**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 27.06.2023

Санкт-Петербург

2023

**Оглавление**

[Введение 3](#_Toc138778310)

[Анализ предметной области 4](#_Toc138778311)

[Описание проблемы 4](#_Toc138778312)

[Обзор литературы 4](#_Toc138778313)

[Выбор методов и технологий реализации 6](#_Toc138778314)

[Данные 6](#_Toc138778315)

[Vision Transformer 7](#_Toc138778316)

[Слой Embedding 7](#_Toc138778317)

[Transformer Encoder 8](#_Toc138778318)

[MLP Head 8](#_Toc138778319)

[Self-attention 8](#_Toc138778320)

[Проектирование и разработка модели 11](#_Toc138778321)

[Предобработка данных 11](#_Toc138778322)

[Разработка модели 11](#_Toc138778323)

[Функция потерь 11](#_Toc138778324)

[Исследование и тестирование модели 13](#_Toc138778325)

[Простой ViT 13](#_Toc138778326)

[ViT с Linformer 14](#_Toc138778327)

[ViT с Nystromformer 15](#_Toc138778328)

[ViT с Sinkhorn Transformer 17](#_Toc138778329)

[Выводы 19](#_Toc138778330)

[Список литературы 20](#_Toc138778331)

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.

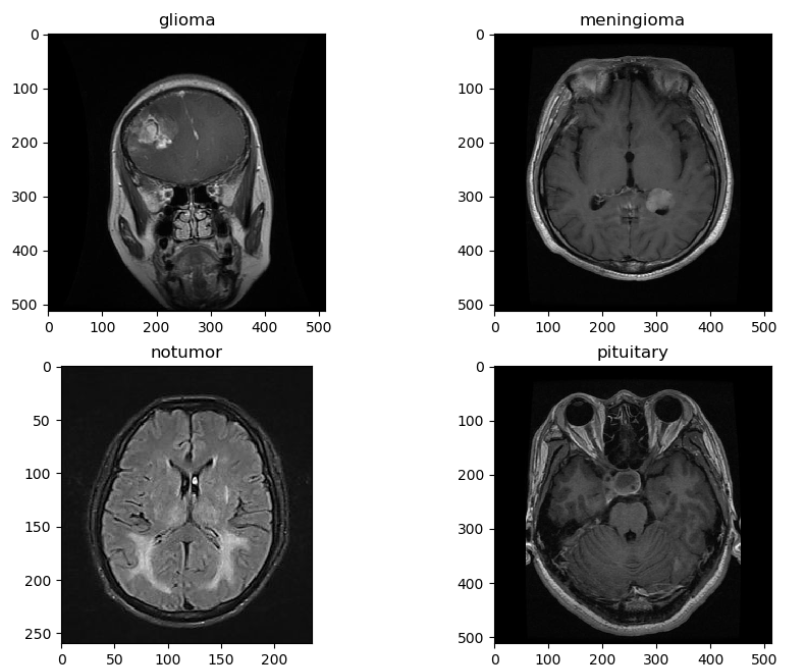


Рис.1 – Снимки с разными классами

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].

## Vision Transformer

Трансформер изначально использовался в области обработки естественного языка (Natural Language Processing, NLP) и оказался больших успехов. [10] 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 2 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.

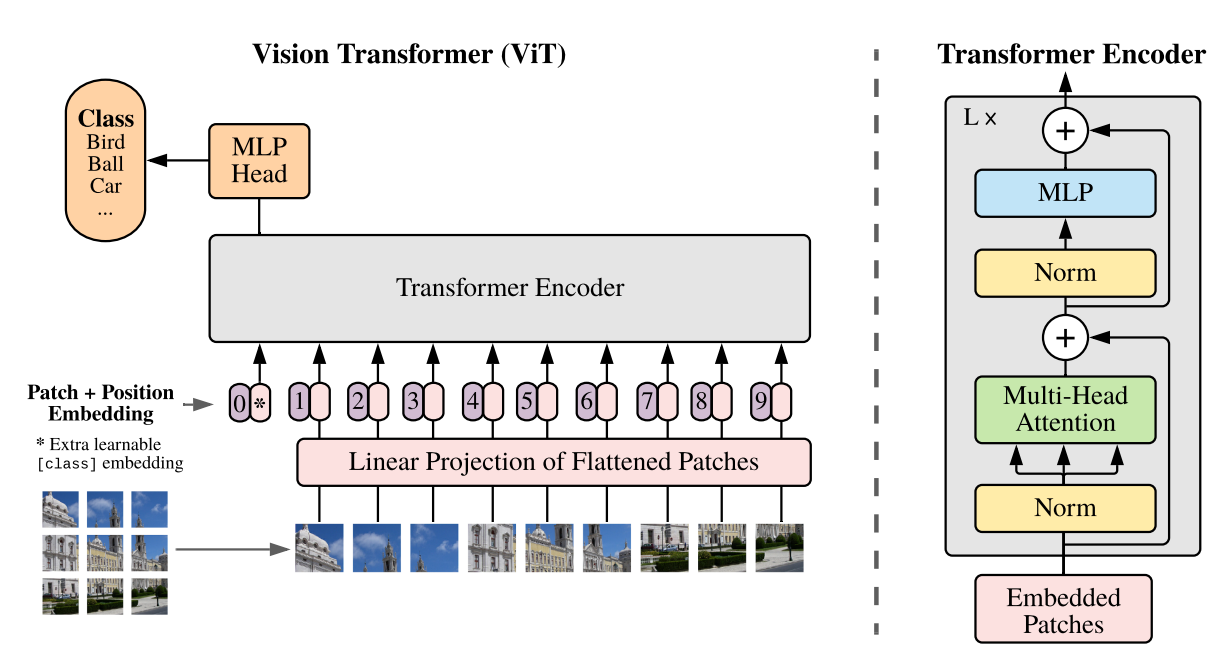


Рис.2 – Архитектура ViT

### 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.

### 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
* dK – Dimension: размерность вложений входной последовательности

Self-attention позволяет модели учитывать различные зависимости в последовательности, однако у него есть квадратичная вычислительная сложность относительно длины входной последовательности, что может вызывать трудности при работе с длинными последовательностями. В связи с этим было предложено множество вариантов Transformer с различными механизмами self-attention.

#### Linformer

Linformer (Linear Tranformer) [11] — это вариант Transformer, предложенный Facebook AI, который облегчает обучение и использование Transformer-моделей для обработки больших последовательностей данных. Формула Linformer выглядит следующим образом:

*,*

где

* Q – Query: слово, с которого смотрим на всё остальное
* K – Key: слово, на которое смотрим
* V – Value: здесь содержится смысл слова
* dK – Dimension: размерность вложений входной последовательности
* E, F – Матрицы проекции, который используются для проецирования входных данных в пространство ключей и значений.

На рисунке 3 показана схема архитектуры Linformer.

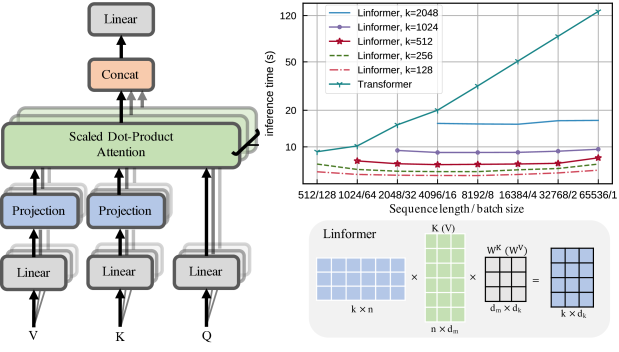


Рис.3 – Linformer

Вместо того чтобы вычислять попарные взаимодействия между всеми элементами входной последовательности, Linformer использует проекции в меньшее пространство, называемое пространством ключей и значений (key-value space). Это достигается с помощью двух матриц проекции E и F, которые умножаются на входные данные для сокращения размерности. Это позволяет сократить вычислительную сложность механизма самовнимания с квадратичной до линейной.

#### Nystromformer

Nystromformer [12] — это вариант Transformer, который достигается путем использования метода Нистрома для аппроксимации матрицы самовнимания. Это позволяет эффективно работать с длинными последовательностями, сокращая вычислительную сложность. На рисунке 4 представлена схема архитектуры Nystromformer.

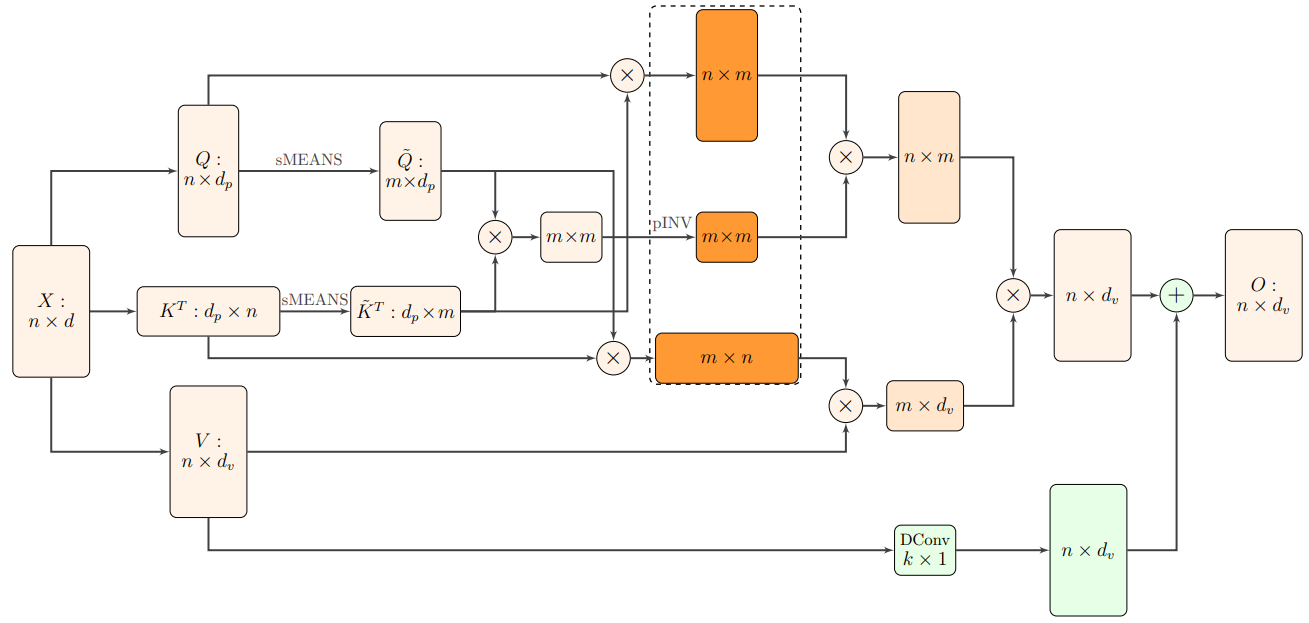


Рис.4 – Nystromformer

#### Sinkhorn Transformer

Sinkhorn Transformer [13] — это вариант архитектуры Transformer, который использует метод Sinkhorn для нормализации матрицы внимания. Это позволяет эффективно работать с длинными последовательностями, сокращая вычислительную сложность и улучшая качество модели. На рисунке 5 представлена схема архитектуры Sinkhorn Transformer. Он включает в себя параметризованную сеть сортировки, использующую нормализацию синхорна для выборки матрицы перестановок, которая сопоставляет наиболее релевантные группы ключей с группами запросов. Он также включает обратимые сети и фрагментацию с прямой связью, чтобы обеспечить дополнительную экономию памяти.

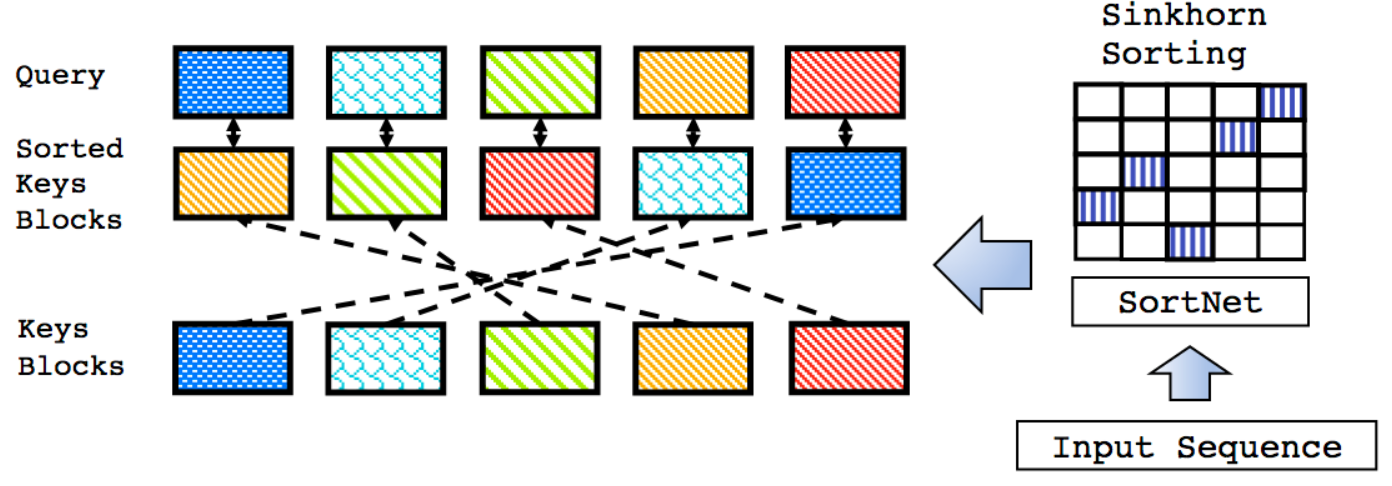


Рис.5 – Sinkhorn Transformer

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 6 показана архитектура модели.

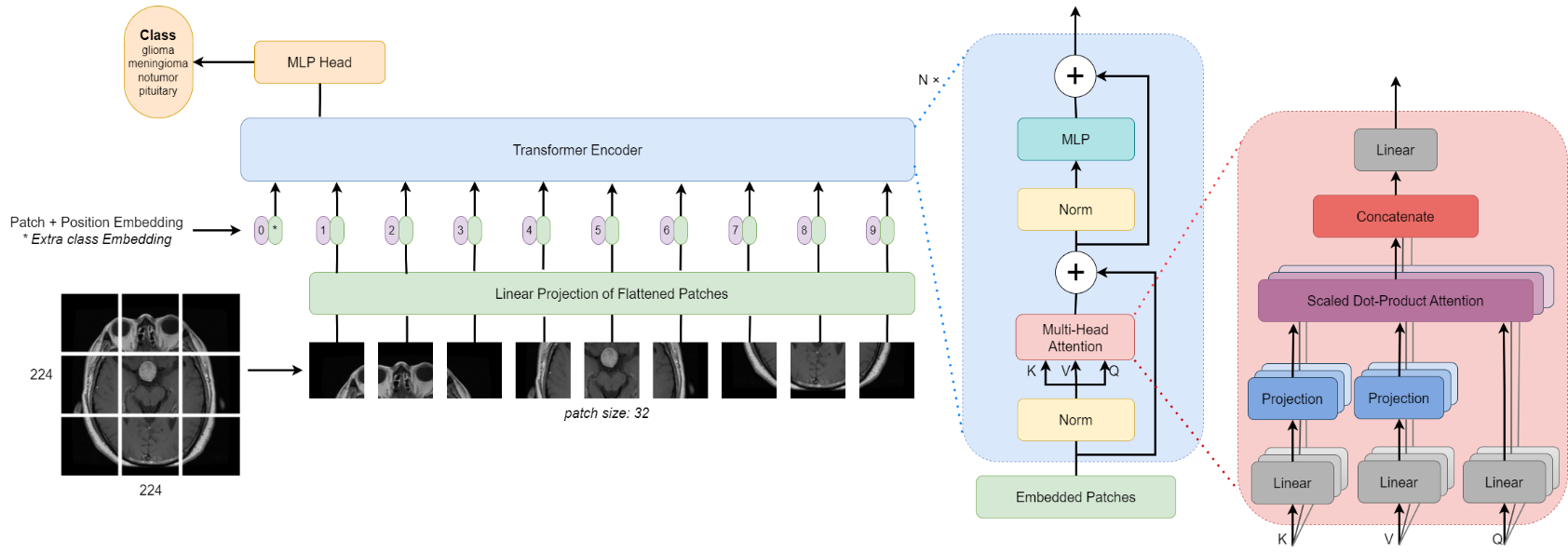


Рис.6 – Архитектура модели

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. В данной работе попробовала multihead linear self-attention, multihead Nystromformer и multihead Sinkhorn transformer. Благодаря их низкому рангу механизма у моделей 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. [14] 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.

Исследование и тестирование модели

В данной работе модели обучаются на графическом процессоре NVIDIA GeForce RTX 3070 8 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.

В ходе экспериментальных исследований были протестированы следующие архитектуры моделей: простой ViT, ViT с применением Linformer, ViT с применением Nystromformer, и ViT с применением Sinkhorn Transformer. Результаты классификации ОГМ по снимкам МРТ оценены с помощью трех мер оценки: Precisiоn (точность), Recall (полнота) и Micro F1-score.

Простой ViT

На рисунке 7 представлены диаграммы кривых функции потерь и кривых точности при обучении и валидации предложенной модели соответственно.

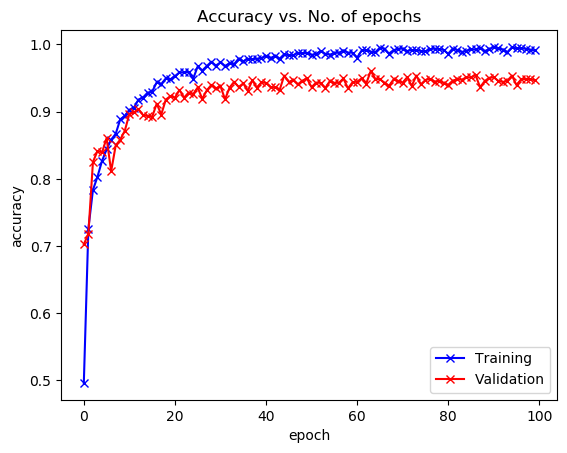
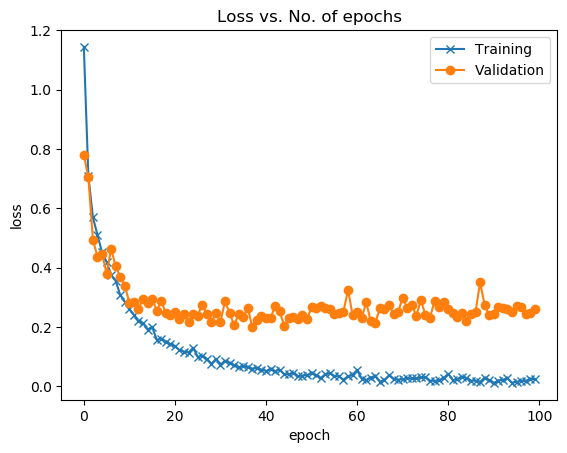


Рис.7 – Диаграммы кривых функции потерь и кривых точности при обучении и валидации модели ViT

Отчеты о результатах классификации, основанные на архитектуре простого ViT в процессе обучения и валидации, приведены в таблицах 1 и 2 соответственно. Настроенная и обученная модель простого ViT на валидационной выборке позволила получить оценку точности 89%, полноты 89% и F1-мер 89%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Классы | Precision | Recall | Micro F1-score | support |
| Глиома | 0.91 | 0.90 | 0.91 | 105700 |
| Менингиома | 0.88 | 0.86 | 0.87 | 107100 |
| Без опухоли | 0.95 | 0.96 | 0.96 | 127600 |
| Гипофиз | 0.94 | 0.96 | 0.95 | 116500 |
| Общая | 0.92 | 0.92 | 0.92 | 456900 |

Таблица 1 – Отчет о результатах классификации на этапе обучении (ViT)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Классы | Precision | Recall | Micro F1-score | support |
| Глиома | 0.89 | 0.85 | 0.87 | 26400 |
| Менингиома | 0.82 | 0.80 | 0.81 | 26800 |
| Без опухоли | 0.93 | 0.94 | 0.93 | 31900 |
| Гипофиз | 0.93 | 0.97 | 0.95 | 29200 |
| Общая | 0.89 | 0.89 | 0.89 | 114300 |

Таблица 2 – Отчет о результатах классификации на этапе валидации (ViT)

Из рисунка 8 интуитивно видно, что данная модель может хорошо классифицировать опухоли и выполнять данную задачу.

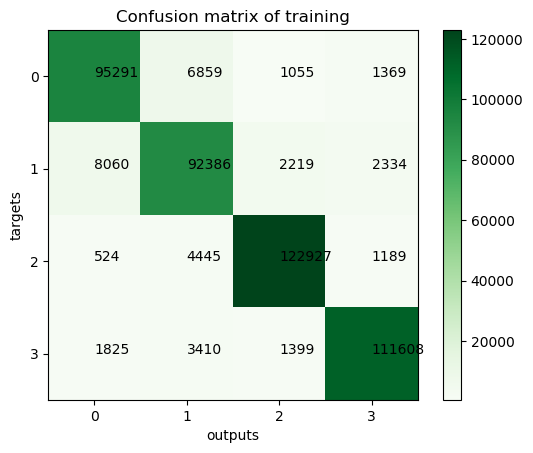
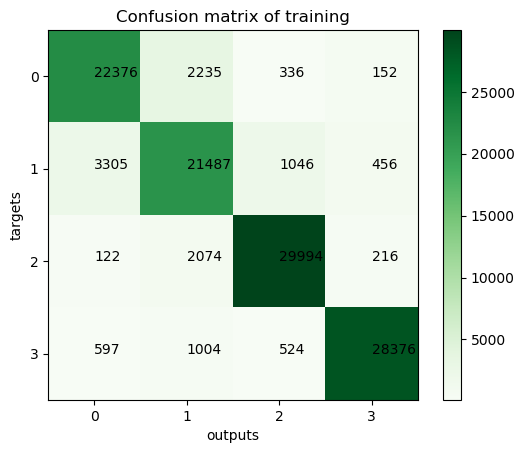
 

Рисунок 8 – Сonfusion matrix (ViT)

ViT с Linformer

На рисунке 9 представлены диаграммы кривых функции потерь и кривых точности при обучении и валидации модели ViT с использованием Linformer соответственно.

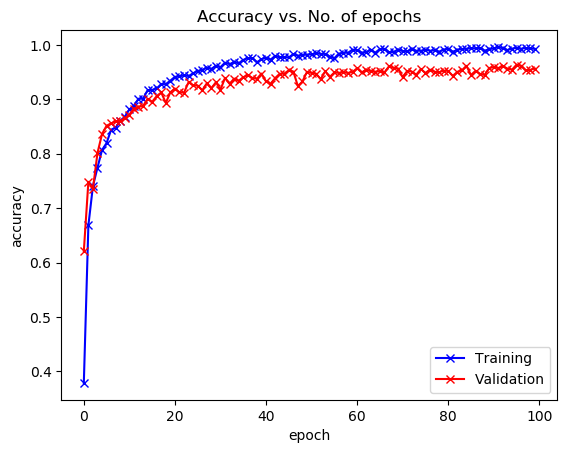
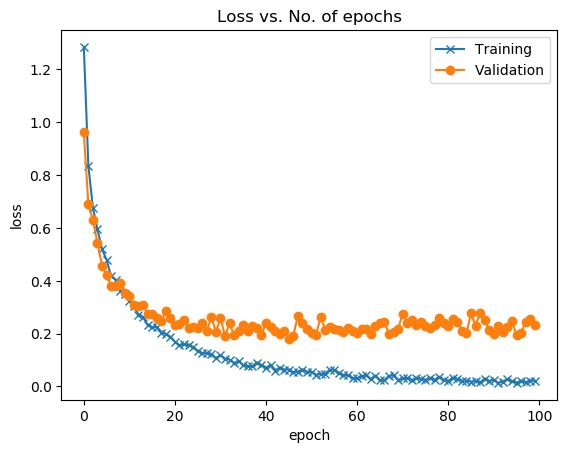


Рис.9 – Диаграммы кривых функции потерь и кривых точности при обучении и валидации модели ViT с Linformer

Отчеты о результатах классификации, основанные на архитектуре простого ViT с применением Linformer в процессе обучения и валидации, приведены в таблицах 3 и 4 соответственно. После настройки и обучения модели ViT с использованием Linformer на валидационной выборке, удалось достичь высоких показателей точности, полноты и F1-меры - соответственно 91,2%, 91% и 91%. Эти результаты говорят о эффективности внутреннего внимания Linformer.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 3 – 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 4 – Отчет о результатах классификации на этапе валидации

На рисунке 10 представлены диаграммы Сonfusion matrix при обучении и валидации модели ViT с Linformer.

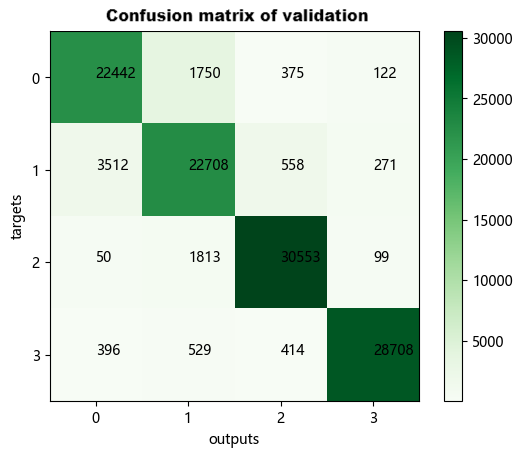


Рисунок 10 – Сonfusion matrix (ViT с Linformer)

ViT с Nystromformer

На рисунке 11 представлены диаграммы кривых функции потерь и кривых точности при обучении и валидации модели ViT с использованием Nystromformer соответственно.

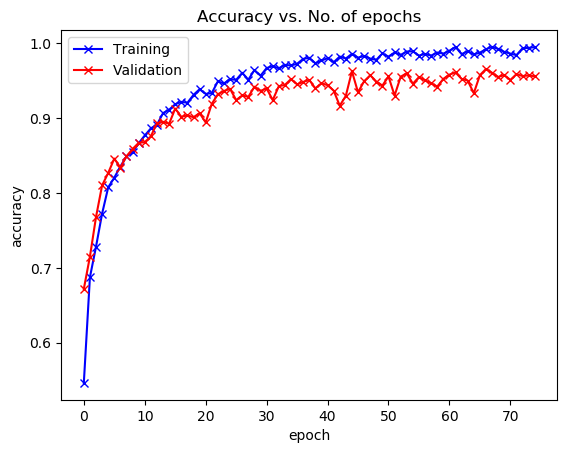
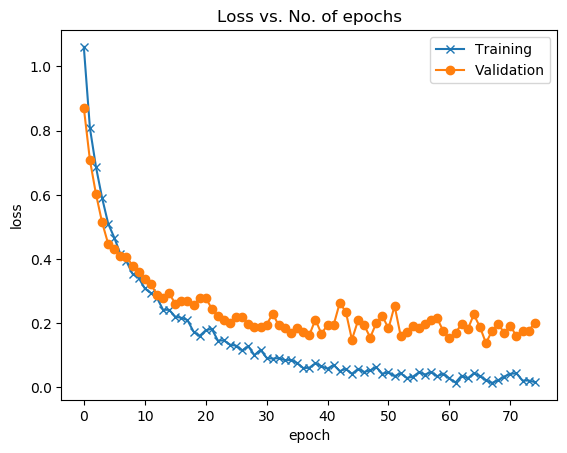


Рис.11 – Диаграммы кривых функции потерь и кривых точности при обучении и валидации модели ViT с Nystromformer

Отчеты о результатах классификации на этапе обучении и валидации представлены в таблице 5 и 6 соответственно. Настроенная и обученная модель ViT с использованием Nystromformer на валидационной выборке позволила получить оценку точности 92%, полноты 92% и F1-мер 92%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Классы | Precision | Recall | Micro F1-score | support |
| Глиома | 0.93 | 0.92 | 0.92 | 105700 |
| Менингиома | 0.90 | 0.89 | 0.90 | 107100 |
| Без опухоли | 0.97 | 0.97 | 0.97 | 127600 |
| Гипофиз | 0.96 | 0.97 | 0.97 | 116500 |
| Общая | 0.94 | 0.94 | 0.94 | 456900 |

Таблица 5 – Отчет о результатах классификации на этапе обучении

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Классы | Precision | Recall | Micro F1-score | support |
| Глиома | 0.91 | 0.87 | 0.89 | 26400 |
| Менингиома | 0.86 | 0.85 | 0.85 | 26800 |
| Без опухоли | 0.94 | 0.96 | 0.95 | 31900 |
| Гипофиз | 0.96 | 0.98 | 0.97 | 29200 |
| Общая | 0.92 | 0.92 | 0.92 | 114300 |

Таблица 6 – Отчет о результатах классификации на этапе валидации

На рисунке 12 представлены диаграммы Сonfusion matrix при обучении и валидации модели ViT с использованием Nystromformer.

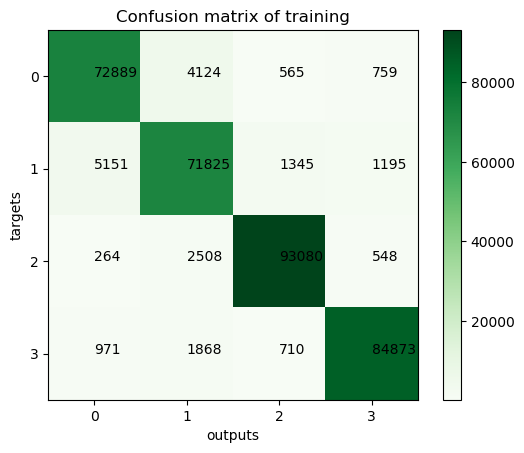
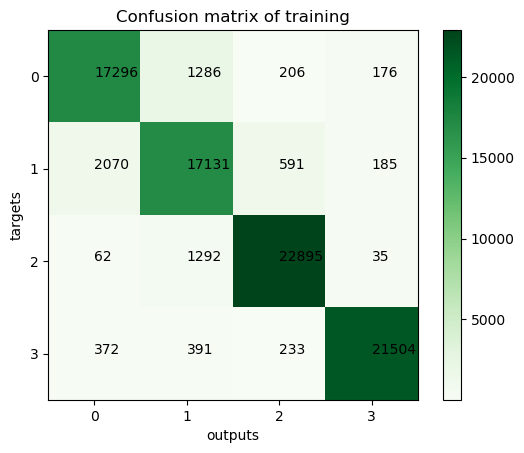
 

Рисунок 12 – Сonfusion matrix (ViT с Nystromformer)

ViT с Sinkhorn Transformer

На рисунке 13 представлены диаграммы кривых функции потерь и кривых точности при обучении и валидации модели ViT с использованием Sinkhorn Transformer соответственно.

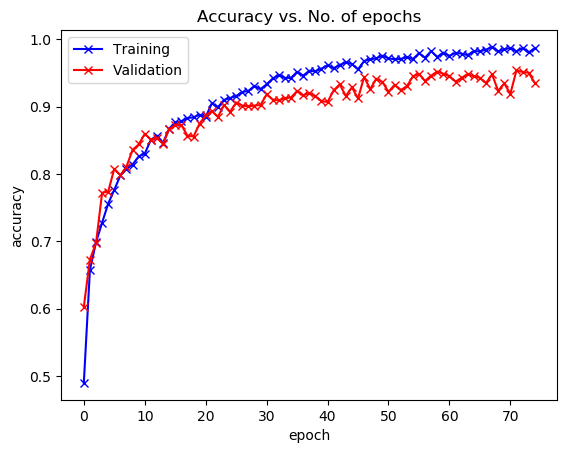
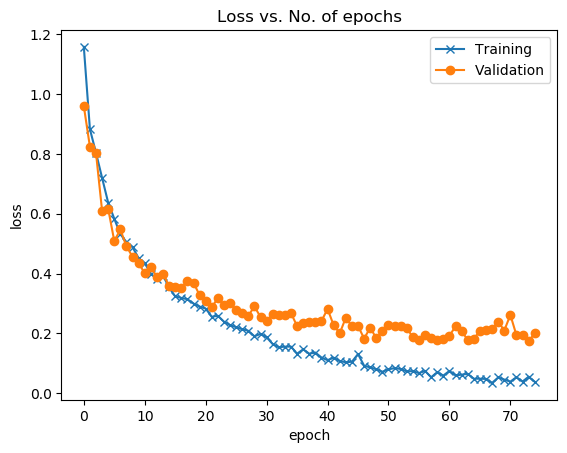


Рис.13 – Диаграммы кривых функции потерь и кривых точности при обучении и валидации модели ViT с Sinkhorn Transformer

Отчеты о результатах классификации на этапе обучении и валидации представлены в таблице 7 и 8 соответственно. После настройки и обучения модели ViT с использованием Sinkhorn Transformer на валидационной выборке были получены следующие показатели: точность - 89%, полнота - 89% и F1-мера - 89%. Стоит отметить, что F1-мера модели ViT с использованием Sinkhorn Transformer оказалась такой же, как у модели ViT без применения Sinkhorn Transformer. Это может указывать на то, что внутреннее внимание Sinkhorn Transformer не дало значительного прироста в эффективности модели.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Классы | Precision | Recall | Micro F1-score | support |
| Глиома | 0.90 | 0.89 | 0.90 | 105700 |
| Менингиома | 0.87 | 0.85 | 0.86 | 107100 |
| Без опухоли | 0.94 | 0.96 | 0.95 | 127600 |
| Гипофиз | 0.94 | 0.95 | 0.95 | 116500 |
| Общая | 0.91 | 0.91 | 0.91 | 456900 |

Таблица 7 – Отчет о результатах классификации на этапе обучении

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Классы | Precision | Recall | Micro F1-score | support |
| Глиома | 0.89 | 0.85 | 0.87 | 26400 |
| Менингиома | 0.81 | 0.81 | 0.81 | 26800 |
| Без опухоли | 0.93 | 0.93 | 0.93 | 31900 |
| Гипофиз | 0.94 | 0.98 | 0.95 | 29200 |
| Общая | 0.89 | 0.89 | 0.89 | 114300 |

Таблица 8 – Отчет о результатах классификации на этапе валидации

На рисунке 14 представлены диаграммы Сonfusion matrix при обучении и валидации модели ViT с использованием Sinkhorn Transformer.

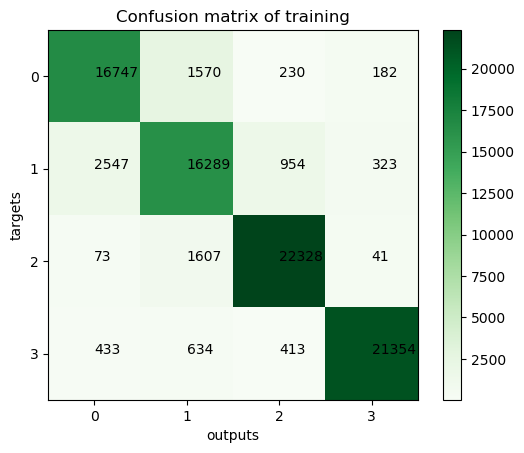
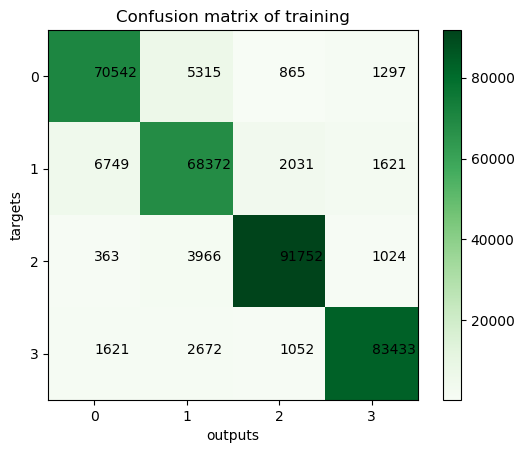


Рисунок 14 – Сonfusion matrix (ViT с Sinkhorn Transformer)

Выводы

В результате данной работы изучены различные механизмы внутреннего внимания для архитектуры ViT. Реализованы и обучены модели для мультиклассовой классификации опухолей головного мозга на основе архитектуры: простой ViT, ViT с применением Linformer, ViT с применением Nystromformer, и ViT с применением Sinkhorn Transformer.

Модель ViT с использованием Nystromformer продемонстрировала наивысшие значения F1-меры среди всех рассмотренных моделей. Это указывает на эффективность метода Nystromformer для задачи классификации опухолей головного мозга на основе снимков МРТ. Модель ViT с Linformer также показала хорошие результаты, но незначительно уступает по сравнению с моделью, использующей Nystromformer. Модель ViT с Sinkhorn Transformer продемонстрировала результаты, сопоставимые с моделью простого ViT. Это говорит о том, что внутреннее внимание данного метода не привнесло существенного улучшения в производительность модели для данной задачи.

В целом, результаты указывают на то, что Nystromformer является самым эффективным методом для улучшения классификации опухолей головного мозга по снимкам МРТ. Однако, дополнительные исследования и сравнения с другими вариантами ViT могут быть полезны для более глубокого понимания и подтверждения этих выводов.

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. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. // Advances in neural information processing systems. 2017. 30.
11. Wang S, Li B Z, Khabsa M, et al. Linformer: Self-attention with linear complexity[J] // arXiv preprint arXiv:2006.04768. 2020. URL: <https://arxiv.org/abs/2006.04768>
12. Xiong Y, Zeng Z, Chakraborty R, et al. Nyströmformer: A nyström-based algorithm for approximating self-attention[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(16): 14138-14148. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/17664>
13. Tay Y, Bahri D, Yang L, et al. Sparse sinkhorn attention[C] // International Conference on Machine Learning. PMLR, 2020: 9438-9447. URL: <https://proceedings.mlr.press/v119/tay20a.html>
14. Функция потерь перекрестной энтропии: Обзор. URL: <https://wandb.ai/wandb_fc/russian/reports/---VmlldzoxNDI4NjAw>