

Нейросетевые модели для анализа МР-изображений аденомы гипофиза

М.Ю. Назарько, К.С. Зайцев, Л.Д. Ковалевич, Н.В. Гарбаева

Аннотация. Целью настоящей работы является исследование методов интеллектуального анализа МРТ хиазмально-селлярной области для диагностики заболеваний гипофиза. Для решения этой задачи предлагается создание нейросетевой архитектуры глубокого обучения на основе фреймворка PyTorch. Для обучения, тестирования и валидации полученной архитектуры были использованы данные из оцифрованных МР-изображений (постконтрастные T1-взвешенные изображения (T1-ВИ) и T2-ВИ) аденом гипофиза с различной степенью инвазии в кавернозные синусы, в количестве 282 единиц, полученные из «Национального медицинского исследовательского центра эндокринологии» Минздрава России («НМИЦ эндокринологии»). Используя полученные данные, были проведены сравнения предложенной архитектуры нейронной сети с известными реализациями архитектур машинного обучения при решении задач классификации по международной шкале Knosp и сегментации аденом гипофиза. Целевой метрикой для оценки качества классификации, была выбрана recall, а для сегментации - Intersection over Union (IoU). Полученные результаты подтвердили эффективность предложенного решения. Итогом работы стало создание эффективной архитектуры системы машинного обучения, применяемой для классификации и сегментации аденом гипофиза.

Ключевые слова — нейронные сети, сегментация, классификация, гипофиз, аденома, МР-изображения

1. ВВЕДЕНИЕ

Аденомы гипофиза являются доброкачественными опухолями, на долю которых приходится 15-20% всех внутричерепных новообразований, с частотой 80-90 случаев на 100 000 населения земли [1]. Выделяют микроаденомы и макроаденомы. Микроаденома доброкачественная опухоль с максимальным диаметром менее 10мм, опухоли с большим диаметром относят к макроаденомам. Магнитно-резонансная томография (МРТ) используется как основной метод диагностики и оценки аденом гипофиза [2]. Для определения агрессивного течения опухолей гипофиза сегодня используют такой признак, как инвазия. Разработаны специальные шкалы, для оценки степени инвазии аденом гипофиза по МРТ, например, шкала Knosp [3].

Для решения задачи выявления аденом можно использовать, как классические методы машинного обучения, так и нейросетевые алгоритмы.

Классические методы основаны на ручном сборе 17 основных признаков. Обычно используются

измеряемые врачами числовые и качественные признаки, такие, например, как размеры гипофиза, его расположение и пр. геометрические характеристики гипофиза, что занимает большую часть времени исследователя [4]. Полученные данные могут обрабатываться классическими алгоритмами машинного обучения, такими, как Decision Tree, Random Forest, Logistic regression, Naïve Bayes, Gradient Boosting Decision Tree, Adaptive Boost, Extreme Gradient Boost. Лучшие результаты среди перечисленных алгоритмов продемонстрировал Random Forest с показателем качества ROC AUC равным 0.78 [4]. Несмотря на то, что эти методы в какой-то степени позволяют врачам самим отбирать признаки, а также хорошо интерпретируются, они все еще глубоко задействуют информационных аналитиков в данной области, и при этом не обладают достаточной точностью.

Также опубликовано исследование применения сверточной нейронной сети (CNN) для диагностики инвазии кавернозного синуса (CSI) при аденоме гипофиза [5]. В проведенное ретроспективное исследование были включены более 370 пациентов, разделенных на инвазивную и неинвазивную группы на основе хирургически подтвержденного CSI. Для анализа использовались T1-взвешенные изображения МРТ хиазмально-селлярной области. Данные были разделены на тренировочные и тестовые выборки для кросс-валидации и обучения нейросетевой модели Resnet50. Однако, хотя оценки Knosp (4 категории) используются в аннотациях, они не являются целью прогнозирования. Модель выполняет простую бинарную классификацию результатов, полученных при МР-сканировании, на инвазивные и неинвазивные группы. Результаты показали, что метрика accuracy предсказаний модели варьировалась от 0.80 до 0.96, в то время как точность системы Knosp была значительно ниже. Этот метод диагностики является промежуточным и не подходит врачам, так как игнорирует шкалу Knosp.

В работе [6] была исследована возможность различения различных патологических типов кистозных поражений в области турецкого седла с помощью предоперационной МРТ. Использовались подходы радиомики для извлечения признаков из контрастных МРТ по 399 пациентам. Выполнено сравнение изображений четырех подтипов кистозных поражений: апоплексия гипофиза,

кистозная аденома гипофиза, киста кармана Ратке и кистозная краниофарингиома. Модель показала среднее значение $AUC = 0.7685$.

В настоящей работе предлагается для решения задачи классификации и сегментации заболеваний гипофиза использовать архитектуру, основанную на работе не с одним МР-срезом, а с серией изображений, как с единым видеорядом, где каждый кадр – это срез, и классифицировать его по шкале Knosp. Также будет использована модель сегментации для выделения зон интереса - срезов с аденомой гипофиза.

2. СОПУТСТВУЮЩИЕ РАБОТЫ

Рассмотрим основные подходы для мультиклассовой классификации и сегментации макроаденом гипофиза на классы Knosp. Их три: UNet-like, Conv-like и Video-like.

а) *UNet-like*. Основная архитектура для сегментации изображений, которая имеет U-образную структуру, состоящую из кодировщика и декодера [7]. Особенностью U-Net является использование "сквозных соединений" (skip connections), которые передают информацию с кодировщика напрямую на соответствующие уровни декодера, что обеспечивает лучшее восстановление деталей на семантической карте. U-Net демонстрирует хорошие результаты в задачах сегментации, особенно при работе с медицинскими изображениями.

Также существуют различные модернизации UNet подобных сетей для повышения точности сегментации. Архитектура Unet++ включает дополнительные пути для передачи информации между энкодером и декодером. Эти дополнительные связи способствуют более глубокому обмену данными на разных уровнях абстракции, что, в свою очередь, может значительно улучшить качество сегментации и повысить устойчивость модели к изменениям в данных. Такая оптимизация позволяет более эффективно использовать информацию на всех этапах обработки, улучшая результаты даже на сложных и разнообразных наборах данных. [8].

В этом подходе следует также рассмотреть архитектуру Feature Pyramid Network (FPN) [9], основанную на идее пирамидального анализа изображений, что позволяет модели учитывать информацию разных масштабов и извлекать более обширный контекст для сегментации объектов различных размеров на изображении. Такая архитектура особенно полезна в задачах компьютерного зрения, требующих учета как мелких деталей, так и глобального контекста.

б) *Conv-like*. Для решения задачи классификации заболеваний гипофиза, по МР-изображениям, применимы и архитектуры, использующие сверточные слои.

Структура сверточных нейронных сетей (CNN) похожа на воронку: всё начинается с общей картины, а потом фокус смещается на детали. Сверточные нейронные сети - это архитектуры,

состоящие из нескольких слоев, которые работают вместе для обучения и распознавания изображений. Они используют различные операции на каждом слое, чтобы извлечь полезные признаки из изображений.

При их использовании создается ядро, проходящее по всему изображению и преобразующее его в меньшую размерность с помощью обученных фильтров. На каждом шаге такого слоя выбирается ядро, с которым для прохода по оставшемуся вектору. Далее производится свертка этого ряда с помощью обученных фильтров. На последнем шаге полученный вектор признаков передается в слои, идущие за свёрточным слоем, для классификации.

Для доменов, основанных на изображениях, такой слой - это эффективное решение, но для классификации целой серии снимков одного пациента такой подход часто малоэффективен, так как не учитывает временную зависимость и информативность снимков, что приводит к ухудшению результатов классификации.

в) *Video-like*. Основной проблемой подхода глубокого обучения для классификации снимков, а в данном случае серии снимков, является то, что при получении данных, эти алгоритмы не выявляют закономерности между соседними снимками. 3D CNN (трехмерные сверточные нейронные сети) расширяют концепцию CNN, добавляя временное измерение к сверткам. Это позволяет сети изучать как пространственные, так и временные особенности видео. Преимущества 3D CNN: способность захватывать движение и временные зависимости, эффективное извлечение пространственно-временных признаков, возможность работы с различными длительностями видео. Но они требуют значительных вычислительных ресурсов и большое количество примеров для обучения.

LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit) - рекуррентные нейронные сети способные запоминать информацию на длительных промежутках времени, что делает их полезными для обработки последовательностей кадров. Их преимущество: эффективная работа с долгосрочными зависимостями, меньшее количество параметров по сравнению с 3D CNN, но они менее эффективны в захвате пространственных особенностей.

Комбинация CNN для извлечения пространственных признаков и LSTM для обработки временных зависимостей обладает преимуществами обоих подходов. Она эффективно захватывает как пространственные, так и временные особенности.

Transformer-based подходы используют архитектуры трансформеров для обработки видео, например, TimeSformer или ViViT [10].

Все эти методы направлены на решение проблемы выявления закономерностей между соседними кадрами и позволяют эффективно работать с видеоданными. Выбор конкретного метода зависит

от специфики задачи, доступных вычислительных ресурсов и требований к модели.

3. ПОДГОТОВКА ДАННЫХ

Для тестирования и сравнения результатов работы построенных архитектур был использован реальный анонимированный набор данных, полученный из «Национального медицинского исследовательского центра эндокринологии» («НМИЦ эндокринологии») Министерства здравоохранения России. Он состоит из МР-изображений хиазмально-селлярной области с разными режимами сканирования (постконтрастные T1-ВИ и T2-ВИ), с комментариями экспертов в данной области. Также для дальнейшего исследования были взяты данные из источника [11]. Эти данные использовались для сегментации аденомы гипофиза и представляют собой МР-изображения в коронарной проекции, постконтрастные T1-ВИ и T2-ВИ, сегментированную маску аденомы гипофиза и внутренних сонных артерий (ВСА). Так как в сегментационной части стоит только задача сегментации аденом, то метки ВСА заменяются нулевыми метками.

В исследование было включено 141 уникальных пациентов/случаев от «НМИЦ эндокринологии» и 13 дополнительных снимков для сегментации из источника [11]. Распределение Knosp grade для левой и правой части снимков пациентов от «НМИЦ эндокринологии» показаны в таблице 1.

Таблица 1. Распределение Knosp grade в левой и правой части гипофиза

Сторона	Knosp 0	Knosp I	Knosp II	Knosp IIIa	Knosp IIIb	Knosp IV
Лев. (S)	51	33	16	12	5	24
Прав.(D)	59	25	21	11	2	23

Для каждого пациента от «НМИЦ эндокринологии» доступна следующая информация:

- изображения хиазмально-селлярной области, то есть МР-изображения в коронарной, аксиальной и сагиттальной проекциях области турецкого седла в режимах постконтрастные T1-ВИ и T2-ВИ.

- аннотации, то есть файл с перечнем оценок Knosp для отдельных срезов хиазмально-селлярной области, полученных при сканировании отдельно для левой и правой стороны.

Маска сегментации - это созданное вручную изображение одинаковых размеров, которое присваивает каждому пикселю один из интересующих объектов (опухоль) или фон с четкой меткой (номером). Маски сегментации есть для всех срезов: коронарного, аксиального, сагиттального.

Для семантической сегментации аденом гипофиза набор данных при обучении и тестировании состоял из коронарных постконтрастных T1-ВИ и T2-ВИ (228 ед.): обучающая выборка — 182 ед. или 1280 срезов (80%), тестовая выборка — 46 ед. или 342 срезов (20%).

Для классификации - обучающая выборка — 230 ед. (80%), тестовая выборка — 52 ед. (20%).

Для локализации и предварительной обработки исходные изображения преобразовывались из формата tiff в формат PNG. Была выполнена нормализация интенсивности изображений, чтобы обеспечить среднее значение интенсивности пикселя для каждого субъекта, равное 0, и стандартное отклонение, равное 1. Так как при МР-сканировании хиазмально-селлярной области обычно фокус оказывается на области турецкого седла, мы грубо локализовали гипофиз, обрезав среднюю часть изображения, а затем изменили размер изображений и масок.

Также для классификации и сегментации снимков проводилась аугментация обучающего набора данных путем аффинных преобразований, отдаления и приближения снимка, добавления вырезов на снимке.

4. МОДЕЛИ И МЕТОДЫ

В представленной работе предлагается использовать подход CNN - Join-net для классификации заболеваний гипофиза и UNet++ с Efficient Net V6 для сегментации.

В серии МР-изображений не все серии равнозначны, поэтому нецелесообразно утверждать, что каждый скан содержит необходимую информацию для классификации. Хотя можно использовать лишь наиболее информативное изображение с наибольшей маской, такой подход утратит пространственную информацию об опухоли. Вместо этого предлагается рассматривать серию изображений. Так как значимая информация сосредоточена неравномерно, как видно на рисунке 1, то нужно выбирать только тот промежуток изображений, где плотность распределения пикселей максимальная.

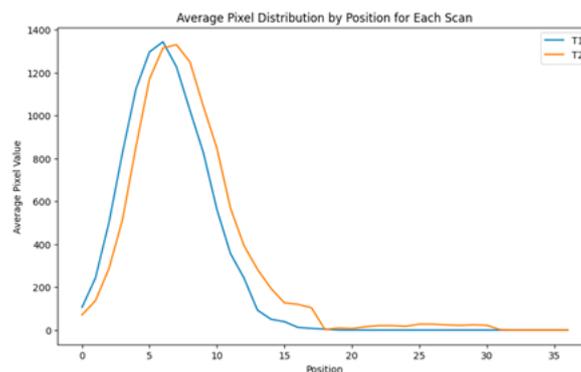


Рисунок 1 - Распределение пикселей в зависимости от среза

Для подобного подхода предлагается архитектура сети представленная на рисунке 2. В качестве энкодера срезов предлагается любая предобученная Conv-like архитектура, а для сети объединения полученных векторов признаков используется join-net: AVG, RNN/LSTM, SUM, ATTENTION. Далее выходы join-net слоя поступают на полносвязные слои, которые, используя полную активацию, пропускают через себя данные и затем на выходе взвешивают их с помощью функции softmax.

Результатом работы становится вероятностная классификация. В этом алгоритме подразумевается использование кросс энтропии [12].

$$Hx = - \sum p_i * \log_2 p_i \quad (1)$$

Такой подход зарекомендовал себя достаточно надежным при классификации видеоряда состоящего из 10-20 кадров, которыми и являются МР-изображения заболеваний гипофиза.

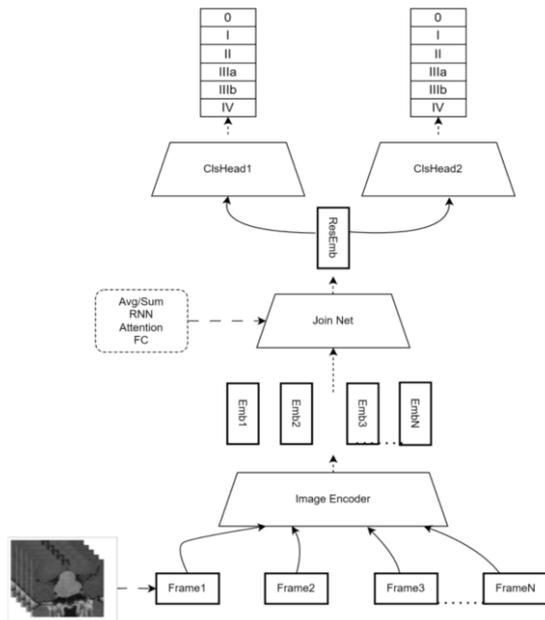


Рисунок 2 - архитектура сети классификации

В качестве Image Encoder использовалась модифицированная нейронная сеть с архитектурой ResNet-50+attention. В роли JoinNet использовались как LSTM, так и AVG. Заключительная часть архитектуры нейронной сети состоит из двух классификационных голов. В каждой голове последовательности полносвязных слоев. Эти слои имеют следующую структуру: 512, 256, 128 и 6 нейронов соответственно.

В качестве функции активации для этих слоев была выбрана стандартная relu. Выходной слой сети содержит 6 нейронов, соответствующих шести классам Knosp grade. Для этого слоя используется функция активации softmax, которая преобразует выходные значения в вероятностное распределение по классам. Такой подход необходим, поскольку в качестве функции потерь используется кросс-энтропия, которая ожидает на входе вероятностные оценки принадлежности к каждому классу. Это позволяет эффективно оптимизировать веса сети в процессе обучения для решения задачи многоклассовой классификации.

В качестве сети сегментации были проанализированы различные SOTA архитектуры: Unet, Unet++, DeepLabV3, DeeplabV3+, PANet, PSPNet, FPN, MANet, Linknet. В качестве Loss функции линейная композиция кросс энтропии и Tversky Loss.

При обучении и тестировании нейронных сетей критически важно применять соответствующие

метрики, которые могут точно оценить качество нейросетевых моделей и способствуют сравнению разных моделей.

Введем обозначения основных характеристик, получаемых при применении классификатора:

TP (True Positive) - количество верно предсказанных положительных объектов;

TN (True Negative) - количество верно предсказанных отрицательных объектов;

FP (False Positive) - количество отрицательных объектов, предсказанных как положительные (ошибки первого рода);

FN (False Negative) - количество положительных объектов, предсказанных как отрицательные (ошибки второго рода).

Recall (чувствительность, полнота) представляет собой долю верно предсказанных положительных объектов относительно всех истинно положительных объектов. Эта метрика измеряет способность модели обнаруживать положительные объекты. Recall рассчитывается по формуле:

$$recall = \frac{TP}{TP + FN} \quad (2)$$

Precision (точность) представляет собой долю верно предсказанных положительных объектов относительно всех объектов, предсказанных как положительные. Эта метрика измеряет способность модели избегать ложных срабатываний, т.е. ошибок первого рода. Precision рассчитывается по формуле:

$$precision = \frac{TP}{TP + FP} \quad (3)$$

F1-score представляет собой гармоническое среднее между Recall и Precision. Эта метрика используется для оценки агрегированной производительности классификатора, и она стремится к балансу между точностью и полнотой. Таким образом, последовательности с несбалансированными классами могут быть лучше предсказаны. F1-score рассчитывается по формуле:

$$f1\ score = \frac{2 * recall * precision}{recall + precision} \quad (4)$$

А так же метрика ассигасу. рассчитывается она по формуле:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Достоинство ассигасу заключается в том, что она легко интерпретируется и понятна. Она представляет собой простую долю правильных предсказаний модели в общем числе предсказаний.

Например, в медицинской диагностике часто важнее Recall, так как ложные отрицательные результаты могут привести к существенным последствиям, которые могут быть гораздо более важными, чем ложные положительные результаты.

В задаче сегментации изображений наиболее значимыми метриками являются коэффициенты сходства и ошибок, которые могут количественно оценивать степень соответствия между предсказанными и истинными областями на изображении. Ниже приведены некоторые часто используемые метрики сегментации.

Коэффициент Жаккара (Jaccard index) или IoU (Intersection over Union) представляет собой отношение пересечения между предсказанной и истинной областями к объединению обеих областей. Значения этой метрики варьируются в интервале от 0 до 1, где 1 - идеальное совпадение, а 0 - полное расхождение.

$$IoU = \frac{\text{intersection of predicted and ground truth}}{\text{union of predicted and ground truth}} \quad (6)$$

Коэффициент Dice (Dice coefficient) аналогичен Jaccard index, но уделяет больше внимания различиям между двумя областями. Значения этой метрики также варьируются в интервале от 0 до 1.

$$DC = \frac{2 * \text{intersection of predicted and ground truth}}{\text{size of predicted} + \text{size of ground truth}} \quad (7)$$

5. РЕЗУЛЬТАТЫ СРАВНЕНИЯ МОДЕЛЕЙ

5.1. Модели сегментации

Для моделей сегментации измерялись коэффициенты Дайса и Жаккара (IoU). В качестве обучающего и валидационного наборов данных, для предсказания использовались только репрезентативные коронарные T1+C и T2 снимки в количестве 192 штук. Из них 150 снимков использовались для обучения и 42 для валидации. Сравнение проводилось на 60 эпохах с backbone ResNet-18. Результаты обучения вышеуказанных моделей представлены в таблице 2.

Таблица 2. Результаты сравнения различных архитектур сегментации.

Модель	Loss val	DICE val	IoU val
Unet	0.455	0.423	0.313
Unet++	0.430	0.479	0.355
DeepLabV3	0.446	0.441	0.326
DeeplabV3+	0.431	0.484	0.356
PANet	0.617	0.344	0.226
PSPNet	0.568	0.311	0.228
FPN	0.486	0.431	0.315
Linknet	0.551	0.407	0.297
MANet	0.533	0.321	0.237

Анализируя полученные результаты, можно сделать вывод, что качество всех моделей зависит от небольшого размера обучающего набора данных и низкой вариабельности уникальных случаев. Фаворитами для дальнейшего исследования являются UNet++ , DeepLabV3+ и FPN.

На архитектуре Unet++ были проведены эксперименты с различными базовыми сетями (backbone) для извлечения признаков из поступающего на вход изображения. Для определения лучшего для данной задачи. Всего снимков - 284, из которых 242 использовались для

обучения и 42 - для валидации. Размер пакета был равен 64. Сравнение проводилось на 50 эпохах. Результаты обучения моделей представлены в таблице 3.

Таблица 3. Результаты сравнения различных backbone для сегментации.

Модель	b1	b6	res50	rxt50	effv2_xl	mb one s4
Params,M	4	40	23	22	208.12	13.6
val IoU	0.40	0.42	0.393	0.385	0.404	0.45

Обучение модели mobileone s4 не является стабильным, в отличие от модели efficient net b1 и b6, метрики которых не флуктуируют и монотонно растут на обучении и валидации. Были обучены сети сегментации с backbone efficient b6 и архитектурами декодера UNet++ и FPN. Обе сети обучались на 200 эпохах с размером пакета 64. В обучении было представлено 228 уникальных пациентов, из них 46 на валидацию, и 182 на обучение. Всего снимков с маской на обучение 1280, а на валидацию 342. Оптимизатор AdamW с lr=1e-4 и wd=1e-6. Так же для повышения эффективности обучения использовался отжиг Cosine Annealing, температура которого линейно понижалась в зависимости от эпохи обучения. Результаты обучения вышеуказанных моделей представлены в таблице 4.

Таблица 4. Результаты обучения лучших сетей сегментации

Модель	Loss val	DICE val	IoU val
Unet++	0.190	0.8442	0.7610
FPN	0.193	0.839	0.7499

5.2. Модели классификации

Для дальнейшей работы с классификацией МР-изображений были проанализированы различные подходы к формированию обучающей выборки. Для упрощения обучения мы классифицировали только одну сторону хиазмально-селлярной области. Основные действия в экспериментах варьировались и были направлены на использование всех изображений или только размеченных врачом (с маской), на классификацию одной стороны или максимального класса из двух сторон, инвертирования изображений по вертикали, закрепление класса или нет, так как некоторых классов в наборе данных крайне мало.

По результатам серии экспериментов было принято решение объединить классы 3a и 3b, инвертировать изображения по вертикали, а также брать только те изображения, где есть зона интереса. С предложенным подходом датасет формировался из 10 значимых подряд идущих изображений, и обрабатывался моделью с предложенной выше архитектурой.

Сравнение различных архитектур проводилось на 50 эпохах. В обучении было представлено 282

уникальных пациента из них 52 на валидацию и 230 на обучение. Оптимизатор AdamW с $lr=1e-4$ и $wd=1e-6$. Так же для повышения эффективности обучения использовался отжиг Cosine Annealing, температура которого линейно понижалась в зависимости от эпохи обучения. Результаты сравнения обученных моделей представлены в таблице 5.

Видно, что предлагаемая архитектура JoinNet превосходит как классические архитектуры сверточных нейронных сетей, по всем целевым метрикам, так и 3D CNN, но эти результаты получены на моделях, которые обучались на сравнительно небольшом количестве уникальных пациентов.

Результаты, полученные в настоящей работе, можно было бы сравнить с результатами, полученными в предыдущие годы на ранее используемых моделях для мульти классовой классификации заболеваний гипофиза в целом.

Таблица 5. Результаты сравнения различных архитектур для классификации.

Метрика	recall	f1	accuracy	precision
Модель				
ResNet18JMean	0.60	0.66	0.60	0.73
ResNet50JMean	0.75	0.73	0.72	0.75
ResNet101JMean	0.64	0.68	0.64	0.74
ResNet50JMean+LSTM	0.50	0.56	0.50	0.67
ResNet50JMean+2 ch. sagittal+cor)	0.53	0.54	0.53	0.56
ResNet50	0.30	0.27	0.484	0.24
3DResNet18	0.50	0.54	0.49	0.60

Однако, учитывая современные технологические достижения, такое сравнение нецелесообразно из-за того, что более ранние модели машинного обучения уступают глубоким нейронным сетям в способности к обобщению данных и эффективности решения поставленных задач работы с медицинскими изображениями.

6. ЗАКЛЮЧЕНИЕ

В настоящей работе рассмотрены подходы для решения задач сегментации и классификации заболеваний гипофиза по МР-изображениям, основанные на архитектурах глубокого обучения, пришедших на смену более простым алгоритмам классического машинного обучения.

Для классификации аденом гипофиза была предложена архитектура, позволяющая работать со срезами МР-сканов, улавливая пространственную взаимосвязь. При этом эта архитектура не обладает большим количеством параметров и способна обучаться на сравнительно небольших наборах данных.

В работе был проведен сравнительный анализ для получения рациональных нейросетевых архитектур сегментации и классификации МР-изображений заболеваний гипофиза.

Для проведения обучения, тестирования и валидации был сформирован набор МР-изображений, полученных от «НМИЦ эндокринологии» Министерства здравоохранения Российской Федерации и нескольких размеченных открытых снимках.

Для создания архитектуры машинного обучения были выбраны такие средства, как PyTorch, которые подробно задокументированы, и с помощью которых созданы многочисленные проекты, использующие архитектуры глубокого обучения.

Во время создания собственной глубокой архитектуры были подобраны слои и их гиперпараметры для получения наилучших результатов на основе целевой метрики recall для классификации и IoU для сегментации.

После анализа результатов экспериментов, можно утверждать, что при классификации МР-изображений с использованием классических CNN архитектур можно добиться точности не более 0.4 на реальных данных. Также стоит сказать, что при небольшом обучающем наборе данных LSTM слой плохо обучается и выдает результаты на порядок хуже чем простое агрегирование по среднему всех эмбедингов. 3DCNN работает лучше Conv+LSTM, но все еще проигрывает предложенному подходу.

Подводя итог, можно сказать, что для решения задачи классификации и сегментации аденом гипофиза предложенные архитектуры нейронных сетей подходят наилучшим образом по критерию «оценка качества» при ориентации на разные метрики качества. Предложенная модель классификации превзошла своих конкурентов за счет меньшей вычислительной сложности, но при этом с возможностью сохранения пространственной структуры, учитывающей разные части МР-изображения.

БЛАГОДАРНОСТИ

Авторы выражают благодарность Высшей инженеринговой школе НИЯУ МИФИ за помощь в возможности опубликовать результаты выполненной работы.

БИБЛИОГРАФИЯ

- [1] Emmett, S.R., Hill, N., Dajas-Bailador, F. Endocrinology, 2019. DOI: 10.1093/oso/9780199694938.003.0012. Дата обращения: 20.01.2023. URL: <http://dx.doi.org/10.1093/oso/9780199694938.003.0012>.
- [2] Overview of the pituitary gland, 2023. [Электронный ресурс]. URL: <https://www.msmanuals.com/home/hormonal-and-metabolic-disorders/pituitary-gland-disorders/overview-of-the-pituitary-gland> (дата обращения 20.01.2024).
- [3] Anatomy Of The Nervous System, 2017. [Электронный ресурс]. URL: https://www.sharetechnote.com/html/Neuroscience/Neuroscience_BrainStructure.html (дата обращения 20.01.2024).
- [4] Liu, Y., Liu, X., Hong, X., Liu, P., Bao, X., Yao, Y., Xing, B., Li, Y., Huang, Y., Zhu, H., Lu, L., Wang, R., Feng, M. Prediction of Recurrence after Transsphenoidal Surgery for Cushing's Disease: The Use of Machine Learning Algorithms. Neuroendocrinology; 2019; 108: 201-210. DOI: 10.1159/000496753.
- [5] Fang Y, Wang H, Feng M, Chen H, Zhang W, Wei L, Pei Z, Wang R, Wang S. Application of Convolutional Neural Network in the Diagnosis of Cavernous Sinus Invasion in Pituitary Adenoma. Front Oncol. 2022 Apr 14;12:835047. doi: 10.3389/fonc.2022.835047. PMID: 35494041; PMCID: PMC9047893.

- [6] Jiang C, Zhang W, Wang H, Jiao Y, Fang Y, Feng F, Feng M, Wang R. Machine Learning Approaches to Differentiate Sellar-Suprasellar Cystic Lesions on Magnetic Resonance Imaging. *Bioengineering* (Basel). 2023 Nov 8;10(11):1295. doi: 10.3390/bioengineering10111295. PMID: 38002419; PMCID: PMC10669327.
- [7] Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241), October 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
- [8] Zhou Z. et al. Unet++: A nested u-net architecture for medical image segmentation // *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4.* – Springer International Publishing, 2018. – С. 3-11.
- [9] Kirillov A. et al. Panoptic feature pyramid networks // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* – 2019. – С. 6399-6408.
- [10] Видеоаналитика: Разбор VideoMAE, ViViT и TimeSFormer – *Habr*, 2024 <https://habr.com/ru/articles/827474/>
- [11] Černý, M., Kybic, J., Májovský, M., Sedlák, V., Pirgl, K., Misiorzová, E., Lipina, R., Netuka, D. Fully automated imaging protocol independent system for pituitary adenoma segmentation: a convolutional neural network-based model on sparsely annotated MRI. *Neurosurg Rev*, 46(1):116, 2023. DOI: 10.1007/s10143-023-02014-3. PMID: 37162632.00166-4
- [12] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951

Статья получена 2 октября 2024.

Назарько Михаил Юрьевич, Национальный Исследовательский Ядерный Университет МИФИ, бакалавр, mikhail.nazarrko@gmail.com

Зайцев Константин Сергеевич, Национальный Исследовательский Ядерный Университет МИФИ, профессор, KSZajtsev@mephi.ru

Ковалевич Лилия Дмитриевна, ФГБУ «НМИЦ эндокринологии» Минздрава России, врач-рентгенолог референс-центра лучевых методов диагностики, liliyakovalevich@gmail.com

Тарбаева Наталья Викторовна, ФГБУ «НМИЦ эндокринологии» Минздрава России, руководитель референс-центра лучевых методов диагностики, кандидат медицинских наук, Ntarbaeva@inbox.ru

Neural network models for analyzing MRI images of pituitary adenoma

M.Yu. Nazarko, K.S. Zaytsev, L.D. Kovalevich, N.V. Tarbaeva

Abstract. The purpose of this work is to study methods for intelligent analysis of MRI images of the sellar region for diagnosing pituitary diseases. To solve this problem, it is proposed to create a deep learning neural network architecture based on the PyTorch framework. To train, test and validate the resulting architecture, data from digitized MRI images of the brain with varying degrees of invasion, a contrast-enhanced T1-weighted (T1-CE), a T2-weighted (T2) in the amount of 282 units, obtained from the National Medical Research Center of Endocrinology of the Ministry of Health of Russia (NMRC of Endocrinology) were used. Using the data obtained, comparisons were made of the proposed neural network architecture with known implementations of machine learning architectures when solving problems of classification according to the international Knosp scale and segmentation of pituitary adenomas. Recall was chosen as the target metric for assessing the quality of classification, and Intersection over Union (IoU) was chosen for segmentation. The results obtained confirmed the effectiveness of the proposed solution. The result of the work was the creation of an effective architecture for a machine learning system used for classification and segmentation of pituitary adenomas.

Keywords — neural networks, segmentation, classification, pituitary gland, adenoma, MRI images

REFERENCES

- [1] Emmett, S.R., Hill, N., Dajas-Bailador, F. Endocrinology, 2019. DOI: 10.1093/oso/9780199694938.003.0012. Дата обращения: 20.01.2023. URL: <http://dx.doi.org/10.1093/oso/9780199694938.003.0012>.
- [2] Overview of the pituitary gland, 2023. [Электронный ресурс]. URL: <https://www.msmanuals.com/home/hormonal-and-metabolic-disorders/pituitary-gland-disorders/overview-of-the-pituitary-gland> (дата обращения 20.01.2024).
- [3] Anatomy Of The Nervous System, 2017. [Электронный ресурс]. URL: https://www.sharetechnote.com/html/Neuroscience/Neuroscience_BrainStructure.html (дата обращения 20.01.2024).
- [4] Liu, Y., Liu, X., Hong, X., Liu, P., Bao, X., Yao, Y., Xing, B., Li, Y., Huang, Y., Zhu, H., Lu, L., Wang, R., Feng, M. Prediction of Recurrence after Transsphenoidal Surgery for Cushing's Disease: The Use of Machine Learning Algorithms. Neuroendocrinology; 2019; 108: 201-210. DOI: 10.1159/000496753.
- [5] Fang Y, Wang H, Feng M, Chen H, Zhang W, Wei L, Pei Z, Wang R, Wang S. Application of Convolutional Neural Network in the Diagnosis of Cavernous Sinus Invasion in Pituitary Adenoma. Front Oncol. 2022 Apr 14;12:835047. doi: 10.3389/fonc.2022.835047. PMID: 35494041; PMCID: PMC9047893.
- [6] Jiang C, Zhang W, Wang H, Jiao Y, Fang Y, Feng F, Feng M, Wang R. Machine Learning Approaches to Differentiate Sellar-Suprasellar Cystic Lesions on Magnetic Resonance Imaging. Bioengineering (Basel). 2023 Nov 8;10(11):1295. doi: 10.3390/bioengineering10111295. PMID: 38002419; PMCID: PMC10669327.
- [7] Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241), October 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
- [8] Zhou Z. et al. U-net++: A nested u-net architecture for medical image segmentation //Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. – Springer International Publishing, 2018. – C. 3-11.
- [9] Kirillov A. et al. Panoptic feature pyramid networks //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – C. 6399-6408.
- [10] Video analytics: Analysis of VideoMAE, ViViT and TimeSFormer – Habr, 2024 <https://habr.com/ru/articles/827474/>
- [11] Černý, M., Kybic, J., Májovský, M., Sedlák, V., Pírgl, K., Misiorzová, E., Lipina, R., Netuka, D. Fully automated imaging protocol independent system for pituitary adenoma segmentation: a convolutional neural network-based model on sparsely annotated MRI. Neurosurg Rev, 46(1):116, 2023. DOI: 10.1007/s10143-023-02014-3. PMID: 37162632.00166-4
- [12] S. Kullback and R. A. Leibler, "On information and sufficiency," The annals of mathematical statistics, vol. 22, no. 1, pp. 79-86, 1951