

Стетоскоп 3.0

Трансформеры, VLM и Latent-Diffusion

Мультимодальный конвейер машинного обучения для
быстрого скрининга бронхиальной астмы по дыхательным
шумам

Докладчик: Аптекарев Ф. А., соискатель степени PhD, НИУ ВШЭ, Нижний Новгород

Авторы:

- Аптекарев Федор (НИУ ВШЭ),
- Соколовский Владимир (Ben-Gurion University),
- Фурман Евгений (ПГМУ им. ак. Е.А. Вагнера),
- Калинина Наталья (Пермская Краевая Клиническая Больница),
- Фурман Григорий (Ben-Gurion University)



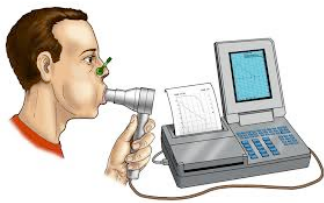
**ПЕРМСКИЙ
МЕДИЦИНСКИЙ**
УНИВЕРСИТЕТ АКАДЕМИКА ВАГНЕРА



Ключевые проблемы

- 250–350 млн пациентов с БА¹
- Нет "золотого стандарта" диагностики; до 30% неверных диагнозов²
- Ограничения спирометрии/аускультации: субъективность, трудности у детей
- Дефицит доступа к специалистам и оборудованию

Мотивация: объективизировать и стандартизовать первичный скрининг



Цель

Разработать и экспериментально обосновать ML-конвейер для скрининга/поддержки диагностики бронхиальной астмы по респираторным шумам

Задачи

- Сформировать корпус дыхательных шумов
- Построить конвейер признаков (мел-спектрограммы)
- Выбрать и обучить модели
- Разработать схему синтетической аугментации
- Обеспечить интерпретируемость
- Разработать модель интеграции в систему здравоохранения

1. - Global Initiative for Asthma, 2018 ; GBD 2019 Diseases and Injuries Collaborators, 2020

2. - "Deep learning facilitates the diagnosis of adult asthma"; Tomita et al., 2019

Эволюция методов анализа респираторных шумов (2015-2025)

До 2017

- Классические методы цифровой обработки сигналов
- Традиционный ML (SVM, GMM, HMM с ручными признаками)
- Первые эксперименты с DL

2017-2020 (до COVID19)

- Расширение экспериментов с DL
- Использование CNN, RNN
- Публикация ICBHI датасета

2020-2025

- Современный DL
- Больше open access датасетов
- Dense CNNs, Трансформеры, VLM, и т.д.

Год	Автор	Название	Что примечательного
2016	Chamberlain et al.	Application of semi-supervised deep learning to lung sound analysis	AUC 0,86/0,74 с неразмеченными данными
2017	Pramono et al.	Automatic adventitious respiratory sound analysis: A systematic review	Систематический обзор
2017	Aykanat et al.	Classification of lung sounds using convolutional neural networks	86% accuracy, сопоставимо с SVM
2017	Rocha et al.	An open access database for the evaluation of respiratory sound classification algorithms	Open access ICBHI 2017 датасет
2019	Chen et al.	Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual Networks	~98% accuracy
2019	Tomita et al.	Deep learning facilitates the diagnosis of adult asthma	~98% accuracy, AUC ~0,99
2021	Kim et al.	Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning	~86,5% accuracy + проверили в клинических условиях
2021	Hsu et al.	Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database – HF_Lung_V1	Open access HF_Lung_V1 датасет
2023	Aptekarev et al.	Application of deep learning for bronchial asthma diagnostics using respiratory sound recordings	~87% accuracy, 93% precision

Датасет

Собирался на базе ПГМУ им. ак. Е.А. Вагнера и Пермской Краевой Клинической Больнице в 2014-2020 гг.

Точка записи	Астма	Здоровый	Больной (не астма)
Трахея	563	123	101
Второе межреберье	285	9	10
Грудная клетка сзади	256	0	12
Ротовая полость	9	1	2
NA	0	0	242
Total	1113	133	367

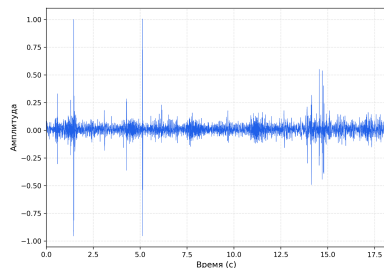
Возрастная группа	Астма	Здоровый	Больной(не астма)
0-2	14	6	1
2-4	8	0	39
4-13	515	32	221
13-20	516	13	106
20	60	82	0
Total	1113	133	367

Всего обследуемых: 1613
Пациенты с подтвержденным диагнозом: 1480
Пациенты с астмой: 1113
Единица анализа: 5-сек фрагменты
Предобработка: фильтрация/нормализация

Диагноз	F	M	Всего
Астма (неполная ремиссия)	113	307	420
Астма (обострение)	53	131	184
Астма (ремиссия)	77	166	243
Астма (другое)	70	196	266
Здоров	52	81	133
БЛД	1	0	1
Муковисцидоз	153	90	243
ХНЗЛ	0	1	1
Пневмония	10	0	10
РОБЛ	13	84	97
Аплазия легкого	0	15	15
Total	542	1071	1613

Предварительная обработка

Исходный сигнал



1.

Фильтрация коротких записей

Устанавливаем порог по хронометражу

$$CHRONO_THRESHOLD = 14c$$

Считаем хронометраж по частоте дискретизации

$$duration = \frac{len(waveform)}{sample_rate}$$

Всё что меньше - бракуем

$$technical_defect = \begin{cases} True, & \text{если } duration < 14c \\ False, & \text{в противном случае} \end{cases}$$

2.

Фильтрация клипинг артефактов

Считаем 95й перцентиль абсолютных значений амплитуд

$$Threshold = P_{95}(| waveform |)$$

Считаем процент сэмплов попадающих в 95й перцентиль

$$Clipping\% = \frac{N_{samples \geq Threshold}}{N_{total}} \cdot 100$$

Бракуем если процент сэмплов попадающих в 95й перцентиль больше чем

$$QUALITY_THRESHOLD = 2$$

3.

Обрезка сигнала

Определяем сколько будем обрезать с начала и конца

$$t_{trim} = 2c$$

Считаем сколько сэмплов отрезать по частоте дискретизации

$$N_{trim} = t_{trim} \cdot f_s$$

Обрезаем нужное кол-во сэмплов с начала и конца

$$waveform[N_{trim} : (len(waveform) - N_{trim})]$$

4.

Сегментация сигнала

Определяем перекрытие и шаг смещения между клипами

$$L = T \cdot f_s \quad S = L \cdot \left(1 - \frac{\alpha}{100}\right)$$

Рассчитываем число целых клипов которые можно получить

$$C = \left\lfloor \frac{L_{total} - L}{S} \right\rfloor + 1$$

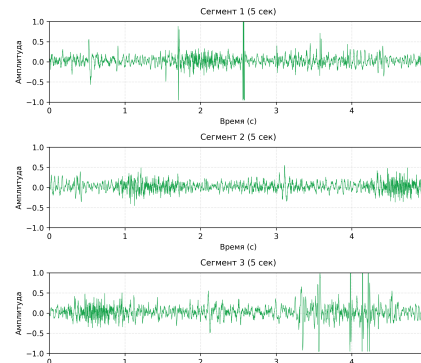
Итерационно вырезаем нужное число целых клипов из середины

$$T_{span} = (C - 1) \cdot S + L$$

$$T_{lower} = \left\lfloor \frac{L_{total} - T_{span}}{2} \right\rfloor, \quad T_{upper} = T_{lower} + T_{span}$$

5.

Нормализация клипов



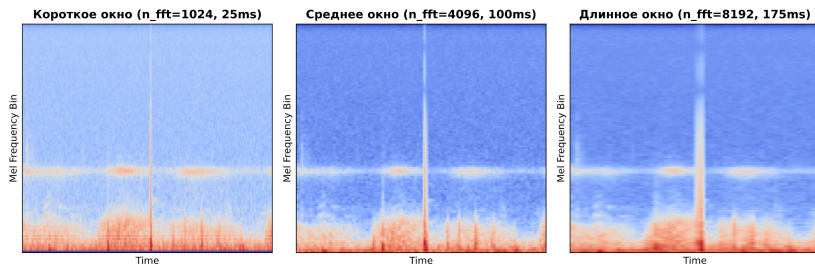
6.

ISU

Модели и обучение

Диагностические модели

- **DenseNet201**: CNN используемая как baseline
- **AST**: Основная модель, инициализированная ViT.
- **Moondream**: Мультимодальная (спектрограмма + текст)



Извлечение спектральных признаков

3 STFT Окна:

- короткое (25 мс)
- среднее (100 мс)
- длинное (175 мс)

Мел-фильтры: Частотная ось преобразуется в нелинейную мел-шкалу (128 мел-коэффициентов, полоса 0-8 кГц)

- “Densely Connected Convolutional Network” (Huang et al, 2017, <<https://doi.org/10.48550/arXiv.1608.06993>>)
- “AST: Audio Spectrogram Transformer (Gong et al, 2021 <<https://doi.org/10.48550/arXiv.2104.01778>>)
- <<https://github.com/vikhyat/moondream>>

Обучение и валидация

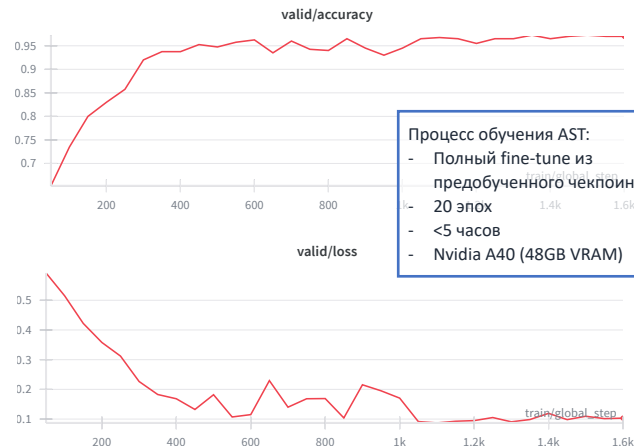
- Базовая классификация: “Астма | Не астма”
- Бутылочное горлышко при создании обучающего корпуса: записи здоровых
- Мониторинг validation loss/accuracy
- Ранняя остановка при отсутствии улучшений
- Метрики: Accuracy, Sensitivity, Youden Index

model_name: "MIT/ast-finetuned-audioset-10-10-0.4593"

```
# Training parameters
num_train_epochs: 25
per_device_train_batch_size: 16
learning_rate: 0.00008
warmup_ratio: 0.25
weight_decay: 0.03
logging_steps: 1
eval_steps: 10
save_steps: 50
gradient_accumulation_steps: 4
max_grad_norm: 1.0
fp16: true
```

```
# Early stopping
early_stopping_patience: 5
```

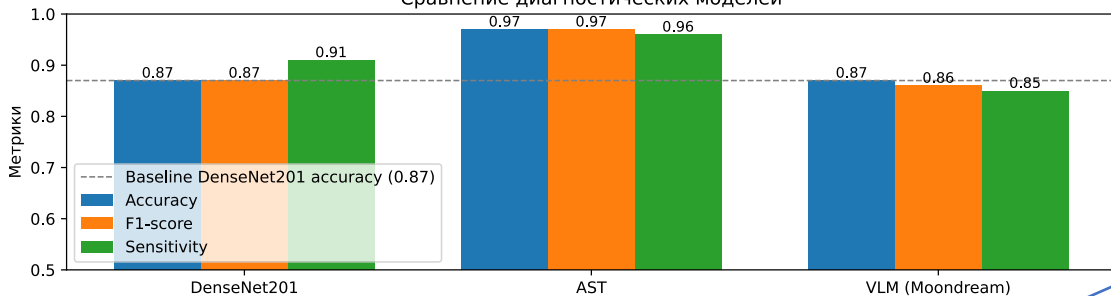
```
# Scheduler parameters
scheduler_type: "cosine"
scheduler_warmup_steps: 150
scheduler_decay_steps: null
scheduler_decay_rate: 0.1
```



Результаты бинарной классификации

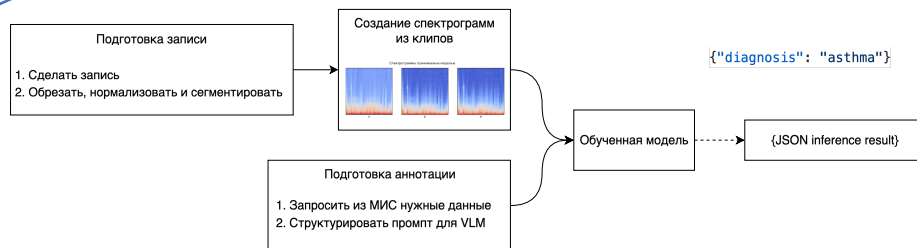
Астма | Не астма

Сравнение диагностических моделей



Оценка на 5-сек клипах; кросс-валидация

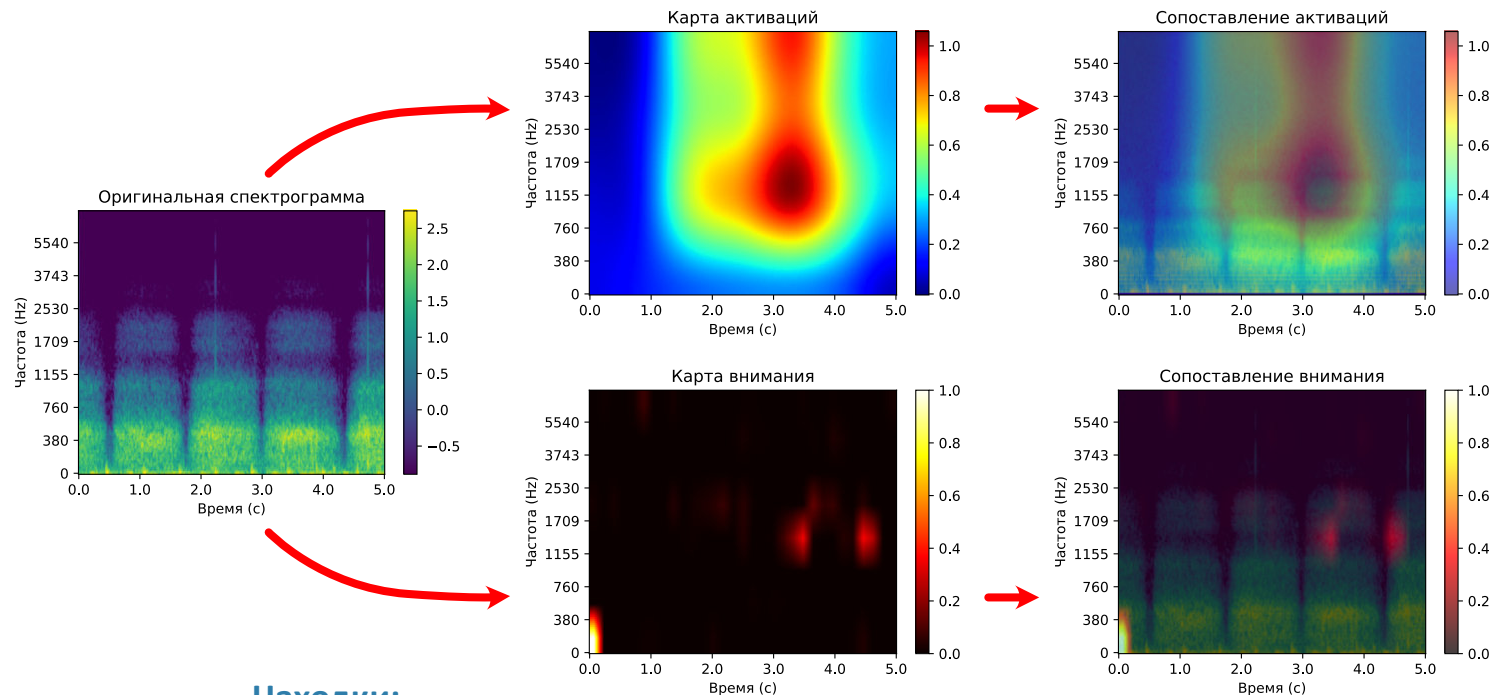
Процесс inference на VLM



```
{  
  "task": "Write a diagnosis for this patient by analyzing the respiratory sound spectrogram.",  
  "patient_sex": "M",  
  "patient_age": 9,  
  "record_point": "posterior thoracic rib",  
  "spectrogram_bins_width": "[0.025, 0.1, 0.175]"  
}
```

На что смотрит модель?

Интерпретируемость: необходимое условие клинического внедрения

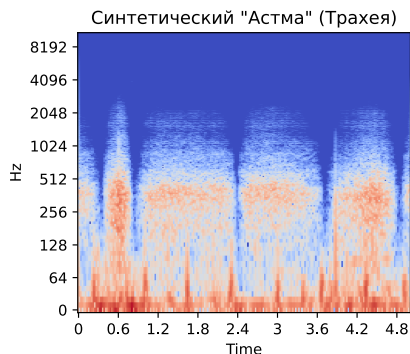
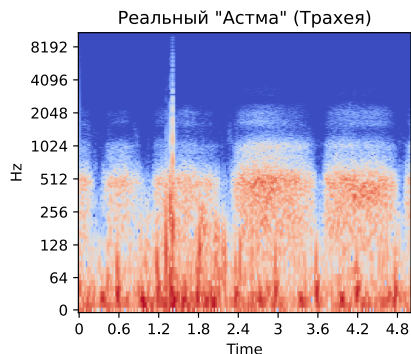
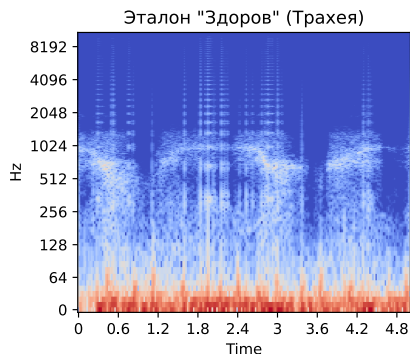
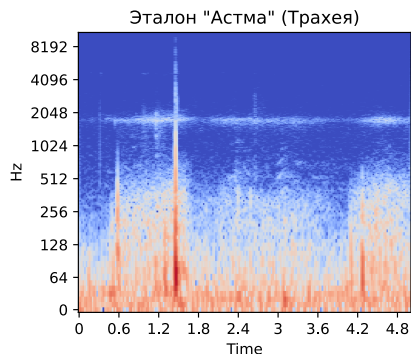


Находки:

- Совпадение фокусов на удлинённом выдохе и характерных зонах
- Подтверждение представлений о клинических проявлениях патологии

Синтетическая аугментация

```
{
  "prompt": "Breath sample from F-8yo diagnosed with bronchial asthma",
  "seconds_start": 0,
  "seconds_total": 5
},
```



Вопрос: как валидировать синтетику?

Точность диагностического
классификатора: $\approx 95.2\%$

Идея: ИИ проверяет ИИ

Цель: борьба с дисбалансом и дефицитом данных

Процесс генерации

Текстовые/акустические условия



Stable Audio Open (адаптация)



Синтетические дыхательные сигналы

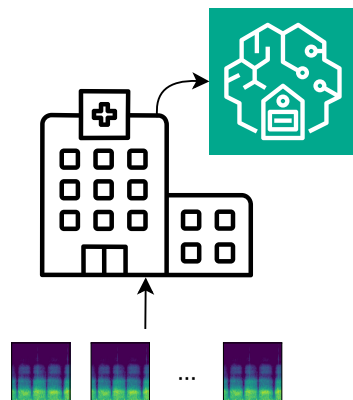
Parameter	Pretraining	Asthma FT	Healthy FT
Steps	250	200	200
CFG scale	[3,6,9]	6.5	5.5
Temperature	1	0.9	0.95
Top-k	100	50	100

Интеграция в систему здравоохранения

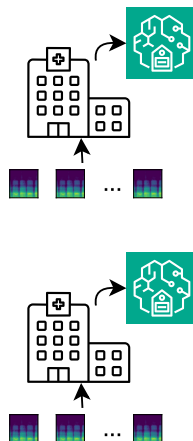
Федеративная схема

Центральная модель (НИИ) → распределение в клиники → локальное дообучение → агрегация весов (FedAvg) → обновленная базовая модель

Специализированный центр
(НИИ, Федеральный центр, ...)



Региональные центры
(РКБ, ОКБ, ...)



Клиническая практика



Преимущества подхода

- Конфиденциальность: данные не покидают учреждение
- Переносимость: локальная адаптация под популяцию/оборудование
- Обновляемость: периодическая синхронизация

Ограничения, выводы и следующие шаги

Ограничения

- Нужна экспертная валидация синтетики и клиническая верификация
- Расширение за пределы бинарной задачи (ХОБЛ, муковисцидоз)
- Стандартизация отчета для врача

Выводы

- ML-конвейер демонстрирует высокую точность (AST $\approx 97\%$) и объяснимость
- Синтетика помогает с дефицитом данных ($\approx 95.2\%$)
- Федеративная схема обеспечивает приватность и адаптацию

Благодарю за внимание! Вопросы приветствуются.

aptekarev@gmail.com



Ссылка на репозитория с демонстрацией и весам