

# MediScan

Сервис для ранней диагностики  
злокачественных новообразований кожи,  
определяющий тип новообразования по  
фотографии

Выполнено:

Алексей Калягин, Мария Герасимова, Дмитрий Саперов,  
Студенты 1-го курса магистратуры МОВС, ФКН, НИУ ВШЭ



# Команда



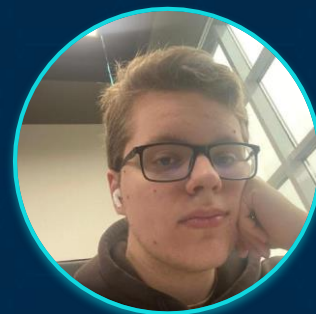
**Алексей  
Калягин**



**Мария  
Герасимова**



**Дмитрий  
Саперов**

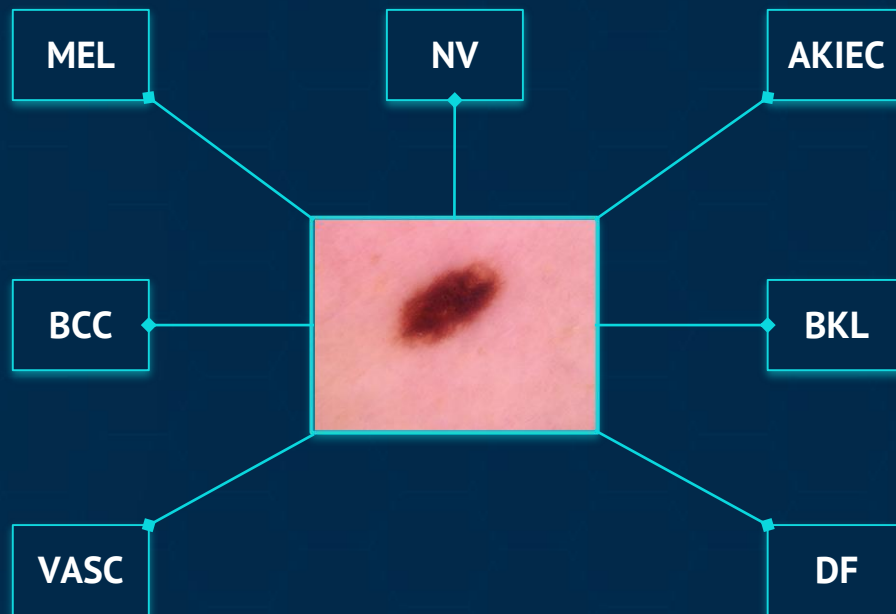


**Михаил  
Никифоров**

Куратор

# Описание задачи

Создание веб-сервиса с  
возможностью определения  
типа новообразования кожи  
по фотографии.



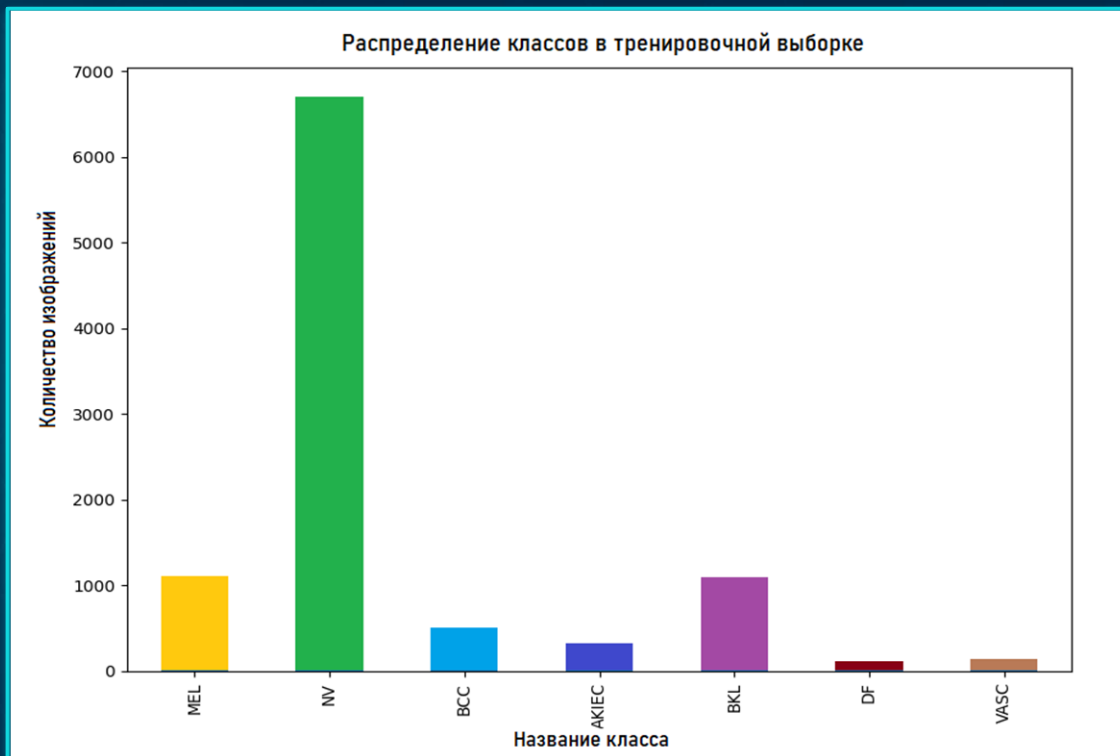
# Данные

## Датасет - ISIC2018 Task 3

- Тренировочная выборка – 10 015 изображений
- Тестовая выборка – 1 512 изображений
- Валидационная выборка – 193 изображения

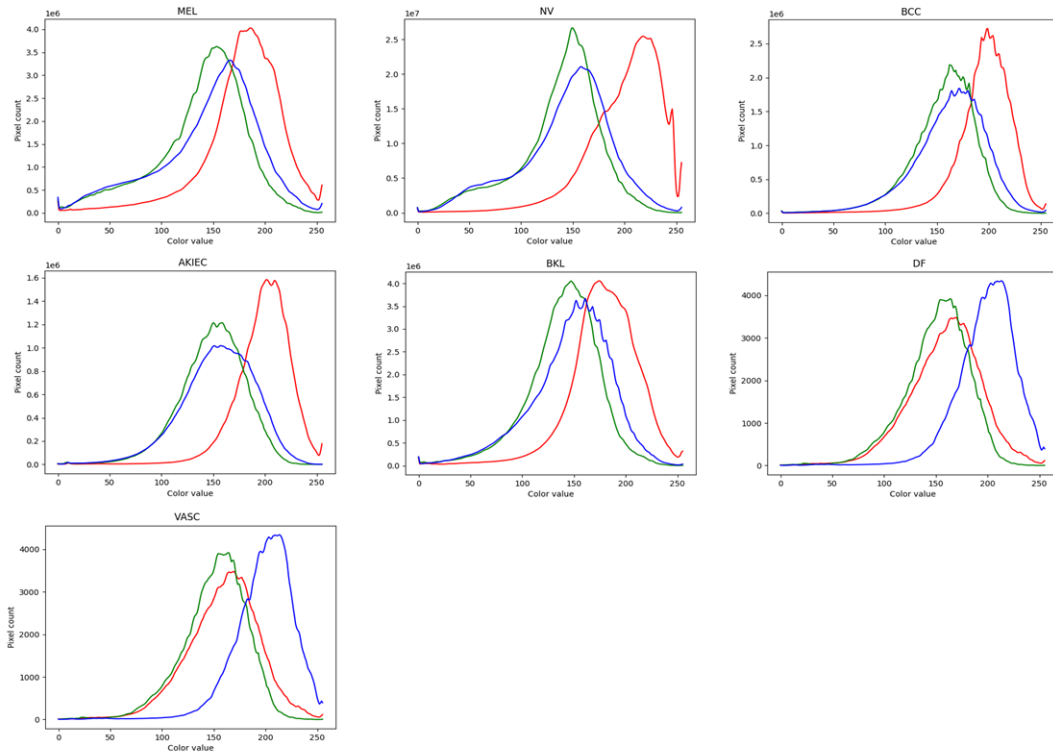
	MEL	NV	BCC	AKIEC	BKL	DF	VASC
Train	1113	<u>6705</u>	514	327	1099	115	142
Test	171	<u>909</u>	93	43	217	44	35
Val	21	<u>123</u>	15	8	22	1	3

# Данные

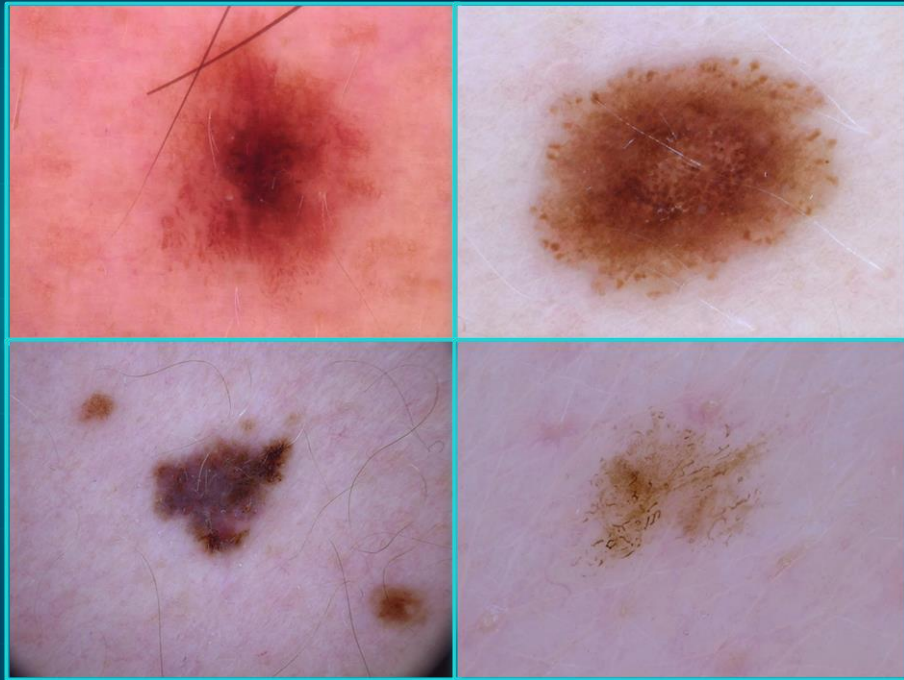


# Данные

Распределение значений пикселей в изображениях



# Данные



Изображения  
центрированы, но на  
многих присутствуют  
**артефакты** в виде волос.

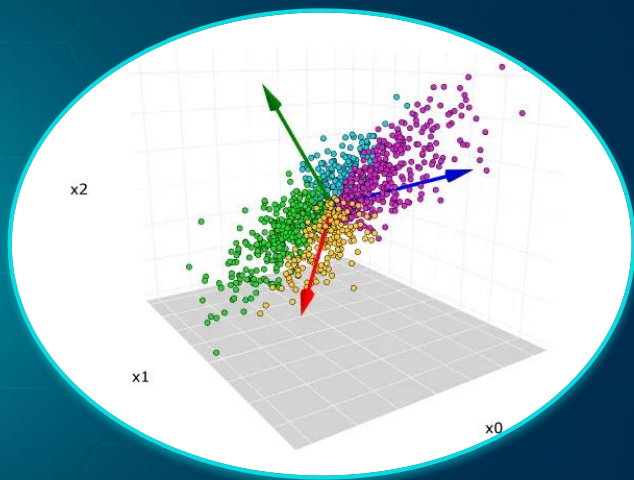
# Обработка изображений

450x600

Размер стандартного изображения в датасете

От 270к до 810к

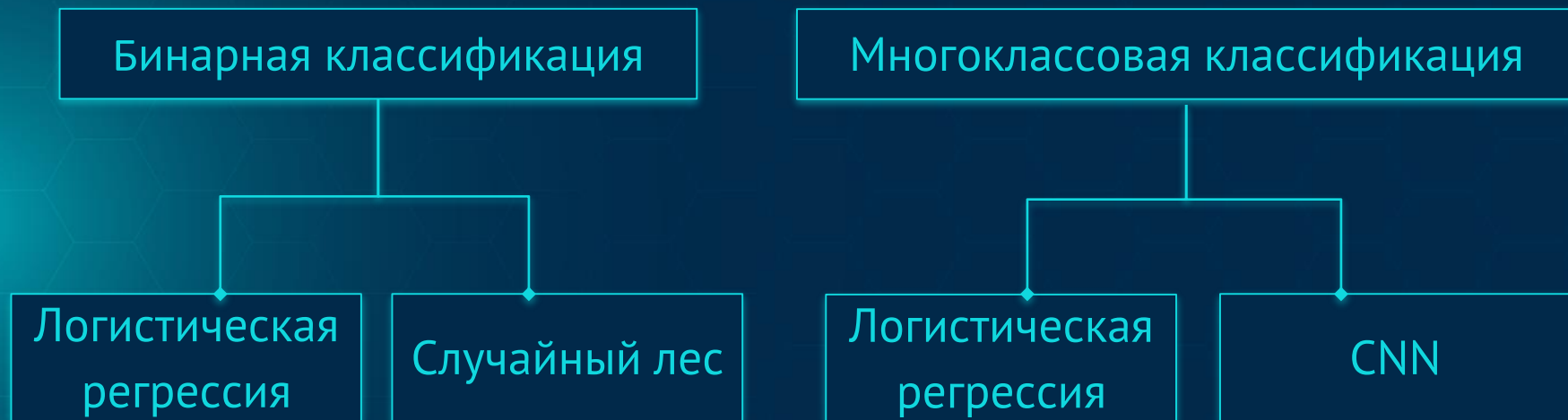
Исходное количество признаков



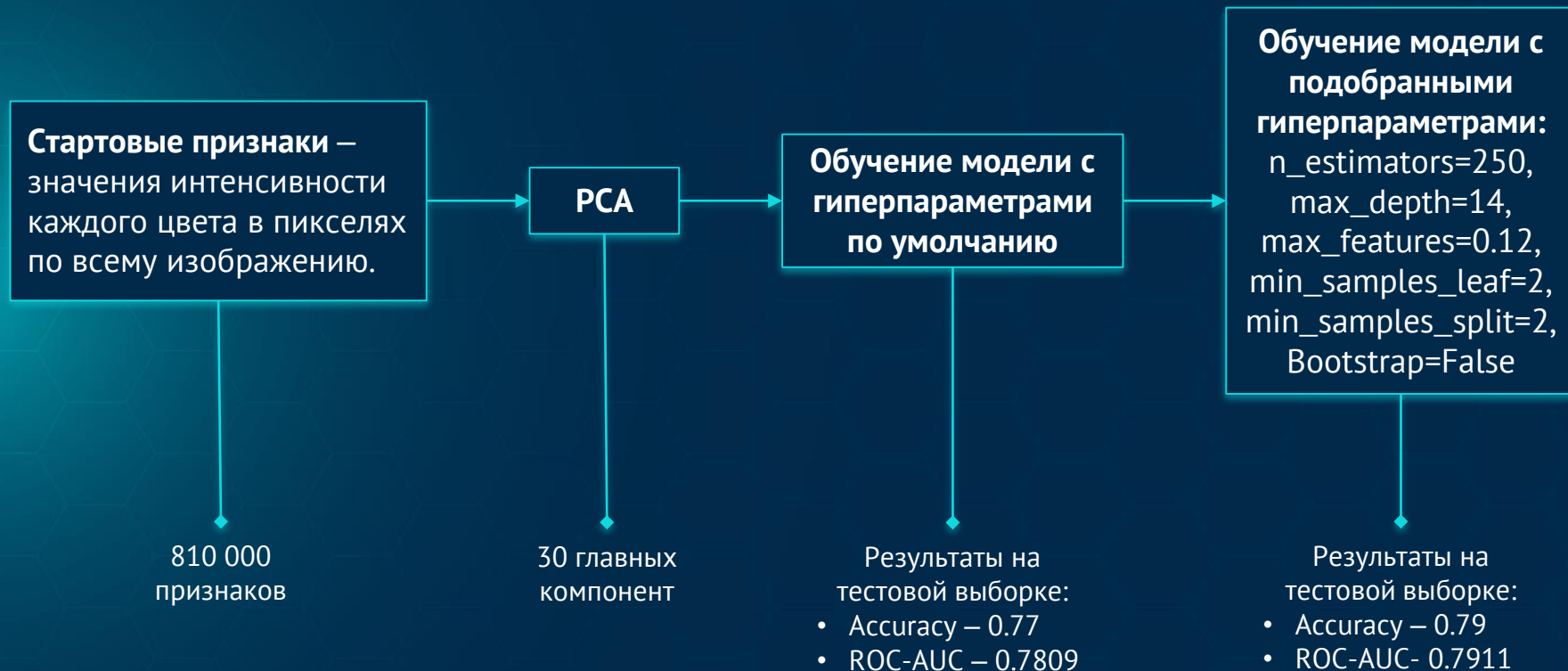
Для уменьшения размерности был применен **PCA**



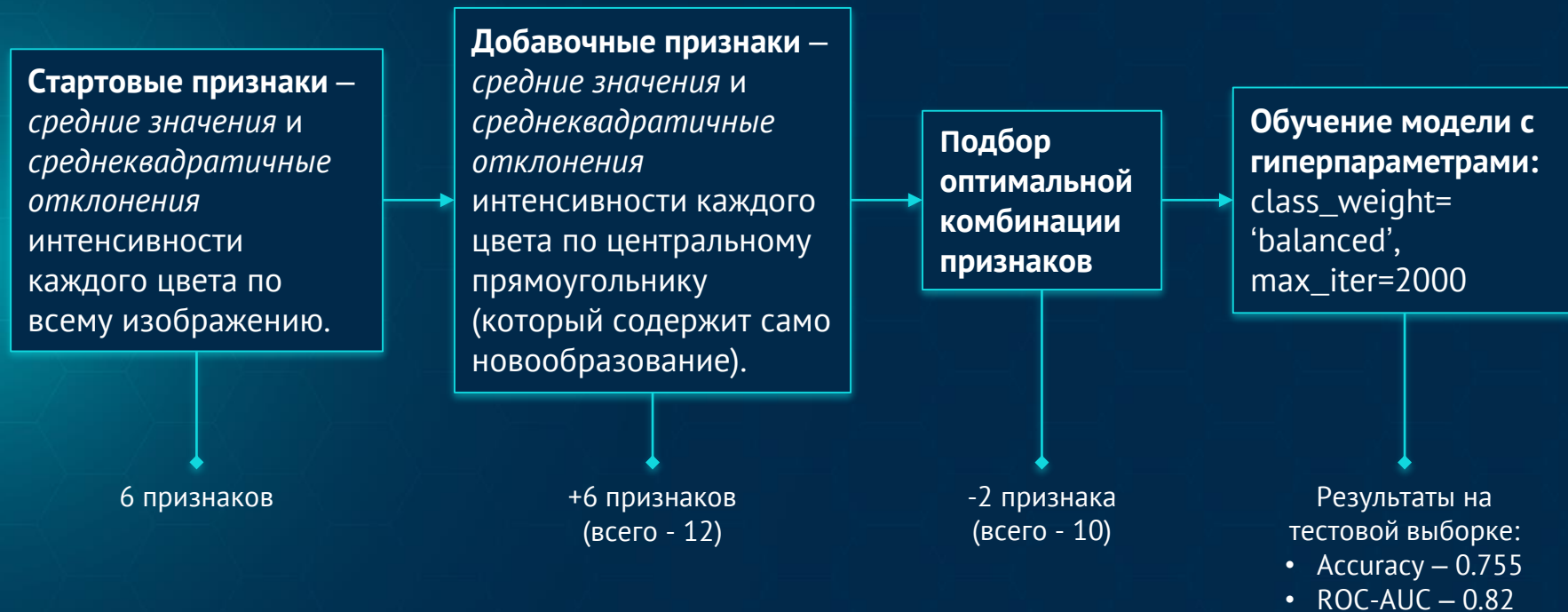
# Рассмотренные подходы



# Бинарная классификация - случайный лес

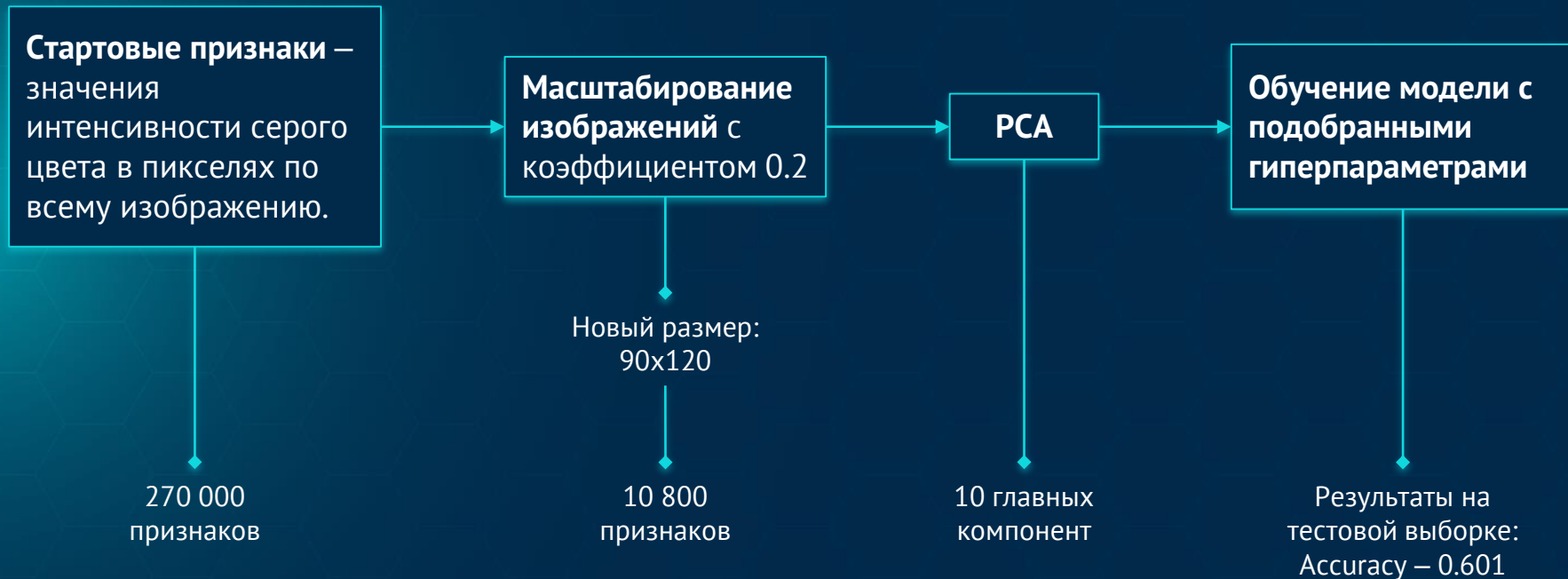


# Бинарная классификация - логистическая регрессия



## Многоклассовая классификация — логистическая регрессия

**Модель** — логистическая регрессия с использованием **SGD** в качестве алгоритма оптимизации.



# Convolutional neural network

**Снижение размерности** изображений до 0.3 от исходной (большой процент приводил к чрезмерно долгой обработке изображений)

Новый размер:  
135x180

Подбор гиперпараметров не проводился

**Обучение модели**

5 эпох (большее количество приводило к чрезмерно большому времени обучения)

Результаты на тестовой выборке:

- Accuracy – 0.618
- Precision – 0.704
- Recall – 0.556
- ROC-AUC – 0.884

Данная модель продемонстрировала наилучшие показатели качества среди рассмотренных и была выбрана в качестве **лучшей модели**.

# Convolutional neural network

3

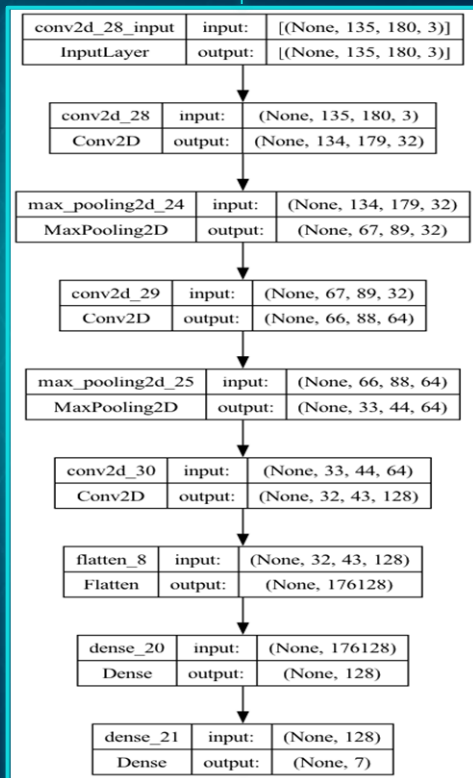
Количество сверток

2

Количество полносвязных слоев

22,600,000

Общее количество параметров



# Telegram бот

Основа — фреймворк **aiogram**.

**/help**

Выводит список доступных команд

**/predict**

Выдает наиболее вероятный тип новообразования на изображении, используя модель логистической регрессии

**/predictmel**

Выдает вероятность того, что на загруженном изображении меланомы

**/predictcnn**

Выдает наиболее вероятный тип новообразования на изображении, используя CNN

**/rate**

Предлагает оценить бота по шкале от 1 до 5

**/showrating**

Выводит накопившуюся усредненную оценку

# Развертывание

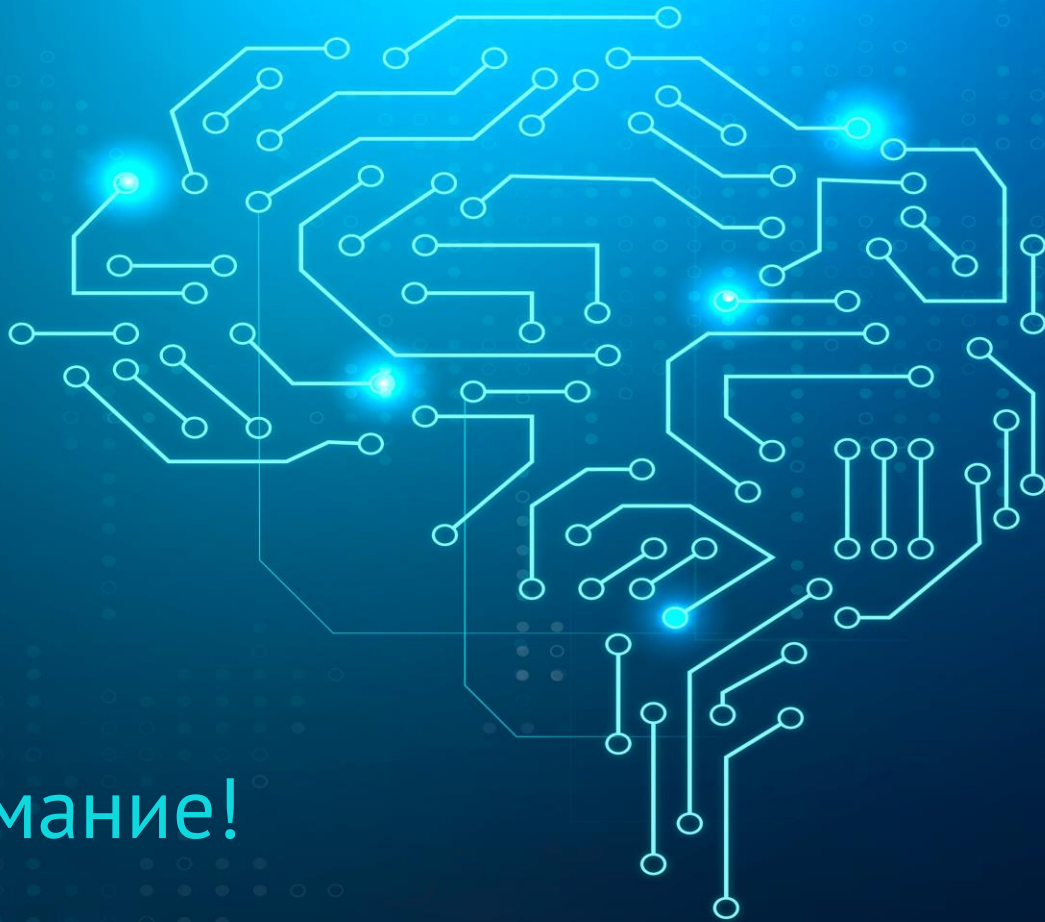


В качестве хостинга для бота  
используется облачная платформа  
**DigitalOcean.**



# Дальнейшие шаги





Спасибо за внимание!