

УДК 004.8

DOI: [10.26102/2310-6018/2025.49.2.023](https://doi.org/10.26102/2310-6018/2025.49.2.023)

Платформа для тестирования программного обеспечения с использованием искусственного интеллекта в лучевой диагностике

А.Ю. Ковальчук[✉], А.П. Пономаренко, К.М. Арзамасов

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

Резюме. Количество программного обеспечения с использованием искусственного интеллекта, применяемого в лучевой диагностике, в последние годы стремительно увеличивается, и эффективность таких ИИ-сервисов должна тщательно оцениваться для обеспечения проверки качества разработанных алгоритмов. Ручная оценка таких систем является трудоемким процессом. В связи с этим, актуальной задачей является разработка специализированной единой платформы, предназначенной для автоматизированного тестирования ИИ-алгоритмов, используемых с целью анализа медицинских изображений. Предлагаемая платформа состоит из трех основных модулей: модуль тестирования, обеспечивающий взаимодействие с тестируемым программным обеспечением и сбор результатов обработки данных; модуль просмотра, предоставляющий инструменты для визуальной оценки полученных графических серий и структурированных отчетов; модуль расчета метрик, позволяющий вычислять диагностические характеристики эффективности работы алгоритмов искусственного интеллекта. В ходе разработки применялись такие технологии как Python 3.9, Apache Kafka, PACS и Docker. Разработанная платформа успешно прошла апробацию на реальных данных. Полученные результаты свидетельствуют о перспективности использования разработанной платформы для повышения качества и надежности ИИ-сервисов в лучевой диагностике, а также для облегчения процесса их внедрения в клиническую практику.

Ключевые слова: платформа, лучевая диагностика, тестирование, медицинские изображения, искусственный интеллект.

Благодарности: Авторы выражают благодарность Васильеву Ю.А., Владимирскому А.В., Омелянской О.В., Кирпичеву Ю.С. и Савкиной Е.Ф. за научно-организационную деятельность. Данная статья подготовлена авторским коллективом в рамках НИОКР «Разработка платформы повышения качества ИИ-Сервисов для медицинской диагностики» (№ ЕГИСУ: 123031400006-0) в соответствии с Приказом от 22.12.2023 г. № 1258 «Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счет средств бюджета города Москвы государственным бюджетным (автономным) учреждениям подведомственным Департаменту здравоохранения города Москвы, на 2024 год и плановый период 2025 и 2026 годов» Департамента здравоохранения города Москвы.

Для цитирования: Ковальчук А.Ю., Пономаренко А.П., Арзамасов К.М. Платформа для тестирования программного обеспечения с использованием искусственного интеллекта в лучевой диагностике. *Моделирование, оптимизация и информационные технологии.* 2025;13(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1917> DOI: 10.26102/2310-6018/2025.49.2.023

Platform for testing radiological artificial intelligence-powered software

A.Yu. Kovalchuk[✉], A.P. Ponomarenko, K.M. Arzamasov

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

Abstract. The amount of AI-based software used in radiology has been rapidly increasing in recent years, and the effectiveness of such AI services should be carefully assessed to ensure the quality of the developed algorithms. Manual assessment of such systems is a labor-intensive process. In this regard, an urgent task is to develop a specialized unified platform designed for automated testing of AI algorithms used to analyze medical images. The proposed platform consists of three main modules: a testing module that ensures interaction with the software being tested and collects data processing results; a viewing module that provides tools for visually evaluating the obtained graphic series and structured reports; a metrics calculation module that allows calculating diagnostic characteristics of the effectiveness of artificial intelligence algorithms. During the development, such technologies as Python 3.9, Apache Kafka, PACS and Docker were used. The developed platform has been successfully tested on real data. The obtained results indicate the potential of using the developed platform to improve the quality and reliability of AI services in radiation diagnostics, as well as to facilitate the process of their implementation in clinical practice.

Keywords: platform, diagnostic imaging, testing, medical images, artificial intelligence.

Acknowledgements: The authors express their gratitude to Vasilev Yu.A., Vladzymyrskyy A.V., Omelyanskaya O.V., Kirpichev Yu.S, and Savkina E.F. for their scientific and administrative work. This paper was prepared by a team of authors as a part of the research and development project "Development of a platform for improving quality of AI services for medical diagnostics" (EGISU No.: 123031400006-0) in accordance with Order No. 1258 dated December 22, 2023: "On approval of state assignments funded from the Moscow city budget to state budgetary (autonomous) institutions subordinated to the Moscow Healthcare Department for 2024 and the planning period of 2025 and 2026" of the Moscow Department of Healthcare.

For citation: Kovalchuk A.Yu., Ponomarenko A.P., Arzamasov K.M. Platform for testing radiological artificial intelligence-powered software. *Modeling, Optimization and Information Technology*. 2025;13(2). (In Russ.). URL: <https://moitvvt.ru/ru/journal/pdf?id=1917> DOI: 10.26102/2310-6018/2025.49.2.023

Введение

Лучевая диагностика признана одной из наиболее перспективных областей для внедрения технологий искусственного интеллекта (ИИ) в медицине [1]. Этот интерес обусловлен растущим спросом на высококачественную медицинскую визуализацию, особенно в условиях глобального дефицита квалифицированных врачей-рентгенологов [2]. Внедрение программного обеспечения с использованием искусственного интеллекта (ПО с ИИ) в лучевую диагностику становится ключевым направлением исследований, привлекающим значительное внимание как научного сообщества, так и практикующих врачей [3]. Тем не менее, несмотря на лидирующую роль лучевой диагностики в разработке и внедрении ИИ-решений, сохраняется значительный разрыв между теоретическим потенциалом, описанным в научной литературе, и их применением в клинической практике [4, 5].

Одной из обязательных процедур перед интеграцией ПО с ИИ в медицинские организации является его валидация, с помощью которой происходит оценка соответствия и рассчитываются необходимые метрики для подтверждения достижения

намеченных результатов [6]. Для реализации комплексной оценки ПО с ИИ необходимо уделить внимание клиническим и техническим параметрам. Основные этапы валидации включают отправку набора данных на ПО с ИИ, получение и анализ результатов, а также расчет метрик [7]. Однако перечисленные процессы являются трудоемкими, требуя значительных усилий со стороны квалифицированных специалистов и приводя к увеличению сроков. В настоящее время тестирование в основном проводится ручными методами, путем сравнения выходных данных ПО с ИИ и разметки экспертов или другой ground truth (GT). При этом для расчета метрик формируются таблицы, а методы загрузки данных и подходы к подключению различаются для разных ПО с ИИ, что усложняет выполнение клинической валидации. Это особенно важно при необходимости быстрого принятия решений об эксплуатации ПО с ИИ в рамках мониторинга.

Набор метрик, которые необходимо рассчитать, зависит от назначения конкретного ПО с ИИ. Например, если ПО с ИИ предназначено для определения класса заболевания, то для целостной оценки работы классификатора важно использовать комбинацию нескольких метрик. Наиболее распространенными из них являются чувствительность, специфичность, точность, а также анализ характеристической кривой [8].

В настоящее время предпринимаются попытки автоматизации процесса валидации. Разработаны отдельные инструменты для автоматического расчета метрик [9, 10], однако создание комплексной платформы, объединяющей этапы тестирования ПО на основе ИИ, визуализации результатов и расчета метрик, представляется перспективным направлением для значительного сокращения временных затрат исследователей и врачей. Существующие решения для валидации медицинского ПО на основе ИИ ограничены, одним из наиболее известных примеров является платформа AI-LAB, разработанная Американским Колледжем Радиологии [11]. Тем не менее, AI-LAB предполагает интеграцию в существующую медицинскую информационную систему организации, ограничивая доступность ее инструментов рамками локальной сети.

Таким образом, целью данной работы является разработка единой мультифункциональной платформы для тестирования и расчета метрик медицинского ПО с ИИ в лучевой диагностике.

Материалы и методы

Платформа для тестирования и расчета метрик медицинского ПО на основе ИИ представляет собой модульное решение, разработанное для автоматизации и масштабирования процесса валидации. Платформа состоит из трех ключевых модулей: модуля тестирования, модуля просмотра и модуля расчета метрик. Каждый из этих модулей выполняет отдельные функции в процессе валидации ПО с ИИ, при этом их интеграция обеспечивает комплексное решение задач, связанных с оценкой производительности и эффективности алгоритмов искусственного интеллекта в условиях клинической практики. Платформа разработана с использованием языка программирования Python 3.9.

Модуль тестирования. Модуль тестирования представляет из себя многослойный сервис, в основе которого лежит Flask-сервер, предоставляющий API для получения запросов от пользователя. Авторизация в модуле и распределение ролей реализовано с помощью библиотеки JWT, упрощающей процесс токенизации и идентификации пользователей. Данные пользователей хранятся в отдельном контейнере Postgres.

Для хранения медицинских исследований в формате DICOM в системе используется PACS-сервер. Данная технология реализует взаимодействие по протоколу DICOM, включает в себя набор стандартных команд для передачи файлов и обеспечивает безопасность путем шифрования и изолированности от запросов неизвестных пользователей. PACS-сервер является наиболее распространенным хранилищем данных в контексте машинного обучения в лучевой диагностике. Для взаимодействия с базой данных в платформе используется библиотека `pyetdicom 2.0`.

Важной особенностью модуля и платформы в целом является использование Apache Kafka. Реализованные Kafka broker и Kafka publisher стандартизируют и автоматизируют процесс тестирования ПО с ИИ посредством отправки в топик шаблонного сообщения формата json, содержащего уникальный id, присвоенный ПО с ИИ, для которого данное сообщение было отправлено, а также массив study uid – идентификаторов исследований, подлежащих обработке. Использование Apache Kafka также унифицирует входную точку ПО с ИИ. Унификация достигается путем предоставления шаблонного решения, включающего в себя Kafka consumer, который отправляет запрос на передачу DICOM файлов, после считывания сообщения из топика Kafka Broker и вызывает функции обработки исследований моделью машинного обучения.

Таким образом, данная схема позволяет горизонтально масштабировать модуль, проводя тестирование одновременно нескольких сервисов. Этот результат достигается благодаря независимости тестируемых ПО с ИИ друг от друга и различности потоков, в которых происходит обработка исследований. Также стандартизация входного потока данных предлагает возможность тестирования моделей вне зависимости от модальности и алгоритма обработки исследований.

Модуль просмотра. Основой для реализации модуля просмотра послужил Orthanc osimis web viewer¹ – плагин для базы данных PACS Orthanc. Для отображения подробной информации о структурированных отчетах была создана HTML страница, включающая в себя представление серии из osimis и преобразующая данные из отчета в читаемый вид. Визуализация содержимого просмотрщика на HTML странице осуществляется с помощью языка программирования JavaScript.

Модуль расчета метрик. Для реализации функций расчета метрик и построения ROC-кривой используется библиотека sklearn с встроенными метриками metrics. Визуальный графический интерфейс (GUI) написан с применением библиотеки Python dash. Используемая библиотека ускоряет разработку веб-интерфейса, предоставляя возможность быстрого построения статических html страниц.

Параметры времени, включающие среднее, медианное, минимальное и максимальное время обработки исследований с помощью протестированного ИИ-сервиса сохраняются в отчете. Для формирования отчёта используется библиотека reportlab, включающая в себя функции, специализированные для создания pdf-отчетов.

Апробация. Платформа была апробирована в июле 2024 года с использованием ПО с ИИ для обнаружения кист печени на основе КТ-исследований области брюшной полости.

Результаты

Платформа выполнена в формате веб-приложения, при этом взаимодействие пользователей с платформой осуществляется посредством визуального графического интерфейса. Платформа позволяет проводить валидацию ПО с ИИ для всех видов

¹ Orthanc – DICOM Server. URL: <https://www.orthanc-server.com/> (дата обращения: 15.04.2025).

модальностей медицинских изображений в лучевой диагностике формата DICOM. Архитектура платформы представлена на Рисунке 1.

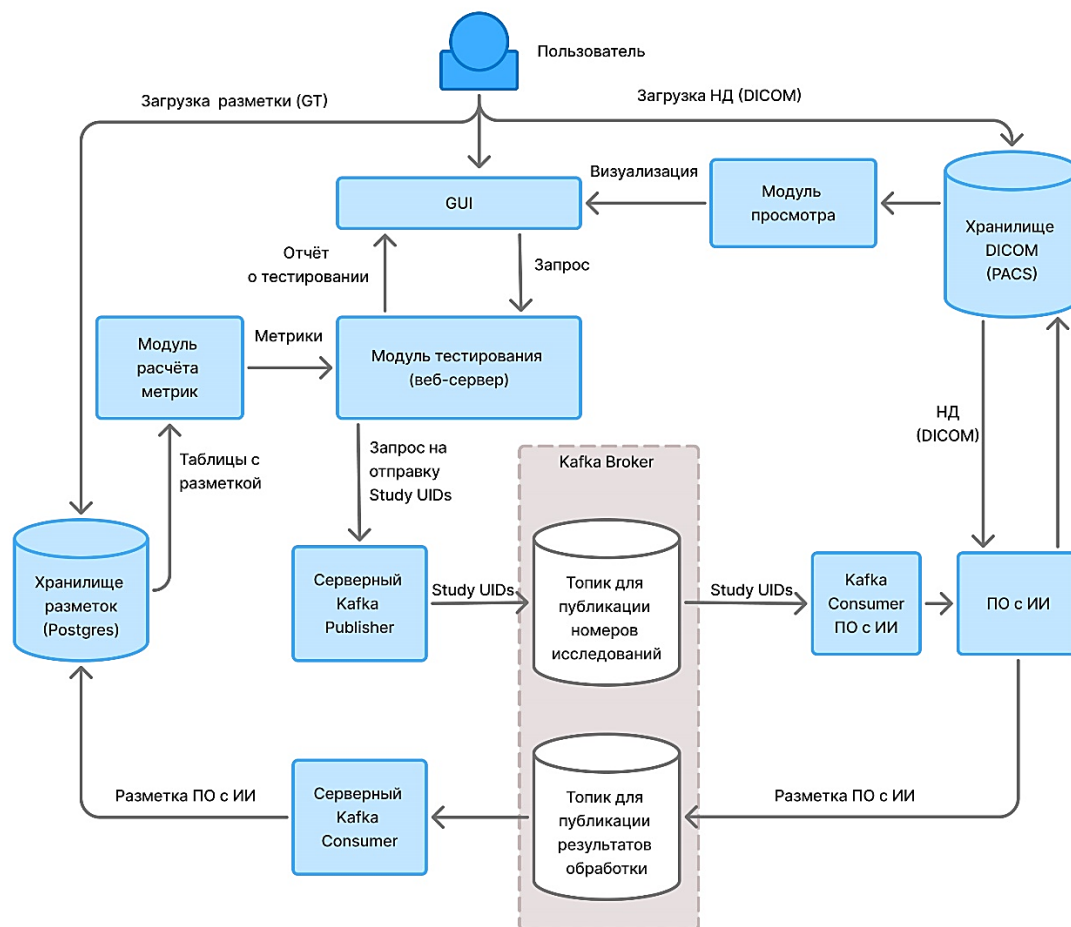


Рисунок 1 – Архитектура платформы
Figure 1 – Platform architecture

Платформа поддерживает одновременное тестирование нескольких ПО с ИИ – каждый процесс тестирования происходит в отдельном канале и не влияет на параллельные потоки.

С целью минимизации риска нарушения конфиденциальности персональных данных для платформы разработана система ролевого разграничения прав, накладывающая ограничения на специализацию сотрудников, имеющих доступ к части функциональных возможностей ПО (Таблица 1).

Таблица 1 – Роли пользователей платформы и соответствующий им функционал
Table 1 – Roles of platform users and their corresponding functionality

Роль		Функционал
Суперпользователь		Управление, техническая поддержка платформы
Руководитель		Назначение ответственных, распределение ролей
Сотрудник	Администратор	Подключение моделей к платформе, загрузка набора данных, выполнение тестирования, контроль процесса тестирования

Таблица 1 (продолжение)
Table 1 (continued)

Аналитик	Оценка отчётов о тестировании
Эксперт	Экспертиза разметки в дополнительных сериях. Оценка клинической корректности данных в структурированном отчёте

Далее подробно рассмотрены результаты разработки и функциональные возможности каждого модуля.

Модуль тестирования. Модуль тестирования является ключевым компонентом платформы, обеспечивающим отправку медицинских изображений на обработку в ПО с ИИ и получение результатов анализа. Перед началом тестирования необходимо определить набор данных, который будет использован для обработки ПО на основе ИИ [12]. Пользователь выбирает этот набор с помощью GUI модуля (Рисунок 2). Интерфейс предоставляет возможность как отбора конкретных исследований, хранящихся на PACS-сервере, так и формирования случайной выборки исследований.

The screenshot shows the 'AI-TEST' interface. At the top, there's a search bar with 'Введите UID' and 'Модальность' fields. Below this is a section titled 'Выборить все исследования' with a list of studies. Each study entry includes a checkbox, a study UID, a series UID, a body part, a study description, a series modality, a series description, and a quantity. The studies listed are: 1. Исследование 1.2.40.0.13.1.6031266466915007344757745322814756283 (LIVER, SR, Dicom SR CT HepatoScan AI 1.0, 1), 2. Исследование 1.2.40.0.13.1.2348329887678229848297423006995729548 (CHEST_TO_PELVIS, CT, Arterial FC01 Vol., 932), 3. Исследование 1.2.40.0.13.1.2348329887678229848297423006995729548 (LIVER, SR, Dicom SR CT HepatoScan AI 1.0, 1), 4. Исследование 1.2.40.0.13.1.2348329887678229848297423006995729548 (CHEST_TO_PELVIS, CT, Dicom S CT HepatoScan AI 1.0, 932), 5. Исследование 1.2.40.0.13.1.2348329887678229848297423006995729548 (CHEST_TO_PELVIS, CT, Dicom S CT HepatoScan AI 1.0, 932). At the bottom, there are buttons for 'Отправка' and 'Просмотр результатов'.

Рисунок 2 – Интерфейс страницы выбора исследований для тестирования
Figure 2 – Interface of the page for selecting studies for testing

После инициации процесса тестирования пользователем модуль тестирования автоматически генерирует сообщение, содержащее информацию о наборе данных, и отправляет его в Kafka Broker (Рисунок 3). Технологии Kafka требуют дополнительной технической квалификации для стандартных сценариев использования, поэтому в модуле предусмотрен функционал облегченного заполнения полей генерируемого сообщения посредством GUI.

x

Введите данные

Содержимое сообщения

Study UID

1.2.40.0.13.1.2348329887678229848297423006995729548, 1.2.40.0.13.1.60312

Model ID

1

Настройка KAFKA-топика

Server Address

127.0.0.1:9092

Group ID

rgorkgroup

Topic Name

RgogkInputTest

Отправка сообщения

Рисунок 3 – Отправка сообщений для тестирования
Figure 3 – Sending messages for testing

После получения сообщения из топика Kafka Broker, ПО с ИИ инициирует запрос на получение медицинских изображений, составляющих набор данных для тестирования. Затем запускается процесс обработки данных, по окончании которого графическая серия с разметкой и структурированный отчет отправляется обратно в PACS. Результат обработки записывается в следующий топик Kafka Broker, откуда собирается таблица результатов для дальнейшей оценки и расчета метрик (Рисунок 4).

Результаты обработки ПО

Выбор тестируемого ПО Тестируемое ПО ▾

Скачать таблицу

StudyUID	Result	Time
1.2.40.0.13.1.6031266466915007344757745322814756283	0	12:35:30
1.2.840.113619.2.428.3.2785674369.852.1711943103.812	1	12:35:40

Рисунок 4 – Таблица с результатами обработки исследований ПО с ИИ
Figure 4 – Table showing the results of the processing of software studies with AI

Помимо вышеперечисленных функций, модуль предоставляет инструменты для оценки технических аспектов работы ПО с ИИ. В частности, для оценки повторяемости, одни и те же исследования повторно подаются на вход ПО с ИИ, с целью сравнения полученных результатов. Для проверки наличия контроля входных данных, на вход тестируемого ПО подаются исследования, модальность которых не соответствует заявленным требованиям. Для оценки быстродействия ПО в модуле предусмотрена

функция измерения времени обработки исследований. Наконец, модуль позволяет фиксировать ситуации, когда ПО не предоставляет ожидаемые выходные данные, такие как дополнительная серия изображений или текстовое заключение.

В итоге работы модуля текстовое заключение и дополнительная графическая серия отправляются в PACS для проверки врачами, а таблица с результатами направляется в модуль расчета метрик.

Модуль просмотра. Модуль просмотра является удобным инструментом для оценки графических серий, полученных в результате обработки исследований ПО с ИИ. Модуль имеет доступ к базе данных PACS, откуда он получает файлы формата DICOM и отображает их содержимое в удобном для оценки виде (Рисунок 5). Слева представлена исходная графическая серия, справа-серия с разметкой ПО с ИИ. Используя данный модуль, врач-эксперт может сравнить между собой оригинальные серии исследования и размеченные моделью машинного обучения. Просмотрщик включает в себя такие стандартные механизмы как изменение яркости, контрастности и приближение изображений. Структурированные отчёты также отображаются в модуле в текстовом формате, располагаясь в отдельном окне на экране с привязкой к исследованию, по которому они были созданы. Также при помощи данного модуля возможно обнаружение различного рода дефектов в работе ПО с ИИ: отсутствие необходимых надписей на графической серии (название ПО, номер версии), обрезка изображений, нарушение яркости или контраста².

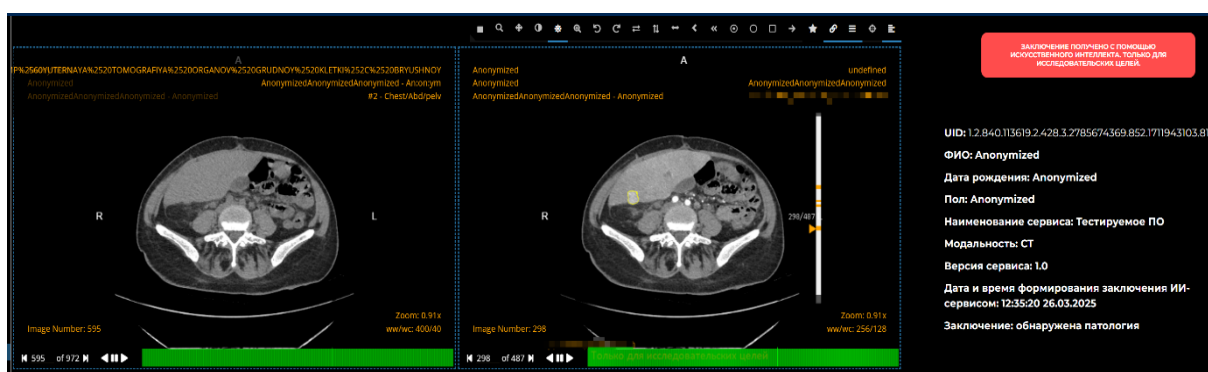


Рисунок 5 – Интерфейс модуля просмотра. Слева представлена исходная графическая серия, справа-серия с разметкой ПО с ИИ

Figure 5 – Viewer module interface. On the left is the original graphical series, on the right is the series with AI software markup

Модуль расчета метрик. Модуль расчета метрик позволяет рассчитывать метрики для оценки эффективности работы ПО с ИИ. На вход модуля подаются таблицы с значениями GT и таблица с результатами обработки ПО с ИИ. Список рассчитываемых метрик включает показатели для ПО с ИИ, решаемого задачу классификации: специфичность, чувствительность, точность, отношение правдоподобия положительного результата, отношение правдоподобия отрицательного результата, прогностическая ценность положительного результата, прогностическая ценность отрицательного результата, частота ложных срабатываний и F-мера. На выходе модуля расчета метрик формируется отчёт о результатах валидации – файл в формате pdf, который отправляется в базу данных. В отчет включены: ROC-кривая, четырехпольная

² Васильев Ю.А., Владимирский А.В., Омелянская О.В. и др. *Оценка зрелости технологий искусственного интеллекта для здравоохранения*. Москва: Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы; 2023. 28 с.

таблица и рассчитанные метрики с указанием 95 %-ого доверительного интервала. Отчет о результатах валидации позволит сделать вывод об эффективности работы ПО с ИИ.

Апробация. Для апробации платформы был использован предварительно сформированный набор данных компьютерной томографии области брюшной полости, включающий 5 исследований, не содержащих искомые признаки (изображения без кист печени), и 5 исследований с наличием искомого признака (изображения с кистами печени). Набор данных был предварительно загружен в PACS. С целью сравнительной оценки результатов, данные были обработаны с использованием платформы и без нее. Результаты апробации представлены в Таблице 2.

Таблица 2 – Результаты апробации платформы

Table 2 – Results of platform testing

Модули	Функционал	Метод апробации	Результат
Модуль тестирования	Подключение ПО к платформе	Проверка совместимости платформы и ПО	Соответствует
	Отправка DICOM из PACS в ПО	Проверка корректности отправки исследований в сервис из PACS	Соответствует
	Получение выходных данных ПО	Проверка корректности сохранения исследований	Соответствует
Модуль просмотра	Отображение дополнительных графических серий (SC)	Проверка наличия и корректного отображения дополнительных графических серий	Соответствует
	Отображение структурированных отчётов (SR)	Проверка наличия и корректного отображения информации из структурированного отчёта	Соответствует
	Отображение оригинальных исследований	Проверка наличия и корректного отображения оригинальных исследований из PACS	Соответствует
Модуль расчета метрик	Расчёт метрик	Проверка корректности проведения расчетов	Соответствует
	Построение ROC-кривой	Проверка корректности построения и отображения графика ROC	Соответствует
	Генерация pdf файла с результатами	Проверка полноты отчета	Соответствует

Апробация платформы продемонстрировала, что значения основных диагностических метрик ПО с ИИ для анализа КТ-исследований области брюшной полости, полученные с использованием платформы, не отличаются от результатов, полученных с помощью ручного тестирования. Апробация подтвердила корректное функционирование всех заявленных возможностей платформы. Простота использования платформы для тестирования ПО на основе ИИ была подтверждена успешным проведением апробации администратором без технической квалификации. Врач-рентгенолог был привлечен для тестирования модуля просмотра.

Обсуждение

В данной статье представлена разработка платформы для тестирования ПО с ИИ, назначением которого является диагностика различных патологических состояний по медицинским изображениям лучевой диагностики формата DICOM. Платформа, обладая модульной архитектурой, обеспечивает гибкость и масштабируемость процесса тестирования, позволяя адаптировать его к различным задачам и типам ПО с ИИ.

В отличие от существующих решений, таких как платформа, предложенная Gidde et al., которая ориентирована на валидацию ИИ-алгоритмов с возможностью разметки данных и сравнения результатов с GT [13], или AI-LAB [11], позволяющая пользователю просматривать наборы данных в формате DICOM, размечать их и оценивать ПО с ИИ, разработанная в этом исследовании платформа устраняет ключевые ограничения. В частности, существующие системы позволяют тестировать только одно ПО с ИИ за раз и требуют развертывания в локальной сети медицинского учреждения, что затрудняет выполнение клинической оценки нескольких ПО с ИИ.

В контексте разработки платформы, описанной в данной статье, был учтен опыт реализации проектов по клиническим испытаниям ПО с ИИ в лучевой диагностике, проводимых в ГБУЗ «НПКЦ ДиТ ДЗМ»³, а также проекта по практическому использованию ИИ в организациях здравоохранения города Москвы [14, 15]. Этот опыт способствовал формированию целостной методологии разработки и позволил достичь высокого качества итогового продукта.

Разработка платформы велась при получении постоянной обратной связи пользователей и других заинтересованных лиц для обеспечения удобства интерфейса, включения всего необходимого функционала. Также были предусмотрены технические возможности для универсальности подключения ПО с ИИ с разными протоколами. Платформа была разработана таким образом, чтобы поддерживать одновременное тестирование нескольких алгоритмов. Такое параллельное тестирование позволяет затратить гораздо меньше времени, чем было бы затрачено при использовании платформ-конкурентов, либо ручном тестировании.

Проведенная апробация подтвердила устойчивую функциональность созданной платформы и достоверность получаемых диагностических показателей, что свидетельствует о ее готовности к внедрению в клиническую практику.

Модульное построение платформы предусматривает простое внесение изменений в функциональную часть платформы путем включения в архитектуру новых программных элементов без необходимости изменять или отключать старые. В рамках дальнейшего развития разработанной платформы планируется внедрение функционала обучения и дообучения ИИ-моделей, что позволит проводить как тестирование, так и обучение сервисов централизовано в рамках одной инфраструктуры. В модуле просмотра также планируется реализовать возможность изменения просмотрщика, который будет учитывать все требования к тестированию ИИ-решений в диагностике, на усмотрения пользователя. Такая возможность позволит устранить зависимость от сторонних решений, обеспечить полный контроль над функционалом просмотра и анализа DICOM-данных, а также позволит создать полностью автономную коробочную версию платформы, не требующей интеграции с внешними коммерческими продуктами. Такой подход повысит гибкость развертывания системы и расширит возможности ее адаптации под специфические исследовательские и клинические задачи.

³ Васильев Ю.А., Владимирский А.В., Шарова Д.Е. и др. *Клинические испытания систем искусственного интеллекта (лучевая диагностика)*. Москва: Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы; 2023. 40 с.

Разработанная платформа соответствует приоритетам Российской Федерации в области цифровизации здравоохранения и развития медицинских технологий⁴. Ее внедрение способствует реализации государственных задач по повышению доступности и качества диагностики за счет применения ИИ, что особенно актуально в рамках цифровой трансформации здравоохранения⁵.

Заключение

Разработана единая модульная платформа, обеспечивающая проведение валидации медицинского ПО с ИИ. Платформа позволяет оптимизировать процессы и снижать риски возникновения ошибок при валидации ПО с ИИ, назначением которого является диагностика различных патологических состояний по медицинским изображениям лучевой диагностики формата DICOM. Платформа открывает возможности стандартизации тестирования ПО с ИИ как для целей клинической валидации, так и для контроля качества на этапе эксплуатации системы. На первом этапе апробация платформы выполнена для ПО с ИИ, предназначенного для решения задачи классификации. Однако были заложены возможности для адаптации модулей для ПО с ИИ, применяемых в других задачах (локализация патологии на изображении и ее мониторинг).

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Alowais S.A., Alghamdi S.S., Alsuhebany N., et al. Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice. *BMC Medical Education*. 2023;23(1). <https://doi.org/10.1186/s12909-023-04698-z>
2. Тюрин И.Е. Лучевая диагностика в Российской Федерации. *Онкологический журнал: лучевая диагностика, лучевая терапия*. 2018;1(4):43–51.
Tyurin I.E. Radiology in the Russian Federation. *Journal of Oncology: Diagnostic Radiology and Radiotherapy*. 2018;1(4):43–51. (In Russ.).
3. Strohm L., Hehakaya Ch., Ranschaert E.R., Boon W.P.C., Moors E.H.M. Implementation of Artificial Intelligence (AI) Applications in Radiology: Hindering and Facilitating Factors. *European Radiology*. 2020;30(10):5525–5532. <https://doi.org/10.1007/s00330-020-06946-y>
4. Kelly B.S., Judge C., Bollard S., 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>
5. Васильев Ю.А., Владимировский А.В., Арзамасов К.М., Шулькин И.М., Астапенко Е.В., Пестренин Л.Д. Ограничения при применении сервисов искусственного интеллекта для анализа рентгенограмм органов грудной клетки. *Digital Diagnostics*. 2024;5(3):407–420. <https://doi.org/10.17816/DD626310>
Vasilev Yu.A., Vladzimirsky A.V., Arzamasov K.M., Shulkin I.M., Astapenko E.V., Pestrenin L.D. Limitations of Using Artificial Intelligence Services to Analyze Chest X-Ray Imaging. *Digital Diagnostics*. 2024;5(3):407–420. (In Russ.). <https://doi.org/10.17816/DD626310>

⁴ Указ Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» (с изменениями и дополнениями от 15 февраля 2024 г.). Гарант. URL: <https://base.garant.ru/72838946/> (дата обращения: 15.04.2025).

⁵ Распоряжение Правительства Российской Федерации от 17 апреля 2024 г. № 959-р «Об утверждении стратегического направления в области цифровой трансформации здравоохранения». Гарант. URL: <https://base.garant.ru/408913257/> (дата обращения: 15.04.2025).

6. Гусев А.В., Морозов С.П., Кутичев В.А., Новицкий Р.Э. Нормативно-правовое регулирование программного обеспечения для здравоохранения, созданного с применением технологий искусственного интеллекта, в Российской Федерации. *Медицинские технологии. Оценка и выбор*. 2021;43(1):36–45. <https://doi.org/10.17116/medtech20214301136>
Gusev A.V., Morozov S.P., Kutichev V.A., Novitsky R.E. Legal Regulation of Artificial Intelligence Software in Healthcare in the Russian Federation. *Medical Technologies. Assessment and Choice*. 2021;43(1):36–45. (In Russ.). <https://doi.org/10.17116/medtech20214301136>
7. Васильев Ю.А., Владимирский А.В., Омельянская 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., Vladzimirsky 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>
8. Erickson B.J., Kitamura F. Magician's Corner: 9. Performance Metrics for Machine Learning Models. *Radiology: Artificial Intelligence*. 2021;3(3). <https://doi.org/10.1148/ryai.2021200126>
9. Hicks S.A., Strümke I., Thambawita V., et al. On Evaluation Metrics for Medical Applications of Artificial Intelligence. *Scientific Reports*. 2022;12. <https://doi.org/10.1038/s41598-022-09954-8>
10. Müller D., Soto-Rey I., Kramer F. Towards a Guideline for Evaluation Metrics in Medical Image Segmentation. *BMC Research Notes*. 2022;15. <https://doi.org/10.1186/s13104-022-06096-y>
11. Brink L., Coombs L.P., Veettil D.K., et al. ACR's Connect and AI-LAB Technical Framework. *JAMIA Open*. 2022;5(4). <https://doi.org/10.1093/jamiaopen/ooac094>
12. Павлов Н.А., Андрейченко А.Е., Владимирский А.В., Ревазян А.А., Кирпичев Ю.С., Морозов С.П. Эталонные медицинские датасеты (MosMedData) для независимой внешней оценки алгоритмов на основе искусственного интеллекта в диагностике. *Digital Diagnostics*. 2021;2(1):49–66. <https://doi.org/10.17816/DD60635>
Pavlov N.A., Andreychenko A.E., Vladzimirsky A.V., Revazyan A.A., Kirpichev Y.S., Morozov S.P. Reference Medical Datasets (MosMedData) for Independent External Evaluation of Algorithms Based on Artificial Intelligence in Diagnostics. *Digital Diagnostics*. 2021;2(1):49–66. <https://doi.org/10.17816/DD60635>
13. Gidde P.S., Prasad Sh.S., Singh A.P., et al. Validation of Expert System Enhanced Deep Learning Algorithm for Automated Screening for COVID-Pneumonia on Chest X-Rays. *Scientific Reports*. 2021;11. <https://doi.org/10.1038/s41598-021-02003-w>
14. Владимирский А.В., Васильев Ю.А., Арзамасов К.М. и др. *Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента*. Москва: Издательские решения; 2023. 388 с.
15. Артюкова З.Р., Петрайкин А.В., Кудрявцев Н.Д. и др. Опыт применения сервисов искусственного интеллекта для диагностики компрессионных переломов тел позвонков по данным компьютерной томографии: от тестирования до апробации. *Digital Diagnostics*. 2024;5(3):505–518. <https://doi.org/10.17816/DD624250>
Artyukova Z.R., Petraikin A.V., Kudryavtsev N.D., et al. Experience with Artificial Intelligence Algorithms for the Diagnosis of Vertebral Compression Fractures Based on Computed Tomography: From Testing to Practical Evaluation. *Digital Diagnostics*. 2024;5(3):505–518. (In Russ.). <https://doi.org/10.17816/DD624250>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Ковальчук Анна Юрьевна, младший научный сотрудник, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, Российская Федерация.

e-mail: KovalchukAY2@zdrav.mos.ru

ORCID: [0009-0006-7975-8974](https://orcid.org/0009-0006-7975-8974)

Anna Y. Kovalchuk, junior researcher, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, the Russian Federation.

Пономаренко Артем Павлович, инженер, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, Российская Федерация.

e-mail: PonomarenkoAP@zdrav.mos.ru

ORCID: [0009-0000-7716-6015](https://orcid.org/0009-0000-7716-6015)

Artem P. Ponomarenko, engineer, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, the Russian Federation.

Арзамасов Кирилл Михайлович, кандидат медицинских наук, руководитель отдела, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, Российская Федерация.

e-mail: ArzamasovKM@zdrav.mos.ru

ORCID: [0000-0001-7786-0349](https://orcid.org/0000-0001-7786-0349)

Kirill M. Arzamasov, Candidate of Medical Sciences, Head of Department, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, the Russian Federation.

Статья поступила в редакцию 21.04.2025; одобрена после рецензирования 08.05.2025; принята к публикации 16.05.2025.

The article was submitted 21.04.2025; approved after reviewing 08.05.2025; accepted for publication 16.05.2025.