

Диагностика внутричерепных кровоизлияний по данным компьютерной томографии головного мозга с помощью искусственного интеллекта

А.Н. Хоружая¹, К.М. Арзамасов¹, М.Р. Коленко¹, Е.И. Кремнёва^{1,2}, Д.В. Буренчев¹

¹ Научно-практический клинический центр диагностики и телемедицинских технологий, Москва, Россия;

² Научный центр неврологии, Москва, Россия

АННОТАЦИЯ

Обоснование. Внутричерепные кровоизлияния характеризуются высокой летальностью и риском инвалидизации, что обуславливает необходимость оперативной и точной диагностики, особенно в первые 24 часа. Использование технологий искусственного интеллекта для анализа компьютерных томограмм головного мозга позволяет сократить время диагностики и улучшить её качество. Актуальность работы подчеркнута ограниченным числом сертифицированных в России сервисов искусственного интеллекта для выявления внутричерепных кровоизлияний, а также отсутствием данных о их долгосрочной эффективности, что обуславливает необходимость многоцентрового мониторинга для оценки устойчивости и точности таких систем в реальной клинической практике.

Цель — оценить диагностическую точность и устойчивость сервиса искусственного интеллекта для диагностики внутричерепных кровоизлияний по данным нативной компьютерной томографии головного мозга в условиях многоцентрового клинического мониторинга на протяжении 18 месяцев.

Методы. Для анализа использовали анонимизированные компьютерные томограммы головного мозга. Сервис искусственного интеллекта прошёл трёхэтапное тестирование для оценки его точности и клинической производительности на ограниченных наборах данных. В течение 18 месяцев два врача-рентгенолога, специализирующиеся на нейровизуализации, ежемесячно оценивали 80 компьютерно-томографических исследований головного мозга, предварительно обработанных сервисом искусственного интеллекта и случайным образом выбранных из клинического потока. Результаты проанализировали методом ROC-анализа с вычислением таких метрик, как чувствительность, специфичность, точность, площадь под характеристической кривой.

Результаты. При клиническом мониторинге проанализировано 1200 компьютерных томограмм головного мозга, из которых признаки внутричерепного кровоизлияния выявлены в 48,3% случаев. По результатам их бинарной классификации на наличие внутричерепных кровоизлияний, выполненной сервисом искусственного интеллекта, получены следующие диагностические метрики: чувствительность — 97,4% (95,8–98,5), специфичность — 75,4% (71,8–78,7), точность — 86,0% (83,9–87,9), площадь под характеристической кривой — 94% (92,6–95,3). Со временем наблюдали статистически значимую умеренную положительную корреляцию между большинством диагностических метрик и временным показателем, за исключением чувствительности, что обусловлено сменой версии сервиса. Однако полное совпадение разметки и описания с заключением врача в выявленных сервисом искусственного интеллекта случаях внутричерепного кровоизлияния достигнуто в 28,5%, а различные расхождения найдены в 71,5%. Уточнённые метрики для случаев с полным соответствием заключения врача составили: чувствительность, специфичность, точность и площадь под характеристической кривой — 26,6, 73,8, 50,1 и 49,6% соответственно.

Заключение. Текущая конфигурация сервиса искусственного интеллекта позволяет исключить кровоизлияние с очень высокой вероятностью, что может быть полезно для первичной сортировки пациентов в условиях неотложной помощи. Однако низкие значения уточнённых метрик указывают на значительные расхождения между заключениями рентгенологов и результатами сервиса в аспектах детальной интерпретации патологии.

Ключевые слова: внутричерепное кровоизлияние; компьютерная томография; искусственный интеллект; диагностическая точность.

КАК ЦИТИРОВАТЬ:

Хоружая А.Н., Арзамасов К.М., Коденко М.Р., Кремнёва Е.И., Буренчев Д.В. Диагностика внутричерепных кровоизлияний по данным компьютерной томографии головного мозга с помощью искусственного интеллекта // Digital Diagnostics. 2025. Т. 6, № 2. С. XX–XX. DOI: 10.17816/DD645364 EDN: RFYVMC

Рукопись получена: 13.01.2025

Рукопись одобрена: 05.02.2025

Опубликована online: 02.06.2025

Статья доступна по лицензии CC BY-NC-ND 4.0 International

© Эко-Вектор, 2025

Diagnosis of intracranial haemorrhage based on brain computed tomography with artificial intelligence

Anna N. Khoruzhaya¹, Kirill M. Arzamasov¹, Maria R. Kodenko¹, Elena I. Kremneva^{1,2}, Dmitriy V. Burenchev¹

¹ Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow, Russia;

² Research Center of Neurology, Moscow, Russia

ABSTRACT

BACKGROUND: Intracranial haemorrhage (ICH) is associated with high mortality and disability and therefore requires prompt and accurate diagnosis, especially in the first 24 hours. The introduction of artificial intelligence (AI) in the analysis of brain CT studies has the potential to reduce diagnostic time and improve its quality, but existing solutions face problems of reproducibility of results and insufficient validation in real clinical settings. The relevance of the work is emphasised by the limited number of AI services certified in Russia for the detection of HFD and the lack of long-term data on their efficacy, which makes multicentre monitoring necessary to assess the stability and accuracy of such systems in dynamics.

AIM: To evaluate the diagnostic accuracy and stability of an artificial intelligence service for the diagnosis of HFD using native brain CT data in multicentre clinical monitoring for 18 months.

METHODS: This retrospective study, approved by the Moscow Radiological Society (ClinicalTrials NCT044899992), analyzed data from the Unified Radiology Information Service of Moscow (URIS EMIAS), covering 56 inpatients with anonymized brain CT scans. The AI underwent three testing phases: self-testing on 10 CTs, functional testing on 5 CTs, and calibration testing on 100 CTs with a balanced normal/pathology dataset. Monthly, two neuroradiologists assessed 80 randomly selected brain CTs processed by the AI. Metrics such as sensitivity (Se), specificity (Sp), accuracy (Ac), and area under the receiver operating characteristic curve (AUROC) were derived from ROC analysis.

RESULTS: Out of 1200 evaluated CTs, 580 (48.3%) showed ICH features. The AI's diagnostic metrics were Se 97.4% (95.8-98.5), Sp 75.4% (71.8-78.7), Ac 86.0% (83.9-87.9), and AUROC 94% (92.6 - 95.3)%. A moderate positive correlation was observed for all metrics over time, except Se, which changed with AI updates. However, only 162 CTs (28.5%) matched radiologist descriptions completely, while discrepancies occurred in 404 CTs (71.5%). Adjusted metrics for fully matching cases were Se 26.6%, Sp 73.8%, Ac 50.1%, and AUROC 49.6%.

CONCLUSION: The AI service effectively identifies hemorrhages, aiding urgent patient triage in emergency departments due to its high probability of exclusion for hemorrhage cases. However, the low refined metrics indicate significant discrepancies between radiologists and the AI regarding pathology details such as segmentation and description, highlighting areas for improvement in AI diagnostics.

Keywords: intracranial haemorrhage; computed tomography; artificial intelligence; diagnostic efficiency.

TO CITE THIS ARTICLE:

Khoruzhaya AN, Arzamasov KM, Kodenko MR, Kremneva EI, Burenchev DV. Diagnosis of intracranial haemorrhage based on brain computed tomography with artificial intelligence. *Digital Diagnostics*. 2025;6(2):XX–XX. DOI: 10.17816/DD645364 EDN: RFYVMC

Submitted: 13.01.2025

Accepted: 05.02.2025

Published online: 02.06.2025

The article can be used under the CC BY-NC-ND 4.0 International License

© Eco-Vector, 2025

ОБОСНОВАНИЕ

Внутричерепное кровоизлияние (ВЧК) — потенциально опасное для жизни острое состояние, связанное с экстравазацией крови в ткани головного мозга, которое может возникнуть либо спонтанно, либо в случае травмы головы или хирургического вмешательства. Частота случаев ВЧК составляет более 25 на 100 тыс. человеко-лет по данным мировой статистики [1]. В Российской Федерации (РФ) ежегодно диагностируют геморрагический инсульт у 43 тыс. пациентов, он составляет от 10 до 15% общего числа случаев острых нарушений мозгового кровообращения [2]. Этиология нетравматического ВЧК разнообразна и включает артериальную гипертензию, инсульт, разрыв аневризмы, васкулопатию, тромбоз венозного синуса, артериовенозную фистулу, злокачественные новообразования, применение антикоагулянтов и реже — воспалительные заболевания [1]. ВЧК связано с высоким риском смерти (до 40–50% случаев) в первые сутки и инвалидизацией выживших пациентов [1, 2]. Обычно среди ВЧК различают эпидуральную, субдуральную, субарахноидальную и паренхиматозную (внутримозговую) локализации. Их дифференцируют с учётом клинической картины, особенностей визуализации и прогноза [2]. Компьютерная томография головного мозга (КТ ГМ) — основной диагностический инструмент, который применяют для обследования пациентов, поступающих в отделения неотложной помощи с головной болью или очаговыми неврологическими нарушениями [2]. Это относительно доступный и быстрый метод, результаты которого позволяют оценить тяжесть патологического процесса и определить тактику лечения [3]. Важно отметить, что диагностика ВЧК в первые 24 часа имеет решающее клиническое значение для снижения ранней смертности. Это достигают за счёт оценки вовлечённых областей мозга, объёма кровоизлияния, наличия высокорисковых признаков визуализации (симптомы «вихря», «пятна»), а также учёта клинического контекста, что позволяет своевременно определить тактику ведения пациента и организовать последующее наблюдение [4, 5].

Технологии искусственного интеллекта (ИИ) активно и успешно применяют в неотложной нейровизуализации для диагностики острого ишемического инсульта, нейроинфекций или компрессии спинного мозга [6]. ИИ востребован и для задач первичной диагностики ВЧК, поскольку данная патология имеет хорошо дифференцируемые диагностические признаки [7]. Целесообразность внедрения технологий ИИ для диагностики ВЧК обусловлена необходимостью ускорения (с 512 до 19 мин) [7] процесса первичной диагностики и повышения его точности [8], в частности, за счёт сортировки пациентов и ранжирования рабочих списков рентгенологов [9, 10]. Однако применение ИИ в клинической практике возможно только при обеспечении высокой диагностической точности, которая определяется качеством обучающих и валидационных выборок [11], а также подтверждением полученных результатов на независимых внешних данных, что остаётся одной из ключевых проблем [12]. Одно из возможных её решений — предварительные и промежуточные проверки диагностических метрик на данных той медицинской информационной системы, где предполагают использовать сервис на основе технологий ИИ. Такие проверки позволяют проводить контроль качества для подтверждения диагностической эффективности сервисов ИИ с течением времени и дорабатывать их на основании регулярного анализа ошибок [13, 14]. Однако в зарубежной литературе представлено

лишь ограниченное число исследований, посвящённых длительному наблюдению за жизненным циклом сервисов на основе ИИ для диагностики ВЧК с регулярной оценкой их эффективности, тогда как в отечественных источниках такие данные отсутствуют.

В Москве с 2020 года проводят эксперимент по использованию технологий компьютерного зрения для анализа медицинских изображений в интересах городской системы здравоохранения (Эксперимент), в котором задействованы более 50 сервисов на основе ИИ по 41 диагностическому направлению¹. В РФ статусом «медицинское изделие» обладает 4 программных продукта на основе технологий ИИ, предназначенных для автоматического анализа цифровых медицинских изображений КТ ГМ на предмет наличия ВЧК. В апреле 2022 г. в Эксперименте начал работать первый такой диагностический сервис, который долгое время был единственным.

ЦЕЛЬ

Оценить диагностическую точность и устойчивость сервиса ИИ для диагностики ВЧК по данным нативной КТ ГМ в условиях многоцентрового клинического мониторинга на протяжении 18 мес.

МЕТОДЫ

ДИЗАЙН ИССЛЕДОВАНИЯ

Проведён ретроспективный многоцентровой клинический мониторинг (18 мес.)

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Объект исследования — программное обеспечение ЦЕЛЬС® (ПО ЦЕЛЬС®, Россия) ООО «Медицинские скрининг системы» (ИИ-сервис). На этапе подачи заявки на участие в Эксперименте в апреле 2022 г. данный ИИ-сервис обучили на массиве из более 15 тыс. анонимизированных диагностических результатов исследования, полученных из двух медицинских учреждений. Их ранжировали по признаку «норма» и «патология» (ВЧК). В обучающую выборку включали все результаты КТ со следующими типами кровоизлияний:

- субдуральное;
- внутримозговое;
- эпидуральное;
- субарахноидальное.

Количество изображений «с патологией» для обучения составило 60% общего объёма выборки. С целью определения исходных метрик точности работы ИИ-сервис прошёл предварительные клинично-технические испытания, что было необходимо для участия в Эксперименте. Для их проведения сформирован независимый набор данных, не использовавшийся в обучающей выборке, включающий 260 исследований КТ (КТ-исследование): 130 с выявленной патологией и 130 — с «нормой». На основе полученных данных сформировали таблицу с распределением исследований по типам диагностических результатов (ложноположительные, ложноотрицательные, истинно положительные и истинно отрицательные), после чего рассчитали метрики аналитической валидации. Получены следующие средние показатели диагностической точности: площадь под кривой (Area Under the Curve, AUROC) — 0,89; чувствительность — 0,84; специфичность — 0,74; точность — 0,79.

Входные данные (КТ-исследования ГМ) для обработки представлены в формате DICOM (Digital Imaging and Communications in Medicine). Результаты обработки включали:

- текстовое описание (DICOM SR);
- изображения с разметкой патологических областей (DICOM SC);
- числовую оценку вероятности наличия патологии в исследовании.

Результаты анализа, выполненного ИИ-сервисом, автоматически интегрированы и доступны вместе с исходными изображениями в Единой радиологической информационной системе (ЕРИС), которая входит в состав государственной информационной системы в сфере здравоохранения субъекта Российской Федерации — «Единая медицинская информационно-аналитическая система города Москвы» (ЕМИАС).

¹ Технологии искусственного интеллекта в здравоохранении. В: Центр диагностики и телемедицины [интернет]. Москва: Центр диагностики и телемедицины, 2020–2024. Режим доступа: <https://mosmed.ai/> Дата обращения: 13.12.2024.

Для исследования использовали бинарную классификацию: определяли вероятность наличия патологии (ВЧК). Оценку её корректности в отношении отдельных типов кровоизлияний, а также точности сегментации и подсчёта объёмов в настоящем исследовании не проводили.

ТЕСТИРОВАНИЕ СЕРВИСА НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Согласно разработанной и апробированной в рамках Эксперимента методологии тестирования и мониторинга программного обеспечения на основе технологий ИИ [15], ИИ-сервис прошёл трёхэтапную проверку перед подключением к основному контуру ЕРИС ЕМИАС для обработки КТ-исследований.

- На этапе самотестирования успешно проведено тестирование технической совместимости алгоритма с обрабатываемыми данными.
- При функциональном тестировании оценивали полноту и достаточность инструментария ИИ-сервиса, а также возможность выполнения диагностической задачи.
- Калибровочное тестирование проводили для оценки клинической производительности и диагностических метрик сервиса.

На этапе функционального тестирования проверку ИИ-сервиса осуществляли с технической и клинической точки зрения, оценивали наличие и работоспособность его функционала в соответствии с базовыми диагностическими² и функциональными требованиями³, которые разработали в рамках Эксперимента. Базовые диагностические требования включают обязательное и опциональное содержание ответа ИИ-сервиса, а также форму его представления. При решении клинической задачи диагностики ВЧК по данным КТ ГМ ИИ-сервис должен обязательно указывать вероятность наличия кровоизлияний в виде числового значения, тип/типы кровоизлияния (эпидуральное, субдуральное, субарахноидальное, внутримозговое), их объёмы (кроме субарахноидального), сегментировать зоны патологической плотности на самих исследованиях. Базовые функциональные требования определяют, какие изображения должен обрабатывать ИИ-сервис, а также содержательные и формальные характеристики предоставляемых результатов.

Этап калибровочного тестирования, следовавший за функциональным, направлен на подтверждение или опровержение заявленных разработчиком метрик эффективности работы ИИ-сервиса. Подсчитывали количество истинно положительных и отрицательных, ложноположительных и -отрицательных ответов в виде четырёхпольной таблицы, из чего рассчитывали основные метрики для оценки — AUROC, чувствительность, специфичность, точность, удельный вес ложноотрицательных и -положительных результатов. Кроме того, фиксировали минимальное, среднее и максимальное время анализа одного КТ-исследования. В качестве эталона в данной клинической задаче принято:

- значение AUROC — не менее 0,81;
- время, затрачиваемое на принятие, обработку изображения и передачу результатов анализа — не более 6,5 мин.;
- удельный вес успешно обработанных исследований — не менее 90% [16, 17].

В течение 18 мес. провели суммарно три калибровочных тестирования с использованием набора данных, содержащего КТ-исследования с балансом классов 1:1. После каждого калибровочного тестирования ИИ-сервиса формировали протокол, в который входили информация о его наименовании, типе, компании-поставщике, а также данные об обработанных изображениях, полученные метрики и решение о том, соответствует ли данный ИИ-сервис принятым эталонным значениям для допуска к дальнейшей работе в ЕРИС ЕМИАС.

Самотестирование проводили на предоставленных в свободном доступе обезличенных диагностических изображениях в формате DICOM с приложением файла в формате Excel, где дополнительно были указаны модальность, тип диагностической процедуры, производитель и модель диагностического устройства. Функциональное и калибровочное тестирования проводили

² Базовые диагностические требования к результатам работы ИИ-сервисов [интернет]. В: Научно-практический клинический центр диагностики и телемедицинских технологий, 2024–2024. Режим доступа: <https://mosmed.ai/ai/docs/>
Дата обращения: 13.12.2024.

³ Базовые функциональные требования к результатам работы ИИ-сервисов [интернет]. В: Научно-практический клинический центр диагностики и телемедицинских технологий, 2024–2024. Режим доступа: <https://mosmed.ai/ai/docs/>
Дата обращения: 13.12.2024.

с использованием эталонного набора данных MosMedData⁴. Для функционального тестирования использовали 5 КТ-исследований (с патологией, «нормой» и артефактом — 2, 2, 1 соответственно), для калибровочного — 100 КТ-исследований (по 50 с патологией и без).

КЛИНИЧЕСКИЙ МОНИТОРИНГ

ИИ-сервис подключили на потоковую обработку КТ-исследований ГМ 28.04.2022 из 56 медицинских организаций стационарной медицинской помощи. Данные о результатах их обработки собирали с апреля 2022 по сентябрь 2023 г. Общее число обработанных КТ-исследований ГМ составило 191 928. Каждый месяц для экспертной оценки случайным образом отбирали 80 КТ-исследований с балансом классов 70:30 (с предполагаемой патологией и нормой — 70 и 30% соответственно по оценке ИИ), согласно разработанной методологии [18].

Два врача-рентгенолога, специализирующиеся на нейровизуализации, со стажем более 3 лет оценивали эти КТ-исследования по двум главным критериям:

- соответствие трактовки (заключения);
- соответствие локализации (маркировки) патологической зоны.

Каждый из критериев мог иметь четыре вариации ответов:

- полное соответствие;
- частично корректная оценка;
- ложноположительные результаты — когда сервис обнаружил кровоизлияние там, где его нет;
- ложноотрицательные результаты — когда сервис не обнаружил кровоизлияние при его наличии.

ЭТИЧЕСКАЯ ЭКСПЕРТИЗА

Дизайн Эксперимента одобрен независимым этическим комитетом Московского рентгенологического общества (выписка из протокола № 2 НЭК МРО РОПР от 20.02.2020), также зарегистрирован на ClinicalTrials (NCT04489992).

СТАТИСТИЧЕСКАЯ ОБРАБОТКА

Для обработки полученных данных использовали метод построения и анализа характеристической кривой (ROC-анализ) посредством специально разработанного Web-инструмента⁵. Согласно результатам эмпирического исследования, минимальный объём набора данных для тестирования ИИ-сервиса в условиях периодического мониторинга должен составлять 400 КТ-исследований, при этом доля патологических случаев — не менее 10% [19]. Однако фактический объём набора данных оказался больше и составил 1200 КТ-исследований ГМ, при доле патологии 48,3%, что соответствовало нашим задачам. Для ИИ-сервиса определены метрики диагностической точности: чувствительность, специфичность, точность, AUROC. Ввиду наличия бинарного ответа от сервиса AUROC рассчитывали с учётом полученных значений чувствительности и специфичности. При расчёте метрик за ложноположительный принимали ответ сервиса о наличии ВЧК в исследовании при отсутствии данной патологии по мнению врача-эксперта, за ложноотрицательный — ответ сервиса об отсутствии ВЧК в исследовании при его наличии по мнению врача-эксперта. Общие метрики ИИ-сервиса, представленные в результатах, рассчитаны с 95% доверительным интервалом (ДИ) методом биномиального теста, поскольку исследуемые выборки содержат бинаризованные значения. Для оценки наличия и характера взаимосвязи между значением диагностических метрик и временем работы сервиса использовали коэффициент корреляции Пирсона (r). С целью сравнения значений диагностических метрик между калибровочными тестированиями применяли тест Манна–Уитни. При сопоставлении использовали одностороннюю версию теста с альтернативной гипотезой «медианное значение метрики до третьего калибровочного тестирования меньше, чем после него». Мы ожидали статистически

⁴ Свидетельство о государственной регистрации базы данных № 2022620559/ 16.03.2022. Бюл. № 3. Морозов С.П., Павлов Н.А., Петрайкин А.В., и др. MosMedData: набор диагностических компьютерно-томографических изображений головного мозга с наличием и отсутствием признаков внутричерепного кровоизлияния. Режим доступа: <https://www.elibrary.ru/item.asp?id=48137428> Дата обращения: 13.12.2024.

⁵ Инструмент для ROC-анализа диагностических тестов [интернет]. В: Научно-практический клинический центр диагностики и телемедицинских технологий, 2022–2024. Режим доступа: <https://roc-analysis.mosmed.ai/> Дата обращения: 13.12.2024.

значимый прирост метрических показателей после третьего калибровочного тестирования. Уровень значимости принятия статистических гипотез был равным 0,05.

РЕЗУЛЬТАТЫ

КАЛИБРОВОЧНОЕ ТЕСТИРОВАНИЕ

Количество КТ-исследований, не прошедших обработку по причине технической ошибки «сбой отправки исследования на тестирование» составило 9, 2 и 1 на первом, втором и третьем этапах калибровочного тестирования соответственно. По результатам анализа корректно обработанных КТ-исследований ГМ построены характеристические кривые, представленные на [рис. 1](#). В [табл. 1](#) представлены численные метрики, которые мы получили при первом, втором и третьем калибровочных тестированиях.

Необходимость их повтора была обусловлена доработками программного обеспечения, при котором его метрики эффективности могли ухудшиться. Калибровочные тестирования проводили после каждой доработки, которая сопровождалась изменениями ядра ИИ-сервиса. Однако во всех трёх случаях отличия в метриках эффективности были статистически незначимыми ($p > 0,05$).

КЛИНИЧЕСКИЙ МОНИТОРИНГ

Результаты клинического мониторинга представлены в [табл. 2](#).

Из представленных данных следует, что ИИ-сервис стабилизировал свои показатели после третьего месяца. Первые 3 мес. (апрель-июнь 2022 г.) были периодом апробации, в ходе которого разработчик дорабатывал своё решение — обработка исследований велась нестабильно, присутствовало большое количество технических ошибок, что влияло на результаты работы. Именно поэтому мы решили не включать данный период в дальнейший анализ.

Общая схема проведения клинического мониторинга ИИ-сервиса представлена на [рис. 2](#).

С момента стабилизации его работы суммарно проанализировано 1200 нативных изображений КТ ГМ (см. [рис. 2](#)). Средний возраст пациентов составил $61,2 \pm 18,6$ года, из них 39% женщин. По результатам экспертной оценки (Ground Truth) признаки ВЧК содержали 580 КТ-исследований (48,3%). Результаты расчёта диагностических метрик за весь период с 3 по 18 мес. включительно представлены на [рис. 3](#). По результатам бинарной классификации, проведённой ИИ-сервисом для выявления ВЧК, получены следующие значения метрик эффективности:

- чувствительность — 97,4% (95,8–98,5);
- специфичность — 75,4% (71,8–78,7);
- точность — 86,0% (83,9–87,9);
- AUROC — 94% (92,6–95,3).

Корреляционный анализ выявил статистически значимую умеренную положительную корреляционную связь между временем работы и значениями диагностических метрик — специфичность и точность (для обеих метрик: $r=0,5$; $p=0,04$), AUROC ($r=0,6$; $p=0,03$). Для чувствительности статистически значимой взаимосвязи между её значением и временем работы ИИ-сервиса не выявлено.

Сравнение значений диагностических метрик между вторым и третьим калибровочными тестированиями, а также последующим этапом показало статистически значимый прирост только для чувствительности и специфичности ($p=0,04$).

Кроме того, мы оценили уточнённые диагностические метрики. Для их расчёта дополнительно проверено наличие не только факта обнаружения патологии, но и характер её локализации, а также соответствие описания врача ответу сервиса. В этом случае истинно положительным считали любой ответ, в котором присутствовало совпадение и по локализации, и по описанию детектированной патологии. Полное соответствие разметки и описания при наличии ВЧК достигнуто в 28,5% случаях (162 КТ-исследования). Соответственно, различия между разметкой и описанием отмечены в 71,5% (404 КТ-исследования). Результаты для уточнённых метрик эффективности следующие:

- чувствительность — 26,6% (22,9–30,4);
- специфичность — 73,8% (70,0–77,4);
- точность — 50,1% (47,1–53,0);
- AUROC — 49,6% (44,1–55,1).

Неточности в описании обнаружены в 61 случае — ИИ-сервис корректно идентифицировал наличие патологии на снимках, но либо пропускал отдельные очаги кровоизлияний при их множественности, либо ошибочно классифицировал его тип. Неточности разметки зафиксированы в 64 КТ-исследованиях, что включало некорректную сегментацию участков кровоизлияний. Однако наиболее часто неточность наблюдали одновременно и в описаниях, и в разметке (279 КТ-исследований) (см. рис. 2).

Одним из наиболее частых примеров частично корректного срабатывания можно назвать случаи, когда при множественных кровоизлияниях ИИ-сервис выделял один тип кровоизлияния, но пропускал другой. Так, на рис. 4, а слева пропущено внутримозговое кровоизлияние в левом полушарии, а на рис. 4, с слева ИИ-сервис корректно его сегментировал в левом полушарии, но не отметил внутрижелудочковое кровоизлияние в правом полушарии, а также межполушарное субарахноидальное кровоизлияние и по бороздам в обоих полушариях. Также частым случаем некорректных срабатываний является частичная сегментация кровоизлияния с ошибочным определением его типа (см. рис. 4 с) — справа представлено субдуральное кровоизлияние, которое ИИ-сервис идентифицировал как субарахноидальное. Гораздо реже среди примеров частично корректного срабатывания можно отметить полностью корректную сегментацию патологической зоны кровоизлияния, но неверное определение его/их типа (см. рис. 4, b) — справа внутримозговое кровоизлияние ошибочно классифицировано как субдуральное, а слева субдуральное кровоизлияние отнесено к субарахноидальному).

Всего отмечено 152 ложноположительных результата. Наиболее частыми их причинами являлись сегментации магистральных артерий, венозных синусов и частично кальцифицированных оболочек мозга (рис. 5, a).

ИИ-сервис не выявил патологические изменения на 14 изображениях КТ ГМ. Наиболее часто встречали ложноотрицательные результаты при субарахноидальных кровоизлияниях — 8 случаев (см. рис. 5, b, в центре и справа). Реже наблюдали пропуски внутрижелудочковых кровоизлияний — 2 случая (см. рис. 5, b, слева), внутримозговых кровоизлияний — 2 случая, а также по одному случаю субдурального и эпидурального кровоизлияния.

ОБСУЖДЕНИЕ

Проведённое исследование не является первым среди клинической апробации ИИ-сервисов для выявления ВЧК по данным бесконтрастной КТ ГМ. Однако его отличительной особенностью является длительность (18 мес.) и многоцентровой характер (56 стационаров, 248 врачей-рентгенологов, составивших первичное заключение).

В нашем исследовании в рамках клинического мониторинга отмечен статистически значимый прирост медианных значений для двух из четырёх основных диагностических метрик. Причём следует отметить различную величину прироста. Изначально довольно высокие показатели чувствительности (Me 95,8%) со временем наблюдения статистически значимо ($p=0,04$) повысились (Me 97,4%). Для специфичности были характерны сравнительно низкие начальные значения (Me 69,4%) и статистически значимый ($p=0,04$) прирост (Me 76,2%). Для точности и AUROC статистически значимых изменений не наблюдали ($p=0,1$).

Стоит отметить различия в значениях метрик эффективности, полученных при калибровочных тестированиях и клиническом мониторинге. Их расчёт при калибровочных тестированиях проводили с использованием индекса Юдена при установленном оптимальном пороге в 75%. Выявлена статистически значимая положительная корреляция между временем работы сервиса и специфичностью, точностью, AUROC (r от 0,5 до 0,6), при этом статистически значимый прирост значения зафиксирован только для специфичности и чувствительности.

ВЧК — неотложная патология, в большинстве случаев требующая быстрого реагирования и повышающая риски ухудшения состояния пациентов. Именно поэтому приоритетной задачей является настройка ИИ-сервиса на максимальную чувствительность к наличию патологии, что позволяет в первую очередь привлекать внимание врача к пациентам с подозрением на ВЧК. Однако автоматическая подстройка порога для оптимизации всех диагностических параметров должна происходить при условии, что это не приводит к критическому снижению специфичности. В нескольких одноцентровых исследованиях продемонстрирована высокая специфичность, варьирующая от 91 до 98%, однако чувствительность оказывалась ниже — в диапазоне от 81 до 94% [20–23].

В единичных исследованиях, проводимых на базе нескольких клинических центров и сопоставимых с настоящим по условиям и объёму, также отмечают более низкие значения чувствительности при высокой специфичности [24–27]. Так, J. McLouth и соавт. [24] при использовании коммерчески доступного сервиса ИИ CINA[®] v1.0 (Avicenna.ai, Франция) сообщили о чувствительности и специфичности — 91,4 и 97,5% соответственно, полученных на выборке 814 пациентов с долей патологии 31%. Похожие показатели эффективности продемонстрированы в работе A. Kundisch и соавт. [25], где при использовании коммерческого программного обеспечения для выявления ВЧК (AIDOC, Израиль) получены значения чувствительности и специфичности — 87,6 и 92,8% соответственно. Исследование провели на выборке из 4946 пациентов с долей патологии 5%. Два крупных недавних исследования — A. J. Del Gaizo и соавт. [26] с выборкой 58 321 (доля патологии 2,7%) и G. Pettet и соавт. [27] с выборкой 1315 (доля патологии 8,5%) — проведённые также с применением коммерческих сервисов ИИ [CINA[®] v1.0 (Avicenna.ai, Франция) и qER[®] v2.0 (Qure.ai, Индия)], предоставили следующие значения чувствительности и специфичности: 75,6 и 92,1; 85,7 и 94,3% соответственно.

По данным систематического обзора S.M. Mäenpää и соавт. [28], посвящённого оценке диагностической точности моделей ИИ для экстренного анализа результатов КТ ГМ в условиях внешней клинической апробации, установлено, что большинство коммерчески доступных сервисов ИИ имели, как правило, более низкую чувствительность и положительную прогностическую значимость, которая отражает количество ложноположительных ответов. Это свидетельствует о более слабой обобщаемости и не совсем подходит для сценария сортировки и приоритизации рабочего списка посредством пометки ВЧК-положительных сканов из-за более высокого риска «утомления бдительности». В нашем исследовании мы получили высокие показатели чувствительности, однако специфичность на уровне 75,4% обусловлена большим количеством ложноположительных результатов и говорит о вероятной дополнительной нагрузке на внимание врача. В некоторых исследованиях снижение прогностической значимости объясняли более низкой распространённостью целевого состояния, что отражает естественное влияние базовой частоты патологии на метрики [23, 29]. Приемлемый уровень ложноположительных ответов ИИ-сервиса на потоке требует дальнейшего изучения.

Об ограничениях в использовании всего функционала ИИ-сервиса (корректное определение типа кровоизлияний и их локализация) говорят и довольно низкие (за исключением специфичности) значения уточнённых метрик. Тем не менее нужно отметить, что представленные метрики рассчитаны исключительно для случаев, в которых ответ ИИ-сервиса полностью совпадал с мнением врача-эксперта как по разметке, так и по типу кровоизлияния. Общие же показатели метрик эффективности достаточно высокие, в частности, чувствительность превышает значения, описанные в литературе.

Некоторые авторы отмечают превосходную согласованность разметки объёмов патологических областей плотности вследствие ВЧК в ГМ между коммерчески доступным сервисом ИИ и полуавтоматической разметкой. Однако N. Schmitt и соавт. [30] отмечают, что при чувствительности и специфичности 91 и 89% соответственно (на выборке с долей патологии 50%) ИИ может служить «вторым мнением» для врача, но не подходит для самостоятельного использования, с чем согласны и другие авторы. Наши данные демонстрируют, что полное совпадение описаний и разметки патологических областей, выполненных исследуемым ИИ-сервисом, с заключениями врачей невозможно достигнуть даже в 1/3 всех случаев с патологией, что в значительной степени связано с сочетанием различных типов кровоизлияний. Именно поэтому ИИ-сервис можно применять как вспомогательный инструмент для первичного выявления патологии, однако пока он не будет существенно доработан в части качества разметки и классификации кровоизлияний, нецелесообразно его использование в виде полноценной системы поддержки принятия врачебных решений, предназначенной для углублённой диагностики типов кровоизлияний и их объёмов.

Постоянная доработка ИИ-сервисов и адаптация к изменяющимся клиническим условиям необходима и технически возможна [31]. Использование в качестве основы глубоких свёрточных нейронных сетей позволяет этому программному обеспечению эффективнее извлекать и анализировать сложные признаки изображений, которые недоступны глазу человека, сопоставляя их в контексте отличной от человеческой логики, что действительно способно приводить к повышению точности диагностики [26, 32]. Обучение на новых данных способствует улучшению производительности с течением времени [33], и в нашем исследовании с периодом наблюдения

более года это продемонстрировано. Кроме того, необходимо обеспечить регулярный — желательно ежемесячный — контроль производительности ИИ-сервиса на независимых клинических данных [21], а также организовать систематическую обратную связь от врачей. Это позволит разработчикам определить необходимость в дополнительных обучающих данных и откорректировать пороговые значения, оптимизируя баланс между чувствительностью и точностью [32].

Существует ещё одна важная причина, по которой необходим непрерывный клинический мониторинг качества работы ИИ-сервиса. Наши данные показывают, что он обеспечивает более объективную оценку, чем тестирование в лабораторных условиях, даже при использовании внешних валидационных данных (в нашем случае — калибровочные тестирования). Кроме того, относительно низкий показатель специфичности должен побуждать врачей к осторожному использованию ИИ-сервиса в реальной клинической практике без должного контроля, поскольку это может привести к увеличению числа необоснованных госпитализаций или даже к ненужным хирургическим вмешательствам [34]. Одновременно с этим высокая чувствительность даёт возможность с высокой точностью исключать кровотечения в случае острого ишемического инсульта, что определяет пригодность пациента для проведения тромболитической терапии [27]. Таким образом, оптимальное использование ИИ рентгенологами может быть дополнено пониманием сценариев, в которых он, вероятно, будет генерировать неточные выходные данные. Кроме того, подсчёт объёмов кровоизлияний с помощью ИИ-сервиса (при наличии соответствующего функционала) может служить эффективным способом объективизации корректности разметки. Данную задачу необходимо рассматривать в рамках дальнейших исследований.

ОГРАНИЧЕНИЯ ИССЛЕДОВАНИЯ

Наше исследование имеет несколько ограничений. Во-первых, мы не анализировали диагностические метрики ИИ-сервиса по каждому типу кровоизлияний и точность площади маркировки патологических областей. Наша концепция наблюдательного ретроспективного многоцентрового анализа направлена на выявление изменений в его производительности во времени на основе фактической клинической работы. Во-вторых, наша выборка для клинического мониторинга была насыщена ВЧК (~50%) и не соответствовала реальной распространённости патологии в популяции (~8–12%), что могло способствовать увеличению количества ложноположительных результатов и снижению специфичности по сравнению с теми метриками эффективности, которые заявлены производителем. Это подчёркивает необходимость стандартизации клинически ориентированного обучения и проверки ИИ в соответствующих условиях. Тем не менее сохраняющаяся высокая чувствительность ИИ-сервиса, даже при увеличении доли случаев с патологией в выборке, свидетельствует о его высокой способности выявлять критически значимые патологические изменения, что в аспекте неотложной медицинской помощи следует рассматривать как преимущество.

ЗАКЛЮЧЕНИЕ

В ходе 18-месячного ретроспективного наблюдения за работой ИИ-сервиса для обнаружения ВЧК при данным неконтрастной КТ ГМ в 56 стационарах г. Москвы удалось продемонстрировать многообещающие результаты с очень высокой чувствительностью (97,4%) и разумной специфичностью (75,4%), которые с течением времени улучшались. Однако низкие значения уточнённых метрик (чувствительность и точность — 26,6 и 50,1% соответственно) указывают на значительные расхождения между оценками рентгенологов и ИИ-сервиса, связанные с неполной сегментацией патологических зон и ошибками в классификации типов ВЧК. Рентгенологам необходимо понимать особенности работы ИИ в клинической практике и учитывать, что положительный результат не всегда свидетельствует о наличии кровоизлияния, а выявленное кровоизлияние может быть не единственным и не всегда точно сегментировано. Разработчикам подобного программного обеспечения необходимо сосредоточиться на снижении количества ложноположительных ответов и повышении качества работы ИИ-сервиса, чтобы его функции были клинически полезными. Тем не менее текущая конфигурация позволяет исключать кровоизлияние с очень высокой вероятностью, что особенно полезно для неотложной сортировки пациентов в приёмных отделениях.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Вклад авторов. А.Н. Хоружая — сбор и анализ литературных данных, тестирование сервиса искусственного интеллекта, анализ данных мониторинга, написание и редактирование текста рукописи; К.М. Арзамасов — концепция исследования, организация тестирования сервиса искусственного интеллекта, сбор данных для мониторинга, редактирование текста рукописи; М.Р. Коденко — статистическая обработка данных, написание и редактирование текста рукописи; Е.И. Кремнева — сбор и анализ литературных данных, анализ данных мониторинга, редактирование текста рукописи; Д.В. Буренчев — концепция исследования, анализ литературных данных, редактирование текста рукописи. Все авторы одобрили рукопись (версию для публикации), а также согласились нести ответственность за все аспекты работы, гарантируя надлежащее рассмотрение и решение вопросов, связанных с точностью и добросовестностью любой её части.

Благодарности. Авторы выражают благодарность главному научному сотруднику Научно-практического клинического центра диагностики и телемедицинских технологий, д.м.н. А.В. Петряйкину за помощь в проведении исследования.

Этическая экспертиза. Дизайн Эксперимента одобрен независимым этическим комитетом Московского рентгенологического общества (выписка из протокола № 2 НЭК МРО РОПР от 20.02.2020), также зарегистрирован на ClinicalTrials (NCT04489992).

Источники финансирования. Данная статья подготовлена авторским коллективом в рамках научно-исследовательской работы «Научные методологии устойчивого развития технологий искусственного интеллекта в медицинской диагностике», (ЕГИСУ: № 123031500004-5).

Раскрытие интересов. Авторы заявляют об отсутствии отношений, деятельности и интересов за последние три года, связанных с третьими лицами (коммерческими и некоммерческими), интересы которых могут быть затронуты содержанием статьи.

Оригинальность. При создании настоящей работы авторы не использовали ранее опубликованные сведения (текст, иллюстрации, данные).

Доступ к данным. Редакционная политика в отношении совместного использования данных к настоящей работе не применима.

Генеративный искусственный интеллект. При создании настоящей статьи технологии генеративного искусственного интеллекта не использовали.

Рассмотрение и рецензирование. Настоящая работа подана в журнал в инициативном порядке и рассмотрена в соответствии с процедурой fast-track. В рецензировании участвовали три внешних рецензента и научный редактор издания.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

1. Li X, Zhang L, Wolfe CDA, Wang Y. Incidence and long-term survival of spontaneous intracerebral hemorrhage over time: a systematic review and meta-analysis. *Frontiers in Neurology*. 2022;13:819737. doi: [10.3389/fneur.2022.819737](https://doi.org/10.3389/fneur.2022.819737) EDN: [MLOQRJ](https://doi.org/10.17816/DD645364)
2. Hemorrhagic stroke: clinical guidelines. Moscow: Ministry of Health of the Russian Federation; 2022. (In Russ.) [cited 2024 Dec 12]. Available from: <https://ruans.org/Text/Guidelines/hemorrhagic-stroke-2022.pdf>
3. Hostettler IC, Seiffge DJ, Werring DJ. Intracerebral hemorrhage: an update on diagnosis and treatment. *Expert Review of Neurotherapeutics*. 2019;19(7):679–694. doi: [10.1080/14737175.2019.1623671](https://doi.org/10.1080/14737175.2019.1623671) EDN: [JWSYUZ](https://doi.org/10.17816/DD645364)
4. Woo D, Comeau ME, Venema SU, et al. Risk factors associated with mortality and neurologic disability after intracerebral hemorrhage in a racially and ethnically diverse cohort. *JAMA Network Open*. 2022;5(3):e221103. doi: [10.1001/jamanetworkopen.2022.1103](https://doi.org/10.1001/jamanetworkopen.2022.1103) EDN: [BVHNLU](https://doi.org/10.17816/DD645364)
5. Yaghi S, Dibu J, Achi E, et al. Hematoma expansion in spontaneous intracerebral hemorrhage: predictors and outcome. *International Journal of Neuroscience*. 2014;124(12):890–893. doi: [10.3109/00207454.2014.887716](https://doi.org/10.3109/00207454.2014.887716)
6. Gong B, Khalvati F, Ertl-Wagner BB, Patlas MN. Artificial intelligence in emergency neuroradiology: current applications and perspectives. *Diagnostic and Interventional Imaging*. 2025;106(4):135–142. doi: [10.1016/j.diii.2024.11.002](https://doi.org/10.1016/j.diii.2024.11.002) EDN: [DHXSGS](https://doi.org/10.17816/DD645364)

7. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *npj Digital Medicine*. 2018;1(1):9. doi: [10.1038/s41746-017-0015-z](https://doi.org/10.1038/s41746-017-0015-z) EDN: [BORIWC](#)

8. Seyam M, Weikert T, Sauter A, et al. Utilization of artificial intelligence-based intracranial hemorrhage detection on emergent noncontrast CT images in clinical workflow. *Radiology: Artificial Intelligence*. 2022;4(2):e210168. doi: [10.1148/ryai.210168](https://doi.org/10.1148/ryai.210168) EDN: [HEPSBX](#)

9. Davis MA, Rao B, Cedeno PA, et al. machine learning and improved quality metrics in acute intracranial hemorrhage by noncontrast computed tomography. *Current Problems in Diagnostic Radiology*. 2022;51(4):556–561. doi: [10.1067/j.cpradiol.2020.10.007](https://doi.org/10.1067/j.cpradiol.2020.10.007) EDN: [NHQFYC](#)

10. O'Neill TJ, Xi Y, Stehel E, et al. Active reprioritization of the reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. *Radiology: Artificial Intelligence*. 2021;3(2):e200024. doi: [10.1148/ryai.2020200024](https://doi.org/10.1148/ryai.2020200024) EDN: [LCDGTM](#)

11. Smorchkova AK, Khoruzhaya AN, Kremneva EI, Petryaikin AV. Machine learning technologies in CT-based diagnostics and classification of intracranial hemorrhages. *Burdenko's Journal of Neurosurgery*. 2023;87(2):85. doi: [10.17116/neuro20238702185](https://doi.org/10.17116/neuro20238702185) EDN: [JVZDST](#)

12. Yu KH, Kohane IS. Framing the challenges of artificial intelligence in medicine. *BMJ Quality & Safety*. 2018;28(3):238–241. doi: [10.1136/bmjqs-2018-008551](https://doi.org/10.1136/bmjqs-2018-008551)

13. Allen B, Dreyer K, Stibolt R, et al. Evaluation and real-world performance monitoring of artificial intelligence models in clinical practice: try it, buy it, check it. *Journal of the American College of Radiology*. 2021;18(11):1489–1496. doi: [10.1016/j.jacr.2021.08.022](https://doi.org/10.1016/j.jacr.2021.08.022) EDN: [NMKGVD](#)

14. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *European Radiology*. 2020;30(6):3576–3584. doi: [10.1007/s00330-020-06672-5](https://doi.org/10.1007/s00330-020-06672-5) EDN: [WWDEXB](#)

15. Vasiliev YuA, Vlazimirskyy AV, Omelyanskaya OV, et al. Methodology for testing and monitoring artificial intelligence-based software for medical diagnostics. *Digital Diagnostics*. 2023;4(3):252–267. doi: [10.17816/DD321971](https://doi.org/10.17816/DD321971) EDN: [UEDORU](#)

16. Morozov SP, Vladzimirsky AV, Klyashtornyy VG, et al. Clinical acceptance of software based on artificial intelligence technologies (radiology). Moscow: Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 2019. EDN: [GWJIMI](#)

17. Morozov SP, Vladzimirsky AV, Andreychenko AE, et al. Regulations for the preparation of data sets with a description of approaches to the formation of a representative data sample. Moscow: Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 2022. (In Russ.) EDN: [XENAJE](#)

18. Chetverikov SF, Arzamasov KM, Andreichenko AE, et al. Approaches to sampling for quality control of artificial intelligence in biomedical research. *Sovremennye tehnologii v medicine*. 2023;15(2):19. doi: [10.17691/stm2023.15.2.02](https://doi.org/10.17691/stm2023.15.2.02) EDN: [FUKXYC](#)

19. Kodenko MR, Bobrovskaya TM, Reshetnikov RV, et al. Empirical approach to sample size estimation for testing of AI algorithms. *Doklady Mathematics*. 2024;110(S1):S62–S74. doi: [10.1134/S1064562424602063](https://doi.org/10.1134/S1064562424602063) EDN: [VJHJRD](#)

20. Salehinejad H, Kitamura J, Ditkofsky N, et al. A real-world demonstration of machine learning generalizability in the detection of intracranial hemorrhage on head computerized tomography. *Scientific Reports*. 2021;11(1):17051. doi: [10.1038/s41598-021-95533-2](https://doi.org/10.1038/s41598-021-95533-2) EDN: [SXMLCH](#)

21. Zia A, Fletcher C, Bigwood S, et al. Retrospective analysis and prospective validation of an AI-based software for intracranial haemorrhage detection at a high-volume trauma centre. *Scientific Reports*. 2022;12(1):19885. doi: [10.1038/s41598-022-24504-y](https://doi.org/10.1038/s41598-022-24504-y) EDN: [IWNBET](#)

22. Ginat DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. *Neuroradiology*. 2019;62(3):335–340. doi: [10.1007/s00234-019-02330-w](https://doi.org/10.1007/s00234-019-02330-w) EDN: [WTOITQ](#)

23. Voter AF, Meram E, Garrett JW, Yu JPY. Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of intracranial hemorrhage. *Journal of the American College of Radiology*. 2021;18(8):1143–1152. doi: [10.1016/j.jacr.2021.03.005](https://doi.org/10.1016/j.jacr.2021.03.005) EDN: [GPJYDS](#)

24. McLouth J, Elstrott S, Chaibi Y, et al. Validation of a deep learning tool in the detection of intracranial hemorrhage and large vessel occlusion. *Frontiers in Neurology*. 2021;12:656112. doi: [10.3389/fneur.2021.656112](https://doi.org/10.3389/fneur.2021.656112) EDN: [FFIXVV](#)

25. Kundisch A, Hönning A, Mutze S, et al. Deep learning algorithm in detecting intracranial hemorrhages on emergency computed tomographies. *PLOS ONE*. 2021;16(11):e0260560. doi: [10.1371/journal.pone.0260560](https://doi.org/10.1371/journal.pone.0260560) EDN: QPACKZ

26. Del Gaizo AJ, Osborne TF, Shahoumian T, Sherrier R. Deep learning to detect intracranial hemorrhage in a national teleradiology program and the impact on interpretation time. *Radiology: Artificial Intelligence*. 2024;6(5):e240067. doi: [10.1148/ryai.240067](https://doi.org/10.1148/ryai.240067) EDN: EHHAOO

27. Pettet G, West J, Robert D, et al. A retrospective audit of an artificial intelligence software for the detection of intracranial haemorrhage used by a teleradiology company in the United Kingdom. *BJR|Open*. 2023;6(1):tzae033. doi: [10.1093/bjro/tzae033](https://doi.org/10.1093/bjro/tzae033) EDN: DWNYCF

28. Mäenpää SM, Korja M. Diagnostic test accuracy of externally validated convolutional neural network (CNN) artificial intelligence (AI) models for emergency head CT scans – A systematic review. *International Journal of Medical Informatics*. 2024;189:105523. doi: [10.1016/j.ijmedinf.2024.105523](https://doi.org/10.1016/j.ijmedinf.2024.105523) EDN: HLVVYQ

29. Eldaya RW, Kansagra AP, Zei M, et al. Performance of automated RAPID intracranial hemorrhage detection in real-world practice: a single-institution experience. *Journal of Computer Assisted Tomography*. 2022;46(5):770–774. doi: [10.1097/ret.0000000000001335](https://doi.org/10.1097/ret.0000000000001335) EDN: GRDZTF

30. Schmitt N, Mokli Y, Weyland CS, et al. Automated detection and segmentation of intracranial hemorrhage suspect hyperdensities in non-contrast-enhanced CT scans of acute stroke patients. *European Radiology*. 2021;32(4):2246–2254. doi: [10.1007/s00330-021-08352-4](https://doi.org/10.1007/s00330-021-08352-4) EDN: OLFWXI

31. Warman R, Warman A, Warman P, et al. Deep learning system boosts radiologist detection of intracranial hemorrhage. *Cureus*. 2022;undefined:.. doi: [10.7759/cureus.30264](https://doi.org/10.7759/cureus.30264) EDN: IRZKDY

32. Buchlak QD, Tang CHM, Seah JCY, et al. Effects of a comprehensive brain computed tomography deep learning model on radiologist detection accuracy. *European Radiology*. 2023;34(2):810–822. doi: [10.1007/s00330-023-10074-8](https://doi.org/10.1007/s00330-023-10074-8) EDN: ZHIEOG

33. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*. 2019;20(5):e262–e273. doi: [10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4)

34. Kiefer J, Kopp M, Ruettinger T, et al. Diagnostic accuracy and performance analysis of a scanner-integrated artificial intelligence model for the detection of intracranial hemorrhages in a traumatology emergency department. *Bioengineering*. 2023;10(12):1362. doi: [10.3390/bioengineering10121362](https://doi.org/10.3390/bioengineering10121362) EDN: EPL1BY

ОБ АВТОРАХ/ AUTHORS' INFO

Автор, ответственный за переписку:	
* Хоружая Анна Николаевна; адрес: Россия, 127051, Москва, ул. Петровка 24; ORCID: 0000-0003-4857-5404; eLibrary SPIN: 7948-6427; e-mail: KhoruzhayaAN@zdrav.mos.ru	* Anna N. Khoruzhaya, MD; address: 24 Petrovka st, Moscow, Russia, 127051; ORCID: 0000-0003-4857-5404; eLibrary SPIN: 7948-6427; e-mail: KhoruzhayaAN@zdrav.mos.ru
Соавторы:	
Арзамасов Кирилл Михайлович, д-р мед. наук; ORCID: 0000-0001-7786-0349 ; eLibrary SPIN: 3160-8062; e-mail: ArzamasovK@zdrav.mos.ru	Kirill M. Arzamasov, MD, Dr. Sci. (Medicine); ORCID: 0000-0001-7786-0349 ; eLibrary SPIN: 3160-8062; e-mail: ArzamasovK@zdrav.mos.ru
Коденко Мария Романовна, канд. техн. наук; ORCID: 0000-0002-0166-3768 ; eLibrary SPIN: 5789-0319; e-mail: KodenkoM@zdrav.mos.ru	Maria R. Kodenko, Cand. Sci. (Engineering); ORCID: 0000-0002-0166-3768 ; eLibrary SPIN: 5789-0319; e-mail: KodenkoM@zdrav.mos.ru
Кремнева Елена Игоревна, д-р мед. наук. ORCID: 0000-0001-9396-6063 ; eLibrary SPIN: 8799-8092; e-mail: KremnevaE@zdrav.mos.ru	Elena I. Kremneva, MD, Dr. Sci. (Medicine); ORCID: 0000-0001-9396-6063 ; eLibrary SPIN: 8799-8092; e-mail: KremnevaE@zdrav.mos.ru

Буренчев Дмитрий Владимирович , д-р мед. наук; ORCID: 0000-0003-2894-6255; eLibrary SPIN: 2411-3959; e-mail: BurenchevD@zdrav.mos.ru	Dmitriy V. Burenchev , MD, Dr. Sci. (Medicine); ORCID: 0000-0003-2894-6255; eLibrary SPIN: 2411-3959; e-mail: BurenchevD@zdrav.mos.ru
--	---

Accepted for publication

ТАБЛИЦЫ

Таблица 1. Метрики эффективности сервиса искусственного интеллекта, предназначенного для автоматического анализа медицинских компьютерно-томографических изображений головного мозга на наличие внутричерепных кровоизлияний, полученные по результатам трёх калибровочных тестирований

Параметры	Калибровка 1 0.6.1 ¹	Калибровка 2 0.7.1 ¹	Калибровка 3 0.8.3 ¹
AUROC, % (95% ДИ)	96 (92–100)	98 (96–99)	96 (91–99)
Чувствительность, % (95% ДИ)	89 (80–98)	90 (81–98)	84 (73–94)
Специфичность, % (95% ДИ)	96 (90–100)	98 (94–100)	98 (94–100)
Точность, % (95% ДИ)	92 (87–98)	94 (89–99)	91 (85–97)
Удельный вес ложноотрицательных результатов, %	11	10	16
Удельный вес ложноположительных результатов, %	4	2	2
Время обработки, с	73	73	85

Примечание. Калибровка 1, 2, 3 соответствует первому, второму и третьему калибровочным тестированиям, проведенным на 0-, 3- и 7-м месяцах работы; 1 — версия сервиса искусственного интеллекта; AUROC — площадь под характеристической кривой.

Таблица 2. Матрица ошибок сервиса искусственного интеллекта и метрики эффективности по каждому месяцу

Месяц	Se, %	Sp, %	Ac, %	AUROC, %	ИП, n	ИО, n	ЛП, n	ЛО, n
<i>Калибровка 1</i>								
1	100	39,4	50,0	93,1	14	26	40	0
2	100	46,0	57,5	94,7	17	29	34	0
3	100	42,2	53,7	99,4	16	27	37	0
<i>Калибровка 2</i>								
4	97,5	82,5	90,0	94,6	39	33	7	1
5	93,5	61,2	73,8	86,2	29	30	19	2
6	94,6	72,1	82,5	90,2	35	31	12	2
7	97,1	66,7	80,0	90,0	34	30	15	1
<i>Калибровка 3</i>								
8	100	75,6	87,5	93,9	39	31	10	0
9	95,3	79,5	87,8	92,8	41	31	8	2
10	100	71,1	83,8	92,8	35	32	13	0
11	100	68,9	82,5	92,2	35	31	14	0
12	97,6	76,9	87,5	93,1	40	30	9	1
13	100	78,0	88,8	94,5	39	32	9	0
14	97,6	73,7	86,3	92,2	41	28	10	1
15	97,4	83,3	90,0	94,7	37	35	7	1
16	95,6	82,9	90,0	93,8	43	29	6	2
17	94,4	77,8	85,0	91,7	34	34	10	2
18	100	88,9	95,0	97,2	44	32	4	0

Примечание. Калибровка 1, 2, 3 соответствует первому, второму и третьему калибровочным тестированиям. Серым цветом обозначены данные, исключённые из дальнейшего анализа. Se — чувствительность; Sp — специфичность; Ac — точность; AUROC — площадь под характеристической кривой; ИП — истинно положительный ответ; ИО — истинно отрицательный ответ; ЛП — ложноположительный ответ; ЛО — ложноотрицательный ответ.

РИСУНКИ

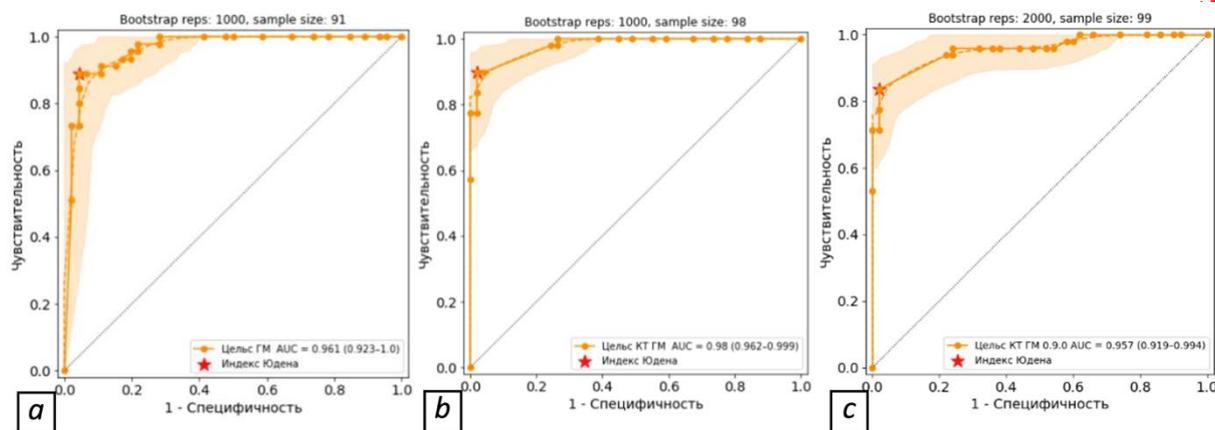


Рис. 1. Характеристические кривые калибровочных тестирований сервиса искусственного интеллекта, предназначенного для автоматического анализа медицинских компьютерно-томографических изображений головного мозга на наличие внутрочерепных кровоизлияний: а — первое; б — второе; с — третье.

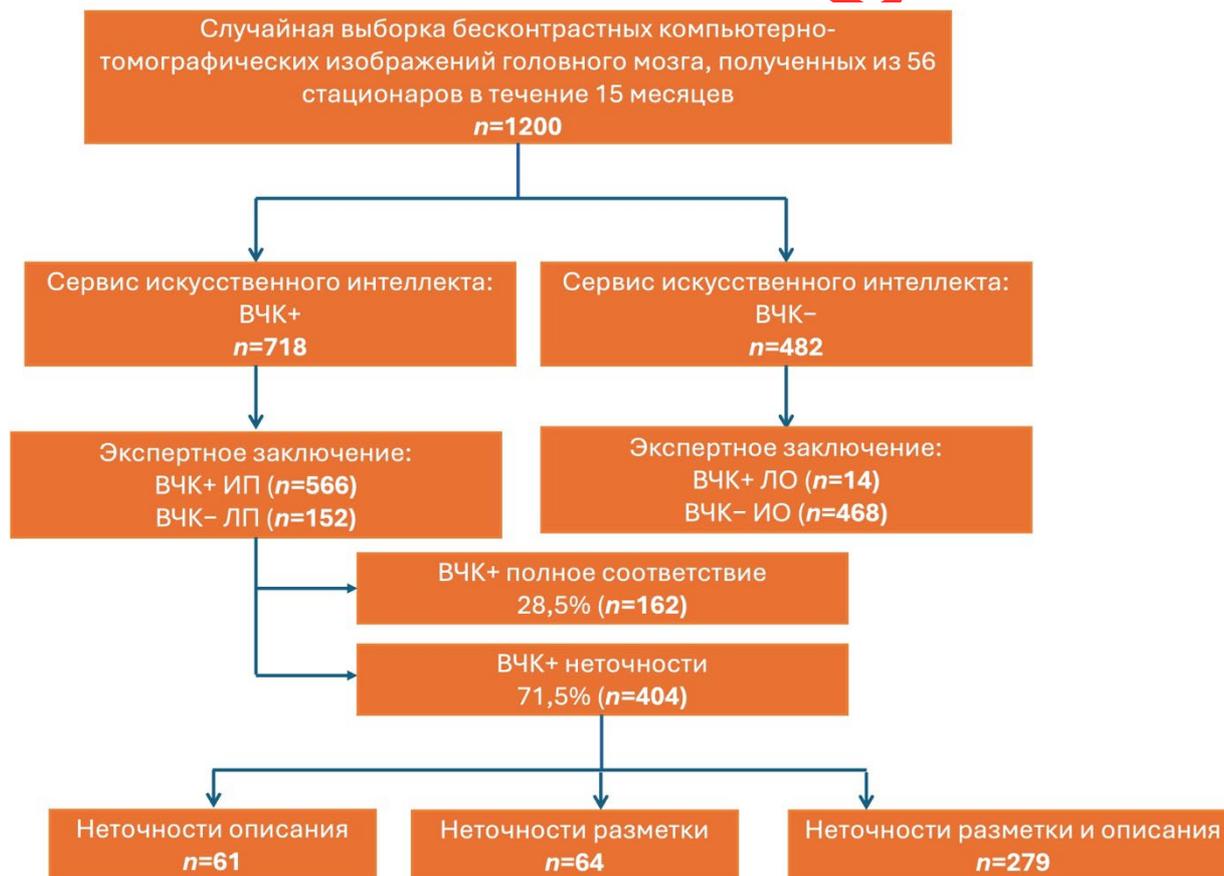


Рис. 2. Схема анализа компьютерно-томографических исследований головного мозга при экспертной оценке в условиях клинического мониторинга: ВЧК+ — наличие внутрочерепных кровоизлияний; ВЧК- — отсутствие внутрочерепных кровоизлияний; ИП — истинно положительный результата; ЛП — ложноположительный результат; ЛО — ложноотрицательный результат; ИО — истинно отрицательный результат.

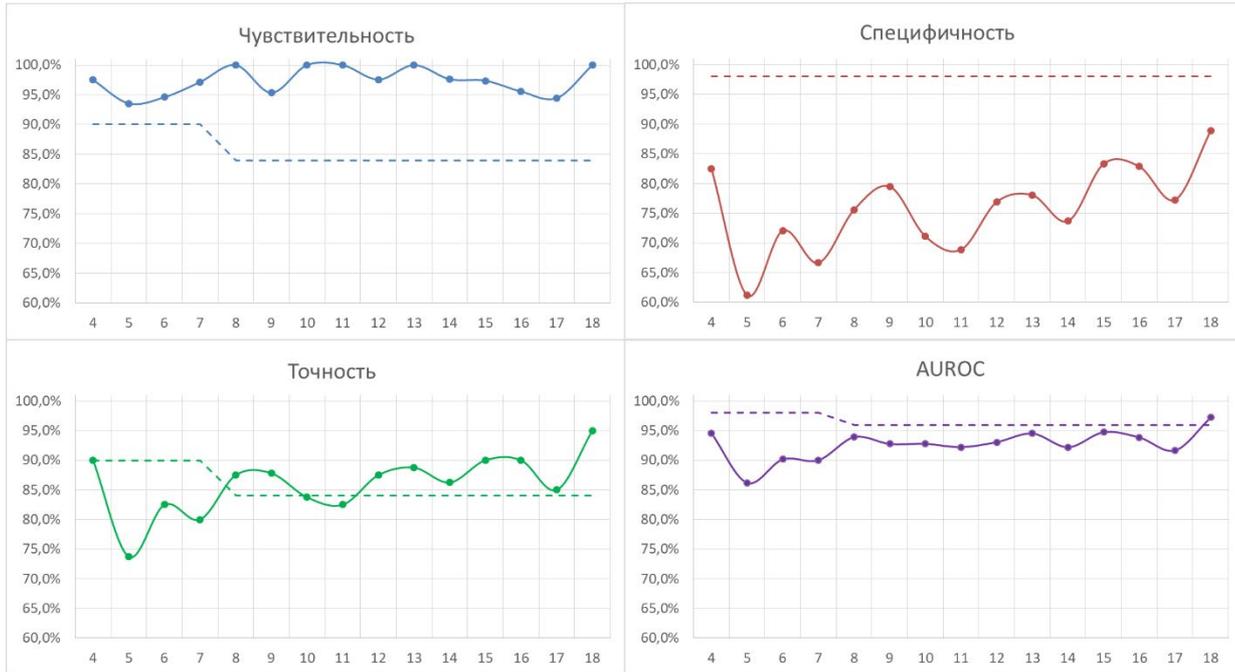


Рис. 3. Динамика диагностических метрик работы сервиса искусственного интеллекта относительно результатов двух калибровочных тестирований: ось абсцисс — значения метрик, ось ординат — месяцы. Пунктиром обозначены результаты метрик, полученные в ходе калибровочных тестирований.

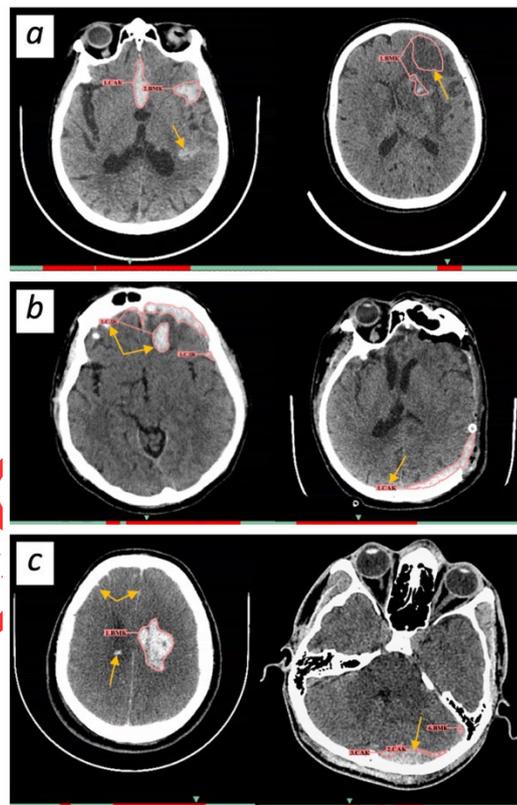


Рис. 4. Примеры частично корректных срабатываний сервиса искусственного интеллекта: *a* — корректное определение типа кровоизлияния и некорректная их сегментация; *b* — корректная сегментация областей кровоизлияния, ошибочное определение их типа; *c* — частичное выделение одних кровоизлияний и пропуск других, некорректные как сегментация, так и определение типа.

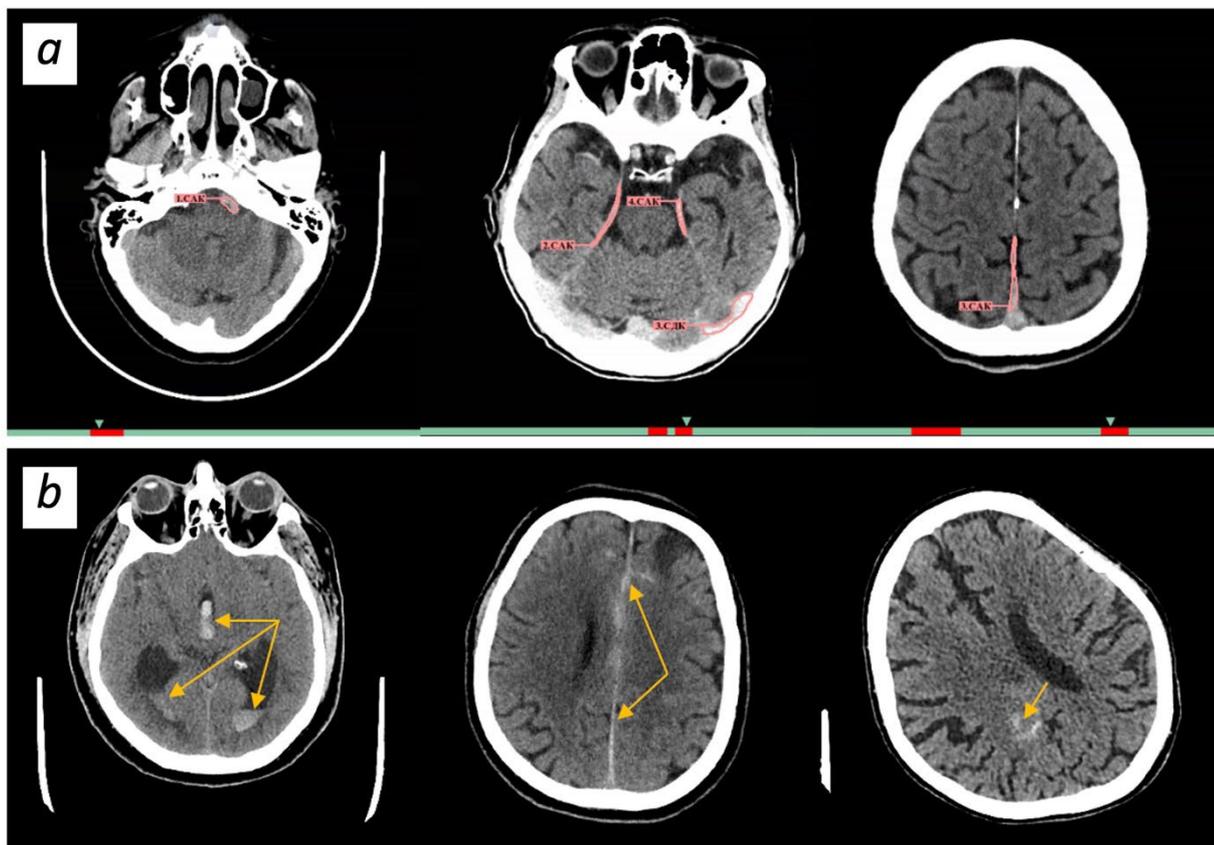


Рис. 5. Примеры ложноположительных (a) и ложноотрицательных (b) срабатываний сервиса искусственного интеллекта.

Accepted for