УДК 616-71+616-079.1 DOI: 10.18699/SSMJ20250214
Оригинальное исследование / Research article

Влияние метаданных компьютерно-томографических исследований головного мозга на производительность работы сервисов искусственного интеллекта: пилотное исследование

А.Н. Хоружая, Д.В. Кулиговский, Ю.А. Васильев

Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы
127051, г. Москва, ул. Петровка, 24, стр. 1

Резюме

Цель исследования – оценить влияние метаданных, добавляемых к обучающему набору данных, на производительность работы системы искусственного интеллекта, направленной на работу с диагностическими изображениями головного мозга. **Материал и методы.** В качестве основы использовали расширенный набор компьютерных томограмм головного мозга с наличием и отсутствием признаков внутричерепного кровоизлияния, дополненный клиническими и техническими параметрами. Из набора выбрали 176 исследований (106 – для обучения, 70 – для тестирования), которые специальным образом подготовили для подачи в нейронную сеть: сегментировали с выделением области интереса (головной мозг), нормализовали, удалили фон. В качестве базовой нейронной сети использовали нейросетевую архитектуру ResNet10, способную анализировать трехмерные медицинские изображения. Она была объединена с другой нейронной сетью для формирования архитектуры ResNet10_Meta, которая с помощью подхода One-Hot-Enconding (OHE) подключает к анализу непосредственно изображений другие метаданные. В качестве метаданных выбрали четыре параметра: Manufacturer, SliceThickness, PatientAge, XrayTubeCurrent. **Результаты.** Проведено 12 экспериментов по обучению нейронной сети с поочередным присоединением одного, двух или всех дополнительных параметров. Модель с добавлением параметра XTube продемонстрировала наиболее высокую специфичность (82,3 %, 95%-й доверительный интервал (95 % ДИ) [69,6–94,9]), превосходя базовую модель Baseline без подключения дополнительных метаданных (специфичность 79,4 %, 95 % ДИ [66,0–92,8]). Чувствительность модели с добавлением нескольких метаданных (XTube + SliceT) была сравнительно больше (69,7 %, 95 % ДИ [54,5–84,9]), чем модели Baseline. Модель с добавлением всех метаданных «All params» существенных улучшений не продемонстрировала. Однако все найденные различия оказались статистически незначимыми ($p > 0,05$). **Заключение.** Наши данные продемонстрировали отсутствие статистически значимых различий в производительности нейронной сети, анализирующей диагностические изображения без учета или с учетом метаданных, добавляемых к обучающему набору данных. Тем не менее данное исследование пилотное и проведено на ограниченной выборке, поэтому еще предстоит выяснить, могут ли коммерчески доступные инструменты искусственного интеллекта быть абсолютно нечувствительными к техническим характеристикам изображений.

Ключевые слова: компьютерная томография, искусственный интеллект, внутричерепные кровоизлияния, головной мозг, наборы данных.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Финансирование. Работа выполнена в рамках НИОКР «Разработка платформы подготовки наборов данных лучевых диагностических исследований» (№ ЕГИСУ: 123031500003-8) в соответствии с Приказом от 21.12.2022 № 1196 «Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счет средств бюджета города Москвы государственным бюджетным (автономным) учреждениям подведомственным Департаменту здравоохранения города Москвы, на 2023 год и плановый период 2024 и 2025 годов» Департамента здравоохранения города Москвы.

Автор для переписки. Хоружая А.Н., e-mail: khoruzhayaAN@zdrav.mos.ru

Для цитирования. Хоружая А.Н., Кулиговский Д.В., Васильев Ю.А. Влияние метаданных компьютерно-томографических исследований головного мозга на производительность работы сервисов искусственного интеллекта: пилотное исследование. *Сиб. науч. мед. ж.* 2025;45(2):132–141. doi: 10.18699/SSMJ20250214

The impact of brain computed tomography metadata on the performance of diagnostic artificial intelligence services: a pilot study

A.N. Khoruzhaya, D.V. Kuligovskiy, Yu.A. Vasilev

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department
127051, Moscow, Petrovka st., 24/1

Abstract

The aim of the study is to evaluate the impact of metadata added to the training dataset on the performance of an artificial intelligence system aimed at working with diagnostic brain images. **Material and Methods.** An expanded set of computed tomography scans of the brain with and without signs of intracranial haemorrhage, supplemented with clinical and technical parameters, was used as a basis. From the dataset, 176 studies were selected (106 for training, 70 for testing), which were specially prepared for input into the neural network: they were segmented with the region of interest (the brain) highlighted, normalized, and the background was removed. The ResNet10 neural network architecture, which is capable of analyzing 3D medical images, was used as the underlying neural network. It was combined with another neural network to form the ResNet10_Meta architecture, which, using a One-Hot-Enconding (OHE) approach, connects other metadata to analyze images directly. Four parameters were selected as metadata: 'Manufacturer', 'SliceThickness', 'PatientAge', and 'XrayTubeCurrent'. **Results.** Twelve experiments were conducted to train the neural network with one, two, or all of the additional parameters added alternately. The model with the addition of the 'XTube' parameter showed the highest specificity (82.3 %, 95 % confidence interval (95 % CI) [69.6–94.9]), outperforming the Baseline model without additional metadata (specificity 79.4 %, 95 % CI [66.0–92.8]). The 'XTube + SliceT' model with the addition of multiple metadata showed comparatively higher sensitivity (69.7 %, 95 % CI [54.5–84.9]) relative to Baseline model. The model with the addition of all metadata 'All params' did not show significant improvements. However, all differences found were statistically insignificant ($p > 0.05$). **Conclusions.** Our data demonstrated that there were no statistically significant differences in the performance of a neural network analyzing diagnostic images without or with metadata added to the training dataset. However, this study is pilot and was conducted on a limited sample, so it remains to be seen whether commercially available artificial intelligence tools can be completely insensitive to image specifications.

Key words: computed tomography, intracranial hemorrhage, artificial intelligence, brain, training dataset.

Conflict of interest. The authors declare no conflict of interest.

Financing. This work was carried out as part of the research and development effort titled "Development of a platform to generate data sets containing diagnostic imaging studies" (USIS No.: 123031500003-8) in accordance with the Order No. 1196 dated December 21, 2022 "On approval of state assignments funded by means of allocations from the budget of the city of Moscow to the state budgetary (autonomous) institutions subordinate to the Moscow Health Care Department, for 2023 and the planned period of 2024 and 2025" issued by the Moscow Health Care Department.

Correspondence author. Khoruzhaya A.N., e-mail: khoruzhayaAN@zdrav.mos.ru

Citation. Khoruzhaya A.N., Kuligovskiy D.V., Vasilev Yu.A. The impact of brain computed tomography metadata on the performance of diagnostic artificial intelligence services: a pilot study. *Sibirskij nauchnyj medicinskij zhurnal = Siberian Scientific Medical Journal.* 2025;45(2):132–141. [In Russian]. doi: 10.18699/SSMJ20250214

Введение

Нагрузка на врачей-рентгенологов непрерывно увеличивается, превышая рост количества новых специалистов данного профиля. Для того чтобы рентгенологам было проще поддерживать высокую производительность труда, активно разрабатываются и внедряются в клиническую деятельность различные системы поддержки принятия решений с использованием искусственного интеллекта (системы искусственного интеллекта, СИИ) [1]. В своей типичной реализации СИИ, которые обучены искать определенную патологию, обрабатывают медицинские изображения сразу после того как они были сделаны, размечают их, выделяя патологические области, и формируют текстовый ответ, который приходит рентгенологу вместе с обработанной СИИ и нативной сериями [2–4], ранжируют списки исследований для описания, подсвечивая те, где есть неотложная патология (т.е. осуществляют триаж) [5]. Также СИИ могут производить разметку естественных ана-

томических структур (измерять размеры, объем, степень смещения и прочее) – то, что рентгенолог обычно делает вручную [6, 7].

Несмотря на большие ожидания от подобных СИИ и потенциальные преимущества их использования, из-за некоторых ограничений их не сразу можно встроить в рутинную практику [8]. Например, чтобы диагностическая ценность СИИ была высокой, метрики их работы должны достигать оптимальных значений, что требует использования больших и разнообразных обучающих наборов данных [9, 10]. Влияют и погрешности при составлении наборов данных [11], и плохая обобщаемость [12], и ограниченное количество клинических валидационных исследований, где в режиме реальной клинической практики СИИ проходили бы «боевое тестирование» [13]. Некорректный анализ СИИ диагностического исследования вместо помощи может затруднить работу врача, вынуждая его уделять дополнительное внимание ложноположительным результатам или меньшее – ложноотрицательным, что может привести к большему количеству пропусков патологий [14].

Помимо чисто технических и инфраструктурных сложностей имеется множество данных, когда СИИ с изначально высокими метриками точности работы встраивался в клинический поток исследований, и его показатели падали. Особенно часто это наблюдается в отношении специфичности [15, 16]. Причин может быть несколько. Во-первых, в условиях рутинной практики встречается гораздо больше, чем в обучающих или валидационных наборах данных, исследований с различными артефактами, что осложняет корректную обработку изображений для СИИ. Во-вторых, исследования, выполненные на разных моделях диагностического оборудования от разных производителей, имеют технические характеристики, отличающиеся друг от друга. Особенно это заметно для компьютерной томографии (КТ). Разнообразие фильтров реконструкций, значений силы тока трубки, толщины срезов и других метаданных может влиять на метрики работы СИИ [17], поскольку они изменяют характеристики самих изображений [18–20]. Тем не менее достоверной исчерпывающей информации на этот счет пока нет.

Мы предположили, что добавление к наборам данных, предназначенных для обучения СИИ и содержащих диагностические изображения с наличием или отсутствием патологии, некоторых метаданных, которые относятся к техническим параметрам КТ, помогут улучшить метрики точности. В качестве наиболее показательного и простого для обучения нейронных сетей примера была выбрана КТ головного мозга с признаками внутричерепных кровоизлияний [21]. Таким образом, целью нашей работы стала оценка влияния метаданных, добавляемых к обучающему набору данных, на производительность работы СИИ, направленной на диагностику внутричерепных кровоизлияний.

Материал и методы

Формирование наборов данных для обучения и тестирования

В качестве основы использован расширенный набор КТ головного мозга с наличием и отсутствием признаков внутричерепного кровоизлияния, дополненный клиническими и техническими параметрами [22]. Набор состоял из 800 КТ, из которых 400 исследований содержали кровоизлияния и 400 были без признаков патологии. Среди технических метаданных, которые теоретически подходили для дальнейшего анализа, содержались следующие параметры: фильтр реконструции (ConvolutionKernel), описание серии (SeriesDescription), толщина среза (SliceThickness), производитель (Manufacturer), напряжение на трубке (KVP), сила тока трубки (XRayTubeCurrent), возраст пациента (Patient Age).

Для формирования из данного набора обучающей и тестовой выборок мы исключили серии, содержащие артефакты от движений и металлических конструкций, с толщиной срезов более 2,5 мм, а также реконструированные с использованием «костного» фильтра. В тренировочную выборку вошло 106 исследований с соотношением классов 50 на 50 (53 исследования без патологии, 53 исследования с кровоизлияниями), из них 90 использовались для обучения и 16 – для проверки. В тестовую выборку вошло 70 исследований (35 исследований без патологий, 35 исследований с кровоизлияниями). Тренировочная и тестовая выборки также балансировались с учетом метаданных. По этой причине было взято только 176 КТ из 800, содержащихся в исходном наборе данных. В нем преобладал производитель Toshiba (596 исследований), другие же производители были представлены в гораздо меньшем количестве: GE Medical Systems – 95 исследований, Siemens – 87 исследований, Philips – 22 исследования. Перевес томографов производителя Toshiba, в целом, соответствует российскому ландшафту диагностического оборудования, поэтому было принято решение выполнить балансировку обучающей и тестовой выборок по параметру Manufacturer (рис. 1).

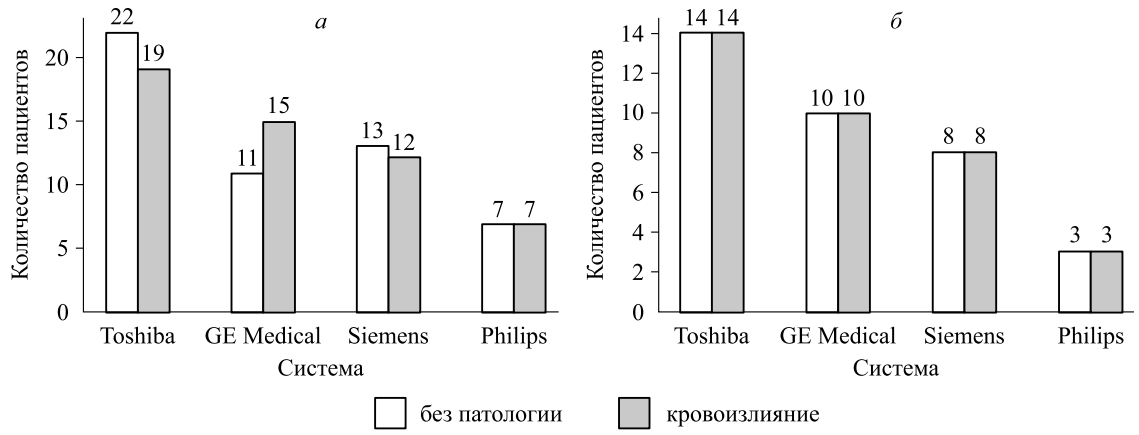

Рис. 1. *Разбиение исследований по параметру «Manufacturer» в обучающей (а) и тестовой (б) выборке*

Fig. 1. *Characterisation of brain computed tomography scans according to the parameter 'Manufacturer' in the training (a) and test samples (б)*

Подготовка КТ-исследований для подачи в нейронную сеть

Для выделения области интереса (ROI) и стандартизации исследований применяли сегментацию мозга. Использовали предобученную нейронную сеть U-net, которая способна выделять на бесконтрастной КТ одновременно 11 внутричерепных структур. Код модели и веса, полученные после обучения, доступны в репозитории Github [23]. Процесс перевода масок из детального представления в бинарное показан на рис. 2. После наложения масок перед подачей в нейронную сеть на томограммах удаляли фоновые структуры (CropForeground), масштабировали значения пикселей в диапазоне [–20, 200] (ScaleIntensityRange), выходное расстояние между вокселами приводили к значению 2.0 для высоты и ширины, 1.5 – для оси глубины (Spacing), а также нормализовали значения пикселей в диапазоне [0, 1] (Normalize) перед входом в нейронную сеть. Далее полученные предобработанные исследования (рис. 3) использовали для обучения нейронной сети.

Нейронная сеть для анализа диагностических изображений с классификатором

Для того чтобы исследовать влияние метаданных на работу СИИ, направленной на анализ диагностических изображений, выбрана нейросетевая архитектура ResNet10 (3D). ResNet10 состоит из четырех блоков, размер каждого из которых составляет 64, 128, 256, 512 фильтров. На вход

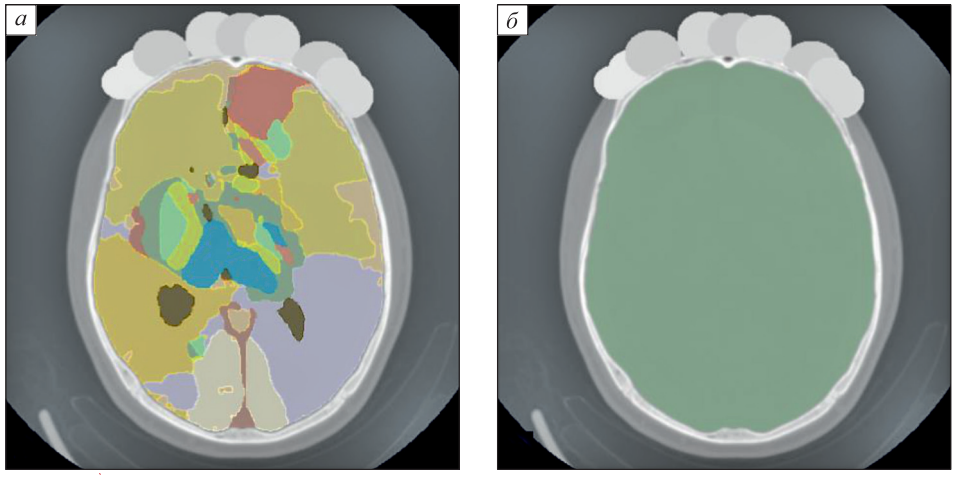

Рис. 2. *Перевод маски (а) в бинарное представление (б)*

Fig. 2. *Translating a mask (a) into a binary representation (б)*

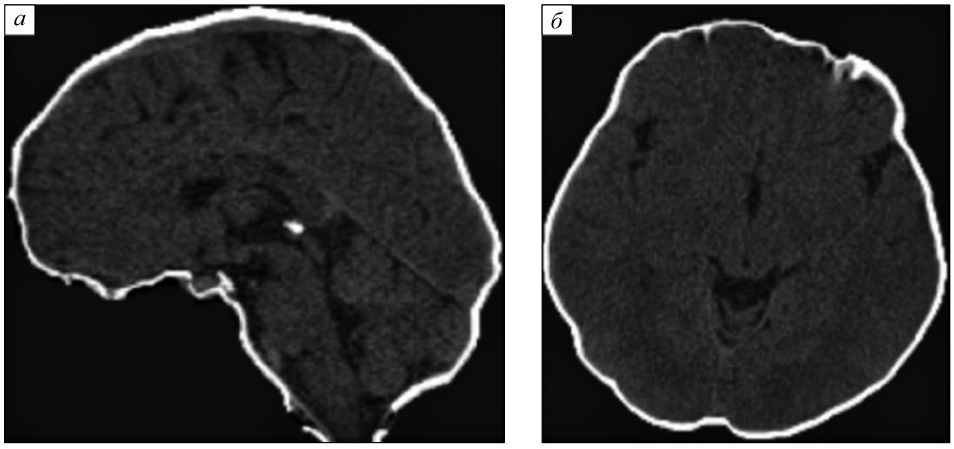

Рис. 3. *КТ-исследование (сагиттальная (а) и аксиальная (б)) плоскость, готовое для подачи на нейронную сеть после предобрабработки*

Fig. 3. *A computed tomography scan (sagittal (a) and axial planes (б)) ready to feed to the neural network after preprocessing*

подается одноканальное трехмерное изображение, на выходе модель прогнозирует два класса (норма или кровоизлияние). Обучение построено на PyTorch, в качестве оптимизатора использовали RAdam, начальная скорость обучения составляла 5e-5, каждые две эпохи она снижалась на 10 %, всего прошло 12 эпох обучения.

Выбор метаданных

Из всех представленных в наборе технических метаданных мы выбрали четыре параметра для проверки их влияния на работу нейронной сети: «Manufacturer», «SliceThickness», «PatientAge», «XrayTubeCurrent». Тег «Manufacturer» содержит наименование компании-производителя оборудования, на котором выполнено исследование, т.е., по сути, уже объединяет в себе информацию, которая находится в тегах «ConvolutionKernel» и «KVP», поскольку эти значения у каждого производителя свои и почти не варьируют от модели к модели. Тег «SliceThickness» описывает толщину среза, с которой выполнено сканирование, «PatientAge» содержит информацию о возрасте пациента, «XrayTubeCurrent» показывает силу тока, протекающего через рентгеновскую трубку КТ-аппарата.

Кодирование метаданных для подключения

Для кодирования метаданных применен подход One-Hot-Enconding (OHE), который основывается на создании бинарных признаков, показывающих принадлежность к уникальному значению. Возраст мы разделили на четыре интервала, согласно классификации ВОЗ: 0–44 (молодой возраст), 44–59 (средний возраст), 60–75 (пожилой возраст), более 75 лет (старческий возраст). Предположим, возраст пациента 56 лет, он попадает во второй интервал. Тогда число кодируется в OHE следующим образом: [0., 1., 0., 0.] (для второго интервала 1, для других 0). Для тега «Manufacturer» также выбрано четыре интервала, поскольку в наборе данных представлено четыре компании-производителя. Тег «SliceThickness» разделен на четыре интервала для балансировки по частоте встречаемости: в первый интервал вошли срезы толщиной 0,5 мм, во второй – 0,625 и 0,75 мм, в третий – 1 мм, в четвертый – 1,25, 1,5, 2,0 и 2,5 мм. Для тега «XRayTubeCurrent» выбраны интервалы 0–110, 111–200, 201–305 и более 305 мА с целью сформировать равномерные по количеству исследований интервалы для тренировочной и тестовой выборок (т.е. также сбалансировать) (рис. 4).

Подключение метаданных к нейронной сети

Мы объединили в единую архитектуру ResNet10_Meta две нейронные сети, одна из которых, ResNet10, анализирует диагностические исследования в 3D, а вторая принимает закодированные в формат OHE метаданные, соответствующие обрабатываемому исследованию. На вход ResNet10_Meta подаются диагностическая серия и метаданные. Признаки, полученные от двух нейронных сетей, объединяются в один вектор, и предсказывается класс (кровоизлияние или норма). Этот подход позволил реализовать и оценить способность нейронной сети находить взаимосвязь между изображением и метаданными.

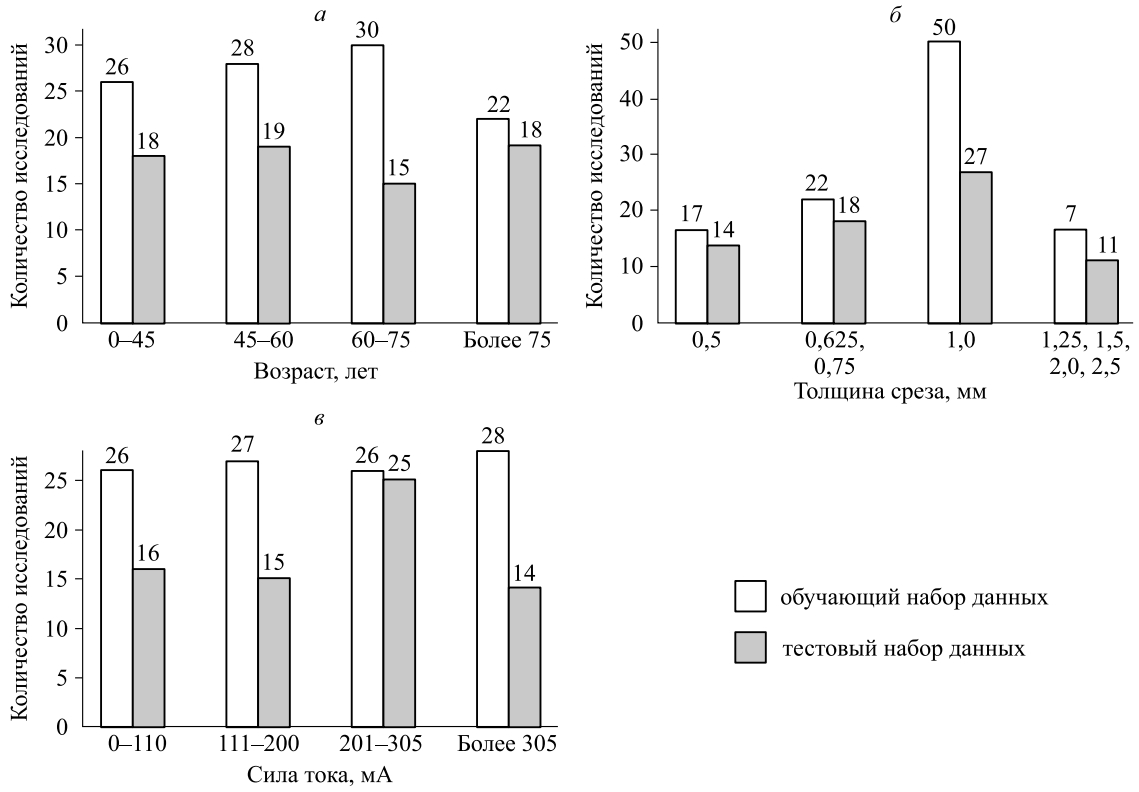

Рис. 4. *Распределение возрастов (а), толщин срезов (б) и сил тока (в) в обучающем и тестовом наборе данных*
Fig. 4. *Distribution of age (a), slice thickness (б) and current strength (в) in the training and test datasets*

Статистический анализ

Все статистические показатели, представленные в результатах, рассчитаны с 95 %-м доверительным интервалом (95 % ДИ) методом случайных выборок с возвратом (так называемый бутстреппинг, от англ. bootstrapping) с количеством итераций 1000. Оценка качества работы алгоритмов проводилась при помощи функции classification_report. Для статистического анализа использовался критерий МакНемара.

Результаты и их обсуждение

Выполнено 12 экспериментов по обучению нейронной сети, после каждого из которых проводилось тестирование и фиксировались метрики чувствительности (Sensitivity), специфичности (Specificity), аккуратности (Accuracy) и точности (Precision). В первом эксперименте обучение и тестирование проходила нейронная сеть ResNet10 без подключения к ней каких-либо дополнительных метаданных (Baseline), в каждом последующем подключались один, два, либо все метаданные (таблица).

Модель с добавлением параметра «XTube» демонстрирует более высокую специфичность (82,3 %), превосходя базовую модель «baseline» (79,4 %), но самую низкую чувствительность (58,7 %) среди всех моделей, что говорит о большом количестве ложноотрицательных случаев (пропусков целевой патологии в тех случаях, где она присутствует). Модель с добавлением нескольких метаданных «XTube+SliceT» продемонстрировала сравнительно большую чувствительность (69,7 %) относительно baseline. Модель с добавлением всех метаданных «All params» существенных улучшений не показала. Хотя некоторые модели показывают улучшения в отдельных метриках в пределах случайных флуктуаций, базовая модель без добавления любых метаданных сохраняет сбалансированные показатели. Также, несмотря на отличия в метриках моделей с добавлением метаданных, ни одна из них не показала значимых отличий от метрик, которые продемонстрировала базовая модель ($p > 0{,}05$).

Для проверки гипотезы об изменении метрик с использованием другой нейросетевой архитектуры был проверен пул экспериментов с нейросетью EfficientNet, полностью подобный тому, который осуществлялся для нейросети ResNet10. Выполнено также 12 стадий обучения и тестирования (базовая и 11 с дополнением метаданными). Сравнение полученных для EfficientNet метрик (как в случае базовой конфигурации, так и дополненной метаданными) с аналогичными

Метрики, полученные в результате обучения модели ИИ без добавления дополнительных параметров (Baseline) и с добавлением единичных дополнительных параметров, а также их сочетаний, 95 % ДИ

Metrics obtained from training the AI model without adding metadata (Baseline) and with single metadata, as well as their combinations, 95 % CI

Показатель	Чувствительность	Специфичность	Точность	Аккуратность
Baseline	62,3 [46,23–78,34]	79,4 [66,04–92,82]	75,2 [59,45–90,89]	70,9 [60,21–81,51]
Один параметр				
Возраст	64 [48,1–79,9]	78,3 [64,63–91,95]	74,7 [59,11–90,23]	71,1 [60,53–81,75]
Производитель	67,4 [51,9–82,9]	73,1 [58,46–87,82]	71,5 [56,12–86,92]	70,3 [59,58–81,0]
Толщина среза	66,9 [51,3–82,5]	75,4 [61,17–89,69]	73,1 [57,76–88,48]	71,1 [60,53–81,75]
Сила тока	58,9 [42,55–75,16]	82,3 [69,64–94,94]	76,9 [60,91–92,83]	70,6 [59,89–81,25]
Два параметра				
Сила тока, толщина среза	69,7 [54,49–84,94]	71,4 [56,46–86,4]	70,9 [55,76–86,1]	70,6 [59,89–81,25]
Сила тока, возраст	61,7 [45,61–77,82]	79,4 [66,04–92,82]	75 [59,19–90,81]	70,6 [59,89–81,25]
Сила тока, производитель	61,1 [44,99–77,29]	76 [61,85–90,15]	71,8 [55,66–87,96]	68,6 [57,69–79,45]
Толщина среза, возраст	65,1 [49,36–80,93]	78,3 [64,63–91,95]	75 [59,61–90,39]	71,7 [61,16–82,26]
Толщина среза, производитель	69,1 [53,84–84,45]	76,6 [62,54–90,6]	74,7 [59,72–89,66]	72,9 [62,44–83,28]
Производитель, возраст	66,3 [50,62–81,95]	77,1 [63,23–91,05]	74,4 [59,04–89,68]	71,7 [61,16–82,26]
Все параметры				
	61,1 [44,99–77,29]	81,1 [68,18–94,1]	76,4 [60,71–92,15]	71,1 [60,53–81,75]

метриками, полученными для ResNet10, не продемонстрировало статистически значимых различий.

В литературе появляется все больше подтверждений тому, что добавление к обучающим наборам данных, которые предназначены для СИИ, направленных на анализ медицинских изображений, дополнительной клинической информации о пациенте может существенно улучшить производительность подобных систем [25]. Но, возможно, стоит обучать СИИ распознавать патологию на КТ-изображениях, используя не только дополнительную клиническую информацию, но и техническую, поскольку реконструируемое изображение может обладать такими особенностями, на которые в том числе может опираться нейронная сеть при обработке [15]. Точных данных, которые бы подтверждали или опровергали этот тезис, в литературе практически нет.

В 2019 г. опубликован приказ правительства города Москвы о начале проведения Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и их дальнейшего применения в системе здравоохранения. На сегодняшний момент в Эксперименте участвует 54 ИИ-сервиса по 34 направлениям. Среди всех сервисов анализ КТ проводят около трети, обрабатывая по большей части данные КТ органов грудной клетки и головного мозга [26].

Внутричерепные кровоизлияния – острая неотложная патология, которая при задержке диагностики или пропуске имеет высокий риск неблагоприятного исхода для пациента [27]. Поэтому более качественное обучение систем, направленных на поддержку врачей в диагностике данной патологии, имеет очевидные позитивные последствия. Тем не менее полученные нами данные не продемонстрировали статистически значимых различий в производительности между моделями, обученными только на КТ головного мозга с наличием и отсутствием кровоизлияний,

либо на тех же томограммах с дополнительно подключаемыми к ним метаданными, содержащими технические характеристики изображений. Подобные выводы делают в своей работе A. Kundisch et al. [24], отмечая, что качество изображения, так же как и размер кровоизлияния, не были существенно связаны с вероятностью неверных результатов СИИ. J. McLouth et al. также сообщают о высокой обобщаемости результатов вне зависимости от моделей и настроек диагностического оборудования [28].

При этом все же, вероятно, имеются некоторые технические особенности сканирования, которые могут влиять на показатели СИИ. Например, в исследовании A. Zia et al. диагностическая точность коммерчески доступной СИИ оказалась самой низкой среди амбулаторной популяции [17]. Авторы отмечают, что существует систематическая разница не только в распространенности, но и в фактическом содержании амбулаторных КТ-сканирований – возможно, имеют значение особенности сканеров или меньший объем кровоизлияний.

Наше пилотное исследование имеет ряд ограничений. К ним относится выбор нейронной сети с открытым исходным кодом, которая не была предобучена и настроена специально для диагностики внутричерепных кровоизлияний на КТ головного мозга, хоть и имела преимущества для анализа непосредственно изображений. Также в анализируемую выборку были включены КТ-данные, полученные только из Единого радиологического информационного сервиса Москвы без учета частных медицинских организаций. В свою очередь учитывались данные только четырех производителей компьютерных томографов, при этом отдельный анализ по моделям не проводился. Данные причины могут ограничить воспроизводимость исследования на других наборах данных из других медицинских информационных систем с другим распределением моделей томографов. Кроме того, выборка ограничилась небольшим набором данных, что обусловлено в том числе недостаточной производительностью компьютеров, на которых происходило обучение и тестирование нейронных сетей. С увеличением размеров выборки возможно преодоление порога статистически значимых отличий для модели с метаданными «XTube+Manuf», поэтому планируется продолжить работу и проверить закономерности на большем количестве исследований.

Заключение

Наши данные продемонстрировали отсутствие достоверных различий в производительности нейронной сети, анализирующей диагностические изображения без учета или с учетом метаданных, добавляемых к обучающему набору данных. Таким образом, кажется маловероятным, что качество изображения само по себе является значимым фактором для СИИ. Хотя еще предстоит выяснить, могут ли коммерчески доступные инструменты ИИ быть абсолютно нечувствительными к техническим характеристикам изображений. Особенности компьютерных томограмм могут довольно сильно отличаться в интерпретации человеком и ИИ.

Список литературы / References

1. Kwee T.C., Kwee R.M. Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. *Insights Imaging*. 2021;12(1):88–101. doi: 10.1186/s13244-021-01031-4

2. Arzamasov K., Vasilev Y., Vladzymyrskyy A., Omelyanskaya O., Shulkin I., Kozikhina D., Goncharova I., Gelezhe P., Kirpichev Y., Bobrovskaya T., Andreychenko A. An international non-inferiority study for the benchmarking of ai for routine radiology cases: chest X-ray, fluorography and mammography. *Healthcare (Basel)*. 2023;11(12):1684. doi: 10.3390/healthcare11121684

3. Соловьёв А.В., Васильев Ю.А., Синицын В.Е., Петряйкин А.В., Владзимирский А.В., Шулькин И.М., Шарова Д.Е., Семенов Д.С. Вклад систем искусственного интеллекта в улучшение выявления аневризм аорты по данным компьютерной томографии грудной клетки. *Digital Diagnostics*. 2024;5(1):29–40. doi: 10.17816/DD569388

Solovev A.V., Vasilev Yu.A., Sinitsyn V.E., Petryaykin A.V., Vladzimirskiy A.V., Shulkin I.M., Sharova D.E., Semenov D.S. Improving aortic aneurysm detection with artificial intelligence based on chest computed tomography data. *Digital Diagnostics*. 2024;5(1):29–40. [In Russian]. doi: 10.17816/DD569388

4. Bizzo B.C., Almeida R.R., Michalski M.H., Alkasab T.K. Artificial Intelligence and Clinical Decision Support for Radiologists and Referring Providers. *J. Am. Coll. Radiol*. 2019;16(9 Pt B):1351–1356. doi: 10.1016/j.jacr.2019.06.010

5. Jha S. Value of triage by artificial intelligence. *Acad. Radiol*. 2020;27(1):153–155. doi: 10.1016/j.acra.2019.11.002

6. Alukaev D., Kiselev S., Mustafaev T., Ainur A., Ibragimov B., Vrtovec T. A deep learning framework for vertebral morphometry and Cobb angle measurement with external validation. *Eur. Spine J*. 2022;31(8):2115–2124. doi: 10.1007/s00586-022-07245-4

7. Hamyoon H., Yee Chan W., Mohammadi A., Yusuf Kuzan T., Mirza-Aghazadeh-Attari M., Leong W.L., Murzoglu Altintoprak K., Vijayananthan A, Rahmat K.,

Ab Mumin N., … Abbasian Ardakani A. Artificial intelligence, BI-RADS evaluation and morphometry: A novel combination to diagnose breast cancer using ultrasonography, results from multi-center cohorts. *Eur. J. Radiol*. 2022;157:110591. doi: 10.1016/j.ejrad.2022.110591

8. Tariq A., Purkayastha S., Padmanaban G.P., Krupinski E., Trivedi H., Banerjee I., Gichoya J.W. Current clinical applications of artificial intelligence in radiology and their best supporting evidence. *J. Am. Coll. Radiol*. 2020;17(11):1371–1381. doi: 10.1016/j.jacr.2020.08.018

9. Alwosheel A., van Cranenburgh S., Chorus C.G. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *J. Choice Mod*. 2018;28:167–182. doi: 10.1016/j.jocm.2018.07.002

10. Васильев Ю.А., Бобровская Т.М., Арзамасов К.М., Четвериков С.Ф., Владзимирский А.В., Омелянская О.В., Андрейченко А.Е., Павлов Н.А., Анищенко Л.Н. Основополагающие принципы стандартизации и систематизации информации о наборах данных для машинного обучения в медицинской диагностике. *Менеджер здравоохр*. 2023;(4):28–41. doi: 10.21045/1811-0185-2023-4-28-41

Vasilev Yu.A., Bobrovskaya T.M., Arzamasov K.M., Chetverikov S.F., Vladzimirskiy A.V., Omelyanskaya O.V., Andreychenko A.E., Pavlov N.A., Anishchenko L.N. Medical datasets for machine learning: fundamental principles of standartization and systematization. *Menedzher zdravookhraneniya = Manager of Health Care*. 2023;(4):28–41. [In Russian]. doi: 10.21045/1811-0185-2023-4-28-41

11. Park S.H., Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018;286(3):800–809. doi: 10.1148/radiol.2017171920

12. Zech J.R., Badgeley M.A., Liu M., Costa A.B., Titano J.J., Oermann E.K. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med*. 2018;15(11):e1002683. doi: 10.1371/journal.pmed.1002683

13. Васильев Ю.А., Владзимирский А.В., Омелянская О.В., Решетников Р.В., Блохин И.А., Коденко М.Р., Нанова О.Г. Обзор метаанализов о применении искусственного интеллекта в лучевой диагностике. *Мед. визуализ*. 2024;28(3):22–41. doi: 10.24835/1607-0763-1425

Vasilev Yu.A., Vladzimirskiy A.V., Omelyanskaya O.V., Reshetnikov R.V., Blokhin I.A., Kodenko M.R., Nanova O.G. Review of meta-analyses on the use of artificial intelligence in radiology. *Meditsinskaya vizualizatsiya = Medical Visualization*. 2024;28(3):22–41. [In Russian]. doi: 10.24835/1607-0763-1425

14. Morozov S., Vladzimirskiy A., Ledikhova N., Andreychenko A., Arzamasov K., Omelyanskaya O., Reshetnikov R., Gelezhe P., Blokhin I., Turavilova E., … Bondarchuk D. Diagnostic accuracy of artificial intelligence for analysis of 1.3 million medical imaging studies: the Moscow experiment on computer vision technologies. *medRxiv*. Preprint. Available at: https://doi.org/10.1101/2023.08.31.23294896

15. Seyam M., Weikert T., Sauter A., Brehm A., Psychogios M.N., Blackham K.A. Utilization of artificial intelligence-based intracranial hemorrhage detection on emergent noncontrast CT images in clinical workflow. *Radiol. Artif. Intell*. 2022;4(2):e210168. doi: 10.1148/ryai.210168

16. Voter A.F., Meram E., Garrett J.W., Yu J.J. Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of intracranial hemorrhage. *J .Am. Coll. Radiol*. 2021;18(8):1143–1152. doi: 10.1016/j.jacr.2021.03.005

17. Zia A., Fletcher C., Bigwood S., Ratnakanthan P., Seah J., Lee R., Kavnoudias H., Law M. Retrospective analysis and prospective validation of an AI-based software for intracranial haemorrhage detection at a high-volume trauma centre. *Sci. Rep*. 2022;12(1):19885. doi: 10.1038/s41598-022-24504-y

18. Schaller S., Wildberger J.E., Raupach R., Niethammer M., Klingenbeck-Regn K., Flohr T. Spatial domain filtering for fast modification of the tradeoff between image sharpness and pixel noise in computed tomography. *IEEE Trans. Med. Imaging*. 2003;22(7):846–853. doi: 10.1109/TMI.2003.815073

19. Wang Y., de Bock G.H., van Klaveren R.J., van Ooyen P., Tukker W., Zhao Y., Dorrius M.D., Proença R.V., Post W.J., Oudkerk M. Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability. *Eur. Radiol*. 2010;20(5):1180–1197. doi: 10.1007/s00330-009-1634-9

20. Behrendt F.F., Das M., Mahnken A.H., Kraus T., Bakai A., Stanzel S., Günther R.W., Wildberger J.E. Computer-aided measurements of pulmonary emphysema in chest multidetector-row spiral computed tomography: effect of image reconstruction parameters. *J. Comput. Assist. Tomogr*. 2008;32(6):899–904. doi: 10.1097/RCT.0b013e31815ade64

21. Yeo M., Tahayori B., Kok H.K., Maingard J., Kutaiba N., Russell J., Thijs V., Jhamb A., Chandra R.V., Brooks M., Barras C.D., Asadi H. Review of deep learning algorithms for the automatic detection of intracranial hemorrhages on computed tomography head imaging. *J. Neurointerv. Surg*. 2021;13(4):369–378. doi: 10.1136/neurintsurg-2020-017099

22. Khoruzhaya A.N., Bobrovskaya T.M., Kozlov D.V., Kuligovskiy D., Novik V.P., Arzamasov K.M., Kremneva E.I. Expanded brain Ct dataset for the development of AI systems for intracranial hemorrhage detection and classification. *Data*. 2024;9(2):30–36. doi: 10.3390/data9020030

23. Cai J.C., Akkus Z., Philbrick K.A., Boonrod A., Hoodeshenas S., Weston A.D., Rouzrokh P.,

Conte G.M., Zeinoddini A., Vogelsang D.C., Huang Q., Erickson B.J. Fully automated segmentation of head ct neuroanatomy using deep learning. *Radiol. Artif. Intell*. 2020;2(5):e190183. doi: 10.1148/ryai.2020190183

24. Kundisch A., Hönning A., Mutze S., Kreissl L., Spohn F., Lemcke J., Sitz M., Sparenberg P., Goelz L. Deep learning algorithm in detecting intracranial hemorrhages on emergency computed tomographies. *PLoS One*. 2021;16(11):e0260560. doi: 10.1371/journal.pone.0260560

25. Schutte K., Brulport F., Harguem-Zayani S., Schiratti J.B., Ghermi R., Jehanno P., Jaeger A., Alamri T., Naccache R., Haddag-Miliani L., … Lassau N. An artificial intelligence model predicts the survival of solid tumour patients from imaging and clinical data. *Eur. J. Cancer*. 2022;174:90–98. doi: 10.1016/j.ejca.2022.06.055

26. Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы. Режим доступа: https://mosmed.ai/ai/

Experiment on the use of innovative technologies in the field of computer vision to analyse medical images and further application in the health care system of Moscow. Available at: https://mosmed.ai/ai/ [In Russian].

27. van Asch C.J., Luitse M.J., Rinkel G.J., van der Tweel I., Algra A., Klijn C.J. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. *Lancet Neurol*. 2010;9(2):167–176. doi: 10.1016/S1474-4422(09)70340-0

28. McLouth J., Elstrott S., Chaibi Y., Quenet S., Chang P.D., Chow D.S., Soun J.E. Validation of a deep learning tool in the detection of intracranial hemorrhage and large vessel occlusion. *Front Neurol*. 2021;12:656112. doi: 10.3389/fneur.2021.656112

Сведения об авторах:

Хоружая Анна Николаевна, ORCID: 0000-0003-4857-5404, e-mail: khoruzhayaAN@zdrav.mos.ru
Кулиговский Дмитрий Вадимович, e-mail: KuligovskiiDV@zdrav.mos.ru
Васильев Юрий Александрович, к.м.н., ORCID: 0000-0002-5283-5961, e-mail: VasilevYA1@zdrav.mos.ru

Information about the authors:

Anna N. Khoruzhaya, ORCID: 0000-0003-4857-5404, e-mail: khoruzhayaAN@zdrav.mos.ru
Dmitriy V. Kuligovskiy, e-mail: KuligovskiiDV@zdrav.mos.ru
Yuriy A. Vasilev, candidate of medical sciences, ORCID: 0000-0002-5283-5961, e-mail: VasilevYA1@zdrav.mos.ru

Поступила в редакцию 08.11.2024
После доработки 26.12.2024
Принята к публикации 17.02.2025

Received 08.11.2024
Revision received 26.12.2024
Accepted 17.02.2025