

DOI: <https://doi.org/10.17816/vto633860>

Применение искусственного интеллекта в диагностике деформирующего остеоартроза крупных суставов нижних конечностей: оценка диагностической точности в реальных клинических условиях

А.В. Владзимирский, Ю.А. Васильев, К.М. Арзамасов, В.Е. Казаринова, Е.В. Астапенко

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

АННОТАЦИЯ

Обоснование. Развитие математических методов, цифровизация медицинской диагностической аппаратуры и рост вычислительных возможностей компьютеров создали условия для появления новых средств автоматизированного анализа биомедицинских данных — технологий искусственного интеллекта (ТИИ). В клинической практике среди перспективных ТИИ наибольшее распространение получило компьютерное зрение. С 2023 г. в Московском Эксперименте применяются ИИ-сервисы для диагностики травм и заболеваний опорно-двигательной системы, что позволило впервые масштабнo изучить качество соответствующих инструментов.

Цель. Изучить диагностическую значимость программного обеспечения на основе технологий искусственного интеллекта для диагностики деформирующего остеоартроза крупных суставов нижних конечностей.

Материалы и методы. Научная работа, выполненная в дизайне диагностического исследования по методологии STARD 2015, включала два этапа — ретроспективный и проспективный. Ретроспективный этап представлял собой расчёт показателей диагностической ценности (AUROC, чувствительность, специфичность и точность). Проспективный этап заключался в регулярном мониторинге диагностического качества работы ИИ-сервиса при анализе реального потока результатов рентгенографии ($n=198\ 821$). Проводился расчёт согласия врача-эксперта с решением ИИ-сервиса, а также интегральная клиническая оценка. Продолжительность исследования — 1 год и 8 месяцев.

Результаты. Были изучены 5 российских программных решений на основе ТИИ для выявления признаков деформирующего остеоартроза. Из них только два ИИ-сервиса успешно прошли этап ретроспективной оценки диагностической точности и были допущены к проспективному этапу. При работе в клинических условиях оба ИИ-сервиса продемонстрировали достаточную техническую надёжность. Для одного из ИИ-сервисов был установлен средне-высокий уровень диагностической ценности (медиана клинической оценки составила более 88,0%), для второго — высокий уровень диагностической ценности (медиана клинической оценки составила более 93,0%).

Заключение. Достигнутый уровень развития программного обеспечения на основе ТИИ позволяет применять соответствующие разработки для повышения точности и производительности труда врачей-рентгенологов при описании результатов рентгенографии крупных суставов нижних конечностей (в контексте диагностики деформирующего остеоартроза).

Ключевые слова: деформирующий остеоартроз; рентгенография; искусственный интеллект.

Как цитировать:

Владзимирский А.В., Васильев Ю.А., Арзамасов К.М., Казаринова В.Е., Астапенко Е.В. Применение искусственного интеллекта в диагностике деформирующего остеоартроза крупных суставов нижних конечностей: оценка диагностической точности в реальных клинических условиях // Вестник травматологии и ортопедии им. Н.Н. Приорова. 2025. Т. 32, № 1. С. 95–105. DOI: <https://doi.org/10.17816/vto633860>

DOI: <https://doi.org/10.17816/vto633860>

Artificial intelligence in the diagnostics of deforming osteoarthritis of large joints of the lower limbs — diagnostic accuracy assessment in the real clinical settings

Anton V. Vladzymyrskyy, Yuriy A. Vasilev, Kirill M. Arzamasov, Veronika E. Kazarinova, Elena V. Astapenko

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

ABSTRACT

BACKGROUND: A development of mathematical methods, digitalization of medical diagnostic equipment, and growth of computing capabilities have created conditions for the emergence of new tools for automated analysis of biomedical data — artificial intelligence (AI) technologies. In clinical practice, computer vision has become the most widespread among promising AI technologies. Since 2023, the Moscow Experiment has been using AI services to diagnose injuries and conditions of the musculoskeletal system, which has allowed studying a quality of the relevant tools for the first time on a large-scale.

AIM: To study a diagnostic significance of software based on artificial intelligence technologies for the diagnosis of deforming osteoarthritis of large joints of the lower limbs.

MATERIALS AND METHODS: The scientific work performed in the design of the diagnostic study according to the STARD 2015 methodology included two stages — retrospective and prospective. The retrospective stage was a calculation of diagnostic accuracy indicators (AUROC, sensitivity, specificity, and accuracy). The prospective phase consisted of regular monitoring of the diagnostic quality of the AI service while analysing the actual flow of radiography results ($n=198,821$). A match of radiologist and AI service results was calculated, as well as an integral clinical evaluation was performed. The duration of the study was 1 year and 8 months.

RESULTS: Five Russian AI-based software for detecting signs of deforming osteoarthritis were studied. Only two of them successfully passed the stage of retrospective diagnostic accuracy assessment and were accepted to participate in the prospective stage. Both AI services demonstrated the sufficient technical reliability in clinical conditions. One of the AI services had a medium-high level of diagnostic value with the median clinical score of more than 88.0%, while the other had a high level of diagnostic value with the median clinical score of more than 93.0%.

CONCLUSIONS: The achieved level of AI-based software development allows applying them to improve the accuracy and productivity of radiologists when providing radiology reports of large joints of the lower limbs (in the context of diagnosing deforming osteoarthritis).

Keywords: osteoarthritis; radiography; artificial intelligence.

To cite this article:

Vladzymyrskyy AV, Vasilev YuA, Arzamasov KM, Kazarinova VE, Astapenko EV. Artificial intelligence in the diagnostics of deforming osteoarthritis of large joints of the lower limbs — diagnostic accuracy assessment in the real clinical settings. *N.N. Priorov Journal of Traumatology and Orthopedics*. 2025;32(1):95–105. DOI: <https://doi.org/10.17816/vto633860>

ОБОСНОВАНИЕ

Теоретические и практические аспекты автоматизированного (компьютерного) анализа биомедицинских данных разрабатываются во всех ведущих странах мира примерно с 1960-х гг. Давно стали обыденностью средства автоматической расшифровки электрокардиограммы и разнообразные лабораторные анализаторы. Вместе с тем развитие математических методов (прежде всего нейронных сетей), цифровизация медицинской диагностической аппаратуры и стремительный рост вычислительных возможностей стандартных компьютеров создали условия для появления следующего поколения средств автоматизированного анализа. В настоящее время такие средства называют технологиями искусственного интеллекта (ТИИ).

Искусственный интеллект не является таковым по своей сути. Множество функций и возможностей мозга человека — интуиция, клиническое мышление, творчество, создание новых знаний — невозпроизводимы технически. Однако ТИИ как способ автоматизации рутинных, механических трудовых процедур могут стать эффективными помощниками в разных сферах жизнедеятельности, в том числе в здравоохранении.

В медицине очень много говорится и пишется о перспективах и потенциальных возможностях искусственного интеллекта — для создания лекарств, прогноза, диагностики, поддержки принятия решений о тактике лечения и т.д. [1]. Однако на практике наибольшее распространение получило компьютерное зрение, то есть ТИИ для распознавания и анализа различных изображений, чаще всего — результатов лучевых исследований.

Благоприятный контекст и появление новых инструментов сочетаются с крайне поверхностными знаниями о реальных возможностях и ограничениях ТИИ в каждой отдельной клинической области. Очевидно, что в этой ситуации требуются системные научные исследования медицинского искусственного интеллекта как нового средства автоматизации.

В последние годы появились разработки на основе ТИИ и для диагностики травм и заболеваний опорно-двигательной системы, потенциально применимые как в практике врачей-рентгенологов, так и в работе ортопедов-травматологов. Создаются нейросети для выявления переломов длинных костей конечностей и таза, диагностики сколиоза, дисплазии, выбора тактики лечения при политравме, а также для текстурного анализа с целью прогноза дегенеративно-дистрофических процессов в крупных суставах [2]. Вместе с тем сведения о качестве, точности, безопасности и эффективности таких разработок носят крайне ограниченный характер.

С 2020 года в г. Москве проводится крупнейшее в мире научное проспективное многоцентровое клиническое исследование безопасности и качества ТИИ для лучевой диагностики — Эксперимент по использованию

инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы (mosmed.ai; далее — Московский Эксперимент). В его рамках программные решения на основе ТИИ (далее — ИИ-сервисы) интегрируются в Единый радиологический информационный сервис Единой медицинской информационно-аналитической системы г. Москвы (ЕРИС ЕМИАС), фактически — в централизованный архив медицинских изображений столицы. После специальных проверочных процедур они допускаются к применению в условиях практического здравоохранения. Работа ИИ-сервисов подвергается постоянному и всестороннему научному изучению по оригинальным методологиям. Это позволяет накапливать принципиально новые знания о реальных возможностях ТИИ, научно обосновывать сценарии применения, создавать методики внедрения и контроля качества [3]. С 2023 г. в Московском Эксперименте применяются ИИ-сервисы для диагностики травм и заболеваний опорно-двигательной системы, что позволило впервые масштабно изучить качество соответствующих инструментов.

Цель исследования — изучить диагностическую значимость программного обеспечения на основе технологий искусственного интеллекта для диагностики деформирующего остеоартроза (ДОА) крупных суставов нижних конечностей.

МАТЕРИАЛЫ И МЕТОДЫ

Дизайн исследования

Научная работа, выполненная в дизайне диагностического исследования по методологии STARD 2015 [4], включала два этапа — ретроспективный и проспективный (обработка всех исследований на потоке). Исследование наблюдательное, многоцентровое. Схема проведения исследования представлена на рис. 1.

Критерии соответствия

Процедурами Московского Эксперимента подразумевается установление порогового значения показателя «площадь под характеристической кривой» (на момент проведения нашего исследования — 0,81). В случае недостижения порогового уровня ИИ-сервисом на ретроспективном этапе он не допускался к следующему, проспективному этапу, а направлялся на доработку и дополнительное обучение.

Условия проведения

Данное исследование проводилось в рамках Московского Эксперимента, участие в нём приняли 410 врачей и 190 медицинских организаций, подведомственных Департаменту здравоохранения города Москвы (ДЗМ), — городские больницы, поликлиники и клиничко-диагностические центры.



Рис. 1. Схема проведения исследования.

Fig. 1. Scheme for conducting the study.

Продолжительность исследования

Продолжительность исследования — 1 год и 8 месяцев.

Описание медицинского вмешательства

Клиническая задача для ИИ-сервисов включала выявление рентгенографических признаков деформирующего артроза суставов с определением стадии патологического процесса по классификации Н.С. Косинской [5].

Основной исход исследования

Показатели диагностической ценности — площади под характеристической кривой (далее — AUROC, от англ. area under receiver operating characteristic curve), чувствительности, специфичности и точности, значение согласованности врача-эксперта с решением ИИ-сервиса в трактовке и локализации, интегральная клиническая оценка.

Значения всех показателей диагностической точности/качества интерпретировали следующим образом: менее 0,7 — неприемлемое, 0,7–0,8 — приемлемое, 0,81–0,9 — хорошее, более 0,9 — высокое [6].

Анализ в подгруппах

Разделение на подгруппы в рамках данного исследования не проводилось. В анализ включались все пациенты из медицинских организаций, подведомственных ДЗМ, которым было проведено рентгенографическое исследование коленного или тазобедренного сустава.

Методы регистрации исходов

Ретроспективный этап представлял собой расчёт показателей диагностической ценности (AUROC,

чувствительности, специфичности и точности). Порог активации определялся по индексу Юдена [7].

Индекс-тест: ИИ-сервис, интегрированный в ЕРИС ЕМИАС. Референс-тест: набор данных ($n=100$), состоящий из результатов рентгенографии с разметкой квалифицированными врачами-рентгенологами по научно обоснованной и утверждённой методологии [8].

В работе использовались два набора данных ($n=200$) с деперсонализированными результатами рентгенографии коленного и тазобедренного сустава — с признаками деформирующего остеоартроза.

Исследуемый ИИ-сервис получал на вход конкретный набор данных, выполнял его анализ и возврат соответствующих результатов в тестовый контур ЕРИС ЕМИАС. Результаты машинного анализа сравнивали с референс-тестом, то есть эталонной разметкой, после чего определяли количество истинно положительных, истинно отрицательных, ложноположительных и ложноотрицательных диагностических решений, составляли четырёхпольную таблицу и вычисляли показатели диагностической ценности.

Проспективный этап заключался в регулярном мониторинге технической надёжности и диагностического качества работы ИИ-сервиса при анализе реального потока результатов рентгенографии ($n=198\ 821$).

Методология такого мониторинга включает случайный выбор результатов лучевых исследований, формирование набора данных (с распределением норма : патология = 25 : 75), пересмотр набора данных врачами-экспертами. По итогам пересмотра для каждого случая врач-эксперт указывает своё согласие с решением ИИ-сервиса:

а) об отсутствии, наличии и характере патологии (трактовка); б) о локализации патологических изменений. Своё мнение каждый эксперт предоставляет в виде унифицированной формулировки (полное соответствие, некорректная оценка, ложноположительный и ложноотрицательный результат), что даёт возможность определить качество ИИ-сервиса в диапазоне от 0 до 100%. Отдельно проводится расчёт согласия для трактовки, локализации, а также интегральная клиническая оценка. Подробно методология приведена в нормативных документах Московского Эксперимента и опубликована в рецензируемом журнале [9].

Все процессы получения, обработки, хранения, использования данных и прочие были выполнены в соответствии с действующим законодательством.

Этическая экспертиза

Данное исследование основано на результатах Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы, утверждённого этическим комитетом (выписка из протокола № 2 НЭК МРО РОПР от 20 февраля 2020 года), также зарегистрированного на ClinicalTrials (NCT04489992).

Статистический анализ

Принципы расчёта размера выборки

На ретроспективном этапе исследования использовались два набора данных: рентгенографические изображения тазобедренного сустава ($n=100$) и рентгенографические изображения коленного сустава ($n=100$). Наборы данных сбалансированы по патологии (норма : патология = 50 : 50). В рамках проспективного этапа мониторинг работы ИИ-сервисов проводился ежемесячно, в каждый соответствующий набор данных случайным образом отбирались результаты 80 рентгенографий. Такой объём выборки был установлен и опубликован нами ранее [10].

Методы статистического анализа данных

Расчёт всех показателей проводился на языке Python с использованием ранее разработанного нами веб-инструмента для выполнения ROC-анализа (свидетельство № RU 2022617324). Формат представления количественных данных табличный. Целью настоящей работы не являлось сопоставление сервисов между собой, поэтому мы использовали описательную статистику и расчёт показателей диагностической точности.

РЕЗУЛЬТАТЫ

Объекты (участники) исследования

Изучены пять российских программных решений на основе технологий искусственного интеллекта для выявления признаков деформирующего остеоартроза коленного сустава — «Care Mentor Knee Arthrosis» (ООО

«КэреМенторЭйАй»), «Цельс РГ ОДА» (ООО «Медицинские скрининг системы»), «CVisionRad — Knee Arthrosis» (ООО «СиВиЖинЛаб»), «Oxytech РГ коленный сустав» (ООО «Оксиджен Технолджис Рус») и тазобедренного сустава — «Oxytech РГ тазобедренный сустав» (ООО «Оксиджен Технолджис Рус»).

Для обеспечения непредвзятого изложения в тексте статьи перечисленные программные продукты анонимизированы. Вместе с тем исчерпывающая информация о качестве каждого отдельного продукта представлена на официальном сайте Московского Эксперимента — mosmed.ai.

Основные результаты исследования

Сводные данные о результатах ретро- и проспективной оценки ИИ-сервисов для диагностики травм и заболеваний крупных суставов нижних конечностей представлены в табл. 1.

Ретроспективный этап тестирования

С использованием размеченных наборов данных проведено тестирование пяти ИИ-сервисов для автоматизированного анализа результатов рентгенографии коленного и тазобедренного суставов в условиях тестового контура ЕРИС ЕМИАС.

Диагностическая ценность ИИ-1 оказалась самой низкой и несоответствующей пороговым значениям для допуска в Эксперимент. Площадь под характеристической кривой составила всего 0,58, чувствительность — 0,76, специфичность — 0,41, точность — 0,58. Отклонения от заявленных разработчиком значений значительно превысили пороговый уровень в 10,0% и составили для AUROC -31,6%, для специфичности — -50,6%, для точности — -30,1%. Порог не был превышен только для чувствительности (-7,3%). В соответствии с полученными результатами к проспективному этапу ИИ-1 не допущен.

Аналогичная результативность получена для ИИ-сервиса ИИ-2. Показатели AUROC, чувствительности, специфичности и точности составили 0,70, 0,53, 0,85 и 0,68 соответственно. При этом значения площади под характеристической кривой, чувствительности и точности были ниже заявленных разработчиком на 20,3, 33,8 и 15,0% соответственно. Лишь специфичность продемонстрировала незначительный рост. Из-за несоответствия показателей диагностической ценности пороговым значениям ИИ-2 к проспективному этапу также не допущен.

ИИ-5 трижды проходил ретроспективный этап калибровочного тестирования. Первоначальной проблемой стала крайне низкая специфичность (0,27), обусловившая генерацию колоссального количества ложноположительных решений; в дальнейшем — наоборот, крайне низкая чувствительность (0,31), приводящая к пропускам случаев с патологическими отклонениями. В соответствии с полученными результатами к проспективному этапу ИИ-5 не допущен.

Таблица 1. Оценка диагностической точности и качества ИИ-сервисов для анализа результатов рентгенографии тазобедренного и коленного суставов

Table 1. Evaluation of diagnostic accuracy and quality of AI services for analyzing hip and knee radiography results

Показатели	ИИ-1	ИИ-2	ИИ-3	ИИ-4	ИИ-5
<i>Заявленные значения</i>					
Точность	0,83	0,8	0,84	0,91	0,87
Чувствительность	0,82	0,8	0,84	0,91	0,87
Специфичность	0,83	0,8	0,84	0,92	0,87
AUROC	0,85	0,88	0,91	0,93	0,92
<i>Ретроспективный этап</i>					
Точность	0,76	0,68	0,88	0,8	0,62
Чувствительность	0,41	0,53	0,8	0,61	0,98
Специфичность	0,58	0,85	0,96	0,98	0,27
AUROC	0,76	0,70	0,92	0,84	0,69
<i>Проспективный этап</i>					
Согласие с заключением, %	–	–	88,8	91,3	–
Согласие с локализацией, %	–	–	89,7	93,8	–
Клиническая оценка, %	–	–	88,0	93,0	–

Для ИИ-3 получены достаточно высокие значения показателей диагностической ценности. В частности, AUROC составил 0,92, чувствительность — 0,8, общая точность — 0,88; обращает на себя внимание высокая специфичность (показатель точности определения именно вида патологии) — 0,96. Примечательно, что полученные значения в целом соответствовали заявленным производителем, а прирост специфичности составил +14,3%.

Своеобразная ситуация сложилась для ИИ-4. При ретроспективном тестировании AUROC составил 0,84, специфичность — 0,98, точность — 0,8. На этом фоне показатель чувствительности был низок — 0,61 (на 33,0% ниже заявленного). Это обуславливало риск ложноотрицательных решений, однако не противоречило правилам Эксперимента (пороговое значение площади под характеристической кривой было преодолено). Несмотря на некоторую парадоксальность ситуации, по формальным признакам ИИ-4 допущен к проспективному этапу.

Таким образом, из пяти ИИ-сервисов для анализа результатов рентгенографии коленного и тазобедренного суставов успешно прошли этап ретроспективной оценки диагностической точности только два, то есть 40,0%.

Проспективный этап тестирования

Технически и организационно ИИ-3 и ИИ-4 переведены в промышленный контур ЕРИС ЕМИАС для выполнения автоматизированного анализа реального потока результатов рентгенографии коленного сустава (рис. 2).

При работе в клинических условиях оба ИИ-сервиса продемонстрировали достаточную техническую надёжность и высокую скорость обработки данных; в подавляющем большинстве случаев автоматизированный

анализ результатов рентгенографии выполнялся менее чем за 60 секунд.

Для ИИ-3 период наблюдения был установлен с 12.08.2022 по 30.04.2024 г. За это время ИИ-сервис автоматически проанализировал результаты 170 362 рентгенографий коленного сустава с целью выявления признаков остеоартроза. Выполнено 20 клинических мониторингов с анализом качества машинной обработки суммарно 1600 рентгенографий. Согласие врачей-экспертов с заключением ИИ-сервиса об отсутствии, наличии и характере патологии колебалось от 66,9 до 98,3%, при этом медиана составила 88,8%. Согласие с локализацией патологических изменений колебалось от 56,9 до 98,7%, медиана — 89,7%. Суммарная клиническая оценка варьировала от 61,9 до 98,5%, медиана — 88,1%. Таким образом, в ходе проспективного изучения установлен средне-высокий уровень диагностической ценности ИИ-3: медиана клинической оценки (включая согласия с заключением и локализацией) составила более 88,0%.

Для ИИ-4 период наблюдения был установлен с 01.09.2023 по 30.04.2024 г. За это время ИИ-сервис автоматически проанализировал результаты 28 459 рентгенографий коленного сустава с целью выявления признаков остеоартроза. Проведено 9 клинических мониторингов с оценкой качества автоматизированного анализа результатов 720 рентгенографий. Нижний уровень согласий врачей-экспертов с заключением ИИ-сервиса и локализацией, а также клинической оценки составил 88,8%. При этом верхняя граница согласия с заключением достигла 93,8%, с локализацией — 98,1%. Соответствующие медианы составили 91,3 и 93,8% соответственно. Зафиксирован максимальный уровень клинической оценки в 94,7%. Таким

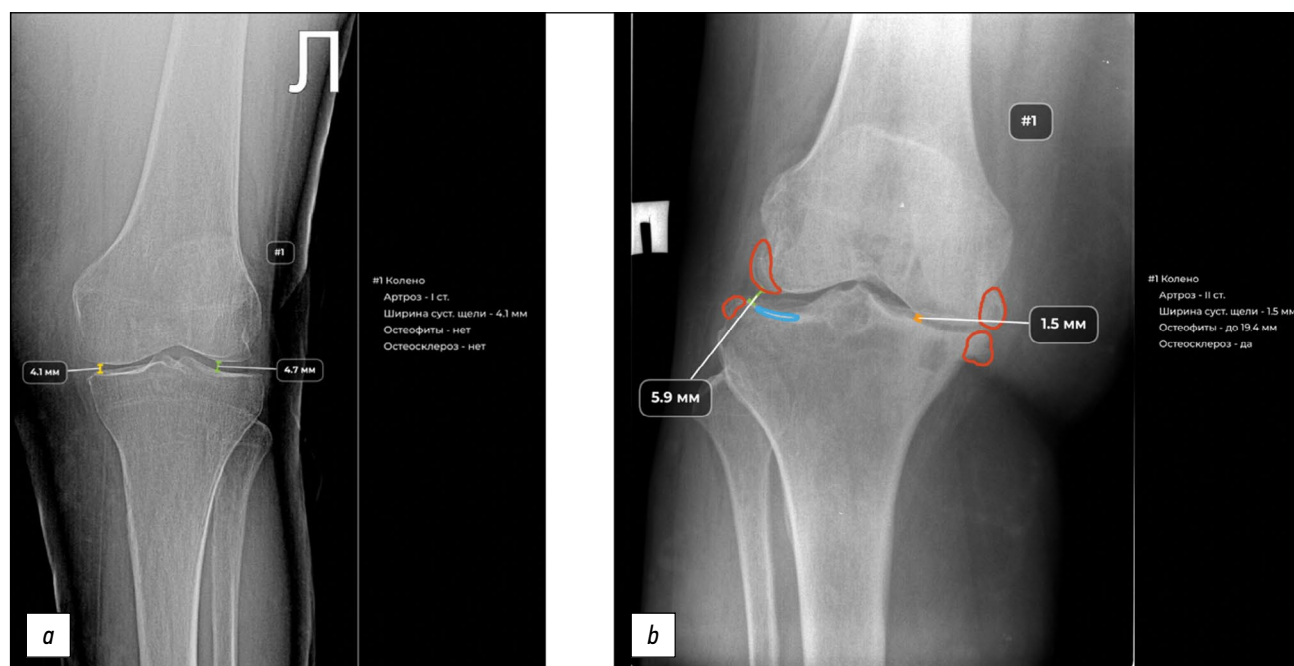


Рис. 2. Пример обработки результатов рентгенографии коленного сустава ИИ-сервисом: *a* — артроз коленного сустава I стадии, *b* — артроз коленного сустава II стадии.

Fig. 2. Example of knee joint radiographs processing by AI-service: *a* — knee arthrosis stage I, *b* — knee arthrosis stage II.

образом, в ходе проспективного изучения установлен высокий уровень диагностической ценности ИИ-4: медиана клинической оценки (включая согласия с заключением и локализацией) составила более 93,0%.

ОБСУЖДЕНИЕ

Резюме основного результата исследования

Были изучены пять российских программных решений на основе ТИИ для выявления признаков деформирующего остеоартроза. Из них только два ИИ-сервиса успешно прошли этап ретроспективной оценки диагностической точности и были допущены к проспективному этапу. При работе в клинических условиях оба ИИ-сервиса продемонстрировали достаточную техническую надёжность и высокую скорость обработки данных. Для ИИ-3 был установлен средне-высокий уровень диагностической ценности, для ИИ-4 — высокий уровень.

Обсуждение основного результата исследования

Вопросы применения технологий искусственного интеллекта в контексте диагностики ДОО коленного и тазобедренного сустава затронуты в ряде зарубежных публикаций; приходится констатировать тот факт, что российские специалисты фактически не занимаются данной проблематикой. Наибольшее внимание сосредоточено на создании ТИИ для выявления признаков ДОО именно коленного сустава, а также определения тяжести заболевания по классификации Келлгрена–Лоуренса (Kellgren–Lawrence). Точность предложенных разработок колеблется от 76,93 до 98,9%, коэффициент согласованности (каппа

Коэна) диагностических решений врачей и ТИИ составляет 0,768–0,815. Чувствительность машинного выявления категорий тяжести от 0 до 4 составляет 1,0, 0,972, 0,979, 0,983 и 0,989, специфичность — 0,992, 0,997, 0,994, 0,991 и 0,995 соответственно [11].

В публикациях заявлены довольно высокие показатели, свидетельствующие о значительной диагностической ценности ТИИ. Нами на ретроспективном этапе получены в целом аналогичные значения. Это говорит об идентичном исходном уровне диагностической точности зарубежных и отечественных разработок.

Однако оптимизм процитированных авторов сильно преувеличен. Все исследования выполнены в ретроспективном одноцентровом дизайне, то есть тестирование нейросетей проведено на тех же наборах данных, на которых проводилось обучение соответствующих математических моделей. Авторы действуют с применением общепринятого среди математиков подхода, когда один набор данных случайным образом разделяют на три части — для обучения, валидации и тестирования. С точки зрения статистики подход вполне может обеспечить достаточный уровень показателей точности (что наглядно подтверждают процитированные выше работы). Но для медицинской науки и практики этого недостаточно. Обязательно требуется проверка точности и качества работы ТИИ на новых данных, полученных из иных медицинских организаций и не использованных для обучения модели. Только при проведении подобной, так называемой внешней, валидации можно получить действительно объективные показатели диагностической точности ТИИ для практической медицины. В идеальной ситуации валидация выполняется

ещё и в виде проспективного мультицентрового клинического исследования.

Нами выявлено только две публикации, в которых проведена внешняя валидация ТИИ для выявления и определения тяжести ДОО коленного сустава. Получены достаточно высокие значения показателей точности (точность — 97,8%, согласованность — от 0,8 до 0,96). Однако и эти работы имели выраженные ограничения. Во-первых — малый объём выборки: в первом исследовании использованы результаты 246 рентгенограмм, во втором — 99. Во-вторых, валидация проведена в варианте мультицентрового, но ретроспективного исследования. То есть ТИИ работали не с потоком данных (как в нашем исследовании), а с отдельными наборами данных, которые не использовались для обучения и тестирования на его этапах [12].

Проблеме автоматизированного анализа результатов рентгенографии тазобедренного сустава посвящены единичные исследования. Показано, что нейросети для выявления признаков деформирующего остеоартроза тазобедренного сустава на рентгенограммах достигают следующих показателей: точность — 90,2–92,8%, чувствительность — 95,0–97,6%, специфичность — 83,0–90,7%. Точнее всего нейросети выявляют субхондральные кисты, а хуже всего — остеофиты в области суставной впадины. Однако эти результаты тоже получены в ограниченных исследованиях — ретроспективных, не имеющих внешней валидации [13]. Отдельные авторы разрабатывают ТИИ для выявления признаков дисплазии как состояния, обуславливающего в том числе развитие ДОО [14]. Эти исследования также носят ограниченный характер.

Качественным отличием нашего исследования стало проведение проспективной мультицентральной внешней валидации на реальном потоке клинических данных ($n=198\ 821$). Достижение высоких показателей точности ($>0,91$) на ограниченных наборах данных не имеет значимости для медицинской науки. Напротив, нами получены данные о точности ТИИ в реальных клинических условиях.

В литературе практически игнорируется вопрос сценария применения ТИИ в контексте диагностики ДОО. Искусственный интеллект, будучи одной из многочисленных информационных технологий, безусловно, не может применяться как «вещь в себе». ТИИ — это инструмент, призванный решать конкретные проблемы клинических производственных процессов.

В настоящее время, с одной стороны, появилась техническая возможность создавать нейросети для выявления признаков ДОО, с другой — роль и место таких продуктов в реальной клинической работе (как врачей-рентгенологов, так и клинических специалистов) не определены.

Некоторые авторы делают акцент на устранении субъективизма при определении тяжести ДОО, то есть ТИИ должны сделать классификацию максимально точной и не зависящей от опыта конкретного врача [15]. С данной точкой зрения мы полностью согласны. Автоматизация

определения степени тяжести ДОО крупных суставов нижних конечностей направлена на повышение объективности диагностических решений, устранение негативных аспектов человеческого фактора (усталости, невнимательности, отсутствия времени на выполнение измерений и пр.).

Интересный способ применения ТИИ предложен в контексте отбора данных для научных исследований (не связанных с искусственным интеллектом) и образовательных курсов. Посредством автоматизированного анализа, в соответствии с критериями включения выполняется отбор релевантных клинических случаев. В частности, применение ТИИ для формирования выборки из результатов 25 778 рентгенограмм коленного сустава с признаками ДОО позволило сэкономить 800 человеко-часов «ручного» труда врача-рентгенолога [16].

В целом приходится сделать неутешительный вывод о том, что врачи практически не участвуют в разработке ТИИ как постановщики задач, тем более не определяют область и задачи применения ТИИ в составе клинических производственных процессов. Наглядный тому пример — крайне сомнительный клинический сценарий использования ТИИ для выявления гонартроза, суть которого заключается в следующем. Недостаточно квалифицированный врач-ортопед, находящийся в отдалённой местности, нуждается в ТИИ для определения того, затрагивает ли дегенеративно-дистрофический процесс один отдел коленного сустава (медиальный или латеральный) либо оба. Определить же объём поражения необходимо для выбора вида эндопротезирования — полного или частичного [17]. Достаточно трудно представить ситуацию, когда врач-ортопед не может на рентгенограмме определить объём поражения при ДОО, но достаточно компетентен, чтобы выполнить эндопротезирование коленного сустава. Подобные публикации выглядят как попытка приспособить свою разработку, подобрать для неё подходящую клиническую задачу.

В своей деятельности мы придерживаемся принципиально иного подхода. Всё начинается с клинического целеполагания — обоснованной постановки задачи, включающей модальность и вид лучевого исследования для автоматизированного анализа, а также стандартизированные требования к результатам работы ИИ. Соответственно, выбор рентгенографии крупных суставов нижних конечностей с целью выявления и определения степени тяжести деформирующего остеоартроза был обусловлен частотой выполнения таких исследований в сети медицинских организаций г. Москвы и распространённостью указанного заболевания. В контексте производственного процесса автоматизация направлена на устранение либо рутинных операций по измерению суставной щели, либо её субъективной оценки; в любом случае ИИ-сервис должен повысить производительность труда врача-рентгенолога.

Вместе с тем открытым остаётся вопрос применения изучаемых ИИ-сервисов в работе врачей

ортопедов-травматологов. Потенциально здесь открываются интересные перспективы оптимизации производственных процессов медицинских организаций, однако этот аспект требует дальнейшего изучения.

Также остаётся неизученным вопрос эффективности внедрения ТИИ для диагностики ДОО крупных суставов нижних конечностей. Представляет интерес публикация, доказывающая достижение идентичности диагностических решений врачей-ортопедов с разным стажем (опытного врача-ортопеда и молодого резидента) относительно наличия и степени выраженности гонартроза. Такой результат обеспечивается за счёт применения технологий искусственного интеллекта (автоматизированного анализа — классификации результатов рентгенографии по Келлгрэну–Лоуренсу, измерения суставной щели, выявления остеосклероза и остеофитов) [18]. Нередко звучат заявления о том, что ИИ снижает нагрузку на врача-рентгенолога. Однако они не подкрепляются конкретными данными, измеримыми результатами [19]. Очевидно, здесь требуются дальнейшие углублённые исследования.

Наконец, третье направление перспективных научных изысканий состоит в разработке ТИИ для одновременной обработки разных видов медицинских данных — собственно рентгенограммы, данных анамнеза, физикальных, лабораторных исследований и т.д. Опубликован первый опыт создания таких нейросетей для определения степени риска развития тяжёлого ДОО и соответствующего прогнозирования необходимости тотального эндопротезирования как коленного, так и тазобедренного сустава. Представленные результаты внушают некоторый оптимизм, но явно требуют дальнейшего изучения [20].

Активное развитие ТИИ в рентгенодиагностике тесно связано с появлением в свободном доступе большого объёма данных для обучения соответствующих моделей. В контексте темы статьи таким набором данных является датасет из «Инициативы по остеоартрозу» — десятилетнего многоцентрового обсервационного исследования, проводимого в США (<https://nda.nih.gov/oai>). Этот набор данных стал доступным широкой публике в 2020 году. Подавляющее большинство ТИИ, представленных выше, разработаны и протестированы именно с его использованием. Это является серьёзным ограничением, критично влияющим на возможность масштабирования таких разработок и воспроизводимость результатов их работы на новых данных, то есть в медицинских организациях иных стран и регионов мира.

Важно отметить, что для развития ТИИ большое значение имеет предоставление наборов данных по научно обоснованной методологии [8]. Таким образом, для осуществления следующего скачка в развитии ИИ и его приближения к работе врача-рентгенолога крайне важна разработка методологии подготовки динамических наборов данных, а также наборов данных, обогащённых дополнительной клинической информацией.

Ограничения исследования

В рамках настоящего исследования не были сопоставлены результаты работы ИИ-сервисов между собой, так как это не являлось целью данной работы.

ЗАКЛЮЧЕНИЕ

В условиях проспективного многоцентрового клинического исследования (Московского Эксперимента) двумя программными продуктами на основе искусственного интеллекта обработаны результаты 198 821 рентгенографии коленного сустава с целью выявления признаков и определения степени тяжести деформирующего остеоартроза. Соответственно, для двух разработок медианы согласия врачей-экспертов с заключением ИИ-сервиса об отсутствии, наличии и характере патологии составили 88,0 и 93,0%, согласия с локализацией патологических изменений — 89,7 и 91,3%, суммарные клинические оценки — 88,1 и 93,0%.

Достигнутый уровень развития программного обеспечения на основе искусственного интеллекта позволяет применять соответствующие разработки для повышения точности и производительности труда врачей-рентгенологов при описании результатов рентгенографии крупных суставов нижних конечностей (в контексте диагностики деформирующего остеоартроза).

Перспективы дальнейших исследований состоят в изучении медицинской, социальной и экономической эффективности внедрения технологий искусственного интеллекта в процессы лучевой диагностики заболеваний и травм опорно-двигательной системы; в разработке ТИИ для комплексного анализа разных видов медицинских данных, в том числе с целью прогнозирования течения дегенеративно-дистрофических и, возможно, репаративных процессов. Особой научно-практической разработки требует вопрос интеграции ИИ-сервисов для автоматизированного анализа результатов лучевых исследований опорно-двигательной системы в работу врачей ортопедов-травматологов, в том числе с целью оптимизации производственных процессов и кадровых ресурсов медицинских организаций.

ДОПОЛНИТЕЛЬНО

Вклад авторов. Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией). Наибольший вклад распределён следующим образом: А.В. Владимирский, Ю.А. Васильев — разработка концепции, подготовка и написание текста, К.М. Арзамасов — предоставление и обработка данных, В.Е. Казаринова — подготовка финальной версии рукописи, Е.В. Астапенко — анализ изображений, работа с текстом.

Источник финансирования. Данная статья подготовлена авторским коллективом в рамках НИР «Научные методологии

устойчивого развития технологий искусственного интеллекта в медицинской диагностике» (№ ЕГИСУ 123031500004-5) в соответствии с Приказом № 1196 от 21.12.2022 г. «Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счёт 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.

Раскрытие интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с проведённым исследованием и публикацией настоящей статьи.

Рассмотрение и рецензирование. Настоящая работа подана в журнал в инициативном порядке и рассмотрена по обычной процедуре. В рецензировании участвовали два внешних рецензента, член редакционной коллегии и научный редактор издания.

Благодарности. Благодарим Румянцева Дениса Андреевича за помощь в подборе показательных исследований.

ADDITIONAL INFO

Author contribution. All authors confirm that their authorship meets the international ICMJE criteria (all authors have made a significant contribution to the development of the concept, research

and preparation of the article, read and approved the final version before publication). The greatest contribution is distributed as follows: A.V. Vladzimirskyy, Y.A. Vasilev — concept development, text preparation and writing; K.M. Arzamasov — data submission and processing; V.E. Kazarinova — preparation of the final version of the manuscript, E.V. Astapenko — analysis of radiological images, text preparation.

Funding source. This paper was prepared by a group of authors as a part of the research and development effort titled “Evidence-based methodologies for sustainable development of artificial intelligence in medical imaging” (USIS No. 123031500004-5) 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.

Disclosure of interests. The authors declare that they have no competing interests.

Provenance and peer-review. This paper was submitted to the journal on an initiative basis and reviewed according to the usual procedure. Two external reviewers, a member of the editorial board and the scientific editor of the publication participated in the review.

Acknowledgments. We thank Denis A. Rumyantsev for help in the selection of demonstration studies.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

1. Khan SD, Hoodbhoy Z, Raja MHR, et al. Frameworks for procurement, integration, monitoring, and evaluation of artificial intelligence tools in clinical settings: A systematic review. *PLOS Digit Health*. 2024;3(5):e0000514. doi: 10.1371/journal.pdig.0000514
2. Nowroozi A, Salehi MA, Shobeiri P, et al. Artificial intelligence diagnostic accuracy in fracture detection from plain radiographs and comparing it with clinicians: a systematic review and meta-analysis. *Clin Radiol*. 2024;S0009-9260(24)00200-9. doi: 10.1016/j.crad.2024.04.009
3. Vasilev YA, Vladzimirskyy AV, editors. *Computer vision in radiation diagnostics: the first stage of the Moscow experiment: a monograph*. Moscow: Publishing Solutions; 2022. 388 p. (In Russ.).
4. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527. doi: 10.1136/bmj.h5527
5. *Basic recommendations for the work of artificial intelligence services for radial diagnostics: Methodological Recommendations No. 54*. Moscow: Scientific and Practical Clinical Centre for Diagnostics and Telemedicine Technologies of the Moscow City Health Department; 2022. 68 p. (In Russ.).
6. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol*. 2022;75(1):25–36. doi: 10.4097/kja.21209
7. *Clinical trials of artificial intelligence systems (radiation diagnostics)*. Moscow: State budgetary institution of health care of Moscow “Scientific and Practical Clinical Centre for Diagnostics and Telemedicine Technologies of the Department of Health Care of Moscow”; 2023. 40 p. (In Russ.).
8. *Preparation of data set for training and testing of software based on artificial intelligence technology*. (Tutorial) Ridero: Scientific and Practical Clinical Centre for Diagnostics and Telemedicine Technologies of the Moscow City Health Department; 2024. 140 p. (In Russ.).
9. Vasilev YuA, Vladzimirskyy AV, Omelyanskaya OV, et al. Methodology of testing and monitoring of software based on artificial intelligence technologies for medical diagnostics. *Digital Diagnostics*. 2023;4(3):252–267. (In Russ.). doi: 10.17816/DD321971
10. Chetverikov SF, Arzamasov KM, Andreichenko AE, et al. Approaches to sample formation for quality control of artificial intelligence systems in biomedical research. *Modern Technologies in Medicine*. 2023;15(2):19–25. (In Russ.). doi: 10.17691/stm2023.15.2.02
11. Yang J, Ji Q, Ni M, et al. Automatic assessment of knee osteoarthritis severity in portable devices based on deep learning. *J Orthop Surg Res*. 2022;17(1):540. doi: 10.1186/s13018-022-03429-2
12. Wang CT, Huang B, Thogiti N, et al. Successful real-world application of an osteoarthritis classification deep-learning model using 9210 knees—An orthopedic surgeon’s view. *J Orthop Res*. 2023;41(4):737–746. doi: 10.1002/jor.25415
13. von Schacky CE, Sohn JH, Liu F, et al. Development and Validation of a Multitask Deep Learning Model for Severity Grading of Hip Osteoarthritis Features on Radiographs. *Radiology*. 2020;295(1):136–145. doi: 10.1148/radiol.2020190925
14. Magnéli M, Borjali A, Takahashi E, et al. Application of deep learning for automated diagnosis and classification of hip dysplasia on plain radiographs. *BMC Musculoskelet Disord*. 2024;25(1):117. doi: 10.1186/s12891-024-07244-0

15. Pi SW, Lee BD, Lee MS, et al. Ensemble deep-learning networks for automated osteoarthritis grading in knee X-ray images. *Sci Rep*. 2023;13(1):22887. doi: 10.1038/s41598-023-50210-4
16. Lenskjold A, Brejnøbøl MW, Nybing JU, et al. Constructing a clinical radiographic knee osteoarthritis database using artificial intelligence tools with limited human labor: A proof of principle. *Osteoarthritis Cartilage*. 2024;32(3):310–318. doi: 10.1016/j.joca.2023.11.014
17. Naguib SM, Kassem MA, Hamza HM, et al. Automated system for classifying uni-bicompartmental knee osteoarthritis by using redefined residual learning with convolutional neural network. *Heliyon*. 2024;10(10):e31017. doi: 10.1016/j.heliyon.2024.e31017

18. Smolle MA, Goetz C, Maurer D, et al. Artificial intelligence-based computer-aided system for knee osteoarthritis assessment increases experienced orthopaedic surgeons' agreement rate and accuracy. *Knee Surg Sports Traumatol Arthrosc*. 2023;31(3):1053–1062. doi: 10.1007/s00167-022-07220-y
19. Yoon JS, Yon CJ, Lee D, et al. Assessment of a novel deep learning-based software developed for automatic feature extraction and grading of radiographic knee osteoarthritis. *BMC Musculoskelet Disord*. 2023;24(1):869. doi: 10.1186/s12891-023-06951-4
20. Salis Z, Driban JB, McAlindon TE. Predicting the onset of end-stage knee osteoarthritis over two- and five-years using machine learning. *Semin Arthritis Rheum*. 2024;66:152433. doi: 10.1016/j.semarthrit.2024.152433

ОБ АВТОРАХ

Васильев Юрий Александрович, канд. мед. наук;
ORCID: 0000-0002-5283-5961;
eLibrary SPIN: 4458-5608;
e-mail: npcmr@zdrav.mos.ru

Владимирский Антон Вячеславович, д-р мед. наук;
ORCID: 0000-0002-2990-7736;
eLibrary SPIN: 3602-7120;
e-mail: npcmr@zdrav.mos.ru

Арзамасов Кирилл Михайлович, канд. мед. наук;
ORCID: 0000-0001-7786-0349;
eLibrary SPIN: 3160-8062;
e-mail: ArzamasovKM@zdrav.mos.ru

*** Казаринова Вероника Евгеньевна**;
ORCID: 0009-0001-3568-8138;
eLibrary SPIN: 5901-5577;
e-mail: KazarinovaVE@zdrav.mos.ru

Астапенко Елена Васильевна;
ORCID: 0009-0006-6284-2088;
eLibrary SPIN: 7362-8553;
e-mail: AstapenkoEV1@zdrav.mos.ru

AUTHORS' INFO

Yuriy A. Vasilev, MD, Cand. Sci. (Medicine);
ORCID: 0000-0002-5283-5961;
eLibrary SPIN: 4458-5608;
e-mail: npcmr@zdrav.mos.ru

Anton V. Vladzimirsky, MD, Dr. Sci. (Medicine);
ORCID: 0000-0002-2990-7736;
eLibrary SPIN: 3602-7120;
e-mail: npcmr@zdrav.mos.ru

Kirill M. Arzamasov, MD, Cand. Sci. (Medicine);
ORCID: 0000-0001-7786-0349;
eLibrary SPIN: 3160-8062;
e-mail: ArzamasovKM@zdrav.mos.ru

*** Veronika E. Kazarinova**;
ORCID: 0009-0001-3568-8138;
eLibrary SPIN: 5901-5577;
e-mail: KazarinovaVE@zdrav.mos.ru

Elena V. Astapenko;
ORCID: 0009-0006-6284-2088;
eLibrary SPIN: 7362-8553;
e-mail: AstapenkoEV1@zdrav.mos.ru

* Автор, ответственный за переписку / Corresponding author