

УДК 004:614.2

DOI: [10.26102/2310-6018/2025.48.1.027](https://doi.org/10.26102/2310-6018/2025.48.1.027)

Метод оценки автономного программного обеспечения на основе технологий искусственного интеллекта для массовых профилактических исследований

В.В. Зинченко^{1✉}, Р.А. Ерижоков^{1,2,3}, К.М. Арзамасов^{1,4}

¹Научно-практический клинический центр диагностики и телемедицинских технологий
Департамента здравоохранения города Москвы, Москва,
Российская Федерация

²Московский физико-технический институт, Московская область, Долгопрудный,
Российская Федерация

³Первый Московский государственный медицинский университет имени И.М. Сеченова
Министерства здравоохранения Российской Федерации (Сеченовский Университет),
Москва, Российская Федерация

⁴МИРЭА – Российский технологический университет, Москва, Российская Федерация

Резюме. Внедрение технологий искусственного интеллекта (ИИ) в медицинскую практику требует тщательной оценки их эффективности, особенно для систем, работающих в автономном режиме. Предлагаемый в данном исследовании метод основан на синтезе требований национальных стандартов в области медицинского ИИ, разработанных экспертами Научно-практического клинического центра диагностики и телемедицинских технологий, и данных, полученных в рамках «Московского эксперимента» по внедрению инновационных технологий. Апробация проведена на трех программных продуктах ИИ, используемых для анализа флюорографических исследований в период с января по май 2023 года. Оценка включала анализ точности алгоритмов (чувствительность, специфичность), а также их эффективности в реальных клинических условиях. В исследовании приведен сравнительный анализ результатов с количественной интерпретацией данных. Наибольший акцент при оценивании был на обеспечении системой ИИ высокого уровня диагностической чувствительности, которая позволит в массовых профилактических исследованиях избавить врачей от рутинной монотонной работы. Разработанный метод продемонстрировал возможность комплексной оценки автономных систем ИИ, выявив различия в эффективности продуктов по ключевым метрикам. Предложенный метод позволяет систематизировать процесс валидации медицинских ИИ-решений, минимизируя риски их некорректного использования в условиях автономной работы. Результаты исследования могут быть применены для стандартизации оценки ИИ-инструментов в радиологии и других областях медицины, требующих высокого уровня надежности диагностики.

Ключевые слова: искусственный интеллект в медицине, автономные диагностические системы, оценка эффективности, лучевая диагностика, рентгенология.

Благодарности: Статья подготовлена авторским коллективом в рамках НИР «Научные методологии устойчивого развития технологий искусственного интеллекта в медицинской диагностике» (№ ЕГИСУ: 123031500004-5) в соответствии с Приказом от 21.12.2022 г. № 1196 «Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счет средств бюджета города Москвы государственным бюджетным (автономным) учреждениям подведомственным Департаменту здравоохранения города Москвы, на 2023 год и плановый период 2024 и 2025 годов» Департамента здравоохранения города Москвы.

Для цитирования: Зинченко В.В., Ерижоков Р.А., Арзамасов К.М. Метод оценки автономного программного обеспечения на основе технологий искусственного интеллекта для массовых

профилактических исследований. *Моделирование, оптимизация и информационные технологии*. 2025;13(1). URL: <https://moitvvt.ru/ru/journal/pdf?id=1822> DOI: 10.26102/2310-6018/2025.48.1.027

Method of evaluation of autonomous software based on artificial intelligence technologies for mass preventive studies

V.V. Zinchenko¹✉, R.A. Erizhokov^{1,2,3}, K.M. Arzamasov^{1,4}

¹*Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, the Russian Federation*

²*Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, the Russian Federation*

³*Federal State Autonomous Educational Institution of Higher Education I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, the Russian Federation*

⁴*MIREA – Russian Technological University, Moscow, the Russian Federation*

Abstract. The introduction of artificial intelligence (AI) technologies into medical practice requires a thorough assessment of their effectiveness, especially for systems operating autonomously. The method proposed in this study is based on a synthesis of the requirements of national standards in the field of medical AI developed by experts of the Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies and data obtained as part of the "Moscow Experiment" on the introduction of innovative technologies. The testing was carried out on three AI software products used to analyze fluorographic studies in the period from January to May 2023. The evaluation included an analysis of the accuracy of algorithms (sensitivity, specificity), effectiveness in real clinical conditions, as well as a comparative analysis of the results with a quantitative interpretation of the data. The emphasis in the evaluation was on providing the AI system with a high level of diagnostic sensitivity, which will allow doctors to relieve themselves of routine monotonous work in mass preventive studies. The developed method demonstrated the possibility of a comprehensive assessment of autonomous AI systems, identifying differences in the effectiveness of products by key metrics. The proposed method allows systematizing the process of validating medical AI solutions, minimizing the risks of their incorrect use in autonomous operation. The results of the study can be used to standardize the assessment of AI tools in radiology and other areas of medicine that require a high level of diagnostic reliability.

Keywords: artificial intelligence in medicine, autonomous diagnostic systems, efficiency assessment, radiation diagnostics, radiology.

Acknowledgments: The article was prepared by a team of authors within the framework of the research work "Scientific methodologies for sustainable development of artificial intelligence technologies in medical diagnostics" (EGISU No.: 123031500004-5) in accordance with the Order of 21.12.2022 No. 1196 "On approval of state assignments, financial support for which is carried out at the expense of the Moscow city budget to state budgetary (autonomous) institutions subordinate to the Moscow City Department of Healthcare, for 2023 and the planning period of 2024 and 2025" of the Moscow City Department of Healthcare.

For citation: Zinchenko V.V., Erizhokov R.A., Arzamasov K.M. Method of evaluation of autonomous software based on artificial intelligence technologies for mass preventive studies. *Modeling, Optimization and Information Technology*. 2025;13(1). (In Russ.). URL: <https://moitvvt.ru/ru/journal/pdf?id=1822> DOI: 10.26102/2310-6018/2025.48.1.027

Введение

В настоящее время информатизация и автоматизация здравоохранения приводит к крупным изменениям в процессах производства и существующих практиках. Применение различных информационных и телекоммуникационных технологий

охватывает все уровни и этапы медицинской помощи, а значит важно оценивать их результативность и эффективность [1–3]. В соответствии с постановлением Правительства РФ от 03.04.2021 № 542 созданы и внедрены в практическую деятельность ряд инструментов для оценки цифровой зрелости продуктов, на основе искусственного интеллекта (ИИ)^{1, 2}.

На данный момент имеющиеся достижения ИИ в области здравоохранения уже позволяют снижать экономические затраты за счет увеличения диагностической точности и оптимизации работы врачей-рентгенологов [4–5]. Следующим этапом развития в данной области видится именно автономный ИИ, имеющий максимальную чувствительность, что позволит в массовых профилактических исследованиях избавить врачей-рентгенологов от рутинной монотонной работы в части выявления исследований без патологии, которая по разным источникам варьируется на уровне 81,6–99,93 % [6–10]. Однако для реализации данной возможности требуется отобрать программное обеспечение (ПО) на основе технологий искусственного интеллекта (ТИИ) «высокого доверия», а затем регулярно следить за результатами их работы.

Существует множество решений и методик для оценки точности или эффективности отдельного ПО на основе ТИИ [11–13], в том числе разработаны национальные стандарты экспертами Научно-практического клинического центра диагностики и телемедицинских технологий Департамента здравоохранения города Москвы (НПКЦ ДИТ ДЗМ), в основу которых легли методы оценки медицинского искусственного интеллекта для клинической медицины, а также разработаны ряд программ для ЭВМ для оценки качества и эффективности ТИИ³, но при этом не учитывающие специфику автономного ИИ.

Таким образом, учитывая быстрое развития медицинских интеллектуальных систем [14], остро стоит необходимость разработки новых методов для оценки их применимости, качества и эффективности. Однако, существующие инструменты оценки ПО на основе ТИИ часто ограничиваются только аспектами автоматизации, рассматривая с позиции СППВР и оценивая по показателям чувствительности и специфичности, но для принятия или непринятия автономных систем необходима разработка интегрального показателя, демонстрирующего одновременно и эффективность, и результативность функционирования ПО на основе ТИИ [15–17]. Таким образом, цель исследования заключается в разработке и валидации (апробации) метода оценки ПО на основе ТИИ для автономного применения в массовых профилактических исследованиях.

Материалы и методы

Схема работы автономного ИИ представлена на Рисунке 1. Данное исследование направлено на разработку двухэтапного метода оценки автономных систем искусственного интеллекта (СИИ) для анализа флюорографических исследований (ФЛГ) в условиях требования максимальной чувствительности.

¹ Национальный стандарт РФ ГОСТ Р 59921.1-2022. Системы искусственного интеллекта в клинической медицине. Часть 1. Клиническая оценка. Москва: Российский институт стандартизации; 2022. 20 с.

² Национальный стандарт РФ ГОСТ Р 71673-2024. Системы искусственного интеллекта в лучевой диагностике. Алгоритмы анализа медицинских изображений. Методы испытаний на определение точности измерений. Москва: Российский институт стандартизации; 2024. 12 с.

³ Ахмад Е.С., Васильев Ю.А., Гусев А.В., Зинченко В.В., Арзамасов К.М., Сурков А.Г. Программа оценки коэффициентов безопасности, эффективности и качества при выполнении мониторинга системы искусственного интеллекта в лучевой диагностике: опубл. 17.05.2024. Свидетельство о государственной регистрации программы для ЭВМ № 2024618798 Российская Федерация.

Первый этап предполагает оценку точности и клинической целесообразности ПО с использованием четырехпольной таблицы. На момент разработки предлагаемого метода критерием целесообразности являлась способность СИИ относить к классу «норма» не менее 60 % исследований, что обусловлено высокой долей таких случаев в клинической практике. Установление нижнего порога на уровне 60 % направлено на минимизацию избыточной нагрузки на врачей-рентгенологов при автономной работе СИИ. Параметр точности, ограниченный 0,1 % ошибок в категории «норма», отражает требование к максимальной чувствительности алгоритмов: СИИ должны гарантировать, что исследования, классифицированные как «норма», практически не требуют последующего пересмотра специалистом-рентгенологом. Это критически важно для сценариев скрининга, где пропуск патологий недопустим.

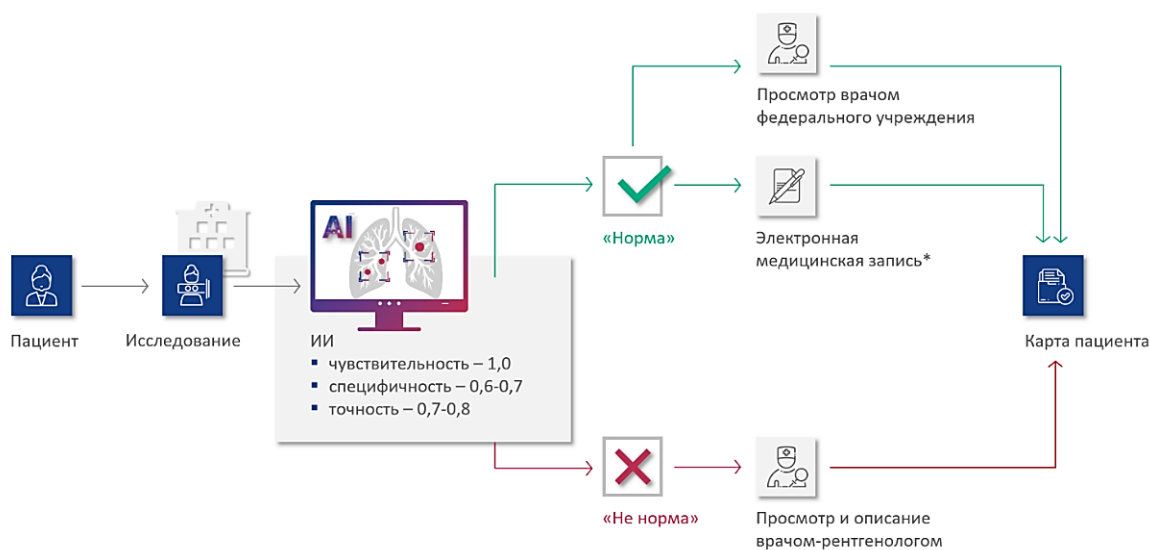


Рисунок 1 – Этапы работы автономного ПО на основе ТИИ
 Figure 1 – Stages of operation of autonomous software based on the AI

На втором этапе работы метода проводится оценка эффективности ИИ-решений, с использованием целевых значений, введенных в работе [18], к которым должны стремиться показатели точности и целесообразности ПО на основе ТИИ.

Данное исследование проводилось в рамках Московского эксперимента – масштабной инициативы по интеграции технологий компьютерного зрения в систему здравоохранения Москвы. В исследовании использованы данные 300000 флюорографических исследований, обработанных тремя анонимизированными в процессе экспериментов СИИ, разработанными ООО «Медицинские скрининг системы», ООО «Платформа Третье Мнение» и ООО «ФтизисБиоМед». Все системы были интегрированы с Единым радиологическим информационным сервисом (ЕРИС ЕМИАС) и протестированы в соответствии с методологией, утвержденной Постановлением Правительства Москвы [19]. Ключевые задачи ПО включали автоматическую сортировку результатов на «норму» и «патологию», маркировку патологических изменений и генерацию шаблонов клинических описаний [20]. После интеграции каждая система прошла технологический и клинический мониторинг, направленный на обеспечение соответствия установленным стандартам качества. Процедура мониторинга, подробно описанная в [21], предусматривала постоянную оценку работы алгоритмов в реальных условиях, включая анализ ошибок и адаптацию к

изменяющимся клиническим требованиям. Схема взаимодействия автономного ИИ с инфраструктурой ЕРИС ЕМИАС представлена на Рисунке 2.

Также в процессе проведения Московского эксперимента был разработан NLP-инструмент для автоматизированного анализа заключения и описания врача-рентгенолога, с последующей классификацией каждого исследования на норму/патологию. Данный инструмент позволяет сравнить результаты анализа исследования врачей с ПО на основе ТИИ. В качестве эталонных результатов (GT) используется заключение врача. В процессе исследования выявляются количество исследований, оцененных: и врачом и ИИ как патология (TP); врачом как норма, ИИ как патология (FP); врачом как патология, ИИ как норма (TN); и врачом и ИИ как норма (FN). Результаты с TN также дополнительно пересматривались врачами-экспертами НПКЦ ДИТ ДЗМ для объективизации результатов оценки. Эти показатели были использованы в качестве базовых для формирования четырехпольной таблицы.



Рисунок 2 – Этапы ввода и тестирований ПО на основе ТИИ в условиях Московского Эксперимента

Figure 2 – Stages of input and testing of software based on the AI under the conditions of the Moscow Experiment

Результаты

Предлагаемый метод оценки программного обеспечения на основе технологий искусственного интеллекта в условиях максимальной чувствительности предусматривает визуализацию получаемых результатов. Для этого построен график, отражающий ежемесячные показатели работы трех ИИ-систем при анализе флюорографических исследований на протяжении пяти последовательных месяцев. По оси абсцисс (X) отложено отношение количества случаев, классифицированных ИИ как «без патологии», к общему числу исследований, обработанных системой $((FN+TN)/N)$. По оси ординат (Y) представлена доля истинно-отрицательных результатов (TN) среди всех случаев, отнесенных ИИ к категории «норма», выраженная в процентах $(TN/(FN+TN) \times 100)$.

На график нанесены две регулируемые границы: горизонтальная линия «0,1%», соответствующая максимально допустимому проценту ошибок в исследованиях, маркированных как «норма» (ложноотрицательные результаты), и вертикальная линия «0,6», определяющая минимальный порог выявления «нормы» в общем потоке данных

(60 %). Эти критерии, заданные в рамках эксперимента, могут быть адаптированы в зависимости от клинических требований к конкретным ИИ-решениям.

Данные, полученные в результате оценки трех ИИ-продуктов для ФЛГ, отображены на Рисунке 3 в виде 15 точек, соответствующих ежемесячным результатам каждого ПО за пятимесячный период. Распределение точек демонстрирует вариабельность показателей между системами, что подчеркивает необходимость индивидуальной настройки алгоритмов под конкретные клинические задачи.

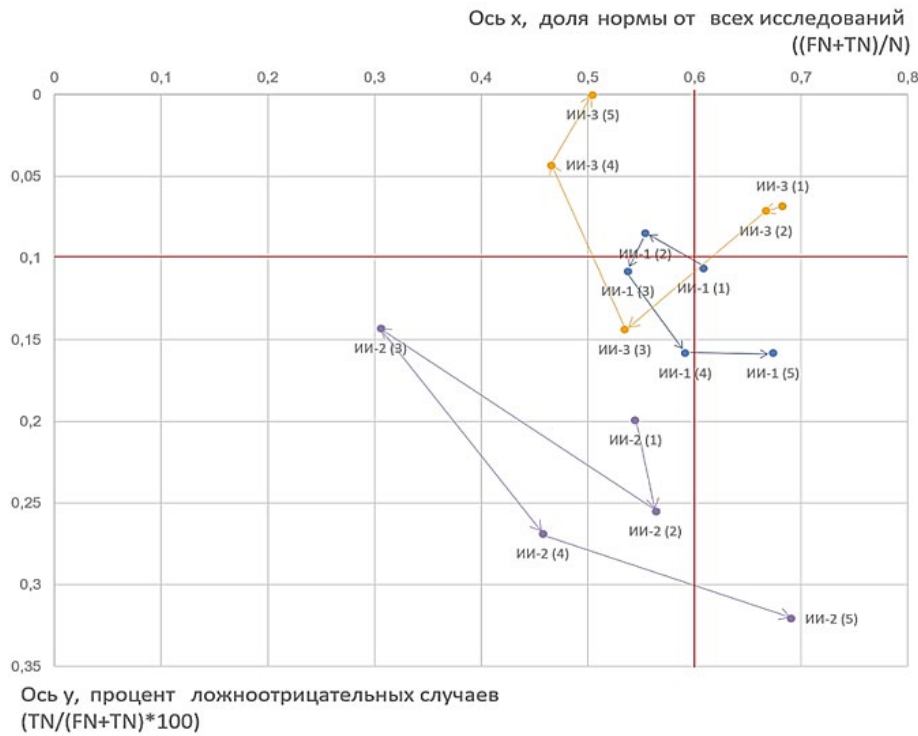


Рисунок 3 – Оценка работы трех ПО на основе ТИИ для модальности ФЛГ
Figure 3 – Evaluation of the performance of three software based on the AI for the radiology

Четырехпольная таблица, представленная на Рисунке 4, служит инструментом для анализа точности и целесообразности применения ИИ-систем в сценариях с максимальной чувствительностью. Она позволяет наглядно сопоставить способность алгоритмов минимизировать ложные заключения при сохранении высокой доли корректно идентифицированных случаев «нормы».

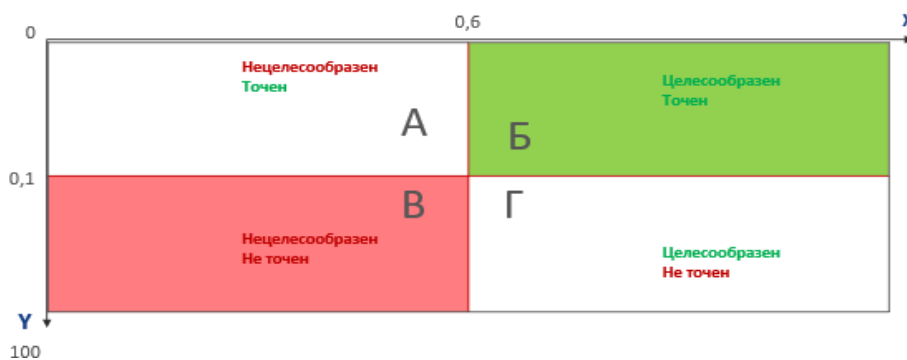


Рисунок 3 – Четырехпольная таблица оценки ПО с ИИ
Figure 3 – Four-field table for evaluating AI software

Предложенный графический подход не только стандартизирует процесс валидации, но и предоставляет разработчикам инструмент для оптимизации параметров ИИ-решений. Анализ положения точек относительно установленных границ позволяет выявить случаи превышения допустимой ошибки или недостаточной доли «нормы», что критически важно для обеспечения безопасности автономной работы систем в условиях массового скрининга. Интеграция четырехпольной таблицы в методологию оценки способствует формированию прозрачных критериев, согласующихся как с техническими возможностями ИИ, так и с клиническими требованиями к диагностической точности.

Также помимо оценки точности и целесообразности использования (в случаях, когда несколько ПО на основе ТИИ попадают в одну зону оценки) для комплексности проведенного анализа и полноценной оценки ПО на основе ТИИ требуется и оценка эффективности. Данную оценку можно также визуализировать на сформированном графике, добавив на него целевое значение, к которому ПО на основе ТИИ следует стремиться. В данном случае, по данным из литературных источников [22] известно, что при ФЛГ в среднем выявляется 90 % нормы (0,9 на оси X). При этом мы ожидаем, что ПО на основе ТИИ не будет совершать ошибок (0 на оси Y) (Рисунок 5).

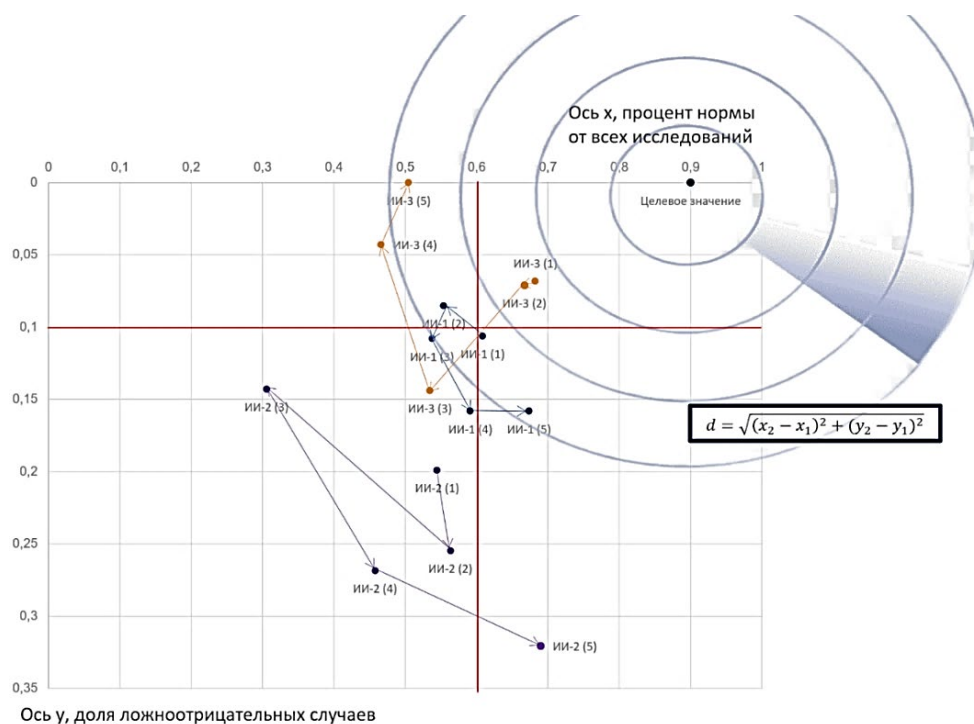


Рисунок 5 – Количественная оценка работы трех ПО на основе ТИИ
Figure 5 – Quantitative assessment of the performance of three software based on the AI

После выявления целевого значения по формуле (1) рассчитывается расстояние от каждой точки А с координатами (x_1, y_1) , представляющей ПО на основе ТИИ, до целевого значения В с координатами (x_2, y_2) :

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

Этот шаг позволяет количественно оценить, насколько близко сервисы находятся к «идеальному» показателю.

Исходя из полученных данных, с использованием формулы (2) вычисляется показатель эффективности ПО на основе ТИИ. Этот показатель разработан для классификации по уровням эффективности или для выдачи более детализированной оценки. Такая шкала может стать полезным инструментом при принятии решений об использовании или доработке ПО.

$$E = \left(\frac{0,9-d}{0,9} \right) \cdot 100\%, \quad (2)$$

где d – расстояние от целевого значения до значения ПО на основе ТИИ.

В результате расчетов выявляется эффективность ПО на основе ТИИ с учетом точности и целесообразности использования (Рисунок 6).

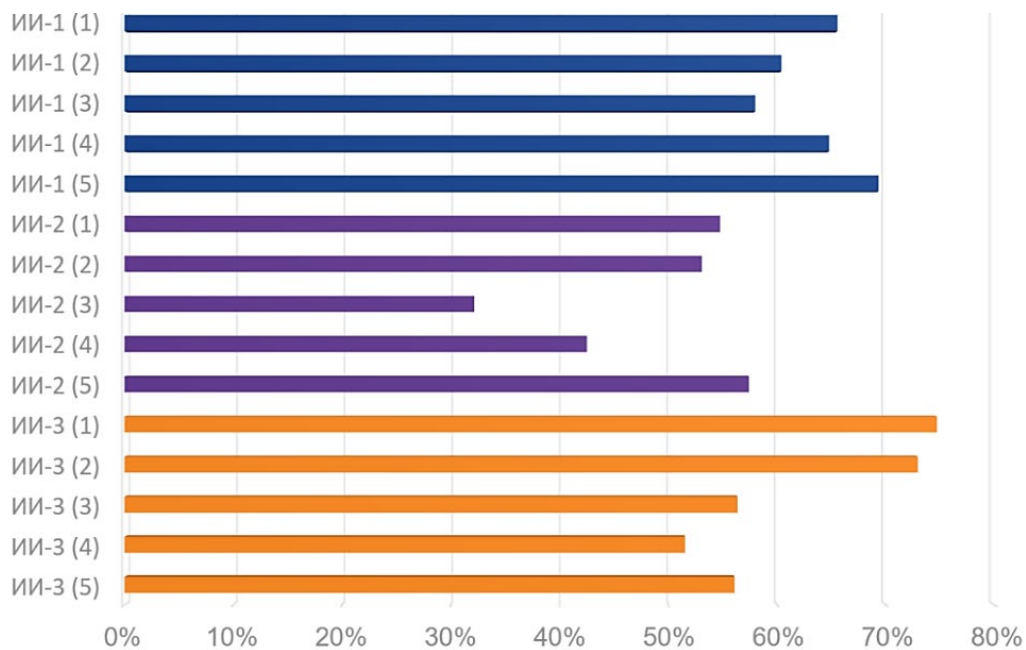


Рисунок 6 – Расчет эффективности для ПО на основе ТИИ
Figure 6 – Calculation of efficiency for software based on the AI

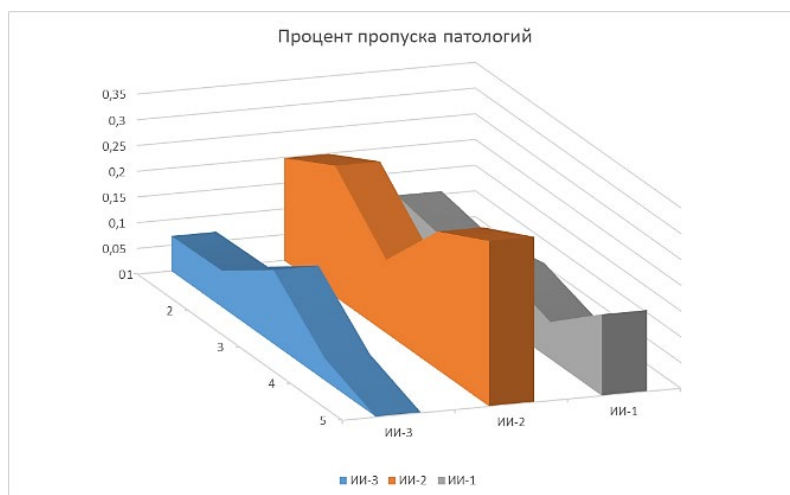


Рисунок 7 – Оценка пропуска патологий при работе ПО на основе ТИИ
Figure 7 – Assessment of missed pathologies when working with software based on the AI

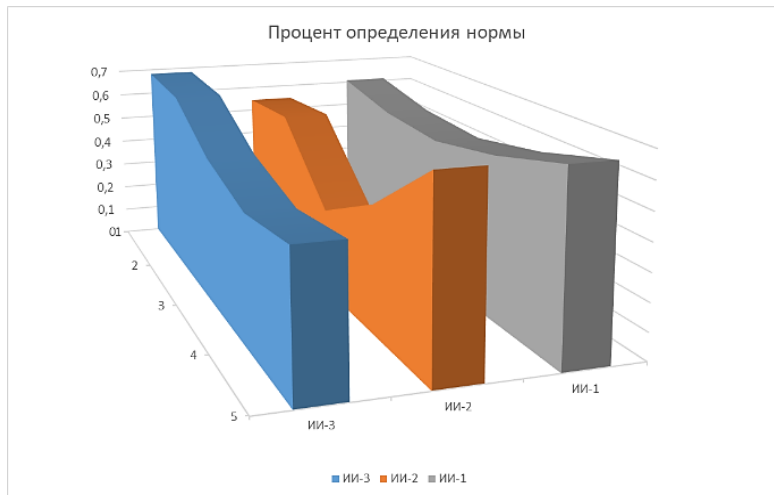


Рисунок 8 – Оценка определения нормы при работе ПО на основе ТИИ
Figure 8 – Evaluation of the definition of the norm when working with software based on the AI

Обсуждение

Рост числа публикаций, посвященных применению искусственного интеллекта (ИИ) в здравоохранении, демонстрирует экспоненциальную динамику [23]. Это отражает не только возрастающий интерес к технологиям, но и острую потребность в методах оценки программного обеспечения (ПО) на основе ИИ, особенно в контексте массовых профилактических исследований. Разработанная методика, направленная на комплексный анализ точности, целесообразности и эффективности автономных ИИ-систем, приобретает ключевое значение для принятия обоснованных решений как разработчиками, так и организаторами здравоохранения.

Результаты апробации метода, представленные на Рисунках 4–8, выявили значительную вариабельность в работе трех тестируемых ИИ-решений (ИИ-1, ИИ-2, ИИ-3) для анализа флюорографических исследований. Так, лишь в 13,3 % случаев (2 из 15 точек на Рисунке 4) системы одновременно соответствовали критериям высокой точности и клинической целесообразности. В 20 % случаев (3 из 15) ПО демонстрировало целесообразность при недостаточной точности, а в аналогичной доле – высокую точность, но нецелесообразность использования. При этом почти половина результатов (46,7 %, 7 из 15) характеризовалась одновременным нарушением обоих критериев.

Динамика работы ИИ-1 выявила регресс в первые три месяца с последующим улучшением показателей идентификации «нормы» на пятом месяце, что, однако, сопровождалось ростом доли пропущенных патологий. Отсутствие обратной связи от разработчика не позволило установить причины таких колебаний. ИИ-2 продемонстрировал схожую нестабильность: резкое снижение доли «нормы» до 20 % на третьем месяце сменилось ее повышением до 70 % к концу периода, что подтвердило отсутствие доработок алгоритма и произвольную регулировку порогов классификации.

Наиболее показательными стали результаты ИИ-3. В первые два месяца система соответствовала ожидаемым параметрам, однако на третьем месяце произошел резкий спад как в доле корректно идентифицированной «нормы», так и в точности. Как выяснилось, это было связано с добавлением нового признака «консолидированные переломы», что потребовало перекалибровки алгоритма. К пятому месяцу ИИ-3 полностью устранил пропуски патологий, вернувшись к приемлемым показателям.

Полученные данные подчеркивают необходимость динамического мониторинга ИИ-систем, особенно в условиях их автономной работы. Предложенный метод,

интегрирующий визуализацию данных (Рисунки 4–8) и анализ временных трендов, позволяет не только выявлять критические отклонения, но и устанавливать их причины – от технических ошибок до некорректной настройки порогов. Регулярная оценка в соответствии с данной методикой способна минимизировать риски, связанные с внедрением ИИ в клиническую практику, обеспечивая баланс между автоматизацией, точностью и экономической эффективностью.

Таким образом, только непрерывная валидация и адаптация алгоритмов, основанная на прозрачном взаимодействии с разработчиками, позволяет реализовать потенциал ИИ в здравоохранении без ущерба для качества диагностики.

Заключение

Предложенный метод по оценке ПО на основе ТИИ может позволить разработчикам осуществлять тонкую настройку ПО на основе ТИИ для оптимизации его работы, а также принимать организаторам здравоохранения решения о дальнейшей работе автономного ПО. Разработанный подход был апробирован для модальности ФЛГ, при этом может быть масштабирован и использоваться для других модальностей в области клинической медицины, так как имеет в своей основе универсальные показатели и параметры.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Пугачев П.С., Гусев А.В., Кобякова О.С. и др. Мировые тренды цифровой трансформации отрасли здравоохранения. *Национальное здравоохранение*. 2021;2(2):5–12. <https://doi.org/10.47093/2713-069X.2021.2.2.5-12>
Pugachev P.S., Gusev A.V., Kobyakova O.S., et al. Global trends in the digital transformation of the healthcare industry. *National Health Care (Russia)*. 2021;2(2):5–12. (In Russ.). <https://doi.org/10.47093/2713-069X.2021.2.2.5-12>
2. Hoc Group on Application of AI Technologies. Artificial Intelligence in Healthcare: Directions of Standardization. In: *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects*. Cham: Springer; 2022. pp. 231–257. https://doi.org/10.1007/978-3-030-83620-7_10
3. Кошечкин К.А., Свистунов А.А., Лебедев Г.С., Фартушный Э.Н. Практика применения систем на основе искусственного интеллекта в сфере обращения лекарственных средств. *Вестник Росздравнадзора*. 2022;(3):27–33.
Koshechkin K.A., Svistunov A.A., Lebedev G.S., Fartushny E.N. Practice of using systems based on artificial intelligence in the field of circulation of medicines. *Vestnik Roszdravnadzora*. 2022;(3):27–33. (In Russ.).
4. Johnson A.E.W., Pollard T.J., Berkowitz S.J., et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data*. 2019;6. <https://doi.org/10.1038/s41597-019-0322-0>
5. Do Hu.M., Spear L.G., Nikpanah M., et al. Augmented Radiologist Workflow Improves Report Value and Saves Time: A Potential Model for Implementation of Artificial Intelligence. *Academic Radiology*. 2020;27(1):96–105. <https://doi.org/10.1016/j.acra.2019.09.014>
6. Ломаков С.Ю., Строгонов Е.А. Современные аспекты организации профилактических лучевых исследований методом флюорографии. *Международный академический вестник*. 2019;(11):4–6.
7. Маркелов Ю.М., Щеголева Л.В. Клинико-экономические аспекты выявления туберкулеза при массовых флюорографических осмотрах населения. *Вестник*

- рентгенологии и радиологии*. 2021;102(3):148–154. <https://doi.org/10.20862/0042-4676-2021-102-3-148-154>
- Markelov Yu.M., Shchegoleva L.V. Clinical and Economic Aspects of Tuberculosis Detection During Mass Fluorographic Examinations of the Population. *Journal of radiology and nuclear medicine*. 2021;102(3):148–154. (In Russ.). <https://doi.org/10.20862/0042-4676-2021-102-3-148-154>
8. Маркелов Ю.М., Щеголева Л.В. Оценка клинико-экономической эффективности и влияния массовых флюорографических осмотров на эпидемиологические показатели по туберкулезу в четырех федеральных округах РФ с различным уровнем охвата флюорографическими осмотрами населения. *Туберкулез и болезни легких*. 2023;101(1):8–16. <https://doi.org/10.58838/2075-1230-2023-101-1-8-16>
- Markelov Yu.M., Schegoleva L.V. Evaluation of Clinical and Economic Efficiency and Impact of Mass Fluorography Screening on Tuberculosis Epidemiological Rates in Four Federal Districts of the Russian Federation with Different Levels of Population Coverage with Mass Fluorography Screening. *Tuberculosis and Lung Diseases*. 2023;101(1):8–16. (In Russ.). <https://doi.org/10.58838/2075-1230-2023-101-1-8-16>
9. Рубис Л.В., Маркелов Ю.М. Оценка эффективности профилактических осмотров взрослого населения с целью ранней диагностики туберкулеза. *Российский медицинский журнал*. 2021;27(3):227–235. <https://doi.org/10.17816/0869-2106-2021-27-3-227-235>
- Rubis L.V., Markelov Y.M. Evaluation of the efficiency of preventive examinations of the adult population, aimed at early tuberculosis diagnosis. *Russian Medicine*. 2021;27(3):227–235. (In Russ.). <https://doi.org/10.17816/0869-2106-2021-27-3-227-235>
10. Kelly B.S., Judge C., Bollard S.M., et al. Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE). *European Radiology*. 2022;32:7998–8007. <https://doi.org/10.1007/s00330-022-08784-6>
11. Tanguay W., Acar Ph., Fine B., et al. Assessment of Radiology Artificial Intelligence Software: A Validation and Evaluation Framework. *Canadian Association of Radiologists Journal*. 2023;74(2):326–333. <https://doi.org/10.1177/08465371221135760>
12. Schaffter Th., Buist D.S.M., Lee Ch.I., et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Network Open*. 2020;3(3). <https://doi.org/10.1001/jamanetworkopen.2020.0265>
13. Васильев Ю.А., Владимировский А.В., Омелянская О.В. и др. Методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики. *Digital Diagnostics*. 2023;4(3):252–267. <https://doi.org/10.17816/DD321971>
- Vasiliev Yu.A., Vlazimirsky A.V., Omelyanskaya O.V., et al. Methodology for testing and monitoring artificial intelligence-based software for medical diagnostics. *Digital Diagnostics*. 2023;4(3):252–267. <https://doi.org/10.17816/DD321971>
14. Гусев А.В., Астапенко Е.М., Иванов И.В., Зарубина Т.В., Кобринский Б.А. Принципы формирования доверия к системам искусственного интеллекта для сферы здравоохранения. *Вестник Росздравнадзора*. 2022;(2):25–33.
- Gusev A.V., Astapenko E.M., Ivanov I.V., Zarubina T.V., Kobrinskii B.A. Principles for building confidence in artificial intelligence systems for healthcare. *Vestnik Roszdravnadzora*. 2022;(2):25–33. (In Russ.).
15. Gómez Ó., Mesejo P., Ibáñez Ó., et al. Evaluating artificial intelligence for comparative radiography. *International Journal of Legal Medicine*. 2024;138(1):307–327. <https://doi.org/10.1007/s00414-023-03080-4>

16. Thodberg H.H., Thodberg B., Ahlqvist J., Offiah A.C. Autonomous artificial intelligence in pediatric radiology: the use and perception of BoneXpert for bone age assessment. *Pediatric Radiology*. 2022;52(7):1338–1346. <https://doi.org/10.1007/s00247-022-05295-w>
17. Daye D., Wiggins W.F., Lungren M.P., et al. Implementation of Clinical Artificial Intelligence in Radiology: Who Decides and How? *Radiology*. 2022;305(3):555–563. <https://doi.org/10.1148/radiol.212151>
18. Çallı E., Sogancioglu E., Van Ginneken B., Van Leeuwen K.G., Murphy K. Deep learning for chest X-ray analysis: A survey. *Medical Image Analysis*. 2021;72. <https://doi.org/10.1016/j.media.2021.102125>
19. Морозов С.П., Владзимирский А.В., Шарова Д.Е., Ахмад Е.С., Зинченко В.В. Первые национальные стандарты Российской Федерации на системы искусственного интеллекта в медицине. *Менеджмент качества в медицине*. 2022;(1):58–62.
Morozov S.P., Vladzimirskyy A.V., Sharova D.E., Akhmad E.S., Zinchenko V.V. The first Russian Federation national standards for artificial intelligence systems in medicine. *Quality management in medicine*. 2022;(1):58–62. (In Russ.).
20. Васильев Ю.А., Арзамасов К.М., Владзимирский А.В. и др. *Подготовка набора данных для обучения и тестирования программного обеспечения на основе технологии искусственного интеллекта*. [б.м.]: Издательские решения; 2024. 140 с.
21. Владзимирский А.В., Васильев Ю.А., Арзамасов К.М. и др. *Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента*. [б.м.]: Издательские решения; 2023. 388 с.
22. Васильев Ю.А., Тыров И.А., Владзимирский А.В., Арзамасов К.М., Пестренин Л.Д., Шулькин И.М. Новая модель организации массовых профилактических исследований, основанная на автономном искусственном интеллекте для сортировки результатов флюорографии. *Здоровье населения и среда обитания – ЗНУСО*. 2023;31(11):23–32. <https://doi.org/10.35627/2219-5238/2023-31-11-23-32>
Vasilev Yu.A., Tyrov I.A., Vladzimirskyy A.V., Arzamasov K.M., Pestrenin L.D., Shulkin I.M. A New Model of Organizing Mass Screening Based on Stand-Alone Artificial Intelligence Used for Fluorography Image Triage. *Public Health and Life Environment – PH&LE*. 2023;31(11):23–32. (In Russ.) <https://doi.org/10.35627/2219-5238/2023-31-11-23-32>
23. Mijwil M.M., Aggarwal K., Sonia S., et al. Has the Future Started? The Current Growth of Artificial Intelligence, Machine Learning, and Deep Learning. *Iraqi Journal for Computer Science and Mathematics*. 2022;3(1). <https://doi.org/10.52866/ijcsm.2022.01.01.013>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Зинченко Виктория Валерьевна, начальник сектора клинических и технических испытаний Научно-практического клинического центра диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, Российская Федерация.
e-mail: ZinchenkoVV1@zdrav.mos.ru
ORCID: [0000-0002-2307-725X](https://orcid.org/0000-0002-2307-725X)

Victoria V. Zinchenko, Head of the Clinical and Technical Trials Branch of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, the Russian Federation.

Ерижоков Рустам Арсеньевич, врач-хирург Первого Московского государственного медицинского университета имени И.М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет), Москва, научный сотрудник Московского физико-технического института, Московская область, Долгопрудный, руководитель отдела стандартизации и контроля качества Научно-практического клинического центра диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, Российская Федерация.

e-mail: npcmr@zdrav.mos.ru
ORCID: [0009-0007-3636-2889](https://orcid.org/0009-0007-3636-2889)

Арзамасов Кирилл Михайлович, кандидат медицинских наук, руководитель отдела стандартизации и контроля качества Научно-практического клинического центра диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, доцент МИРЭА – Российского технологического университета, Москва, Российская Федерация.

e-mail: npcmr@zdrav.mos.ru
ORCID: [0000-0001-7786-0349](https://orcid.org/0000-0001-7786-0349)

Rustam A. Erizhokov, Surgeon Federal State Autonomous Educational Institution of Higher Education I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, research fellow MIPT – Moscow Institute of Physics and Technology, Moscow Region, Dolgoprudny, Head of Standardization and Quality Control of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, the Russian Federation.

Kirill M. Arzamasov, Candidate of Medical Sciences, Head of the Department of Medical Informatics of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Associate Professor MIREA – Russian Technological University, Moscow, the Russian Federation.

Статья поступила в редакцию 14.02.2025; одобрена после рецензирования 25.02.2025; принята к публикации 03.03.2025.

The article was submitted 14.02.2025; approved after reviewing 25.02.2025; accepted for publication 03.03.2025.