

ИСКУССТВЕННЫЙ МЕДИЦИНСКИЙ ИНТЕЛЛЕКТ «ФТИЗИСБИОМЕД»: ПРОГРАММА АВТОМАТИЗИРОВАННОГО АНАЛИЗА ЦИФРОВЫХ РЕНТГЕНОГРАММ ОРГАНОВ ГРУДНОЙ КЛЕТКИ/ФЛЮОРОГРАММ

DOI: 10.17691/stm2023.15.4.01

УДК 617.541–073.756.8–004.891.3

Поступила 13.03.2023 г.

© Ю.Т. Гогоберидзе, старший инженер-разработчик¹;
В.И. Классен, д.т.н., профессор, генеральный директор²; председатель правления¹;
М.Я. Натензон, к.т.н., старший научный сотрудник³; председатель Совета директоров⁴;
И.А. Просвиркин, к.т.н., IT-директор²;
А.В. Владзимирский, д.м.н., зам. директора по научной работе⁵;
Д.Е. Шарова, руководитель отдела инновационных технологий⁵;
В.В. Зинченко, начальник сектора клинических и технических испытаний⁵

¹ООО «ФтизисБиоМед», ул. Карла Маркса, 135, Республика Татарстан, Чистополь, 422980;

²АО «РК «Вектор», ул. Карла Маркса, 135, Республика Татарстан, Чистополь, 422980;

³Российская академия наук, Ленинский проспект, 14, Москва, 119071;

⁴Научно-производственное объединение «Национальное телемедицинское агентство»,
ул. Демьяна Бедного, 9, офис 8, Москва, 123423;

⁵Научно-практический клинический центр диагностики и телемедицинских технологий
Департамента здравоохранения города Москвы, ул. Петровка, 24, стр. 1, Москва, 127051

Увеличивающиеся из года в год объемы диагностических медицинских исследований породили разумное стремление к развитию и внедрению новых технологий в сфере диагностики и анализа медицинских данных. Одним из наиболее перспективных решений этой проблемы, зарекомендовавших себя в ходе массового практического применения, стали алгоритмы искусственного интеллекта (ИИ). В ходе трехлетнего (начиная с 2020 г.) Московского эксперимента удалось выработать методологии использования ИИ и успешно внедрить его в систему здравоохранения регионального уровня.

В данной статье авторы делятся опытом разработки медицинского ИИ на примере ИИ-сервиса «ФтизисБиоМед» и результатами его применения в условиях реальной клинической практики. Данный ИИ-сервис показал свои качество и надежность, что подтверждается технологическим мониторингом.

Клинические испытания ИИ-сервиса «ФтизисБиоМед» проведены на специально подготовленном верифицированном наборе данных (n=1536) с учетом эпидемиологических показателей распространенности основных заболеваний органов грудной клетки. Средняя чувствительность сервиса составила 0,975 (95% CI: 0,966–0,984).

Медицинский ИИ-сервис «ФтизисБиоМед» зарегистрирован в качестве медицинского изделия (регистрационное удостоверение на медицинское изделие №РЗН 2022/17406 от 31 мая 2022 г.) и активно применяется в Российской Федерации как диагностический инструмент, позволяющий снизить нагрузку на врача-рентгенолога и ускорить процесс получения медицинского заключения.

Ключевые слова: искусственный медицинский интеллект; ИИ-сервис «ФтизисБиоМед»; нейронные сети; заболевания органов грудной клетки; СППВР; ЕРИС ЕМИАС.

Как цитировать: Gogoberidze Y.T., Klassen V.I., Natenzon M.Y., Prosvirkin I.A., Vladzimirsky A.V., Sharova D.E., Zinchenko V.V. PhthisisBioMed artificial medical intelligence: software for automated analysis of digital chest X-ray/fluorograms. *Sovremennye tehnologii v medicine* 2023; 15(4): 5, <https://doi.org/10.17691/stm2023.15.4.01>

Для контактов: Гогоберидзе Юрий Тенгизович, e-mail: gut@vector.ru

PhthisisBioMed Artificial Medical Intelligence: Software for Automated Analysis of Digital Chest X-ray/Fluorograms

Y.T. Gogoberidze, Senior Development Engineer¹;
V.I. Klassen, DSc, Professor, General Director²; Board Chairman¹;
M.Y. Natenzon, PhD, Senior Researcher³; Chairman of the Board of Directors⁴;
I.A. Prosvirkin, PhD, IT Director²;
A.V. Vladzimirsky, MD, DSc, Deputy Director for Scientific Research⁵;
D.E. Sharova, Head of Innovative Technologies Department⁵;
V.V. Zinchenko, Head of Clinical and Technical Trials Sector⁵

¹PhthisisBioMed LLC, 135 Karla Marksa St., Chistopol, Republic of Tatarstan, 422980, Russia;

²RK Vector JSC, 135 Karla Marksa St., Chistopol, Republic of Tatarstan, 422980, Russia;

³Russian Academy of Sciences, 14 Leninsky Prospekt, Moscow, 119071, Russia;

⁴Scientific and Production Association National Telemedicine Agency, Office 8, 9 Demyana Bednogo St., Moscow, 123423, Russia;

⁵Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Department of Health, Bldg. 1, 24 Petrovka St., Moscow, 127051, Russia

The scope of diagnostic medical examinations increases from year to year causing a reasonable desire to develop and implement new technologies to diagnostics and medical data analysis. Artificial intelligence (AI) algorithms became one of the most promising solutions to this problem and proved themselves in the course of mass practical application. During the three-year Moscow experiment started in 2020, the possibility was achieved to develop methodologies of AI use and to successfully implement it into the regional level healthcare system.

In this article, the authors share their experience in developing a medical AI service using the example of PhthisisBioMed AI service and the results of its application in real clinical activities environment. This AI service has shown its quality and reliability confirmed by technological monitoring.

Clinical trials of PhthisisBioMed AI service were conducted on a specially prepared verified data set (n=1536) considering epidemiological indicators of the thoracic organs major diseases prevalence. The mean sensitivity of the service was 0.975 (95% CI: 0.966–0.984).

PhthisisBioMed medical AI service is registered as a medical device (medical device registration certificate No.RZN 2022/17406 dated May 31, 2022) and is actively used in the Russian Federation as a diagnostic tool to reduce the burden on radiologists and to accelerate the process of medical report obtaining.

Key words: artificial medical intelligence; PhthisisBioMed AI service; neural networks; thoracic organ diseases; CDSS; URIS UMIAS.

Введение

С развитием здравоохранения и увеличением объема оказываемых медицинских услуг, что в первую очередь обусловлено внедрением массовой диспансеризации и первичной медицинской помощи, неизбежно возрастает количество диагностических данных, которые необходимо принимать в расчет при постановке диагноза и назначении лечения.

Согласно данным Росстата, число ежегодных флюорографических обследований возросло с 70 млн человек в 2017 г. до 87,7 млн в 2023 г. (план Минздрава РФ) [1], в то время как общее количество специалистов-рентгенологов (фтизиатров) сокращалось в среднем примерно на 100 человек в год.

Обработка огромного массива информации требует перехода к цифровому здравоохранению и создания систем поддержки принятия врачебных решений (СППВР) для помощи конкретному больному с использованием аналитики больших данных для оптимизации оказания медицинской помощи.

Создание продуктов искусственного медицинского интеллекта (ИМИ) с глубоким машинным обучением и высокими техническими характеристиками возможно только при тесном взаимодействии медиков и IT-специалистов.

Массовое практическое применение ИМИ дало понимание того, что современное медицинское сообщество резко поляризовано в вопросе доверия к искусственному интеллекту (ИИ): от резкого, полного неприятия до безоговорочной веры и отсутствия минимального скепсиса. Попытка автоматизировать, почти «механизировать» процесс работы врача зачастую не дает желаемого результата из-за недопонимания спектра задач, которые способен решать ИМИ, и набора ограничений, заложенных в саму концепцию ИИ. Стремление некоторых специалистов максимально снять с себя рутинную нагрузку нередко приводит к постоянному неосмысленному согласию с заключениями ИМИ.

В данной статье обсуждаются вопросы внедрения и массовой эксплуатации СППВР в контексте кон-

кретного интеллектуального медицинского сервиса на базе ИИ, а также в свете большого продолжающегося Московского эксперимента, частью которого упомянутый сервис является.

Медицинский ИИ-сервис «ФтизисБиоМед»

Модель и архитектура

Медицинский ИИ-сервис — программный продукт, предоставляющий пользователю обособленную медицинскую цифровую услугу. Это сложная многоблочная структура, содержащая нейросетевые алгоритмы, обученные, как правило, выполнению диагностической задачи.

Медицинский ИИ-сервис, или ИМИ, «ФтизисБиоМед» начал разрабатываться в 2014 г. предприятием ООО «ФтизисБиоМед» и прошел несколько длительных циклов разработки, совершенствования и эксплуатации. В настоящий момент действует версия 3.3.

Ядром системы является интеллектуальный блок.

Интеллектуальный блок. В основе интеллектуального блока ИМИ «ФтизисБиоМед» находится модель глубокой полносверточной нейронной сети, приспособленной для детекции и локализации легочных патологий. Базовой архитектурой для нейронной сети ИМИ «ФтизисБиоМед» послужила архитектура U-net, которая в дальнейшем была модифицирована, чтобы лучше соответствовать поставленной задаче. Подробнее этот вопрос рассмотрен в предыдущей публикации авторов [2]. Здесь же приведем общие данные модели ResUNet (рис. 1). Интеллектуальный блок образуют три такие нейронные сети, формирующие ансамбль.

Дополнительным элементом служит многослойный классификатор, также представляющий собой модель глубокого обучения. Классификатор выполняет функцию разделения и распознавания типов (классов) патологий, детектированных и локализованных

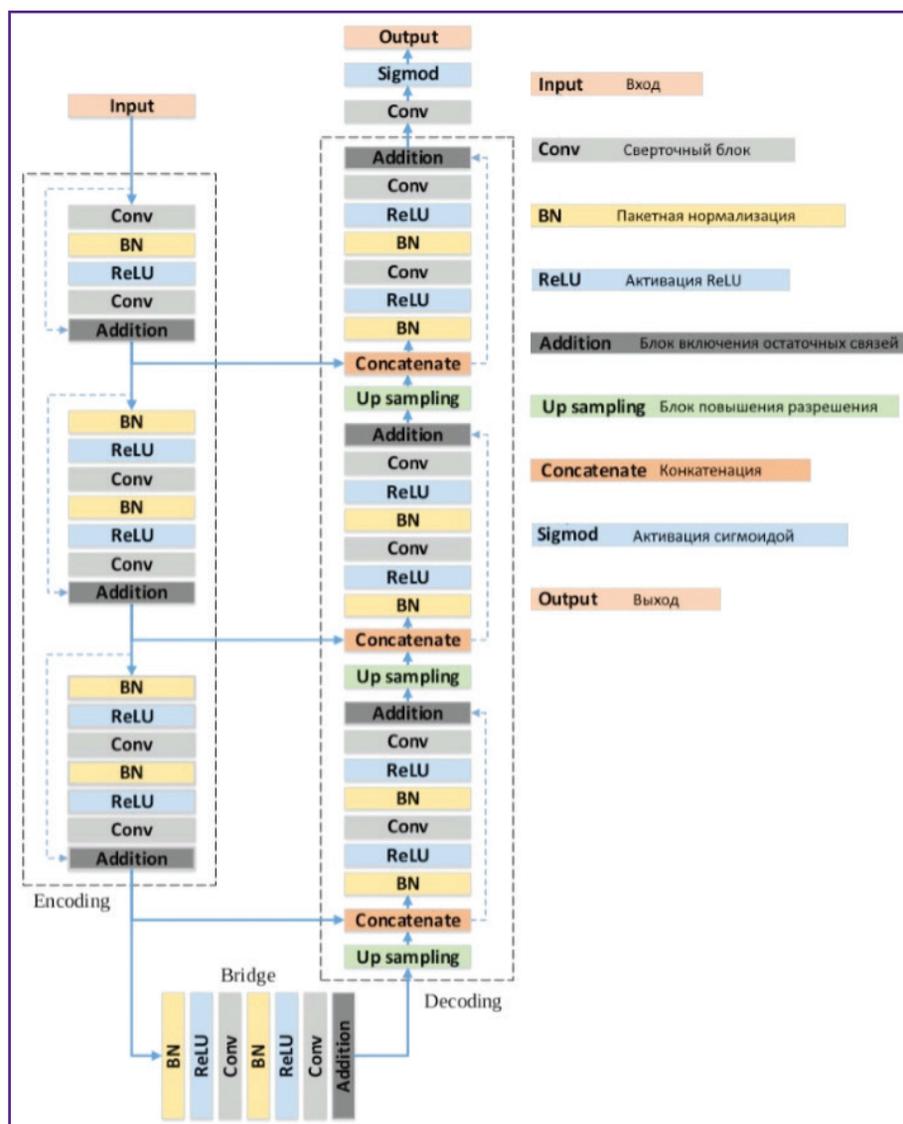


Рис. 1. Архитектура ResUNet

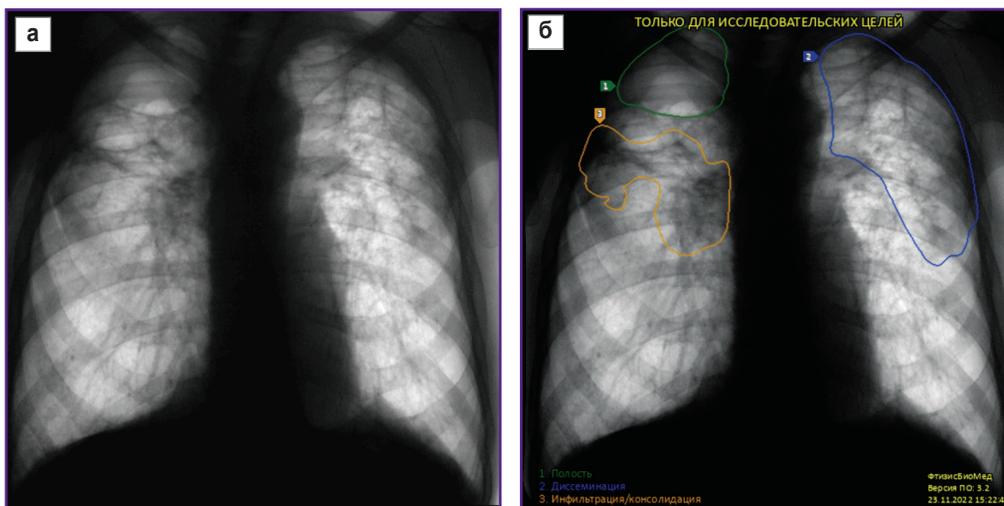


Рис. 2. Результат работы искусственного медицинского интеллекта «ФтизисБиоМед»:
 а — оригинальное изображение, поданное на анализ; б — результат анализа

на изображениях, которые подвергнуты анализу ансамблем локализаторов. Базовой моделью классификатора является DenseNet201 [3]. На рис.2 представлен результат обработки ИИ-сервисом «ФтизисБиоМед» флюорографического снимка.

Следующим по значимости в структуре ИИ-сервиса является надстроечный блок.

Надстроечный блок. Одной из проблем анализа медицинских изображений является отсутствие стандартизации создания самих изображений. Даже если формальные регламенты существуют, зачастую они являются рамочными.

Снимки легких, сделанные в момент, когда пациент стоит, лежит или сидит, обладают различными специфическими свойствами, кроме того, текущее клиническое состояние также имеет значение. Не стоит забывать и о том, что рентгеновские и флюорографические аппараты разных производителей обладают различными свойствами и опциями. Все это осложняет анализ снимков с помощью одних только моделей нейронных сетей для поиска патологий и требует обеспечения сервиса дополнительными алгоритмами, учитывающими эти особенности аппаратуры.

Надстроечный блок служит для предварительной обработки изображения до отправки его на анализ. Это подразумевает следующие манипуляции со снимком:

поиск области интереса на изображении (в данном случае — легких), что предполагает под собой интеллектуальную (нейросетевую) и алгоритмическую сегментацию органа;

предварительный анализ свойств самого изображения, т.е. алгоритмическое определение — представлен снимок в негативе или позитиве, а также нейросетевой анализ — является ли снимок сам по себе корректным образцом для анализа (является ли снимок рентгенографическим изображением грудной клетки в прямой проекции).

Для предварительной обработки данных сервис «ФтизисБиоМед» использует несколько механизмов: фильтрацию, определение яркостных характеристик снимка (негатив/позитив, диапазон), сегментацию области интереса.

Кроме того, важным элементом является выявление инородных тел на изображениях органов грудной клетки, таких как электрокардиостимуляторы или операционные швы, а также любых посторонних объектов, попавших на изображение, например нательных цепочек. Данные объекты нередко приводят к ложным срабатываниям нейронных сетей интеллектуального блока. Чтобы минимизировать подобные срабатывания, ложные активации, образованные в результате возбуждения нейронной сети интеллектуального блока, распознаются и ликвидируются специальной надстроечной моделью (рис. 3).

Последним блоком является блок взаимодействия с клиентом.

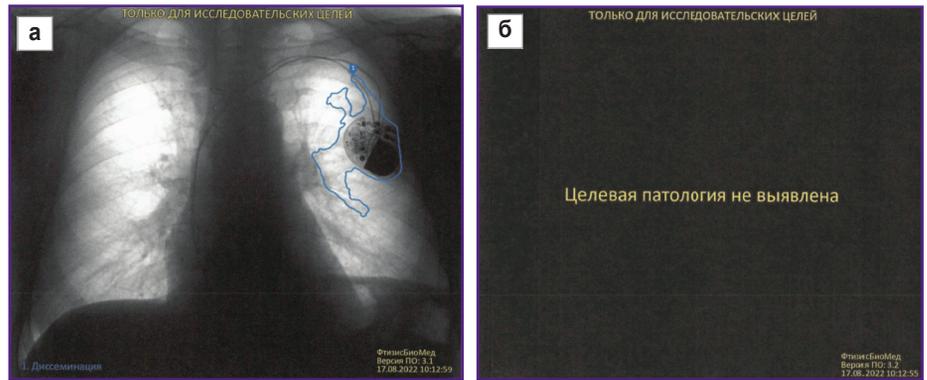
Блок взаимодействия с клиентом. Этот блок представляет собой комплекс мер, алгоритмов и протоколов взаимодействия для автоматизации работы сервиса в условиях реального клинического процесса. В целом блок обеспечивает создание бесперебойных линий обработки исследований, совершаемых параллельно и удаленно относительно физического расположения самого сервера сервиса. Благодаря данному блоку возможны проведение удаленного параллельного анализа и поддержка врачей не только в отдельных медицинских учреждениях, но и в медицинских инфраструктурах регионального масштаба в режиме реального времени.

Обучение

Добиться высокой точности диагностики ИИ невозможно без обучающей выборки, размеченной с высокой достоверностью. Согласно исследованию

Рис. 3. Результат работы надстроечной модели, предназначенной для минимизации ложных активаций на инородных телах:

а — активация интеллектуального блока без надстроечной нейронной сети, распознающей активацию на инородных телах;
б — результат после постобработки



сотрудников Сеульского национального университета [4], посвященному оценке диагностической точности заключений специалистов-рентгенологов, для квалифицированных специалистов площадь под ROC кривой составляет 0,781–0,907 с доверительной вероятностью 0,95 для задач локализации.

Чтобы оценить точность диагностики, выполненной врачами-экспертами, которые осуществляют разметку выборки для машинного обучения сервиса «ФтизисБиоМед», проведено предварительное исследование на пересекающихся выборках. Три врача независимо друг от друга выполняли разметку одной и той же выборки. После этого арбитры Центрального научно-исследовательского института туберкулеза (ЦНИИТ) проводили мониторинг результатов разметки. По каждой найденной патологии врач-арбитр с помощью специального программного обеспечения (рис. 4) принимал решение: присутствует ли данная патология на снимке, и если присутствует, то кто из

трех врачей-экспертов, выполнявших разметку, локализовал ее наиболее точно.

В разметке принимали участие два врача-эксперта районного уровня и один врач-эксперт регионального уровня со стажем более 15 лет.

Два эксперта выполнили разметку 861 снимка каждый, один разметил только 267 снимков из общей выборки (рис. 5). Каждый эксперт выполнял локализацию патологий путем разметки маски на снимке. (На одном снимке может быть размечено несколько масок.)

Результаты арбитража по бинарным заключениям (есть патология/нет патологии) для каждого эксперта приведены на гистограммах (рис. 6). Врач регионального уровня (эксперт №1) показал наилучшие результаты (чувствительность — 73,27%, специфичность — 96,72%); эксперты №2 и 3 продемонстрировали высокую специфичность — 99,34 и 99,47% соответственно. Такие показатели могут быть связаны со специализацией врача.

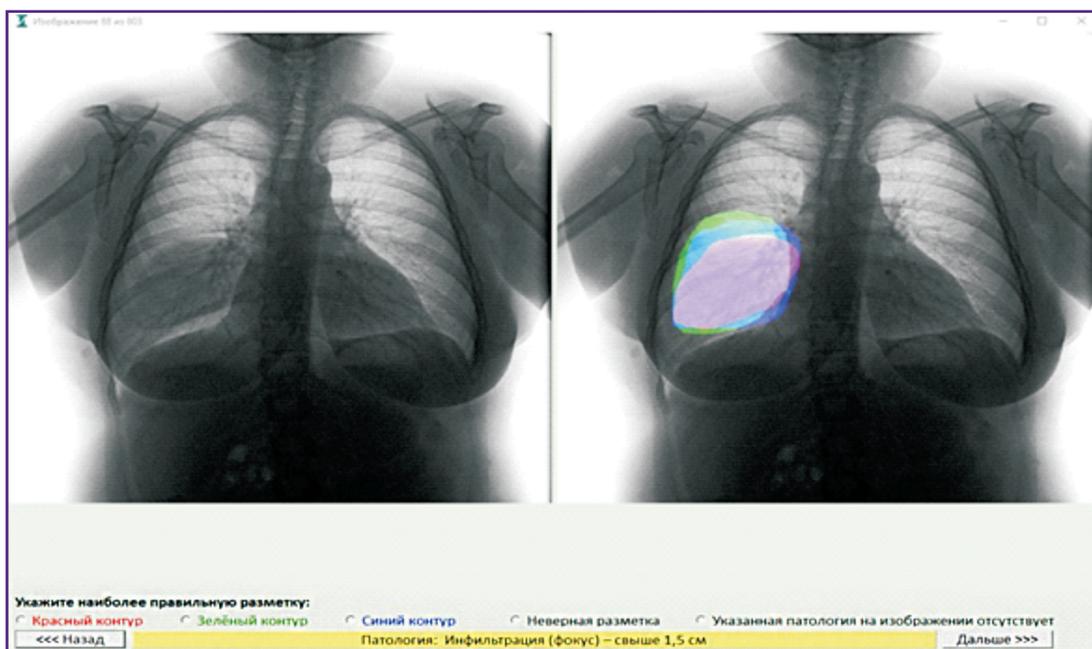


Рис. 4. Программное обеспечение для мониторинга результатов разметки предварительного исследования

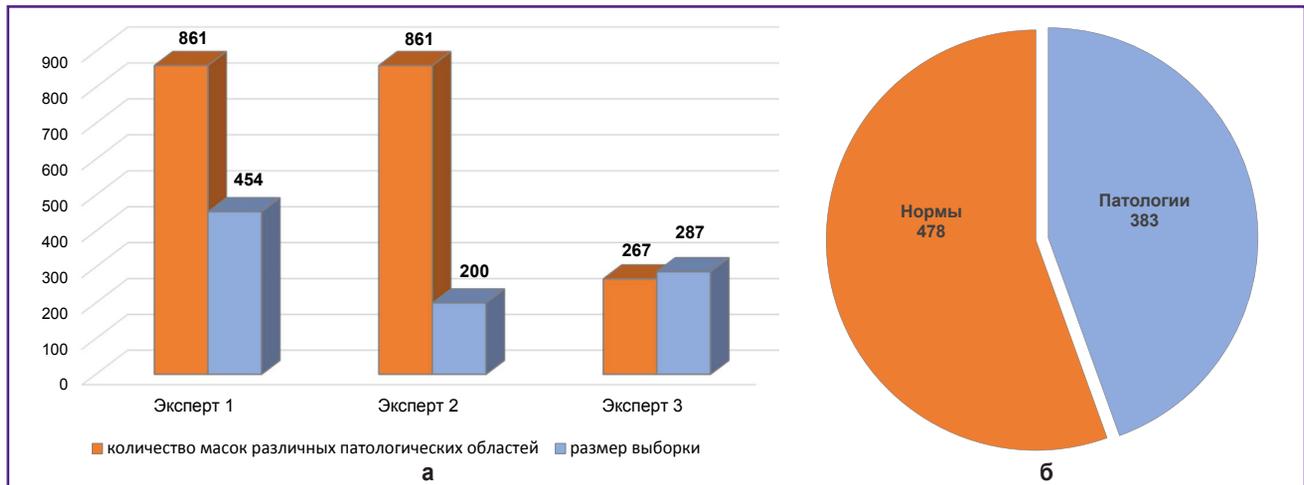


Рис. 5. Структура выборки для предварительного исследования:

а — размер выборки и число размеченных масок; **б** — разбиение выборки (n=861) на нормы и патологии



Рис. 6. Результаты арбитража по бинарным заключениям

Высокая чувствительность требуется при диагностике изображений на потоке, т.е. с низкой претестовой вероятностью, например при флюорографическом скрининге. При массовом скрининге лучше провести дополнительную диагностику здорового пациента, чем пропустить пациента с патологией.

Высокая специфичность свойственна для задач диагностики изображений с высокой претестовой вероятностью. Например, в рентген-кабинет поликлиники поступают пациенты с подозрением на патологию. Недостаточно обоснованное лечение из-за ошибочного заключения в данном случае может привести к побочным эффектам и дополнительным затратам.

Результаты арбитража по качеству разметки приведены на гистограммах (рис. 7). В анализ включены все патологии (маски), найденные хотя бы одним врачом. Оценка «Верно» по маске назначалась в случае, когда в результате арбитража было сделано заключение о

том, что данный эксперт выполнил наиболее точную разметку. Оценка «Частично верно» назначалась в случае, когда размеченная область патологии совпадает с наиболее точной разметкой более чем на 50%. Оценка «Неверно» назначалась в случае, когда размеченная область патологии совпадает с наиболее точной разметкой менее чем на 50% (или не пересекается). Оценка «Патология отсутствует» назначалась в случае, если арбитраж не подтвердил наличия данной патологии на снимке.

Анализ результатов описанного исследования продемонстрировал низкий уровень сходимости заключений врачей как по факту наличия или отсутствия патологии, так и по локализации патологии на снимке. В связи с этим, чтобы максимально сократить количество ложных результатов разметки, был предложен метод тройного чтения врачами различной квалификации с арбитражем их диагнозов врачами наивысшей квалификации.

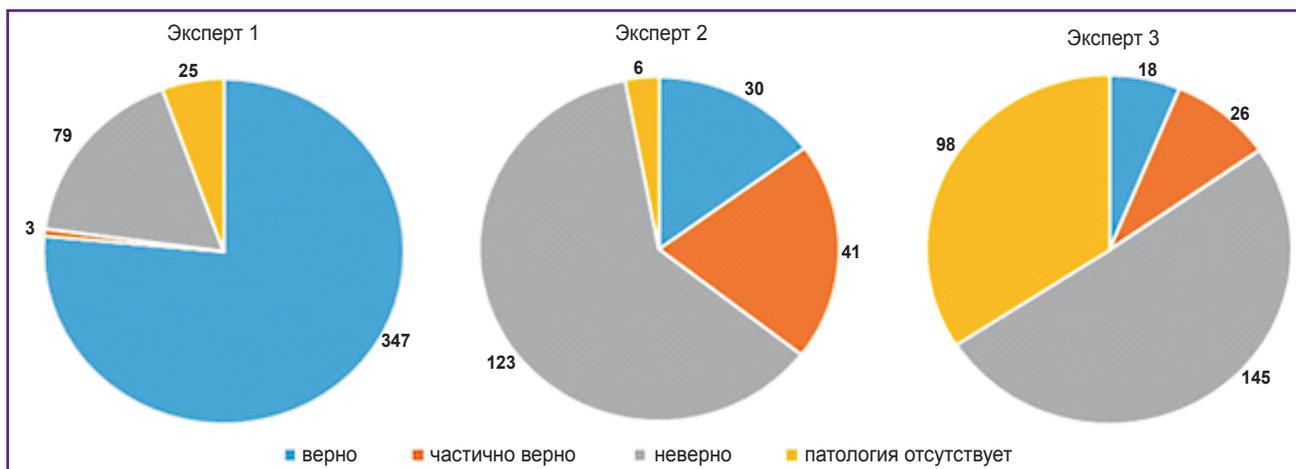


Рис. 7. Результаты арбитража по качеству разметки

Приняв во внимание результаты исследования сходимости мнений врачей-специалистов с большим стажем работы, мы сделали вывод, что первичная выборка данных должна быть собрана из диверсифицированных источников и маркироваться группой квалифицированных специалистов. Должен быть проведен отсев снимков низкого качества, а также несоответствующих иным техническим критериям.

Для обучения сервиса «ФтизисБиоМед» в его базовой версии использовалась выборка из порядка 300 000 медицинских изображений. Однако дальнейшие исследования, в том числе упомянутый выше эксперимент и практические испытания сервиса, показали несостоятельность неизбирательного подхода к обучающим выборкам, и было решено перейти к подходу «качество прежде количества». Таким образом из изначальной выборки автоматизированными методами, а затем методами ручного отбора было получено 9593 снимка, которые были отправлены на переработку квалифицированным специалистам-рентгенологам со стажем работы не менее 14 лет. 4533 из 9593 оценены как непатологические, 5060 — как с возможным наличием патологий. Затем часть снимков была удалена из выборки как вызывающая разногласия у экспертов. Итоговая базовая выборка обучения составила 8662 снимка, из которых 2904 нормы и 5758 патологий. В работе над разметкой снимков принимали участие специалисты Республиканского клинического противотуберкулезного диспансера и другие специалисты-радиологи поликлиник Москвы и Республики Татарстан.

Валидация

Валидация (внутренняя) должна представлять собой тест по выборке, подобранной специальным образом: во-первых, включать в себя все распознаваемые ИМИ патологии; во-вторых — нормы. Соотношение норм и патологий по умолчанию должно быть близко к 50/50 [5] или же соответствовать известному соотно-

шению в популяции, для которой будут оказываться услуги ИИ [6].

Врачи, разметившие выборку, высказывают консолидированное мнение по каждому из снимков выборки и совместно разрабатывают маркировку. Все снимки, по которым мнения разошлись, отбрасываются и заменяются. После чего по заменяющим снимкам тоже проводится процедура согласования. Консилиум врачей-составителей выборки не должен совпадать или пересекаться с составителями выборок обучения. Стаж работы по специальности составителей — не менее 14 лет.

Для внутренней валидации сервиса «ФтизисБиоМед» использовали (на момент написания статьи) выборку из 115 изображений: 52 без признаков патологий и 63 — с различными признаками патологий (виды размеченных патологий: инфильтрация — 35, пневмоторакс — 4, плевральный выпот — 23, очаги — 4, диссеминация — 2, кальцинаты — 13, нарушение целостности кортикального слоя — 3).

Для оценки эффективности сервиса применяли стандартные метрики диагностической точности, такие как площадь под характеристической ROC-кривой (AUCROC), чувствительность, специфичность и другие.

Значимость результатов определяется 95% доверительным интервалом (confidence interval, CI). При расчете CI для AUCROC используется алгоритм DeLong [7].

Приемлемые значения чувствительности и специфичности выбираются стандартно по критерию максимизации индекса Юдена — для максимизации общей валидности или по критерию максимизации значения предсказательной ценности отрицательного результата (предсказательная стандартная метрика диагностической точности, *англ.* negative predictive value, NPV) — для минимизации ложноотрицательных срабатываний.

Порог индекса Юдена — 0,79. Порог max NPV — 0,79. При проведении внутренней валидации на

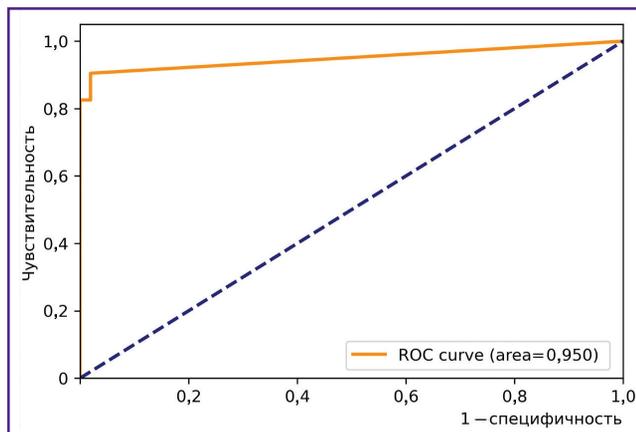


Рис. 8. ROC-кривая, построенная на основании обработки сервисом набора из 115 исследований (версия 3.2 сервиса «ФтизисБиоМед»)

наборе данных из 115 изображений достигнуто значение AUCROC в 95,0% случаев (рис. 8).

Московский эксперимент и внешняя валидация

В начале 2020 г. Департамент здравоохранения Москвы объявил о приеме заявок на участие в «Эксперименте по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения Москвы» (далее — эксперимент) [8]. Нормативной базой для эксперимента стало постановление Правительства Москвы №1543-ПП от 21.11.2019 г. и соответствующий приказ Департамента здравоохранения г. Москвы [9]. Цель эксперимента — научное исследование возможности использования в системе здравоохранения Москвы методов поддержки принятия врачебных решений на основе результатов анализа данных с применением передовых инновационных технологий. Эксперимент проводили на базе платформы «Единый радиологический информационный сервис г. Москвы» (ЕРИС ЕМИАС), объединяющей все медицинские организации Департамента здравоохранения Москвы (около 2000 единиц оборудования для лучевой диагностики всех модальностей) [10]. По правилам, установленным организаторами, участниками эксперимента могли стать юридические лица, предоставляющие сервисы (программное обеспечение) на базе технологий компьютерного зрения, которые предназначены для анализа медицинских изображений по следующим видам исследований:

компьютерная томография и низкодозная компьютерная томография грудной клетки с целью выявления рака легких;

маммография с целью выявления рака молочных желез;

рентгенография легких с целью определения патологии легких;

компьютерная томография грудной клетки с целью выявления новой коронавирусной инфекции.

27.07.2020 г. сервис «ФтизисБиоМед» был допущен в ЕРИС ЕМИАС и поставлен на потоковый анализ исследований. На ранних этапах совместно с организаторами эксперимента велась работа по устранению технических несовершенств интеграции информационных систем, развитию стандартизации, в том числе в части обязательности унифицированного заполнения DICOM-тегов [11].

Для решения технических проблем, а также с целью повышения доступности сервиса медицинским специалистам был разработан интерфейс для интеграции разнородных систем, который включал в себя программно-аппаратные шлюзы, подключаемые к ЕРИС ЕМИАС через разных интернет-провайдеров, что в свою очередь позволило решить задачи аппаратного и программного резервирования, а также распределения нагрузки (рис. 9).

Архитектура ЕРИС ЕМИАС предполагает взаимодействие сервисов ИИ с двумя интерфейсами: единой системой уведомлений для внешних взаимодействий (ЕСУВВ) и системой передачи и архивации медицинских изображений стандарта DICOM (PACS-сервер). ЕСУВВ — подсистема ЕРИС ЕМИАС, предназначенная для взаимодействия и обмена информацией между узлами ЕРИС ЕМИАС, построенная на платформе Apache Kafka [12]. Эта подсистема в режиме реального времени публикует для всех подключенных к ней абонентов задания на обработку вновь поступивших в ЕРИС ЕМИАС исследований. Функциональная схема интерфейса интеграции сервиса с ЕРИС ЕМИАС приведена на рис. 10.

Нагрузку (поток заданий на обработку) между подключенными сетевыми узлами одного абонента автоматически распределяет ЕСУВВ. Система пу-

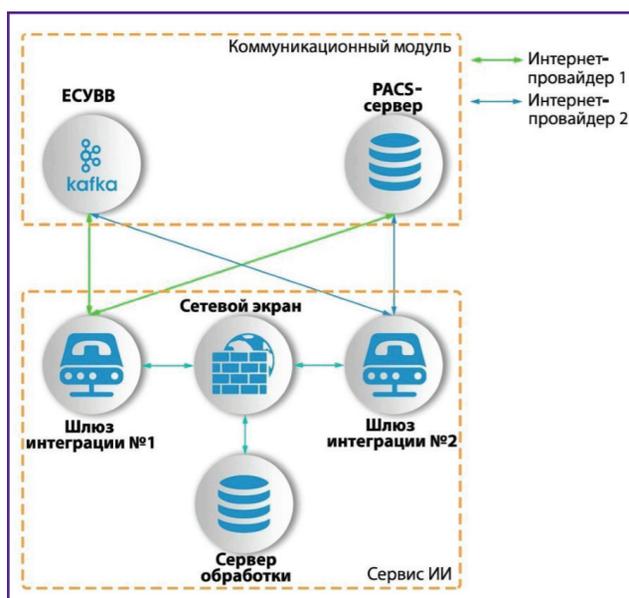


Рис. 9. Аппаратная схема интеграции с ЕРИС ЕМИАС

бликует задания на обработку исследований для всех подключенных к ней абонентов. Для того чтобы сервис мог обрабатывать только предназначенные для него актуальные исследования (давность которых не превышает 6,5 мин к моменту обработки ИИ — таково условие эксперимента), разработан модуль фильтрации. Программный модуль читает все сообщения из ЕСУВВ, выполняет их синтаксический анализ, фиксируя идентификаторы исследований: во-первых, не являющихся устаревшими, во-вторых, предназначенных для сервиса с учетом модальности (типа диагностического оборудования) и возрастной группы пациента.

После синтаксического анализа по идентификатору исследования программно-аппаратный шлюз запрашивает и скачивает исследования из PACS-сервера ЕРИС ЕМИАС.

Полученное исследование формата DICOM может содержать в себе серию снимков разных проекций (прямая или боковая) как в негативе, так и в позитиве. Для того чтобы на обработку ИИ попал правильный снимок, необходимо провести ряд последовательных проверок:

анализ DICOM-тегов, характеризующих модальность изображения;

интеллектуальный анализ изображения с помощью вспомогательной нейронной сети, определяющей, насколько представленное изображение валидно для анализа (т.е. в случае сервиса «ФтизисБиоМед» — является ли представленное изображение рентгенографическим снимком легких в прямой проекции);

классификацию снимка по критерию негатив/позитив с применением детерминированного алгоритма и инверсию негативного снимка.

Позитивная рентгенограмма/флюорограмма грудной клетки в прямой проекции передается на обработку в ИИ, а также на обработку нейронной сетью, решающей задачу сегментации контуров легких. Далее, в случае если ИИ сделал заключение о наличии патологии, строятся контуры выявленных на снимке патологий. Результат построения контуров накладывается на исходное изображение. В противном случае сервис выдает заключение о том, что патологии на снимке не выявлены. Последней стадией обработки медицинского изображения является классификация патологических признаков. Для решения этой задачи используется вспомогательная нейронная сеть (классификатор), на вход которой подаются выявленные в процессе локализации патологические области. Эти области вычлняются из изображения и последовательно подаются на анализ вспомогательной нейронной сетью.

Классификатор решает задачу классификации по 9 патологическим признакам: плевральный выпот, пнев-

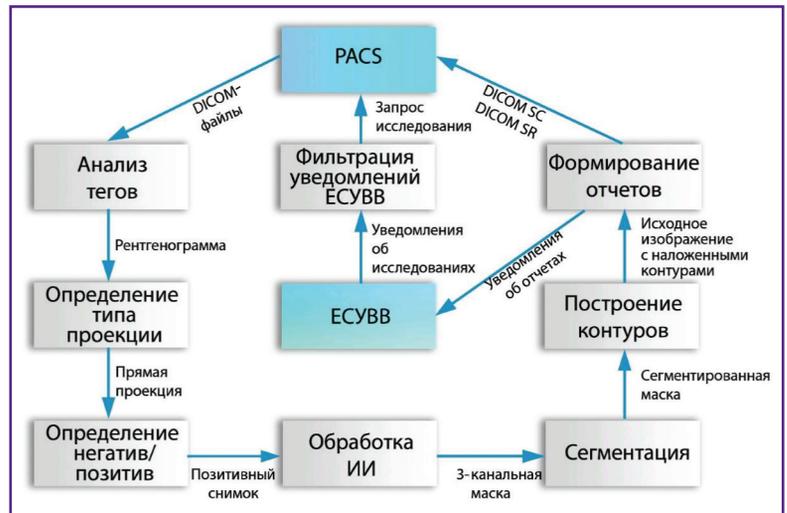


Рис. 10. Функциональная схема интеграции с ЕРИС ЕМИАС

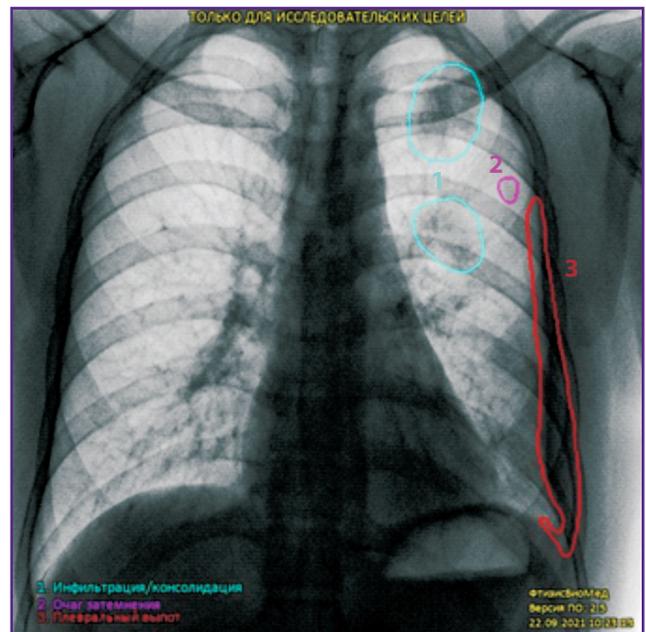


Рис. 11. Отчет DICOM SC

моторакс, ателектаз, очаг затемнения, инфильтрация/консолидация, диссеминация, полость, кальцинат/кальцинированная тень в легких, нарушение целостности кортикального слоя.

По результатам обработки сервисом исследования в радиологическую информационную систему отправляются отчет DICOM SC (рис. 11), отчет DICOM SR (рис. 12) и два сообщения в ЕСУВВ (Kafka), уведомляющих об отправке каждого из DICOM-отчетов.

На финальном этапе интеграции в ЕРИС ЕМИАС был проведен калибровочный тест, по результатам которого принято решение о допуске сервиса на платформу. Значения показателей метрик диагностической ценности [9] сервиса по результатам калибровочного тестирования представлены в табл. 1.

Статус	Частично - Проверка обращена
Модальность	ФЛГ
Область исследования	CHEST
Идентификатор исследования	1.871.3.1859213745.54904.17703.1983172798.2474477225.1
Дата и время формирования заключения ИИ-Сервисом	15.09.2021 15:46
Предупреждение	Только для исследовательских целей
Предупреждение	Заключение получено при поддержке алгоритма искусственного интеллекта
Наименование сервиса	FBM
Версия сервиса	2.5
Назначение сервиса	Сервис определяет наличие патологических изменений
Технические данные	Количество изображений: 1
Данные по качеству серий	Обработана 1 серия
Описание	
Область целевого исследуемого органа: органы грудной клетки (лёгкие)	
Проекция: прямая передняя	
Заключение	
Вероятность наличия патологических изменений: 0.45	
Выявленные патологические области:	
1. Класс патологии не определен	
2. Плевральный выпот	
Количество выявленных патологических изменений: 2	
Руководство пользователя	
ФтизисБиоМед - автоматическая система, выявляющая области предполагаемых патологических признаков на рентгенограммах.	
Области локализуются цветными контурами на изображении. Каждому цвету контура соответствует одно из заключений по одному патологическому признаку, одному из двух возможных патологических признаков или о том, что класс патологии не определен.	
Перечень заключений по каждому выявленному контуру приведен в левом нижнем углу дополнительной серии.	
Диагностируемые патологические признаки:	
-Плевральный выпот	

Рис. 12. Отчет DICOM SR

В результате нескольких этапных тестирований на эталонном наборе данных достигнуто значение AUCROC — 0,965, чувствительности — 0,92 (95% CI: 0,87–0,97), специфичности — 0,94 (95% CI: 0,89–0,99). Этапность тестирований была обусловлена процедурами эксперимента, выпуском новых версий ИИ-сервиса «ФтизисБиоМед».

Далее ИИ-сервис «ФтизисБиоМед» был допущен

Таблица 1

Результаты калибровочного тестирования

Метрика	Полученное значение на выборке данных	
	Индекс Юдена	max NPV
AUCROC	0,965	
Точность, 95% CI	0,93 (0,89–0,97)	0,93 (0,89–0,97)
Чувствительность, 95% CI	0,92 (0,87–0,97)	0,92 (0,87–0,97)
Специфичность, 95% CI	0,94 (0,89–0,99)	0,94 (0,89–0,99)
Удельный вес ложноотрицательных результатов, %	0,08	0,08
Удельный вес ложноположительных результатов, %	0,06	0,06
Оптимальный порог	0,65	0,65

к работе на потоке исследований. Для автоматизированного анализа на сервис маршрутизировали результаты профилактических рентгенологических исследований органов грудной клетки, выполненных в медицинских организациях Департамента здравоохранения Москвы. В условиях реальной клинической работы ИИ-сервис «ФтизисБиоМед» показал свое качество и надежность, что подтверждается процедурами технологического мониторинга (предусмотренными методикой эксперимента) и включением сервиса в тройку лидеров ежемесячного рейтинга участников эксперимента (mosmed.ai/ai).

Клинические испытания ИИ-сервиса «ФтизисБиоМед»

Гипотезы исследования

Проверяемая гипотеза сформулирована в соответствии с форматом, описанным D.A. Korevaar с соавт. [13]. Она обозначена как (с учетом 95% CI) H_0 : {чувствительность < 0,93 и/или специфичность < 0,70}.

Материалы и методы

Проведено ретроспективное диагностическое исследование. Описание дизайна и результатов выполнено в соответствии с методологией STARD 2015 [14].

Методология испытаний [15] сформирована в том

числе на основе ГОСТ Р 59921.1—2022 «Системы искусственного интеллекта в клинической медицине. Часть 1. Клиническая оценка».

Индекс-тест проводили на программном обеспечении «Программа автоматизированного анализа цифровых рентгенограмм органов грудной клетки/флюорограмм по ТУ 62.01.29-001-96876180-2019» производства ООО «ФтизисБиоМед», Россия (далее — ПО «ФтизисБиоМед»). Это программное обеспечение предназначено для использования квалифицированными сотрудниками медицинских организаций. Функциональное назначение позволяет позиционировать его как инструмент поддержки принятия врачебных решений. Потенциально ПО «ФтизисБиоМед» при его клинической реализации может сократить время постановки диагноза врачом, позволит своевременно заметить признаки патологий, что в результате даст дополнительное время врачу (сотруднику медицинской организации) для лечения и реабилитации пациента.

Применение ПО «ФтизисБиоМед» показано для анализа цифровых флюорографических снимков (рентгенограмм) легких в прямой передней проекции и выявления патологий.

Осуществляется автоматическое выявление рентгенологических признаков следующих патологий:

1. Патологии 1-го рода — условно «опасные». Признаки: инфильтрация (фокус) — свыше 1,5 см; полость; пневмоторакс; гидроторакс; очаг; патологические изменения корней лёгких; уровень жидкости; очаги.

2. Патологии 2-го рода — условно «неопасные». Признаки: интерстициальные изменения в легочной паренхиме; цирроз; фиброторакс; изменения плевры; кальцинаты/обызвествления; диафрагмальная грыжа; изменения в костях; цепочки металлических швов; инородные тела; участок повышенной прозрачности (не полость); ателектаз; изменения в органах средостения.

Абсолютные и относительные противопоказания к применению ПО отсутствуют.

Референс-тест представляет собой верифицированный эталонный набор данных с разметкой [16].

Для формирования первоначального набора данных использовали результаты флюорографических исследований из ЕРИС ЕМИАС в соответствии с действующим законодательством. Все результаты исследований подвергнуты процедуре деперсонализации в установленном порядке.

Первоначальный отбор из ЕРИС ЕМИАС проводили по следующим параметрам: наименование процедуры, вид диагностического устройства, вид медицинской организации, дата снимка, возраст пациента. В результате отбора получено около 400 тыс. идентификаторов исследований.

Для отобранных исследований получали заключения врачей-рентгенологов из ЕРИС ЕМИАС. Эти заключения впоследствии анализировали по ключевым словам. В анализ включали только флюорограммы

пациентов старше 18 лет. В результате отбора по ключевым словам получили набор данных из 5000 изображений для последующей верификации и использования в ходе исследования диагностической точности ПО «ФтизисБиоМед».

Набор данных включает в себя результаты флюорографии: диагностические изображения без признаков патологии (норма) и изображения со всеми рентгенологическими признаками патологий, перечисленных выше.

Проводили визуальный анализ отобранных результатов флюорографических исследований, в ходе которого исключали: спорные изображения, не содержащие выраженных признаков патологий, перечисленных в документации производителя, и которые нельзя достоверно отнести к категории «норма»; изображения с недостаточным техническим качеством (низкая контрастность и т.п.).

В результате отбора получено 1536 изображений, содержащих только фронтальные снимки области грудной клетки. Поскольку тестируемое ПО, согласно документации, не обрабатывает боковые проекции, при формировании верифицированного набора данных такие изображения принудительно удаляли.

В итоговый набор данных вошли результаты исследований, выполненные на диагностических аппаратах 18 различных моделей (11 производителей), 4 из которых относятся к категории флюорографов (n=670), остальные 14 — к категории рентген-диагностических аппаратов (n=866). Далее выполнена разметка набора данных.

В разметке и верификации набора данных участвовали четыре врача-рентгенолога со стажем работы от 9 до 40 лет. Разметку набора данных выполняли два врача-рентгенолога, которые анализировали фронтальные проекции. При просмотре сопоставляли с каждым снимком один или несколько рентгенологических признаков патологий. Если изображение не содержало признаков патологии, его относили к категории «норма».

По результатам анализа установлено, что 393 изображения из 1536 не содержат признаков патологии. Другие 1143 изображения содержат отметки одного и более патологического признака (всего 3304 отметки признаков на всей группе изображений с признаками патологий). Из них 291 изображение содержит только по одной отметке и, таким образом, эти изображения могут быть точно отнесены к определенному классу признаков патологий.

Верификацию набора данных проводили два врача-рентгенолога, которые имели ученые степени и сертификаты о прохождении курсов надлежащей клинической практики.

Статистический анализ

Расчеты необходимого объема выборки и статистической значимости выполнены согласно стандартным методикам [17].

Проведен расчет метрик диагностической точности (чувствительности и специфичности) по методике, рекомендованной для клинических испытаний программного обеспечения на основе ИИ-технологий [17].

Определение точности работы ПО проводили в два этапа: на первом оценивали интегральную характеристику системы без разделения на рентгенологические признаки, на втором — точность для отдельных рентгенологических признаков.

Статистическая значимость результатов определялась 95% CI.

Результаты

Верифицированный набор данных (n=1536) обработали с помощью ПО «ФтизисБиоМед». Среднее

время обработки одного изображения не превышало 5 с. Для 1517 изображений (98,8%) получены текстовые отчеты, содержащие сведения о вероятности ($p \in [0, 1]$) наличия патологии в каждом конкретном снимке.

Для 19 изображений текстовые отчеты получены не были. В связи с этим предприняли попытку обработать их отдельно. При повторной обработке текстовые отчеты также не получили, поэтому данные изображения исключили из списка верифицированного набора данных и в дальнейшем не анализировали.

Таким образом, уточненный верифицированный набор данных содержит 1517 изображений. Из них 391 (25,8%) не содержит признаков патологии. Распределение рентгенологических признаков представлено в табл. 2.

Соответственно, 1126 изображений (74,2%) содержат 3259 отметок о наличии патологий. По одной патологической отметке содержится только в 289 изображениях; 837 изображений содержат более чем по одной отметке на изображение (2970 отметок, в среднем по 3 или 4 отметки на изображение).

Определено число истинно-положительных, истинно-отрицательных, ложноположительных, ложноотрицательных результатов, построена и проанализирована четырехпольная таблица, а также характеристическая кривая (ROC-кривая, рис. 13), при этом установлено оптимальное значение порога активации (cut-off) — $T=0,1$ (оптимальным считали значение, на фоне которого достигнута максимальная чувствительность при специфичности не ниже 0,3). На первом этапе выполнена интегральная оценка показателей точности, на втором — по отдельным рентгенологическим признакам.

Оценка интегральных показателей чувствительности и специфичности. При заданном пороге получили следующие значения показателей точности: чувствительность — 0,876 (95% CI: 0,85–0,89); специфичность — 0,75 (95% CI: 0,71–0,78).

Все 94 ложноположительных и 139 ложноотрицательных случаев передали для повторной экспертной оценки врачам-рентгенологам. Повторную разметку

Таблица 2

Тестовый набор данных изображений с признаками патологии (n=1126), успешно обработанных ПО «ФтизисБиоМед»

Рентгенологический признак	Количество изображений, содержащих только этот признак	Количество изображений, содержащих этот признак наряду с другими
Тотальное (субтотальное) затемнение (уплотнение) легочного поля, в том числе:	2	80
инфильтрация	1	48
уровень жидкости	1	32
Ограниченное (долевое, сегментарное, субсегментарное, полисегментарное) затемнение легочного поля (инфильтрация), в том числе:	18	348
инфильтрация	15	186
уровень жидкости	1	126
фиброторакс	1	20
цирроз	1	16
Круглая тень легочной локализации	33	87
Кольцевидная тень, в том числе:	20	110
диафрагмальная грыжа	14	44
полость	6	66
Воздух в плевральной полости	2	74
Очаги	17	214
Очаговые диссеминации	1	79
Усиление и деформация легочного рисунка (интерстициальные изменения в легочной паренхиме, пневмофиброз, пневмосклероз)	76	755
Утолщение листков плевры	18	300
Расширение (\pm деформация) тени корня/корней легких или смещение тени корня/корней легких	4	283
Кальцинаты/обызвествления легочной и внелегочной локализации	17	147
Травматические изменения костных структур	24	79
Инородные тела	34	216
Изменение тени средостения	22	438
Участок повышенной прозрачности	1	49
Итого отметок	289	3259

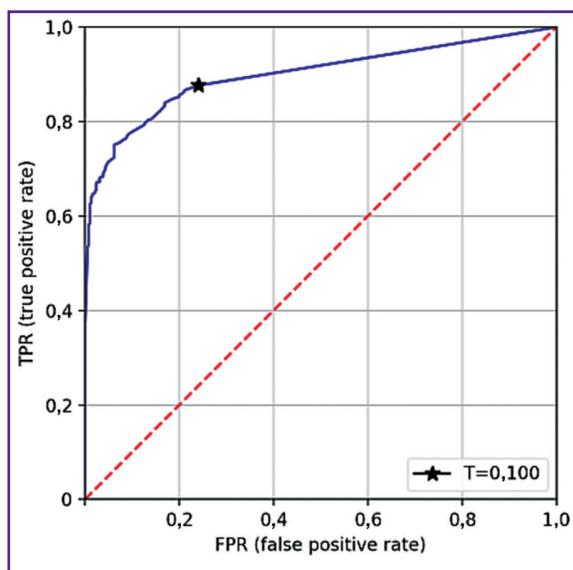


Рис. 13. ROC-кривая при анализе набора данных без разбиения на признаки (состояние порога активации 0,1)

проводили «вслепую»: не сообщали врачам никакой дополнительной информации о ранее выполненной разметке этих изображений и никаких сведений о результатах работы ПО «ФтизисБиоМед».

Экспертную оценку спорных случаев выполняли члены рабочей группы, имеющие ученую степень в области медицины и соответствующую специализацию. Эксперты подтвердили, что все случаи, не распознан-

ные программным обеспечением, действительно корректно отнесены к классам «норма» или «патология» при первоначальной разметке. Таким образом, все ошибки распознавания отнесли на счет некорректной работы ПО «ФтизисБиоМед».

Оценка показателей по отдельным рентгенологическим признакам. Поскольку для подтверждения большинства признаков количества уникальных изображений недостаточно, применяли итерационный подход:

1. Выбирали рентгенологические признаки, по которым чувствительность по результатам тестирования не достигает порогового значения.

2. Среди них выбирали признак, содержащий наибольшее число уникальных изображений (на первой итерации это был признак «Усиление и деформация легочного рисунка» ($n=76$)).

3. Определяли чувствительность для выбранной группы изображений. Показатель составил 0,81. На основании этого признали, что чувствительность диагностики названного выше признака недостаточна.

4. Исключили из анализа 76 изображений, уникальных по данному признаку.

5. Исключили этот признак из анализа. При этом изображения, ранее включавшие данный признак наряду с каким-либо другим, теперь считали включающими только другой признак.

6. После исключения изображений и признака, отобранного на шаге 2, возвращались к шагу 1.

7. Действия 1–6 повторяли до тех пор, пока интегральная чувствительность по оставшимся признакам

Таблица 3

Признаки с интегральной чувствительностью $\geq 0,93$

Рентгенологический признак	Количество изображений	Чувствительность P1 по признаку	95% CI P1	Статистическая значимость α при $P_0=0,93$
Тотальное (субтотальное) затемнение (уплотнение) легочного поля, в том числе:				
инфильтрация	48	1,0	0,997–1,0	0
уровень жидкости	32	1,0	0,996–1,0	0
Ограниченное затемнение легочного поля, в том числе:				
инфильтрация	186	0,989	0,97–1,0	10^{-14}
фиброторакс	20	1	0,996–1,0	0
цирроз	16	1	0,995–1,0	0
уровень жидкости	13	0,923	0,78–1,0	—
Воздух в плевральной полости	74	0,986	0,959–1,0	$4,4 \times 10^{-5}$
Очаги	214	0,958	0,93–0,98	0,04
Очаговые диссеминации	79	1	0,998–1,0	0
Утолщение листков плевры	300	0,97	0,95–0,98	$4,9 \times 10^{-5}$
Кальцинаты/обызвествления легочной и внелегочной локализации	147	0,986	0,97–1,0	$7,7 \times 10^{-9}$
Травматические изменения костных структур	41	0,90	0,80–0,99	—
Кольцевидная тень (полость)	13	0,923	0,78–1,0	—

не достигла заявленной производителем чувствительности, равной 0,93. Всего исключили 7 признаков (кольцевидная тень (диафрагмальная грыжа); усиление и деформация легочного рисунка (интерстициальные изменения в легочной паренхиме, пневмофиброз, пневмосклероз); расширение (деформация) тени корня/корней легких или смещение тени корня/корней легких; инородные тела; изменение тени средостения; круглая тень легочной локализации; участок повышенной прозрачности) и 371 ложноотрицательное исследование.

В табл. 3 представлены признаки с интегральной чувствительностью 0,93.

Средняя чувствительность по оставшейся выборке (1183 отметки) составила 0,975 (95% CI: 0,966–0,984). Соответственно, специфичность тестируемого ПО «ФтизисБиоМед» — 0,75 с уровнем значимости 0,023.

Рентгенологические признаки, представленные в табл. 3, и ассоциированные с ними патологии диагностируются с высокой чувствительностью. Статистическая значимость составила менее 0,05, что говорит о высокой достоверности оценки.

Таким образом, нулевая гипотеза H_0 : {чувствительность $< 0,93$ и/или специфичность $< 0,70$ }:

1) отклонена для следующих признаков: инфильтрация — при частичном и тотальном затемнении; гидроторакс (уровень жидкости) — при частичном и при тотальном затемнении; цирроз; фиброторакс; ателектаз — при частичном и тотальном затемнении; очаг; полость; пневмоторакс; очаги (диссеминация); изменения плевры; кальцинаты; изменения костных структур;

2) принята для признаков следующих признаков: инфильтрация (фокус); диафрагмальная грыжа; интерстициальные изменения в легочной паренхиме; патологические изменения корней легких; цепочки металлических швов, инородные тела; изменения в органах средостения; участок повышенной прозрачности (не полость).

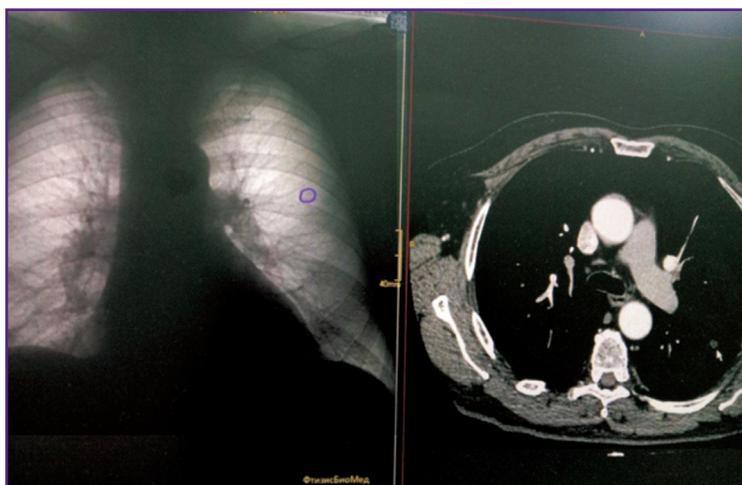


Рис. 14. Обработка профилактической флюорографии сервисом «ФтизисБиоМед» (слева) и результаты КТ-исследования (справа)

Частные примеры анализа

Для прояснения аспектов работы сформулируем несколько тезисов.

1. ИИ не руководствуется вспомогательными факторами, кроме тех, которые ему представлены, и даже из них он оперирует только теми, которые способен интерпретировать. Иными словами, если ИИ предназначен для анализа изображений, то только это он и будет делать. По умолчанию возраст, пол, результаты анализа крови, история болезни пациента и другие факторы в расчет не принимаются. При использовании более совершенных моделей или же алгоритмических надстроек некоторые из этих факторов могут быть учтены. Однако за универсальность и широкий охват анализируемых признаков зачастую приходится платить точностью системы. Таким же образом приходится платить, когда ИИ приспособливают для решения более одной задачи, например задач локализации и классификации.

2. Мнение ИИ — это математический результат, поэтому оно зачастую сильно подвержено искажению в зависимости от того, при каких условиях анализ был произведен. Большое значение имеют исходные данные. Искаженные, зашумленные изображения или же частичные изображения анализируемых органов приводят к «мусорным» результатам инференции. Например, огромное значение для ИИ-анализа рентгенографических изображений имеет правильная укладка пациента. Кроме того, даже проанализировав снимки одного и того же пациента, полученные разными рентген-аппаратами или даже одним рентген-аппаратом, но с измененными настройками, ИИ может дать несколько различающиеся заключения.

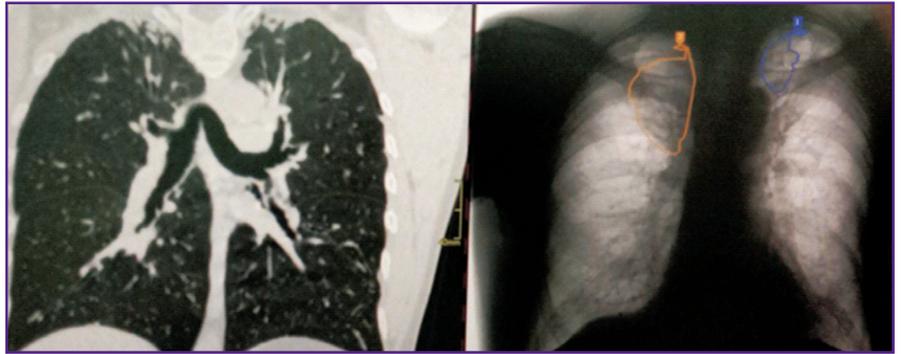
3. Сколько бы раз ИИ-сервис «ФтизисБиоМед» ни анализировал одно и то же исследование в одних и тех же условиях, он будет приходить к одинаковому заключению внутри одной итерации обучения.

С учетом вышесказанного предлагаем обратить внимание на несколько результатов обработки ИИ исследований реальных пациентов.

Пример 1. На профилактической флюорографии (рис. 14) ИИ «ФтизисБиоМед» выявил округлое новообразование диаметром 8 мм или туберкулезный очаг. Дальнейшее дообследование в виде КТ с внутривенным контрастированием показало отсутствие туберкулеза и исключило злокачественность образования. Наиболее вероятный диагноз по результатам КТ — «доброкачественное врожденное образование — гамартома легкого» (нет патологической опухолевой сосудистой сети, питающей образование).

Пример 2. ИИ «ФтизисБиоМед» на флюорографии (рис. 15) маркирует измене-

Рис. 15. Обработка потоковой флюорографии сервисом «ФтизисБиоМед» (справа) и результаты КТ-исследования (слева)



ния на вершинах обоих легких. Для исключения активного туберкулеза врач обратился к флюорографическому архиву: выявлено отсутствие динамики. На КТ (ранее) обнаружены массивные плевроапикальные наслоения с двух сторон. Активного туберкулеза нет.

Пример 3. ИИ «ФтизисБиоМед» на флюорографии (рис. 16) выявил затемнение бокового плеврального синуса слева. Необходимо в первую очередь исключить наличие воспалительной жидкости в плевральной полости. У данного пациента женского пола была операция с частичной резекцией лёгкого, что вызвало поднятие диафрагмы слева, укорочение и затемнение синуса слева. Заметно сместились срединные структуры (сердце и сосуды) влево вследствие объемного уменьшения легкого после операции. С учетом операции в анамнезе и стабильной рентгенологической картины на протяжении трех месяцев можно заключить, что воспалительная жидкость в плевральном синусе слева исключена.

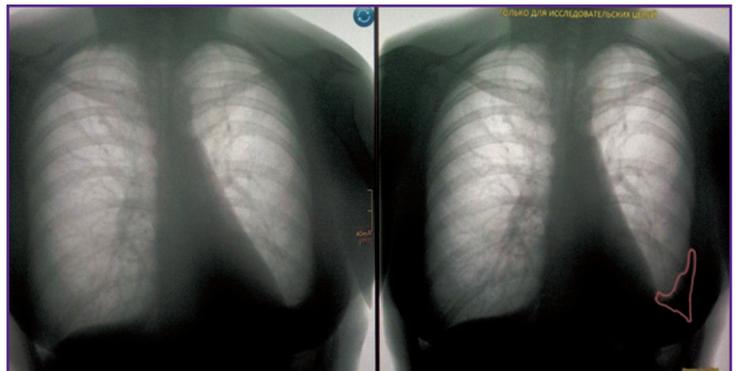


Рис. 16. Обработка потоковой флюорографии сервисом «ФтизисБиоМед» (справа), исходное изображение (слева)

Пример 4. ИИ «ФтизисБиоМед» на флюорографии (рис. 17) маркирует изменение в виде врожденного шейного ребра слева (чаще — скомпенсированная патология, однако изредка могут быть неврологические симптомы).

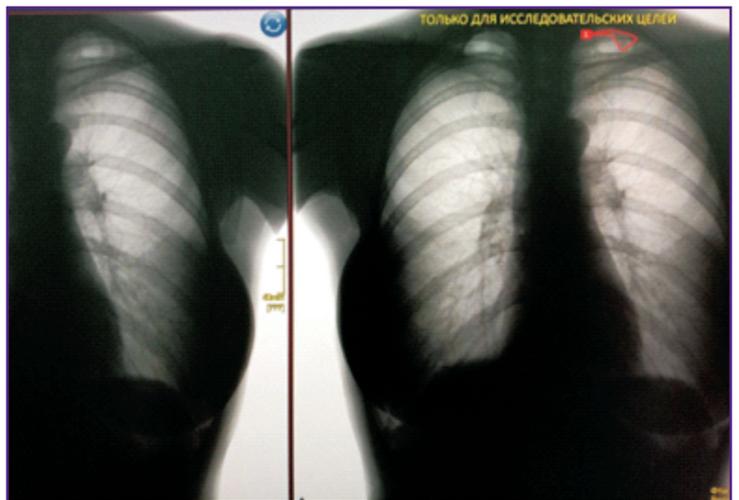


Рис. 17. Обнаружение с помощью искусственного интеллекта «ФтизисБиоМед» врожденного шейного ребра (справа) и исходное изображение (слева)

Пример 5. ИИ «ФтизисБиоМед» на флюорографии (рис. 18) выявил последствия перенесенной пневмонии в виде линейной локальной деформации легочного рисунка в базальном отделе левого легкого. По базе данных эти рубцовые изменения — без динамики в течение 6 мес.

Пример 6. ИИ «ФтизисБиоМед» на флюорографии (рис. 19) маркирует участок линейной деформации легочного рисунка и предлагает дифференцировать дисковидный ателектаз и постпневмонический фиброз. КТ из архива показывает степень поражения легких КТЗ при COVID-19 за два года до настояще-

го времени. Вывод: на флюорографии постпневмонический фиброз (хотя возможно даже неполное расправление участка легкого с формированием фиброателектаза после массивного воспаления с формированием грубых рубцовых изменений — фиброза в легких).

Пример 7. ИИ «ФтизисБиоМед» выявил значительные остаточные изменения после перенесенного



Рис. 18. Обнаружение с помощью искусственного интеллекта «ФтизисБиоМед» последствий пневмонии (справа) и исходное изображение (слева)

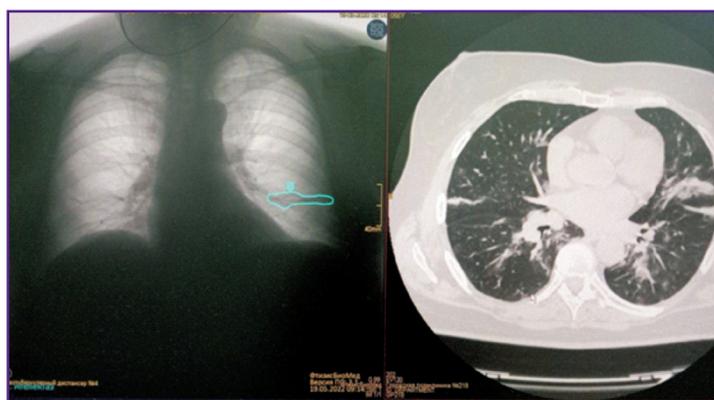


Рис. 19. Обнаружение с помощью искусственного интеллекта «ФтизисБиоМед» дисковидного ателектаза (слева) и результаты КТ-исследования (справа)

туберкулеза (рис. 20). Дальнейшее обследование на КТ подтверждает наличие изменений и исключает активность инфекции.

Результаты настоящего исследования показали, что, поскольку диагностические метрики врачей могут быть сильно разнородными, обусловленными их специализацией и опытом, использование заключения отдельного врача как эталонного мнения при оценке диагностических метрик ИИ-сервиса нецелесообразно. Объективная оценка требует коллективного мнения.

Природа ошибок врача и ИИ различна, при этом метрики сопоставимы. С одной стороны, ошибки ИИ зачастую очевидны для врача, с другой — ИИ-сервис может сфокусировать внимание врача на неочевидной патологической области, которая потенциально могла быть им пропущена. Объединение врача и ИМИ в систему способно породить синергетический эффект, выраженный в повышении диагностических метрик системы «врач + ИМИ» относительно метрик врача и ИМИ в отдельности.

Исследования, проведенные как в рамках Московского эксперимента (на 2023 г. обработано бо-

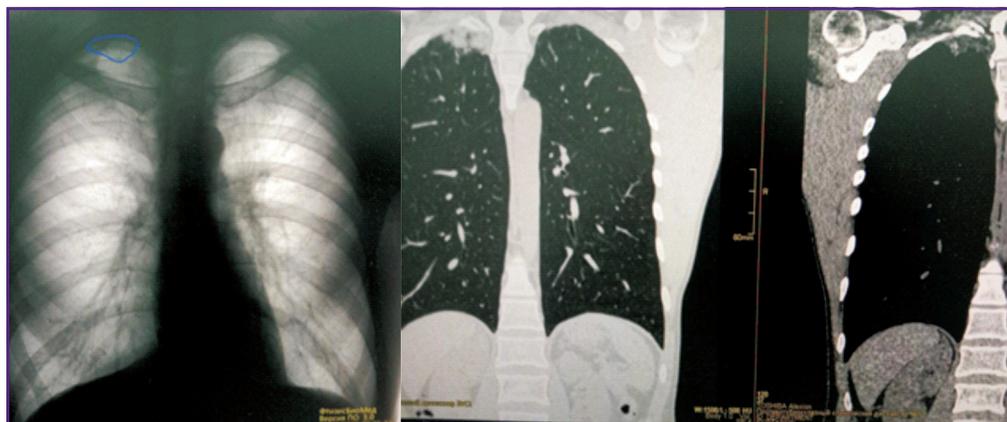


Рис. 20. Обнаружение с помощью искусственного интеллекта «ФтизисБиоМед» остаточных изменений (слева) и результаты КТ-исследования (справа)

лее 8 млн медицинских изображений), так и международными экспертами в последние годы наглядно демонстрируют, что качество анализа ИИ медицинских изображений вплотную приближается к качеству работы опытных врачей-диагностов, и это позволяет смотреть в будущее технологий ИМИ со сдержанным оптимизмом. В результате многолетнего периода исследований, разработок, тестирований выявлена и доказана на практике жизнеспособность СППВР на базе технологий ИИ.

Для развития и увеличения эффективности технологий ИМИ разработаны и проверены на практике методологии обучения и испытаний продуктов ИИ, а также внедрение ИМИ в клинические процессы медицинских учреждений.

Заключение

В рамках Московского эксперимента ИИ-сервис «ФтизисБиоМед» прошел этапные процедуры тестирования на эталонных наборах данных. Достигнуто высокое значение площади под характеристической ROC-кривой — 0,965, чувствительности — 0,92 (95% CI: 0,87–0,97) и специфичности — 0,94 (95% CI: 0,89–0,99). В условиях реальной клинической практики (обработка потока результатов профилактических рентгенологических исследований органов грудной клетки, выполненных в медицинских организациях Департамента здравоохранения Москвы) «ФтизисБиоМед» показал свои качество и надежность, что подтверждается процедурами технологического мониторинга (предусмотренными методикой эксперимента).

Клинические испытания медицинского ИИ-сервиса «ФтизисБиоМед» проведены на верифицированном наборе данных (n=1536), подготовленном с учетом эпидемиологических показателей распространенности основных заболеваний органов грудной клетки. В процессе тестирования установлено, что диагностическая точность «ФтизисБиоМед» не равнозначна для разных патологических признаков, поэтому несколько таковых были исключены из возможностей ИИ-сервиса. В результате клинических испытаний определена средняя чувствительность медицинского ИИ-сервиса «ФтизисБиоМед» — 0,975 (95% CI: 0,966–0,984).

«ФтизисБиоМед» зарегистрирован в качестве медицинского изделия (Регистрационное удостоверение на медицинское изделие №РЗН 2022/17406 от 31 мая 2022 г.) и активно применяется в Российской Федерации как диагностический инструмент, позволяющий снизить нагрузку на врача-рентгенолога и ускорить процесс получения медицинского заключения.

Благодарности. Авторы выражают благодарность академику РАН А.Г. Чучалину, члену-корреспонденту РАН А.Э. Эргешову, директору ЦНИИ туберкулеза,

д.м.н. Р.Б. Амансахедову, к.т.н. А.В. Мальцеву, к.т.н. Н.Г. Андрианову, А.П. Алексееву, С.И. Коновалову, к.м.н. А.Ф. Алиеву за полезные дискуссии и предоставленные клинические результаты.

Финансирование исследования. Наибольшая часть финансирования разработки ИИ-сервиса «ФтизисБиоМед» была получена от Департамента здравоохранения Москвы. Важной составляющей была и финансовая (грантовая) помощь Фонда содействия инновациям. Постоянной на протяжении семи лет была поддержка АО «РК "Вектор"».

Конфликт интересов. Авторы заявляют об отсутствии конфликтов интересов.

Литература/References

1. Федеральная служба государственной статистики (Росстат). *Здравоохранение*. URL: <https://rosstat.gov.ru/folder/13721>.
2. Federal State Statistics Service (Rosstat). *Zdravookhraneniye* [Healthcare]. URL: <https://rosstat.gov.ru/folder/13721>.
3. Гогоберидзе Ю.Т., Классен В.И., Натензон М.Я., Просвиркин И.А., Сафин А.А. Особенности имплементации систем искусственного интеллекта в задаче анализа двумерных радиологических изображений. В кн.: *Математические методы распознавания образов*. М: Российская академия наук; 2019; с. 307–308.
4. Gogoberidze Yu.T., Klassen V.I., Natenzon M.Ya., Prosvirkin I.A., Safin A.A. Osobennosti implementatsii sistem iskusstvennogo intellekta v zadache analiza dvukhmernykh radiologicheskikh izobrazheniy. V kn.: *Matematicheskie metody raspoznavaniya obrazov*. [Features of the implementation of artificial intelligence systems in the problem of analyzing two-dimensional radiological images. In: Mathematical methods of pattern recognition]. Moscow: Rossiyskaya akademiya nauk; 2019; p. 307–308.
5. Huang G., Liu., van der Maaten L., Weinberger K.Q. Densely connected convolutional networks. *arXiv*; 2018, <https://doi.org/10.48550/arXiv.1608.06993>.
6. Hwang E.J., Park S., Jin K.N., Kim J.I., Choi S.Y., Lee J.H., Goo J.M., Aum J., Yim J.J., Cohen J.G., Ferretti G.R., Park C.M.; DLAD Development and Evaluation Group. Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2019; 2(3): e191095, <https://doi.org/10.1001/jamanetworkopen.2019.1095>.
7. Морозов С.П., Владзимирский А.В., Кляшторный В.Г., Андрейченко А.Е., Кульберг Н.С., Гомболевский В.А., Сергунова К.А. *Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика)*. Серия «Лучшие практики лучевой и инструментальной диагностики». Вып. 57. М; 2019.
8. Morozov S.P., Vladzimirskiy A.V., Klyashtorny V.G., Andreychenko A.E., Kul'berg N.S., Gombolevskiy V.A., Sergunova K.A. *Klinicheskie ispytaniya programmnogo obespecheniya na osnove intellektual'nykh tekhnologiy (luchevaya diagnostika)*. Seriya "Luchshie praktiki luhevoy I instrumental'noy diagnostiki". Vyp. 57 [Clinical trials of software based on intelligent technologies (radiology). Series "Best practices of radiological and instrumental diagnostics". Issue 57]. Moscow; 2019; 51 p.
9. Rebrova O.Yu., Gusev A.V. Sample size calculation for

clinical trials of medical decision support systems with binary outcome. *Sovremennyye tekhnologii v medicine* 2022; 14(3): 6, <https://doi.org/10.17691/stm2022.14.3.01>.

7. Sun X., Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett* 2014; 21(11): 1389–1393, <https://doi.org/10.1109/lsp.2014.2337313>.

8. Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы». *Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения Москвы*. URL: <https://mosmed.ai>.

Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Department of Health. *Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения Москвы* [An experiment on the use of innovative computer vision technologies for medical image analysis and subsequent applicability in the healthcare system of Moscow]. URL: <https://mosmed.ai>.

9. Приказ Департамента здравоохранения города Москвы от 19.02.2020 №142 «Об утверждении Порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы».

Приказ Департамента здравоохранения города Москвы от 19.02.2020 No.142 "Об утверждении Порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы" [Order of the Moscow Department of Health dated February 19, 2020 No.142 "On approval of the Procedure and conditions for conducting. URL: <https://mosmed.ai>.

10. Laval medical systems. *Единый Радиологический Информационный Сервис*. URL: <http://lvmed.ru/eris/>.

Laval medical systems. *Единый Радиологический Информационный Сервис* [Unified Radiological Information Service]. URL: <http://lvmed.ru/eris/>.

11. DICOM® — *Digital Imaging and Communications in Medicine*. Current Edition. URL: <https://www.dicomstandard.org/current/>.

12. Apache Software Foundation. *Apache Kafka*. URL: <https://kafka.apache.org/>.

13. Korevaar D.A., Gopalakrishna G., Cohen J.F., Bossuyt P.M. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses. *Diagn Progn Res* 2019; 3: 22, <https://doi.org/10.1186/s41512-019-0069-2>.

14. Bossuyt P.M., Reitsma J.B., Bruns D.E., Gatsonis C.A., Glasziou P.P., Irwig L., Lijmer J.G., Moher D., Rennie D., de Vet H.C., Kressel H.Y., Rifai N., Golub R.M., Altman D.G., Hooft L., Korevaar D.A., Cohen J.F.; STARD Group. STARD 2015: an updated list of essential items for Reporting Diagnostic Accuracy Studies. *Radiology* 2015; 277(3): 826–832, <https://doi.org/10.1148/radiol.2015151516>.

15. Зинченко В.В., Ахмад Е.С., Заянчковский С.Ю., Шарова Д.Е., Владзимирский А.В., Морозов С.П. Методология проведения клинических испытаний медицинского программного обеспечения с искусственным интеллектом. *Медицинская физика* 2022; 1: 32.

Zinchenko V.V., Akhmad E.S., Zayunchkovsky S.Yu., Sharova D.E., Vladzimirsky A.V., Morozov S.P. Methodology for conducting clinical trials of medical software with artificial intelligence. *Медицинская физика* 2022; 1: 32.

16. Морозов С.П., Владзимирский А.В., Андрейченко А.Е., Ахмад Е.С., Блохин И.А., Гомболовский В.А., Зинченко В.В., Кульберг Н.С., Новик В.П., Павлов Н.А. *Регламент подготовки наборов данных с описанием подходов к формированию репрезентативной выборки данных. Часть 1: методические рекомендации. Серия «Лучшие практики лучевой и инструментальной диагностики»*. Вып. 103. М: ГБУЗ «НПКЦ ДиТ ДЗМ»; 2022.

Morozov S.P., Vladzimirsky A.V., Andreychenko A.E., Akhmad E.S., Blokhin I.A., Gombolevskiy V.A., Zinchenko V.V., Kul'berg N.S., Novik V.P., Pavlov N.A. *Reglament podgotovki naborov dannykh s opisaniem podkhodov k formirovaniyu reprezentativnoy vyborki dannykh. Chast' 1: metodicheskie rekomendatsii. Seriya "Luchshie praktiki luchevoy i instrumental'noy diagnostiki"*. Vyp. 103 [Regulations for the preparation of data sets with a description of approaches to the formation of a representative sample of data. Part 1: guidelines. Series "Best practices of radiological and instrumental diagnostics". Issue 103]. Moscow: GBUZ "NPKTs DiT DZM"; 2022.

17. Наркевич А.Н., Виноградов К.А. Методы определения минимально необходимого объема выборки в медицинских исследованиях. *Социальные аспекты здоровья населения* 2019; 65(6): 10.

Narkevich A.N., Vinogradov K.A. Methods for determining the minimum required sample size in medical research. *Social'nye aspekty zdorov'a naselenia* 2019; 65(6): 10.