

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
«Южно-Уральский государственный университет  
(национальный исследовательский университет)»  
Высшая школа электроники и компьютерных наук  
Кафедра системного программирования

**РАЗРАБОТКА МОДЕЛИ ПРОМЕЖУТОЧНОГО СЛИЯНИЯ МОДАЛЬНОСТИ  
РЕНТГЕНОЛОГИЧЕСКИХ СНИМКОВ С МЭПИНГОМ ЭМБЕДДИНГОВ  
ИЗ МОДАЛЬНОСТИ КЛИНИЧЕСКИХ ДАННЫХ  
ДЛЯ КЛАССИФИКАЦИИ ЗАБОЛЕВАНИЙ ЛЕГКИХ**

Рецензент,  
доцент кафедры ИИТиМОИ  
ФГБОУ ВО «ЮУрГГПУ»,  
к.п.н., доцент О.А. Дмитриева

Научный руководитель,  
доцент кафедры СП, к.п.н. О.Н. Иванова  
Автор работы,  
студент группы КЭ-229 Р.Г. Гильманова

Челябинск, 2024 г.

# Актуальность



- 1) улучшение качества диагностики;
- 2) персонализация лечения;
- 3) нагруженность медицинского персонала;
- 4) человеческий фактор.

# Цель и задачи

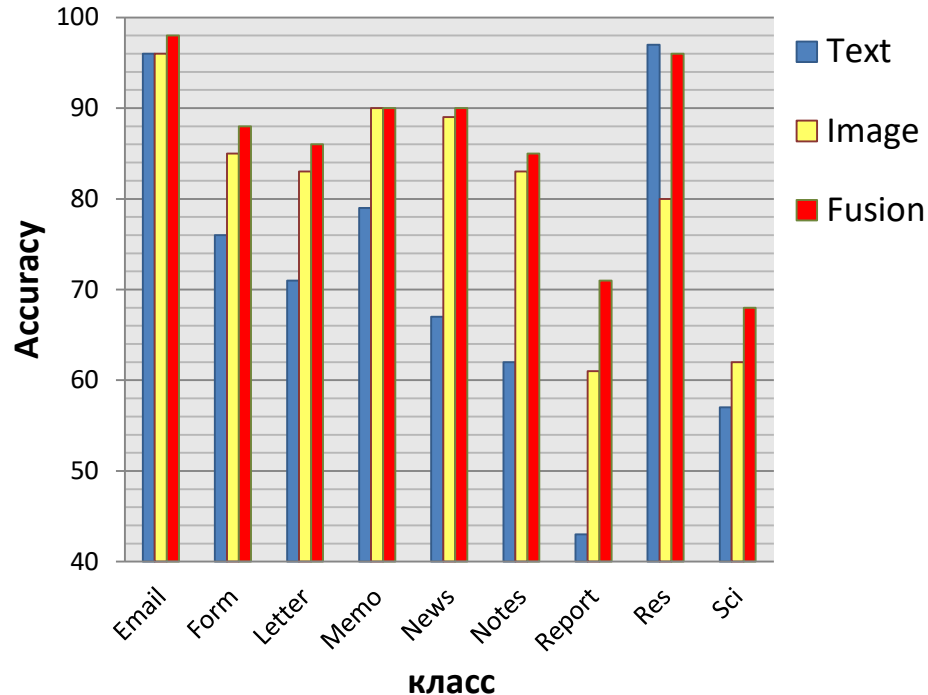
## Цель:

разработка модели промежуточного слияния модальности рентгенологических снимков с мэппингом эмбеддингов из модальности клинических данных для классификации заболеваний легких.

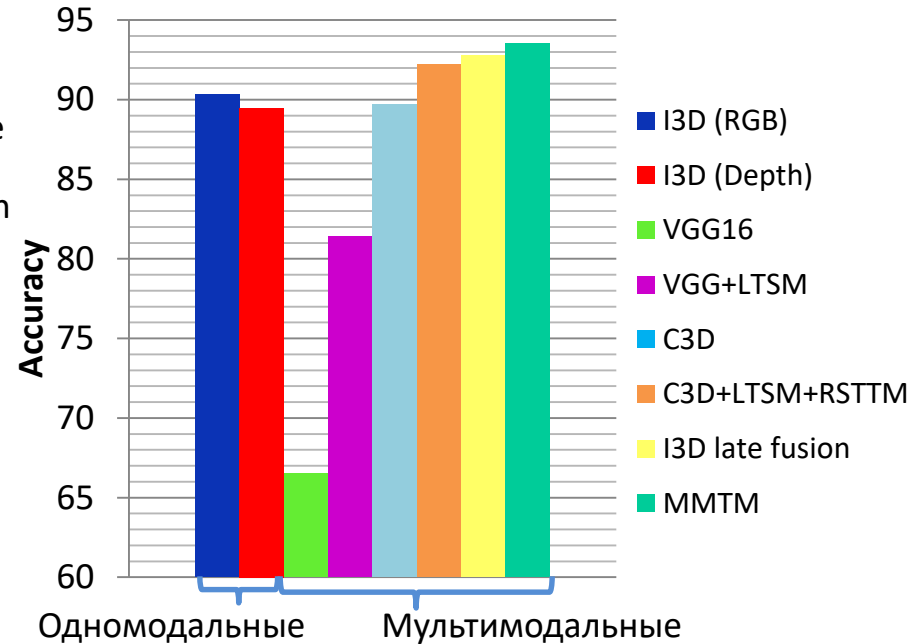
## Задачи:

- 1) сегментация снимков легких из набора данных для обучения;
- 2) обучение модели промежуточного слияния модальности рентгенологических снимков на наборе данных из сегментированных и не сегментированных рентгенологических снимков;
- 3) сравнение точности модели при обучении на сегментированных и не сегментированных рентгенологических снимках;
- 4) мэппинг эмбеддингов из модальности клинических данных в модальность рентгенологических снимков;
- 5) анализ изменения точности модели после мэппинга эмбеддингов модальности клинических данных.

# Обзор близких работ



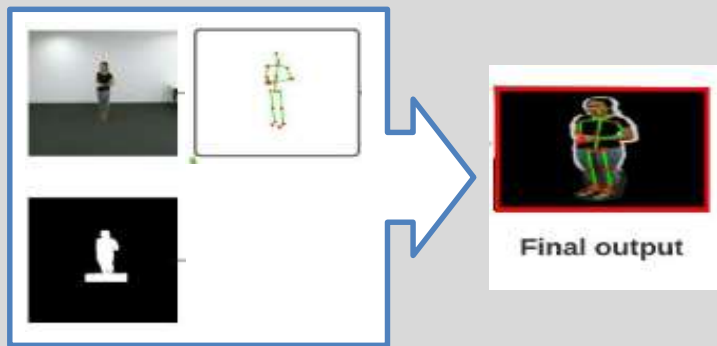
Audebert N., Herold C., Slimani K., Vidal C. Multimodal deep networks for text and image-based document classification // arXiv.org, 2019. Дата обновления: 15.07.2019 г. URL: <https://arxiv.org/abs/1907.06370> (дата обращения: 12.05.2024 г.).



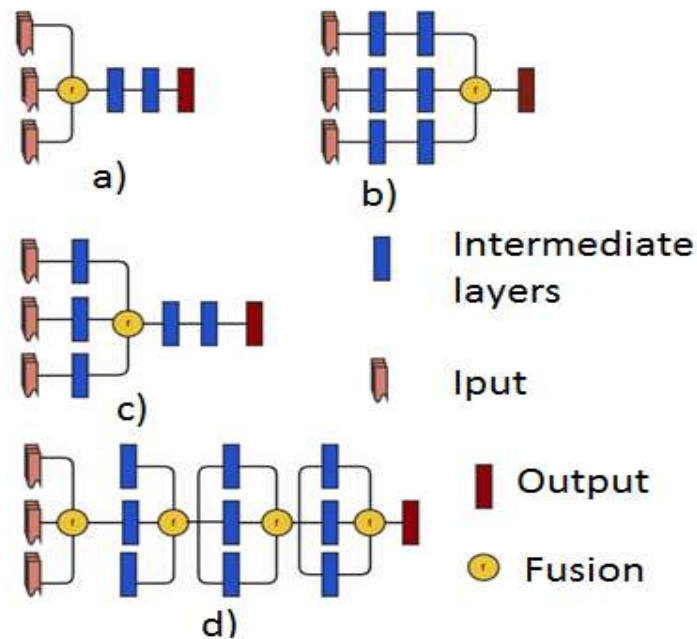
Joze H.R.V., Shaban A., Iuzzolino M.L., Koishida K. MMTM: Multimodal Transfer Module for CNN Fusion // arXiv.org, 2019. Дата обновления: 30.03.2020 г. URL: <https://arxiv.org/abs/1911.08670> (дата обращения: 12.05.2024 г.).

# Мультимодальные нейронные сети

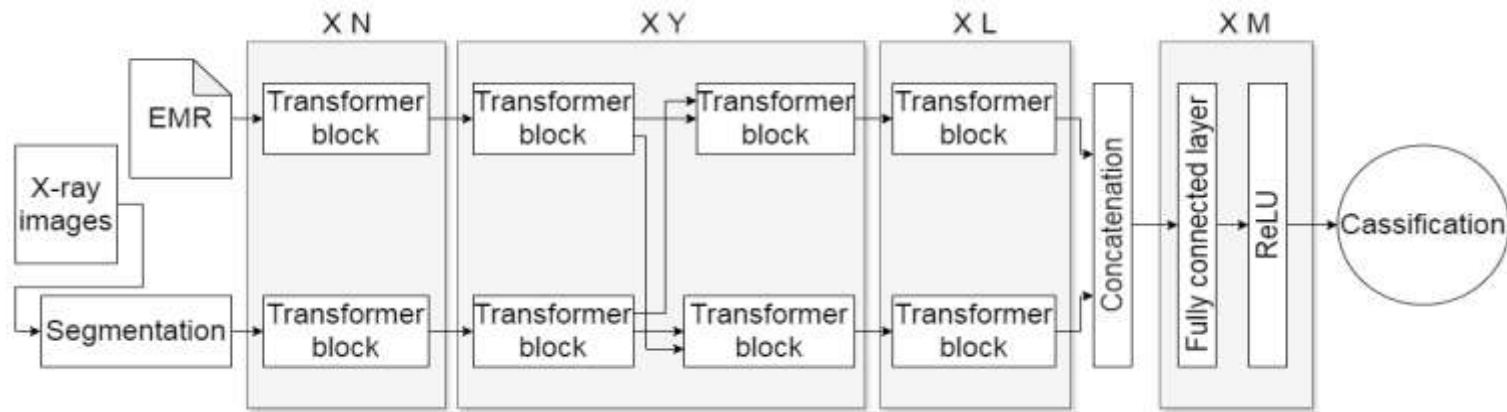
Мультимодальное глубокое обучение – это область искусственного интеллекта, которая включает интеграцию информации из нескольких модальностей



## Схемы слияний модальностей



# Проектирование



**Язык программирования:** Python 3.10.12

**Фреймворк для сегментации:** keras 2.15.0

**Среда разработки:** Google Colaboratory

**Фреймворк для классификации:** PyTorch 2.3.0

# Набор данных для классификации

Примеры снимков из набора обучающих данных  
Набор данных Chest X-rays



**Набор данных:**

chest X-ray collection from Indiana University

**Источник данных:** Open Access Biomedical Image Search Engine

**Количество пар снимок-текст**

Изначальное – 7470

Фронтальные снимки – 3811

Снимки классов normal и lung – 1937

После нормализации – 1116

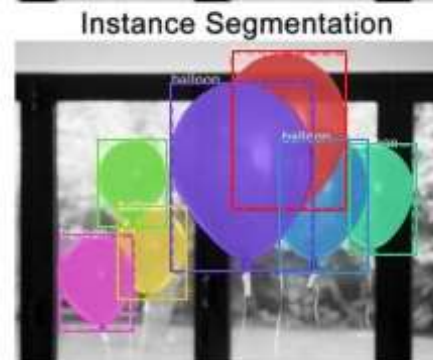
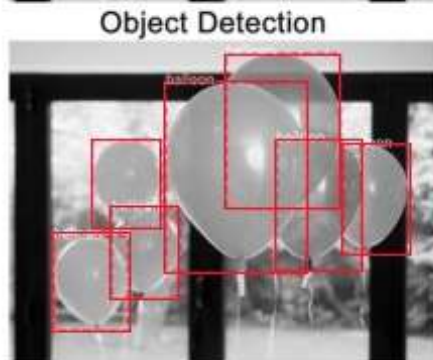
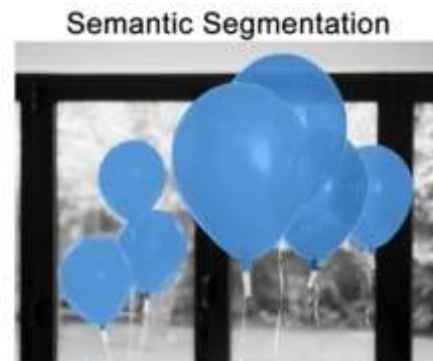
**Классы оригинального набора данных:**

normal, no indexing, lung, calcified granuloma, thoracic vertebrae

# Предобработка рентгенологических снимков

## Задачи компьютерного зрения

|                              |  |
|------------------------------|--|
| <b>Semantic segmentation</b> | определение всех пикселей объектов определённого класса или фона на изображении    |
| <b>Object detection</b>      | обнаружение всех объектов указанных классов и определение рамки для каждого из них |
| <b>Instance segmentation</b> | определение пикселей, принадлежащих каждому объекту каждого класса по отдельности  |



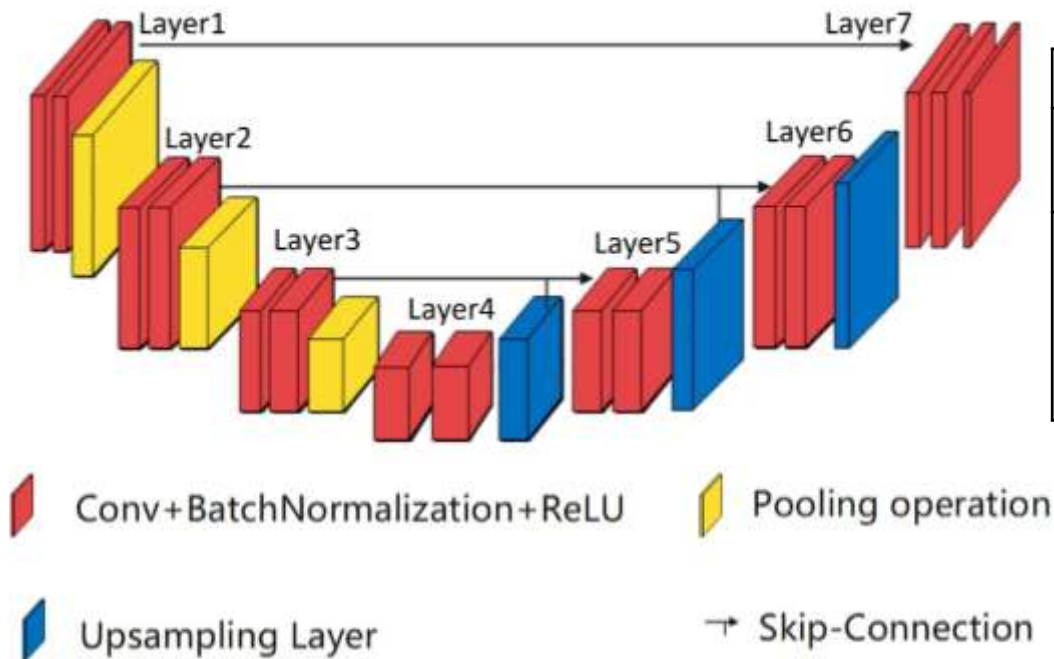
Mask R-CNN: архитектура современной нейронной сети для сегментации объектов на изображениях. [Электронный ресурс]  
URL: <https://habr.com/ru/post/421299/>



# Предобработка рентгенологических снимков

## Архитектура и набор данных

Схема архитектуры U-Net



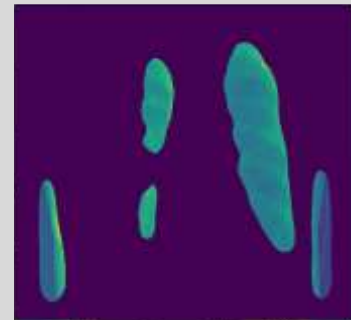
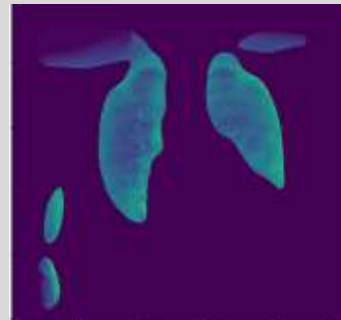
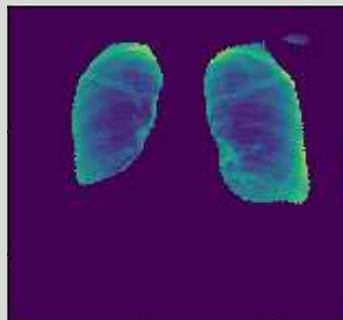
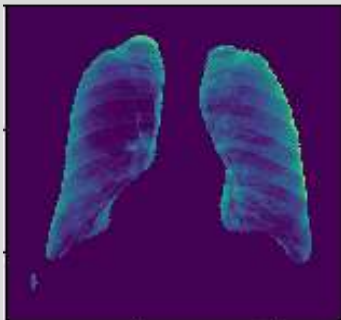
Информация о наборах данных

|           | Название   | СНИМКОВ |
|-----------|--|---------|
| Модель №1 | Chest Xray Masks and Labels                      | 1600    |
| Модель №2 | Lung Masks for Shenzhen Hospital Chest X-ray Set | 566     |
|           | Pulmonary Chest X-Ray Abnormalities              | 566     |

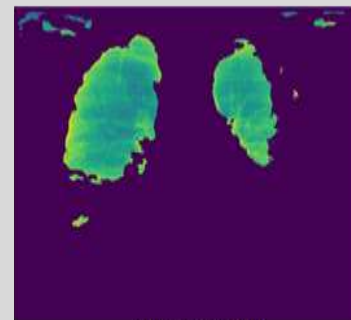
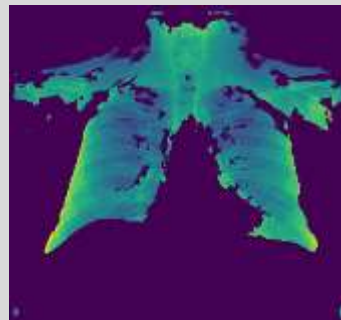
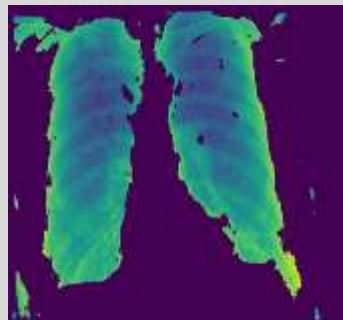
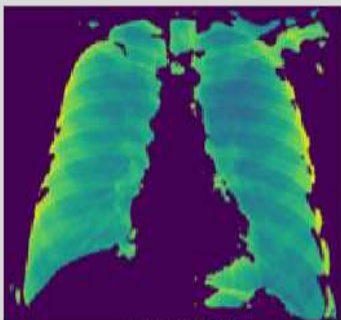
Cai S., Wu Y., Chen G. A Novel Elastomeric UNet for Medical Image Segmentation // frontiersin.org, 2022. URL: <https://www.frontiersin.org/articles/10.3389/fnagi.2022.841297/full>

# Примеры сегментированных снимков

## Первый вариант модели



## Второй вариант модели



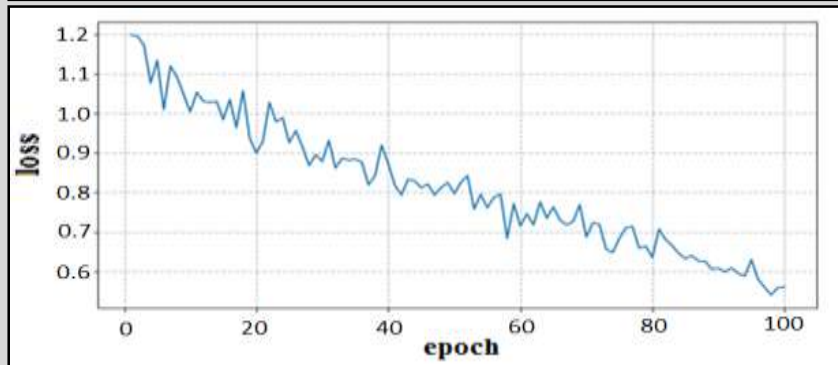
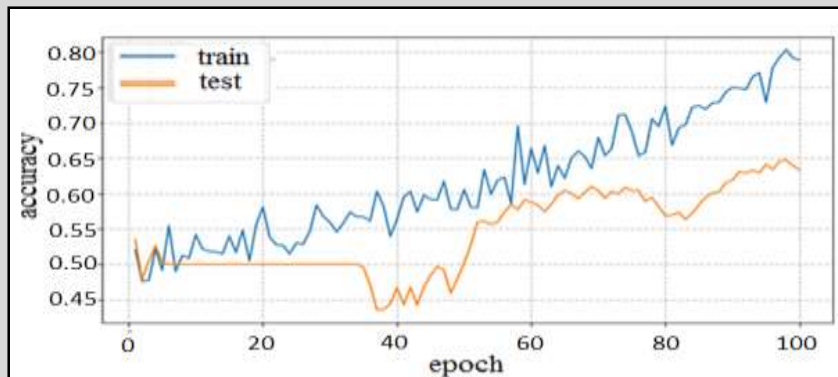
# Анализ влияния сегментации снимков на классификацию

## Accuracy

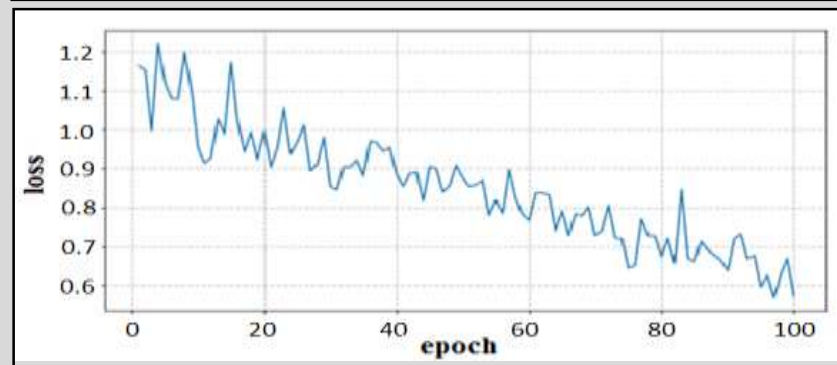
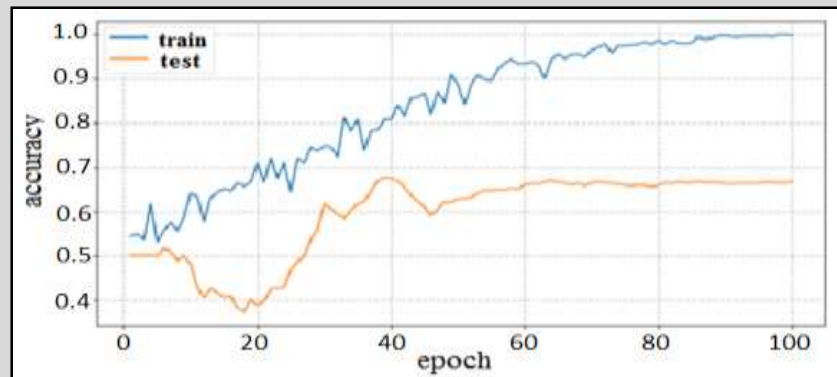
| Архитектура        | Обучающие данные | Эпоха 1 | Эпоха 2 | Эпоха 3 | Эпоха 4 | Эпоха 5 | Изменение accuracy [%] |
|--------------------|------------------|---------|---------|---------|---------|---------|------------------------|
| <b>DenseNet121</b> | обработаны       | 0,6569  | 0,5882  | 0,4118  | 0,7059  | 0,4902  | 4,4                    |
|                    | не обработаны    | 0,5000  | 0,4706  | 0,5441  | 0,3235  | 0,6618  |                        |
| <b>DenseNet169</b> | обработаны       | 0,5000  | 0,5000  | 0,5000  | 0,5000  | 0,5000  | 0                      |
|                    | не обработаны    | 0,2059  | 0,4559  | 0,4559  | 0,4559  | 0,5000  |                        |
| <b>DenseNet201</b> | обработаны       | 0,4020  | 0,6471  | 0,5000  | 0,5000  | 0,5000  | 13,2                   |
|                    | не обработаны    | 0,5000  | 0,5000  | 0,5000  | 0,5147  | 0,5000  |                        |
| <b>VGG19</b>       | обработаны       | 0,5000  | 0,5000  | 0,6765  | 0,5000  | 0,5000  | 17,6                   |
|                    | не обработаны    | 0,2647  | 0,5000  | 0,5000  | 0,5000  | 0,5000  |                        |
| <b>NASNetLarge</b> | обработаны       | 0,6961  | 0,8431  | 0,5686  | 0,4706  | 0,7941  | 18,1                   |
|                    | не обработаны    | 0,2059  | 0,6618  | 0,5000  | 0,5000  | 0,5000  |                        |

# Анализ влияния сегментации снимков на классификацию DenseNet121

## Обучение на несегментированных снимках



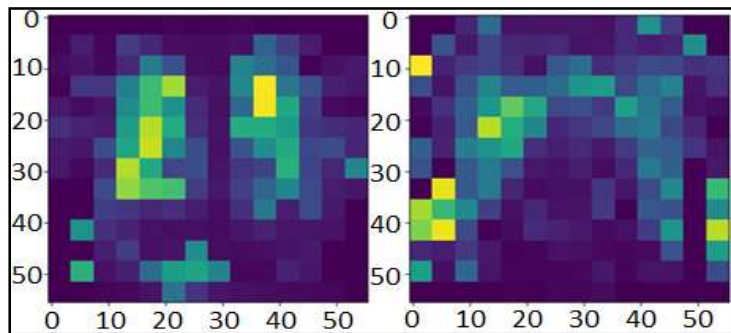
## Обучение на сегментированных снимках



# Реализация классификации

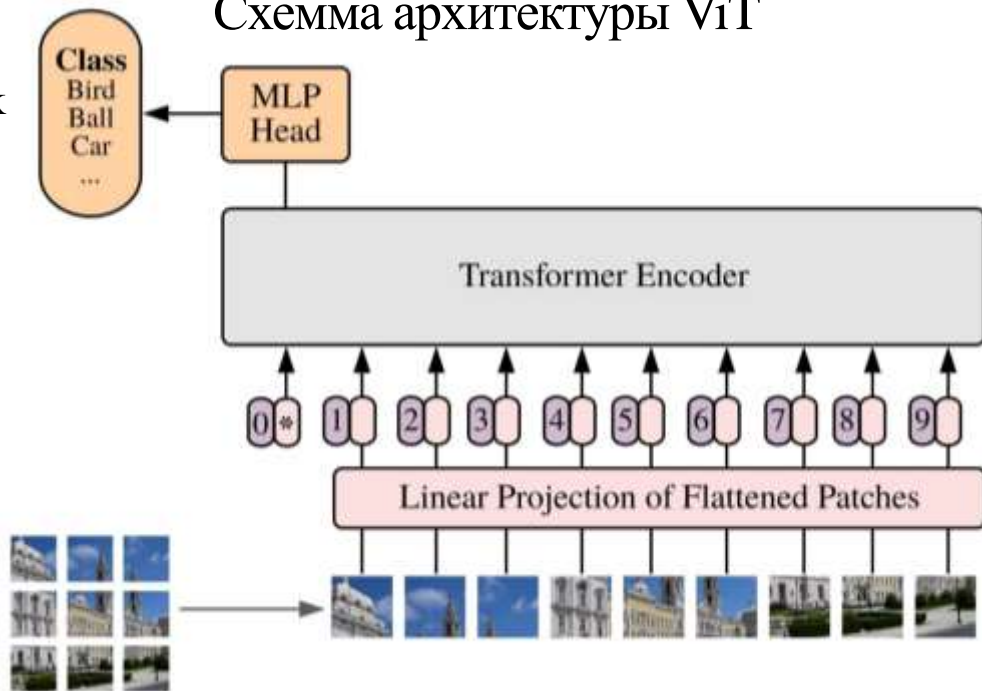
## Классификация модальности рентгенологических снимков

модель vit-base-patch16-224-in2  
предобученна на наборе данных ImageNet-21k  
принимает изображения 224 на 224 пикселя  
размерность эмбедингов составляет 768



Визуализация матриц attention

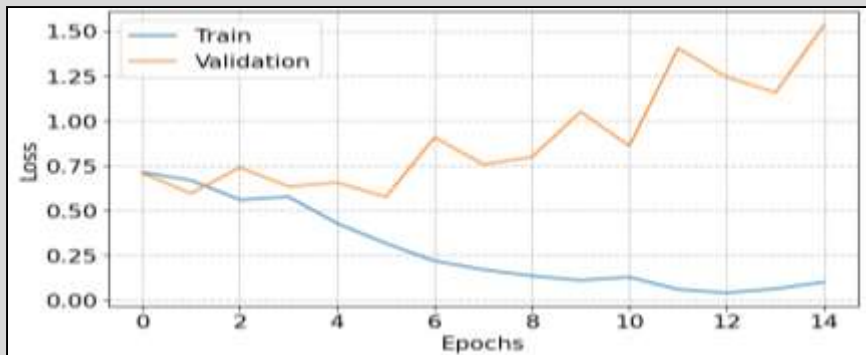
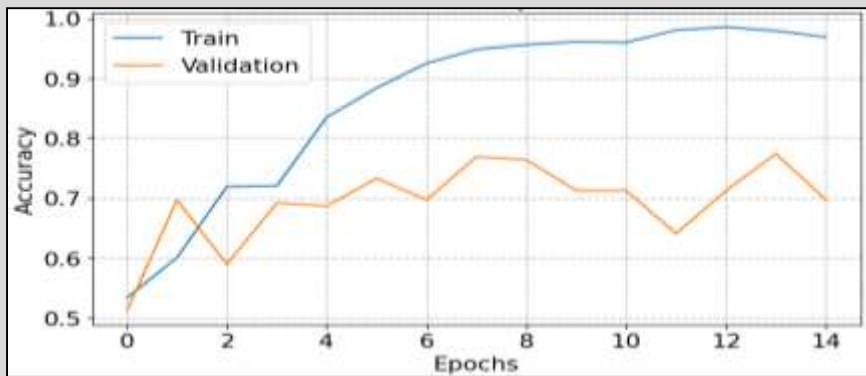
### Схема архитектуры ViT



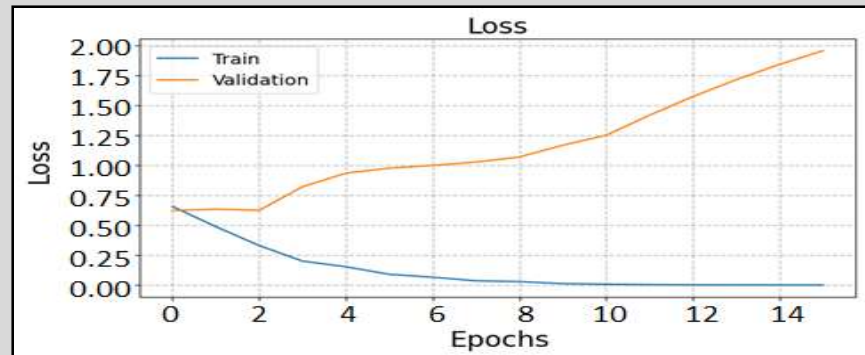
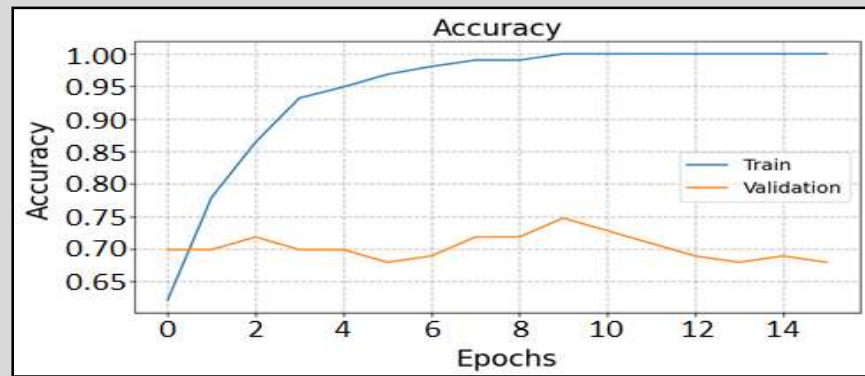
# Анализ влияния сегментации снимков на классификацию

## классификация модальности рентгенологических снимков

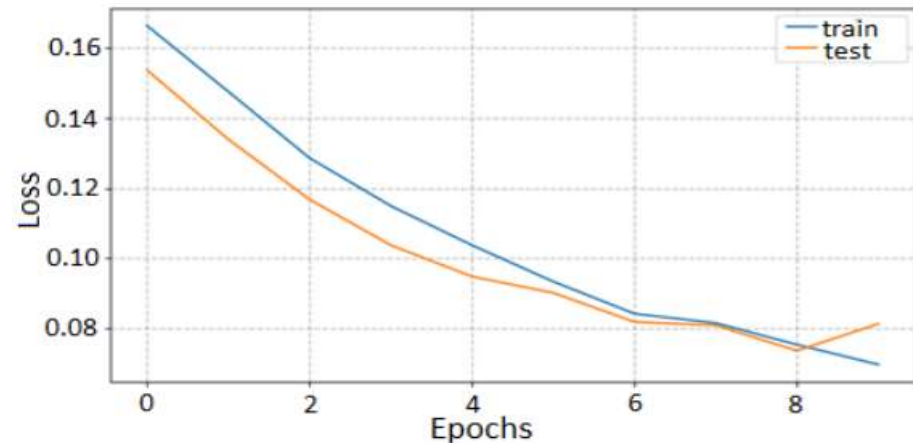
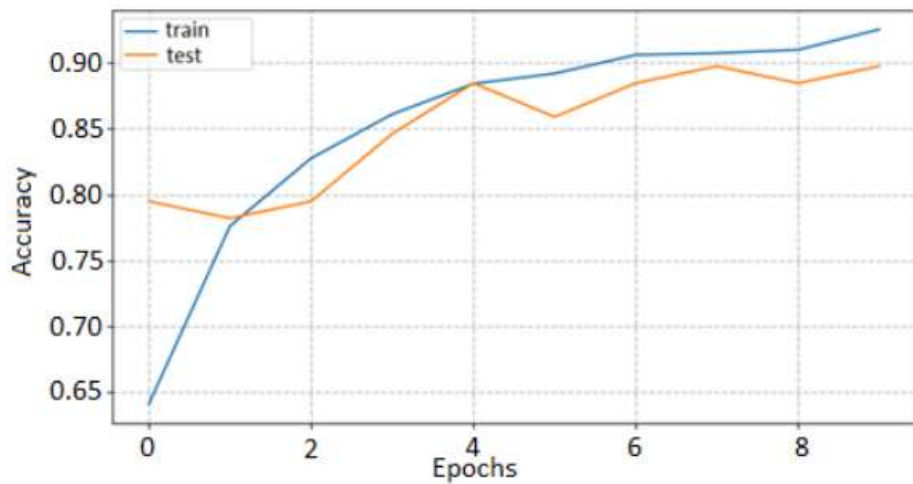
### Обучение на несегментированных снимках



### Обучение на сегментированных снимках



# Классификация модальности клинических данных

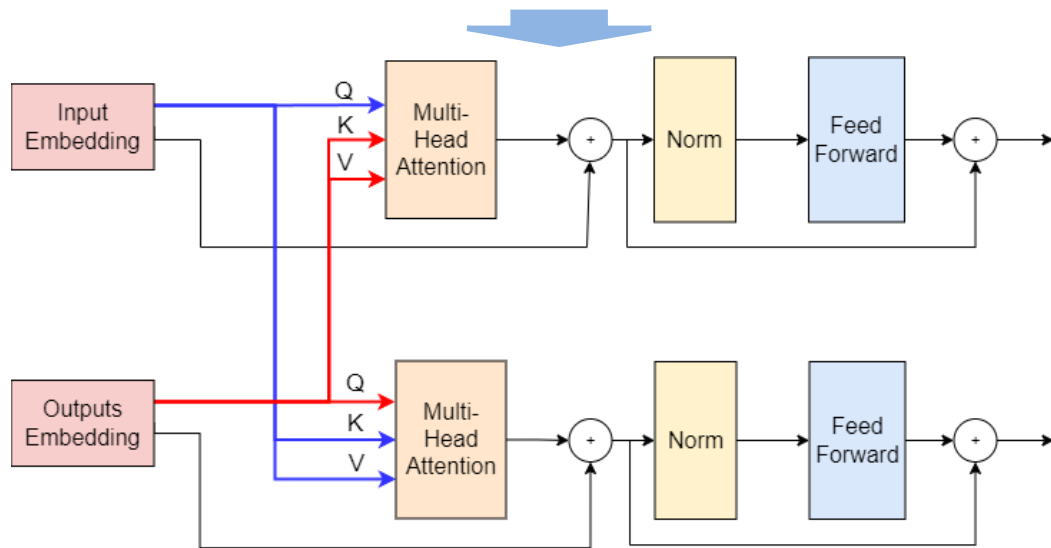
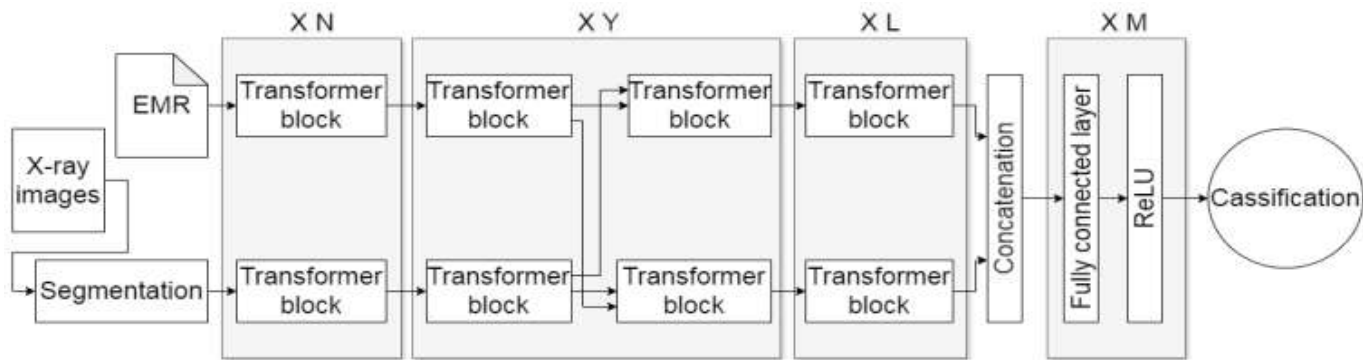


**Модель:** bert-base-uncased

Пример классифицируемых данных

The lungs are clear bilaterally. Specifically, no evidence of focal consolidation, pneumothorax, or pleural effusion.. Cardio mediastinal silhouette is unremarkable. Visualized osseous structures of the thorax are without acute abnormality.

# Реализация слияния модальностей

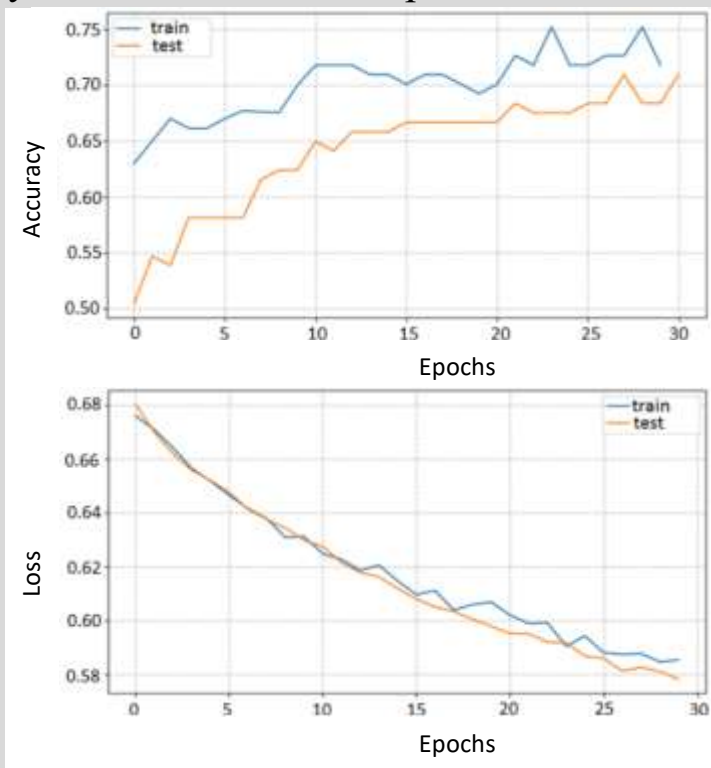




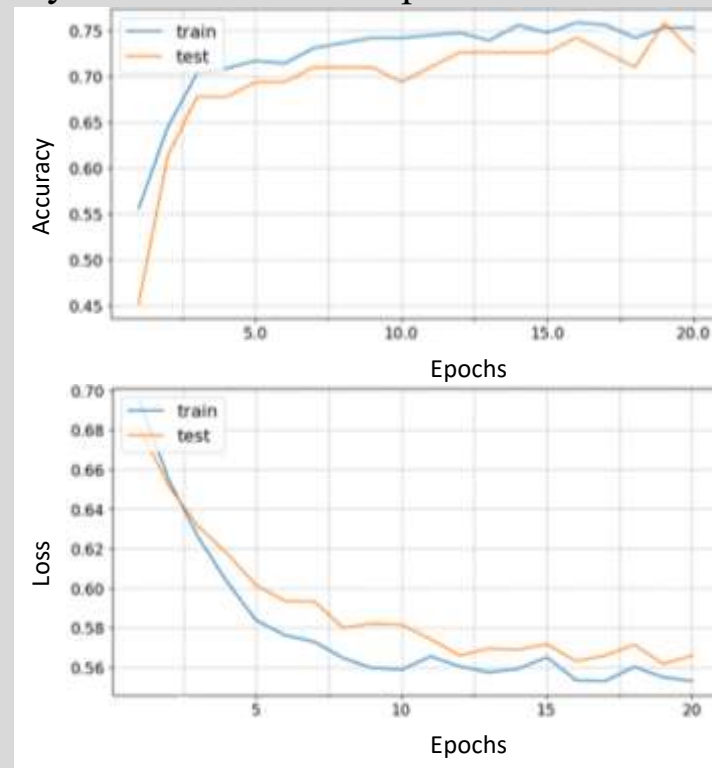
# Анализ влияния сегментации снимков на классификацию

## мульти模альная классификация

### Обучение на несегментированных снимках

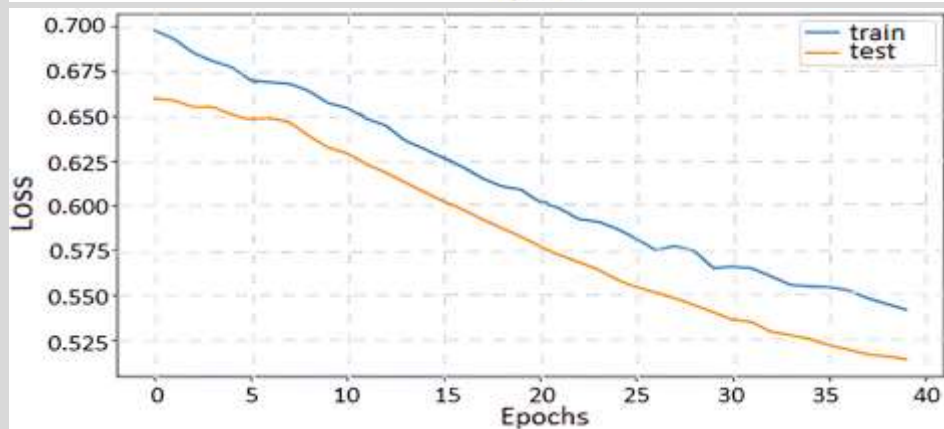


### Обучение на сегментированных снимках



# Обучение на большом количестве эпох

графики обучения мультимодальной модели с использованием сегментированных снимков



сравнение моделей

| Accuracy [%] | Время 1 эпохи [с.] | Модальность                                    |
|--------------|--------------------|--|
| 77           | 3393               | рентгенологические снимки                      |
| 88           | 420                | клинические данные                             |
| 79           | 1563               | рентгенологические снимки + клинические данные |

ССЫЛКА НА  
ИСХОДНЫЙ КОД



# Основные результаты

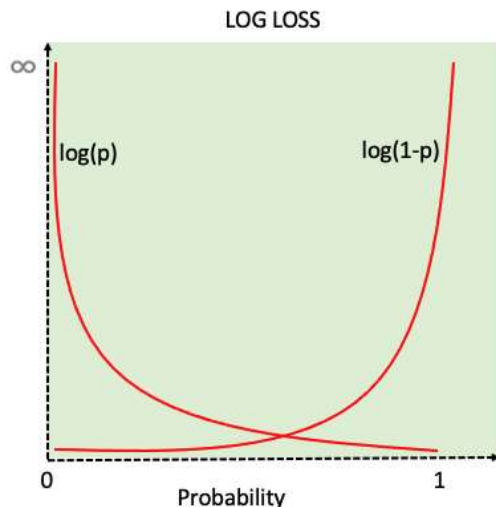
1. Разработана модель промежуточного слияния модальности рентгенологических снимков с мэппингом эмбедингов из модальности клинических данных для классификации заболеваний легких.
2. Проведена сегментация снимков легких из набора данных для обучения.
3. Проведено обучение модели промежуточного слияния модальности рентгенологических снимков на наборе данных из сегментированных и не сегментированных рентгенологических снимков.
4. Проведено сравнение точности модели при обучении на сегментированных и не сегментированных рентгенологических снимках.
5. Произведен мэппинг эмбедингов из модальности клинических данных.
6. Проанализированы изменения точности модели после мэппинга эмбедингов модальности клинических данных.

# Использованные метрики

## 1. Sørensen–Dice coefficient

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|},$$

где  $|A|$  – количество элементов в наборе  $A$ ,  
 $|B|$  – количество элементов в множестве  $B$ ,  
 $|A \cap B|$  – количество элементов, которые  
присутствуют в обоих наборах.



## 2. Accuracy

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

где  $TP$  – количество истинно-положительных результатов,  
 $TN$  – количество истинно-отрицательных результатов,  
 $FP$  – количество ложно-положительных результатов,  
 $FN$  – количество ложно-отрицательных результатов.

## 3. Cross Entropy Loss

$$-\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C (y_{i,j} \cdot \log(p_{i,j}))$$

$C$  – количество классов.

$y_{i,j}$  – истинные метки для класса  $j$  для  $i$

$p_{i,j}$  – прогнозируемая вероятность для класса  $j$  для  $i$

# Предобработка рентгенологических снимков

## Графики обучения

