



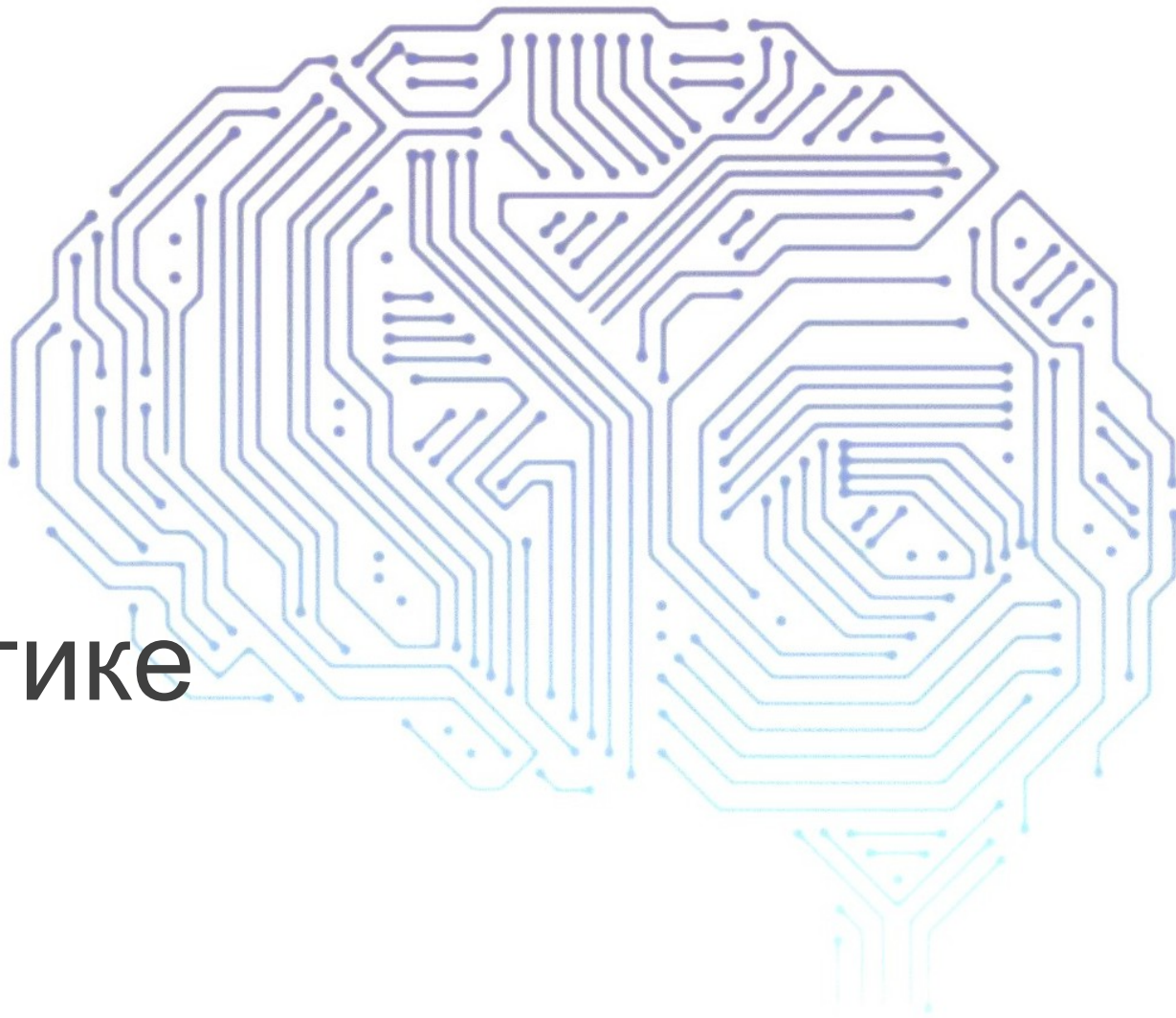
НМИЦ
ОНКОЛОГИИ
им. Н.Н. Блохина

«Использование искусственного интеллекта в диагностике и лечении»

Программа повышения квалификации (36 ак. часов)
2026 г.

Шувалова Екатерина Викторовна

доцент, кандидат физико-математических наук





**НМИЦ
ОНКОЛОГИИ**
им. Н.Н. Блохина

О программе

Цель программы

- Познакомить слушателей с современными возможностями применения ИИ в медицине,
- дать практические навыки работы с медицинскими данными, моделями машинного обучения и инструментами их анализа,
- сформировать понимание этических, правовых и организационных аспектов внедрения ИИ в клиническую практику

Целевая аудитория

- Врачи и специалисты здравоохранения, заинтересованные в использовании ИИ
- Исследователи в области медицинских технологий и искусственного интеллекта
- Руководители медицинских учреждений, планирующие внедрение AI-решений



ФГБНЦ
ВНИИМ
им. Н.Н. Блохина

О программе

Структура курса

Модуль	Тема	часы
1	Введение в искусственный интеллект в медицине	4
2	Медицинские данные: источники, форматы, предобработка	8
3	Основы нейронных сетей для анализа медицинских изображений	8
4	Метрики и оценка качества моделей в медицине	4
5	Продвинутые методы анализа медицинских изображений	6
6	Обработка медицинских текстов с использованием NLP	4
7	Внедрение ИИ в клиническую практику	2



НМИЦ
ОНКОЛОГИИ
им. Н.Н. Блохина

О программе

Особенности работы в российском информационном пространстве

Некоторые зарубежные платформы (Kaggle) могут быть недоступны
Используем альтернативные источники данных и облачные сервисы

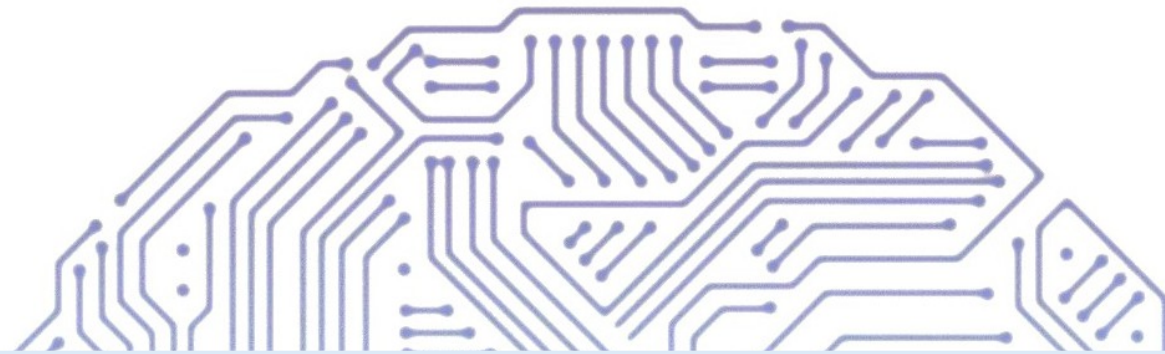
Доступные ресурсы в РФ:

- Облачные платформы: Yandex DataSphere, SberCloud ML Space
- Датасеты: [TCIA](#), NIH ChestX-ray, LUNA16, ISIC, BRATS, [Medical Segmentation Decathlon](#), российские RuMedPrimeData, data.gov.ru, opendata.mos.ru
- Библиотеки и модели: Hugging Face (доступен), DeepPavlov, RuBERT, Natasha
- Инструменты: Python (зеркало PyPI Яндекса), TensorFlow/PyTorch в GoogleCollab



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1



Введение в искусственный интеллект в медицине (4 часа)

Основные темы:

- Роль ИИ в современной медицине: диагностика, лечение, прогнозирование, управление
- Основные понятия машинного и глубокого обучения (обзор для начинающих)
- Этические аспекты: конфиденциальность, информированное согласие, справедливость алгоритмов
- Правовое регулирование в РФ: 152-ФЗ, регистрация медицинских изделий на базе ИИ (Росздравнадзор)

Практическая часть:

- Знакомство с интерфейсом Yandex DataSphere / Google Colab
- Установка необходимых библиотек (через pip с зеркалом Яндекса)



НМИЦ
ОНКОЛОГИИ
им. Н.Н.Блохина

Тема 1.1

Роль ИИ в современной медицине

Диагностика

Анализ медицинских изображений (рентген, КТ, МРТ)

Выявление патологий (опухоли, переломы)

Лечение

Подбор терапии, хирургические роботы

Персонализированная медицина

Прогнозирование

Риск развития заболеваний, исходы госпитализаций

Эпидемиологические прогнозы

Управление

Оптимизация расписания, маршрутизация пациентов

Анализ эффективности работы отделений



ФГБНЦ
ВНИИ
ОНКОЛОГИИ
им.Н.Н.Блохина

Роль ИИ в со

Диагностика

Анализ медицинских изображений
(рентген, КТ, МРТ и
т.д.)

Выявление патологий
(опухоли, переломы)



X-ray



US



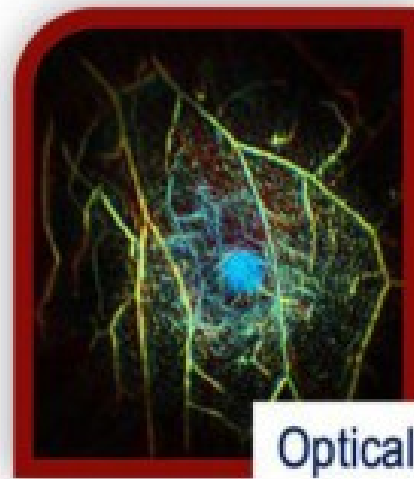
Scinti



PET



CT



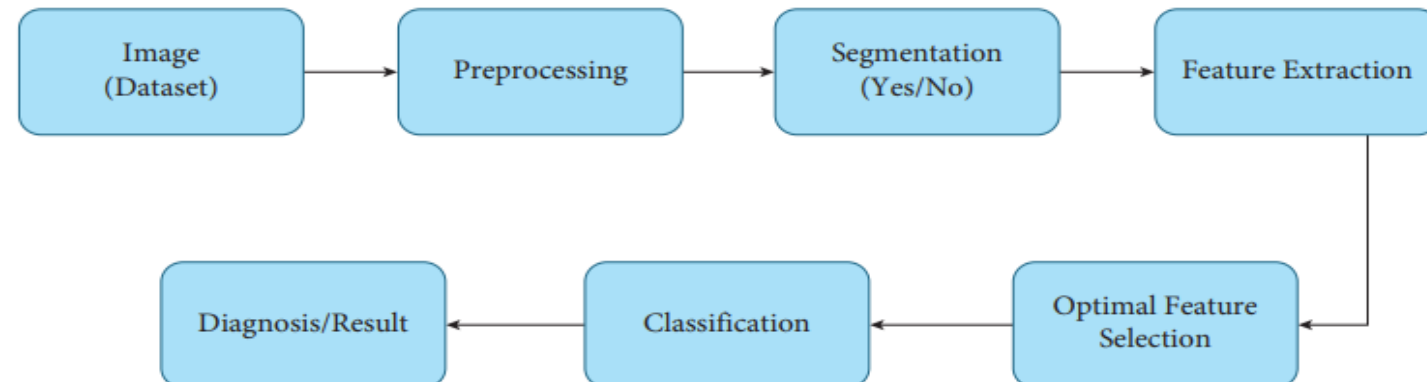
Optical



MRI



SPECT



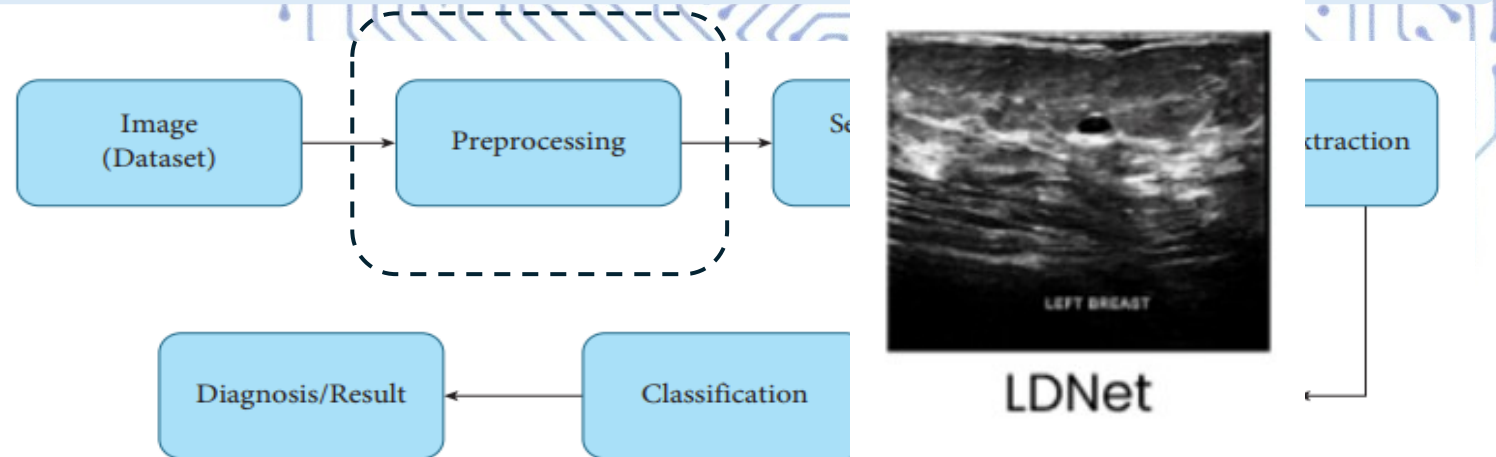
Тема 1.1

Роль ИИ в современной медицине

Диагностика

Анализ медицинских изображений
(рентген, КТ, МРТ)

Выявление патологий
(опухоли, переломы)



Основная задача предобработки заключается в уменьшении шума, улучшении контраста, выравнивании яркости и контрастности, фильтрации для удаления артефактов, а также совмещение одного вида изображений с другим (например, КТ или МРТ) для более точной локализации.

Роль ИИ в с

Диагностика

Анализ медицинских изображений (рентген, КТ, МРТ)

Выявление патологии (опухоли, переломы)

Сегментация изображений позволяет отделить патологические области от здоровых тканей, обеспечить высокую детализацию контуров, выделить органы, опухоли или кровеносные сосуды.

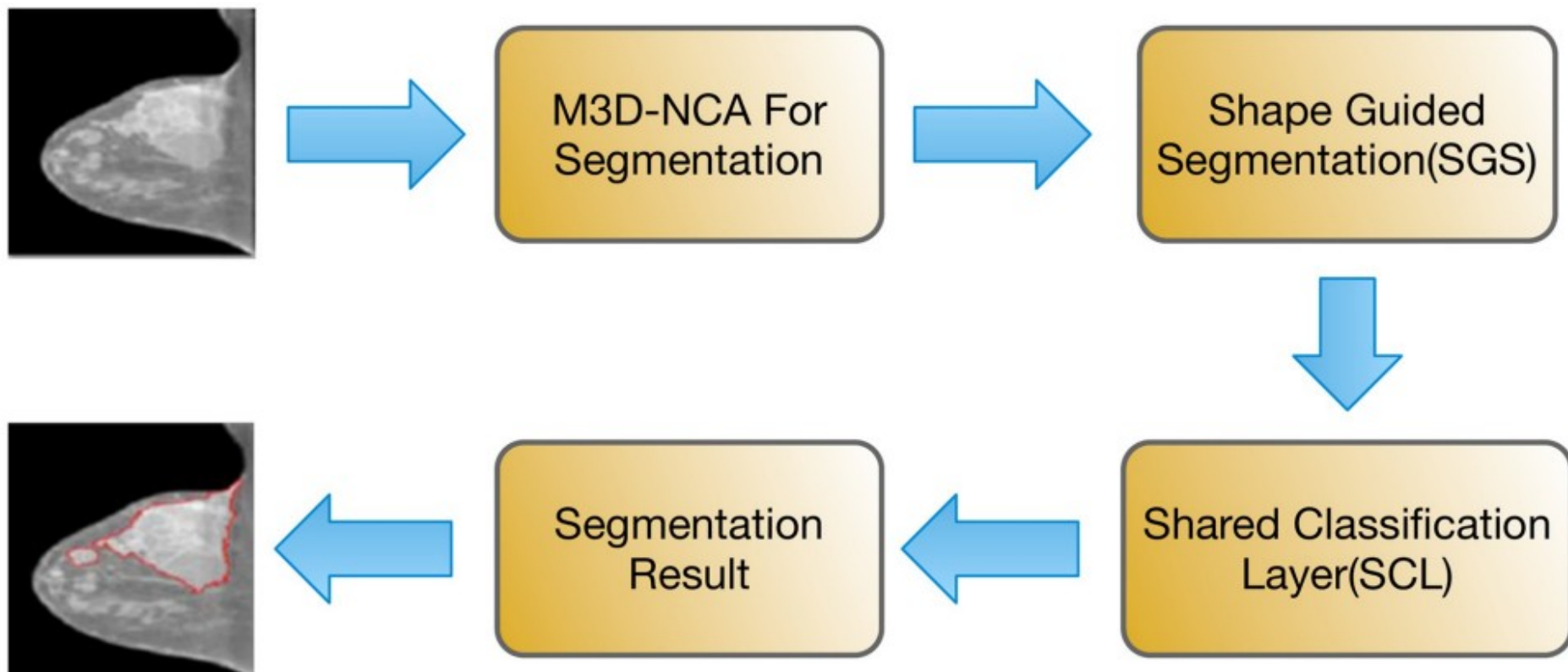


Fig 1. Flowchart of functional units.

<https://doi.org/10.1371/journal.pone.0309421.g001>



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

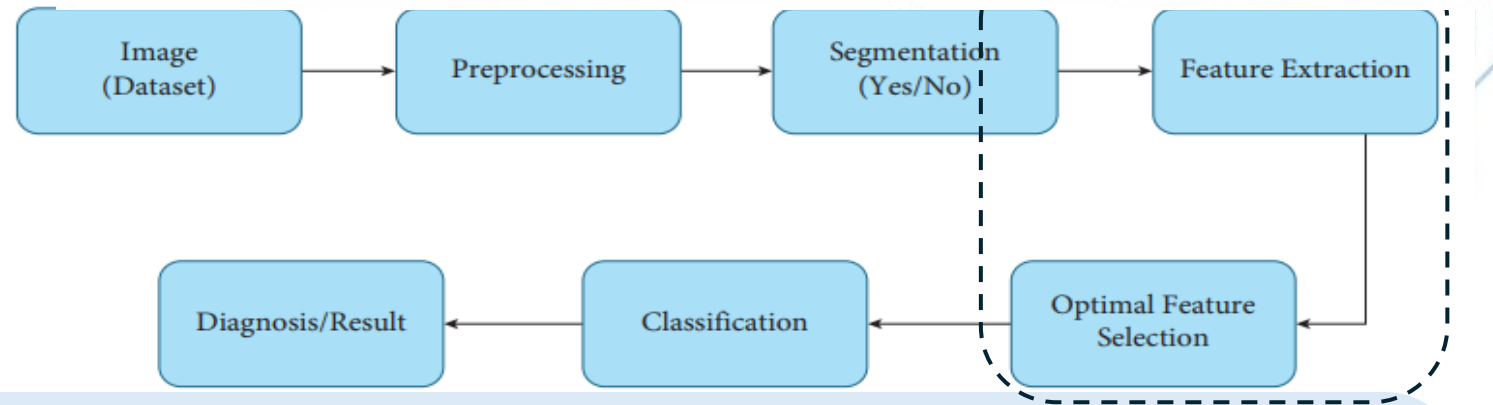
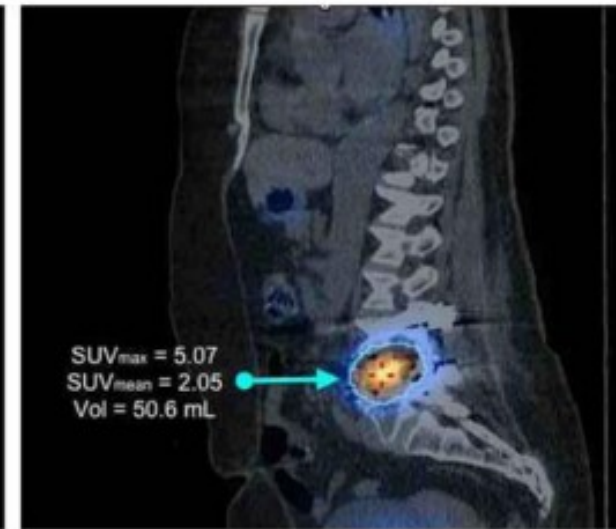
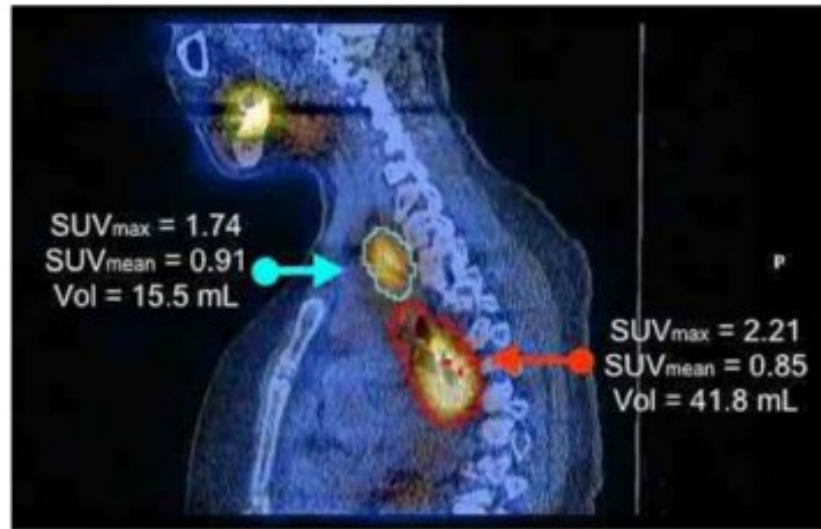
Тема 1

Роль ИИ в современной

Диагностика

Анализ медицинских изображений
(рентген, КТ, МРТ)

Выявление патологий
(опухоли, переломы)



Изображения содержат множество признаков: плотность тканей, контуры, текстуры, геометрические параметры, интенсивность сигнала. Извлечение признаков включает анализ этих характеристик, чтобы определить, какие именно аспекты изображения (форма, текстура и т.п.) наиболее значимы для диагностики и сократить количество ненужных деталей.



ФГБНЦ
ВНИИ
ОБРАЗОВАНИЯ
ИМ. Н.Н. БЛОХИНА

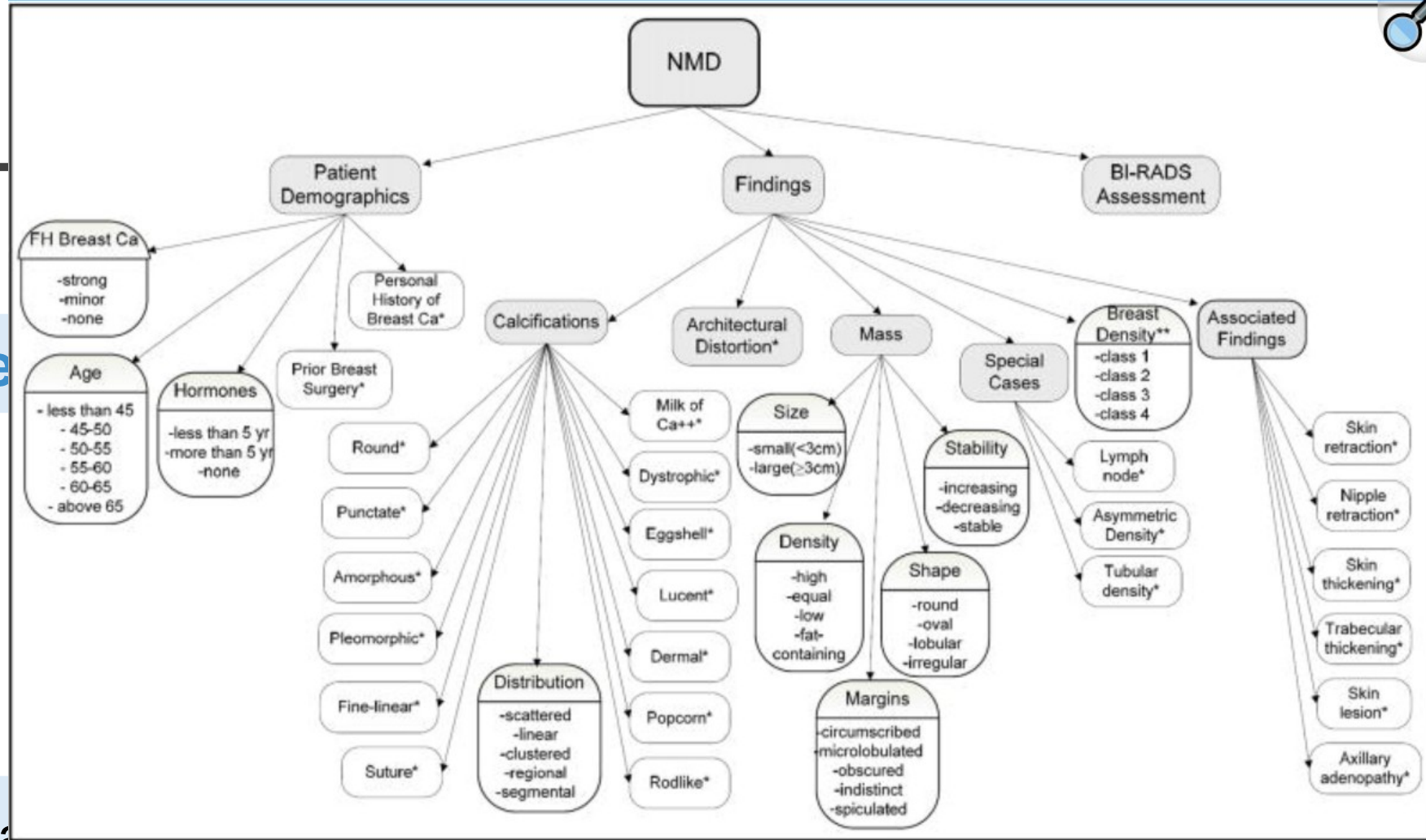
Роль ИИ в современной медицине

Диагностика

Анализ медицинских изображений
(рентген, КТ, МРТ)

Выявление патологий
(опухоли, переломы)

На этапе классификации патологий, определяется, является ли обнаруженное образование злокачественным или доброкачественным. После классификации предоставляется результат диагностики, вероятность наличия заболевания или рекомендации для дальнейшего обследования.





ФГБНЦ
ВНИИ
ОНКОЛОГИИ
им.Н.Н.Блохина

7

Роль ИИ в совр

Диагностика

Анализ медицинских
изображений
(рентген, КТ, МРТ)

Выявление патологий
(опухоли, переломы)



На этапе классификации принимается решение о наличии или отсутствии патологий, определяется, является ли обнаруженное образование злокачественным или доброкачественным. После классификации предоставляется результат диагностики, вероятность наличия заболевания или рекомендации для дальнейшего обследования.



НМИЦ
ОНКОЛОГИИ
им. Н.Н. Блохина

Роль ИИ в сов

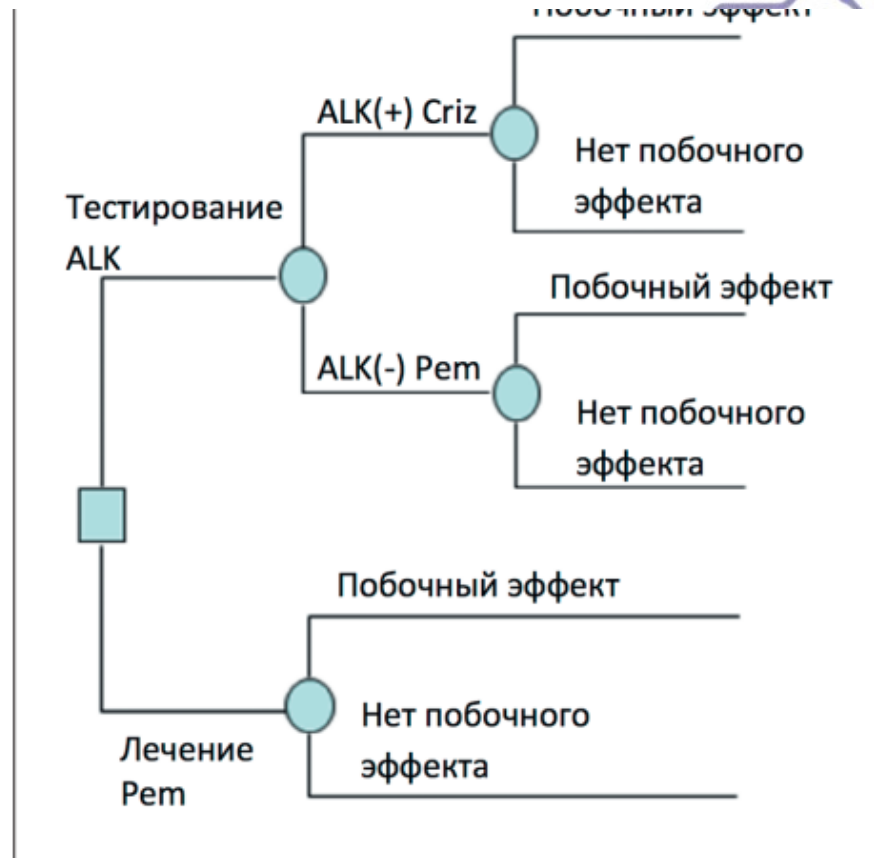
Лечение

Подбор терапии,
хирургические
роботы

Персонализиро-
ванная медицина

Анализ решений в медицине: принципы
построения дерева решений

С.Ч. Джалалов, Д.Х. Джалалова, Д.С. Хоч



Древо решений; сравнение частоты побочных
эффектов при таргетной терапии и при лечении без
предварительного генетического тестирования.

Примечания:

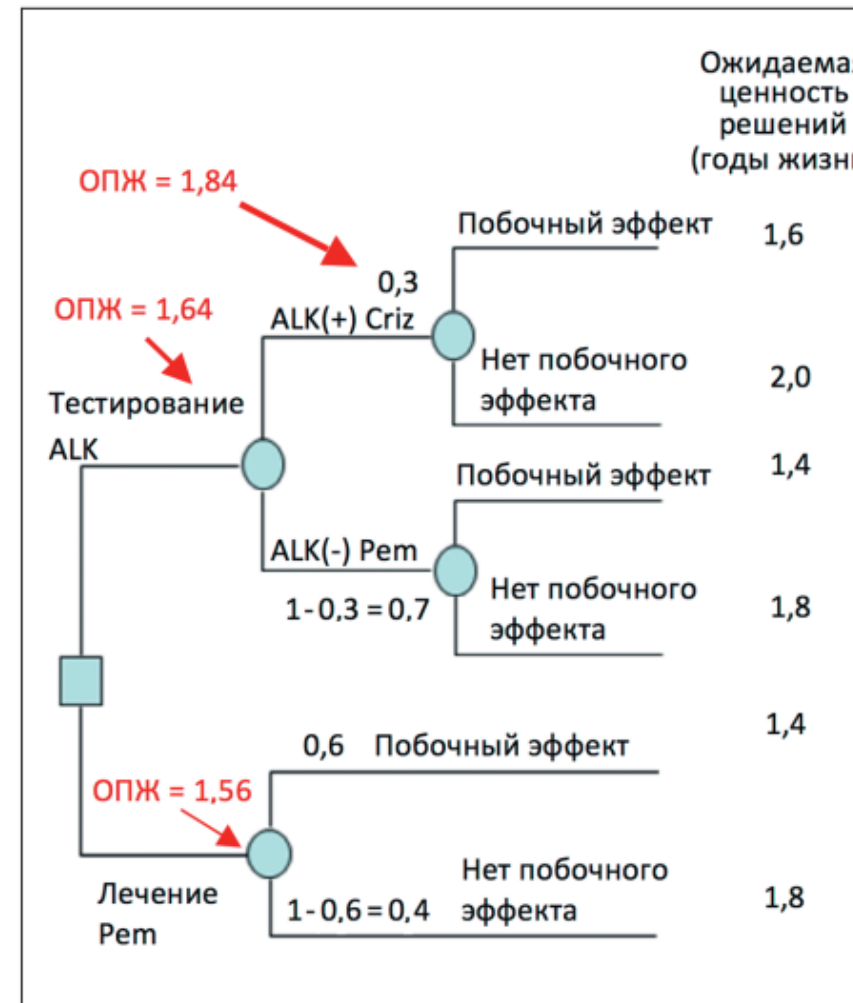
Pem – Pemetrexed

Criz – Crizotinib

ALK(+) – наличие мутации

ALK(-) – отсутствие мутации

Источник: собственная разработка авторов.



Определение вероятности событий и расчет
ожидаемой ценности решений

Тема 1.

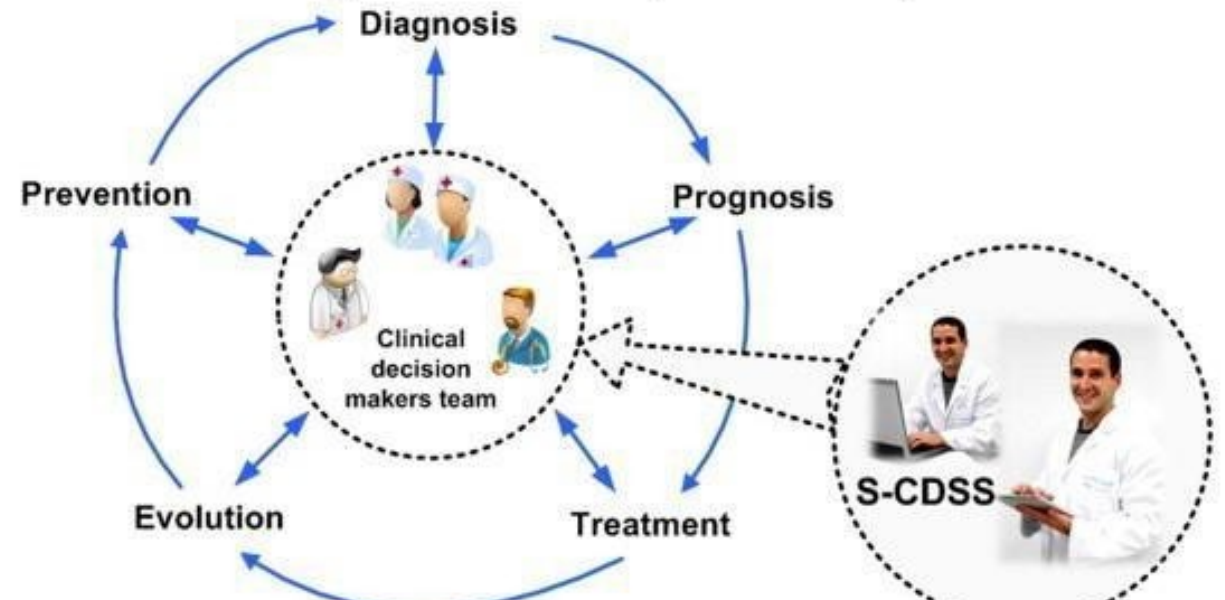
Роль ИИ в современной медицине

Лечение

Подбор терапии,
хирургические
роботы

Персонализиро-
ванная медицина

Clinical Decision Support Systems (CDSS)



Мониторинг состояния пациентов и диагностика

CDSS (Clinical Decision Support System):

- интеграция с электронными медицинскими картами,
- обработка сигналов от мониторов,
- предупреждение о критических показателях (ЧСС, давление),
- рекомендации по лекарствам,
- использование ML-моделей (регрессия, классификация) для прогноза развития осложнений и т.п..



ФГБУ
НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Роль ИИ в современной медицине

Прогнозирование

Риск развития заболеваний,
исходы госпитализаций

Эпидемиологические
прогнозы

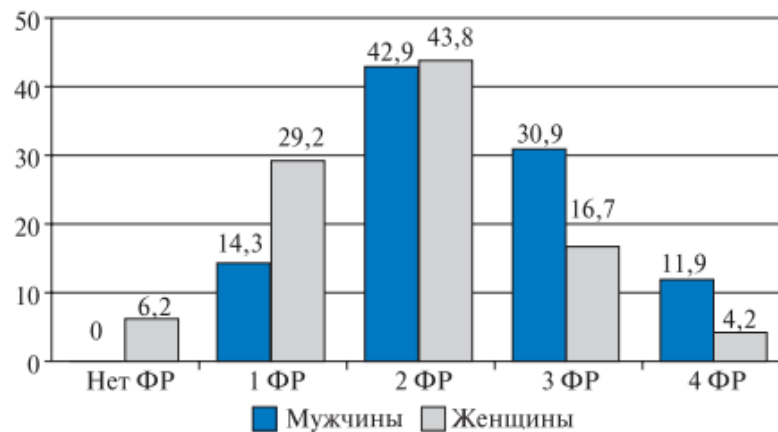


Рис. 1. Частота факторов риска среди мужчин и женщин, %.

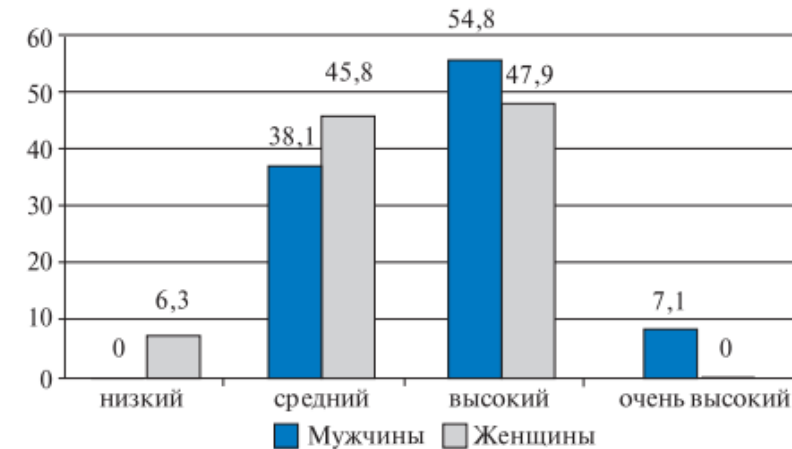


Рис. 2. Градация сердечно-сосудистого риска среди мужчин и женщин с впервые диагностированной АГ по критериям Национальных рекомендаций по диагностике и лечению АГ, %.



ФГБУ
НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема

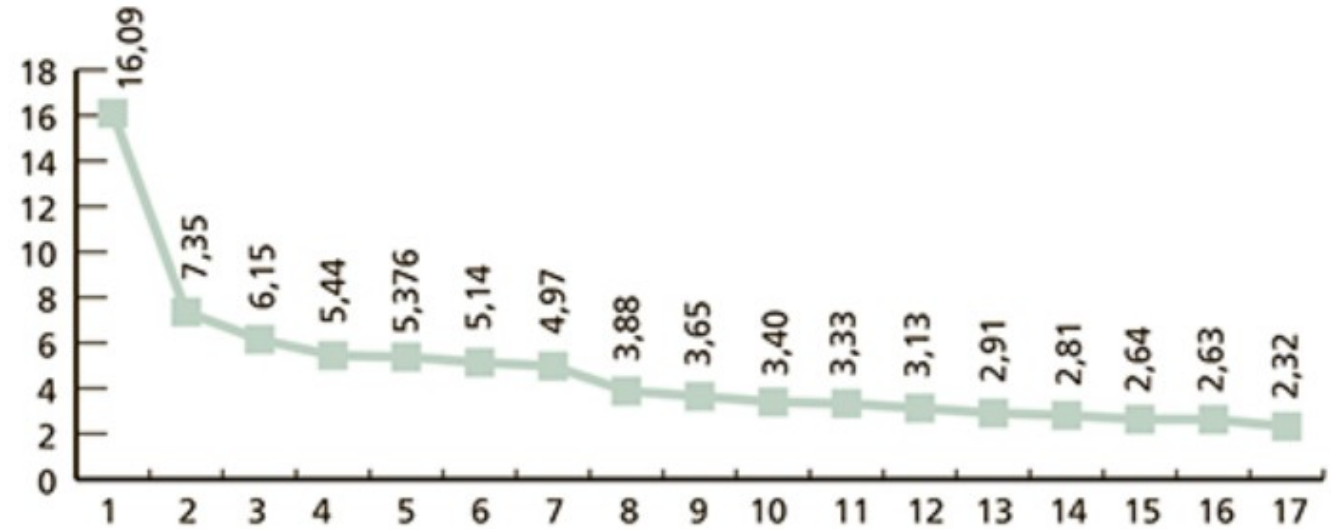
Роль ИИ в современно

Прогнозирование

Риск развития заболеваний,
исходы госпитализаций

Эпидемиологические
прогнозы

Матрица весовых нагрузок для сформированных факторов



1. Фактор I «сосудистый фактор»
2. Фактор II «распространенность туберкулезного процесса»
3. Фактор III «систолическая функция сердца»
4. Фактор IV «ремоделирование сердца»
5. Фактор V «состояние легочной функции»
6. Фактор VI «центральная гемодинамика»
7. Фактор VII «суточный индекс САД и ДАД»
8. Фактор VIII «форма туберкулеза; внелегочная локализация туберкулеза»
9. Фактор IX «осложнения течения туберкулеза»
10. Фактор X «фаза туберкулезного процесса, распад легочной ткани»
11. Фактор XI «лекарственная устойчивость микобактерий туберкулеза»
12. Фактор XII «режим химиотерапии»
13. Фактор XIII «насосная функция сердца»
14. Фактор XIV «скорость утреннего подъема САД»
15. Фактор XV «транспульмональный градиент»
16. Фактор XVI «ОФВ₁/ФЖЕЛ»
17. Фактор XVII «размеры правого желудочка»



НМИЦ
ОНКОЛОГИИ
им. Н.Н.Блохина

Тема 1.1

Роль ИИ в современной медицине

Источники и литература по теме «Управление»

Нормативные документы:

- Национальные стандарты по искусственному интеллекту в здравоохранении (22 ГОСТа)
- Приказ Минздрава России от 31.07.2020 № 785н «Об утверждении Требований к организации и проведению внутреннего контроля качества и безопасности медицинской деятельности»

Научные публикации и обзоры:

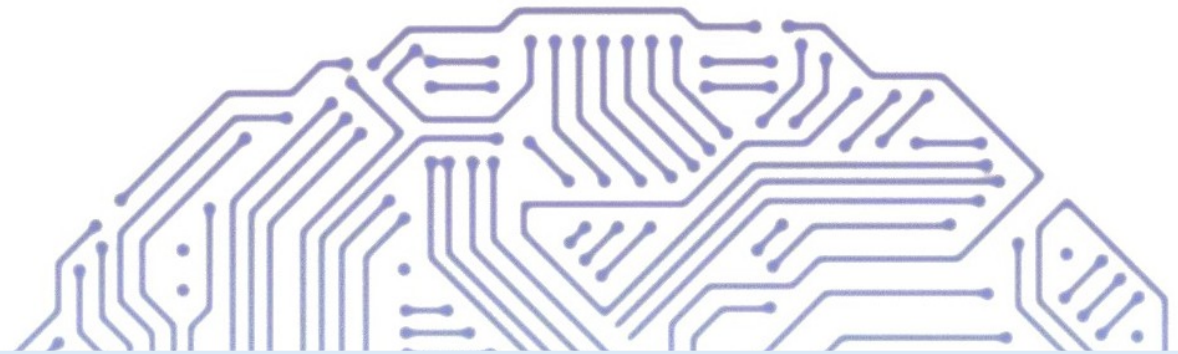
1. Остроухова Н.Г., Гажева А.В., Бурковская Ю.В. «Искусственный интеллект в управлении здравоохранением: эффективные решения для будущего: экспертный обзор». – М.: ГБУ «НИИОЗММ ДЗМ», 2025. – 27 с.
2. Кочетков А.Е. и др. «Опыт многопрофильного стационара по внедрению мероприятий организационного развития с применением технологий искусственного интеллекта». Журнал им. Н.В. Склифосовского «Неотложная медицинская помощь». 2025;14(3):652-658.

Интернет-ресурсы:

- НИИ организации здравоохранения и медицинского менеджмента (niiioz.ru) – раздел новостей и публикаций
- Официальный сайт Минздрава РФ (minzdrav.gov.ru)



Тема 1.1



Роль ИИ в современной медицине

Уровень	Задачи	Примеры
Стратегическое управление	Планирование, прогнозирование, распределение ресурсов	Анализ потребности в кадрах, маршрутизация потоков пациентов, прогноз заболеваемости
Операционное управление	Организация работы медорганизации	Управление расписанием, логистика, закупки
Административные процессы	Снижение нагрузки на персонал	Голосовые помощники, автоматизация документооборота,



НМИЦ
ОНКОЛОГИИ
им. Н.Н.Блохина

Тема 1.1

Роль ИИ в современной медицине

Пример. Саммаризация электронных медкарт (Москва)

- ИИ-агент анализирует весь массив данных электронной медкарты за многие годы
- Извлекает информацию, релевантную цели текущего приема (история обращений, результаты исследований, назначения, вызовы скорой)
- Врач получает персонализированную тактику лечения, не тратя время на поиск в огромном объеме данных

<https://niioz.ru/news/ii-servisy-pomogayut-stolichnym-vracham-v-lechenii-i-ekonomit-vremya-na-prieme/>

ИИ в управлении — это не только экономия, но и высвобождение времени врача для непосредственной работы с пациентом

Российская практика

- 20–30% — сокращение времени диагностики при использовании ИИ-систем
- 40% — снижение нагрузки на сотрудников колл-центров благодаря голосовым помощникам
- 83% — совпадение диагнозов, предложенных ИИ, с заключительным диагнозом врача

<https://minzdravri.ru/minzdrav-ingushetii-pristupil-k-testirovaniyu-ii-sistem-v-mediczine/>



НМИЦ
ОНКОЛОГИИ
им. Н.Н.Блохина

Тема 1.1

Роль ИИ в современной медицине

Москва: платформа «МосМедИИ»

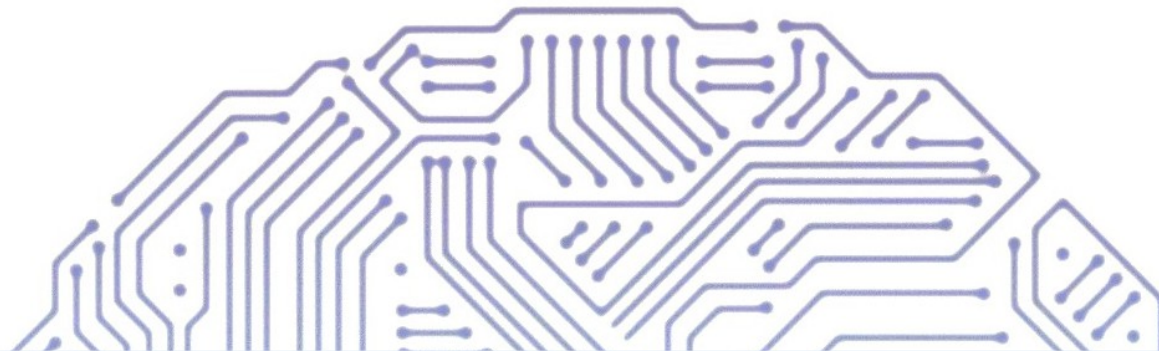
- 22 национальных стандарта для медицинских ИИ-сервисов разработано на базе московского опыта
- 74 региона подключены к платформе
- Более 2000 медицинских организаций используют платформу
- Более 8 миллионов исследований проанализировано

Стандарты регулируют: применение нейросетей для организации и управления, дистанционный мониторинг, прогнозную аналитику, системы поддержки принятия врачебных решений



НМИЦ
ОНКОЛОГИИ
им. Н.Н. Блохина

Тема 1.1



Роль ИИ в современной медицине

Три макротренда в управлении здравоохранением (по итогам стратегической сессии Минздрава России)

«Стратегическое управление ИИ»

Создание Центра ИИ для стратегического управления

Координация внедрения на федеральном уровне

«Универсальные клинические помощники на базе генеративного ИИ»

Интеграция ИИ в повседневную практику врача

Поддержка принятия решений

«ИИ как источник операционной эффективности»

Цифровые двойники медицинских учреждений (пилотные проекты)

Оптимизация всех видов ресурсов

ведущие эксперты из РНИМУ им. Пирогова, AIRI, РЭУ им. Плеханова, Центра диагностики и телемедицины, НМИЦ нейрохирургии им. Бурденко, МФТИ.



НМИЦ
ОНКОЛОГИИ
им. Н.Н. Блохина

Тема 1.1

обзор открытых наборов данных, доступных в РФ

1. **Kaggle Datasets** <https://www.kaggle.com/datasets> (частично доступный)
2. **The Cancer Imaging Archive (TCIA)** <https://www.cancerimagingarchive.net/>
3. **ChestX-ray14 NIH Chest X-rays**
4. **LUNA16** <https://luna16.grand-challenge.org/>
5. **ISIC Archive (International Skin Imaging Collaboration)** <https://www.isic-archive.com/>
6. **DRIVE: Digital Retinal Images for Vessel Extraction**
<http://www.isi.uu.nl/Research/Databases/DRIVE/>
7. **BRATS (Brain Tumor Segmentation)** <https://www.med.upenn.edu/cbica/brats2020/data.html>
8. **COVID-19 Image Data Collection** <https://github.com/ieee8023/covid-chestxray-dataset>
9. **APOLLO Scape: Medical Image Segmentation Dataset** <https://apolloscape.auto/scene.html>
10. **Open Access Series of Imaging Studies (OASIS)** <http://www.oasis-brains.org/>

*** Как использовать эти наборы данных

- **Зарегистрируйтесь** на платформах, где находятся наборы данных.
- **Скачайте данные** на свой компьютер и ли в облако. В некоторых случаях данные могут быть большими.
- **Обработка данных:** Многие медицинские изображения сохраняются в формате **DICOM**, поэтому потребуется специальное ПО или библиотеки, такие как **pydicom** (в Python), для обработки этих файлов.
- **Используйте библиотеку для обработки изображений:** Библиотеки, такие как **OpenCV**, **SimpleITK**, **scikit-image** и **TensorFlow**, помогут вам выполнить предобработку, анализ и построение моделей машинного обучения для работы с изображениями.



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

обзор открыт

THE CANCER IMAGING ARCHIVE										
Submit Your Data			Access The Data			Help				
About Us			Research Activities			News				
Collection	Cancer Type	Locations	Species	Subjects	Data Types	Supporting Data	Access	Status	Updated	
CPTAC-HNSCC	Head and Neck Cancer	Head-Neck	Human	137	CT, PT, MR, RTSTRUCT, Histopathology	Clinical, Genomics, Proteomics, Microarray	Limited	Ongoing	2024/11/18	
APOLLO-5	Breast Cancer, Bladder Urothelial Carcinoma, Cholangiocarcinoma, Colon	Abdomen, Adrenal, Bile duct, Bladder, Brain, Breast, Chest, Chest-	Human	437	MG, MR, NM, PT, CT, US, XA		Limited	Ongoing	2024/10/31	

[The Cancer Imaging Archive \(TCIA\)](https://www.cancerimagingarchive.net/), финансируемый Национальным институтом онкологии США (NCI) — это место хранения в открытом доступе радиологических и гистопатологических снимков, в основном в формате DICOM, представляющих 21 тип рака. Коллекции данных разделены по заболеваниям (рак груди и так далее), модальности снимков (КТ, МРТ и так далее) и задачам исследований. Архив также дополняет снимки имеющимися вспомогательными данными, такими как подробности лечения, результаты лечения и геномика.

<https://www.cancerimagingarchive.net/browse-collections/>

С текущего начала слайда

Онлайн-презентация

Произвольное слайд-шоу

Начать слайд-шоу

Настройка слайд-шоу

Скрыть слайд

Настройка времени

Записать слайд-шоу

Настройка

☒ Воспроизвести закадровый текст

☒ Использовать время показа слайдов

☒ Показать элементы управления проигрывателем

Монитор: Автоматически

☐ Режим докладчика

Мониторы

44



45



46



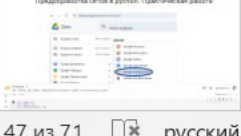
47



48



49



The Cancer Imaging Archive (TCIA)

<https://www.cancerimagingarchive.net/>

[The Cancer Imaging Archive \(TCIA\)](#), финансируемый Национальным институтом онкологии США (NCI) — это место хранения в открытом доступе радиологических и гистопатологических снимков, в основном в формате DICOM, представляющих [21 тип](#) рака. Коллекции данных разделены по заболеваниям (рак груди и так далее), модальности снимков (КТ, МРТ и так далее) и задачам исследований. Архив также дополняет снимки имеющимися вспомогательными данными, такими как подробности лечения, результаты лечения и геномика.

Submit Your Data

Access The Data

Help

CANCER IMAGING ARCHIVE

About Us

Research Activities

News

Collection	Cancer Type	Locations	Species	Subject	Data Types	Supporting Data	Access	Status	Updated
CPTAC-HNSCC	Head and Neck Cancer	Head-Neck	Human	137	CT, PT, MR, RTSTRUCT, Histopathology	Clinical, Genomics, Proteomics, Microarray	Limited	Ongoing	2024/11/18
APOLLO-5	Breast Cancer, Bladder Urothelial Carcinoma, Cholangiocarcinoma, Colon	Abdomen, Adrenal, Bile duct, Bladder, Brain, Breast, Chest, Chest-	Human	437	MG, MR, NM, PT, CT, US, XA		Limited	Ongoing	2024/10/31

<https://www.cancerimagingarchive.net/browse-collections/>

обзор открытых наборов данных, доступных в РФ

Загрузка данных в Google Colab

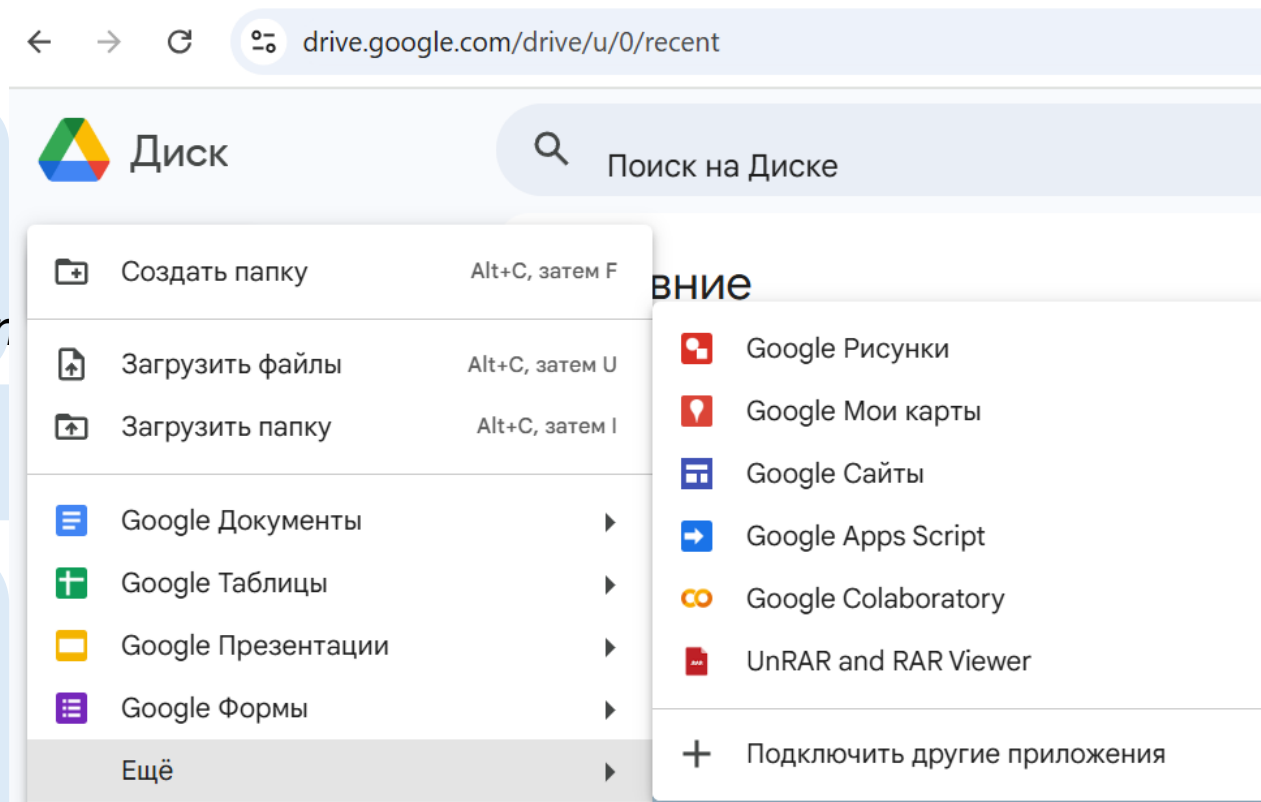
- Напрямую с таких платформ, как Kaggle, TCIA или с других источников, с помощью их API или командной строки
Пример загрузки с TCIA в блокноте Noise.ipynb

- Загрузка с локального компьютера:

```
from google.colab import files  
uploaded = files.upload()
```

- . Использование Google Drive. Можно подключить Google Drive, чтобы получить доступ к большим наборам данных

```
from google.colab import drive  
drive.mount('/content/drive')  
input_folder = '/content/drive/MyDrive/images'
```



Практическая работа. Загрузка и просмотр медицинских изображений

- Создадим запрос API для получения изображения коллекции. Данные будут получены и возвращаются в формате JSON

Данные JSON

SeriesInstanceUID:

Уникальный идентификатор серии изображений.

StudyInstanceUID:

Уникальный идентификатор исследования.

SeriesDescription:

Описание серии.

Modality: Тип модальности (например, CT, MRI, PET).

BodyPartExamined:

Область тела, подвергнутая исследованию.

ProtocolName: Название используемого протокола.

Manufacturer: Производитель оборудования.

The Cancer Imaging Archive (TCIA)

Файлы

Подключение к среде выполнения для просмотра файлов...

```
import requests
```

Пример запроса к TCIA API

The Cancer Imaging Archive (TCIA) предоставляет доступ к медицинским изображениям через API. Вы можете использовать библиотеку requests для загрузки данных из TCIA:

```
import requests

# Пример запроса к TCIA API
url = "https://services.cancerimagingarchive.net/services/v3/TCIA/query/getSeries?Collection=TCGA-BRCA"
response = requests.get(url)

# Проверка ответа
if response.status_code == 200:
    print("Данные получены успешно")
else:
    print("Ошибка получения данных")
```

Данные получены успешно



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практическая работа. Загрузка и просмотр медицинских изображений

- Обработка данных JSON. Получаем заархивированные данные.

The Cancer Imaging Archive (TCIA)

TCIA - The Cancer Imaging Archive

URL: https://services.cancerimagingarchive.net/services/v3/TCIA/query/getImage?SeriesInstanceUID={series_instance_uid}

File: images (1).zip
120 Б • Готово

Files: Noise.ipynb

File Edit View Insert Environment Tools Help Changes saved

Files: .., D:, ronc, images, images.zip, блохина, sample_data

Code Editor: + Код + Текст

использованием SeriesInstanceUID

```
import os
import requests

# Пример использования SeriesInstanceUID
series_instance_uid = "1.3.6.1.4.1.14519.5.2.1.6279.6001.29880613728863345324697563017805357912" # Пример идентификатора

# URL для загрузки изображений
url = f"https://services.cancerimagingarchive.net/services/v3/TCIA/query/getImage?SeriesInstanceUID={series_instance_uid}"
response = requests.get(url, stream=True)

# Путь для сохранения ZIP-файла
output_path = "/content/images.zip"

# Создание необходимых папок, если их нет
folder_path = os.path.dirname(output_path)
os.makedirs(folder_path, exist_ok=True)

# Проверка ответа и сохранение изображения
if response.status_code == 200:
    print(f"Сохранение файла в {output_path}")
    with open(output_path, "wb") as file:
        for chunk in response.iter_content(chunk_size=128):
            file.write(chunk)
    print(f"Изображения успешно загружены и сохранены как {output_path}")
else:
    print("Ошибка при загрузке изображений")
```



Файлы



bin
boot
content
images

1-222.dcm
10-231.dcm
11-232.dcm
12-233.dcm
13-234.dcm
14-235.dcm
15-236.dcm
16-237.dcm
17-238.dcm
18-239.dcm
19-240.dcm
2-223.dcm
20-241.dcm
21-242.dcm
22-243.dcm
23-244.dcm
24-245.dcm



+ Код + Текст

✓
7 сек.

Пример использования SeriesInstanceUID

series_instance_uid = "1.3.6.1.4.1.14519.5.2.1.6450.4002.265963154110062013582166845425" # Пример идентификатора

URL для загрузки изображений

url = f"[https://services.cancerimagingarchive.net/services/v3/TCIA/query/getImage?SeriesInstanceUID={series_instance_uid}](\"https://services.cancerimagingarchive.net/services/v3/TCIA/query/getImage?SeriesInstanceUID={series_instance_uid}\")"

response = requests.get(url, stream=True)

Путь для сохранения ZIP-файла

output_path = "/content/images.zip"

Создание необходимых папок, если их нет

folder_path = os.path.dirname(output_path)

os.makedirs(folder_path, exist_ok=True)

Проверка ответа и сохранение изображения

if response.status_code == 200:

print(f"Сохранение файла в {output_path}")

with open(output_path, "wb") as file:

for chunk in response.iter_content(chunk_size=128):

file.write(chunk)

print(f"Изображения успешно загружены и сохранены как {output_path}")

else:

print("Ошибка при загрузке изображений")



Сохранение файла в /content/images.zip

Изображения успешно загружены и сохранены как /content/images.zip

✓
0 сек.

import zipfile

Распаковка загруженного ZIP-файла

with zipfile.ZipFile("images.zip", 'r') as zip_ref:

zip_ref.extractall("images") # Папка для распакованных изображений

print("Файлы успешно распакованы")



Файлы успешно распакованы



НМИЦ
ОНКОЛОГИИ
им. Н.Н. Блохина

Тема 1.1

Практическая работа. Загрузка и просмотр медицинских изображений

- Установка библиотек и пробное изображение

1. **cv2**

Библиотека OpenCV для компьютерного зрения и обработки изображений. Используется для чтения, обработки, анализа изображений и видео.

3. **numpy**

Библиотека для работы с многомерными массивами и выполнения численных вычислений.

4. **pydicom**

- Библиотека для работы с DICOM-файлами (медицинские изображения). Позволяет загружать, анализировать, модифицировать и сохранять файлы в формате DICOM.

5. **matplotlib.pyplot**

- Модуль библиотеки Matplotlib, предназначенный для визуализации данных. Используется для построения графиков, отображения изображений и других визуальных элементов.

6. **skimage.exposure**

- Модуль из библиотеки **scikit-image** для обработки изображений.

Включает функции для управления экспозицией

The Cancer Imaging Archive (TCIA) *Пример загрузки с TCIA в блокноте Noise.ipynb*

Для чтения DICOM файлов используйте pydicom, а для визуализации изображений — matplotlib.

```
!pip install pydicom
```



```
... Collecting pydicom
  Downloading pydicom-3.0.1-py3-none-any.whl.metadata (9.4 kB)
  Downloading pydicom-3.0.1-py3-none-any.whl (2.4 MB)
  2.4/2.4 MB 30.2 MB/s eta 0:00:00
Installing collected packages: pydicom
Successfully installed pydicom-3.0.1
```



ФГБНЦ
ВНИИ
ООНКОЛОГИИ
им.Н.Н.Блохина

Практическая работа с изображениями

- Установка библиотек и
загрузка изображения

Библиотека OpenCV для работы с изображениями. Используется для загрузки изображений и видео.

3. **numpy**

Библиотека для работы с многомерными численными вычислениями.

4. **pydicom**

- Библиотека для работы с DICOM-изображениями. Позволяет загружать и сохранять файлы в формате DICOM.

5. **matplotlib.pyplot**

- Модуль библиотеки Matplotlib, который используется для построения изображений и других визуализаций.

6. **skimage.exposure**

- Модуль из библиотеки **scikit-image**

Включает функции для управления экспозицией

```
import os
import pydicom
import matplotlib.pyplot as plt

# Папка с распакованными изображениями
dicom_folder = "images"

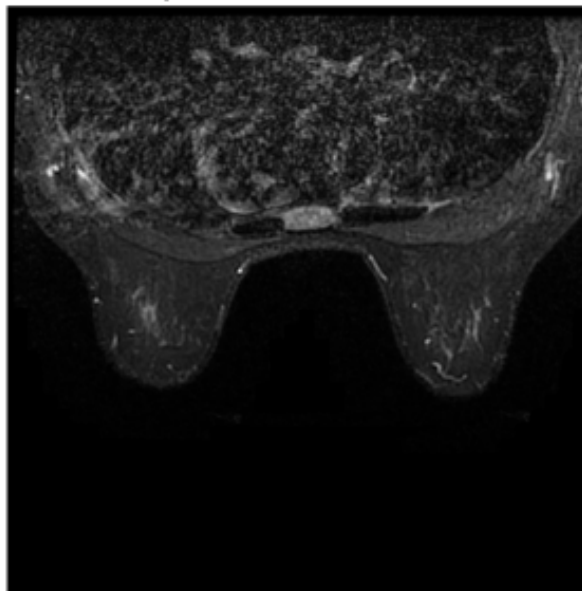
# Чтение всех DICOM-файлов в папке и визуализация первого изображения
for filename in os.listdir(dicom_folder):
    if filename.endswith(".dcm"):
        dicom_path = os.path.join(dicom_folder, filename)
        dicom_data = pydicom.dcmread(dicom_path)

        # Получение массива пикселей изображения
        image = dicom_data.pixel_array

        # Визуализация изображения
        plt.imshow(image, cmap="gray")
        plt.title(f"Пациент ID: {dicom_data.PatientID}")
        plt.axis("off")
        plt.show()
        break # Визуализируем только первое изображение
```



Пациент ID: TCGA-BH-A0H5



СКИХ

Archive (TCIA)
окном Noise.ipynb



Практика изображений Medical S (MSD)

Описание
Это колл
изображе
сегмента
таких как

drive.google.com/drive/folders/1HqEgzS8BV2c7xYNrZdEAnrHk7osJJ--2

Диск

Поиск на Диске

+ Создать

- Личный
- Мой диск
- Компьютеры
- Доступные мне
- Недавние
- Помеченные
- Спам
- Корзина
- Хранилище

Использовано 10,24 ГБ из 15 ГБ

Доступные мне > MSD

выбрано: 1

Завершите оформление подписки Google One Купите больше места для Диска, Gmail и Фото. [Купить](#)

Название	Послед...
Task08_HepaticVessel.tar	2 февр. 2020 г.
Task07_Pancreas.tar	
Task03_Liver.tar	2 февр. 2020 г.
Task01_BrainTumour.tar	14 янв. 2020 г.
Task10_Colon.tar	6 авг. 2018 г.

Открыть с помощью

Скачать

Переименовать Ctrl+Alt+E

Создать копию Ctrl+C Ctrl+V

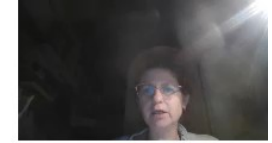
Поделиться

Систематизировать

Информация о файле

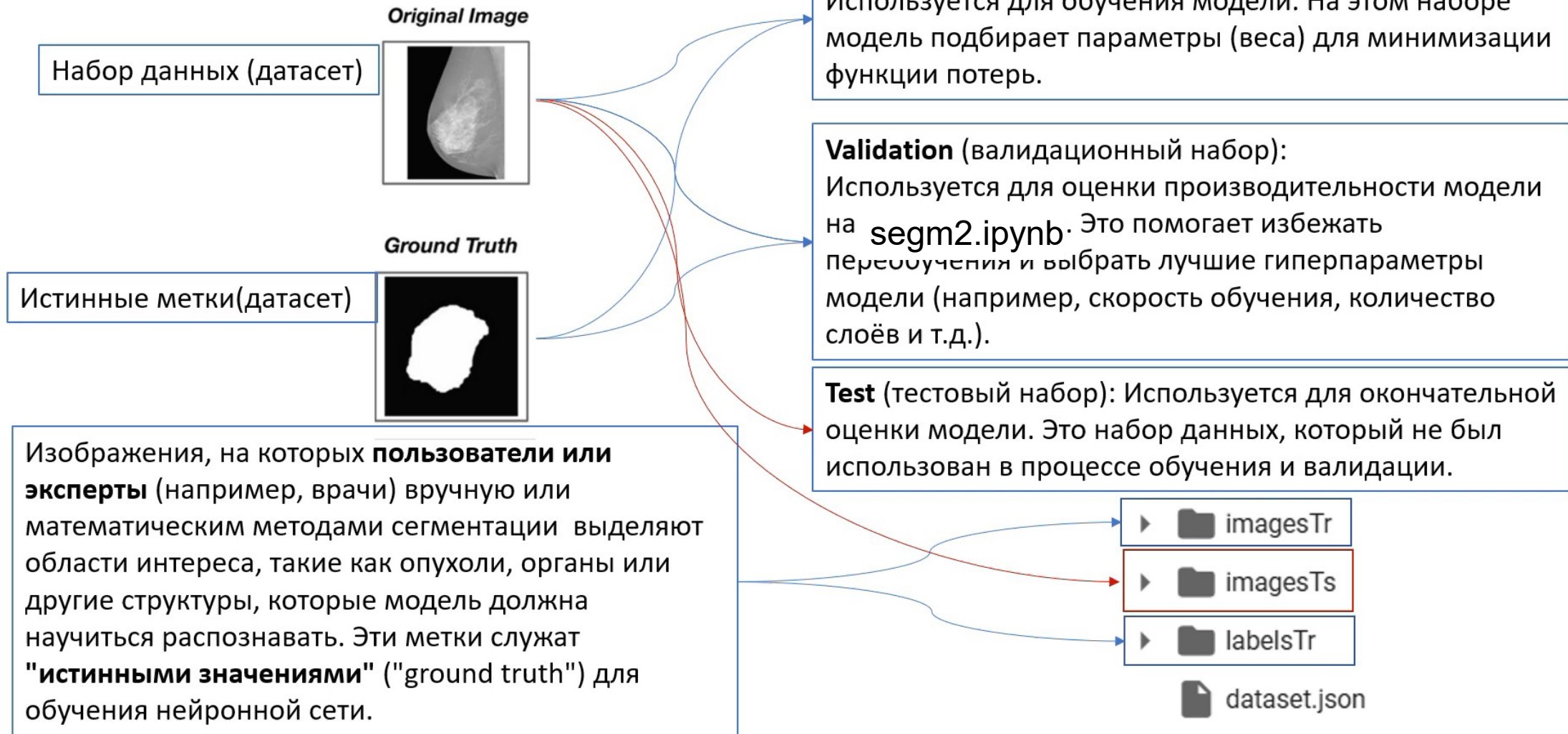
Удалить Delete

Пожаловаться или заблокировать



6. Сегментация на основе машинного обучения

6.1. Подготовка данных для обучения

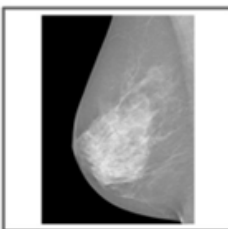


Тема 1.1

Практическая работа. Загрузка и просмотр медицинских изображений

Набор данных (датасет)

Original Image



Ground Truth

Истинные метки(датасет)



Изображения, на которых **пользователи или эксперты** (например, врачи) определяют как **"истинные значения"** ("ground truth") для обучения нейронной

сети

► imagesTr

► imagesTs

► labelsTr

Medical Segmentation Decathlon (MSD)

segm2.ipynb resource2.ipynb

JSON (JavaScript Object Notation) — это формат текстового файла, который используется для хранения и обмена структурированными данными.

Данные хранятся в виде пар "ключ-значение". JSON поддерживает вложенные структуры, что позволяет описывать сложные данные (например, списки, массивы, объекты).

Training (обучающий набор):

Validation (валидационный

набор):

Test (тестовый набор)

Тема 1.1

Практическая работа. Загрузка и п изображений

JSON (JavaScript Object Notation)

1. информация о тренировочных данных, валидационных данных и тестовых данных.
2. Указание путей к изображениям, меткам (Ground Truth), размерам данных и др.
3. Хранение гиперпараметров модели (например, количество слоёв, размер шага обучения и т.д.).
4. Хранение метрик (точности, ошибки), параметров модели, истории обучения.
5. Списки файлов с изображениями и метками, которые используются в задачах классификации, сегментации и др.

```
json
{
  "name": "HepaticVessel",
  "description": "Hepatic Vessels and Tumour Segmentation",
  "modality": {
    "0": "CT"
  },
  "labels": {
    "0": "background",
    "1": "Vessel",
    "2": "Tumour"
  },
  "numTraining": 303,
  "numTest": 140,
  "training": [
    {
      "image": "./imagesTr/hepaticvessel_001.nii.gz",
      "label": "./labelsTr/hepaticvessel_001.nii.gz"
    },
    {
      "image": "./imagesTr/hepaticvessel_002.nii.gz",
      "label": "./labelsTr/hepaticvessel_002.nii.gz"
    }
  ],
  "test": [
    {
      "image": "./imagesTs/hepaticvessel_101.nii.gz"
    }
  ]
}
```



Тема 1.1

Практическая работа. Загрузка и просмотр медицинских изображений

JSON (JavaScript Object Notation)

1. Используем библиотеку gdown
2. Копируем ссылку на нужный файл и извлекаем ID

Medical Segmentation Decathlon
(MSD)

segm2.ipynb

resource2.ipynb

✓
11
сек.



```
!pip install gdown
```



```
Requirement already satisfied: gdown in /usr/local/lib/python3.10/dist-packages (5.2.0)  
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from gdown) (4.12.3)  
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from gdown) (3.16.1)
```

<https://drive.google.com/file/ink>



UuUxL/view?usp=drive_l

ID



Segm2.ipynb ☆

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка Изменения сохранены

Файлы



П {x}

И

1.

2.

3.

- ..
- Task08_HepaticVessel
 - imagesTr
 - imagesTs
 - labelsTr
 - dataset.json
 - sample_data
 - my_file_name.tar

+ Код + Текст

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-pa
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /usr/local/lib/python3.10/dis

```
[2] import gdown

# ID файла с Google Диска
file_id = '1qVrpV7vmhIsUxFiH189LmAn0ALbAPrgS'

# Формируем URL для скачивания
url = f'https://drive.google.com/uc?id={file_id}'

# Скачиваем файл в директорию /content/
file_path = '/content/my_file_name.tar'
gdown.download(url, file_path, quiet=False)

# Распаковываем архив .tar
import tarfile
with tarfile.open(file_path, 'r') as tar:
    tar.extractall(path='/content/')
print("Архив распакован в директорию /content/")
```



Downloading...

From (original): <https://drive.google.com/uc?id=1qVrpV7vmhIsUxFiH189LmAn0ALbAPrgS>From (redirected): <https://drive.google.com/uc?id=1qVrpV7vmhIsUxFiH189LmAn0ALbAPrgS&cc>

To: /content/my_file_name.tar

100%|██████████| 9.35G/9.35G [02:04<00:00, 75.2MB/s]

Архив распакован в директорию /content/



Практика изображений

1. Исполнить код
2. Копировать
3. Исполнить код
расширением
4. Открыть файл
содержимое

Файлы



- ..
- Task08_HepaticVessel
 - imagesTr
 - imagesTs
 - labelsTr
 - dataset.json
- sample_data
- my_file_name.tar

+ Код + Текст

0 сек.

```
import json

# Путь к файлу JSON
json_file_path = '/content/Task08_HepaticVessel/dataset.json'

# Загружаем содержимое JSON файла
with open(json_file_path, 'r') as f:
    dataset_info = json.load(f)

# Просмотрим содержимое JSON
print(json.dumps(dataset_info, indent=4)) # Красиво выводим содержимое
```

```
{
  "name": "HepaticVessel",
  "description": "Hepatic Vessels and Tumour Segmentation",
  "reference": "Memorial Sloan Kettering Cancer Center",
  "licence": "CC-BY-SA 4.0",
  "release": "1.1 14/08/2018",
  "tensorImageSize": "3D",
  "modality": {
    "0": "CT"
  },
  "labels": {
    "0": "background",
    "1": "Vessel",
    "2": "Tumour"
  },
  "numTraining": 303,
  "numTest": 140,
  "training": [
    {
      "image": "./imagesTr/hepaticvessel_001.nii.gz",
      "label": "./labelsTr/hepaticvessel_001.nii.gz"
    },
    {
      "image": "./imagesTr/hepaticvessel_002.nii.gz",
      "label": "./labelsTr/hepaticvessel_002.nii.gz"
    }
  ]
}
```

КАКИХ

ecathlon

ce2.ipynb





ФНЦ
ООНКОЛОГИИ
им. Н.Н. Блохина

Тема 1.1

Практическая работа. Загрузка и изображений

1. Используем библиотеку gdown
2. Копируем ссылку на нужный файл и извлекаем ID
3. Используя библиотеку gdown, ID, скачиваем и распаковываем датасет в папку /content.
4. Открываем и просматриваем json для вывода содержания для дальнейшей работы.
5. Выводим для просмотра 20 случайных объектов с метками из обучающего набора.

```
import random
import nibabel as nib
import matplotlib.pyplot as plt
import numpy as np
import os

# Папки с изображениями и метками
images_dir = '/content/Task08_HepaticVessel/imagesTr/'
labels_dir = '/content/Task08_HepaticVessel/labelsTr/'

# Фильтруем доступные файлы
available_files = sorted([f for f in os.listdir(images_dir) if not f.startswith('.') and f.endswith('.nii.gz')])
images = [os.path.join(images_dir, file) for file in available_files]
labels = [os.path.join(labels_dir, file.replace('imagesTr', 'labelsTr')) for file in available_files]

# Проверяем количество файлов
print(f"Обнаружено {len(images)} изображений и {len(labels)} меток.")

# Случайно выбираем 20 изображений
random_indices = random.sample(range(len(images)), min(20, len(images)))

# Цветовая карта для меток
cm = plt.cm.colors.ListedColormap(['black', 'green', 'red']) # Цвета: фон, сосуды, опухоли
bounds = [0, 1, 2, 3]
norm = plt.cm.colors.BoundaryNorm(bounds, cm.N)

# Визуализация изображений и меток
for idx in random_indices:
    try:
        # Загружаем изображение и метку
        image = nib.load(images[idx]).get_fdata()
        label = nib.load(labels[idx]).get_fdata()

        # Поиск срезов с ненулевыми метками
        non_empty_slices = [z for z in range(label.shape[2]) if np.unique(label[:, :, z]).size > 1]

        if not non_empty_slices:
            print(f"Все срезы в {images[idx]} содержат только фон.")
            continue

        # Выбираем случайный срез с ненулевыми метками
        slice_index = random.choice(non_empty_slices)
        print(f"Выбран срез {slice_index} для изображения {images[idx]}")

        plt.figure(figsize=(10, 5))

        # Отображаем изображение
        plt.subplot(1, 2, 1)
        plt.imshow(image[:, :, slice_index], cmap='gray')
        plt.title(f'Изображение (Slice: {slice_index})')
        plt.axis('off')

        # Отображаем метки
        plt.subplot(1, 2, 2)
        plt.imshow(label[:, :, slice_index], cmap=cm, norm=norm)
        plt.title('Метки')
        plt.axis('off')

        plt.show()

    except Exception as e:
        print(f"Ошибка при обработке файла {images[idx]}: {e}")
```



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практическая работа. Загрузка и просмотр медицинских изображений

```
# Цветовая карта для  
меток  
cmap =  
plt.cm.colors.ListedColo  
rmap(['black', 'green',  
'red']) # Цвета: фон,  
сосуды, опухоли  
bounds = [0, 1, 2, 3]  
norm =  
plt.cm.colors.BoundaryNo  
rm(bounds, cmap.N)
```

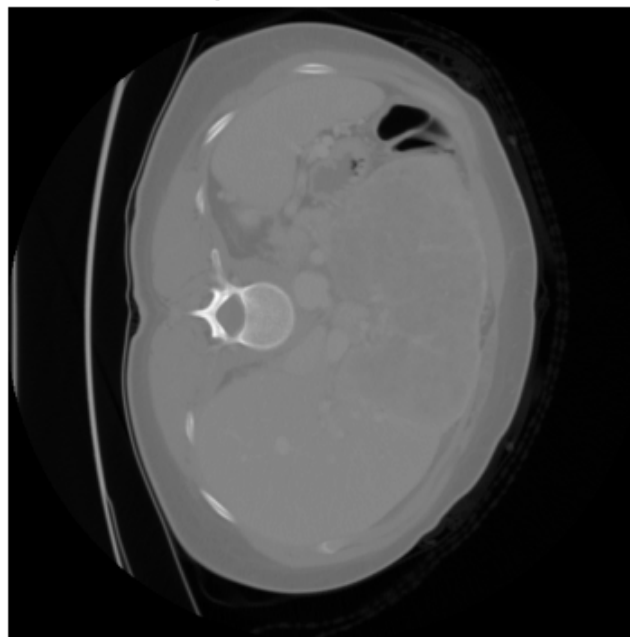
Medical Segmentation Decathlon
(MSD)

segm2.ipynb

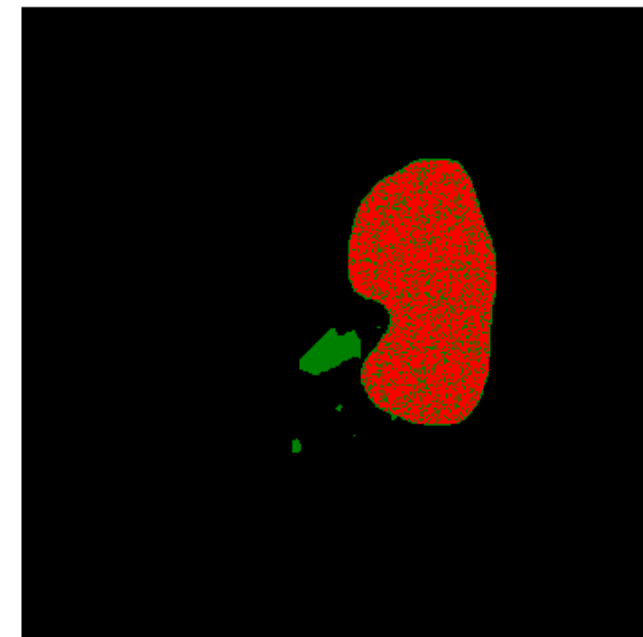
resource2.ipynb

Выбран срез 49 для изображения /content/Task08_HepaticVessel/imagesTr/hepaticvessel_209.nii.gz

Изображение (Slice: 49)



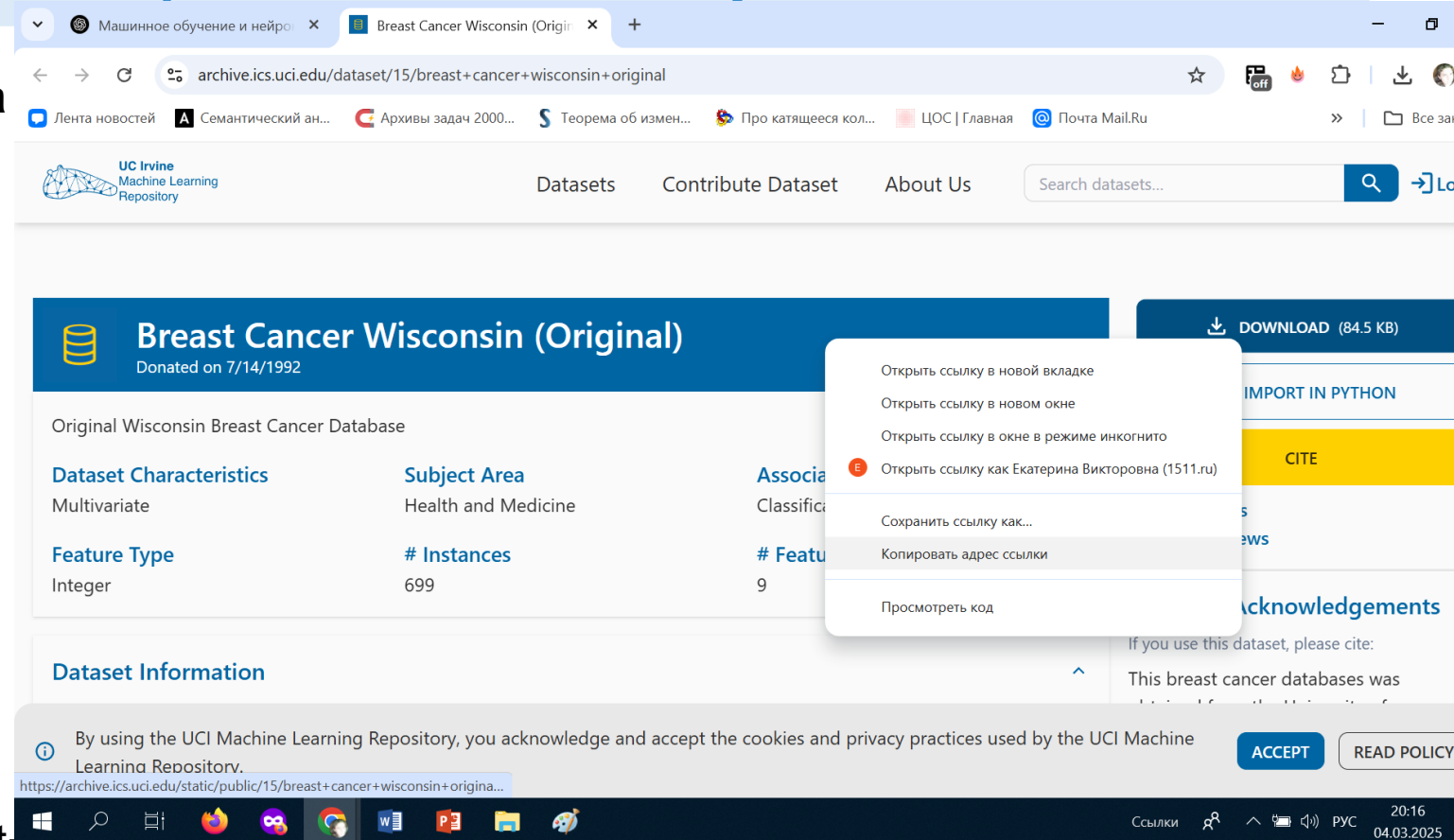
Метки



Практикум. Загрузка с помощью библиотеки pandas.

Breast Cancer Wisconsin (Original) Data

- датасет, используемый для классификации доброкачественных и злокачественных опухолей груди (одна из классических медицинских выборок)



Машинное обучение и нейронные сети

Breast Cancer Wisconsin (Original)

archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original

Лента новостей Семантический анализ Архивы задач 2000... Теорема об изменении Про катящееся колесо ЦОС | Главная Почта Mail.Ru

UC Irvine Machine Learning Repository

Datasets Contribute Dataset About Us Search datasets...

Breast Cancer Wisconsin (Original)
Donated on 7/14/1992

Original Wisconsin Breast Cancer Database

Dataset Characteristics	Subject Area	Association
Multivariate	Health and Medicine	Classification
Feature Type	# Instances	# Features
Integer	699	9

Dataset Information

By using the UCI Machine Learning Repository, you acknowledge and accept the cookies and privacy practices used by the UCI Machine Learning Repository.

https://archive.ics.uci.edu/static/public/15/breast+cancer+wisconsin+original...

Ссылки

20:16 04.03.2025

<https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original>

ФайлГлавнаяВставкаРисованиеКонструкторПереходыАнимацияСлайд-шоуРецензированиеВидСправкаЧто вы хотите сделать?

С текущего начала

С текущего слайда

Онлайн-презентация

Произвольное слайд-шоу

Начать слайд-шоу

Настройка слайд-шоу

Скрыть слайд

Настройка времени

Записать слайд-шоу

☒ Воспроизвести закадровый текст

☒ Использовать время показа слайдов

☒ Показать элементы управления проигрывателем

Монитор: Автоматически

☐ Режим докладчика

Мониторы

34

Тема 3.3
Медицинские данные и загрузка с помощью библиотеки pandas.

35

Тема 3.3
Медицинские данные и загрузка с помощью библиотеки pandas.

36

Тема 3.3
Медицинские данные и загрузка с помощью библиотеки pandas.

37

Тема 3.3
Медицинские данные и загрузка с помощью библиотеки pandas.

38

Тема 3.3
Медицинские данные и загрузка с помощью библиотеки pandas.

39

Тема 3.3
Медицинские данные и загрузка с помощью библиотеки pandas.

16151413121110987654321012345678910111213141516

Тема 3.3

Медицинские данные и загрузка с помощью библиотеки pandas.

Reast Cancer Wisconsin (Original) Data Set — датасет, используемый для классификации доброкачественных и злокачественных опухолей груди (одна из классических медицинских выборок)

Бреast Cancer Wisconsin (Original)

Donated on 7/14/1992

Original Wisconsin Breast Cancer Database

Dataset Characteristics

Multivariate

Feature Type

Integer

Subject Area

Health and Medicine

Instances

699

Associa

Classific

Featu

9

Dataset Information

By using the UCI Machine Learning Repository, you acknowledge and accept the cookies and privacy practices used by the UCJ Machine Learning Repository.

https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original...

DOWNLOAD (84.5 KB)

IMPORT IN PYTHON

CITE

Acknowledgements

If you use this dataset, please cite:
This breast cancer databases was:

ACCEPT

READ POLICY

36

Слайд 36 из 45русский

ЗаметкиПримечания

65 %

11:22
05.03.2025

Тема 1.1

Практикум. Загрузка с помощью библиотеки pandas.

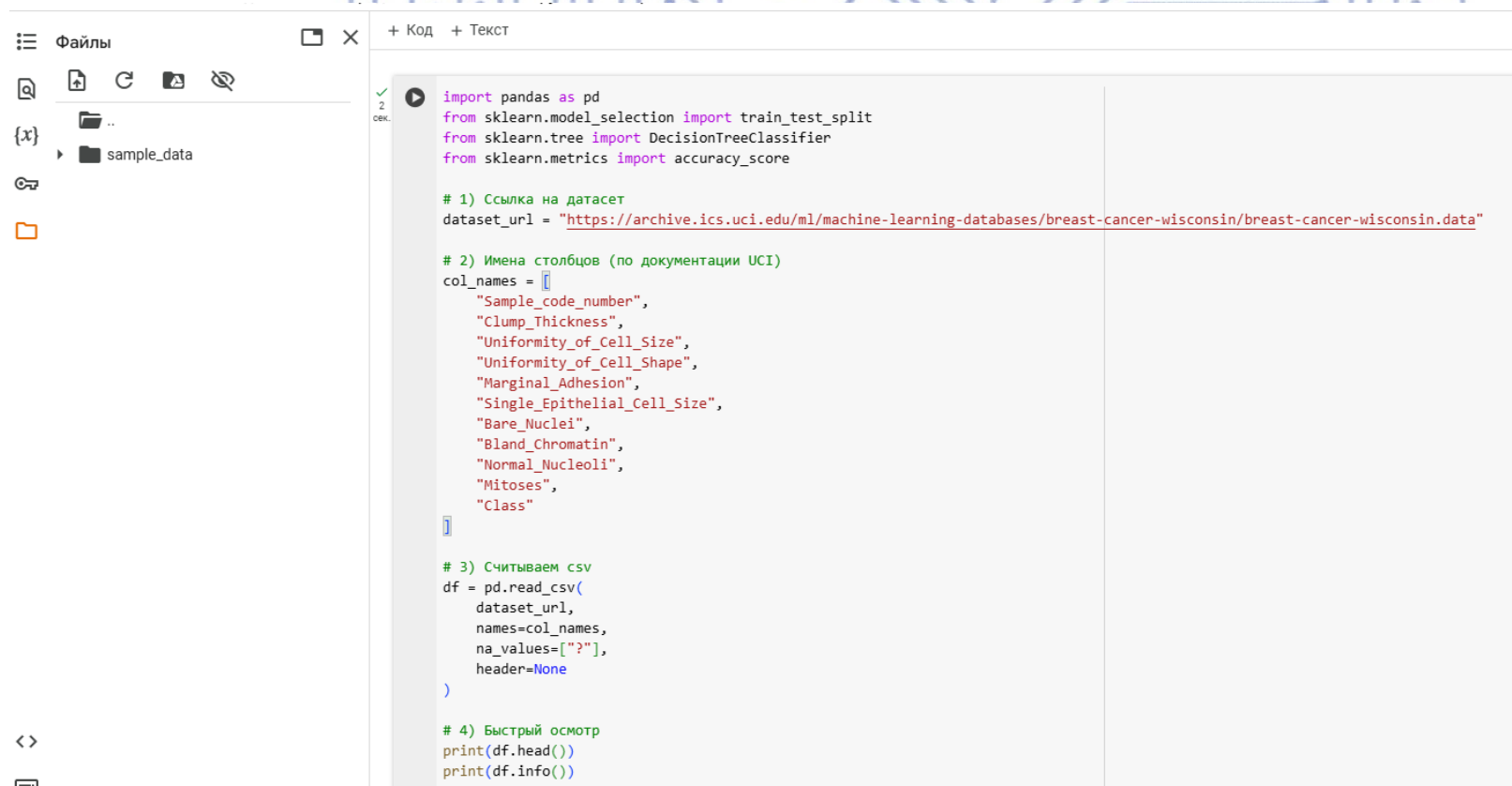
Breast Cancer Wisconsin (Original)

names=col_names —
чтобы присвоить столбцам
понятные имена (так как в
файле нет заголовка).

Загрузка датасета через
pandas.read_csv

na_values=["?"] — все «?»
превращаются в NaN
(пропущенные значения)

Dataset.ipynb



```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

# 1) Ссылка на датасет
dataset_url = "https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data"

# 2) Имена столбцов (по документации UCI)
col_names = [
    "Sample_code_number",
    "Clump_Thickness",
    "Uniformity_of_Cell_Size",
    "Uniformity_of_Cell_Shape",
    "Marginal_Adhesion",
    "Single_Epithelial_Cell_Size",
    "Bare_Nuclei",
    "Bland_Chromatin",
    "Normal_Nucleoli",
    "Mitoses",
    "Class"
]

# 3) Считываем csv
df = pd.read_csv(
    dataset_url,
    names=col_names,
    na_values=["?"],
    header=None
)

# 4) Быстрый осмотр
print(df.head())
print(df.info())
```



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практикум. Загрузка с помощью библиотеки pandas.

Breast Cancer Wisconsin (Original) Data Set

Class принимает значения 2 (Benign) или 4 (Malignant).

```
Sample_code_number  Clump_Thickness  Uniformity_of_Cell_Size  \
0                1000025              5                      1
1                1002945              5                      4
2                1015425              3                      1
3                1016277              6                      8
4                1017023              4                      1

Uniformity_of_Cell_Shape  Marginal_Adhesion  Single_Epithelial_Cell_Size  \
0                        1                  1                          2
1                        4                  5                          7
2                        1                  1                          2
3                        8                  1                          3
4                        1                  3                          2

Bare_Nuclei  Bland_Chromatin  Normal_Nucleoli  Mitoses
0           1.0              3                1        1
1          10.0              3                2        1
2           2.0              3                1        1
3           4.0              3                7        1
4           1.0              3                1        1
<class 'pandas.core.frame.DataFrame'>
```

Dataset.ipynb



ФНЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практикум. Загрузка с помощью библиотеки pandas.

Breast Cancer Wisconsin (Original) Data Set

Разделение признаков (X) и целевой переменной (y)

```
y = df["Class"] # Целевой столбец, где: 2 = benign, 4 = malignant  
X = df.drop(["Sample_code_number", "Class"], axis=1)
```

- Столбец `Class` — это то, что мы предсказываем (вид опухоли: доброкачественная/злокачественная).
- Остальные столбцы — **признаки**, на основе которых строится прогноз (`Clump_Thickness`, `Uniformity_of_Cell_Shape` и т.п.).



НМИЦ
ОНКОЛОГИИ
им. Н.Н. Блохина

Тема 1.1

Практикум. Загрузка с помощью библиотеки pandas.

Breast Cancer Wisconsin (Original) Data Set

Это алгоритм деревьев решений, который последовательно делит пространство признаков (толщина, форма клеток и т.д.) на узлы, чтобы максимально отделять класс «2» (benign) от «4» (malignant).

Dataset.ipynb

Создание и обучение (fit) модели

```
clf = DecisionTreeClassifier(random_state=42)
clf.fit(X_train, y_train)
```

1. Создаём объект класса `DecisionTreeClassifier` (из библиотеки `sklearn.tree`).

Тема 1.1

Практикум. Загрузка с помощью библиотеки pandas.

Breast Cancer Wisconsin (Original) Data Set

2. Обучение (`.fit()`):

Это алгоритм деревьев решений, который последовательно делит пространство признаков (толщина, форма клеток и т.д.) на узлы, чтобы максимально отделять класс «2» (benign) от «4» (malignant).

Dataset.ipynb

- Модель просматривает `X_train` (признаки) и `y_train` (классы опухолей) и пытается найти **правила ветвления** (если `Uniformity_of_Cell_Size > 3`, идём налево и т.д.).
- Цель: **минимизировать** ошибки классификации на тренировочной выборке (т.е. правильно разделить доброкачественные и злокачественные опухоли).

Предсказание на тестовой выборке

```
predictions = clf.predict(X_test)
```

NOTE
Accuracy: 0.942

Получив обученное дерево, мы применяем его к новым примерам из `X_test`. Результат `predictions` — это массив предсказанных классов (2 или 4).



ФНЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практикум. Загрузка с помощью библиотеки pandas.

Парсинг(сбор данных).

Когда готовых выгрузок (CSV, JSON, API) нет, можно парсить (сбирать данные) (scrape) сайты автоматически, например: **requests** (Python) для загрузки HTML-страницы.

BeautifulSoup (bs4) для извлечения нужных HTML-элементов (таблиц, списков).

pandas - `read_html` (если таблицы хорошо структурированы).

```
import requests
from bs4 import BeautifulSoup
import pandas as pd

url = "https://opendata.some-region.ru/datasets/interesting-table"
response = requests.get(url)
soup = BeautifulSoup(response.text, "html.parser")

# Предположим, на странице есть <table> с классом "data-table"
table_html = soup.find("table", {"class": "data-table"})

# Преобразуем в DataFrame, если таблица более-менее стандартная
df_list = pd.read_html(str(table_html)) # возвращает список df
df = df_list[0] # берем первую таблицу

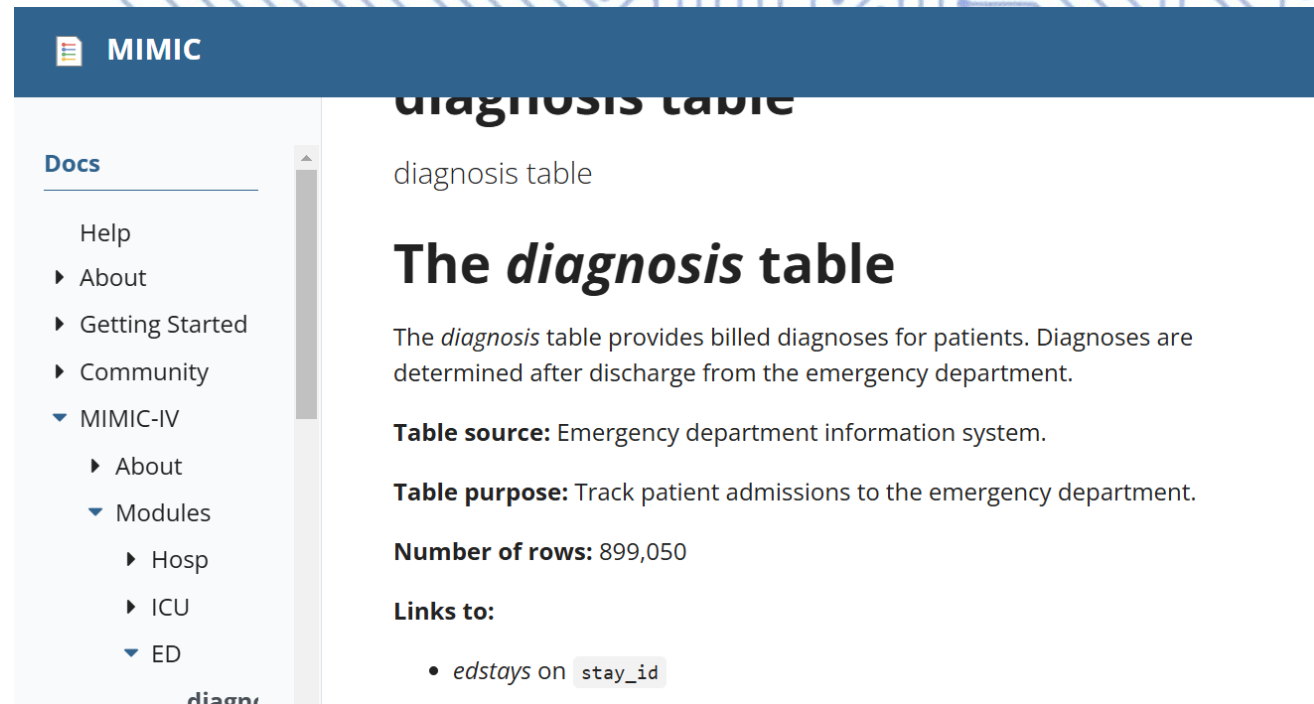
df.head()
```

Другие открытые наборы данных

MIMIC-III / MIMIC-IV (Medical Information Mart for Intensive Care)

- Анонимизированные записи из отделения интенсивной терапии (больница Бет Исраэль Дияконесса, США).
- Содержит не только табличные данные, но и текст (выписки, нотации).
- Можно использовать для разработки методов выделения диагнозов, симптомов и т.д.
- Доступ по определённой процедуре (CITI training + DUA).

<https://mimic.mit.edu/>

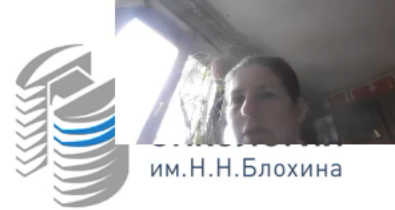


The screenshot shows the MIMIC database website. The header is dark blue with the MIMIC logo. The left sidebar contains a 'Docs' section with links to Help, About, Getting Started, Community, and MIMIC-IV. Under MIMIC-IV, there are links to About, Modules, Hosp, ICU, and ED. The main content area is titled 'diagnosis table' and contains the following information:

- The *diagnosis* table**
- The *diagnosis* table provides billed diagnoses for patients. Diagnoses are determined after discharge from the emergency department.
- Table source:** Emergency department information system.
- Table purpose:** Track patient admissions to the emergency department.
- Number of rows:** 899,050
- Links to:**
 - *edstays* on *stay_id*

Тема 3.3

Открытые наборы данных (Medical NLP)



MIMIC-III / MIMIC-IV (Medical Information Mart for Intensive Care)

- Анонимизированные записи из отделения интенсивной терапии (больница Бет Исраэль Дияконесса, США).
- Содержит не только табличные данные, но и текст (выписки, нотации).
- Можно использовать для разработки методов выделения диагнозов, симптомов и т.д.
- Доступ по определённой процедуре (CITI training + DUA).

<https://mimic.mit.edu/>

The screenshot shows the MIMIC-IV documentation interface. On the left is a navigation menu with 'Docs' at the top, followed by 'Help', 'About', 'Getting Started', 'Community', and 'MIMIC-IV'. Under 'MIMIC-IV', there are links for 'About', 'Modules', 'Hosp', 'ICU', and 'ED'. The 'diagnosis table' is selected. The main content area has a blue header with 'MIMIC' and 'diagnosis table'. Below this, the title 'The diagnosis table' is followed by a description: 'The diagnosis table provides billed diagnoses for patients. Diagnoses are determined after discharge from the emergency department.' It then lists 'Table source: Emergency department information system.', 'Table purpose: Track patient admissions to the emergency department.', and 'Number of rows: 899,050'. Under 'Links to:', there is a bullet point 'edstays on stay_id'.



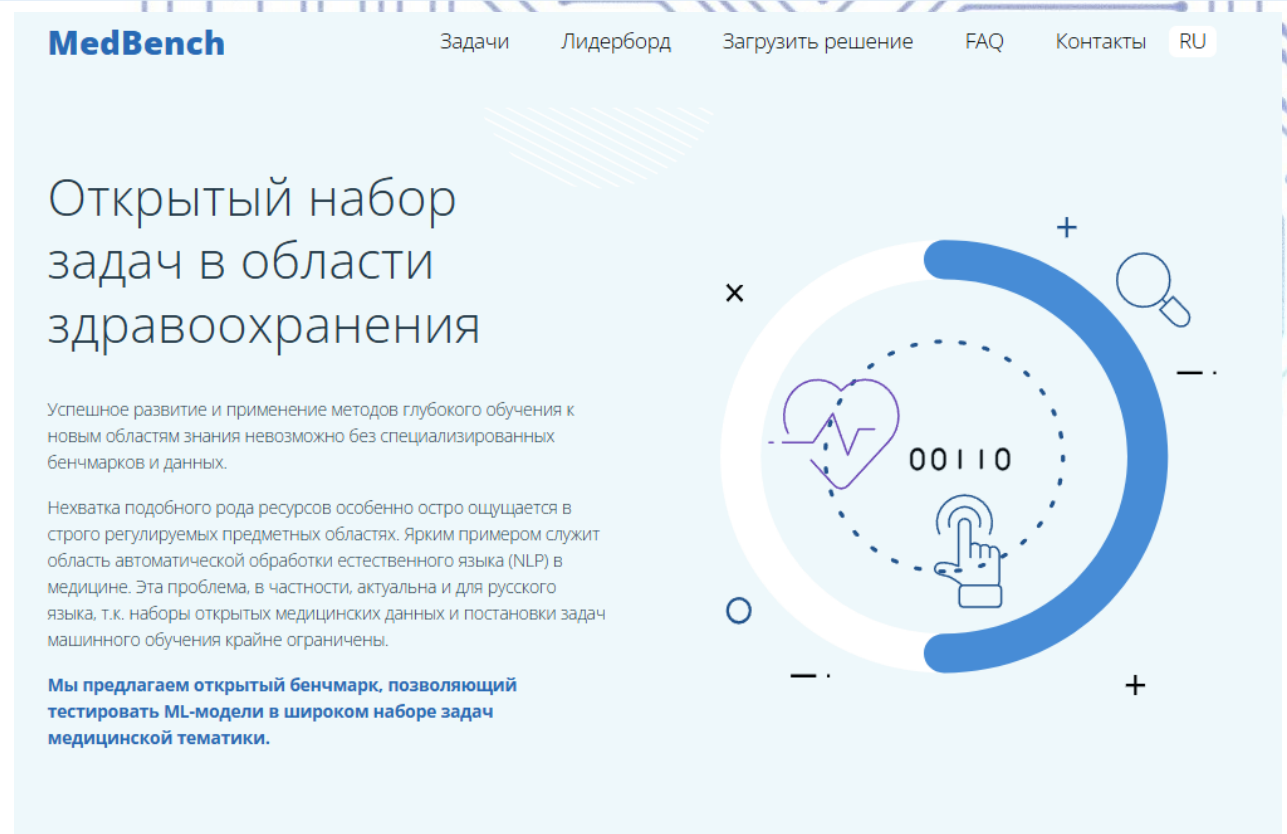
НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Другие открытые наборы данных

MedBench

открытый бенчмарк, позволяющий тестировать ML-модели в широком наборе задач медицинской тематики



The screenshot shows the MedBench website. At the top, there is a navigation bar with links: "Задачи", "Лидерборд", "Загрузить решение", "FAQ", "Контакты", and a language selector "RU". The main heading is "Открытый набор задач в области здравоохранения". Below this, there is a paragraph in Russian about the importance of deep learning in healthcare and the need for specialized benchmarks. Another paragraph discusses the lack of resources in strictly regulated fields like NLP in medicine. At the bottom, a blue box contains the text: "Мы предлагаем открытый бенчмарк, позволяющий тестировать ML-модели в широком наборе задач медицинской тематики." On the right side of the page, there is a large graphic featuring a heart with an ECG line, a hand pointing at a screen, and binary code "00110", all enclosed in a circular frame with mathematical symbols (+, -, x, /) around it.

MedBench Задачи Лидерборд Загрузить решение FAQ Контакты RU

Открытый набор задач в области здравоохранения

Успешное развитие и применение методов глубокого обучения к новым областям знания невозможно без специализированных бенчмарков и данных.

Нехватка подобного рода ресурсов особенно остро ощущается в строго регулируемых предметных областях. Ярким примером служит область автоматической обработки естественного языка (NLP) в медицине. Эта проблема, в частности, актуальна и для русского языка, т.к. наборы открытых медицинских данных и постановки задач машинного обучения крайне ограничены.

Мы предлагаем открытый бенчмарк, позволяющий тестировать ML-модели в широком наборе задач медицинской тематики.

<https://medbench.ru/ru/?page=1>



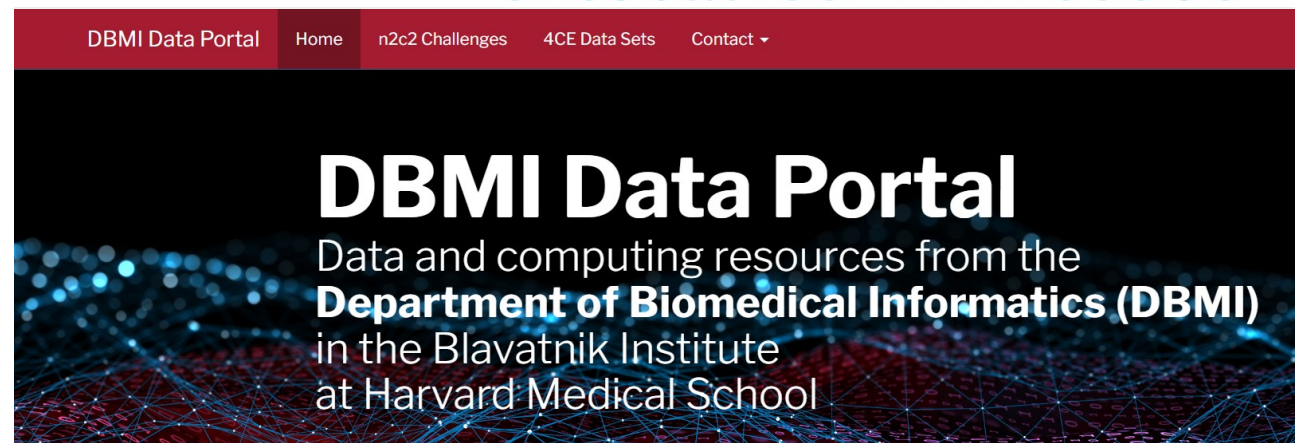
НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Другие открытые наборы данных

n2c2 (National NLP Clinical Challenges)

- Проводит регулярные соревнования / челенджи по клиническому NLP: Medication Extraction, Adverse Drug Events, Risk Factors и т



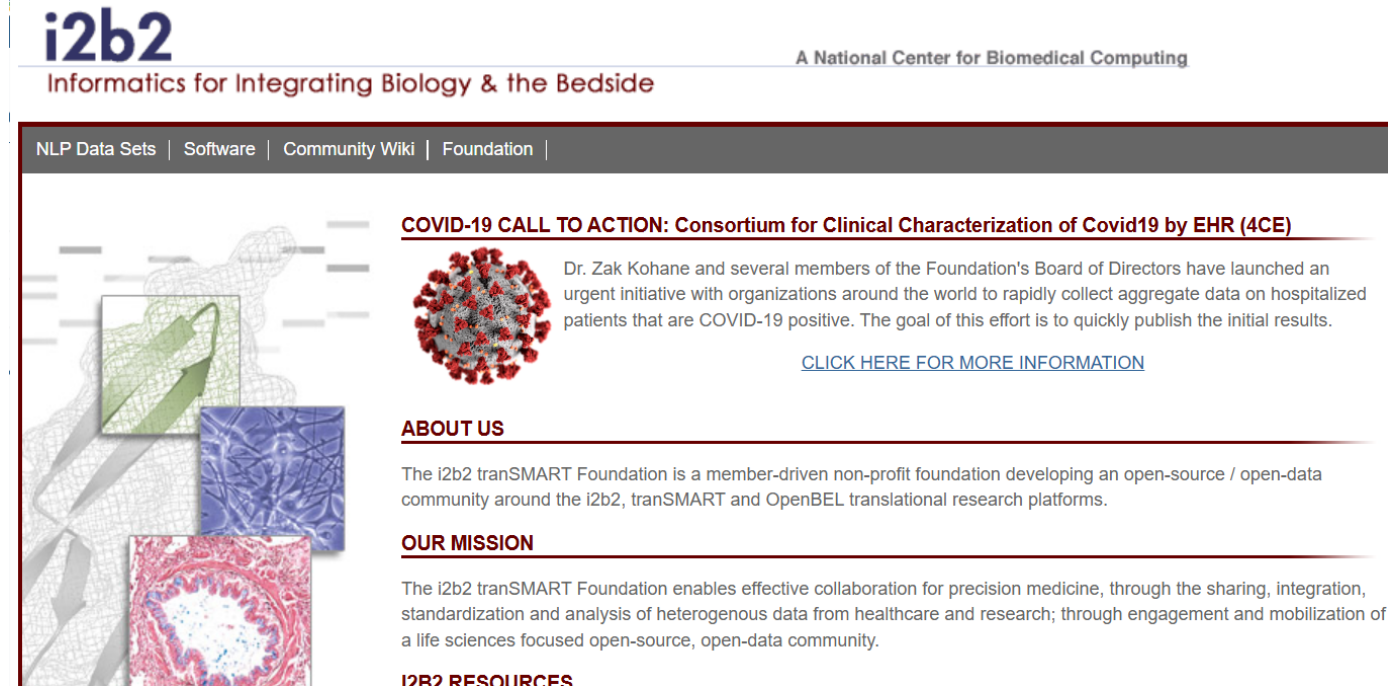
<https://portal.dbmi.hms.harvard.edu/>

Другие открытые наборы данных

i2b2 (Informatics for Integrating Biology and the Bedside) / NLP Challenge Datasets

- Серия открытых клинических текстовых корпусов: задачи по анонимизации, извлечению признаков, определению временных интервалов, классификации диагнозов и др.
- Тексты на английском, можно тренировать системы по извлечению ключевых терминов (диагноз, симптомы).

<https://www.i2b2.org/>



The image shows a screenshot of the i2b2 website. The header features the i2b2 logo and the text 'A National Center for Biomedical Computing'. Below the header is a navigation bar with links: 'NLP Data Sets', 'Software', 'Community Wiki', and 'Foundation'. The main content area includes a section titled 'COVID-19 CALL TO ACTION: Consortium for Clinical Characterization of Covid19 by EHR (4CE)' with a sub-header 'Dr. Zak Kohane and several members of the Foundation's Board of Directors have launched an urgent initiative with organizations around the world to rapidly collect aggregate data on hospitalized patients that are COVID-19 positive. The goal of this effort is to quickly publish the initial results.' and a link 'CLICK HERE FOR MORE INFORMATION'. Below this is an 'ABOUT US' section stating 'The i2b2 transSMART Foundation is a member-driven non-profit foundation developing an open-source / open-data community around the i2b2, transSMART and OpenBEL translational research platforms.' and an 'OUR MISSION' section stating 'The i2b2 transSMART Foundation enables effective collaboration for precision medicine, through the sharing, integration, standardization and analysis of heterogenous data from healthcare and research; through engagement and mobilization of a life sciences focused open-source, open-data community.' At the bottom, there is a section for 'i2b2 RESOURCES'. On the left side of the main content area, there are three overlapping images: a green leaf, a blue micrograph, and a red micrograph.



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Другие открытые наборы дан

DeepPavlov

- Развивается при МФТИ, ориентирован на диалоговые системы и обработку русского языка. Есть модули для **NER (Named Entity Recognition)**, чат-ботов, классификации текста.
- Можно использовать для выделения сущностей (например, «диагноз», «симптом», «медикамент»)

<https://deeppavlov.ai/>

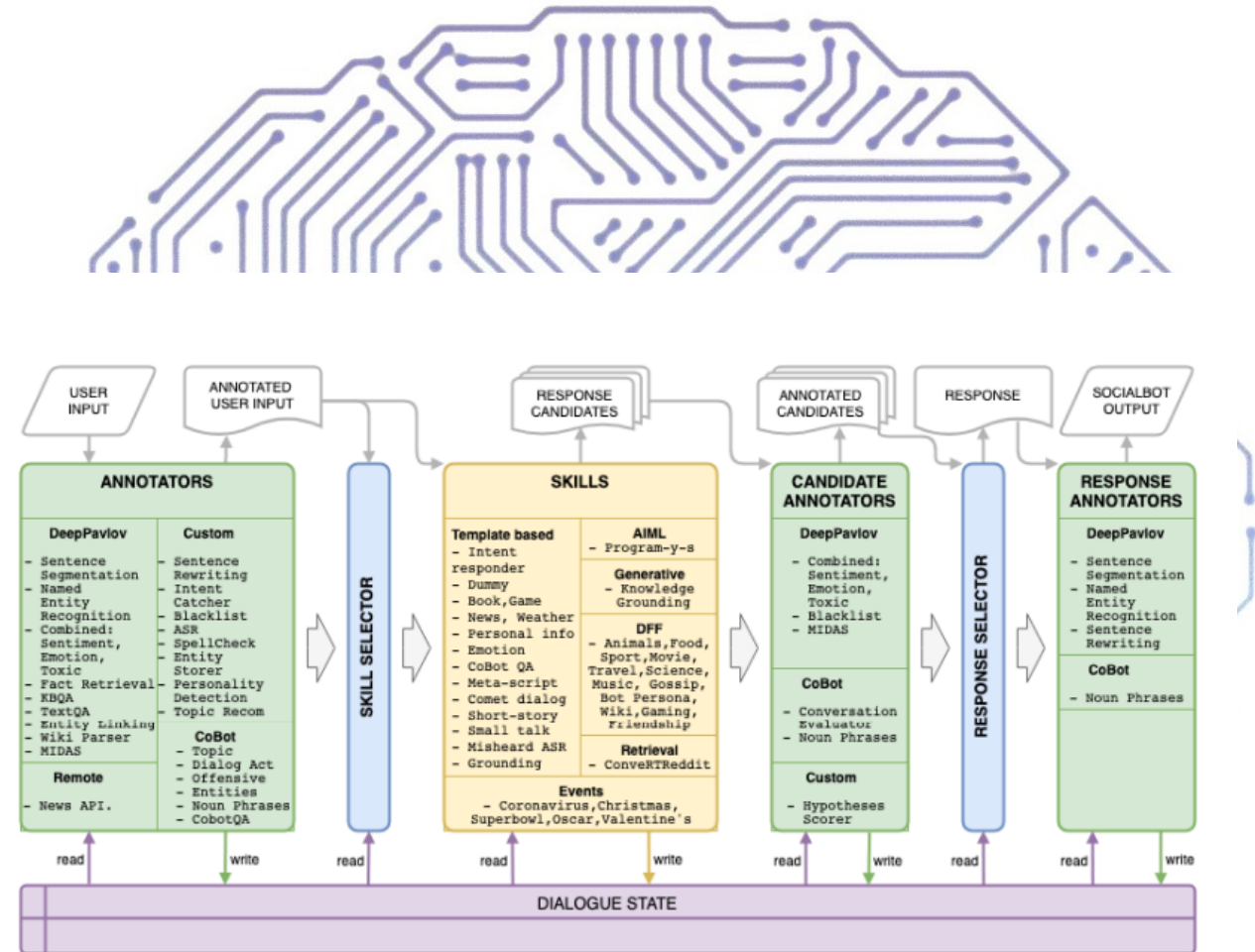
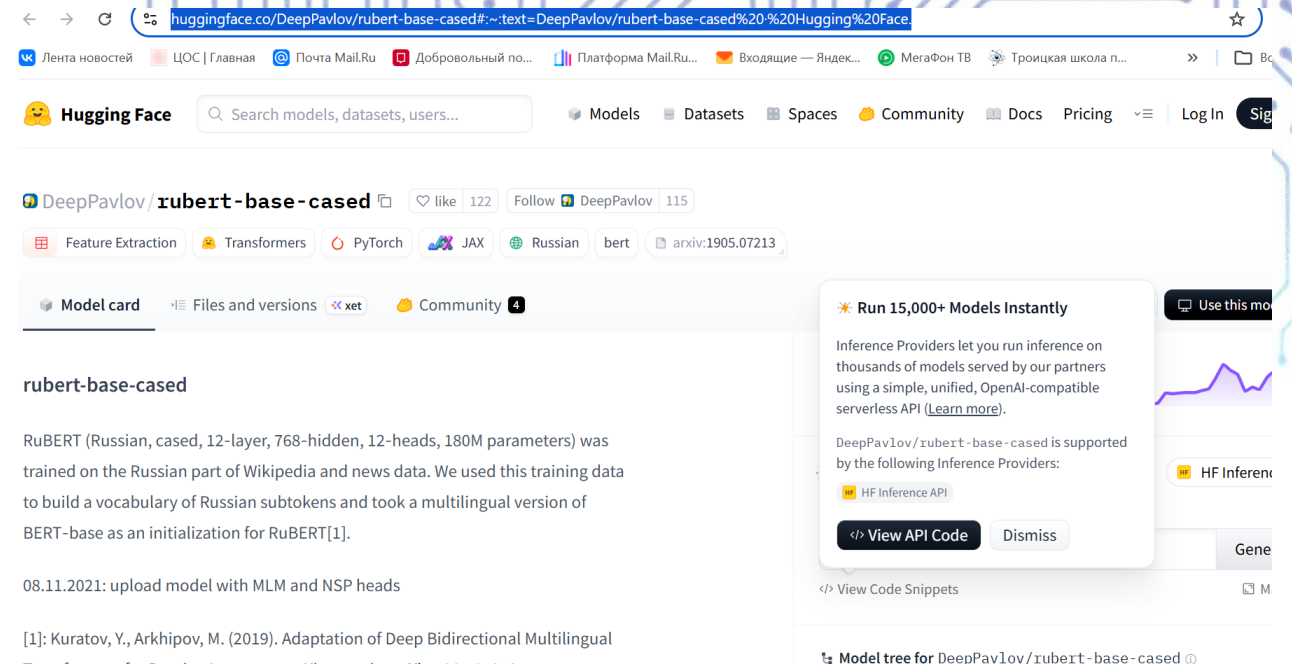


Figure 1: **DREAM socialbot architecture.** Multiple Annotators are used to extract information from the user input. Skill Selector defines a subset of active Skills based on the extracted information. Selected Skills propose their response candidates. Finally, Response Selector picks a response to be sent to Response Annotators and, eventually, to the user. All elements of the pipeline are running asynchronously with two points of synchronization: Skill Selector and Response Selector. Dialogue State serves as a shared memory.

Другие открытые наборы данных

RuBERT / RuGPT / СберБерт

Большие языковые модели, обученные на русскоязычных корпусах (литература, интернет, СМИ). Для автоматизации документации (например, распознавание жалоб, извлечение ICD-10 кода из диагноза). Например, [RuBERT](#) — это специализированная версия нейросетевой архитектуры BERT, обученная [DeepPavlov на Hugging Face](#) для понимания русского языка.



The screenshot shows the Hugging Face interface for the model `DeepPavlov/rubert-base-cased`. The browser address bar contains the URL: `huggingface.co/DeepPavlov/rubert-base-cased#:~:text=DeepPavlov/rubert%2Dbase%2Dcased%20%C2%B7%20Hugging%20Face`. The page header includes the Hugging Face logo and navigation links like Models, Datasets, Spaces, Community, Docs, Pricing, Log In, and Sign Up. The model card for `rubert-base-cased` is displayed, showing it has 122 likes and 115 followers. It lists tags for Feature Extraction, Transformers, PyTorch, JAX, Russian, bert, and an arXiv link (1905.07213). The card includes a 'Model card' tab, 'Files and versions', and a 'Community' section. The main text describes the model as a Russian, cased, 12-layer, 768-hidden, 12-heads, 180M parameters BERT model trained on the Russian part of Wikipedia and news data. It mentions the training data and the use of a multilingual version of BERT-base as an initialization. A date stamp indicates the model was uploaded on 08.11.2021. A citation is provided at the bottom: [1]: Kuratov, Y., Arkhipov, M. (2019). Adaptation of Deep Bidirectional Multilingual. A sidebar on the right promotes 'Run 15,000+ Models Instantly' and lists inference providers, including HF Inference API, with buttons for 'View API Code' and 'Dismiss'.



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Другие открытые наборы данных

RuMedPrimeData

открытый датасет, содержащий деперсонализированные медицинские данные из амбулаторных карт Сибирского государственного медицинского университета (СибГМУ).

The screenshot shows the Zenodo record page for the dataset 'RuMedPrimeData'. The header includes the Zenodo logo, a search bar, and links to 'Communities' and 'My dashboard'. Below the header, the dataset is associated with the 'Siberian State Medical University' and features its logo. The publication date is 'December 1, 2021' and the version is '1.0'. There are buttons for 'Dataset' and 'Open'. The authors listed are Starovoytova Elena A., Kulikov Evgeny S., Fedosenko Sergey V., Shmyrina Alexandra A., Kirillova Natalia A., Vinokurova Daria A., and Balaganskaya Marina A., each with a green ORCID icon. A description states: 'The Dataset contains anonymized data from outpatient visits to SSMU hospital. Each visit has the following fields: new_patient_id – patient identification; new_visit_id – visit identification;'. The URL at the bottom of the screenshot is <https://zenodo.org/records/5765873#.Y6WH2NJBzmy>.

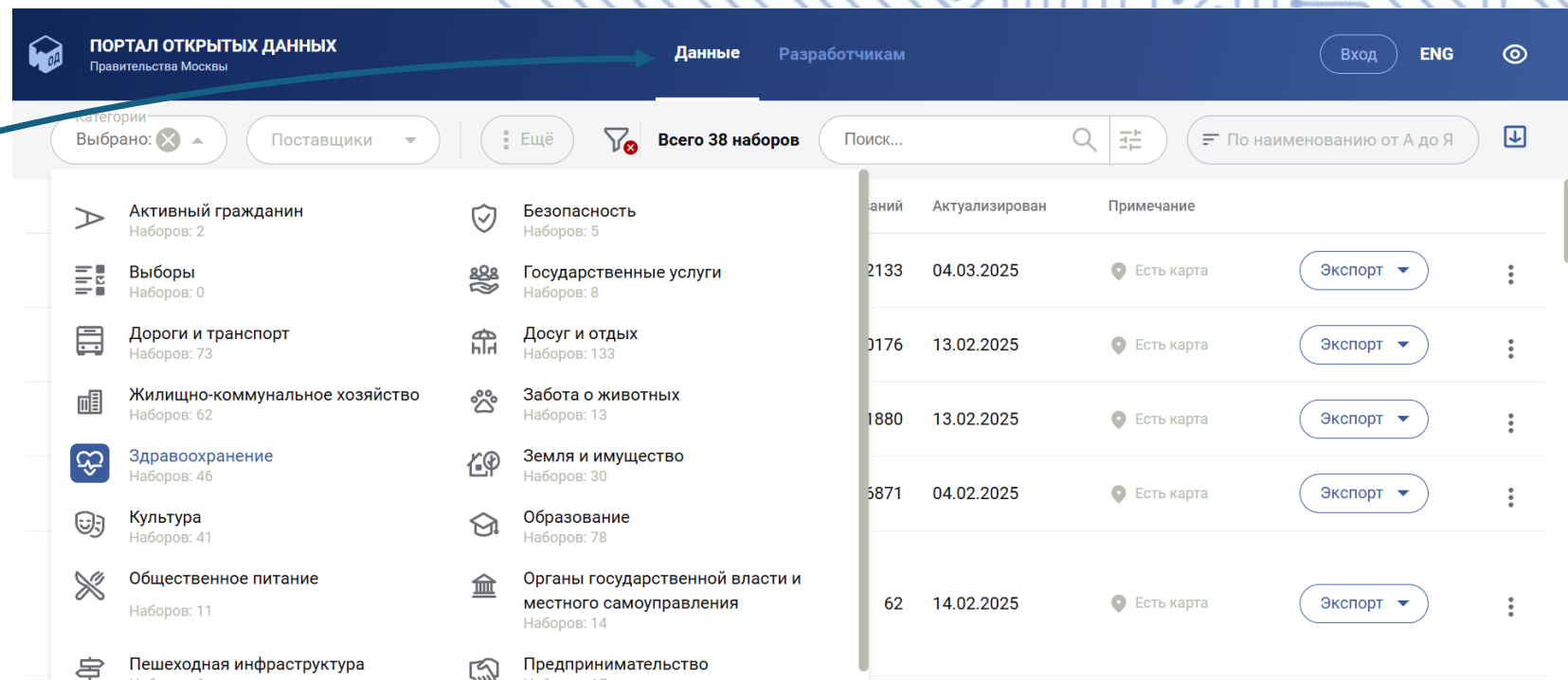
<https://zenodo.org/records/5765873#.Y6WH2NJBzmy>

Другие открытые наборы данных

Портал открытых данных Москвы: opendata.mos.ru.

Шаг 1.

Открыть портал по ссылке и выбрать вкладку «данные»



ПОРТАЛ ОТКРЫТЫХ ДАННЫХ
Правительства Москвы

Данные Разработчикам

Вход ENG

Категории
Выбрано: X

Поставщики

Ещё

Всего 38 наборов

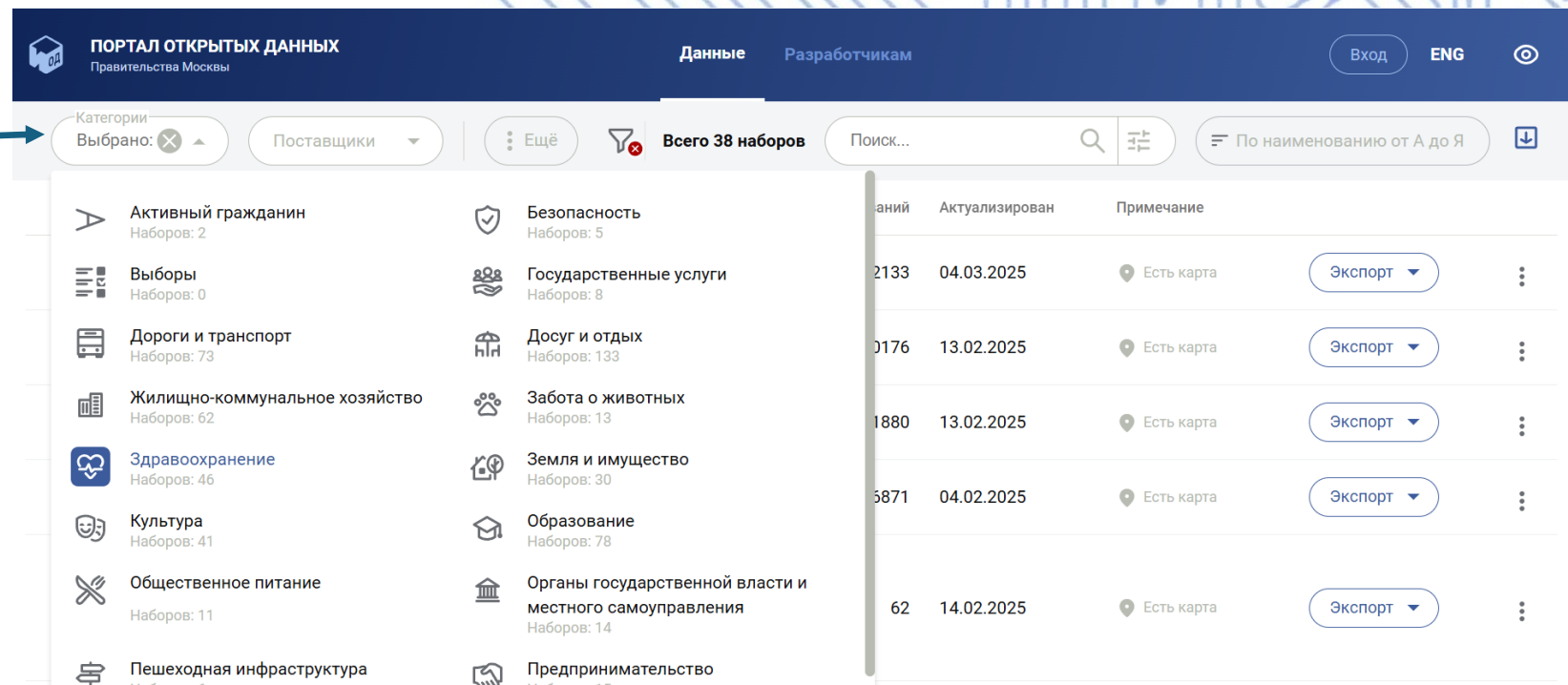
Поиск...

По наименованию от А до Я

Наименование	Актуализирован	Примечание	Экспорт
Активный гражданин Наборов: 2	2133	04.03.2025 Есть карта	Экспорт
Выборы Наборов: 0	0176	13.02.2025 Есть карта	Экспорт
Дороги и транспорт Наборов: 73	1880	13.02.2025 Есть карта	Экспорт
Жилищно-коммунальное хозяйство Наборов: 62	5871	04.02.2025 Есть карта	Экспорт
Здравоохранение Наборов: 46	62	14.02.2025 Есть карта	Экспорт
Культура Наборов: 41			
Общественное питание Наборов: 11			
Пешеходная инфраструктура Наборов: 6			
Безопасность Наборов: 5			
Государственные услуги Наборов: 8			
Досуг и отдых Наборов: 133			
Забота о животных Наборов: 13			
Земля и имущество Наборов: 30			
Образование Наборов: 78			
Органы государственной власти и местного самоуправления Наборов: 14			
Предпринимательство Наборов: 15			

Другие открытые наборы данных

Портал открытых данных Москвы: opendata.mos.ru.



ПОРТАЛ ОТКРЫТЫХ ДАННЫХ
Правительства Москвы

Данные Разработчикам

Вход ENG

Категории
Выбрано: X

Поставщики

Ещё

Всего 38 наборов

Поиск...

По наименованию от А до Я

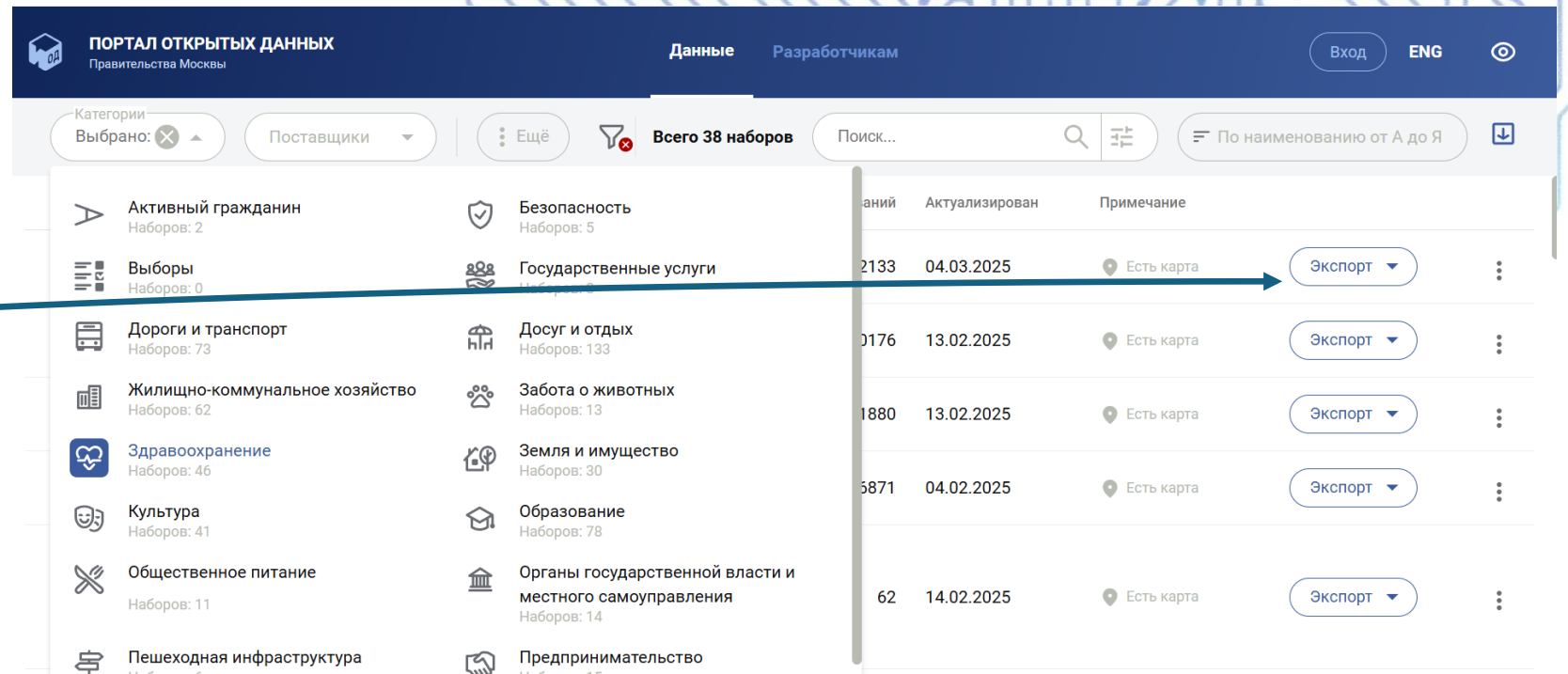
Активный гражданин Наборов: 2	Безопасность Наборов: 5	Актуализирован	Примечание
Выборы Наборов: 0	Государственные услуги Наборов: 8	2133 04.03.2025	Есть карта Экспорт
Дороги и транспорт Наборов: 73	Досуг и отдых Наборов: 133	0176 13.02.2025	Есть карта Экспорт
Жилищно-коммунальное хозяйство Наборов: 62	Забота о животных Наборов: 13	1880 13.02.2025	Есть карта Экспорт
Здравоохранение Наборов: 46	Земля и имущество Наборов: 30	5871 04.02.2025	Есть карта Экспорт
Культура Наборов: 41	Образование Наборов: 78		
Общественное питание Наборов: 11	Органы государственной власти и местного самоуправления Наборов: 14	62 14.02.2025	Есть карта Экспорт
Пешеходная инфраструктура	Предпринимательство		

Шаг 2.

Выбрать категорию
«Здравоохранение»

Другие открытые наборы данных

Портал открытых данных Москвы: opendata.mos.ru.



ПОРТАЛ ОТКРЫТЫХ ДАННЫХ
Правительства Москвы

Данные Разработчикам

Вход ENG

Категории
Выбрано: X

Поставщики

Ещё

Всего 38 наборов

Поиск...

По наименованию от А до Я

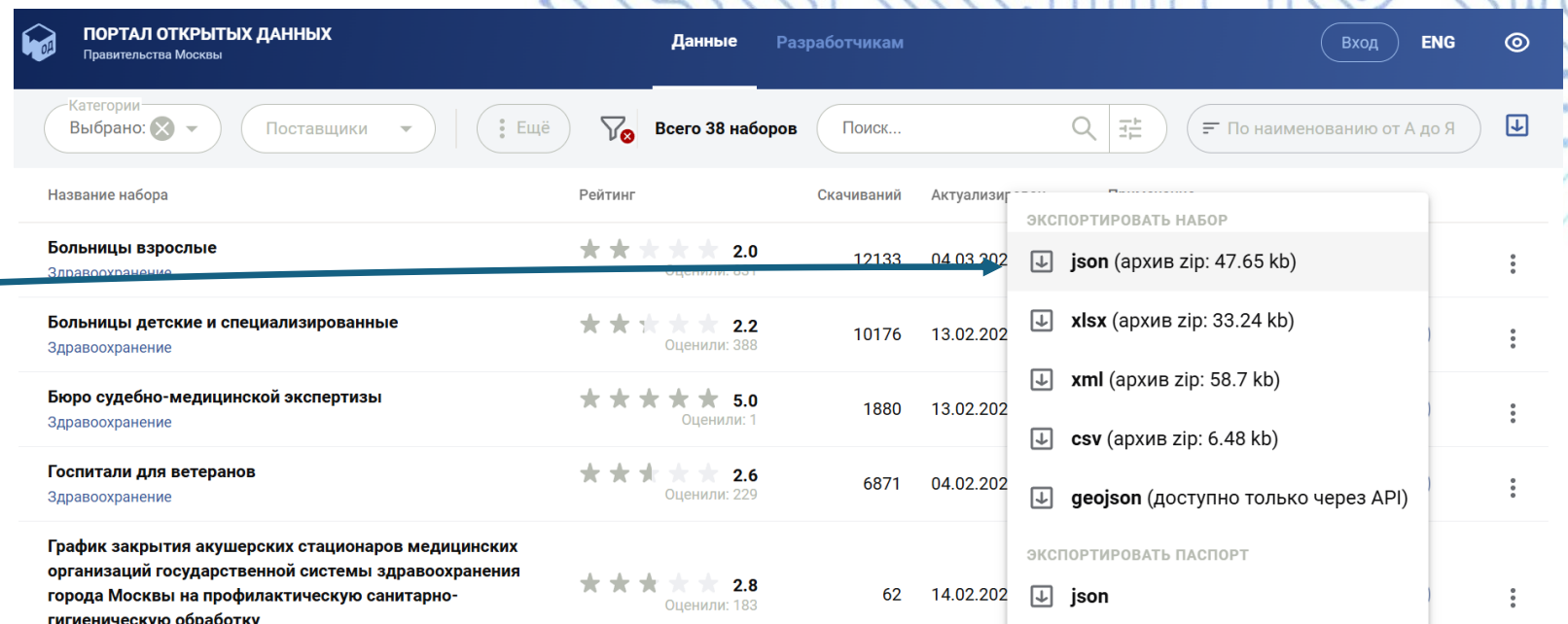
Набор данных	Актуализирован	Примечание	Экспорт
Активный гражданин Наборов: 2			
Безопасность Наборов: 5			
Выборы Наборов: 0			
Государственные услуги Наборов: 2133	04.03.2025	Есть карта	Экспорт
Дороги и транспорт Наборов: 73			
Досуг и отдых Наборов: 133	13.02.2025	Есть карта	Экспорт
Жилищно-коммунальное хозяйство Наборов: 62			
Забота о животных Наборов: 13	13.02.2025	Есть карта	Экспорт
Здравоохранение Наборов: 46			
Земля и имущество Наборов: 30			
Культура Наборов: 41			
Образование Наборов: 78	04.02.2025	Есть карта	Экспорт
Общественное питание Наборов: 11			
Органы государственной власти и местного самоуправления Наборов: 14			
Пешеходная инфраструктура Наборов: 62	14.02.2025	Есть карта	Экспорт
Предпринимательство Наборов: 14			

Шаг 3.

Напротив нужного набора открыть вкладку «Экспорт»

Другие открытые наборы данных

Портал открытых данных Москвы: opendata.mos.ru.



ПОРТАЛ ОТКРЫТЫХ ДАННЫХ
Правительства Москвы

Данные Разработчикам

Вход ENG

Категории
Выбрано:  

Поставщики 

Ещё 

Всего 38 наборов

Поиск...

По наименованию от А до Я 

Название набора	Рейтинг	Скачиваний	Актуализи...	Полнота
Больницы взрослые Здравоохранение	★ ★ ★ ★ ★ 2.0 Оценили: 651	12133	04.03.2022	...
Больницы детские и специализированные Здравоохранение	★ ★ ★ ★ ★ 2.2 Оценили: 388	10176	13.02.2022	...
Бюро судебно-медицинской экспертизы Здравоохранение	★ ★ ★ ★ ★ 5.0 Оценили: 1	1880	13.02.2022	...
Госпитали для ветеранов Здравоохранение	★ ★ ★ ★ ★ 2.6 Оценили: 229	6871	04.02.2022	...
График закрытия акушерских стационаров медицинских организаций государственной системы здравоохранения города Москвы на профилактическую санитарно-гигиеническую обработку	★ ★ ★ ★ ★ 2.8 Оценили: 183	62	14.02.2022	...

ЭКСПОРТИРОВАТЬ НАБОР

- json (архив zip: 47.65 kb)
- xlsx (архив zip: 33.24 kb)
- xml (архив zip: 58.7 kb)
- csv (архив zip: 6.48 kb)
- geojson (доступно только через API)

ЭКСПОРТИРОВАТЬ ПАСПОРТ

- json

Шаг 4.

Скачать данные в нужном формате.

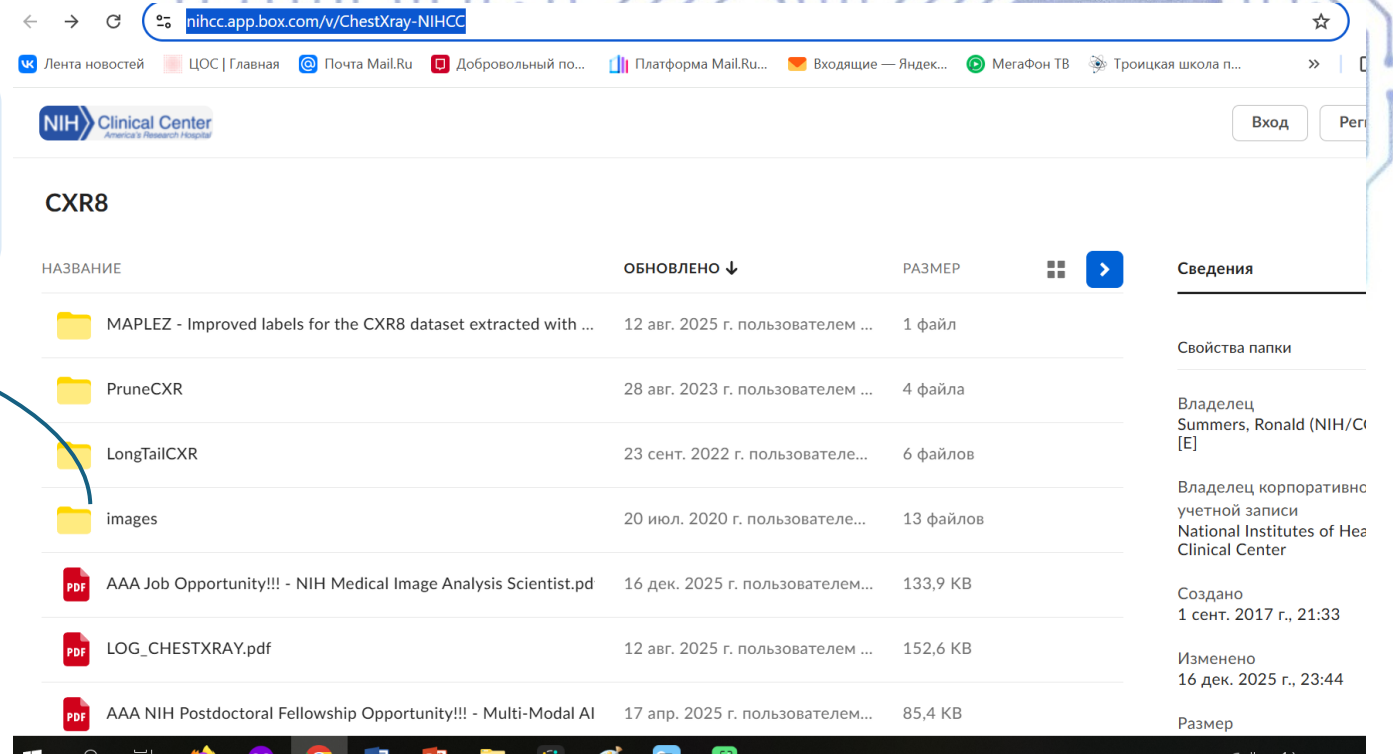
Другие открытые наборы данных

ChestX-ray-NIH

Национальных институтов здоровья США (NIH) на платформе [Box.com](https://nihcc.app.box.com/v/ChestXray-NIHCC). содержит большой объём рентгеновских снимков.

CXR8 > images

НАЗВАНИЕ	ОБНОВЛЕНО ↓	РАЗМЕР
 batch_download_zips.py	20 июл. 2020 г. пользователем...	1,3 KB
 images_004.tar.gz	8 сент. 2017 г. пользователем...	3,6 GB
 images_007.tar.gz	6 сент. 2017 г. пользователем...	3,7 GB
 images_009.tar.gz	5 сент. 2017 г. пользователем...	3,8 GB
 images_010.tar.gz	5 сент. 2017 г. пользователем...	3,9 GB



Скриншот веб-интерфейса NIHCC на платформе Box.com. В адресной строке указан URL: nihcc.app.box.com/v/ChestXray-NIHCC. В меню браузера видны ссылки на Ленту новостей, ЦОС, Главную, Почту Mail.Ru, Добровольный по..., Платформу Mail.Ru..., Входящие — Яндекс..., МегаФон ТВ, Троицкая школа п... В интерфейсе NIHCC Clinical Center America's Research Hospital отображены файлы и папки CXR8. Таблица данных:

НАЗВАНИЕ	ОБНОВЛЕНО ↓	РАЗМЕР	Сведения
MAPLEZ - Improved labels for the CXR8 dataset extracted with ...	12 авг. 2025 г. пользователем ...	1 файл	Свойства папки
PruneCXR	28 авг. 2023 г. пользователем ...	4 файла	Владелец Summers, Ronald (NIH/C[...])
LongTailCXR	23 сент. 2022 г. пользователем ...	6 файлов	Владелец корпоративн[...] учетной записи National Institutes of Hea[...] Clinical Center
images	20 июл. 2020 г. пользователем ...	13 файлов	Создано 1 сент. 2017 г., 21:33 Изменено 16 дек. 2025 г., 23:44 Размер
AAA Job Opportunity!!! - NIH Medical Image Analysis Scientist.pdf	16 дек. 2025 г. пользователем ...	133,9 KB	
LOG_CHESTXRAY.pdf	12 авг. 2025 г. пользователем ...	152,6 KB	
AAA NIH Postdoctoral Fellowship Opportunity!!! - Multi-Modal AI	17 апр. 2025 г. пользователем ...	85,4 KB	

Тема 1.1

Другие открытые наборы данных

ChestX-ray-NIH

Национальных институтов здоровья США (NIH) на платформе [Box.com](https://nihcc.app.box.com). содержит большой объём рентгеновских снимков.

CXR8 > images

НАЗВАНИЕ	ОБНОВЛЕНО ↓	РАЗМЕР
 batch_download_zips.py	20 июл. 2020 г. пользователем...	1,3 KB
 images_004.tar.gz	8 сент. 2017 г. пользователем...	3,6 GB
 images_007.tar.gz	6 сент. 2017 г. пользователем...	3,7 GB
 images_009.tar.gz	5 сент. 2017 г. пользователем...	3,8 GB
 images_010.tar.gz	5 сент. 2017 г. пользователем...	3,9 GB

<https://nihcc.app.box.com/v/ChestXray-NIHCC>



 **batch_download_zips.py**

images · Обновлено 20 июл. 2020 г. пользователем Summers, Ronald (NIH/CC/DRD) [E]

```
#!/usr/bin/env python3
# Download the 56 zip files in Images_png in batches
import urllib.request

# URLs for the zip files
links = [
    'https://nihcc.box.com/shared/static/vfk49d74nhbxq3nqjg0900w5nvkorp5c.gz',
    'https://nihcc.box.com/shared/static/i28rlmbvmfjbl8p2n3ril0pptcmcu9d1.gz',
    'https://nihcc.box.com/shared/static/f1t00wrtkd94satdfb9olcolqx20z2jp.gz',
    'https://nihcc.box.com/shared/static/0aowwzs5lhjrceb3qp67ahp0rd1l1etg.gz',
    'https://nihcc.box.com/shared/static/v5e3goj22zr6h8tzualxfsqaygfbsn.gz',
    'https://nihcc.box.com/shared/static/asi7ikud9jwnkrnkj99jnpfkjdes7l6l.gz',
    'https://nihcc.box.com/shared/static/jn1b4mw4n6lnh74ovmcjb8y48h8xj07n.gz',
    'https://nihcc.box.com/shared/static/tvpxmn7qyrgl0w8wfh9kqfjskv6nmm1j.gz',
    'https://nihcc.box.com/shared/static/upyy3ml7qdumlgk2rfcvlb9k6gvqq2pj.gz',
    'https://nihcc.box.com/shared/static/l6nilvfa9cg3s28tqv1qc1olm3gnz54p.gz',
    'https://nihcc.box.com/shared/static/hhq8fdgvcari67vfhs7ppg2w6ni4jze.gz',
    'https://nihcc.box.com/shared/static/ioqwiy20ihqwyr8pf4c24eazhh281pbu.gz'
]

for idx, link in enumerate(links):
    fn = 'images_002d.tar.gz' % (idx+1)
    print('downloading'+fn+'...')
    urllib.request.urlretrieve(link, fn) # download the zip file
```

— 100% +



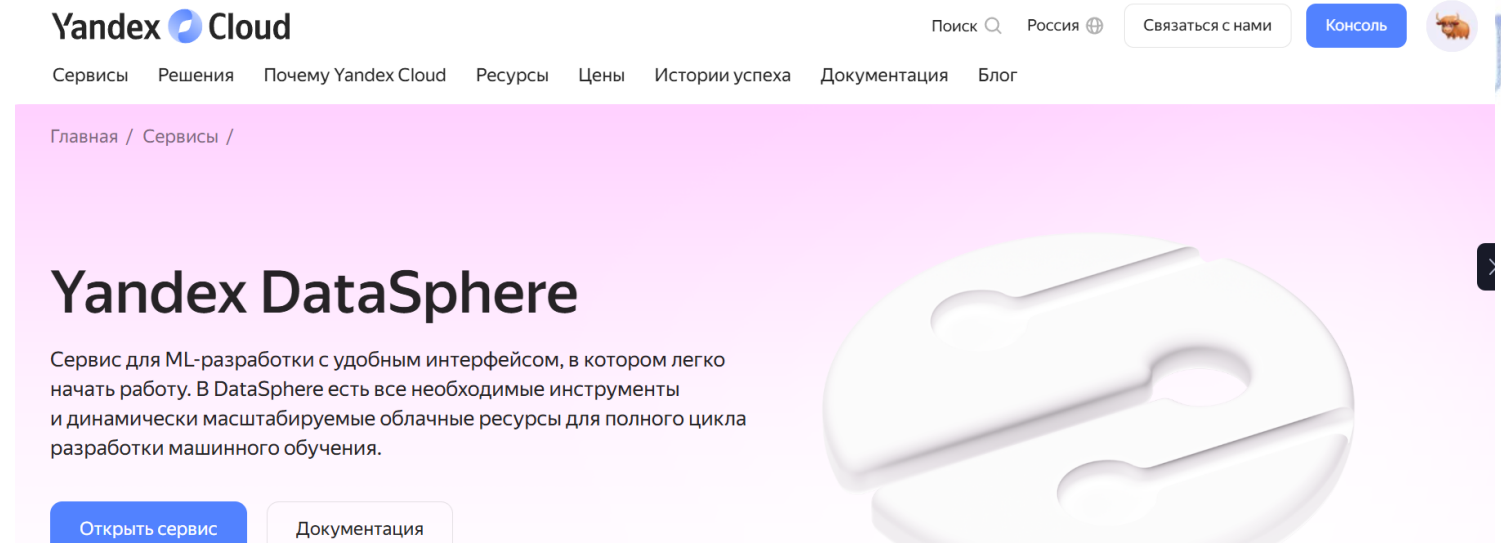


НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практикум. Просмотр изображений ChestX-ray-NIH на платформе в Yandex DataSphere

Перейдите на сайт [Yandex DataSphere](https://datasphere.yandex.cloud) и войдите под своим Яндекс ID.

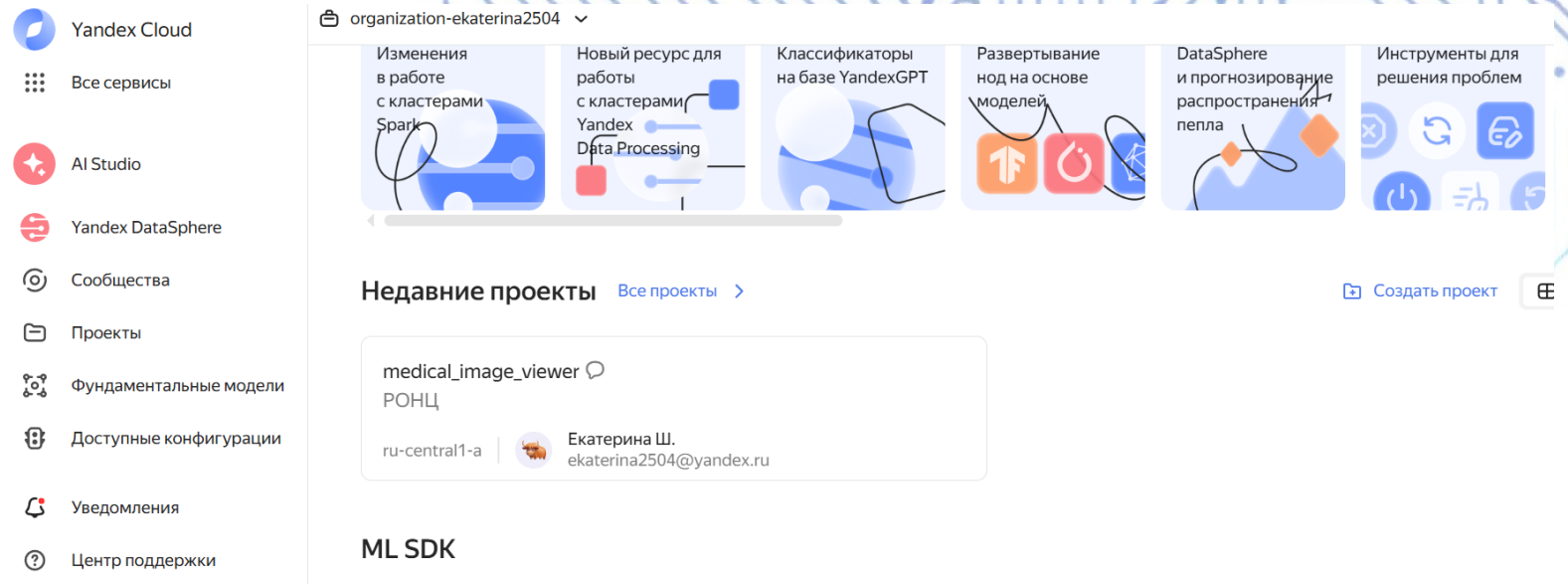


<https://datasphere.yandex.cloud>

<https://nihcc.app.box.com/v/ChestXray-NIHCC>

Практикум. Просмотр изображений ChestX-ray-NIH на платформе в Yandex DataSphere

В консоли нажмите кнопку **«Создать проект»**.



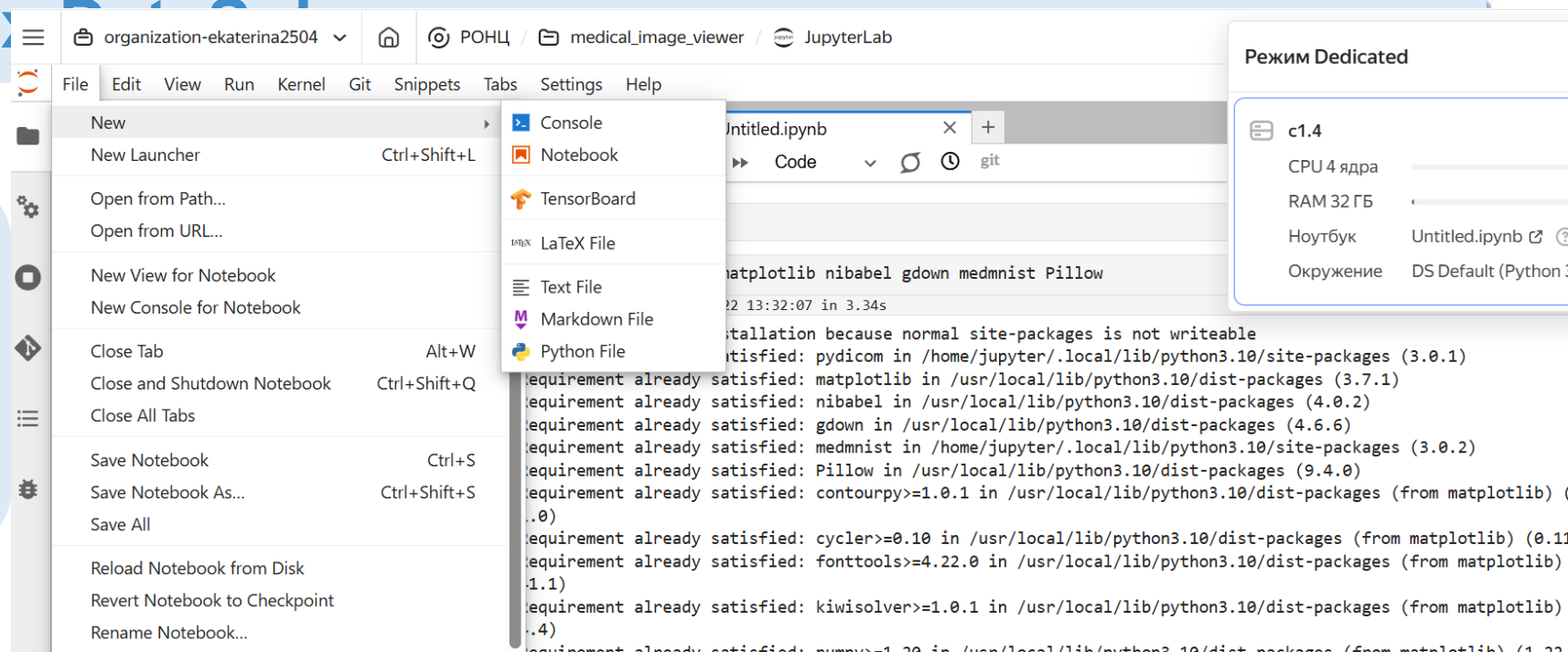
<https://datasphere.yandex.cloud>

<https://nihcc.app.box.com/v/ChestXray-NIHCC>

Тема 1.1

Практикум. Просмотр изображений ChestX-ray-NIH на платформе в Yandex DataSphere

Нажмите на название созданного проекта, чтобы открыть его. Затем нажмите кнопку **«Открыть ноутбук»** или **«Создать ноутбук»**. Откроется знакомая среда Jupyter.



<https://datasphere.yandex.cloud>

<https://nihcc.app.box.com/v/ChestXray-NIHCC>



ФГБНЦ
ООНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практикум. Просмотр изображений ChestX-ray-NIH на платформе в Yandex DataSphere

Загрузите библиотеки

Установка библиотек

```
[2]: %pip install pydicom matplotlib nibabel gdown medmnist Pillow
```

Last executed at 2026-02-22 13:32:07 in 3.34s

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: pydicom in /home/jupyter/.local/lib/python3.9/site-packages
Requirement already satisfied: matplotlib in /home/jupyter/.local/lib/python3.9/site-packages
Requirement already satisfied: nibabel in /home/jupyter/.local/lib/python3.9/site-packages
Requirement already satisfied: gdown in /home/jupyter/.local/lib/python3.9/site-packages
Requirement already satisfied: medmnist in /home/jupyter/.local/lib/python3.9/site-packages
Requirement already satisfied: Pillow in /home/jupyter/.local/lib/python3.9/site-packages

<https://datasphere.yandex.cloud>

<https://nihcc.app.box.com/v/ChestXray-NIHCC>



ФГБНУ
ВНИИМ
им. Н.Н. Блохина

Тема 1.1

Практикум. Просмотр изображений ChestX-ray-NIH на платформе в Yandex DataSphere

Вставьте код для скачивания коллекции – выберите нужную.

<https://datasphere.yandex.cloud>

<https://nihcc.app.box.com/v/ChestXray-NIHCC>

The screenshot shows a JupyterLab environment. On the left, a file explorer displays a directory structure with files like 'images_01', 'Untitled.ipynb', 'med.py', 'images_02.t...', and 'dataset.zip'. The 'Untitled.ipynb' file is selected. On the right, a code editor shows Python code for downloading zip files from a list of URLs. The code includes comments in Russian and uses the 'urllib.request' module.

```
# Download the 56 zip files in Images_png in batches
import urllib.request

# URLs for the zip files
links = [
    'https://nihcc.box.com/shared/static/vfk49d74nhbxq3nqjg0900w5nvkorp5c...',
    'https://nihcc.box.com/shared/static/i28rlmbvmfjb18p2n3ril0pptcmcu9d1.gz',
    'https://nihcc.box.com/shared/static/f1t00wrtdk94satdfb9olcolqx20z2jp.gz',
    'https://nihcc.box.com/shared/static/0aowwzs51hjrceb3qp67ahp0rd111etg.gz',
    'https://nihcc.box.com/shared/static/v5e3goj22zr6h8tzualxfsq1qaygfbns.gz',
    'https://nihcc.box.com/shared/static/asi7ikud9jwnkrnkj99jnpfkjdes7161.gz',
    'https://nihcc.box.com/shared/static/jn1b4mw4n6lnh74ovmcjb8y48h8xj07n.gz',
    'https://nihcc.box.com/shared/static/tvpxmn7qyrg10w8wfh9kqfjkskv6nmm1j.gz',
    'https://nihcc.box.com/shared/static/upyy3m17qdumlgk2rfcvlb9k6gvqq2pj.gz',
    'https://nihcc.box.com/shared/static/l6nilvfa9cg3s28tqv1qc1o1m3gnz54p.gz',
    'https://nihcc.box.com/shared/static/hhq8fdgvcari67vfhs7ppg2w6ni4jze.gz',
    'https://nihcc.box.com/shared/static/ioqwi20ihqwy8pf4c24eazhh281pbu.gz'
]

for idx, link in enumerate(links):
    fn = 'images_%02d.tar.gz' % (idx+1)
    print('downloading'+fn+'...')
```



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практикум. Просмотр изображений ChestX-ray-NIH на платформе в Yandex DataSphere

Выполните распаковку

```
[2]: import tarfile
import os

# Распаковка
with tarfile.open('images_02.tar.gz', 'r:gz') as tar:
    tar.extractall(path='images_01')
print("✅ Распаковано в папку 'images_01'")
print("Количество файлов в папке:", len(os.listdir('images_01')))
```

Last executed at 2026-02-22 13:57:43 in 4m 0.66s

✅ Распаковано в папку 'images_01'

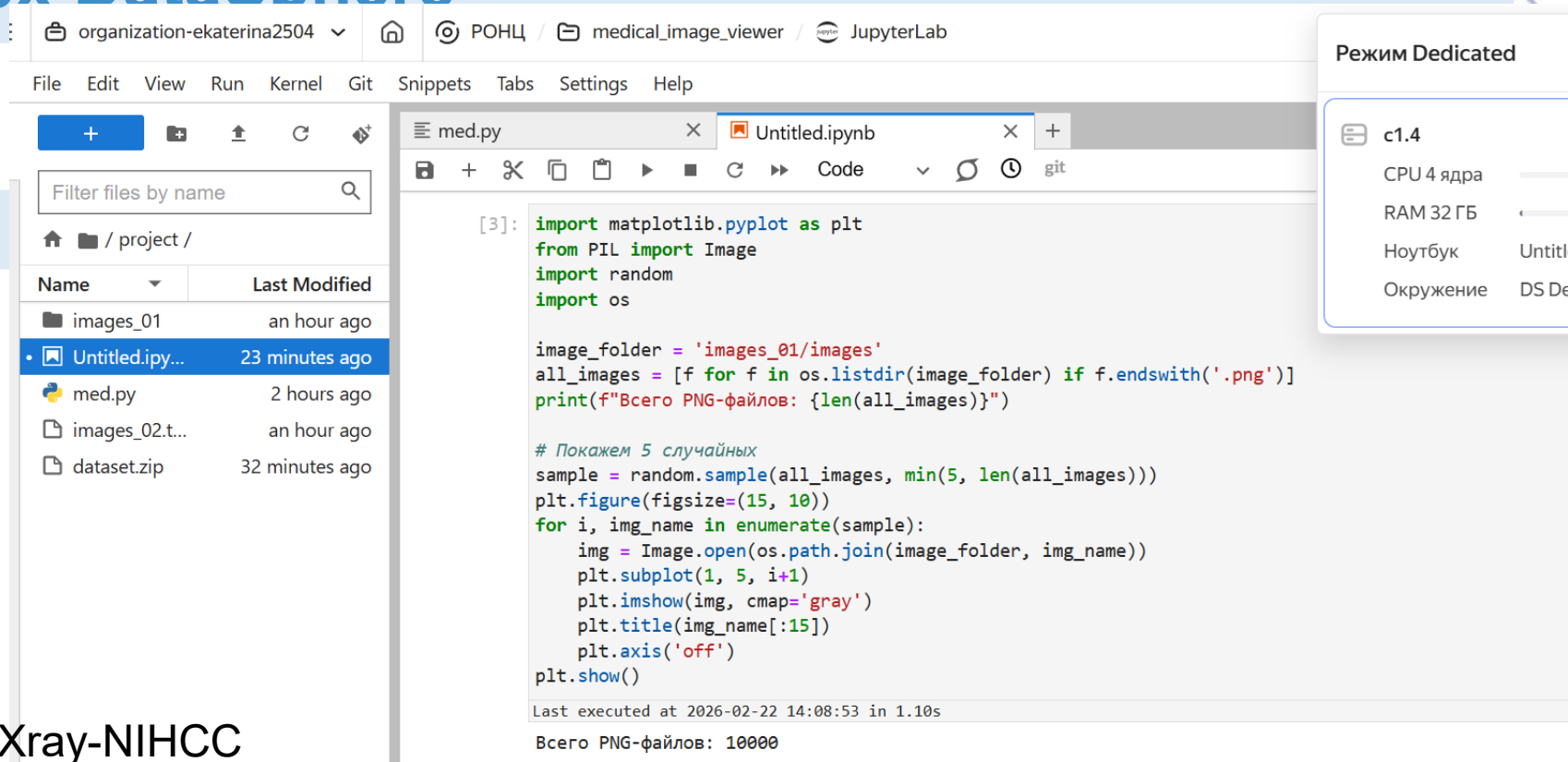
<https://datasphere.yandex.cloud>

<https://nihcc.app.box.com/v/ChestXray-NIHCC>

Тема 1.1

Практикум. Просмотр изображений ChestX-ray-NIH на платформе в Yandex DataSphere

Выполните вывод изображений



The screenshot displays the Yandex DataSphere JupyterLab environment. The top bar shows the user 'organization-ekaterina2504' and the current workspace 'medical_image_viewer'. The left sidebar contains a file explorer with a search bar and a list of files: 'images_01' (an hour ago), 'Untitled.ipynb' (23 minutes ago), 'med.py' (2 hours ago), 'images_02.t...' (an hour ago), and 'dataset.zip' (32 minutes ago). The main area is a code editor for 'Untitled.ipynb' with the following Python code:

```
[3]: import matplotlib.pyplot as plt
from PIL import Image
import random
import os

image_folder = 'images_01/images'
all_images = [f for f in os.listdir(image_folder) if f.endswith('.png')]
print(f"Всего PNG-файлов: {len(all_images)}")

# Покажем 5 случайных
sample = random.sample(all_images, min(5, len(all_images)))
plt.figure(figsize=(15, 10))
for i, img_name in enumerate(sample):
    img = Image.open(os.path.join(image_folder, img_name))
    plt.subplot(1, 5, i+1)
    plt.imshow(img, cmap='gray')
    plt.title(img_name[:15])
    plt.axis('off')
plt.show()
```

The output at the bottom shows 'Всего PNG-файлов: 10000'. On the right, a 'Режим Dedicated' panel displays the environment configuration: 'c1.4', 'CPU 4 ядра', 'RAM 32 ГБ', 'Ноутбук', and 'Окружение DS De'.

<https://datasphere.yandex.cloud>

<https://nihcc.app.box.com/v/ChestXray-NIHCC>



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Практикум. Просмотр изображений ChestX-ray-NIH на платформе в Yandex DataSphere

organization-ekaterina2504 POHЦ / medical_image_viewer JupyterLab

Всего PNG-файлов: 10000

00001819_004.pn



00003460_001.pn



00001409_010.pn



00002051_003.pn



00002451_015.pn



<https://datasphere.yandex.cloud>

<https://nihcc.app.box.com/v/ChestXray-NIHCC>

```
plt.axis('off')  
plt.show()
```

Last executed at 2026-02-22 14:08:53 in 1.10s

Всего PNG-файлов: 10000



НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Другие открытые наборы данных

Mosmed

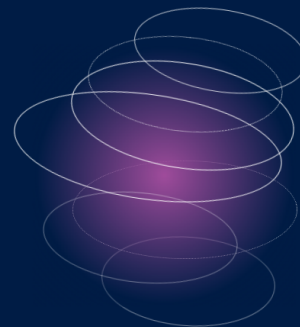
mosmed.ai — это самый прямой аналог NIH, с удобным интерфейсом и разнообразными датасетами. Там же есть и КТ, и МРТ, и маммография .

<https://mosmed.ai/>



Технологии искусственного интеллекта в здравоохранении

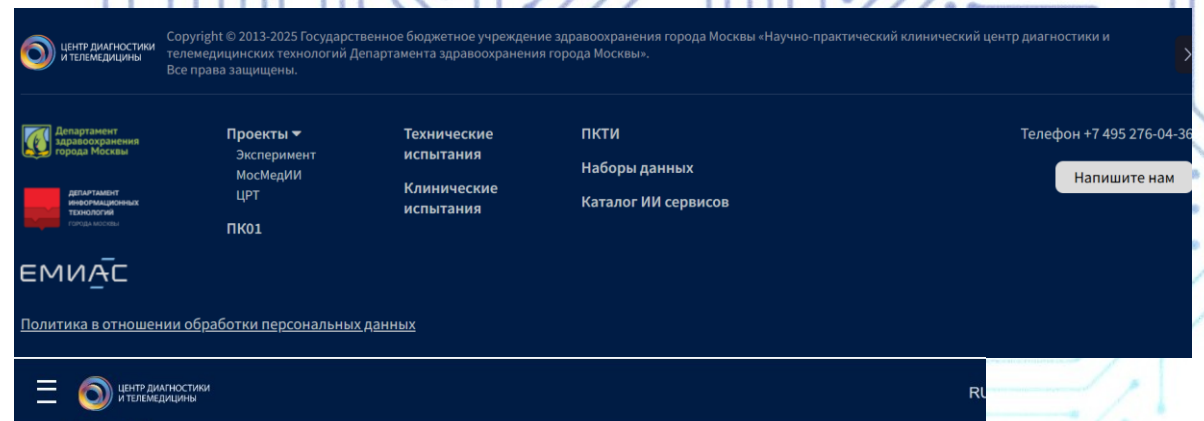
Научное исследование: использование методов поддержки врачебных решений на основе искусственного интеллекта



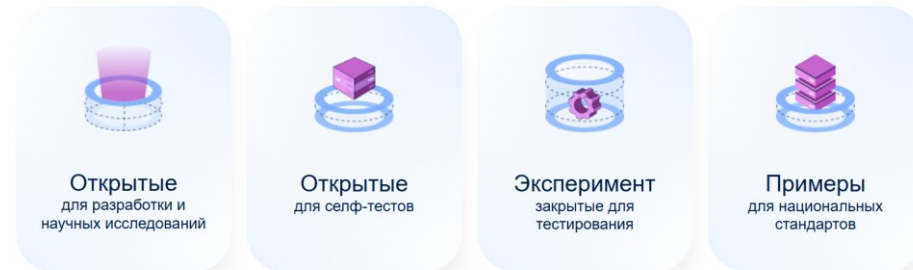
Другие открытые наборы данных

Mosmed

Откройте ссылку на набор данных .
Выберите – Открытые данные



Наборы данных





НМИЦ
ОНКОЛОГИИ
им.Н.Н.Блохина

Тема 1.1

Другие открытые наборы данных

Mosmed

Выберите набор данных и скачайте.

Анатомическая область ▾

Назначение датасета ▾

Метод верификации:

- ☐ Лабораторное исследование
- ☐ Клинический диагноз
- ☐ Исследование той же модальности в динамике
- ☐ Пересмотр специалистом
- ☐ Анализ корреляционных характеристик сигнала
- ☐ Исследование другой модальности

Условия доступа:

Наименование

Целевая
патология

Анатомическая
область

КТ +3  Набор данных КТ, ММГ, РГ/ФЛГ с целью селф-тестирования ИИ-сервисов для поиска признаков приоритетных патологий
Записей: 41  602  7123

Мульти

Мульти

УЗД  Набор данных УЗИ фантома с целью обучения ИИ-сервисов для определения признаков мерцающего артефакта
Записей: 29  135  3730

Мульти

Фантомы

КТ MosMedData: результаты

COVID-19

Грудная



НМИЦ
ОНКОЛОГИИ
им. Н.Н. Блохина

Тема 1.1

Другие открытые наборы данных

Mosmed

Выберите набор данных и скачайте. Либо скопируйте ссылку.

Инструкция по доступу к Набор данных КТ, ММГ, РГ/ФЛГ с целью селф-тестирования ИИ-сервисов для поиска признаков приоритетных патологий



MOSMED mosmed@noreply.telemedai.ru Сегодня в 14:15 ★
Я >

Пересказ от Алисы Про

MOSMED сообщает, что для скачивания набора данных КТ, ММГ, РГ, ФЛГ для тестирования ИИ-сервисов нужно обязательно указать источник при публикации.

✓ Скрыть

📧 Неважное письмо

📧 Плохой пересказ

Название: Набор данных КТ, ММГ, РГ, ФЛГ с целью селф-тестирования ИИ-сервисов для поиска признаков приоритетных патологий

Условия использования: обязательное цитирование при публикации и публичном представлении алгоритмов, программного обеспечения и иных объектов, разработанных с использованием набора данных

Для скачивания датасета перейдите по

[Скачать](https://mosmed.ai/)

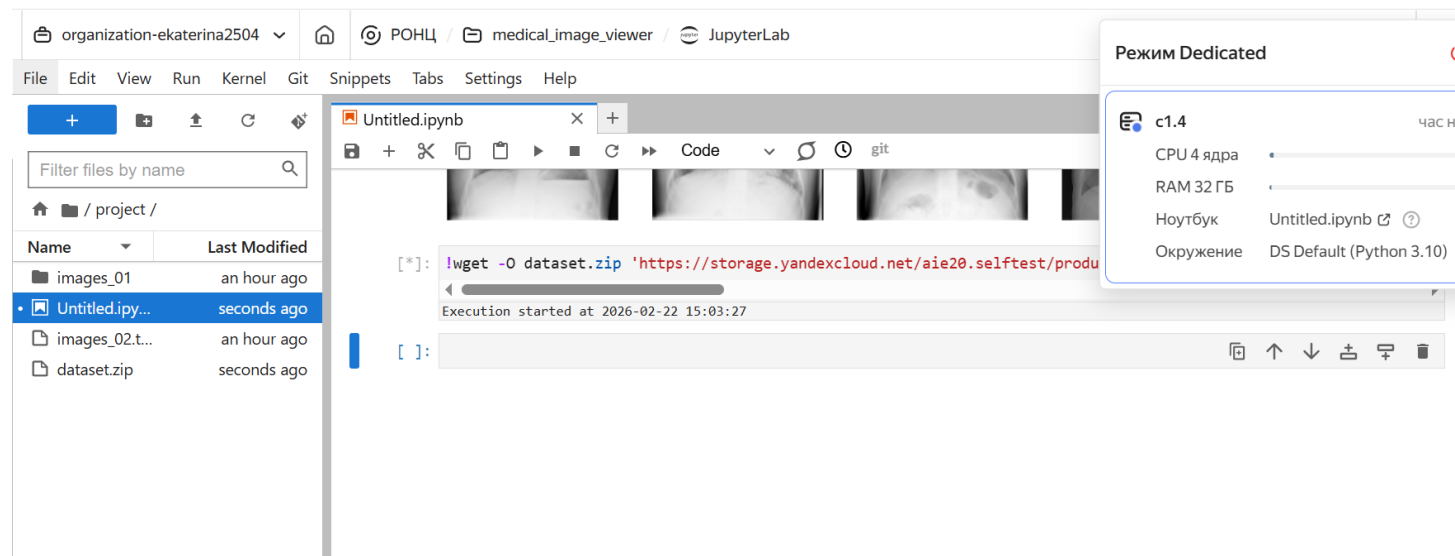
<https://mosmed.ai/>

Тема 1.1

Другие открытые наборы данных

Mosmed

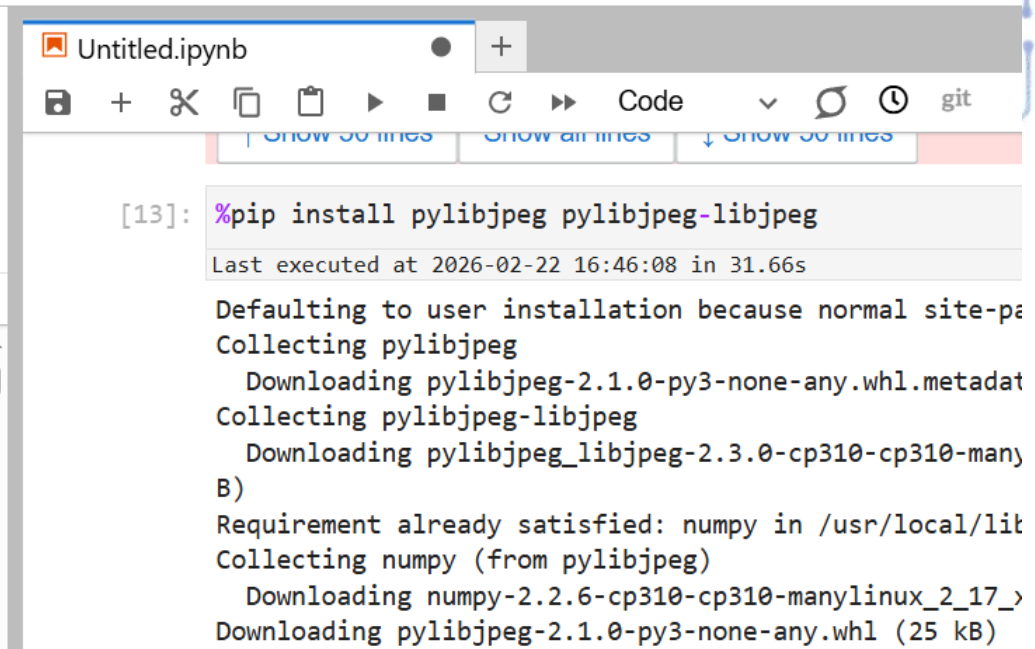
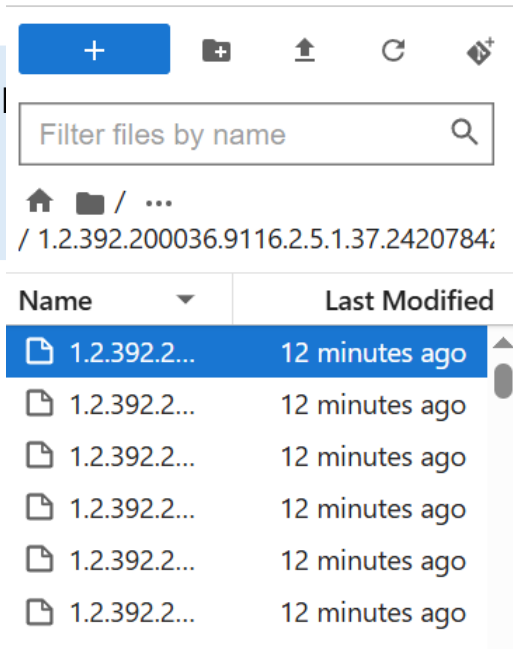
Скачайте коллекцию



Другие открытые наборы данных

Mosmed

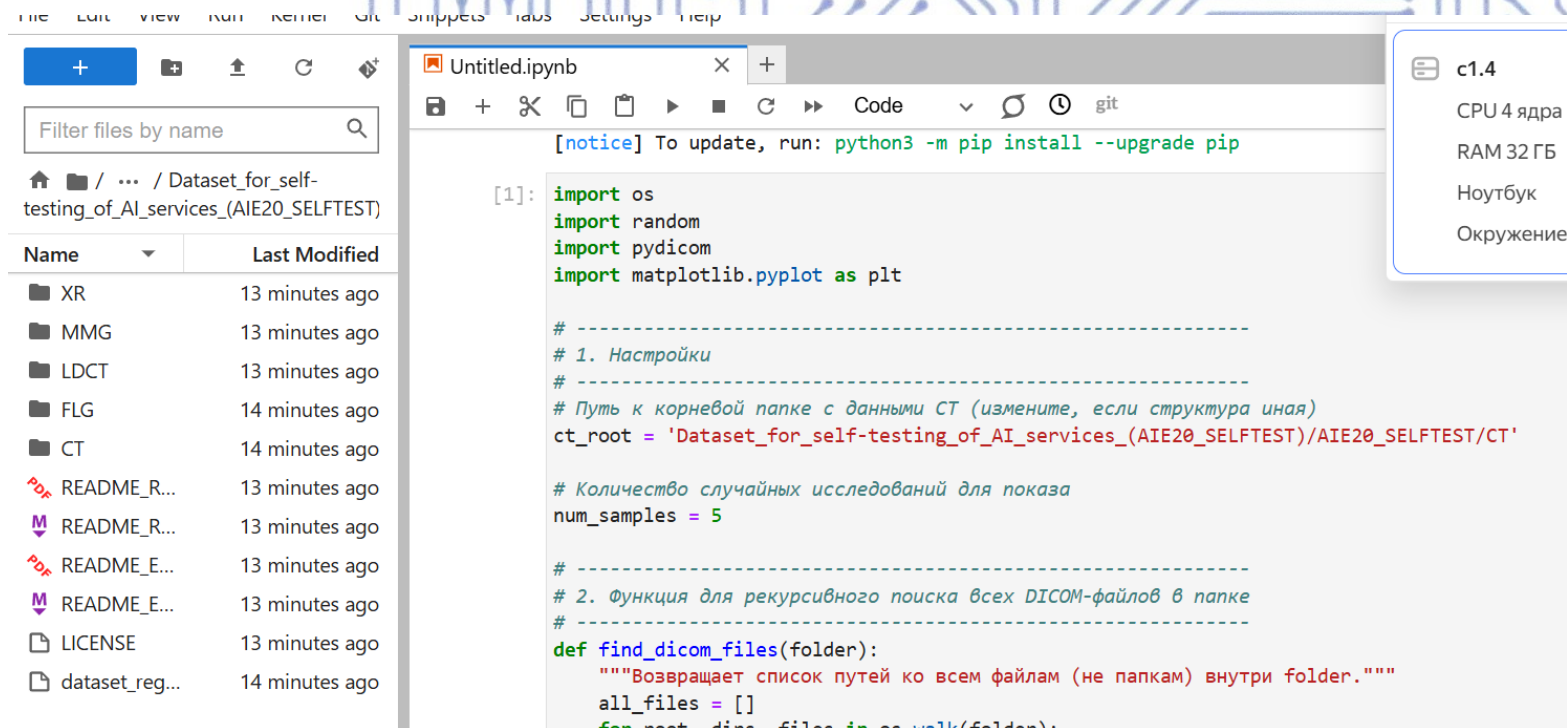
Распакуйте изображения и доустановите библиотеки для работы с данными изображениями



Другие открытые наборы данных

Mosmed

Прочитайте о видах изображений, выберите нужный тип и запустите код просмотра изображений



```
[notice] To update, run: python3 -m pip install --upgrade pip

[1]: import os
import random
import pydicom
import matplotlib.pyplot as plt

# -----
# 1. Настройки
# -----
# Путь к корневой папке с данными CT (измените, если структура иная)
ct_root = 'Dataset_for_self-testing_of_AI_services_(AIE20_SELFTEST)/AIE20_SELFTEST/CT'

# Количество случайных исследований для показа
num_samples = 5

# -----
# 2. Функция для рекурсивного поиска всех DICOM-файлов в папке
# -----
def find_dicom_files(folder):
    """Возвращает список путей ко всем файлам (не папкам) внутри folder."""
    all_files = []
    for root, dirs, files in os.walk(folder):
```



A decorative graphic at the bottom of the page consisting of a complex, symmetrical pattern of blue lines that resemble a circuit board or a stylized mountain range. The lines are interconnected, forming various loops and paths.

Mosmed

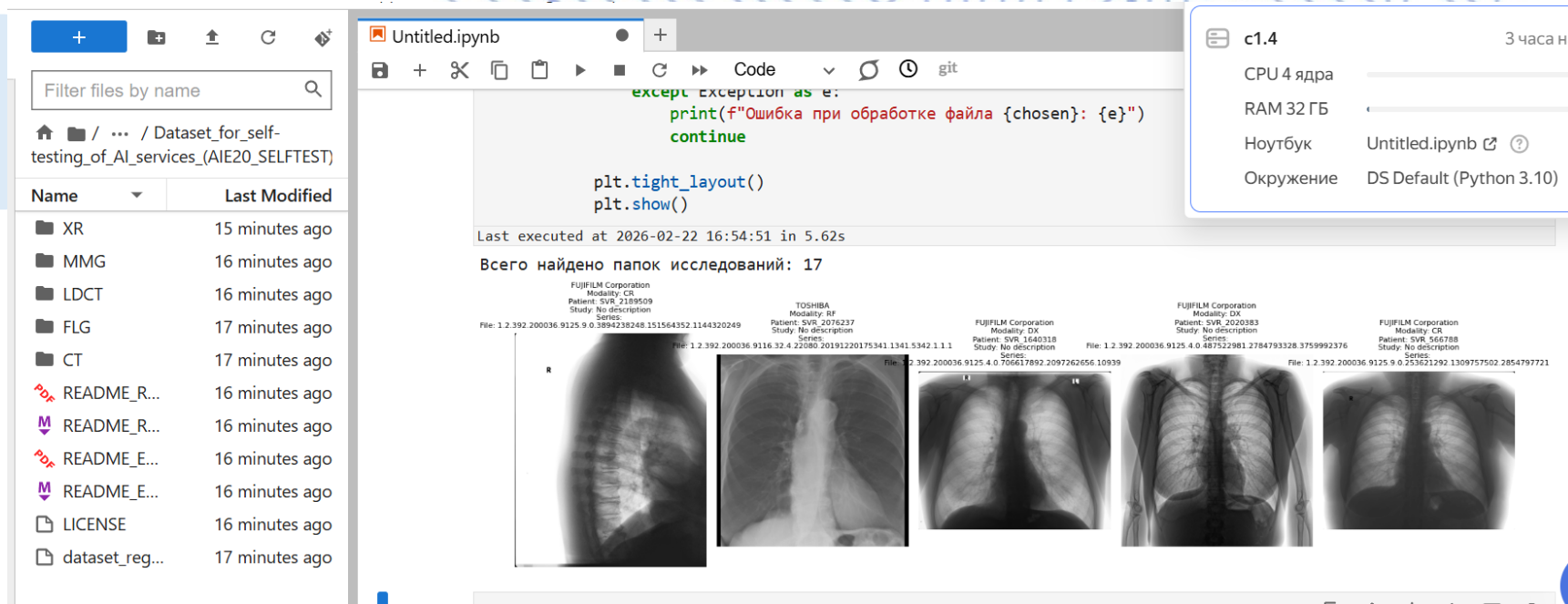
Прочитайте о видах изображений, выберите нужный тип и запустите код просмотра

Тема 1.1

Другие открытые наборы данных

Mosmed

Прочитайте о видах изображений, выберите нужный тип и запустите код просмотра изображений



The screenshot displays the Mosmed web interface. On the left, a file explorer shows a directory structure with folders like XR, MMG, LDCT, FLG, and CT, along with README files and a LICENSE. The main area features a code editor with Python code for handling exceptions and displaying images using Matplotlib. Below the code, a message states "Всего найдено папок исследований: 17". At the bottom, a grid of medical images is shown, including chest X-rays and CT scans, each with associated patient and study information.

```
except Exception as e:
    print(f"Ошибка при обработке файла {chosen}: {e}")
    continue

plt.tight_layout()
plt.show()
```

Last executed at 2026-02-22 16:54:51 in 5.62s

Всего найдено папок исследований: 17

File: 1.2.392.200036.9125.9.0.3894238248.151564352.1144320249
FUJIFILM Corporation
Modality: CR
Patient: SVR_2185509
Study: No description
Series:

File: 1.2.392.200036.9116.32.4.22080.20191220175341.1341.5342.1.1.1
TOSHIBA
Modality: RF
Patient: SVR_2076237
Study: No description
Series:

File: 1.2.392.200036.9125.4.0.706617892.2097262656.10939
FUJIFILM Corporation
Modality: DX
Patient: SVR_1640318
Study: No description
Series:

File: 1.2.392.200036.9125.4.0.487522981.2784793328.3759992376
FUJIFILM Corporation
Modality: DX
Patient: SVR_2020383
Study: No description
Series:

File: 1.2.392.200036.9125.9.0.253621292.1309757502.2854797721
FUJIFILM Corporation
Modality: CR
Patient: SVR_566788
Study: No description
Series:

Спасибо за внимание!



Шувалова
Е.В.
Доцент, к.ф.-м.н.

Контакты:

ekaterina2504@ya.ru