Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

Методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики

Ю.А. Васильев, А.В. Владзимирский, О.В. Омелянская, К.М. Арзамасов, С.Ф. Четвериков, Д.А. Румянцев, М.А. Зеленова
Научно-практический клинический центр диагностики и телемедицинских технологий, Москва, Российская Федерация

АННОТАЦИЯ

Обоснование. Мировая сумма инвестиций в компании по разработке программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики составила 80 млн долларов в 2016 году, 152 млн долларов — в 2017 и, ожидаемо, продолжает расти. Активная деятельность компаний-производителей программного обеспечения должна соответствовать существующим клиническим, биоэтическим, правовым и методологическим основам и стандартам. Как на национальном, так и на международном уровне не существует единых стандартов и протоколов проведения испытаний и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики.

Цель — разработать универсальную методологию тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики, направленную на повышение его качества и внедрение в практическое здравоохранение.

Материалы и методы. В ходе аналитического этапа был проведён обзор литературы по базам данных PubMed и eLIBRARY. Практический этап включал апробацию разработанной методологии в рамках Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы.

Результаты. Разработана методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики, направленная на повышение качества данного программного обеспечения и его внедрение в практическое здравоохранение. Методология состоит из 7 этапов: самотестирование, функциональное тестирование, калибровочное тестирование, технологический мониторинг, клинический мониторинг, обратная связь и доработка.

Заключение. Отличительными особенностями методологии являются цикличность этапов тестирования, мониторинга и доработки программного обеспечения, приводящая к постоянному повышению его качества, наличие подробных требований к результатам его работы, участие врачей в его оценке. Методология позволит разработчикам программного обеспечения достичь высоких результатов и продемонстрировать достижения в различных направлениях, а пользователям — сделать осознанный и уверенный выбор среди программ, прошедших независимую и всестороннюю проверку качества.

Ключевые слова: программное обеспечение; искусственный интеллект; рентгенология; диагностическая визуализация; методология; контроль качества.

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

Как цитировать:
Васильев Ю.А., Владзимирский А.В., Омелянская О.В., Арзамасов К.М., Четвериков С.Ф., Румянцев Д.А., Зеленова М.А. Методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики // Digital Diagnostics. 2023. Т. 4, № 3. С. XX–XX. DOI: https://doi.org/10.17816/DD321971

Рукопись получена: 06.04.2023 **Рукопись одобрена:** 15.06.2023 **Опубликована:** 30.08.2023

Methodology for testing and monitoring AI-based software for medical diagnostics

Yuriy A. Vasilev, Anton V. Vladzymyrskyy, Olga V. Omelyanskaya, Kirill M. Arzamasov, Sergey F. Chetverikov, Denis A. Rumyantsev, Maria A. Zelenova
Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

ABSTRACT

BACKGROUND: The global amount of investment in companies developing software based on artificial intelligence technologies for medical diagnostics was $80 million in 2016, $152 million in 2017 and is expected to continue to grow. Activity of software manufacturing companies should comply with existing clinical, bioethical, legal and methodological frameworks and standards. Both at the national and international levels, there are no uniform standards and protocols for testing and monitoring artificial intelligence based software.

AIM: to develop a universal methodology for testing and monitoring artificial intelligence based software for medical diagnostics, aimed at improving its quality and implementing into practical healthcare.

MATERIALS AND METHODS: During the analytical phase, a literature review was conducted on the PubMed and eLibrary databases. The practical stage included approbation of the developed methodology within the framework of the an Experiment on the use of innovative technologies in the field of computer vision for the analysis of medical images and further application in the health care system of the city of Moscow.

RESULTS: A methodology for testing and monitoring artificial intelligence based software for medical diagnostics has been developed, aimed at improving its quality and introducing it into practical healthcare. The methodology consists of 7 stages: self-testing, functional testing, calibration testing, technological monitoring, clinical monitoring, feedback and refinement.

CONCLUSION: Distinctive features of the methodology are: the cyclical stages of testing, monitoring and software development, leading to continuous improvement of its quality, the presence of detailed requirements for the results of the software work, the participation of doctors in software evaluation. The methodology will allow both software developers to achieve high results and demonstrate achievements in various areas, and users to make an informed and confident choice among software that has passed an independent and comprehensive quality check.

Keywords: software; artificial intelligence; radiology; diagnostic imaging; methodology; quality control.

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

To cite this article:
Vasilev YuA, Vladzymyrskyy AV, Omelyanskaya OV, Arzamasov KM, Chetverikov SF, Rumyantsev DA, Zelenova MA. Methodology for testing and monitoring AI-based software for medical diagnostics. *Digital Diagnostics.* 2023;4(3):XX–XX. DOI: https://doi.org/10.17816/DD321971

Received: 06.04.2023 **Accepted:** 15.06.2023 **Published:** 30.08.2023

ОБОСНОВАНИЕ

Мировая сумма инвестиций в компании по разработке программного обеспечения (ПО) на основе технологий искусственного интеллекта (ТИИ) для медицинской диагностики составила 80 млн долларов в 2016 году, 152 млн долларов — в 2017 и, ожидаемо, продолжает расти [1]. В России ПО на основе ТИИ в медицинской диагностике получило широкое распространение в 2019 году, когда Правительство Москвы приняло решение о проведении масштабного научного исследования, которое продолжается до сих пор (2023 год) — Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы (далее — Эксперимент)[1].

Активная деятельность компаний-производителей ПО должна соответствовать существующим клиническим, биоэтическим, правовым и методологическим основам и стандартам [1]. В соответствии с российским законодательством, для того чтобы ПО на основе ТИИ начало использоваться в практическом здравоохранении, а не в условиях эксперимента, оно должно пройти государственную регистрацию как медицинское изделие в установленном законом порядке, результатом чего будет получение ПО регистрационного удостоверения (РУ) Росздравнадзора[2].

Обязательным подготовительным этапом перед направлением ПО на регистрацию является проведение технических и клинических испытаний с целью проверки заявленного функционала ПО[3]. При этом как на национальном, так и на международном уровне единых стандартов и протоколов проведения испытаний не существует, что обусловлено спецификой ПО на основе ТИИ — отсутствием понятной пользователям информации о порядке работы и принципах принятия им решений [2]. Управление по контролю качества пищевых продуктов и лекарственных средств Соединённых Штатов Америки (Food and Drug Administration, FDA) также находится в процессе разработки точных рекомендаций по оценке и регулированию ПО на основе ТИИ [1]. Отсутствие возможности достоверного подтверждения выполнения ПО предъявляемых к нему требований приводит к крайне негативным последствиям: недоверию пользователей к ПО, торможению его внедрения в практическое здравоохранение, упущенному положительному социально-экономическому эффекту от применения ПО, торможению развития здравоохранения в целом [3].

---

[1] Постановление Правительства Москвы от 21.11.2019 № 1543-ПП «О проведении эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы». Режим доступа: https://docs.cntd.ru/document/563879961.

[2] Постановление Правительства РФ от 24.11.2020 № 1906 «О внесении изменений в Правила государственной регистрации медицинских изделий». Режим доступа: http://publication.pravo.gov.ru/Document/View/0001202011270010.

[3] Федеральный закон от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации. Ст. 38. Медицинские изделия. Режим доступа: https://www.consultant.ru/document/cons_doc_LAW_121895/ddcfddbdbb49e64f085b65473218611b4bb6cd65/.

После получения ПО РУ выполняется его пострегистрационный клинический мониторинг, нацеленный на обеспечение безопасности его применения в практическом здравоохранении[4]. Однако существующие требования по проведению мониторинга являются общими для разных медицинских изделий и также не учитывают специфику ПО на основе ТИИ для медицинской диагностики [4]. Согласно решению Коллегии Евразийской экономической комиссии, мониторинг медицинских изделий третьего класса риска, к котором относится ПО на основе ТИИ, выполняется ежегодно в течение трёх лет после получения РУ[5]. Однако высокая вариабельность медицинских данных и трудность прогнозирования изменений окружающих условий, например, эпидемиологической обстановки, обусловливают необходимость более частого проведения мониторинга [5]. В ходе мониторинга возможно выявление критических замечаний к результатам работы ПО, что будет требовать его доработки, а после доработки ПО будут необходимы повторное тестирование и мониторинг.

ПО на основе ТИИ для медицинской диагностики может изучаться посредством клинического исследования, наиболее подходящим видом которого является ретроспективное когортное исследование [1]. Данный метод оценки ПО также имеет множество недостатков, главным из которых является отличие полученных результатов работы ПО на практике и в исследовании [1]. Распространённым примером несовершенства методик внедрения ПО на основе ТИИ для медицинской диагностики в практическое здравоохранение является негативный опыт внедрения первой системы компьютерной диагностики (computer-aided diagnostic, CAD) для скрининговой маммографии. Результаты масштабных многоцентровых исследований показали повышение выявляемости рака молочной железы на 2–10% благодаря использованию данного ПО [6]. В 1998 году ПО было одобрено регулятором FDA для начала использования в практическом здравоохранении. Однако в условиях практического здравоохранения данное ПО не достигло положительных результатов и даже привело к снижению выявляемости заболевания и повышению уровня ложноположительных результатов интерпретации маммографических исследований [6]. Одно из объяснений, предлагаемых в литературе, заключается в том, что рентгенологи с разным уровнем опыта работы использовали новую технологию по-разному. Более опытные врачи не обращали на неё внимания, а менее опытные — совершали ошибки из-за возникающего за счёт неё ложного чувства безопасности. Второе объяснение заключается в том, что ПО оказалось неэффективным в выявлении определённых форм рака, что не было выявлено в ранее выполненных исследованиях [1].

Таким образом, несмотря на то, что наиболее характерными для сферы ПО на основе ТИИ являются проблемы этического и правового характера, также существует немаловажная методологическая проблема, которая может быть сформулирована как проблема отсутствия универсальной и комплексной методологии тестирования и мониторинга ПО на основе ТИИ для медицинской диагностики, направленной на повышение его качества и дальнейшее внедрение в практическое здравоохранение [7]. На основании всего вышесказанного представляется актуальным создание такой методологии. Стоит отметить, что методология не будет являться заменой существующей методики оценки безопасности и эффективности ПО, утверждённой законодательно, а будет существовать независимо и являться дополнением, позволяющим повысить шансы ПО успешно получить РУ Росздравнадзора, а после его

[4] Приказ Минздрава России от 15.09.2020 № 980н «Об утверждении Порядка осуществления мониторинга безопасности медицинских изделий». Режим доступа: https://docs.cntd.ru/document/566006416.

[5] Решение Коллегии Евразийской экономической комиссии от 22.12.2015 № 174 «Об утверждении Правил проведения мониторинга безопасности, качества и эффективности медицинских изделий». Режим доступа: https://www.alta.ru/tamdoc/15kr0174/.

получения — выполнять дальнейшую оценку и улучшение ПО с целью эффективного внедрения в практическое здравоохранение.

Цель исследования — разработать универсальную методологию тестирования и мониторинга ПО на основе ТИИ для медицинской диагностики, направленную на повышение его качества и внедрение в практическое здравоохранение.

МАТЕРИАЛЫ И МЕТОДЫ

Дизайн исследования

Аналитическое исследование, включающее анализ литературы и собственного опыта, на основании которого была разработана представленная методология.

Разработка методологии

Разработка методологии состояла из двух этапов — аналитического и практического.

В ходе аналитического этапа с целью изучения существующих методологий выполнен обзор литературы, опубликованной в период с 2018 по 2023 год (последние 5 лет) в научных библиотеках PubMed и eLIBRARY, по запросам «methodology for evaluation artificial intelligence in radiology» и «методология оценки искусственного интеллекта в рентгенологии». Работы включались в анализ после оценки их релевантности путём прочтения названия и абстракта. Всего было проанализировано 22 работы [1–22] и 5 нормативно-правовых актов[6].

Далее методология была апробирована рамках Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы, проводимого на базе Единой радиологической информационной сети (ЕРИС) Единой медицинской информационно-аналитической системы (ЕМИАС) города Москвы. Некоторые результаты апробации методологии приведены в данной статье в качестве иллюстрации.

Статистическое обоснование объёма выборок

Определено следующее количество исследований в выборке на разных этапах оценки.

1. На этапе самотестирования объём набора данных (НД) не регламентирован и варьирует в зависимости от клинической задачи, решаемой ПО[7]. НД, используемые на этапах самотестирования, функционального и калибровочного тестирований, подготовлены с учётом данных экспертного консенсуса, в отдельных случаях — с учётом гистологических заключений (например, для

[6] Постановление Правительства Москвы РФ от 21.11.2019 № 1543-ПП (https://docs.cntd.ru/document/563879961); Постановление Правительства РФ от 24.11.2020 № 1906 (http://publication.pravo.gov.ru/Document/View/0001202011270010); Ст. 38 ФЗ от 21.11.2011 № 323-ФЗ (https://www.consultant.ru/document/cons_doc_LAW_121895/); Приказ Минздрава России от 15.09.2020 № 980н (https://docs.cntd.ru/document/566006416); Приказ Департамента здравоохранения г. Москвы от 16.02.2023 № 134 (https://mosmed.ai/documents/227/приказ_ДЗМ__134_от_16.02.2023.pdf).

[7] Центр диагностики и телемедицины. Официальный сайт. Наборы данных. Режим доступа: https://mosmed.ai/datasets/.

оценки злокачественных новообразований). Подробно процесс подготовки НД описан в регламенте его подготовки [19].

2. На этапе функционального тестирования — НД из 5 исследований (на основании ГОСТ Р 8.736-2011 под многократными измерениями понимают не менее четырёх измерений)[8]. Истинным значением считается заключение врача-эксперта. Врач-эксперт — врач, работающий по специальности более 5 лет и описывающий исследования по данному направлению (определённая модальность и целевая патология) на потоке, прошедший инструктаж по работе с ПО на основе ТИИ. В данном этапе участвуют как минимум один технический специалист и один врач-эксперт.
3. На этапе калибровочного тестирования — НД из 100 исследований с балансом классов 50/50 (50% исследований с целевой патологией, 50% — без)[9] [20, 21]. На данном этапе участвуют как минимум один технический специалист и один врач-эксперт.
4. На этапе технологического мониторинга — все исследования, проанализированные ПО за отчётный период для дефектов «а, б» (на основании автоматизации выявления дефектов), и выборка из 80 исследований для дефектов «в–д»[10] [20, 21]. В данном этапе участвует как минимум один технический специалист.
5. На этапе клинического мониторинга — вышеуказанная выборка из 80 исследований, истинным значением считается заключение врача-эксперта[11] [20, 21]. На данном этапе участвует один врач-эксперт.

Этическая экспертиза

Настоящая работа проводилась в рамках ранее одобренного локальным этическим комитетом исследования (№ NCT04489992) «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы» (Московский эксперимент).

РЕЗУЛЬТАТЫ

По результатам анализа литературы были найдены работы, посвящённые отдельным этапам оценки ПО на основе ТИИ для медицинской диагностики: валидации [1, 5, 8, 9], мониторингу [10], а также внедрению [7, 11–13] и нормативному регулированию ПО [14, 15]. При этом не обнаружено единой методологии тестирования и мониторинга ПО на основе ТИИ для медицинской диагностики. Существуют работы, посвящённые жизненному циклу ПО на основе ТИИ [16], однако они посвящены преимущественно ПО, предназначенному не для медицинских, а иных сфер деятельности, и не учитывают особенности ПО на основе ТИИ для медицинской

[8] ГОСТ Р 8.736-2011. Национальный стандарт Российской Федерации. Государственная система обеспечения единства измерений. Измерения прямые многократные. Методы обработки результатов измерений. Основные положения. Режим доступа: https://docs.cntd.ru/document/1200089016.
[9] Приказ Департамента здравоохранения города Москвы от 16.02.2023 № 134 «Об утверждении Порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображения и дальнейшего применения в системе здравоохранения города Москвы». Режим доступа: https://mosmed.ai/documents/227/приказ_ДЗМ__134_от_16.02.2023.pdf.
[10] Там же.
[11] Там же.

диагностики. Кроме того, имеются руководства по выполнению исследований и написанию научных публикаций на тему ПО на основе ТИИ, однако с их помощью невозможно выполнить тестирование и мониторинг ПО [17, 18]. Отдельно стоит подчеркнуть, что также не обнаружено публикаций, посвящённых доработке ПО после его тестирования и мониторинга. При этом именно доработка ПО является необходимой для повышения его качества и успешного внедрения в практическое здравоохранение.

В связи с этим авторами была разработана методология тестирования и мониторинга ПО на основе ТИИ для медицинской диагностики, направленная на повышение его качества и внедрение в практическое здравоохранение. Методология состоит из 7 этапов, представленных на рис. 1. Далее в тексте для каждого этапа описаны его цель, суть и результаты.

Самотестирование

Этап самотестирования предназначен для оценки технической совместимости ПО с входными данными. Разработчикам (или поставщикам) ПО предоставляется доступ к открытому НД, содержащему файлы формата Dicom (Digital Imaging and Communications in Medicine), являющиеся обезличенными примерами диагностических исследований[12]. В наборе данных предусмотрена следующая структура параметров: модальность, тип диагностической процедуры, производитель и модель диагностического устройства [19].

Совместимость ПО с данными позволяет выполнить техническую интеграцию ПО в рентгенологическую информационную сеть медицинского учреждения и приступить к дальнейшей оценке, начинающейся с этапа функционального тестирования[13].

Функциональное тестирование

Функциональное тестирование — этап, в ходе которого осуществляется проверка наличия и работоспособности функций ПО, заявленных компанией-поставщиком. Проверка выполняется с технической и клинической точек зрения. С технической точки зрения ПО оценивается по следующим критериям: приоритизация исследований (триаж); наличие дополнительной серии изображений от ПО; наличие названия дополнительной серии; наличие графического обозначения ПО на изображениях дополнительной серии; наличие предупреждающей надписи «Только для исследовательских целей» на изображениях и в DICOM SR; возможность синхронизации серий; отображение вероятности наличия патологии; указание категории патологии; наличие полной структуры протокола DICOM SR (рис. 2, 3).

Эта часть функционального тестирования выполняется специалистами с техническим образованием в соответствии с базовыми функциональными требованиями, разработанными ГБУЗ города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента

[12] Центр диагностики и телемедицины. Официальный сайт. Наборы данных. Режим доступа: https://mosmed.ai/datasets/.

[13] Приказ Департамента здравоохранения города Москвы от 16.02.2023 № 134 «Об утверждении Порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображения и дальнейшего применения в системе здравоохранения города Москвы». Режим доступа: https://mosmed.ai/documents/227/приказ_ДЗМ__134_от_16.02.2023.pdf.

здравоохранения города Москвы» (НПКЦ ДиТ ДЗМ)[14]. Оценка функций ПО с медицинской точки зрения выполняется врачами-экспертами в соответствии с базовыми диагностическими требованиями, разработанными НПКЦ ДиТ ДЗМ[15]. Базовые диагностические требования включают такие критерии, как обязательное и опциональное содержание ответа ПО, формат и форму представленного ответа. В базовых функциональных и базовых диагностических есть как общие для всех ПО требования, так и специфичные, определённые клинической задачей, решаемой ПО.

При выявлении критических несоответствий тестирование ПО прекращается до момента устранения компанией-поставщиком их причин. Критическими считаются несоответствия базовым функциональным требованиям, поскольку они негативно сказываются на рабочих процессах врача и прямо или косвенно влияют на жизнь и здоровье пациента (рис. 4, 5).

После устранения компанией-поставщиком причин критических несоответствий проводится повторное функциональное тестирование. Претенденту предоставляется возможность повторного прохождения этапа не более 2 раз. Первое повторное прохождение претендент может выполнить спустя неограниченное время после получения протокола с неудовлетворительными результатами тестирования. Второе повторное прохождение осуществляется не ранее чем через 3 месяца после получения последнего протокола с неудовлетворительными результатами тестирования. При неудачном прохождении второго повторного тестирования претенденту может быть предложен альтернативный вариант научно-практического сотрудничества[16]. При отсутствии выявленных критических несоответствий ПО переходит на этап калибровочного тестирования[17].

Калибровочное тестирование

Калибровочное тестирование — этап, в ходе которого определяются показатели диагностической точности ПО. Основным показателем является площадь под ROC-кривой (Area under the ROC Curve, AUC). При анализе ROC-кривой с помощью индекса Юдена, максимизации отрицательной и положительной прогностической ценности определяется оптимальное значение порога активации. Другие определяемые показатели включают чувствительность, специфичность, точность, прогностическую ценность положительного результата, прогностическую ценность отрицательного результата. Определяется также минимальное, среднее и максимальное время анализа одного исследования, в виде четырёхпольной таблицы представляется количество истинно положительных, ложноотрицательных, ложноположительных, истинно отрицательных результатов. Пороговые значения для некоторых показателей следующие: AUC — не менее 0,81 или 0,91 (в зависимости от клинической задачи); время, затрачиваемое на принятие, обработку исследования и передачу результатов анализа, — не более 6,5 минут; удельный вес успешно обработанных исследований — не менее 90%[18] [21].

[14] Базовые функциональные требования к результатам работы ИИ-сервисов. Режим доступа: https://mosmed.ai/documents/218/Базовые функциональные требования 29.11.2022.pdf.
[15] Базовые диагностические требования к результатам работы ИИ-сервисов. Режим доступа: https://mosmed.ai/documents/226/Базовые_диагностические_требования_22_02_2023.pdf.
[16] Приказ Департамента здравоохранения города Москвы от 16.02.2023 № 134 «Об утверждении Порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображения и дальнейшего применения в системе здравоохранения города Москвы». Режим доступа: https://mosmed.ai/documents/227/приказ_ДЗМ__134_от_16.02.2023.pdf.
[17] Там же.
[18] Там же.

Результатом калибровочного тестирования является его протокол (рис. 6), который может содержать критические и некритические несоответствия. Критическими считаются несоответствия вышеуказанным пороговым значениям и существенные отклонения от методических рекомендаций [21]. При их выявлении тестирование ПО прекращается до момента их устранения. В случае их отсутствия ПО получает возможность начать проспективный анализ исследований в рамках этапа периодического мониторинга, включающего технологический и клинический мониторинг[19].

Технологический мониторинг

Технологический мониторинг — этап, который представляет собой периодическую проверку результатов работы ПО с технической точки зрения. Данный этап необходим для оперативного выявления дефектов, оперативного контроля качества и недопущения некорректно функционирующего ПО в практику врачей-рентгенологов. Дефекты, которые могут быть выявлены на данном этапе, разделены на следующие группы:

а) время, затрачиваемое на обработку одного исследования, превышает 6,5 минут;
б) отсутствие результатов проанализированных исследований;
в) некорректная работа заявленного функционала ПО, затрудняющая работу врача-рентгенолога или делающая её выполнение невозможным с надлежащим качеством;
г) дефекты, связанные с отображением области изображений;
д) иные нарушения целостности и содержимого файлов с результатами исследований, обусловливающие ограничение их диагностической интерпретации.

Мониторинг дефектов «а, б» осуществляется автоматически для всех исследований, проанализированных ПО за отчётный период, дефектов «в–д» — полуавтоматически на выборке из 80 исследований. Для корректной оценки дефектов разработана форма внутреннего отчёта проведения мониторинга работы ПО с инструкцией по мониторингу технологических дефектов (рис. 7). На рис. 8 приведена графическая информация о среднем числе технологических дефектов для направления «рентгенография органов грудной клетки», на которой наблюдается тенденция к снижению количества дефектов.

Результатом технологического мониторинга является отчёт по технологическому мониторингу (рис. 9). Если удельный вес выявленных дефектов превышает 10%, то маршрутизация исследований на данное ПО останавливается до момента устранения причин дефектов. Если удельный вес выявленных дефектов не превышает 10%, то функционирование ПО и его периодический мониторинг продолжаются[20].

Клинический мониторинг

В ходе периодического мониторинга выполняется также клиническая оценка результатов работы ПО врачами-рентгенологами. Двумя основными критериями

[19] Там же.
[20] Приказ Департамента здравоохранения города Москвы от 16.02.2023 № 134 «Об утверждении Порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображения и дальнейшего применения в системе здравоохранения города Москвы». Режим доступа: https://mosmed.ai/documents/227/приказ_ДЗМ__134_от_16.02.2023.pdf.

оценки являются трактовка (заключение) и локализация (маркировка) патологии. Вариантами ответа, которые врачи могут выбрать в ходе оценки, являются полное соответствие, некорректная оценка, ложноположительный и ложноотрицательный результаты. Например, формулировка «Трактовка: Полное соответствие» выбирается в случае полного согласия врача с заключением от ПО, а формулировка «Трактовка: Некорректная оценка» выбирается в случае частичного согласия врача с заключением от ПО (например, согласие врача с наличием патологических находок, но несогласие с их детализацией, или, наоборот, согласие с детализацией, но несогласие с общим выводом о вероятности или степени тяжести патологических находок). Формулировки «Трактовка: Ложноположительный» и «Трактовка: Ложноотрицательный» (рис. 10) выбираются в случае полного несогласия врача с заключением от ПО.

Результаты клинической оценки заносятся в вышеуказанную форму внутреннего отчёта по мониторингу, после чего она загружается в программный модуль мониторинга, где в автоматическом режиме формируется итоговый отчёт по мониторингу.

По результатам периодического мониторинга принимается одно из заключений: «Участие ПО в Эксперименте продолжается», «Участнику Эксперимента необходимо внести изменения в работу ПО», «Участие ПО в Эксперименте приостанавливается до внесения изменений в работу ПО»[21].

Обратная связь

Этап обратной связи от врачей-рентгенологов необходим для оценки практической значимости ПО. Форма для обратной связи находится в окне программы на автоматизированном рабочем месте врача-рентгенолога (рис. 11). Последний может согласиться или не согласиться с заключением ПО, в случае несогласия — выбрать причину. Основными причинами являются технологический дефект и диагностическая неточность. Достаточной является обратная связь от врачей по 5% всех проанализированных ПО исследований. Кроме того, обратная связь собирается с помощью анкетирования врачей, что позволяет оценить их удовлетворенность работой ПО[22].

Доработка

При выявлении критического замечания к работе ПО на этапах функционального, калибровочного тестирования и периодического мониторинга тестирование ПО прекращается до момента устранения причин замечания. Доработка происходит на стороне компании-поставщика и является «чёрным ящиком» для медицинской организации. В случае необходимости доработок, не несущих в себе изменений первично заявленных функций, технической архитектуры и не затрагивающих изменений метрик диагностической точности ПО, претендент после внесения доработок может сразу перейти на следующий этап методологии.

В случае осуществления претендентом доработок, несущих в себе изменения первично заявленных функций, технической архитектуры и затрагивающих изменения метрик диагностической точности ПО, проводится повторное функциональное, а затем

[21] Приказ Департамента здравоохранения города Москвы от 16.02.2023 № 134 «Об утверждении Порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображения и дальнейшего применения в системе здравоохранения города Москвы». Режим доступа: https://mosmed.ai/documents/227/приказ_ДЗМ__134_от_16.02.2023.pdf.
[22] Там же.

калибровочное тестирование, независимо от того, на каком этапе методологии ПО находилось ранее[23].

ОБСУЖДЕНИЕ

В данной работе представлена методология тестирования и мониторинга результатов работы ПО на основе ТИИ для медицинской диагностики, направленная на повышение его качества и внедрение в практическое здравоохранение. Необходимость её создания была обусловлена, во-первых, отсутствием конкретных требований к тестированию и мониторингу ПО на основе ТИИ для медицинской диагностики в существующей нормативной документации, во-вторых, отсутствием регламентированных принципов выбора медицинской организацией ПО среди многообразия существующих программ, представленных на рынке. Данная методология не нарушает установленных законодательно требований и при этом учитывает специфику ПО на основе ТИИ для медицинской диагностики. Методология включает 7 уникальных и чётко организованных этапов, обоснованных результатами научных исследований [1–4, 19–21] и подкреплённых законодательными документами[24].

Ценной особенностью методологии является наличие разработанных базовых функциональных и базовых диагностических требований, используемых на этапе функционального тестирования[25]. Систематизация дефектов и требований является уникальной (в рассмотренных источниках не приведены их детальные описания). Особенно важным представляется существующее разделение несоответствий на критические и некритические, что удобно как разработчикам ПО, так и пользователям. На мировом уровне известными являются документы Института наук о данных Американской коллегии радиологов, в которых описаны клинические задачи, решаемые с помощью ПО, ожидаемые входные и выходные данные[26].

Другой важной особенностью методологии являются обязательное проведение калибровки ПО на локальных данных (этап калибровочного тестирования) и последующая валидация на потоке реальных данных (этап периодического мониторинга). По данным зарубежного систематического обзора [22], только 6% ПО на основе ТИИ проходили этапы внешней валидации. Валидация может быть разделена на «широкую» и «узкую» [8]. Целью «узкой» валидации является оценка «правильности» продукта, т.е. насколько результаты его использования соответствуют целям его использования. К ней можно отнести клиническую валидацию и оценку удобства использования. «Широкая» валидация включает «узкую» валидацию, а также является синонимом контроля качества, т.е. гарантией того, что ПО было разработано с помощью лучших практик и методов. К ней можно отнести анализ алгоритма, тестирование ПО, исследование документации. В таком случае оценивается внутренняя структура ПО, и оно становится «белым ящиком» [8].

---

[23] Там же.

[24] Постановление Правительства Москвы от 21.11.2019 № 1543-ПП (https://docs.cntd.ru/document/563879961); Приказ Департамента здравоохранения г. Москвы от 16.02.2023 № 134 (https://mosmed.ai/documents/227/приказ_ДЗМ__134_от_16.02.2023.pdf).

[25] Базовые функциональные требования к результатам работы ИИ-сервисов (https://mosmed.ai/documents/218/Базовые_функциональные_требования_29.11.2022.pdf); Базовые диагностические требования к результатам работы ИИ-сервисов (https://mosmed.ai/documents/226/Базовые_диагностические_требования_22_02_2023.pdf).

[26] ACR Data Science Institute Releases Landmark Artificial Intelligence Use Cases. 2018. Режим доступа: https://www.acr.org/Media-Center/ACR-News-Releases/2018/ACR-Data-Science-Institute-Releases-Landmark-Artificial-Intelligence-Use-Cases.

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

Отдельно стоит отметить наличие в методологии этапа доработки ПО после выявления критических несоответствий. Именно доработка ПО приводит к постепенному снижению количества технологических дефектов и повышению показателей диагностической точности ПО. Таким образом, методология позволит разработчикам ПО на основе ТИИ для медицинской диагностики достичь высоких результатов в различных направлениях, а пользователям — сделать осознанный и уверенный выбор среди ПО, прошедших независимую проверку качества, что в конечном счёте приведёт к внедрению ПО в практическое здравоохранение, снижению трудозатрат рентгенолога и повышению эффективности интерпретации диагностических исследований, т.е. достижению первоначальной цели автоматизации процессов с помощью ТИИ.

Данная методология не является заменой процесса регистрации медицинских изделий в рамках установленных процедур. При этом вся методология или отдельные её этапы и подходы могут быть реализованы для оценки безопасности и эффективности ПО на основе ТИИ со стороны регулирующих органов, а также быть частью системы менеджмента качества производителей. Методология может использоваться как разработчиками ПО, например, в процессе составления плана пострегистрационного клинического мониторинга, который должен быть представлен в комплекте документов при регистрации медицинских изделий, так и медицинскими организациями с целью выбора наиболее подходящего для конкретных условий и целей ПО [4]. Методология может применяться неопределённо долго, удовлетворяя требованиям как Евразийской экономической комиссии по проведению мониторинга в течение 3 лет, так и рекомендациям FDA по проведению мониторинга на протяжении всего периода эксплуатации продукта.

Наличие у ПО на основе ТИИ регистрационного удостоверения не является обоснованием того, что проходить все или отдельные этапы представленной методологии не надо, как минимум по двум причинам. Во-первых, регистрационное удостоверение могло быть получено при испытании на определённом диагностическом оборудовании, и при работе ПО на другом оборудовании результаты работы ПО могут измениться. Во-вторых, регистрационное удостоверение могло быть получено для решения ПО определённой клинической задачи, а в дальнейшем разработчики могут дополнить функции ПО.

Приведённые в данной работе примеры относятся к области рентгенологии, однако методология может быть применима для ПО на основе ТИИ, используемого в других областях клинической медицины. В таком случае потребуется корректировка отдельных форм: списка технологических дефектов, клинической оценки и др.

Ограничения исследования

Ограничением методологии является разобщённость компании-производителя и организации, выполняющей оценку. Во многих методологиях разработка ПО и его оценка выполняются одной компанией (методология «от концепции до внедрения») [16]. В случае представленной методологии оценка выполняется сторонней организацией ближе к этапу внедрения. Ошибки, совершённые разработчиком на ранних этапах разработки, тем не менее могут быть выявлены, но исправить их разработчику может быть сложнее.

На этапе периодического мониторинга ПО анализирует большое количество исследований (более 1000). Возможность контроля качества всех исследований отсутствует по причине ограниченности ресурсов, количества врачей-экспертов и их рабочего времени. Несмотря на автоматизированное формирование репрезентативной

псевдослучайной выборки исследований, на этапе периодического мониторинга возможен пропуск ошибок, обусловленный систематической ошибкой выборки.

Перспективы исследования

1. Публикация результатов оценки ПО с помощью представленной методологии (гипотеза — оценка ПО в соответствии с представленной методологией приводит к повышению показателей диагностической точности и практической значимости ПО на основе ТИИ в медицинской диагностике).
2. Сравнение ПО, получивших и не получивших РУ Росздравнадзора, с помощью представленной методологии.
3. Создание в методологии этапа тестирования, целью которого будет оценка результатов обработки ПО «неудовлетворительных» исследований (исследования с неподходящей для данного ПО анатомической областью, модальностью, артефактами, неправильной укладкой пациента, наличием имплантатов и другого медицинского оборудования).

ЗАКЛЮЧЕНИЕ

Разработана методология тестирования и мониторинга ПО на основе ТИИ для медицинской диагностики, направленная на повышение его качества и внедрение в практическое здравоохранение. Методология состоит из 7 этапов: самотестирование, функциональное тестирование, калибровочное тестирование, технологический мониторинг, клинический мониторинг, обратная связь и доработка. Отличительными особенностями методологии являются цикличность этапов тестирования, мониторинга и доработки ПО, приводящая к постоянному повышению качества ПО; наличие подробных требований к результатам работы ПО; участие врачей в оценке ПО. Методология позволит разработчикам ПО достичь высоких результатов и продемонстрировать достижения в различных направлениях, пользователям — сделать осознанный и уверенный выбор среди ПО, прошедших независимую и всестороннюю проверку качества.

ДОПОЛНИТЕЛЬНО

Источник финансирования. Данная статья подготовлена авторским коллективом в рамках работы № ЕГИСУ: «Разработка платформы повышения качества ИИ-Сервисов для медицинской диагностики», № 123031400006-0, в соответствии с Приказом Департамента здравоохранения города Москвы от 21.12.2022 № 1196 «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».

Конфликт интересов. Авторы заявляют об отсутствии явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

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: Ю.А. Васильев — разработка концепции, утверждение итогового варианта рукописи; А.В. Владзимирский — разработка концепции, утверждение итогового варианта рукописи; О.В. Омелянская — разработка методологии,

утверждение итогового варианта рукописи; К.М. Арзамасов — разработка концепции, проведение исследования, редактирование и утверждение итогового варианта текста рукописи; С.Ф. Четвериков — разработка методологии, проведение исследования; Д.А. Румянцев — анализ литературных данных, написание и редактирование текста статьи; М.А. Зеленова — редактирование текста статьи.

ADDITIONAL INFORMATION

Funding source. This article was prepared by a group of authors as a part of the research and development effort titled "Development of a platform for improving the quality of AI services for clinical diagnostics", No. 123031400006-0 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.

Competing interests. The authors declare that they have no competing interests.

Authors' contribution. All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. Yu.A. Vasiliev — development of the concept, approval of the final version of the manuscript; A.V. Vladzimirsky — development of the concept, approval of the final version of the manuscript; O.V. Omelyanskaya — development of methodology, approval of the final version of the manuscript; K.M. Arzamasov — concept development, research, editing and approval of the final version of the manuscript; S.F. Chetverikov — development of methodology, research; D.A. Rumyantsev — literature review, writing and editing the text of the article; M.A. Zelenova — editing the text of the article.

СПИСОК ЛИТЕРАТУРЫ

1. Oakden-Rayner L., Palme L.J. Artificial intelligence in medicine: Validation and study design. In: Ranschart E., Morozov S., Algra P., eds. Artificial intelligence in medical imaging. Cham: Springer, 2019. P. 83–104.
2. Морозов С.П., Зинченко В.В., Хоружая А.Н., и др. Стандартизация искусственного интеллекта в здравоохранении: Россия выходит в лидеры // Врач и информационные технологии. 2021. № 2. С. 12–19. doi: 10.25881/18110193_2021_2_12
3. Мелдо А.А., Уткин Л.В., Трофимова Т.Н. Искусственный интеллект в медицине: современное состояние и основные направления развития интеллектуальной диагностики // Лучевая диагностика и терапия. 2020. № 1. С. 9–17. doi: 10.22328/2079-5343-2020-11-1-9-17
4. Зинченко В.В., Арзамасов К.М., Четвериков С.Ф., и др. Методология проведения пострегистрационного клинического мониторинга для программного обеспечения с применением технологий искусственного интеллекта // Современные технологии в медицине. 2022. Т. 14, № 5. С. 15–25. doi: 10.17691/stm2022.14.5.02
5. Tanguay W., Acar P., Fine B., et al. Assessment of radiology artificial intelligence software: A validation and evaluation framework // Can Assoc Radiol J. 2023. Vol. 74, N 2. P. 326–333. doi: 10.1177/08465371221135760
6. Kohli A., Jha S. Why CAD failed in mammography // J Am Coll Radiol. 2018. Vol. 15, N 3, Pt. B. P. 535–537. doi: 10.1016/j.jacr.2017.12.029
7. Recht M.P., Dewey M., Dreyer K., et al. Integrating artificial intelligence into the clinical practice of radiology: Challenges and recommendations // Eur Radiol. 2020. Vol. 30, N 6. P. 3576–3584. doi: 10.1007/s00330-020-06672-5

8. Higgins D.C., Johner C. Validation of artificial intelligence containing products across the regulated healthcare industries // Ther Innov Regul Sci. 2023. Vol. 57, N 4. P. 797–809. doi: 10.1007/s43441-023-00530-4
9. Rudolph J., Schachtner B., Fink N., et al. Clinically focused multi-cohort benchmarking as a tool for external validation of artificial intelligence algorithm performance in basic chest radiography analysis // Sci Rep. 2022. Vol. 12, N 1. P. 12764. doi: 10.1038/s41598-022-16514-7
10. 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 // J Am Coll Radiol. 2021. Vol. 18, N 11. P. 1489–1496. doi: 10.1016/j.jacr.2021.08.022
11. Strohm L., Hehakaya C., Ranschaert E.R., et al. Implementation of artificial intelligence (AI) applications in radiology: Hindering and facilitating factors // Eur Radiol. 2020. Vol. 30, N 10. P. 5525–5532. doi: 10.1007/s00330-020-06946-y
12. Sohn J.H., Chillakuru Y.R., Lee S., et al. An open-source, vender agnostic hardware and software pipeline for integration of artificial intelligence in radiology workflow // J Digit Imaging. 2020. Vol. 33, N 4. P. 1041–1046. doi: 10.1007/s10278-020-00348-8
13. Wichmann J.L., Willemink M.J., De Cecco C.N. Artificial intelligence and machine learning in radiology: Current state and considerations for routine clinical implementation // Invest Radiol. 2020. Vol. 55, N 9. P. 619–627. doi: 10.1097/RLI.0000000000000673
14. Larson D.B., Harvey H., Rubin D.L., et al. Regulatory frameworks for development and evaluation of artificial intelligence-based diagnostic imaging algorithms: Summary and recommendations // J Am Coll Radiol. 2021. Vol. 18, N 3, Pt. A. P. 413–424. doi: 10.1016/j.jacr.2020.09.060
15. Milam M.E., Koo C.W. The current status and future of FDA-approved artificial intelligence tools in chest radiology in the United States // Clin Radiol. 2023. Vol. 78, N 2. P. 115–122. doi: 10.1016/j.crad.2022.08.135
16. De Silva D., Alahakoon D. An artificial intelligence life cycle: From conception to production // Patterns (NY). 2022. Vol. 3, N 6. P. 100489. doi: 10.1016/j.patter.2022.100489
17. Cerdá-Alberich L., Solana J., Mallol P., et al. MAIC-10 brief quality checklist for publications using artificial intelligence and medical images // Insights Imaging. 2023. Vol. 14, N 1. P. 11. doi: 10.1186/s13244-022-01355-9
18. Vasey B., Novak A., Ather S., et al. DECIDE-AI: A new reporting guideline and its relevance to artificial intelligence studies in radiology // Clin Radiol. 2023. Vol. 78, N 2. P. 130–136. doi: 10.1016/j.crad.2022.09.131
19. Регламент подготовки наборов данных с описанием подходов к формированию репрезентативной выборки данных. Москва: Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, 2022. 40 с. (Лучшие практики лучевой и инструментальной диагностики; Часть 1).
20. Четвериков С.Ф., Арзамасов К.М., Андрейченко А.Е., и др. Подходы к формированию выборки для контроля качества работы систем искусственного интеллекта в медико-биологических исследованиях // Современные технологии в медицине. 2023. Т. 15, № 2. С. 19–27. doi: 10.17691/stm2023.15.2.02
21. Морозов С.П., Владзимирский А.В., Кляшторный В.Г., и др. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика). Москва: Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, 2019. 33 с.
22. Kim D.W., Jang H.Y., Kim K.W., et al. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images:

Results from recently published papers // Korean J Radiol. 2019. Vol. 20, N 3. P. 405–410. doi: 10.3348/kjr.2019.0025

REFERENCES

1. Oakden-Rayner L, Palme LJ. Artificial intelligence in medicine: Validation and study design. In: Ranschart E, Morozov S, Algra P, eds. Artificial intelligence in medical imaging. Cham: Springer; 2019. P. 83–104.
2. Morozov SP, Zinchenko VV, Khoruzhaya AN, et al. Standardization of artificial intelligence in healthcare: Russia is becoming a leader. *Doctor Inform Technol*. 2021;(2):12–19. (In Russ). doi: 10.25881/18110193_2021_2_12
3. Mello AA, Utkin LV, Trofimova TN. Artificial intelligence in medicine: The current state and main directions of development of intellectual diagnostics. *Radiation Diagnost Therapy*. 2020;(1):9–17. (In Russ). doi: 10.22328/2079-5343-2020-11-1-9-17
4. Zinchenko VV, Arzamasov KM, Chetverikov SF, et al. Methodology of post-registration clinical monitoring for software using artificial intelligence technologies. *Modern Technol Med*. 2022;14(5):15–25. (In Russ). doi: 10.17691/stm2022.14.5.02
5. Tanguay W, Acar P, Fine B, et al. Assessment of radiology artificial intelligence software: A validation and evaluation framework. *Can Assoc Radiol J*. 2023;74(2):326–333. doi: 10.1177/08465371221135760
6. Kohli A, Jha S. Why CAD failed in mammography. *J Am Coll Radiol*. 2018;15(3 Pt B):535–537. doi: 10.1016/j.jacr.2017.12.029
7. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: Challenges and recommendations. *Eur Radiol*. 2020;30(6):3576–3584. doi: 10.1007/s00330-020-06672-5
8. Higgins DC, Johner C. Validation of artificial intelligence containing products across the regulated healthcare industries. *Ther Innov Regul Sci*. 2023;57(4):797–809. doi: 10.1007/s43441-023-00530-4
9. Rudolph J, Schachtner B, Fink N, et al. Clinically focused multi-cohort benchmarking as a tool for external validation of artificial intelligence algorithm performance in basic chest radiography analysis. *Sci Rep*. 2022;12(1):12764. doi: 10.1038/s41598-022-16514-7
10. 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. *J Am Coll Radiol*. 2021;18(11):1489–1496. doi: 10.1016/j.jacr.2021.08.022
11. Strohm L, Hehakaya C, Ranschaert ER, et al. Implementation of artificial intelligence (AI) applications in radiology: Hindering and facilitating factors. *Eur Radiol*. 2020;30(10):5525–5532. doi: 10.1007/s00330-020-06946-y
12. Sohn JH, Chillakuru YR, Lee S, et al. An open-source, vender agnostic hardware and software pipeline for integration of artificial intelligence in radiology workflow. *J Digit Imaging*. 2020;33(4):1041–1046. doi: 10.1007/s10278-020-00348-8
13. Wichmann JL, Willemink MJ, De Cecco CN. Artificial intelligence and machine learning in radiology: Current state and considerations for routine clinical implementation. *Invest Radiol*. 2020;55(9):619–627. doi: 10.1097/RLI.0000000000000673
14. Larson DB, Harvey H, Rubin DL, et al. Regulatory frameworks for development and evaluation of artificial intelligence-based diagnostic imaging algorithms: Summary and recommendations. *J Am Coll Radiol*. 2021;18(3 Pt A):413–424. doi: 10.1016/j.jacr.2020.09.060
15. Milam ME, Koo CW. The current status and future of FDA-approved artificial intelligence tools in chest radiology in the United States. *Clin Radiol*. 2023;78(2):115–122. doi: 10.1016/j.crad.2022.08.135

16. De Silva D, Alahakoon D. An artificial intelligence life cycle: From conception to production. *Patterns (NY)*. 2022;3(6):100489. doi: 10.1016/j.patter.2022.100489
17. Cerdá-Alberich L, Solana J, Mallol P, et al. MAIC-10 brief quality checklist for publications using artificial intelligence and medical images. *Insights Imaging*. 2023;14(1):11. doi: 10.1186/s13244-022-01355-9
18. Vasey B, Novak A, Ather S, et al. DECIDE-AI: A new reporting guideline and its relevance to artificial intelligence studies in radiology. *Clin Radiol*. 2023;78(2):130–136. doi: 10.1016/j.crad.2022.09.131
19. Regulations for the preparation of data sets with a description of approaches to the formation of a representative sample of data. Moscow: Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Department of Health of the City of Moscow; 2022. 40 p. (Best practices in radiological and instrumental diagnostics; Part 1). (In Russ).
20. Chetverikov S, Arzamasov KM, Andreichenko AE, et al. Approaches to sampling for quality control of artificial intelligence systems in biomedical research. *Modern Technol Med*. 2023;15(2):19–27. (In Russ). doi: 10.17691/stm2023.15.2.02
21. Morozov SP, Vladzimirsky AV, Klyashtorny VG, et al. Clinical trials of software based on intelligent technologies (radiation diagnostics). Moscow: Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Department of Health of the City of Moscow; 2019. 33 p. (In Russ).
22. Kim DW, Jang HY, Kim KW, et al. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers. *Korean J Radiol*. 2019;20(3):405–410. doi: 10.3348/kjr.2019.0025

ОБ АВТОРАХ	**AUTHORS' INFO**
*** Румянцев Денис Андреевич;** адрес: Россия, 127051, Москва, ул. Петровка, д. 24, стр. 1; ORCID: 0000-0001-7670-7385; eLibrary SPIN: 734-2085; e-mail: RumyantsevDA3@zdrav.mos.ru	*** Denis A. Rumyantsev;** address: 24/1 Petrovka street, 127051 Moscow, Russia; ORCID:0000-0001-7670-7385; eLibrary SPIN: 734-2085; e-mail: RumyantsevDA3@zdrav.mos.ru
Васильев Юрий Александрович, канд. мед. наук; ORCID: 0000-0002-0208-5218 ; eLibrary SPIN: 4458-5608; e-mail: npcmr@zdrav.mos.ru	**Yuriy A. Vasilev,** MD, Cand. Sci. (Med.); ORCID: 0000-0002-0208-5218; eLibrary SPIN: 4458-5608; e-mail: npcmr@zdrav.mos.ru
Владзимирский Антон Вячеславович, д-р мед. наук; ORCID: 0000-0002-2990-7736 ; eLibrary SPIN: 3602-7120; e-mail: npcmr@zdrav.mos.ru	**Anton V. Vladzymyrskyy,** MD, Dr. Sci. (Med.); ORCID: 0000-0002-2990-7736 ; eLibrary SPIN: 602-7120; e-mail: npcmr@zdrav.mos.ru
Омелянская Ольга Васильевна; ORCID: 0000-0002-0245-4431; eLibrary SPIN: 8948-6152; e-mail: npcmr@zdrav.mos.ru	**Olga V. Omelyanskaya;** ORCID: 0000-0002-0245-4431 ; eLibrary SPIN: 8948-6152; e-mail: npcmr@zdrav.mos.ru
Арзамасов Кирилл Михайлович, канд. мед. наук; ORCID: 0000-0001-7786-0349;	**Kirill M. Arzamasov,** MD, Cand. Sci. (Med.); ORCID: 0000-0001-7786-0349;

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

eLibrary SPIN: 3160-8062; e-mail: ArzamasovKM@zdrav.mos.ru	eLibrary SPIN: 3160-8062; e-mail: ArzamasovKM@zdrav.mos.ru
Четвериков Сергей Федорович, канд. тех. наук; ORCID: 0000-0002-3097-8881; eLibrary SPIN: 3815-8870; e-mail: ChetverikovSF@zdrav.mos.ru	**Sergei F. Chetverikov,** Cand. Sci. (Engin.); ORCID: 0000-0002-3097-8881; eLibrary SPIN: 3815-8870; e-mail: ChetverikovSF@zdrav.mos.ru
Зеленова Мария Александровна; ORCID:0000-0001-7458-5396; eLibrary SPIN: 3823-6872; e-mail: ZelenovaMA@zdrav.mos.ru	**Maria A. Zelenova;** ORCID: 0000-0001-7458-5396; eLibrary SPIN: 3823-6872; e-mail: ZelenovaMA@zdrav.mos.ru
* Автор, ответственный за переписку / Corresponding author	

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

Рис. 1. Методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики.

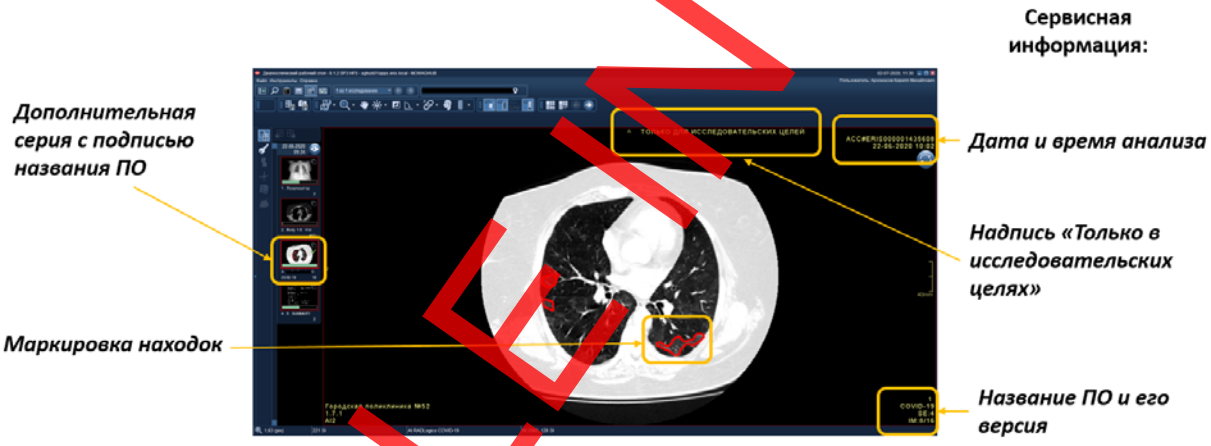


Рис. 2. Основные компоненты результата работы программного обеспечения на основе технологий искусственного интеллекта с изображением: пример эталонной работы.

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

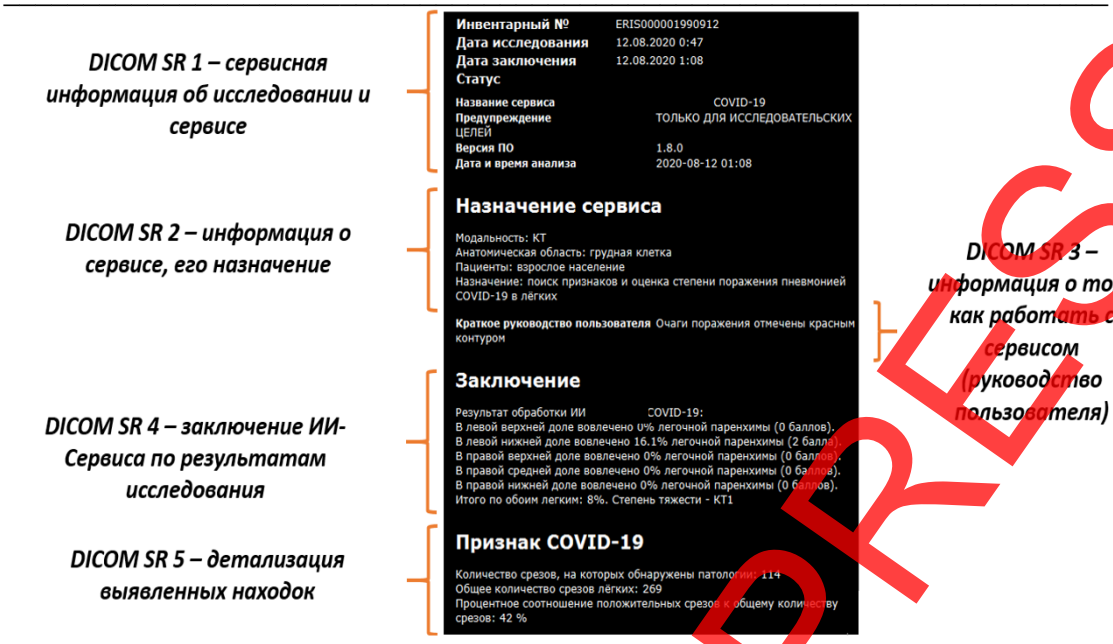


Рис. 3. Основные компоненты результата работы программного обеспечения на основе технологий искусственного интеллекта с DICOM SR: пример эталонной работы.

Рис. 4. Обрезка изображения дополнительной серии программного обеспечения на основе технологий искусственного интеллекта: критическое несоответствие базовым функциональным требованиям.

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

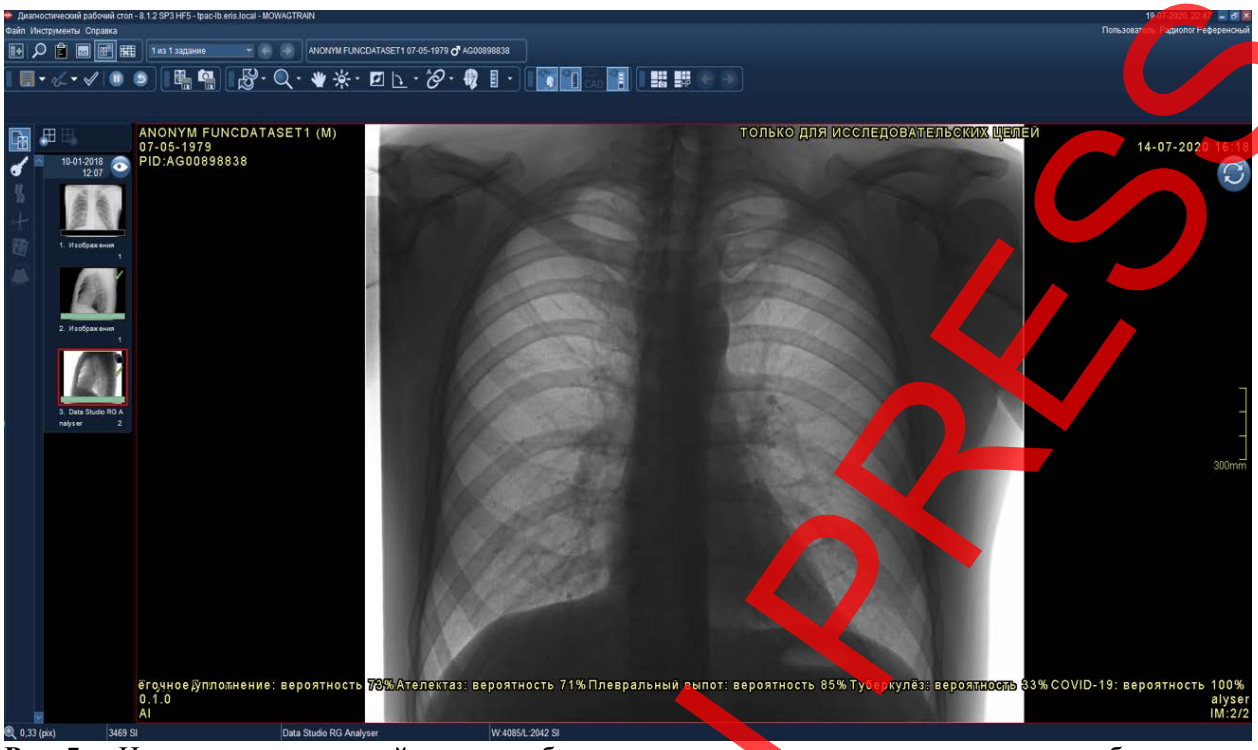


Рис. 5. Наложение подписей на изображение: критическое несоответствие базовым функциональным требованиям.

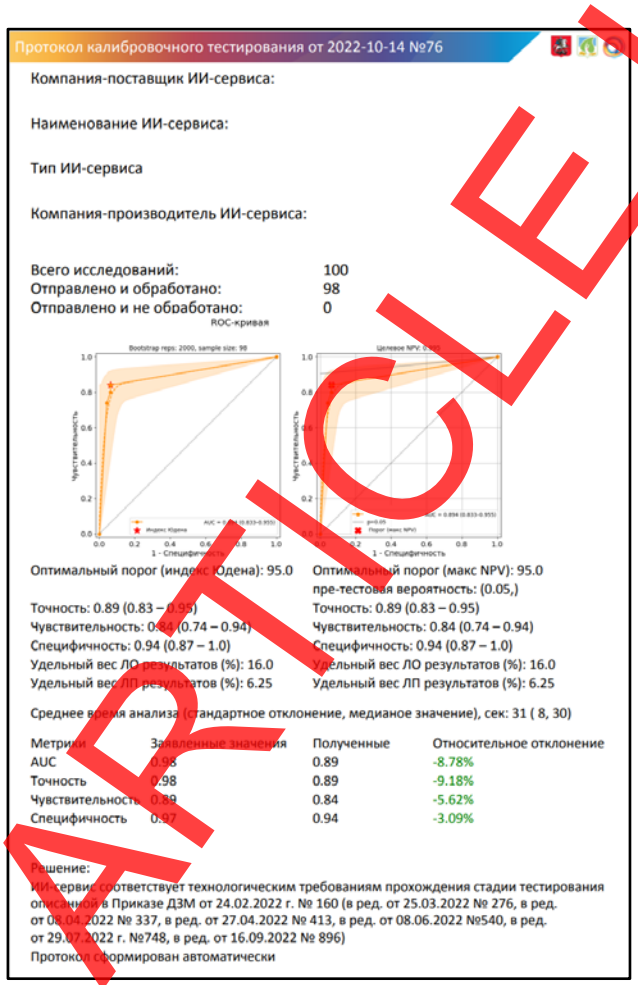


Рис. 6. Пример протокола калибровочного тестирования.

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

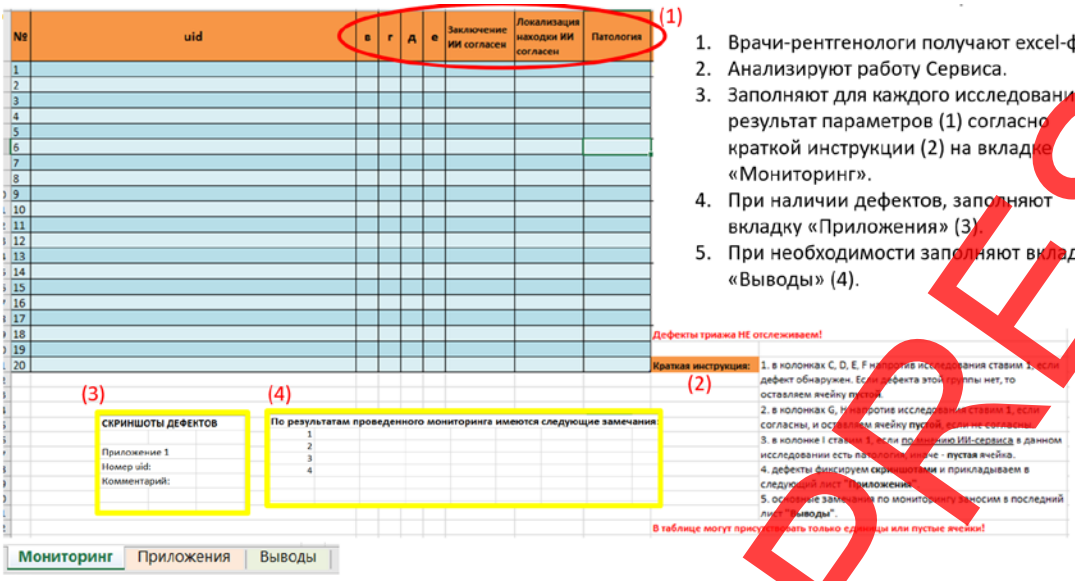
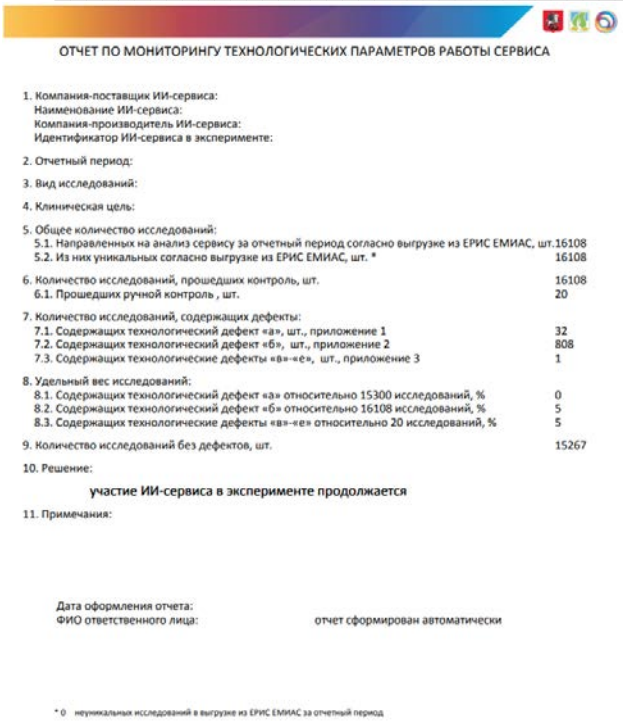


Рис. 7. Форма внутреннего отчёта проведения мониторинга работы программного обеспечения на основе технологий искусственного интеллекта.

Рис. 8. Динамика технологических дефектов для программного обеспечения по модальности «рентгенография органов грудной клетки».

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971



ОТЧЕТ ПО МОНИТОРИНГУ ТЕХНОЛОГИЧЕСКИХ ПАРАМЕТРОВ РАБОТЫ СЕРВИСА

1. Компания-поставщик ИИ-сервиса:
   Наименование ИИ-сервиса:
   Компания-производитель ИИ-сервиса:
   Идентификатор ИИ-сервиса в эксперименте:
2. Отчетный период:
3. Вид исследований:
4. Клиническая цель:
5. Общее количество исследований:
   5.1. Направленных на анализ сервису за отчетный период согласно выгрузке из ЕРИС ЕМИАС, шт. 16108
   5.2. Из них уникальных согласно выгрузке из ЕРИС ЕМИАС, шт. * 16108
6. Количество исследований, прошедших контроль, шт. 16108
   6.1. Прошедших ручной контроль , шт. 20
7. Количество исследований, содержащих дефекты:
   7.1. Содержащих технологический дефект «а», шт., приложение 1 32
   7.2. Содержащих технологический дефект «б», шт., приложение 2 808
   7.3. Содержащих технологические дефекты «в»-«е», шт., приложение 3 1
8. Удельный вес исследований:
   8.1. Содержащих технологический дефект «а» относительно 15300 исследований, % 0
   8.2. Содержащих технологический дефект «б» относительно 16108 исследований, % 5
   8.3. Содержащих технологические дефекты «в»-«е» относительно 20 исследований, % 5
9. Количество исследований без дефектов, шт. 15267
10. Решение:
    участие ИИ-сервиса в эксперименте продолжается
11. Примечания:

Дата оформления отчета:
ФИО ответственного лица: отчет сформирован автоматически

* 0 неуникальных исследований в выгрузке из ЕРИС ЕМИАС за отчетный период

Рис. 9. Пример отчёта по технологическому мониторингу.

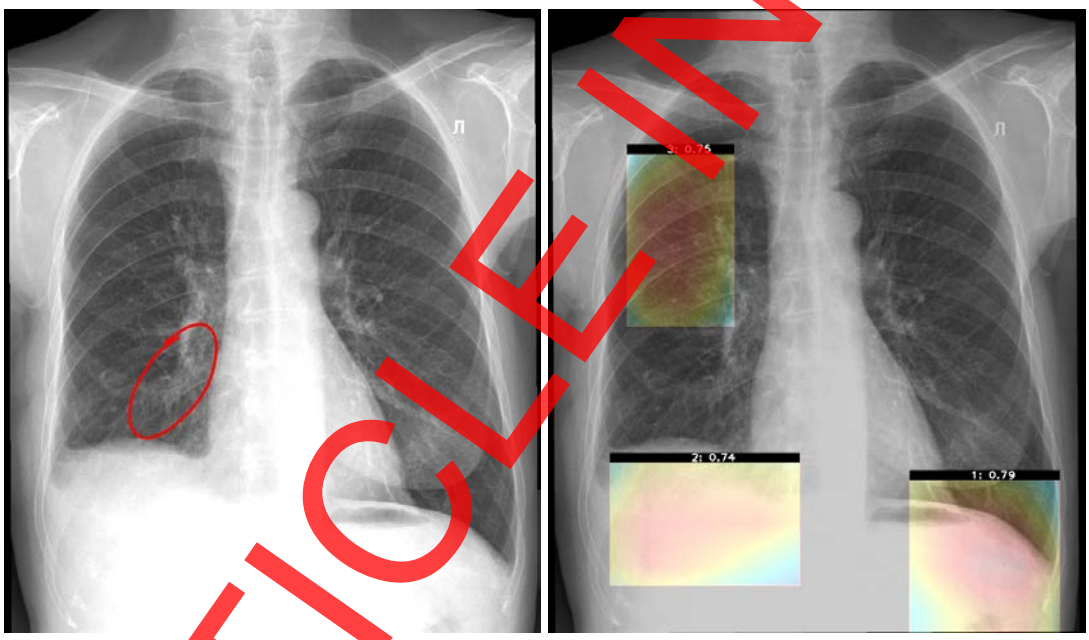


Рис. 10. Ложноотрицательное срабатывание (отсутствие локализации субсегментарного ателектаза нижней доли правого лёгкого): некритическое несоответствие базовым диагностическим требованиям.

Digital Diagnostics. 2023;4(3):XX–XX.
DOI: https://doi.org/10.17816/DD321971

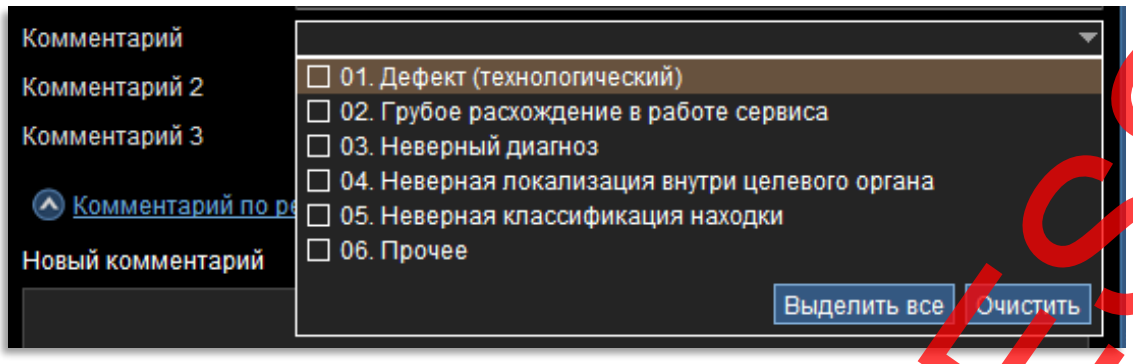


Рис. 11. Содержание окна обратной связи в пользовательском интерфейсе.