**Министерство науки и высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**«Национальный исследовательский университет ИТМО»**

**(Университет ИТМО)**

**Факультет программной инженерии и компьютерной техники (ФПИ и КТ)**

**Образовательная программа Нейротехнологии и программная инженерия**

**Направление подготовки (специальность) 09.04.04 Программная инженерия**

О Т Ч Е Т

о научно-исследовательской работе

Тема задания: **\_\_**Модели на основе нейронных сетей для мультиклассовой классификации

опухолей головного мозга по снимкам магнитно-резонансной томографии**\_\_**

Обучающийся *Ян Цзяфэн, № группы P4123*

Руководитель практики от профильной организации:*Русак Алена Викторовна, преподаватель факультета программной инженерии и компьютерной техники*

Руководитель практики от университета:*Русак Алена Викторовна, преподаватель факультета программной инженерии и компьютерной техники*

Практика пройдена с оценкой

Подписи членов комиссии:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(подпись)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(подпись)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(подпись)

Дата 25.01.2023

Санкт-Петербург

2023

**Оглавление**

[Введение 3](#_Toc125643797)

[Анализ предметной области 4](#_Toc125643798)

[Описание проблемы 4](#_Toc125643799)

[Обзор литературы 4](#_Toc125643800)

[Выбор методов и технологий реализации 6](#_Toc125643801)

[Данные 6](#_Toc125643802)

[Свёрточная нейронная сеть (CNN) 7](#_Toc125643803)

[Слой свёртки 7](#_Toc125643804)

[Слой активации 8](#_Toc125643805)

[Слой пулинга 9](#_Toc125643806)

[Полносвязный слой 9](#_Toc125643807)

[Visual Geometry Group (VGG) 9](#_Toc125643808)

[ResNet 10](#_Toc125643809)

[Vision Transformer 11](#_Toc125643810)

[Self-attention 12](#_Toc125643811)

[Слой Embedding 12](#_Toc125643812)

[Transformer Encoder 13](#_Toc125643813)

[MLP Head 13](#_Toc125643814)

[Сравнение архитектур VGG, ResNet, ViT 13](#_Toc125643815)

[Проектирование и разработка модели 14](#_Toc125643816)

[Предобработка данных 14](#_Toc125643817)

[Разработка модели 14](#_Toc125643818)

[Функция потерь 14](#_Toc125643819)

[Исследование и тестирование модели 16](#_Toc125643820)

[Выводы 17](#_Toc125643821)

[Список литературы 18](#_Toc125643822)

Введение

**Актуальность темы.** Опухоли головного мозга (ОГМ) являются одним из самых смертоносных видов рака, которые серьезно угрожают здоровью человека. Существует более 130 видов ОГМ, наиболее распространенными из которых являются менингиомы, глиомы и опухоли гипофиза. Ручная диагностика по анализу МРТ-изображений требует достаточно высокой квалификации и внимания специалиста, при этом не исключены ошибки. Модели мультиклассовой классификации могут помочь врачам идентифицировать и классифицировать очаги поражения, а также повысить эффективность и точность диагностики. Новая модель нейронных сетей Vision Transformers (ViT) демонстрирует высокие результаты, в задачах классификации изображений. Поэтому данная работа будет направлена на исследование возможности применения моделей нейронных сетей на основе архитектуры Трансформера для анализа и обработки медицинских изображений на примере задачи мультиклассовой классификации опухолей головного мозга по снимкам магнитно-резонансной томографии.

**Целью** данной работы является исследование возможности применения нейросетевых моделей трансформеров в задачах мультиклассовой классификации опухолей головного мозга по снимкам магнитно-резонансной томографии для повышения точности диагностики.

**Задачи**, в соответствии с целью работы, приведены ниже:

1. Обзор и анализ методов мультиклассовой классификации опухолей головного мозга по снимкам МРТ.
2. Разработка подхода к мультиклассовой классификации опухолей головного мозга по снимкам МРТ с использованием архитектуры трансформер.
3. Подбор и обработка данных для моделирования.
4. Реализация модели трансформера для мультиклассовой классификации опухолей головного мозга по снимкам МРТ.
5. Проведение экспериментальных исследований и оценка эффективности разработанной модели на основе выбранных критериев.

В этой работе предлагается архитектура трансформер с линейным механизмом внутреннего внимания для построения модели, предназначенная для задачи мультиклассовой классификации ОГМ по снимкам МРТ, и используется F1-мера для оценки модели. Окончательное значение F1-меры составило 91,35%, поэтому мы можем доказать, что при использовании такого типа модели мы можем эффективно классифицировать ОГМ по снимкам МРТ.

Анализ предметной области

## Описание проблемы

Опухоли головного мозга (ОГМ) являются одним из самых смертоносных видов рака, которые серьезно угрожают здоровью человека. Существует более 130 видов ОГМ, наиболее распространенными из которых являются менингиомы, глиомы и опухоли гипофиза. В настоящее время в клинической диагностике, исследовании и лечении ОГМ наиболее широко используется магнитно-резонансная томография (МРТ). Ручная диагностика по анализу МРТ-изображений требует достаточно высокой квалификации и внимания специалиста, при этом не исключены ошибки. Поэтому при решении задач обработки и анализа медицинских изображений все чаще применяются компьютерные системы диагностики. Учитывая разные планы лечения различных видов ОГМ, алгоритмов бинарной классификации, которые могут только диагностировать наличие или отсутствие опухолей, недостаточно. Модели мультиклассовой классификации могут помочь врачам идентифицировать и классифицировать очаги поражения, а также повысить эффективность и точность диагностики.

В последние десять лет сверточные нейронные сети (CNN) играют очень важную роль в задачах классификации изображений и обладают отличными возможностями извлечения локальных особенностей. Однако, поскольку рецептивное поле операции свертки ограничено размером ядра свертки, сложно получить глобальные особенности изображения с помощью CNN. Появление архитерктуры Transformer с механизмом внимания позволяет эффективно решить данную проблему. Новая модель нейронных сетей Visual Transformers (ViT) демонстрирует высокие результаты, в задачах классификации изображений. Поэтому данная научная работа будет направлена на исследование возможности применения моделей нейронных сетей на основе архитектуры трансформер для анализа и обработки медицинских изображений на примере задачи мультиклассовой классификации опухолей головного мозга по снимкам магнитно-резонансной томографии.

## Обзор литературы

Диссертация [1] посвящена гибридной системе классификации опухолей головного мозга по снимкам МРТ. Данная система состоит из набора статических признаков и глубоких нейронных сетей. Для выделения признаков используются три метода: 2-D DWT (Дискретное вейвлетпреобразование), 2-D фильтр Габора и статистические признаки, представленные первой и второй порядковой статистикой. Система классификатора разработана с использованием двух типов алгоритмов нейронных сетей: Stacked Sparse Autoencoder (SSA) и Softmax Classifier. Для обучения Sparse Autoencoder предлагается регуляризация разреженности и L2- регуляризация. Для оценки и валидации предложенной системы использовались набор данных из китайских больниц и набор данных BRATS, точности которых достигли 94% и 98.8% соответственно.

В главе 2 «Математический аппарат обработки и анализа диагностических изображений» [2] описываются техника применения МРТ, модели и методы обработки биомедицинских изображений, включая краткое описание анализа объектов на изображениях с использованием искусственных нейронных сетей, метода текстурного анализа и вейвлет-анализ.

Авторы [3] предложили модель ViT, которая применяет Transformer для классификации изображений, провели анализ и выявили, что ViT превосходит CNN при достаточном количестве данных для предварительного обучения.

Авторы [4] предлагают точную и полностью автоматическую систему классификации опухолей головного мозга, в которой применена концепция трансферного обучения и используется предварительно обученный GoogLeNet для извлечения признаков из снимков МРТ головного мозга. Данная система прошла перекрестную проверку на наборе данных МРТ из figshare и достигла средней точности классификации 98%.

Авторы [5] разработали пять клинически значимых мультиклассовых наборов данных и сравнили модель CNN, основанную на трансферном обучении, с шестью различными методами классификации машинного обучения (дерево решений, линейный дискриминант, наивный байесовский метод, метод опорных векторов, K-ближайший сосед и ансамбль). Результаты показали, что система ИИ, основанная на трансферном обучении, может использоваться для мультиклассовой классификации опухолей головного мозга и её производительность лучше, чем у систем машинного обучения.

Выбор методов и технологий реализации

## Данные

Набор данных, используемый в этой статье, представляет собой комбинацию трех наборов данных из figshare, SARTAJ и Br35H. [6] Данный набор данных содержит 7022 снимки МРТ, которые классифицируются по 4 классам: глиома, менингиома, без опухоли и гипофиз. Количество изображений для обучения, валидации и тестирования в соотношении 4:1:1. На рисунке 1 показаны 4 примеры снимки с разными классами. Снимки в этом наборе данных бывают разных размеров. Поэтому мы можем изменить их размер до определенного размера на этапе предварительной обработки и удаления лишних полей.

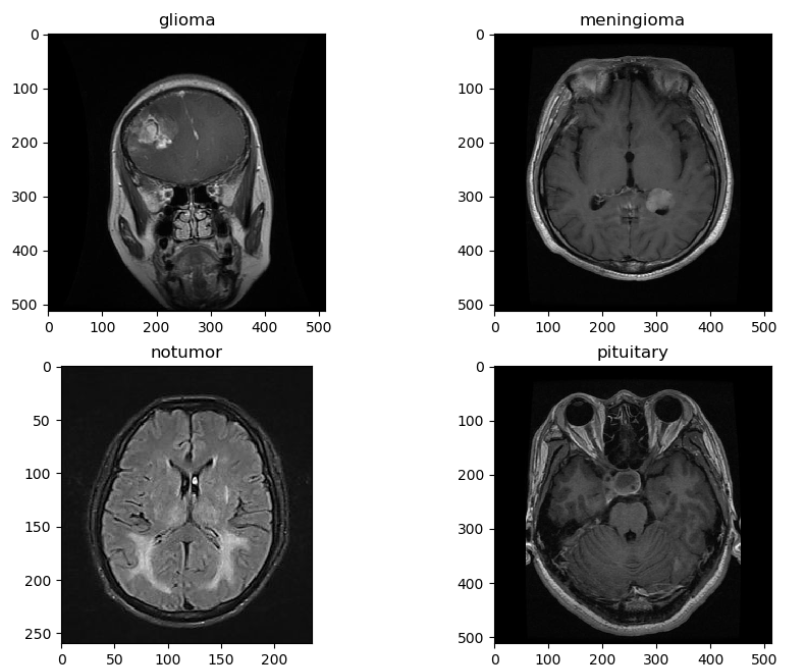


Рисунок 1 – Снимки с разными классами

Менингиомы в основном представляют собой доброкачественные опухоли, происходящие из клеток паутинной оболочки, и часто возникают у пожилых людей и женщин. Эти опухоли составляют 13–26% всех внутричерепных опухолей [7]. Глиомы являются наиболее частыми и первичными внутричерепными опухолями, которые являются злокачественными. Они составляют 81% всех внутричерепных опухолей, которые могут вызывать значительную смертность и заболеваемость [8]. Опухоли гипофиза возникают в гипофизе и в основном являются доброкачественными. Поскольку эта железа регулирует различные гормоны, присутствующие в ней опухоли могут вызывать серьезные изменения в организме. Эти опухоли составляют 10–15% всех внутричерепных опухолей [9].

## Свёрточная нейронная сеть (CNN)

В последние годы сверточные нейронные сети стали основным направлением исследований в области компьютерного зрения. Сверточные нейронные сети считаются одним из лучших алгоритмов обработки изображений и достигли наилучших результатов в классификации, сегментации и обнаружении изображений. Сверточные нейронные сети возникли в результате изучения искусственных нейронных сетей (ANN), обе из которых являются моделями, которые делают прогнозы о входных данных с помощью множества математических операций [10]. Искусственная нейронная сеть рассчитывается полносвязным образом, то есть все нейроны предыдущего слоя используются в качестве входных данных следующего слоя, который в основном используется для подгонки функций. Сверточная нейронная сеть использует сверточные вычисления для извлечения признаков, а область вычислений связана с размером ядра свертки.

Для задач распознавания изображений архитектура сверточной нейронной сети в общем виде показана на рисунке 2. Операции свертки и пулинга могут не только генерировать высокоуровневые карты признаков изображений, но также уменьшать объем вычислений и повышать эффективность работы сети. Они являются двумя важными элементами сверточных нейронных сетей.



Рисунок 2 – Архитектура свёрточной нейронной сети

### Слой свёртки

Слой свертки является основным методом работы CNN, в котором используется метод разделения весов, то есть операция свертки одного и того же канала использует одно и то же ядро свертки. На рисунке 3 показан пример процесса операции двумерной свертки. Рассмотрим изображение размером 4x4, рассчитанное с использованием двух ядер свертки 3x3 с разными весами, на выходе также должна быть карта признаков с двумя каналами. Сначала изображение расширяется нулями, а затем выполняется операция свертки. Процесс вычисления можно понимать как покрытие ядра свертки на изображении, начиная с первого пикселя в верхнем левом углу, пиксель умножается на соответствующий элемент ядра свертки, а затем суммируется. Результатом вывода является вывод центрального пикселя, покрытого ядром свертки. Затем ядро свертки перемещается упорядоченным образом в соответствии с заданным размером шага, пока не будет вычислено все изображение.

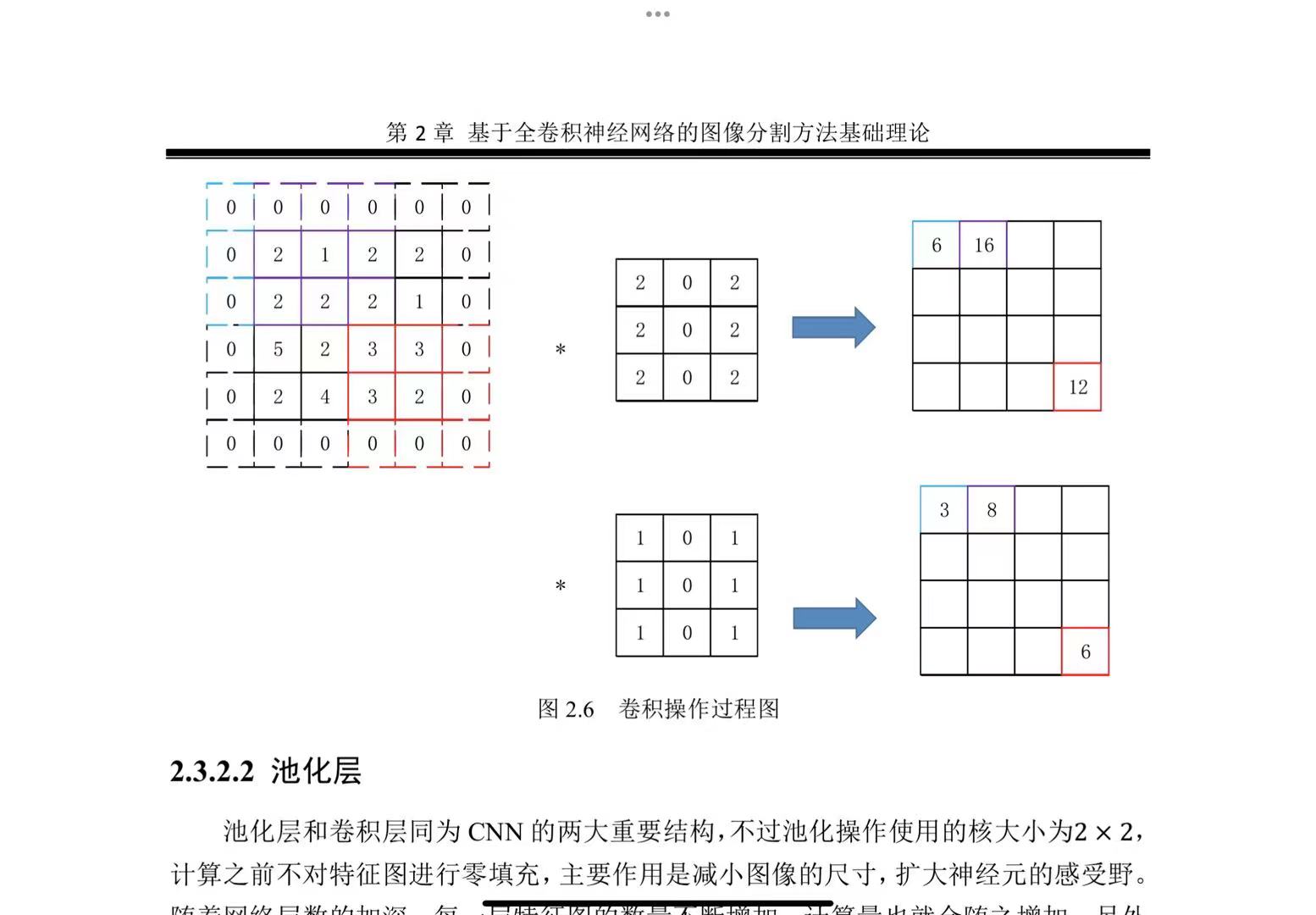


Рисунок 3 – Пример процесса операции двумерной свертки

### Слой активации

Слой активации предназначен для преобразования выходного значения предыдущего слоя в значение в определенном интервале с помощью определенной функции активации. Обычно используемые функции слоя активации включают функцию сигмоида, гиперболический тангенс, линейный выпрямитель (ReLU, Rectified Linear Unit), линейный выпрямитель с «утечкой» (Leaky ReLU), параметрический линейный выпрямитель (PReLU, Parameteric Rectified Linear Unit) и т. д.

Модель в данной работе использует функцию ReLU в качестве функции активации. Эта функция подавляет отрицательную полуось и активирует положительную полуось, что больше соответствует биологическим характеристикам. Она не имеет исчезновения градиента на положительной полуоси, и скорость вычислений примерно в 6 раз выше, чем у функции сигмоида и функции тангенса. На рисунке 4 показана схема функции ReLU, из нее видно, что данная функция возвращает 0, если принимает отрицательный аргумент, в случае же положительного аргумента, функция возвращает само число. То есть её можно записать в виде формулы:

,

где R(z) представляет собой выход ReLU, а z представляет вход ReLU.



Рисунок 4 – Функция ReLU

### Слой пулинга

Слои пулинга, иногда называемые слоями субдискретизации, обычно вычисляются с ядрами размером 2x2 без заполнения нулями карт объектов перед вычислением. Его основная роль заключается в уменьшении размера изображения и расширении рецептивного поля нейронов. По мере увеличения количества слоев сети количество карт признаков в каждом слое увеличивается, а также увеличивается объем вычислений. Кроме того, чем больше размер изображения, тем больше объем вычислений. Следовательно, разрешение изображения должно быть уменьшено с помощью операции пулинга, чтобы уменьшить объем вычислений сети. Слой пулинга, как правило, вставляется после слоя свёртки перед слоем следующей свёртки.

Операции пулинга включают пулинг с функцией максимума, пулинг с функцией среднего значения и т.д. На рисунке 5 показана схема пулинга с функцией максимума. Слева - входное изображение, а справа - результат после пулинга с функцией максимума. Размер ядра пулинга - 2x2. Подобно операции слоя свертки, операция пулинга также начинается с левого верхнего угла, перемещается в соответствии с заданным шагом и подсчитывает максимальное значение окна пулинга по очереди, пока все изображение не будет пройдено.



Рисунок 5 – Пулинг с функцией максимума и фильтром 2×2 с шагом 2

### Полносвязный слой

Полносвязный слой означает, что нейроны связаны со всеми пикселями изображения, что может преобразовать двумерное изображение в одномерные признаки. Из-за полносвязности каждый пиксель будет влиять на вывод полносвязного слоя, а размер ввода должен быть фиксированным. Полносвязный уровень обычно используется в задней части сети для классификации данных.

## Visual Geometry Group (VGG)

VGG [11] было предложено Visual Geometry Group. VGG состоит из слоев conv, pool, fc и softmax. Сверточный слой сети VGG не уменьшает изображение. Каждый слой площадки ценен, а уменьшение изображения реализуется пулом. VGG использует несколько слоев свертки с меньшими ядрами свертки (3x3) вместо одного слоя свертки с большим ядром свертки. С одной стороны, это может уменьшить параметры, с другой стороны, это эквивалентно больше нелинейному отображению, которое может увеличить подгонка сети. На рисунке 6 показана архитектура VGG.

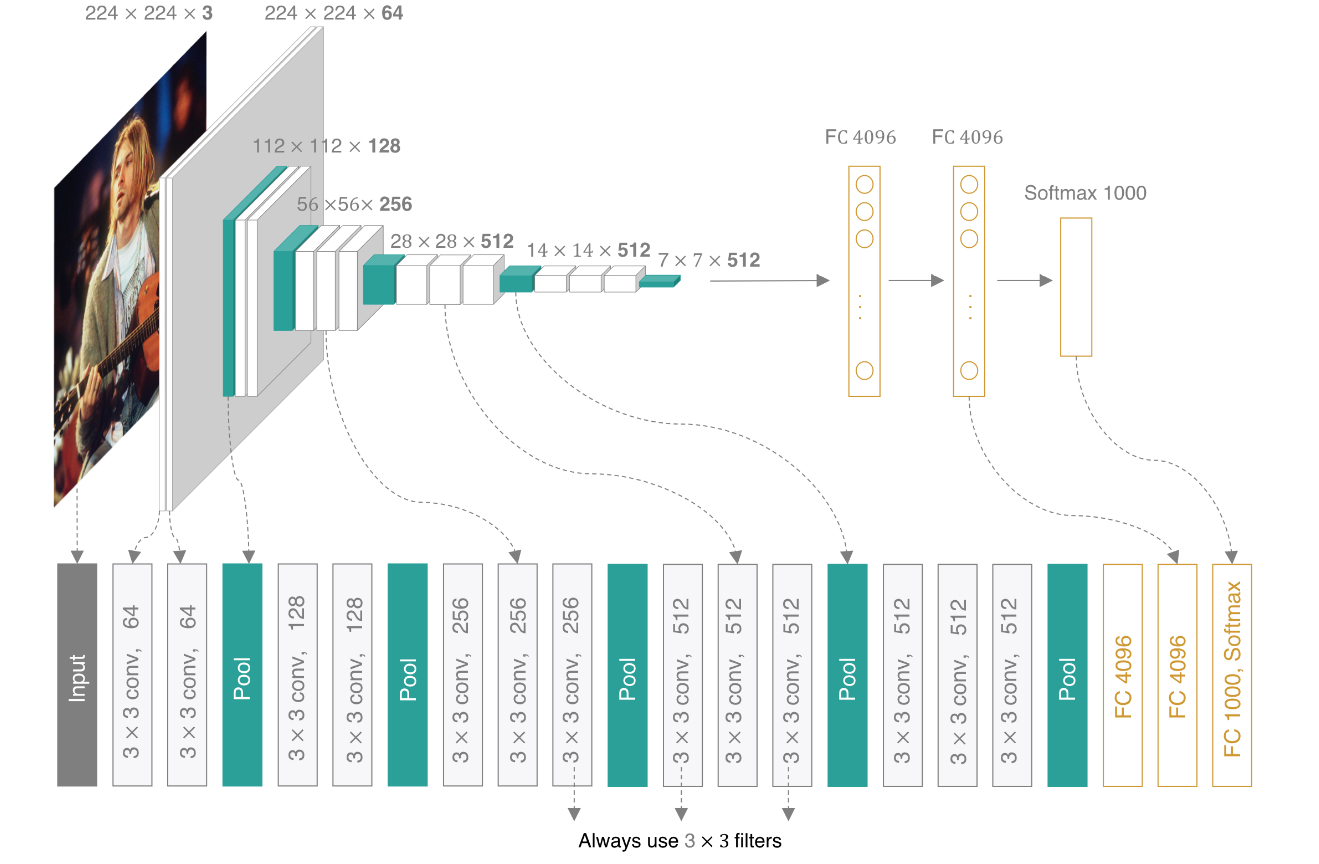


Рисунок 6 – Архитектура VGG

## ResNet

Глубокие сверточные нейронные сети совершили ряд прорывов в распознавании и классификации изображений. Люди пытаются использовать нейронные сети для более глубокого решения более сложных задач. Однако из-за таких проблем, как исчезновение градиента, если количество слоев модели постоянно увеличивается, точность начнет насыщаться и в итоге будет снижаться, что приведет к проблеме деградации, и скорость обучения сети будет постепенно замедляться. K He и др. предложили остаточные сети, которые позволяют обучать более 2000 слоев с более высокой точностью [12].

На рисунке 7 показана архитектура остаточного блока. Предположим, что нейронная сеть имеет x в качестве входных данных и аппроксимирует H (x). Обозначая разницу между ними как F(x), уравнение выглядит следующим образом:

*.*

F(x) — остаточная функция, из-за пропущенных соединений (skip connections) исходная функция становится F(x) + x, входной слой и выходной слой могут быть наложены друг на друга, а больший градиент распространяется на начальный слой, что позволяет обучать более глубокие сети и решать проблему исчезновения градиентов.

Архитектура остаточного блока может быть определена по формуле:

,

где y и x — выход и вход сети соответственно, а — остаточная функция, используемая для обучения. Архитектура остаточного блока, показанная на рисунке 6, имеет два слоя, тогда

,

где представляет собой функцию активации, а вход и выход завершаются short skip connections и не вводят дополнительные параметры и вычисления.

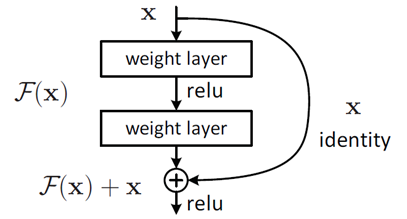


Рисунок 7 – Архитектура остаточного блока

## Vision Transformer

Трансформер изначально использовался в области обработки естественного языка (Natural Language Processing, NLP) и оказался больших успехов. [13] Трансформер отказывается от традиционной последовательной структуры рекуррентной нейронной сети (RNN) и использует механизм self-attention, позволяющий параллельно обучать модель и обрабатывать глобальную информацию. Вдохновленные механизмом self-attention в NLP, некоторые модели на основе сверточных нейронных сетей (CNN) используют пространственное внимание или канальное внимание, или и то, и другое для извлечения значимых признаков в изображениях. Хотя эти модели с механизмом внимания достигают неплохих результатов, они все же уступают моделям CNN.

Vision Transformer (ViT) можно рассматривать как графическую версию Трансформер, где стандартная модель Transformer переносится непосредственно в область компьютерного зрения (CV) с минимально изменениями. Чтобы применить модель Transformer к изображениям, авторы [3] разрезали изображения на подблоки и сформировали эти подблоки в линейные последовательности встраивания, которые затем использовали в качестве входных данных для Transformer, чтобы имитировать последовательность фраз в области NLP. На рисунке 8 представлена архитектура ViT. Верхние слои ViT изучают глобальные объекты, тогда как нижние слои изучают как глобальные, так и локальные объекты. Это позволяет ViT изучать более общие паттерны.

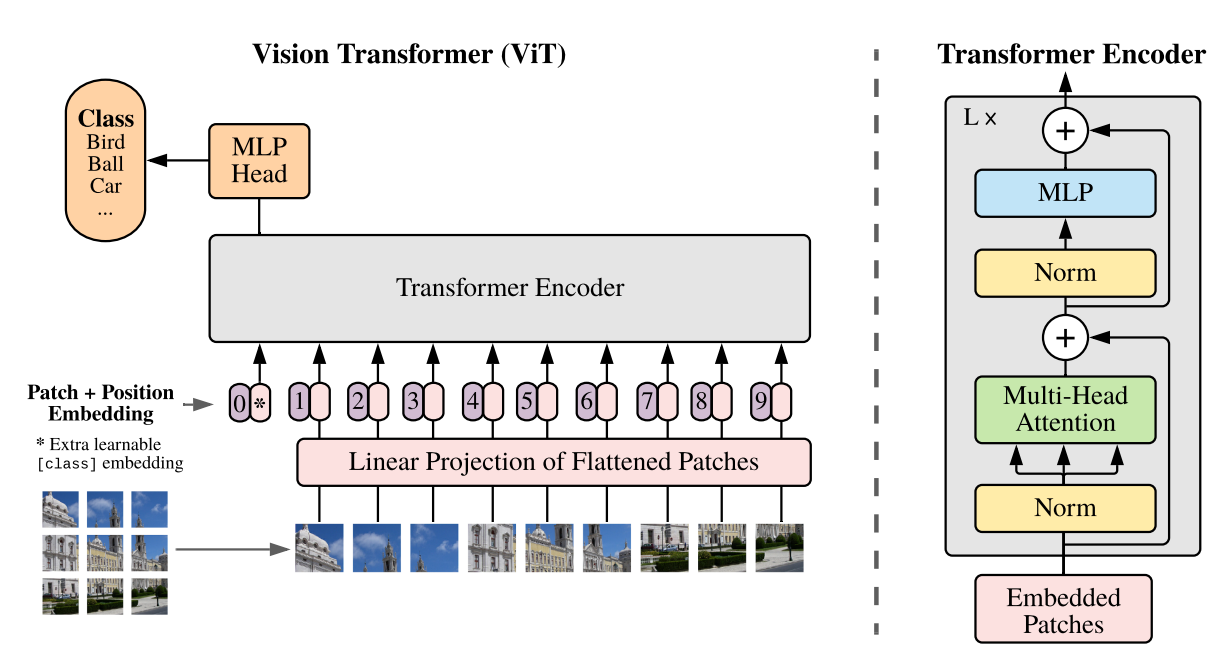


Рисунок 8 – Архитектура ViT

### Self-attention

Self-attention является основным модулем Vision Transformer, и формула его расчета такова:

*,*

где

* Q – Query: слово, с которого смотрим на всё остальное
* K – Key: слово, на которое смотрим
* V – Value: здесь содержится смысл слова

### Слой Embedding

Для стандартного модуля Трансформер требуемый вход — последовательность токенов (векторов), то есть двумерная матрица [num\_token, token\_dim]. Для данных изображения его формат данных — [канал, высота, вес], который представляет собой трехмерную матрицу, что явно не то, что нужно Трансформер. Следовательно, необходимо сначала преобразовать данные через слой Embedding. Во-первых, разделить картинку на кучу патчей определенного размера. В данной работе входное изображение (img\_size - 224x224) разбито на 7\*7 патчей, и 32 патча будет получено после деления. Затем каждый патч сопоставляется с одномерным вектором посредством линейного отображения, а форма данных каждого патча имеет вид [7, 7, 3]. Вектор длиной 147 получается посредством отображения (далее именуемый непосредственно токеном).

Обратите внимание, что [class]token и Position Embedding необходимо добавить перед входом в Transformer Encoder. [class]token — это обучаемый параметр, специально используемый для классификации, а его формат является вектором, как и другие токены, который сращивается вместе с токенами, сгенерированными из изображения ранее.

### Transformer Encoder

Transformer Encoder фактически многократно складывает Encoder Block L раз. Encoder Block состоит из Layer Normalization (LN), Multi-Head Self Attention Layer (MSA), Многослойные персептроны (MLP), Residual connections.

* LN перед каждым блоком в основном предложена для области NLP. LN нормализует каждый токен и помогает улучшить время обучения и производительность обобщения.
* MSA преобразуется на основе self-attention и улучшает использование информации за счет разделения Q, K и V на несколько head. MSA может извлекать локальную информацию и глобальную информацию, содержащуюся во входной последовательности.
* Блок MLP состоит из Fully Connected Layer(FC), функции активации GELU и Dropout.
* Residual connections применяются после каждого блока.

### MLP Head

Для классификации изображений классификация реализована с использованием MLP с одним скрытым слоем. MLP Head состоит из Linear, функции активации tanh и Linear при обучении набора данных, но при переносе обучения других данных достаточно только одного Linear.

## Сравнение архитектур VGG, ResNet, ViT

В результате изучения существующих решений была создана таблица 1.

|  |  |  |
| --- | --- | --- |
| **Метод классификации** | **Достоинства** | **Недостатки** |
| VGG | 1. Простая архитектура; 2. Использует несколько сверточных слоев с меньшими ядрами свертки, что может уменьшить параметры. | 1. Высокие требования к памяти или вычислительным ресурсам; 2. Не может получить глобальные особенности изображения. 3. Проблема исчезновения градиента |
| ResNet | Введен остаточный блок для эффективного устранения проблемы исчезновения градиента и дальнейшего повышения точности классификации. | Не может получить глобальные особенности изображения. |
| ViT | Возможность получить глобальные особенности изображения | Необходимость больших объёмов данных, т.к. на маленьком наборе данных точность низкая. |

Таблица 1 – Сравнительная таблица известных решений

Проектирование и разработка модели

Полный код реализации модели можно найти по ссылке <https://colab.research.google.com/drive/1RWZLcSkmlhDdNN3eJO4zHsFHP2lBXTZ6?usp=sharing>.

## Предобработка данных

Из-за разных размеров изображений на этапе предварительной обработки данных изображения необходимо уменьшить до определенного размера (224x224). Затем можно искусственно увеличить количество изображений, вращая и отразя изображение по горизонтали.

## Разработка модели

В данной работе разработана модель, основанная на архитектуре Vision Transformer с multihead linear self-attention. На рисунке 9 показана архитектура модели.

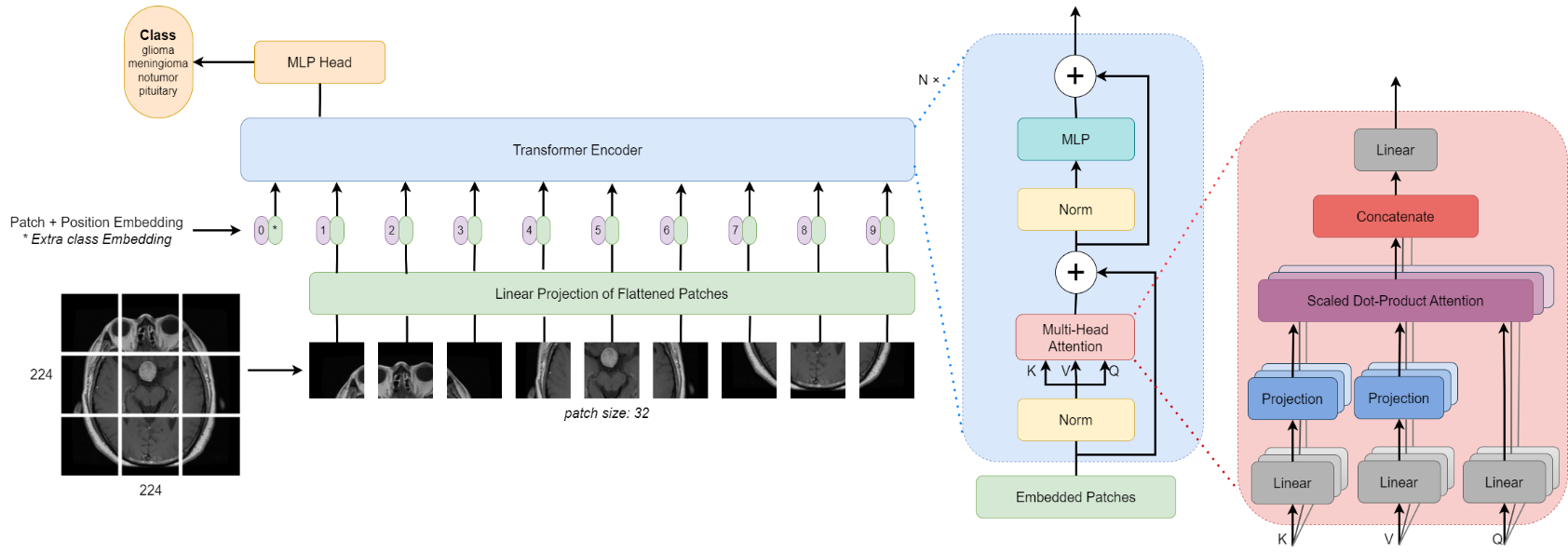


Рисунок 9 – Архитектура модели

Мы разбиваем предобработанные изображения (размер 224x224) на 7x7 патчи с фиксированным размером 32. Затем линейно встраиваем каждый из них, добавляем Position Embedding и передаем полученную последовательность векторов в Transformer Encoder. Чтобы выполнить классификацию, мы добавляем в последовательность дополнительный Classification Token.

В Transformer Encoder самое важное это multihead self-attention. На самом деле мы можем изменить его на собственный механизм. Например, Nystromformer, Routing Transformer и т.д. Здесь я вставила multihead linear self-attention. Благодаря его низкому рангу механизма у него меньше пространственная и временная сложность. После ряда вычислений мы наконец получим классификационный вектор.

## Функция потерь

В данной работе в качестве функции потерь используется функция потерь перекрестной энтропии, часто используемая для измерения разницы между двумя распределениями вероятностей. [15] Формула для расчета бинарной классификации выглядит следующим образом:

,

где y - предсказанная вероятность и - индикатор (0 или 1 в случае бинарной классификации). Общая кросс-энтропия представляет собой сумму кросс-энтропии для каждого класса в задаче мультиклассовой классификации.

На рисунке 10 представлены кривые функции потерь при обучении и валидации предложенной модели соответственно, а на рисунке 11 показана диаграмма точности модели.

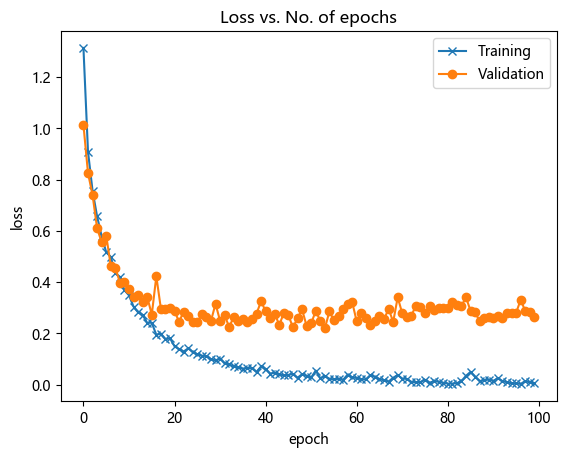


Рисунок 10 – Кривые функции потерь при обучении и валидации модели

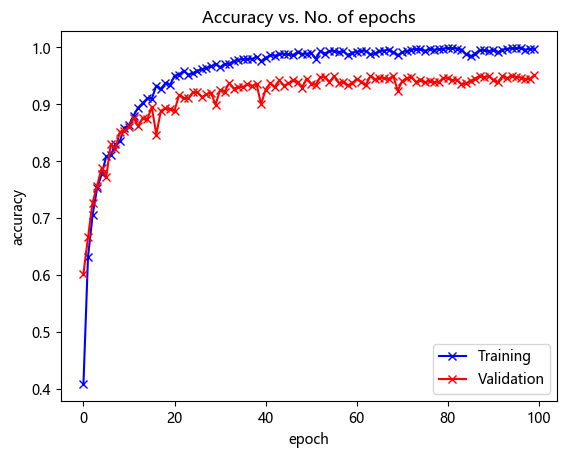


Рисунок 11 – Диаграмма точности модели

Исследование и тестирование модели

В данной работе модель обучается на графическом процессоре NVIDIA GeForce RTX 2060 6 ГБ. Во время обучения набор данных обучается с размером партий – 32, скоростью обучения – 0.00007 и количеством эпох – 100. Оптимизатором для обучения экспериментальной модели в этой работе является Adam, предоставленный pytorch.

Результаты классификации ОГМ по снимкам МРТ оценены с помощью трех мер оценки: Precisiоn (точность), Recall (полнота) и Micro F1-score. Отчеты о результатах классификации на этапе обучении и валидации представлены в таблице 2 и 3 соответственно. Настроенная и обученная модель на валидационной выборке позволила получить оценку точности 91.2%, полноты 91% и F1-мер 91%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Классы | Precision | Recall | Micro F1-score | support |
| Глиома | 0.95 | 0.93 | 0.94 | 105700 |
| Менингиома | 0.93 | 0.92 | 0.92 | 107100 |
| Без опухоли | 0.97 | 0.98 | 0.97 | 127600 |
| Гипофиз | 0.97 | 0.98 | 0.97 | 116500 |
| Общая | 0.955 | 0.952 | 0.95 | 456900 |

Таблица 2 – Отчет о результатах классификации на этапе обучении

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Классы | Precision | Recall | Micro F1-score | support |
| Глиома | 0.91 | 0.85 | 0.88 | 26400 |
| Менингиома | 0.84 | 0.85 | 0.84 | 26800 |
| Без опухоли | 0.94 | 0.96 | 0.95 | 31900 |
| Гипофиз | 0.96 | 0.98 | 0.97 | 29200 |
| Общая | 0.912 | 0.91 | 0.91 | 114300 |

Таблица 3 – Отчет о результатах классификации на этапе валидации

Из рисунка 12 интуитивно видно, что данная сеть может хорошо классифицировать опухоли и выполнять данную задачу.

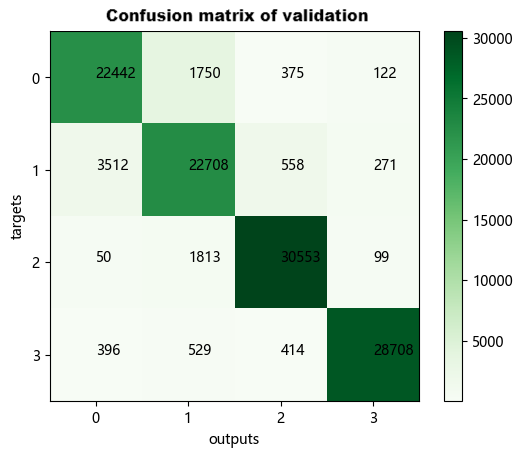


Рисунок 12 – Сonfusion matrix

Выводы

В результате данной работы изучены различные подходы к предобработке медицинских изображений и различные архитектуры нейронных сетей, часто используемые для классификации изображений. Реализована и обучена модель на основе архитектуры трансформер для мультиклассовой классификации опухолей головного мозга по снимкам МРТ. В рамках работы была настроенная и обученная модель на валидационной выборке позволила получить окончательное значение F1-меры 91,35%. Это может продемонстрировать применимость архитектур трансформер для задач мультиклассовой классификации опухолей головного мозга.

Дальнейшее развитие данной НИР связано с повышением точности модели путем изменения различных механизмов внутреннего внимания или дообучения модели.

Список литературы

1. M. R. Ismael. Hybrid Model-Statistical Features and Deep Neural Network for Brain Tumor Classification in MRI Images. — 2018. URL: <https://scholarworks.wmich.edu/dissertations/3291/>
2. Информационные технологии анализа изображений в задачах медицинской диагностики: коллективная монография. — Москва: Радио и связь. Н. Ю. Ильясова, А. В. Куприянов, А. Г. Храмов, 2012. — 424 с.
3. Dosovitskiy A., Beyer L., Kolesnikov A., et al. An image is worth 16x16 words: Transformers for image recognition at scale // arXiv preprint arXiv:2010.11929. 2020. URL: <https://arxiv.org/abs/2010.11929>
4. Deepak, S., Ameer, P. M. Brain tumor classification using deep CNN features via transfer learning // Computers in biology and medicine: 111, 103345. 2019. URL: <https://www.sciencedirect.com/science/article/pii/S0010482519302148>
5. Tandel G. S., Balestrieri A., Jujaray T., et al. Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm // Computers in Biology and Medicine, 2020, 122: 103804. URL: <https://www.sciencedirect.com/science/article/pii/S0010482520301724>
6. Brain Tumor MRI Dataset. URL: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>
7. Marosi, C.; Hassler, M.; Roessler, K.; Reni, M.; Sant, M.; Mazza, E.; Vecht, C. Meningioma // Crit. Rev. Oncol. Hematol. 2008. 67, 153–171.
8. Ostrom, Q.T.; Gittleman, H.; Stetson, L.; Virk, S.M.; Barnholtz-Sloan, J.S. Epidemiology of gliomas // Cancer Treat. Res. 2015. 163, 1–14.
9. Chintagumpala, M.; Gajjar, A. Brain tumors. // Pediatr. Clin. N. Am. 2015. 62, 167–178.
10. Lundervold A. S., Lundervold A. An overview of deep learning in medical imaging focusingon MRI // Zeitschrift fur Medizinische Physik. – 2019. – Т. 29. – No. 2. – С. 102-127.
11. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J] // arXiv preprint arXiv:1409.1556, 2014.
12. K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks // European conference on computer vision. – 2016. – P. 630–645.
13. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. // Advances in neural information processing systems. 2017. 30.
14. Функция потерь перекрестной энтропии: Обзор. URL: <https://wandb.ai/wandb_fc/russian/reports/---VmlldzoxNDI4NjAw>