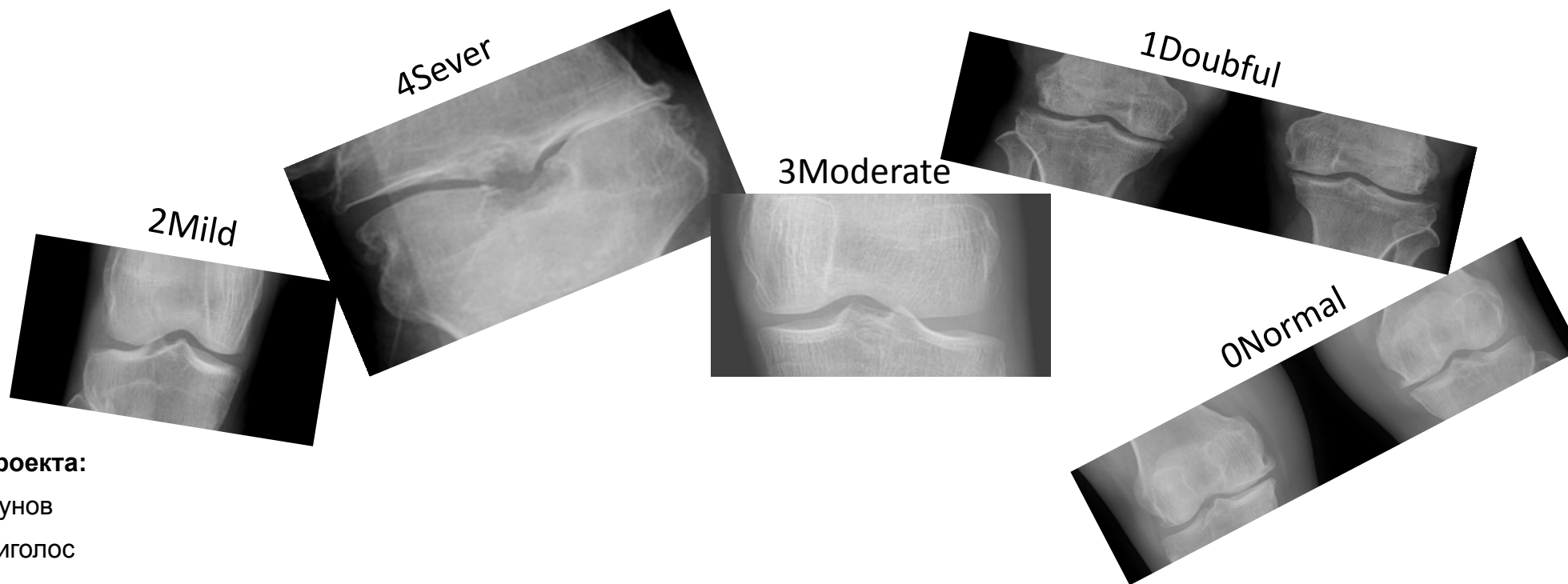


Классификация цифровых рентгеновских снимков суставов по степени поражения остеоартритом



Команда проекта:

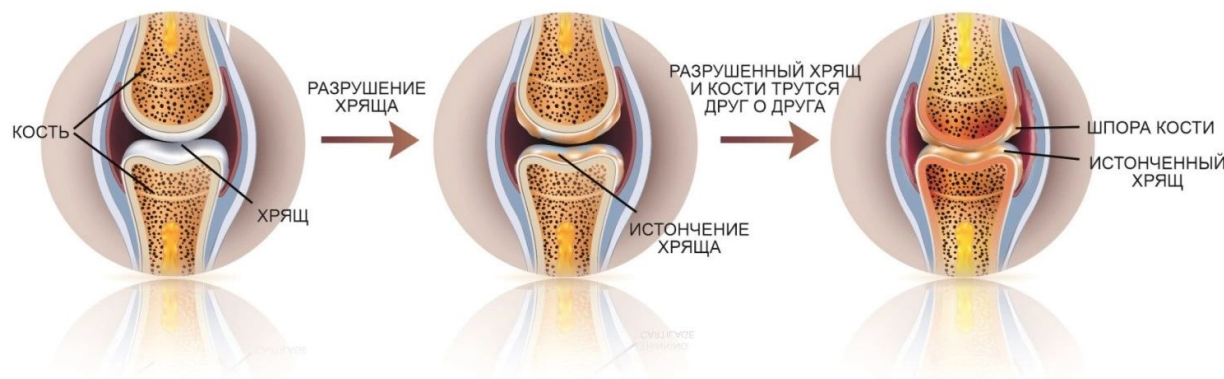
Николай Тиунов
Алексей Триголос
Иван Юминов

Куратор: Марк Блуменау

Описание проекта

- **Остеоартрит** – воспаление оболочки сустава, вследствие которого разрушаются хрящевая и костная ткани
- **Остеоартрит** – одно из ведущих заболеваний по распространенности среди ревматических
- Причиной 54% жалоб на боль в крупных суставов является **остеоартрит**
- 70% лиц старше 55 лет сталкиваются с заболеванием

ОСТЕОАРТРИТЫ



Цели и задачи проекта

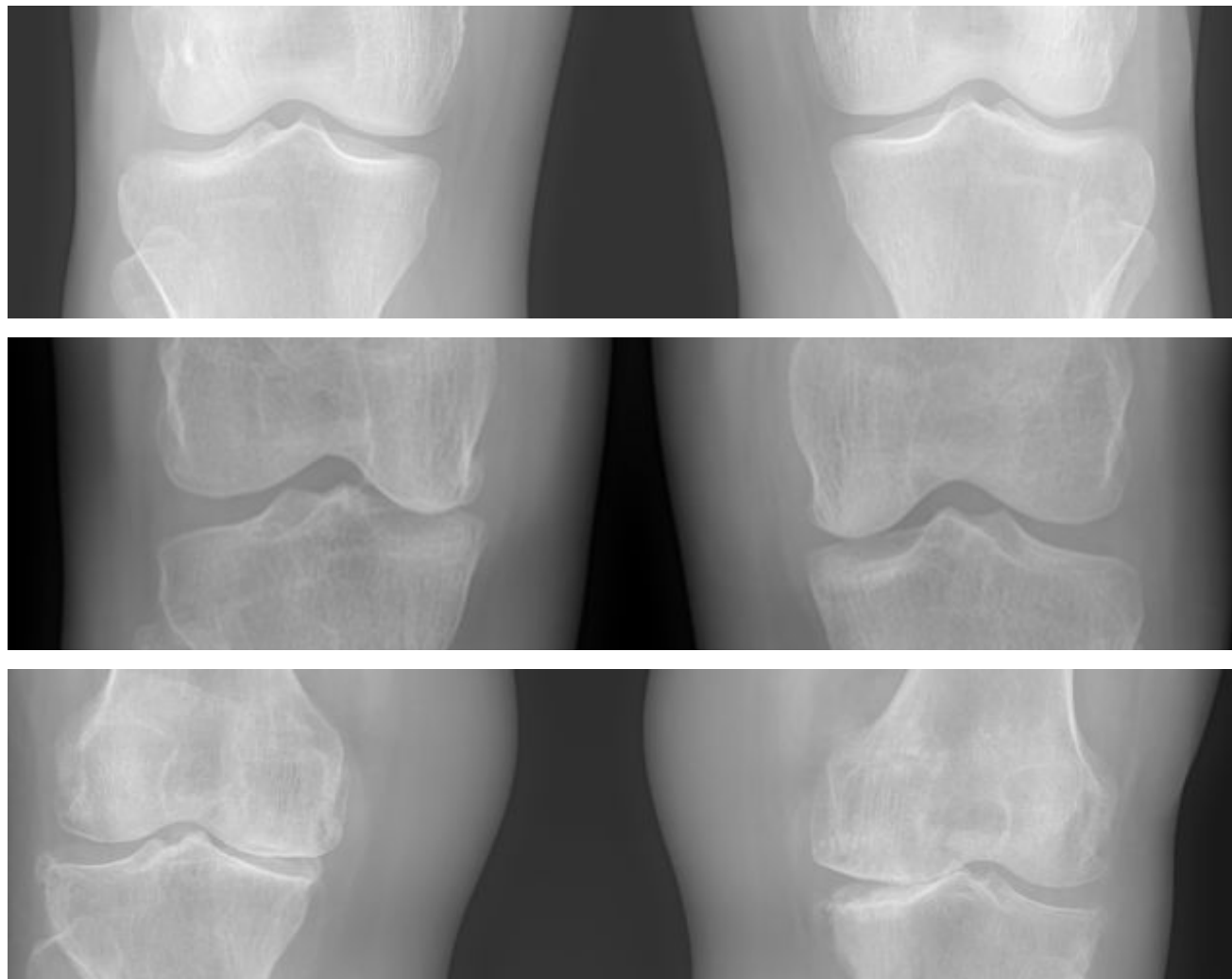
Цель проекта: создать сервис классификации рентгеновских снимков коленных суставов по степени поражения остеоартритом

Задачи:

1. Исследовать и подготовить данные
2. Построить модель для предсказания степеней поражения суставов
3. Создать пользовательский интерфейс
4. Реализовать интерпретацию предсказания модели для пользователя

Описание данных

- набор данных – это цифровые рентгеновские изображения коленного сустава
- одно наблюдение – это 8-битное изображение в оттенках серого
- набор данных собран специалистами университета Рани Чаннамма (Индия)
- изображения маркированы в соответствии с оценками Келлгрена и Лоуренса двумя медицинскими экспертами



Критерий оценивания снимков

Классификация Келлгрена и Лоуренса оценивает тяжесть остеоартрита коленного сустава

Создание критерия - 1957 г.

Утверждение критерия ВОЗ - 1961 г.

Усовершенствование критерия - 1982 г.

Категории критерия:

0 стадия – норма

1 стадия – сомнительная:

- сомнительное сужение высоты суставной щели
- сомнительные мелкие наросты (остеофиты)

2 стадия - мягкая:

- достоверные мелкие наросты,
- небольшие сужения суставной щели

3 стадия – умеренная:

- выраженные множественные наросты
- сужение высоты суставной щели
- возможная деформацией поверхностей

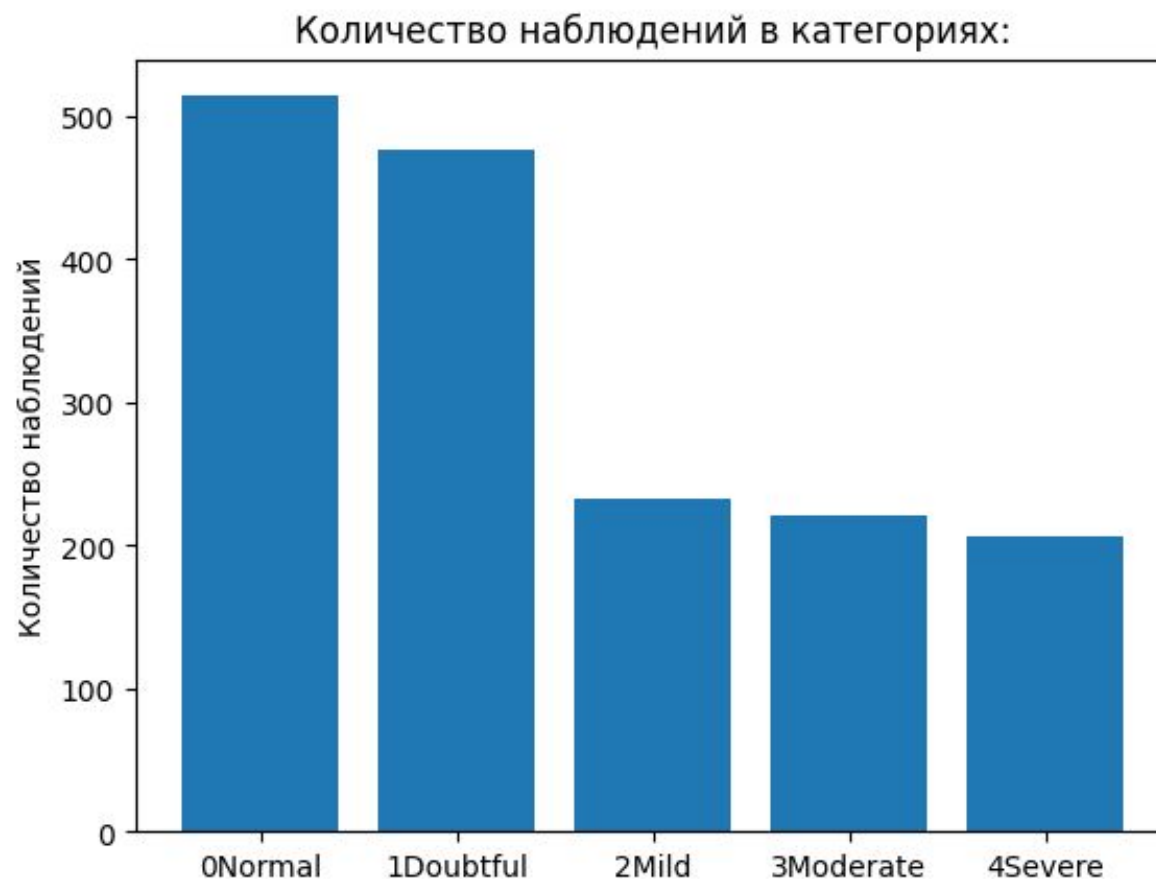
4 стадия – серьезная:

- крупные остеофиты
- выраженное сужение суставной щели
- деформация контуров кости



Результаты EDA

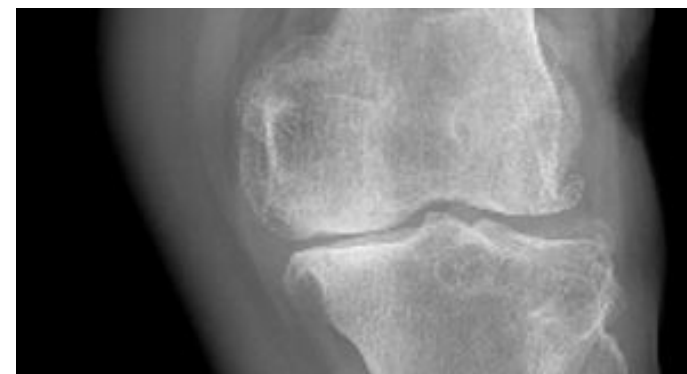
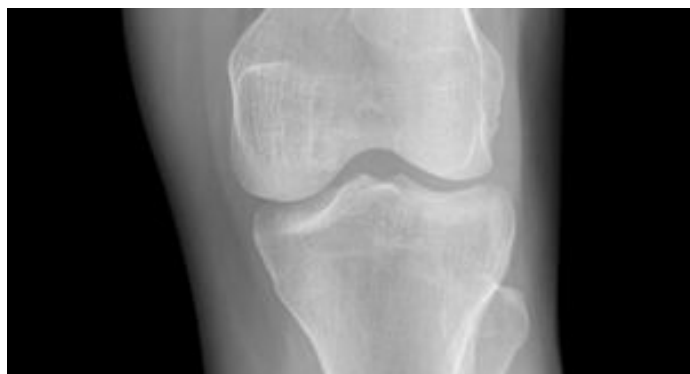
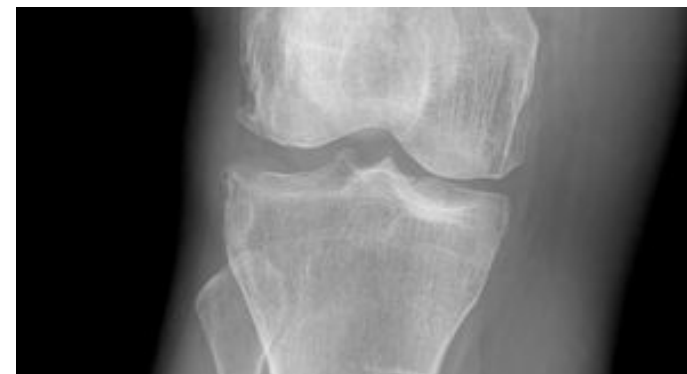
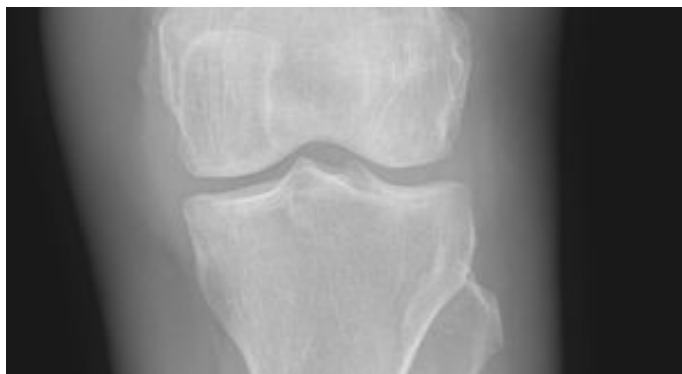
1. Набор данных состоит из 1650 наблюдений, которые классифицировались по 5-ти категориям двумя экспертами
2. Данные распределились по категориям следующим образом:



Результаты EDA

МОBC 2024

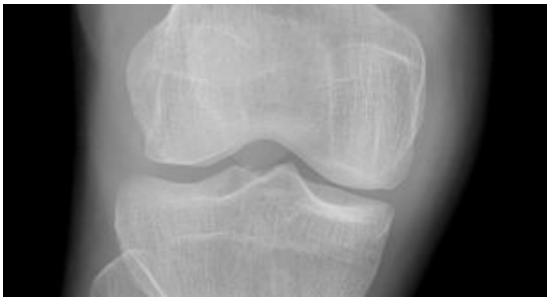
3. Выявлено 30 дубликатов
4. Снимки имеют разный угол поворота, например:



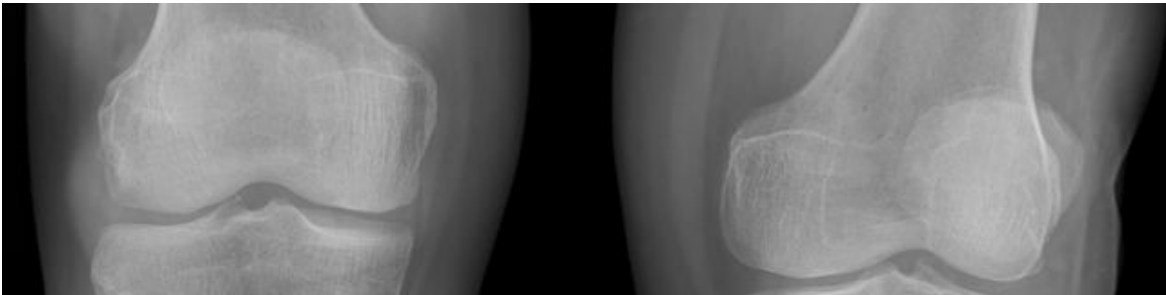
Результаты EDA

5. В наборе данных представлены снимки разного размера:

162x300 – снимок одного колена

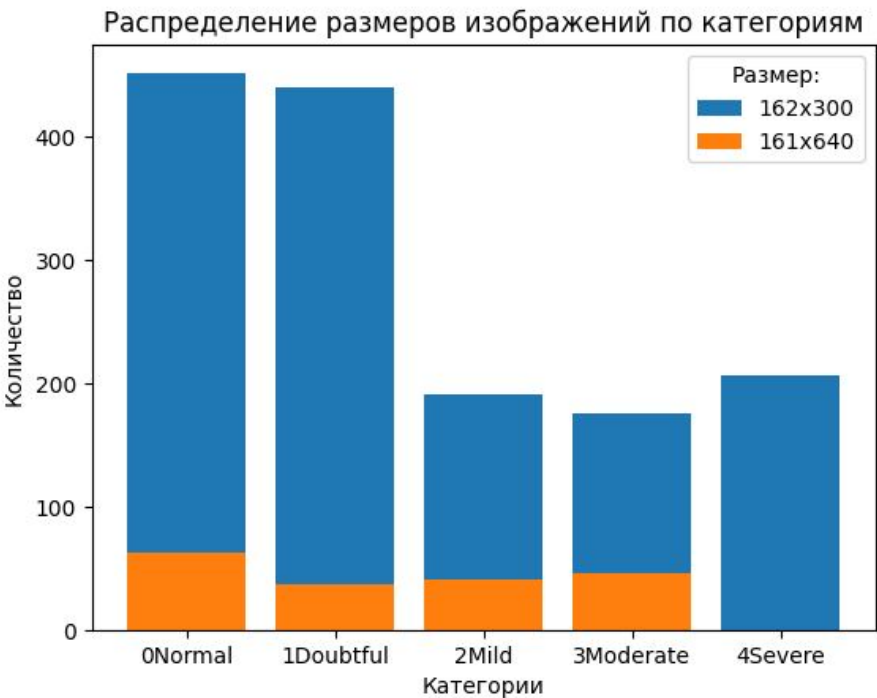


161x640 – снимок двух колен



Размеры изображений по категориям:

		0Normal	1Doubtful	2Mild	3Moderate	4Severe
162	300	452	440	191	175	206.0
161	640	62	37	41	46	0.0



Результаты EDA

6. Были определены средние изображения для каждой категории:

0 стадия



1 стадия



2 стадия



3 стадия



4 стадия

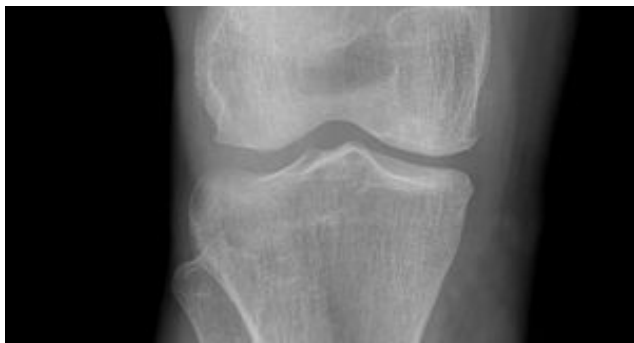


Результаты EDA

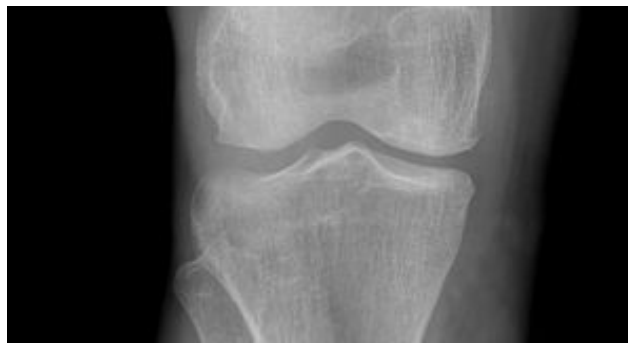
MOBC 2024

7. 11 наблюдений имеют разную классификацию экспертов. Например:

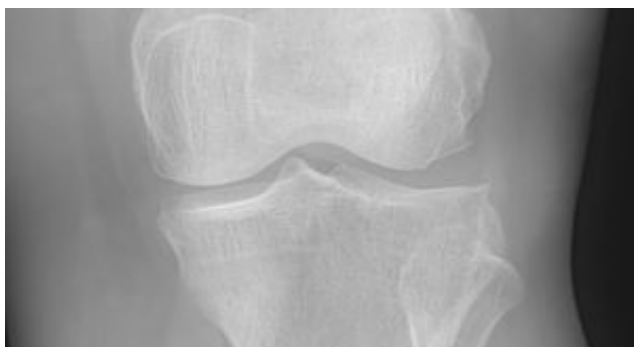
Medical Expert-I 0Normal



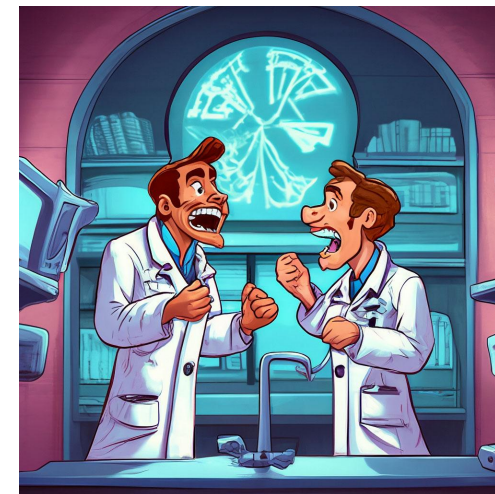
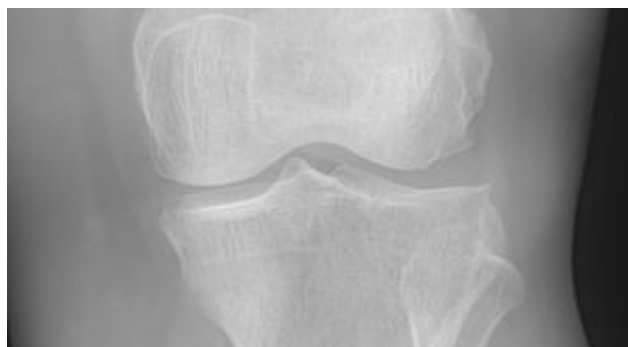
Medical Expert-II 1Doubtful



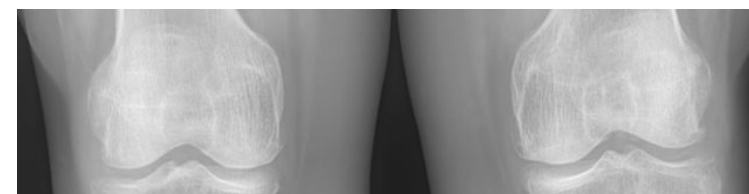
Medical Expert-I 0Normal



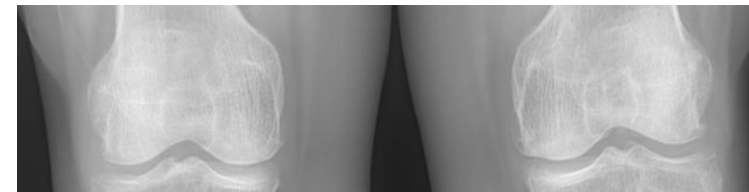
Medical Expert-II 1Doubtful



Medical Expert-I 0Normal

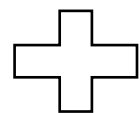


Medical Expert-II 1Doubtful

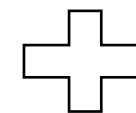


Модель HOG + CatBoost

HOG



CatBoost



HYPEROPT

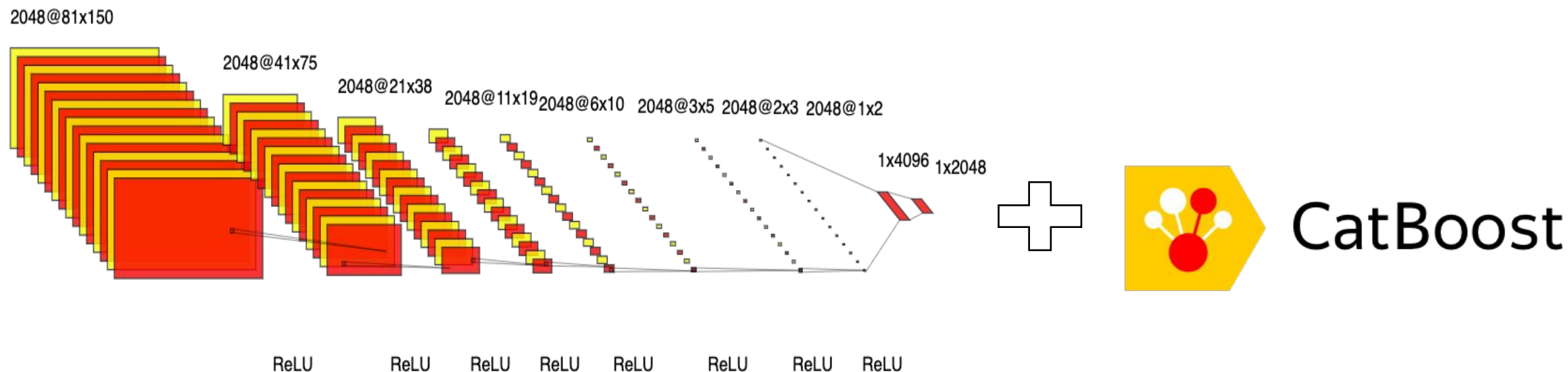
Реализация HOG: <https://scikit-image.org/docs/stable/api/skimimage.feature.html#skimimage.feature.hog>

Ссылка на модель: https://storage.yandexcloud.net/tnn-hse-medtech/models/hog_cat_boost_v3.cbm

Ссылка на код: https://github.com/TiunovNN/hse-knee-xray-research/blob/master/models/ML_hog/hog.ipynb

ROC-AUC: 0.98 / F1-score macro: 0.89

Модель Autoencoder + CatBoost



Ссылка на модель: https://storage.yandexcloud.net/tnn-hse-medtech/models/autoencoder_catboost.cbm

Ссылка на код:

https://github.com/TiunovNN/hse-knee-xray-research/blob/master/models/ML_autoencoder/Autoencoders.ipynb

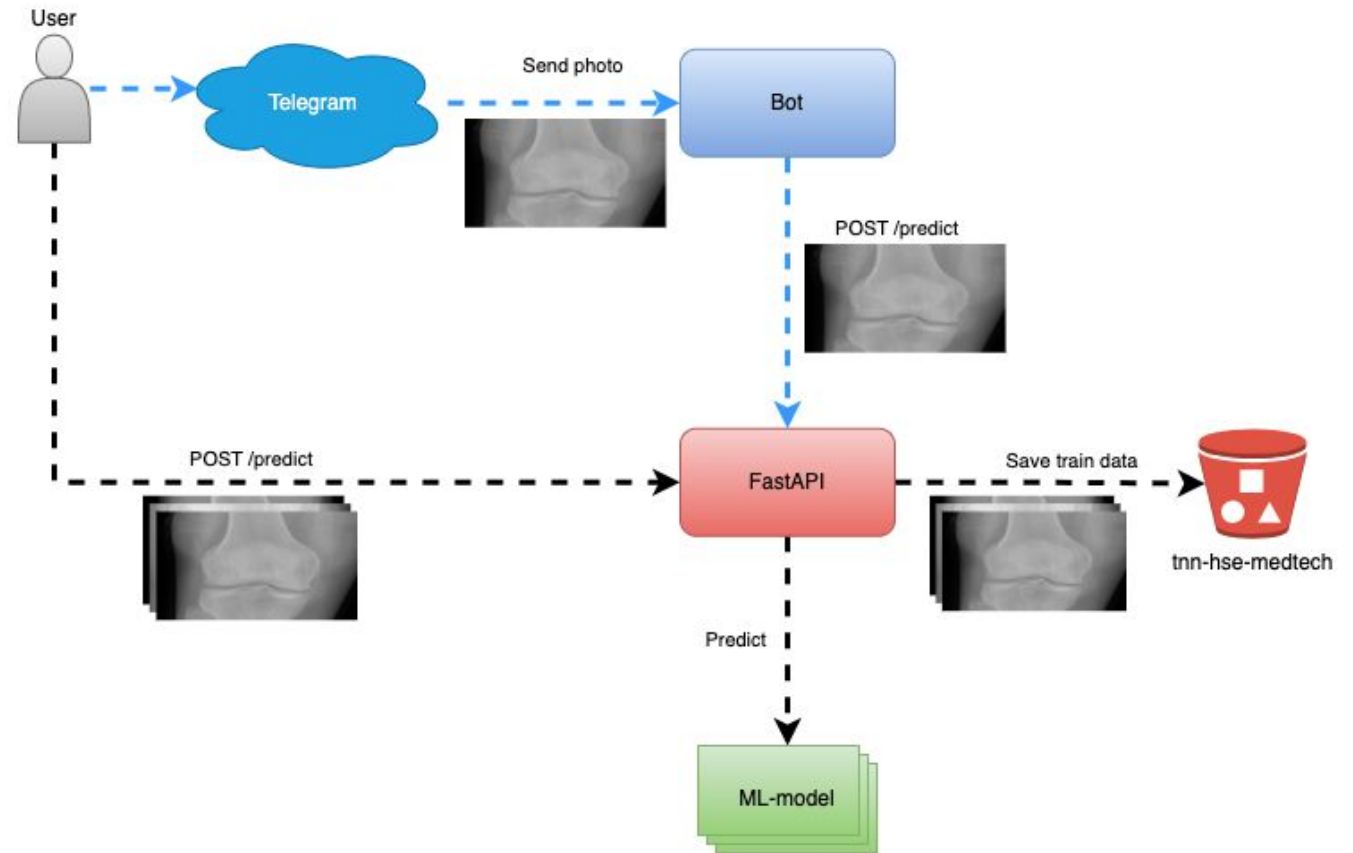
ROC-AUC: 0.84 / F1-score macro: 0.54

Описание сервиса

MOBC 2024

Компоненты:

1. Telegram-bot:
<https://t.me/KneeXrayBot>
2. API:
<http://hse-knee.tiunovnn.ru/docs>
3. S3 bucket (yandex cloud):
<https://tnn-hse-medtech.storage.yandexcloud.net>

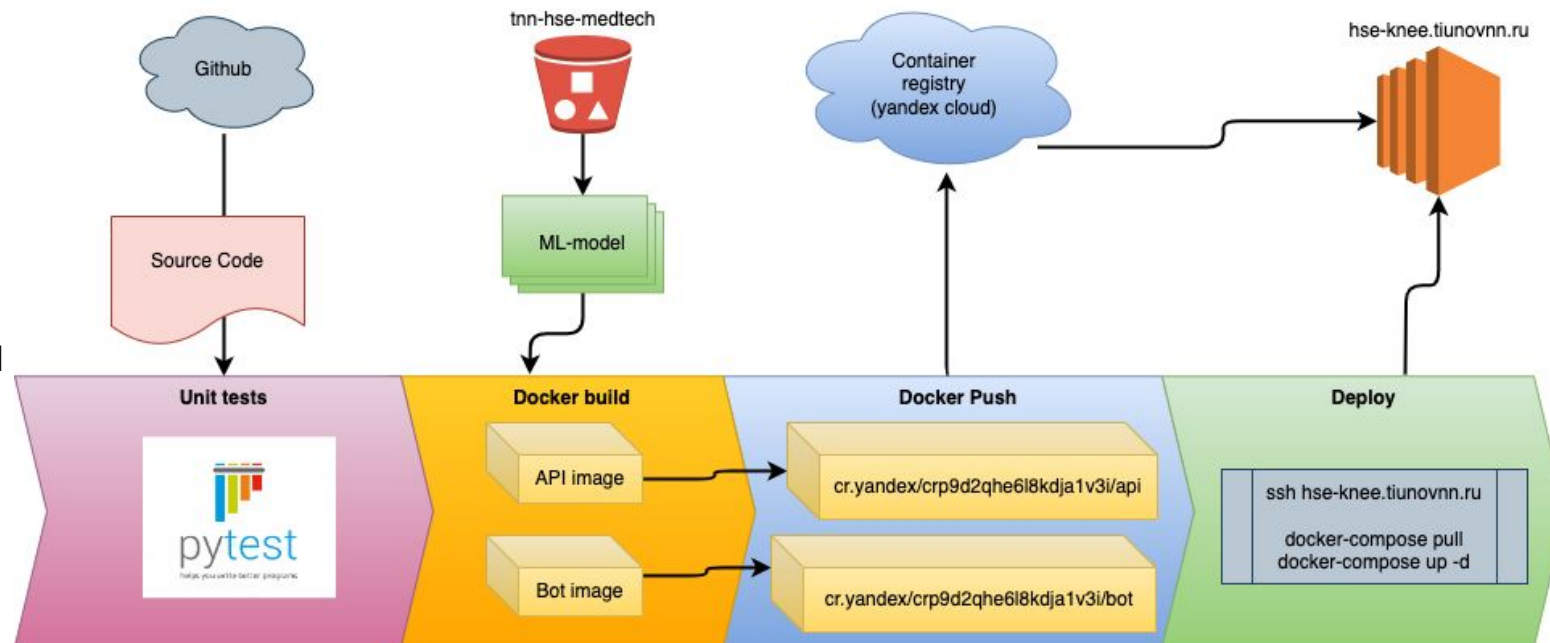


CI & CD

MOBC 2024

В репозитории настроен CI &CD:

1. На созданном Pull Request запускаются Unit-тесты
2. После мержа кода в master запускается сборка docker-образов
3. Далее эти образы заливаются в Container Registry расположенном в Yandex облаке
4. После заливка выполняется подключение в хосту, где работают сервисы и обновляются образа

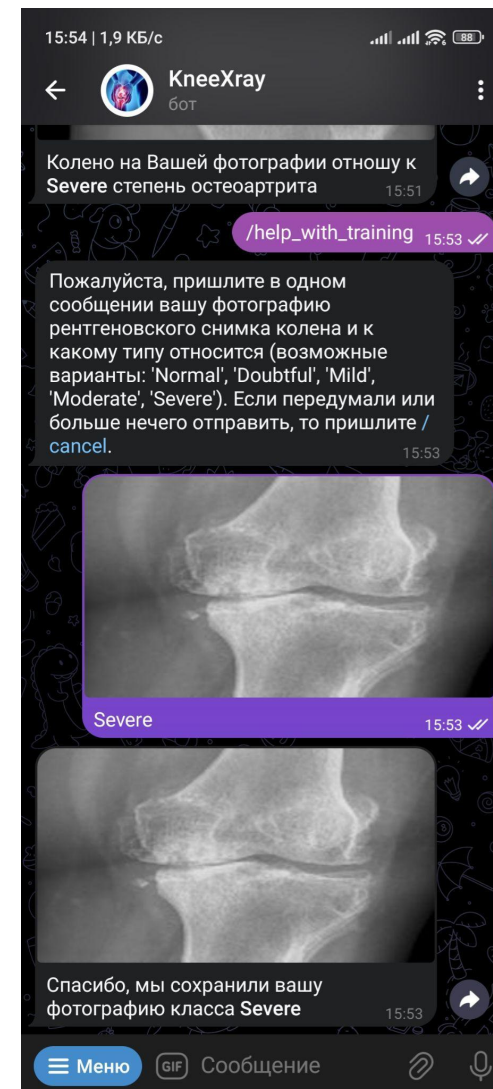
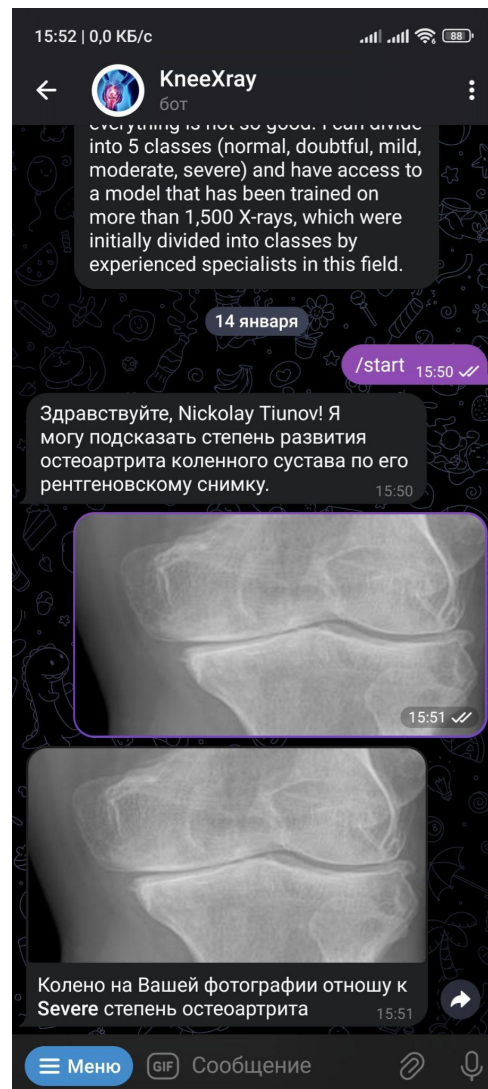


Компоненты:

1. ML-модель: https://storage.yandexcloud.net/tnn-hse-medtech/models/hog_cat_boost_v3.cbm
2. Код бота: https://github.com/TiunovNN/hse-knee-xray-research/tree/master/tg_bot
3. Код API: <https://github.com/TiunovNN/hse-knee-xray-research/tree/master/api>

Демонстрация бота

МОBC 2024



Демонстрация API

MOBC 2024

Curl

```
curl -X 'POST' \  
  'http://hse-knee.tiunovnn.ru/predict' \  
  -H 'accept: application/json' \  
  -H 'Content-Type: multipart/form-data' \  
  -F 'files=@DoubtfulG1 (95).png;type=image/png' \  
  -F 'files=@ModerateG3 (97).png;type=image/png'
```

Request URL

`http://hse-knee.tiunovnn.ru/predict`

Server response

Code

Details

200

Response body

```
[  
  {  
    "filename": "DoubtfulG1 (95).png",  
    "severity": 1  
  },  
  {  
    "filename": "ModerateG3 (97).png",  
    "severity": 3  
  }  
]
```



Download

Response headers

```
content-length: 97  
content-type: application/json  
date: Sun, 14 Jan 2024 13:00:09 GMT  
server: uvicorn
```


Прогресс

К защите в январе было сделано:

1. EDA
2. Обучена модель HOG+CatBoost
3. Написаны и развернуты Telegram Bot и API проекта
4. Настроена инфраструктура в yandex cloud: S3, Compute Cloud

К 1 марта было сделано:

1. Написаны тесты на бота
2. Настроен CI для прогона тестов
3. Настроен CI&CD для автоматической сборки и деплоя сервиса
4. Обучена модель на Autoencoder CNN + CatBoost

Планы

1. Продолжить эксперименты с автокодировщиками для улучшения качества предсказания
2. Написать тесты на API
3. Добавить автоматический запуск линтера в CI
4. Применение нейросетевой модели для классификации изображений
5. Применение методов интерпретации результатов модели (LIME, SHAP, GSM и тд.)