

Первые результаты неинвазивного типирования глиом головного мозга по данным магнитно-резонансной томографии с помощью машинного обучения

Г.В. Данилов, И.Н. Пронин, В.В. Королев, Н.Г. Малоян, Д.Ф. Саада,
Е.А. Ильюшин, М.А. Шифрин, Т.А. Ишанкулов, Р.М. Афандиев, А.М. Шевченко,
Т.А. Конакова, Т.В. Цуканова, С.В. Шугай, А.А. Потапов

*Лаборатория биомедицинской информатики и искусственного интеллекта,
Национальный медицинский исследовательский центр нейрохирургии имени академика Н.Н. Бурденко,
Москва, Российская Федерация*

Автор для переписки: Г.В. Данилов, ФГАУ «НМИЦ нейрохирургии им. ак. Н.Н. Бурденко»
Минздрава России,
ул. 4-я Тверская-Ямская 16, Москва, 125047, Российская Федерация;
эл. почта: gdanilov@nsi.ru

Резюме

Глиомы - наиболее распространенные нейроэпителиальные опухоли головного мозга, отличающиеся по структуре и биологическим свойствам. Современная классификация опухолей нервной системы и подходы к их лечению основаны на понимании тканевых и молекулярных особенностей конкретного новообразования. Однако гистологическое и молекулярно-генетическое типирование опухолей сегодня может быть проведено только путем инвазивных исследований. В этой связи интересно совершенствование неинвазивных методов дооперационной диагностики в нейроонкологии. Одно из перспективных направлений - применение технологий искусственного интеллекта к данным нейровизуализации для выявления диагностически значимых паттернов, связанных с гистологическим и молекулярным профилями опухоли и недоступных в явном виде лучевому диагносту.

Цель исследования: Наше пилотное исследование было направлено на оценку диагностической точности методов глубокого обучения (ГО) для типирования глиом в соответствии с классификацией ВОЗ 2007 года по данным предоперационной магнитно-резонансной томографии (МРТ).

Материалы и методы:

В исследование включали МР-томограммы пациентов с глиальными опухолями головного мозга, проходивших нейрохирургическое лечение в НМИЦ нейрохирургии им. Н.Н. Бурденко, которым на дооперационном этапе госпитализации была выполнена МРТ с контрастным усилением. Данные МРТ в формате плоских срезов (2D) и трехмерных объектов (3D) были использованы для обучения искусственных нейронных сетей с двумя архитектурами (Resnest200e и DenseNet соответственно) классифицировать опухоли по 4 категориям (I-IV степеням злокачественности по ВОЗ). Обучение проводили на 80% исследований, отобранных случайным образом. Метрики качества классификации оценивали на оставшихся 20% исследований (валидационной и тестовой выборках).

Результаты:

В анализ были включены 707 МРТ-исследований в модальности T1 с контрастным усилением. Подход 3D-классификации с помощью модели DenseNet показал наилучший результат предсказания степени злокачественности опухоли по ВОЗ (точность = 83%, ROC AUC = 0,95), что согласуется с результатами других авторов, использовавших иные методики.

Выводы:

Первые результаты наших исследований подтвердили принципиальную возможность градации аксиальных T1-изображений с контрастным усилением по классам ВОЗ 2007 г. с помощью моделей глубокого обучения.

Ключевые слова: нейроонкология, нейрохирургия, искусственный интеллект, глиома, машинное обучение, глубокое обучение, биомедицинская информатика

Введение

Глиомы являются наиболее распространенными нейроэпителиальными опухолями головного мозга, различающимися по гистологическому, молекулярному типу и прогнозу (1). Широко используемая классификация опухолей центральной нервной системы, разработанная Всемирной организацией здравоохранения (ВОЗ) в 2007 году, была основана на гистогенезе и микроскопическом сходстве новообразований (2). Согласно классификации ВОЗ 2007 года глиомы подразделяли на четыре степени злокачественности (WHO Grade) (2). Эксперты ВОЗ применяют I-II категории к глиомам низкой степени злокачественности (Low-Grade Glioma, LGG), а III-IV — к глиомам высокой степени злокачественности (High-Grade Glioma, HGG). Эта классификация была пересмотрена в 2016 году: новая система оказалась более сложной и основывалась на молекулярных подтипах опухолей, определяющих тактику и результаты лечения (3). Обновленная версия классификации глиом была выпущена в 2021 году (4). Несмотря на обоснованность последней классификации, ее внедрение во всем мире является ресурсозатратным, требующим дорогостоящих инфраструктуры и исследований. Именно по этой причине старая классификация ВОЗ все еще активно используется, особенно в странах с ограниченными ресурсами.

Появление новых неинвазивных методов гистологической и молекулярной диагностики новообразований нервной системы может революционно изменить эффективность и безопасность медицинской помощи пациентам с этой патологией и радикально сократить расходы системы здравоохранения. Одно из перспективных направлений - это применение технологий искусственного интеллекта к данным нейровизуализации для выявления диагностически значимых паттернов, связанных с гистологическим и молекулярным профилями опухоли и недоступных в явном виде лучевому диагносту. Потенциал этих технологий связывают с возможностями распознавать сложные закономерности в данных больших объемов.

Текущие успехи технологий искусственного интеллекта в нейровизуализации открывают новые перспективы в развитии неинвазивной и персонализированной диагностики (5,6). Активно развиваются методы радиомики — извлечения большого количества признаков из медицинских изображений (7). Эти признаки могут содержать информацию для описания опухолей и структур мозга, которые не видны невооруженным глазом (8). Предполагается, что правильное представление и анализ изображения с нейровизуализационными характеристиками поможет различать типы опухолей и соотносить их с клиническими проявлениями заболевания, прогнозом и наиболее эффективным лечением. Технологии, оценивающие взаимосвязь между особенностями визуализации опухолей и экспрессией генов, получили название «радиогеномики». Эти методы направлены на создание визуализационных биомаркеров, которые могут идентифицировать генетические признаки заболевания без проведения биопсии.

Таким образом, исследование данных нейровизуализации в нейроонкологии, в первую очередь — магнитно-резонансной томографии (МРТ), с применением методов радиомики, радиогеномики и искусственного интеллекта является актуальным трендом. Тем не менее, большинство таких исследований ограничены небольшим объемом анализируемых данных, что препятствует применению наиболее эффективных технологий

анализа и машинного обучения, в том числе – глубокого, то есть с использованием искусственных нейронных сетей.

ФГАУ «НМИЦ нейрохирургии им. ак. Н.Н. Бурденко» Минздрава России – головное нейрохирургическое учреждение Российской Федерации, один из крупнейших нейрохирургических центров в мире. За период более 20 лет использования электронных медицинских карт и PACS-систем (*англ. Picture Archiving and Communication System* - система передачи и архивации изображений), НМИЦ нейрохирургии накопил большой архив изображений нервной системы, которые можно использовать в исследованиях с применением радиомики и искусственного интеллекта. Целью данного исследования было оценить диагностическую точность глубокого обучения (ГО) в типировании глиом по степени злокачественности ВОЗ (I-IV) по данным предоперационной МРТ, выполненной в НМИЦ нейрохирургии им. Н.Н. Бурденко.

Материалы и методы

Критериями включения в исследование были:

- госпитализация в ФГАУ «НМИЦ нейрохирургии им. ак. Н.Н. Бурденко» Минздрава России в период с 2009 г. по 2018 г.;
- диагноз глиальной опухоли, верифицированный с помощью гистологического исследования;
- нейрохирургическое лечение в ФГАУ «НМИЦ нейрохирургии им. ак. Н.Н. Бурденко» Минздрава России;
- предоперационная МРТ головного мозга с контрастным усилением, выполненная в ФГАУ «НМИЦ нейрохирургии им. ак. Н.Н. Бурденко» Минздрава России.

Данные МРТ для каждого пациента получали в виде серий аксиальных двумерных изображений в формате DICOM. Отбирали исследования в режиме T1 с контрастным усилением.

В исследовании решали техническую задачу классификации данных МРТ по четырем градациям степени злокачественности ВОЗ (I-IV). Использовали два базовых подхода. Первый подход – 3D-классификация – обучение математической модели определять степень злокачественности опухоли по набору всех срезов головного мозга из одного исследования – как единому объекту. В этом случае метку класса (степень от I до IV) присваивали целому МР-исследованию. Второй подход – «2D-классификация» - это «посрезовая» классификация, для которой в качестве объекта анализа использовали один аксиальный срез. Таким образом, класс (I-IV) присваивали каждому срезу.

В случае 3D-классификации мы применяли следующие методы предобработки данных для их нормализации и унификации. DICOM-изображения представляли в виде массивов вещественных чисел. Каждое число отражало интенсивность соответствующего пикселя в исходном снимке. Данные нормализовали путем вычитания среднего по массиву и деления на стандартное отклонение:

$$arr = \frac{arr - arr.mean()}{arr.std()}$$

Разрешение каждого изображения устанавливали равным 512 x 512. Для всех пациентов число срезов приводили к 32, используя алгоритм интерполяции. Таким образом, в исследования с меньшим количеством срезов добавляли дополнительные изображения, в то время как для пациентов с избыточным числом срезов некоторые соседние срезы усредняли.

В случае 2D-классификации предварительная подготовка изображений отличалась. Данные каждого среза также записывали в числовой массив. Исходное одноканальное

изображение трансформировали в трехканальное с дублированием одних и тех же значений по всем каналам. Такой подход не нарушал структуру изображения, но был необходим для форматирования данных для подачи на вход нейронной сети. Изображения приводили к одному размеру 512 x 512 и нормализовали с помощью среднего = (0,485; 0,456; 0,406) и стандартного отклонения = (0,229; 0,224; 0,225). Эти значения были получены из открытого набора данных ImageNet. Мы использовали несколько методов аугментации с вероятностью 0,2: поворот на угол ≤ 10 градусов, масштабирование на 0,9-1,1 от исходного изображения и зеркальное отображение изображения МР-снимков для искусственного увеличения размера обучающей выборки.

И в 3D, и в 2D случаях исходные данные делили на обучающую, валидационную и тестовую выборки со стратификацией по целевой переменной. В качестве базовой целевой переменной использовали категориальную величину – степень злокачественности опухоли, принимавшую значения (метки классов) «I», «II», «III» и «IV». Обучающую выборку формировали случайным образом так, чтобы она составляла 80% от исходной. Валидационная и тестовая выборки составили по 10% от общего набора данных.

Основной моделью глубокого обучения на 3D-данных стала искусственная нейронная сеть с архитектурой DenseNet, адаптированная для работы с 3D-изображениями. В двумерном случае использовали архитектуру нейронной сети Resnet200e. Для всех моделей использовали алгоритм оптимизации Adam со скоростью обучения $1e-4$. В качестве функции потерь была выбрана кросс-энтропия, реализованная по следующей формуле, где P - распределение истинного ответа, Q – распределение прогнозируемых моделью вероятностей, x – МР-изображение:

$$H(P, Q) = - \sum_x P(x) \log Q(x)$$

Обучение проводили для классификации глиальных опухолей как по всем 4 типам злокачественности, так и в бинарном случае, когда для каждого объекта определяли только низкую (I и II) и высокую (III и IV) степень злокачественности.

Процесс обучения реализовывали на суперкомпьютере ФГАУ «НМИЦ нейрохирургии им. ак. Н.Н. Бурденко» Минздрава России с использованием 8 графических процессоров NVIDIA A100-SXM4-40GB. Вычисления были распараллелены. Программный код для процедуры машинного обучения был написан с помощью языка программирования Python (version 3.8) с использованием библиотек PyTorch (версия 1.7).

Результаты

Первичный набор данных для экспериментов по машинному обучению был собран за период с 2009 г. по 2018 г. из PACS ФГАУ «НМИЦ нейрохирургии им. ак. Н.Н. Бурденко» для 1280 пациентов с верифицированными глиомами головного мозга, которым на дооперационном этапе выполнили МРТ головного мозга. МР-изображения в формате DICOM были получены с разных томографов. Наблюдали различия в количестве срезов (от 18 до 30), режимах сканирования, разрешении изображений для каждого исследования. Из всего массива данных были выбраны аксиальные изображения в последовательности T1 с контрастным усилением ($n = 707$), подавляющее большинство которых было создано на одном и том же устройстве.

Результаты бинарной классификации МР-изображений на опухоли низкой и высокой степени злокачественности с помощью 3D- и 2D-подходов представлены в Таблице 1.

Таблица 1. Метрики качества бинарной классификации МР-изображений с помощью 3D- и 2D-подходов.

Метрика качества	3D (DenseNet)	2D (Resnest200e)
Точность	67%	61%
ROC AUC	76%	73%
Чувствительность	58%	44%
Специфичность	78%	81%

В экспериментах с четырехуровневой целевой переменной удалось достигнуть более высоких показателей качества при использовании 3D-подхода (см. Таблицу 2). Общую точность рассчитывали как отношение числа верно определенных классов к общему числу классифицируемых объектов (исследований МРТ - в случае 3D и единичных срезов - в случае 2D). ROC AUC вычисляли как среднюю ROC AUC при бинарной детекции каждого класса по принципу «один против всех».

Таблица 2. Обобщенные метрики качества мультиклассовой классификации МР-изображений с помощью 3D- и 2D-подходов.

Метрика качества	3D (DenseNet)	2D (Resnest200e)
Точность	83%	50%
ROC AUC	95%	72%

Метрики классификации по каждому из 4 классов для 3D и 2D подходов показаны в таблицах 3 и 4 соответственно.

Таблица 3. Метрики качества мультиклассовой классификации МР-изображений с помощью 3D-подхода.

Степень злокачественности глиомы (ВОЗ)	Положительное прогностическое значение	Чувствительность	F1
I	0,79	1,00	0,88
II	0,97	0,63	0,76
III	0,50	1,00	0,67
IV	0,95	0,85	0,90

Таблица 4. Метрики качества мультиклассовой классификации МР-изображений с помощью 2D-подхода.

Степень злокачественности глиомы (ВОЗ)	Положительное прогностическое значение	Чувствительность	F1
I	0,60	0,56	0,58
II	0,11	0,45	0,17
III	0,02	0,32	0,04
IV	0,85	0,47	0,61

Обсуждение

Надежная неинвазивная персонализированная диагностика гистологического и молекулярно-генетического типа глиом по данным лучевых методов на дооперационном этапе может существенно улучшить выбор тактики лечения и повысить безопасность оказания медицинской помощи пациентам. На сегодняшний день «неинвазивная биопсия» головного мозга все еще является фантастикой. Однако, исследования, направленные на развитие принципов минимальной инвазии и персонализированного подхода в диагностике, являются стратегически приоритетными.

В цифровых архивах ФГАУ «НМИЦ нейрохирургии им. ак. Н.Н. Бурденко» Минздрава России накоплен большой объем информации о случаях лечения пациентов с глиомами головного мозга. Ежегодно в НМИЦ нейрохирургии выполняют более 1000 операций по поводу этой патологии. Именно поэтому такой большой массив накопленных данных может быть использован вторично в исследовательских целях. Проанализировать и извлечь информацию из наборов данных такого масштаба сегодня помогают наука о данных и технологии «искусственного интеллекта».

В настоящей работе представлен наш первый опыт типирования МР-изображений глиальных опухолей по степени злокачественности ВОЗ (классификация 2007 г.) с использованием искусственных нейронных сетей. Эту же задачу, по сути, интуитивно решает лучевой диагност или нейрохирург, впервые знакомясь с МР-изображением опухоли. Для условной оценки степени злокачественности глиальных опухолей с 2007 г. применяется четырехуровневая градация, предложенная Всемирной организацией здравоохранения (ВОЗ). Несмотря на кардинальный пересмотр системы классификации опухолей нервной системы в 2016 г. и ее модификация в 2021 г., классификация 2007 года еще продолжает широко применяться во многих клиниках.

Полученные нами результаты не противоречат данным научной литературы об эффективности машинного обучения в разделении изображений глиом на классы злокачественности. Например, Н. Cho и соавт. (2018) показали потенциал применения трех моделей (логистической регрессии, метода опорных векторов и случайного леса) в бинарном разделении изображений глиом на классы низкой и высокой степеней злокачественности со средней площадью под кривой (AUC) 0.9030 для тестовой выборки и средними точностью - 88%, чувствительностью - 95% и специфичностью - 70% (9). Авторы проанализировали данные МРТ 285 пациентов с опухолями головного мозга, используя T1-взвешенные изображения, T1-изображения с контрастным усилением, T2-взвешенные сканы и FLAIR. Извлечение признаков было выполнено с использованием алгоритма минимальной избыточности и максимальной релевантности.

С. Su и соавт. (2019) изучили диагностическую ценность методов радиомики в приложении к структурной, диффузионной и перфузионной МРТ для дифференцирования типов глиом и прогнозирования прогрессии опухоли (10). Наилучшее значение площади под ROC-кривой (AUC) в исследованиях авторов достигал для глиом II-III степени злокачественности по ВОЗ - 0,896, II-IV степени – 0,997 и для III-IV степени – 0,881.

Упомянутые выше исследования опирались, в основном на радиомику, или инженерию признаков МР-изображений. Z. Ning и соавт. (2021) попытались объединить локальные МР-признаки с признаками, извлеченными в процессе глубокого обучения с использованием сверточной нейронной сети в задаче неинвазивной классификации типов глиом (8). Площадь под ROC-кривой, чувствительность и специфичность моделей, обученных на комбинации таких свойств изображений, составили для тестовой выборки 0,88 (95% CI: 0,84; 0,91), 88% (95% CI: 80%; 93%) и 81% (95% CI: 76%; 86%) соответственно. Авторы показали, что разработанная модель имела большую эффективность, чем модели обученные только на какой-либо одной категории признаков

($p < 0.001$). При этом результаты работы модели оказались сопоставимыми с таковыми для лучевых диагностов.

В нашем исследовании применен изолированный подход к классификации МР-изображений на основе глубокого обучения. Согласно данным литературы и нашему первому опыту, бинарное разделение МРТ глиальных опухолей на низко- и высокозлокачественные классы является более решаемой задачей, чем дифференцирование подтипов глиом. Это вполне объяснимо, поскольку опытный врач, как правило, в типичных случаях видит различие между доброкачественной и злокачественной опухолью по данным МРТ. Таблица 1 показывает, что модель диагностирует I и IV классы достаточно уверенно, при этом точность дифференцирования между II и III классами меньше. Это согласуется и с опытом врачей, оценивающих сигналы МРТ.

Сегодня появляется все больше информации о том, что применение машинного обучения к предоперационным МР-данным позволяет с достаточно высокой вероятностью предсказать мутацию IDH I, метилирование промотора MGMT и делецию 1p/19q (5). Результаты нашей работы, выполненной на значительном объеме данных, говорят в пользу перспективности развития данного направления исследований.

Тем не менее, в настоящий момент не существует достаточной доказательной базы для того, чтобы ожидать появления инструментов неинвазивной лучевой «биопсии» по данным какой-либо одной модальности. Вероятнее, что наилучшее качество диагностики будет достигнуто за счет комбинации разных модальностей, включая МР-спектроскопию, КТ-перфузию и ПЭТ. Однако, сбор такого набора данных для каждого пациента представляет значительные сложности по множеству причин.

Также нет оснований полагать, что какая-либо одна технология сможет обеспечить максимальную эффективность анализа данных. Однако, методы искусственного интеллекта имеют потенциал занять в диагностике глиом важное место. Решения, создаваемые даже на основе одного метода и для какой-либо одной модальности, вносят свой вклад в базу знаний о возможностях неинвазивной диагностики опухолей с помощью нейровизуализации и искусственного интеллекта.

Ограничения нашего исследования связаны с гетерогенностью данных МРТ, полученных от разных МР-томографов (с разным количеством срезов, разным разрешением), отсутствием разметки каждого среза по наличию и степени злокачественности глиомы, отсутствием учета клинических факторов при анализе данных. Количество доступных для анализа МР-объектов также является лимитирующим фактором. В то же время наш набор данных существенно превышает серии наблюдений многих других авторов.

Заключение

Наши предварительные результаты показывают принципиальную возможность разделения T1-взвешенных МР-изображений глиальных опухолей с контрастным усилением по типам злокачественности с помощью моделей глубокого обучения. Диагностическая точность предложенного решения может быть улучшена за счет добавления к данным других модальностей МРТ, тестирования новых методов пред- и постобработки изображений и алгоритмов машинного обучения.

Конфликт интересов:

Работы по глубокому машинному обучению были поддержаны грантом Министерства науки высшего образования Российского № 075-15-2021-1343.

Литература:

1. Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018. *Neuro Oncol* [Internet]. 2021 Oct 1 [cited 2022 Jun 30];23(12 Suppl 2):III1–105. Available from: <https://pubmed.ncbi.nlm.nih.gov/34608945/>
2. Louis DN, Ohgaki H, Otmar ·, Wiestler D, Cavenee WK, Burger PC, et al. The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta Neuropathol*. 2007;114:97–109.
3. Louis DN, Perry A, Reifenberger · Guido, Von Deimling A, Figarella-Branger D, Webster ·, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol*.
4. Louis D, Perry A, Wesseling P, Brat D, Cree I, Figarella-Branger D, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* [Internet]. 2021 Aug 1 [cited 2021 Oct 28];23(8):1231–51. Available from: <https://pubmed.ncbi.nlm.nih.gov/34185076/>
5. Jian A, Jang K, Manuguerra M, Liu S, Magnussen J, Di Ieva A, et al. Machine Learning for the Prediction of Molecular Markers in Glioma on Magnetic Resonance Imaging: A Systematic Review and Meta-Analysis. *Neurosurgery* [Internet]. 2021 Apr 7 [cited 2021 Apr 24]; Available from: <https://academic.oup.com/neurosurgery/advance-article/doi/10.1093/neuros/nyab103/6214540>
6. Habib A, Jovanovich N, Hoppe M, Ak M, Mamindla P, Colen RR, et al. MRI-Based Radiomics and Radiogenomics in the Management of Low-Grade Gliomas: Evaluating the Evidence for a Paradigm Shift. *J Clin Med*. 2021;10.
7. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Stiphout RGPM van, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* [Internet]. 2012 Mar [cited 2021 Oct 28];48(4):441. Available from: <https://pubmed.ncbi.nlm.nih.gov/22449941/>
8. Ning Z, Luo J, Xiao Q, Cai L, Chen Y, Yu X, et al. Multi-modal magnetic resonance imaging-based grading analysis for gliomas by integrating radiomics and deep features. *Ann Transl Med* [Internet]. 2021 [cited 2021 May 11];9(4). Available from: <http://dx.doi.org/10.21037/atm-20-4076>
9. Cho H-H, Lee S-H, Kim J, Park H. Classification of the glioma grading using radiomics analysis. 2018 [cited 2021 May 11]; Available from: <http://doi.org/10.7717/peerj.5982>
10. Su C, Jiang J, Zhang S, Shi J, Xu K, Shen N, et al. Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour. 2019 [cited 2021 May 12]; Available from: <https://doi.org/10.1007/s00330-018-5704-8>