. P. W., Freedman D. et al. Audio set: An ontology and human-labeled dataset for audio events. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, 2017, pp. 776–780.

Human breathing sounds play an important role in the diagnosis of lung diseases. Different pathological characteristics of lung diseases have different characteristics of the sound of breathing. A CNN (convolutional neural network) model was developed and trained using a data set on the breathing sounds of ICHL patients. This neural network is included in the Internet of Things (IoT) system, which includes a smartphone and a server. During the experiment, the patient's breathing was entered (via a smartphone), signs were isolated and optimized. For recognition, they are sent to the VGGish network (VGG classifier located on the server). The result was displayed on the smartphone. The system is implemented on the basis of Android OS. The experimental results show that the implementation of the proposed recognition algorithm has achieved 83% accuracy in the classification of diseases associated with breathing sounds.

Keywords: lung diagnostics, voice sound processing, machine learning, neural network, Internet of Things network.

Статья поступила в редакцию 05.09.2025.