



ЦЕНТР ДИАГНОСТИКИ
И ТЕЛЕМЕДИЦИНЫ

КОМПЬЮТЕРНОЕ ЗРЕНИЕ В ЛУЧЕВОЙ ДИАГНОСТИКЕ: ПЕРВЫЙ ЭТАП МОСКОВСКОГО ЭКСПЕРИМЕНТА

Монография под редакцией
Ю.А. Васильева, А.В. Владзимирского

Москва, 2023

Государственное бюджетное учреждение здравоохранения города Москвы
«Научно-практический клинический центр диагностики и телемедицинских технологий
Департамента здравоохранения города Москвы»

Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента

Монография. 2-е издание,
переработанное и дополненное

Издательские решения
По лицензии Ridero
2023

УДК 61
ББК 53
К63

Авторы: Васильев Ю. А., Владимирский А. В., Арзамасов К. М.,
Андрейченко А. Е., Гомболевский В. А., Кульберг Н. С., Омелянская О. В.,
Павлов Н. А., Решетников Р. В., Сергунова К. А., Шарова Д. Е.,
Шулькин И. М.

Шрифты предоставлены компанией «ПараТайп»

Компьютерное зрение в лучевой диагностике: первый этап
К63 **Московского эксперимента** : Монография. 2-е издание,
переработанное и дополненное / Ю. А. Васильев [и д. р.]. — [б. м.] :
Издательские решения, 2023. — 376 с.
ISBN 978-5-0059-3043-9

Систематизированы результаты первого этапа (2019–2020 гг.) крупнейшего в мире научного «Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы». Представлены методологии испытаний и применения ИИ в лучевой диагностике. Для организаторов здравоохранения, научных и медицинских работников, специалистов по разработке ИИ, сотрудников и учащихся учебных заведений.

УДК 61
ББК 53

12+ В соответствии с ФЗ от 29.12.2010 №436-ФЗ

ISBN 978-5-0059-3043-9

Оглавление

ПРЕДИСЛОВИЕ	8
РЕЦЕНЗЕНТЫ	11
БЛАГОДАРНОСТИ	12
ГЛОССАРИЙ	13
СПИСОК СОКРАЩЕНИЙ	15
ВВЕДЕНИЕ	17
Часть первая. КЛИНИЧЕСКАЯ МЕТОДОЛОГИЯ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВОЙ ДИАГНОСТИКЕ	24
Глава 1. СИСТЕМНЫЙ ПОДХОД К ПРОБЛЕМЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВОЙ ДИАГНОСТИКЕ	25
§1. Стратегия развития технологий искусственного интеллекта	25
§2. Научно-методологическое обеспечение реализации положений Национальной стратегии развития искусственного интеллекта в сфере лучевой диагностики ..	27
Глава 2. ТЕХНОЛОГИЧЕСКАЯ ОСНОВА ДЛЯ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВОЙ ДИАГНОСТИКЕ	46
Глава 3. КЛИНИЧЕСКИЙ КОНТЕКСТ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА	55
§1. Основной производственный процесс службы лучевой диагностики	57
§2. Запрос системы здравоохранения Российской Федерации	67
§3. Эпидемиологический и организационный аспекты	68
§4. Ожидания врачебного сообщества	72
§5. Клинический контекст	74
Глава 4. КЛАССИФИКАЦИЯ И МЕТОДОЛОГИЯ ПОДГОТОВКИ ЭТАЛОННЫХ НАБОРОВ ДАННЫХ	77
§1. Требования к эталонным наборам данных	83
§2. Классификация наборов данных	86
§3. Методология формирования наборов данных	90
§4. Процесс создания наборов данных	98
§5. Практическая реализация подготовки наборов данных	110

Глава 5. МЕТОДОЛОГИЯ КЛИНИЧЕСКИХ ИСПЫТАНИЙ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА	115
§1. Разработка методологии	115
§2. Общие положения	116
§3. Клинические испытания	123
§4. Метрики качества	142
§5. Оформление результатов	148
Часть вторая. ВНЕДРЕНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВУЮ ДИАГНОСТИКУ	150
Глава 6. ОРГАНИЗАЦИОННАЯ, ТЕХНИЧЕСКАЯ И МЕТОДОЛОГИЧЕСКАЯ ПОДГОТОВКА ЭКСПЕРИМЕНТА	153
§1. Основные организационные аспекты Эксперимента	154
§2. Технологические аспекты Эксперимента	160
§3. Обеспечение компетенций	164
§4. Специальные методологии и процедуры эксперимента	168
§5. Подготовка наборов данных в рамках Эксперимента ...	207
§6. Общий ход Эксперимента	210
Глава 7. НАУЧНАЯ МЕТОДОЛОГИЯ ЭКСПЕРИМЕНТА	223
Глава 8. ОЦЕНКА ДИАГНОСТИЧЕСКОЙ ТОЧНОСТИ СЕРВИСОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ	253
§1. Первый этап оценки	253
§2. Второй этап оценки	256
§3. Сравнение результатов первого и второго этапов оценки	260
§4. Третий этап оценки	264
Глава 9. ОЦЕНКА ЦЕЛЕСООБРАЗНОСТИ ПРИМЕНЕНИЯ ИИ- СЕРВИСОВ В ЛУЧЕВОЙ ДИАГНОСТИКЕ	268
§1. Социологический опрос пользователей (врачей- рентгенологов) ИИ-сервисов	268
§2. Результаты технологического мониторинга	275
§3. Сбор и анализ обратной связи по конкретным клиническим случаям от пользователей (врачей- рентгенологов)	286
§4. Оценка технологической готовности	292
§5. Хронометраж	293
§6. Обобщение результатов оценки целесообразности применения ИИ-сервисов	298
ЗАКЛЮЧЕНИЕ	301
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	305

ПРИЛОЖЕНИЯ	315
Приложение 1	315
Приложение 2	320
Приложение 3	322
Приложение 4	327
Приложение 5	328
Приложение 6	333
Приложение 7	335

Научные редакторы: Ю. А. Васильев, А. В. Владзимирский

Авторы: Ю. А. Васильев, А. В. Владзимирский, К. М. Арзамасов, А. Е. Андрейченко, В. А. Гомболевский, Н. С. Кульберг, О. В. Омелянская, Н. А. Павлов, Р. В. Решетников, К. А. Сергунова, Д. Е. Шарова, И. М. Шулькин

Организаторы и активные участники первого этапа (2019–2020 гг.) Эксперимента (сотрудники ГБУЗ «НПҚЦ ДиТ ДЗМ» и медицинских организаций ДЗМ): Е. С. Ахмад, Р. Н. Ахметов, А. С. Бардин, Е. Г. Бахтеева, И. А. Блохин, О. С. Бухтиярова, Р. И. Волошин, Е. В. Дмитриева, В. А. Дроговоз, С. О. Ермолаев, Д. И. Живоглядов, В. С. Живоденко, С. М. Зайцева, Г. В. Иванова, Е. М. Ильина, С. Г. Киреев, Т. Г. Киреев, Ю. С. Кирпичев, В. Г. Кляшторный, Д. Ю. Кокина, Е. М. Корепина, Н. Д. Кудрявцев, Е. С. Кузьмина, Д. А. Курятников, Н. В. Ледихова, Т. А. Логунова, М. Ю. Малова, В. И. Мельников, Ю. Е. Мешалкин, О. А. Мокиенко, С. П. Морозов, Е. Л. Морозова, А. Н. Мухортова, Ю. В. Никишова, В. П. Новик, В. А. Нуждина, Е. В. Панина, А. В. Петрайкин, Н. С. Полищук, Е. В. Попов, В. Г. Раковчен, А. А. Ревазян, Л. Г. Родионова, С. С. Семенов, А. В. Слепушкина, И. В. Смирнов, И. В. Солдатов, Е. С. Соломатина, А. В. Титова, И. А. Трофименко, Е. В. Туравилова, А. В. Усков, С. С. Федоров, А. Н. Хоружая, С. Ф. Четвериков, Н. В. Шипилова

Технический редактор А. И. Овчарова
Компьютерная верстка Е. Д. Бугаенко

ПРЕДИСЛОВИЕ

В ноябре 2019 года Указом Президента Российской Федерации утверждена Национальная стратегия развития искусственного интеллекта на период до 2030 года. Искусственный интеллект должен внести колоссальный вклад в реализацию приоритетных направлений научно-технологического развития России. Одним из таких направлений стало создание условий для улучшения уровня жизни населения за счет повышения качества услуг в сфере здравоохранения, включая профилактические обследования, диагностику, прогнозирование возникновения и развития заболеваний, подбор оптимальных дозировок лекарственных препаратов, сокращение угроз пандемий, автоматизацию и повышение точности хирургических вмешательств.

Как подчеркивается в Национальной стратегии стремительное развитие технологий искусственного интеллекта сопровождается существенным ростом как государственных, так и частных инвестиций в их развитие и разработку прикладных технологических решений. Громадный интерес и значительные потоки средств не должны быть расхищены безрезультатно. Критично важно сформировать адекватный, эффективный, справедливый рынок. Важно наладить коммуникации между индустрией информационных технологий, компьютерных наук и сферой здравоохранения, клинической медициной. Разработать общие правила, стандарты, которые бы в равной степени отвечали потребностям и признавались бы всеми заинтересованными сторонами. Критично необходима методическая поддержка разработчиков, ведь «механический» перенос принципов разработки искусственного интеллекта из промышленности или транспорта на медицину невозможен.

Залогом формирования рынка, прежде всего, служит создание спроса. Это означает необходимость научных исследований по трансформации диагностических и клинических процессов медицинских организаций на основе технологий искусственного интеллекта. Важно научно проверить гипотезы о том, как действительно автоматизация влияет на качество, производительность труда в здравоохранении; как помогает быстрее и точнее ставить диагноз, назначать и осуществлять лечение. Понимание транс-

формации производственных процессов позволит создать новые медицинские услуги, оказываемые с применением технологий искусственного интеллекта. Не исключая врача, но дополняя и усиливая его возможности, а также — повышая эффективность и производительность всей системы здравоохранения.

Безусловно, формирование рынка крайне важно для развития экономики столицы и государства, но не является самоцелью. Развитие искусственного интеллекта должно изменить способы и качество оказания медицинских услуг. Автоматизация может освободить медицинский персонал от рутинных, повторяющихся действий, чреватых дефектами и рисками. Вместо этого, врачи и медицинские сестры смогут физически больше времени уделять общению с пациентами, широко и полноценно используя клиническое мышление принимать сложные решения, недоступные компьютерам. Выполнение отдельных функций искусственным интеллектом позволит сократить время ожидания медицинской помощи, например — ускорит предоставление гражданам результатов диагностических исследований. Автоматизированный контроль не допустит ошибок в назначениях или маршрутизации, а значит медицинская помощь станет более безопасной.

Особенность состоит в том, что для медицины нужна не только разработка технологий, но и их всесторонняя научная оценка.

Научный подход к развитию искусственного интеллекта в здравоохранении не только дополняет экономические и инженерно-технические его аспекты, но, прежде всего, создает основу для безопасности пациента. Цена ошибки или неточности в медицине критично высока. Поэтому неукоснительным требованием к любой новой технологии в здравоохранении является безвредность для пациента. Полномасштабное обеспечение безопасности при применении новых технологий обеспечивается научными исследованиями на принципах доказательной медицины.

В целях реализации Национальной стратегии уже в ноябре 2019 года Правительство Москвы обеспечило организацию и проведение крупнейшего в мире научного исследования искусственного интеллекта в медицине — Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы.

Эксперимент направлен на получение знаний об эффективной и безопасной трансформации производственных процессов медицинских организаций на основе искусственного интеллекта. На их основе формируются условия для эффективного взаимодействия государства (органов исполнительной власти, регуляторов), медицинских и научных организаций, бизнеса и индустрии, а также граждан. Это позволит российским технологиям искусственного интеллекта занять значительную долю мирового рынка, значительно повысив качество и безопасность медицинской помощи.

РЕЦЕНЗЕНТЫ

Нуднов Николай Васильевич — д.м.н., профессор, заместитель директора по научной работе, заведующий НИО комплексной диагностики заболеваний и радиотерапии ФГБУ «Российский научный центр рентгенорадиологии» Минздрава России

Лебедев Георгий Станиславович — д.т.н., профессор, директор Института цифровой медицины, заведующий кафедрой информационных и интернет-технологий ФГАОУ ВО «Первый Московский государственный медицинский университет им. И. М. Сеченова» Минздрава России (Сеченовский Университет)

БЛАГОДАРНОСТИ

Авторы монографии выражают искреннюю благодарность и глубочайшее профессиональное уважение руководителям и сотрудникам Департамента здравоохранения города Москвы, Департамента информационных технологий города Москвы, медицинских организаций государственной системы здравоохранения города Москвы, членам Междисциплинарной рабочей группы по разработке и внедрению системы поддержки врачебных решений на этапе первичной диагностики и лечения: Агафоновой Олесе Алексеевне, Антиповой Анне Михайловне, Барышову Владимиру Ивановичу, Бойко Светлане Валентиновне, Вакуленко Елене Анатольевне, Васильевой Вере Игоревне, Вдовиной Александре Евгеньевне, Григель Вере Владимировне, Гулиевой Марии Ахадовне, Гуревичу Александру Борисовичу, Иманбердиной Айман Толеповне, Летучиной Татьяне Александровне, Лобанову Михаилу Николаевичу, Мокиенко Олеся Александровне, Никитину Борису Сергеевичу, Никишовой Юлии Вячеславовне, Пермогорской Анне Сергеевне, Поливанову Гайку Эдуардовичу, Самбурскому Станиславу Евгеньевичу, Смирновой Евгении Игоревне, Соколиной Ирине Александровне, Стецюк Лидии Дмитриевне, Трофимовой Марине Владимировне, Хомякову Александру Константиновичу, Хохловой Ирине Вячеславовне, Шепелевой Елене Николаевне, а также всем сотрудникам ГБУЗ «НПКЦ ДиТ ДЗМ» г. Москвы, всем замечательным коллегам, благодаря которым успешно стартовал первый в мире эксперимент по применению технологий искусственного интеллекта в лучевой диагностике.

Авторы

ГЛОССАРИЙ

АНАЛИТИЧЕСКАЯ ВАЛИДАЦИЯ — оценка корректности обработки входных данных программным обеспечением для создания надежных выходных данных; оценивается с применением эталонных наборов размеченных данных.

ИИ-СЕРВИС — программное обеспечение на базе технологий компьютерного зрения, предназначенное для анализа медицинских изображений и предоставляемое юридическим лицом (разработчиком или имеющим право на предоставление).

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ (ИИ) — комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе то, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений.

КЛИНИЧЕСКАЯ АПРОБАЦИЯ — оценка эффективности медицинского изделия (в данном контексте — программного обеспечения) путем использования в рамках стандартного производственного процесса. Состоит из двух компонентов: клинической корреляции и клинической валидации.

КЛИНИЧЕСКИЕ ИСПЫТАНИЯ — разработанное и запланированное систематическое научно-практическое исследование, принятое для оценки безопасности и эффективности медицинского изделия.

КЛИНИЧЕСКИЙ КОНТЕКСТ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА — единый комплекс специфической базовой информации для осознанного и эффективного применения конкретной технологии искусственного интеллекта в практическом здравоохранении, в том числе включающий цель, задачи (запросы), конкретные процессы и операции, нозологии, виды данных, функции, способы и формы представления результатов анализа, измеримые метрики качества.

КОМПЬЮТЕРНОЕ ЗРЕНИЕ — технология искусственного интеллекта для формирования полезных выводов относительно объектов и сцен реального мира на основе анализа изображений, полученных с помощью датчиков.

НАБОР ДАННЫХ — совокупность данных, прошедших предварительную подготовку (обработку) в соответствии с требованиями законодательства РФ об информации, информационных технологиях и о защите информации и необходимых для разработки программного обеспечения на основе искусственного интеллекта.

НАБОР ДАННЫХ ВЕРИФИЦИРОВАННЫЙ — это эталонный набор данных для обучения или тестирования алгоритмов искусственного интеллекта, подготовленный проспективно и содержащий данные из медицинской документации об окончательном и/или патологоанатомическом диагнозе.

ОБЪЯСНИМЫЙ ИИ — (от англ. explainable AI), функциональная возможность программного обеспечения на основе технологий ИИ объяснить/интерпретировать свое решение и степень уверенности в этом решении человеку.

ПОДГОТОВКА ДАННЫХ — процесс выгрузки структурированных и неструктурированных данных из медицинских информационных систем по заданным критериям (фильтрам).

РАЗМЕТКА ДАННЫХ — этап обработки структурированных и неструктурированных данных, в процессе которого данным (в том числе текстовым документам, фото- и видеоизображениям) присваиваются идентификаторы, отражающие тип данных (классификация данных), и (или) осуществляется интерпретация данных для решения конкретной задачи, в том числе с использованием методов машинного обучения.

ТЕХНИЧЕСКИЕ ИСПЫТАНИЯ — разработанное и запланированное исследование соответствия характеристик медицинского изделия требованиям нормативной документации, технической и эксплуатационной документации производителя и принятия последующего решения о возможности проведения клинических испытаний.

ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА — технологии, основанные на использовании искусственного интеллекта, включая компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта.

СПИСОК СОКРАЩЕНИЙ

АРМ — автоматизированное рабочее место врача.

БД — база данных.

ГБУЗ «НПКЦ ДиТ ДЗМ» — государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы».

ДЗМ — Департамент здравоохранения города Москвы.

ДУ — диагностическое устройство.

ЕМИАС — Единая медицинская информационно-аналитическая система.

ЕРИС — Единый радиологический информационный сервис.

ЕСУВВ — подсистема ЕМИАС, Единая система уведомлений для внешних взаимодействий.

ЗНО — злокачественное новообразование.

ИИ — искусственный интеллект.

ИО — истинно отрицательный.

ИП — истинно положительный.

ИТ — информационная технология или интеллектуально-технологический.

КЗ — компьютерное зрение.

КТ — компьютерная томография.

КТТ — контрольно-техническое тестирование.

ЛО — ложноотрицательный.

ЛП — ложноположительный.

ММГ — маммография.

МО — медицинская организация.

МРТ — магнитно-резонансная томография.

НДКТ — низкодозная компьютерная томография.

ОЛД — отделение лучевой диагностики.

ПКТИ — предварительные клинико-технические испытания.

ПО — программное обеспечение.

ППАК — продуктивный программно-аппаратный комплекс, продуктивная среда ЕРИС.

ПУМ — подсистема ЕМИАС, продукт управления моделями.

ПЭТ/КТ — позитронно-эмиссионная компьютерная томография.

РИС — радиологическая информационная система.

РГ — рентгенография.

РЛ — рак легкого.

РФ — Российская Федерация.

ТПАК — тестовый программно-аппаратный комплекс, тестовая среда ЕРИС.

ФЛГ — флюорография.

AUC — англ. Area Under Curve; площадь под характеристической кривой.

CAD — англ. Computer-Aided Detection and Diagnosis system; программное обеспечение, система компьютерного обнаружения и диагностики.

CE — фр. Conformité Européenne; маркировка «европейское соответствие».

CONSORT — англ. Consolidated Standards of Reporting Trials; диаграмма потока участников исследования.

DICOM — англ. Digital Imaging and Communication in Medicine; Международный стандарт создания, хранения, визуализации и передачи медицинских файлов, которые хранят информацию о проведенных исследованиях. В данном документе под этими файлами подразумеваются файлы исследований, включающих цифровые медицинские изображения, сформированные на оборудовании в отделениях лучевой диагностики, документы пациентов и протоколы хранения/передачи информации.

FDA — англ. Food and Drug Administration; Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США.

PACS — программное обеспечение, система передачи и архивации медицинских изображений.

ROC — англ. Receiver Operating Characteristic; характеристическая кривая.

ВВЕДЕНИЕ

*Чем грандиознее идея и ее польза,
тем слабее бывает первое исполнение.
Причина понятна.
Это – трудность ее реализации.*

Константин Эдуардович Циолковский

Искусственный интеллект (ИИ). Всего несколько лет назад это словосочетание буквально потрясло мир лучевой диагностики. «Робот заменит врача!», «Радиология исчезает!», «Безработные врачи заполняют улицы!» — каких только лозунгов и предвидений не довелось услышать. Тысячи инженеров и математиков по всему миру вдруг бросились создавать стартапы для разработки очередного «искусственного интеллекта», в надежде лишить непонятно, чем провинившихся рентгенологов куска хлеба. Транснациональные вендоры диагностической аппаратуры, неторопливо оглядевшись с высоты своего положения, побежали вдогонку, наверстывая упущенное. Появились разнообразные «экосистемы» и «платформы» для совместной работы разработчиков и врачей. Рекой потекли инвестиции — для наиболее быстрых аббревиатура «AI» в названии гипотетического продукта стала залогом получения финансовых средств (впрочем, это продолжалось недолго). Под напором ИТ-индустрии врачи «ушли в глухую оборону», поглядывая на происходящее со стороны, иногда иронично, иногда нервно, но чаще всего — безучастно. Тем не менее на сотнях научных конференций по всему миру закипели дискуссии. Визионеры рассказывали с трибун о грядущей роботизации лучевой диагностики. Прагматики требовали предъявить работающие технологии. Популисты публиковали в социальных сетях свои рассуждения.

Что же послужило причиной такого взрывного интереса и кипучей активности? Идею усиления человеческого разума за счет возможностей особых «интеллектуальных машин» сформулировал в 1832 году ветеран Наполеоновских войн, чиновник Министерства внутренних дел Семен Николаевич Корсаков (1787–1853). В своем потрясающем труде, основанном на принципах механи-

стического материализма, Семен Николаевич предложил пять «машин, сравнивающих идеи», а еще — перфокарты, метод многокритериального поиска с использованием весовых коэффициентов и первый способ обработки больших данных (предтечу современных алгоритмов)¹. Увы, блестящие открытия не прошли строгого рецензирования академиков того времени и почти на сто лет были забыты. Лишь с развитием кибернетики Норберта Виннера, Сергея Соболева, Виктора Глушкова и многих других ученых вновь вспомнились идеи «интеллектуальных машин».

В 1950 году английский математик Алан Тьюринг (1912–1954) впервые систематизировал рассуждения на тему возможности машинного мышления и предложил тест, позволяющий различить (или не различить) результаты интеллектуального труда машины и человека². Впрочем, в своих трудах Тьюринг, по-видимому, опирался на идеи философа Алфреда Айера, писавшего об эмпирических тестах для поиска различий между «разумным существом» и «глупой машиной» для определения «наличия или отсутствия сознания»³. В 1956 году американский ученый Джон Маккарти (1927–2011) предложил термин «искусственный интеллект» как резюме двухмесячной дискуссии на тему, возможно ли моделировать рассуждения, интеллект и творческие процессы с помощью вычислительных машин. При этом под интеллектом понималась только «вычислительная составляющая способности достигать целей в мире». В дальнейшем периоде времени кибернетика, философия и инженерия шли разном, создавая компьютеры, алгоритмы, языки программирования, а «искусственный интеллект» стал отдельной научной дисциплиной.

Но причем же здесь медицина? Достаточно быстро — примерно к началу 1970-х годов — появились компьютерные технологии анализа медицинских данных. Наибольшей эффективностью отличались средства автоматизированного анализа электрокардиогра-

¹ Корсаков С. Н. Начертание нового способа исследования при помощи машин, сравнивающих идеи / Пер. с франц. под ред. А. С. Михайлова. М.: МИФИ, 2009. 44 с.

² Turing A. Computing Machinery and Intelligence // Mind. 1950. Vol. 59, №236. P. 433–460.

³ Ayer A.J. Language, Truth, and Logic. 2nd edition. London: Gollancz, 1946. 160 p.

фических (ЭКГ) исследований. Были разработаны нужный математический аппарат, принципы алгоритмизации анализа, создан ряд программных продуктов, часть из которых даже применялась в контексте коммерческих услуг по дистанционной автоматической расшифровке, а также — в рамках космических программ.

В СССР наибольший вклад в развитие «вычислительной ЭКГ-диагностики» внесли Юрий Исаакович Неймарк (1920–2011) и Александра Петровна Матусова (1919–2010), в США — Цезарь Касерес (1927–2020)¹. Примерно через 10 лет появились «экспертные системы» — это были первые попытки создать автоматизированные системы поддержки принятия врачебных решений (СППВР). В научных журналах 80-х годов XX века можно найти большое количество публикаций на эту тему. Экспертные системы были предложены для множества дисциплин — кардиологии, хирургии, эндокринологии, нейрохирургии и т. д. Однако значительного, действительно масштабного развития эти СППВР не получили, так как основывались на жестких правилах. Фактически в основе их алгоритмов лежали структурированные «деревья решений»; они могли быть очень обширными, но предсказуемыми и полностью статичными. Обучаться или предлагать нетривиальные подходы эти системы еще не умели. Интерес к автоматизированному анализу медицинских данных резко снизился. Этот период иногда называют «зимой ИИ». Но в конце 1990-х — начале 2000-х гг. началась «весна». Своим возрождением ИИ обязан новому математическому аппарату — искусственным нейронным сетям — и стремительному наращиванию вычислительных возможностей компьютерных технологий.

Для создания алгоритмов на основе нейронных сетей сформировались способы обучения соответствующих моделей, в том числе — машинное, глубинное и проч. Появилась *техническая* возможность создавать алгоритмы для быстрой обработки огромных массивов данных в условиях низкой определенности; причем такие алгоритмы могут дообучаться и работать на обычных персональных компьютерах. Вот тут-то на медицину и обрушилась волна искусственно создаваемого ажиотажа. И первый ее удар приняла именно лучевая диагностика. Для разработчиков ИИ здесь все было очевидно:

¹ Владимирский А. В. История телемедицины: стоя на плечах гигантов (185–1979). М.: Де`Либри, 2019. 410 с.

— врач смотрит на снимок и ставит диагноз — значит алгоритм может сделать эту работу за человека;

— рентгенограмм и томограмм делают много — значит продукт на основе алгоритма будет востребован.

Данные тезисы звучали логично, а главное — выглядели очень простыми по сравнению со сложными комбинациями симптомов и десятков лабораторных показателей, которые нужно анализировать, например, для прогноза сепсиса или коронарного синдрома. Увы, такая логичная и простая конструкция полностью оказалась оторванной от жестких реалий настоящей медицины.

Вдоволь наслушавшись прогнозов и угроз, врачи подошли к новой проблеме стандартным прагматичным путем — с позиций медицинской науки и доказательной практики. Для автоматизированного анализа, как и для любого иного метода/средства/технологии в здравоохранении, должны быть научным путем доказаны: безопасность, эквивалентность существующим аналогам, эффективность (превосходство над существующими аналогами). По установленным правилам и принципам научных исследований должны быть обеспечены **методологические** возможности для практического применения технологий автоматизированного анализа медицинских данных. Все перечисленное обусловило формирование искусственного интеллекта в качестве объекта профессиональных научных исследований.

Диалектические единство и борьба противоположностей в отношении медицинского искусственного интеллекта наглядно иллюстрируются двумя цитатами. Сооснователь стартапа в области онлайн-обучения «Coursera» и профессор технологий машинного обучения Эндрю Ын утверждает: «Предприниматели XIX века в короткие сроки поставили электричество на службу человечеству, чтобы тот мог готовить пищу, освещать помещения и приводить в действие промышленное оборудование. Точно также современные предприниматели ставят на службу человеку машинное обучение и искусственный интеллект»¹. Это ярко демонстрирует решительность и быстроту действий ИТ-индустрии.

Но как ярко и четко возражает (возражает ли?) ему врач-радио-

¹ Цитируется по: Ли Кай-фу. Сверхдержавы искусственного интеллекта. Китай, кремниевая долина и новый мировой порядок. М.: Манн, Иванов и Фербер, 2019. 240 с.

лог и медицинский блогер Хью Харви: «Если вы планируете инвестировать в ИИ, то выбирайте компании, которые сосредоточены на решении клинически значимой проблемы, обладают значительным объемом верифицированных клинических данных, нацелены на получение регистрации продукта как медицинского изделия и не чрезмерно распиарены. Все остальное — хайп»¹. Данное выражение подчеркивает взвешенность и системность со стороны сферы здравоохранения. Обе отрасли (здравоохранение и ИТ-индустрию) роднит целеустремленность, вот только цели у всех в этом процессе разные.

После 2010 года стало появляться все больше научных статей, посвященных не только теории, но и практике применения искусственного интеллекта в здравоохранении. На фоне потока публикаций — статей и препринтов — от математиков и программистов особую ценность представляли работы, написанных именно врачами, пусть даже и вовлеченными в разработку того или иного решения.

Большинство практических публикаций было направлено на изучение диагностической и прогностической точности алгоритмов автоматизированного анализа, использовавшихся в лучевой и лабораторной диагностике, кардиологии, дерматологии и ряде иных дисциплин. Впрочем, именно рентгенология и радиология оставались здесь бесспорными лидерами — применению ИИ в этих сферах были посвящены более 70% публикаций². Примерное распределение предпочтений модальностей среди создателей алгоритмов было следующим: магнитно-резонансная томография (МРТ) — 40% разработок, компьютерная томография (КТ) — 27%, ультразвуковые исследования (УЗИ) — 6%, маммография — 4%, рентгенография — 3%, гибридные лучевые методы — 1%. Среди анатомических локализаций, чьи лучевые изображения подвергались автоматизированному анализу наиболее часто, лидировали центральная нервная система — 40%, опорно-двигательная систе-

¹ Цитируется по: Hugh Harvey. URL: <https://medium.com/@DrHughHarvey>.

² Kim D.W., Jang H.Y., Kim K.W., Shin Y. et al. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers // Korean J Radiol. 2019. Mar. Vol. 20, №3. P. 405–410. URL: <https://doi.org/10.3348/kjr.2019.0025>.

ма — 8%, сердечно-сосудистая система и молочная железа — по 7% каждая, мочевыводящая система, грудная полость и брюшная полость — по 6% каждая¹. Итак, большинство разработчиков трудились над алгоритмами для анализа диагностических изображений, а врачи «отвечали» изучением качества этих алгоритмов.

В 2019 году был подведен первый итог научных исследований искусственного интеллекта с позиций доказательной медицины. Результаты были катастрофическими: из 516 проанализированных статей клинические аспекты применения ИИ описывали... 0 (ноль!) из них. Лишь 1% статей соответствовал стандартам дизайна диагностического исследования, обязательного для медицинской науки. При обучении моделей применяли медицинские данные из нескольких клиник, то есть придерживались мультицентрового подхода, лишь 2,9% разработчиков. Валидировали свои алгоритмы на новых, не использованных для обучения данных — 6% команд. Итоговая картина выглядела следующим образом. В подавляющем большинстве ситуаций алгоритмы ИИ для медицины:

- создавали на ограниченных наборах данных;
- проверяли на тех же самых данных (в итоге алгоритм просто «не знал», что бывают другие данные со своими особенностями);
- при проверке использовали нестандартный статистический аппарат, доказывая лишь принципиальную потенциальную пригодность ИИ для решения некой — часто довольно абстрактной — медицинской задачи.

Результаты работы алгоритмов оказывались невоспроизводимыми на новых данных и в условиях новых медицинских организаций; для врачебного сообщества доказательность таких статей была нулевой. ИТ-сообщество продолжало твердить о консерватизме медиков, а на самом деле абсолютное большинство врачей не хотело применять непроверенное сомнительное средство, руководствуясь древним принципом медицины: *Noli Nocere!*²

Важно отметить, что в том же 2019 году появились первые международные обобщения списков программного обеспечения на ос-

¹ Pesapane F., Codari M., Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine // Eur Radiol Exp. 2018. Oct 24. Vol. 2, №1. P.35. URL: <https://doi.org/10.1186/s41747-018-0061-6>.

² Не навреди! (лат.). См.: Душенко К. В., Багриновский Г. Ю. Большой сло-

нове искусственного интеллекта, сертифицированного в качестве медицинского изделия. Наиболее показателен в этом ключе список Института науки о данных Американского колледжа радиологии. Из него узнаем, что среди сертифицированных ИИ лидируют алгоритмы для нейровизуализации — 38%, для работы с изображениями органов грудной клетки (включая сердце и сосуды) — 30%, молочной железы — 15%³. Казалось бы, наличие государственной сертификации в качестве медицинского изделия — это отличная гарантия безопасности и качества. Но, увы, это не так. Под давлением ИТ-рынка в США установлена практика сертификации программного обеспечения на основе искусственного интеллекта по принципу прецедентов; при этом отсутствует система клинических испытаний ИИ (в отличие от медикаментов, инструментов и прочих иных технологий для медицины). С точки зрения безопасности для пациента, минимизации рисков ошибки для медицинского работника, качества в реальном клиническом контексте, такой подход не выдерживает критики.

Таким образом, **методологические** возможности для практического применения технологий искусственного интеллекта так и оставались *terra incognita*⁴; напряженность между ИТ-индустрией и врачебным сообществом нарастала...

варь латинских цитат и выражений. М.: Эксмо, 2013. 976 с.

³ URL: <https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms>.

⁴ Неизвестная земля (лат.); перен. — нечто совершенно неизвестное. См.: Душенко К. В., Багриновский Г. Ю. Указ. раб.

Часть первая. КЛИНИЧЕСКАЯ МЕТОДОЛОГИЯ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВОЙ ДИАГНОСТИКЕ

Глава 1. СИСТЕМНЫЙ ПОДХОД К ПРОБЛЕМЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВОЙ ДИАГНОСТИКЕ

§1. Стратегия развития технологий искусственного интеллекта

В 2019 году Указом Президента Российской Федерации от 10.10.2019 №490 утверждена Национальная стратегия развития искусственного интеллекта на период до 2030 года¹, в которой определяются цели и основные задачи развития искусственного интеллекта в Российской Федерации, а также меры, направленные на его использование в целях обеспечения национальных интересов и реализации стратегических национальных приоритетов, в том числе в области научно-технологического развития.

Основные принципы развития и использования технологий ИИ в России таковы:

- защита прав и свобод человека;
- безопасность;
- прозрачность;
- технологический суверенитет;
- целостность инновационного цикла;
- разумная бережливость;
- поддержка конкуренции.

Национальная стратегия прямо определяет, что использование технологий искусственного интеллекта в социальной сфере способствует созданию условий для улучшения уровня жизни насе-

¹ Указ Президента Российской Федерации от 10.10.2019 №490 «О развитии искусственного интеллекта в Российской Федерации».

ления, в том числе за счет повышения качества услуг в сфере здравоохранения (включая профилактические обследования, диагностику, основанную на анализе изображений, прогнозирование возникновения и развития заболеваний, подбор оптимальных дозировок лекарственных препаратов, сокращение угроз пандемий, автоматизацию и точность хирургических вмешательств). Нельзя не добавить, что указанное в Стратегии повышение качества услуг в сфере образования также имеет непосредственное отношение к здравоохранению, непрерывному профессиональному развитию медицинских работников.

Стратегией установлено, что использование технологий искусственного интеллекта во всех отраслях носит общий («сквозной») характер и способствует созданию условий для улучшения эффективности и формирования принципиально новых направлений деятельности. Развивая положения Стратегии, утверждаем, что в здравоохранении сказанное может достигаться за счет:

а) повышения эффективности процессов планирования, прогнозирования и принятия управленческих решений — создания и мониторинга достижения целевой модели здравоохранения, в том числе в особых условиях (пандемий, чрезвычайных ситуаций и проч.);

б) автоматизации рутинных (повторяющихся) производственных операций — оптимального использования ресурсов, высвобождения медицинского персонала для когнитивно сложных трудовых задач;

в) использования автономного интеллектуального оборудования и робототехнических комплексов, интеллектуальных систем управления логистикой — роботизации диагностических и инвазивных процедур, роботической телемедицины для повышения доступности медицинской помощи, минимизации дефектов, связанных с человеческим фактором;

г) повышения безопасности сотрудников при выполнении бизнес-процессов — реализации постоянного (фонового) контроля качества, снижающего риски и частоту дефектов и ошибок в работе медицинского персонала;

д) повышения лояльности и удовлетворенности потребителей (в том числе направления им персонализированных предложений и рекомендаций, содержащих существенную информацию) — обеспечения приверженности к сохранению здоровья, выполне-

нию программ диспансерного наблюдения, лечению; реализации в лечении каждого пациента уникального сочетания персонализированного подхода и принципов доказательной медицины;

е) оптимизации процессов подбора и обучения кадров, составления оптимального графика работы сотрудников с учетом различных факторов — реализации индивидуальных траекторий развития специалистов с учетом показателей контроля качества, индивидуальных предпочтений, потребностей системы здравоохранения.

Благодаря реализации Стратегии должны быть созданы условия для эффективного взаимодействия государства, организаций, в том числе научных, и граждан в сфере развития искусственного интеллекта, что позволит российским технологиям искусственного интеллекта занять значительную долю мирового рынка.

§2. Научно-методологическое обеспечение реализации положений Национальной стратегии развития искусственного интеллекта в сфере лучевой диагностики

К историческому моменту принятия Национальной стратегии развития искусственного интеллекта в Российской Федерации уже была проведена значительная подготовительная научная, аналитическая и методическая работа, позволившая в сфере здравоохранения (точнее — лучевой диагностики) сразу перейти к реализации стратегических задач.

Еще в 2015 году в Научно-практическом клиническом центре диагностики и телемедицинских технологий Департамента здравоохранения города Москвы (НПКЦ ДиТ ДЗМ) впервые в Российской Федерации начались системные научные исследования применения технологий искусственного интеллекта в здравоохранении, в частности — в лучевой диагностике.

По своей научной тематике исследования полностью соответствовали следующей наиболее значимой международной повестке:

1. Стандартизация и методология подготовки данных для обучения алгоритмов.

2. Клинический контекст применения ИИ, в том числе как основы для продуктивной его разработки.

3. Оценка качества и методология клинических испытаний технологий ИИ.

В глобальной перспективе научным врачебным сообществом перечисленные направления (или «вызовы») были заявлены как наиболее актуальные, требующие немедленных научных изысканий.

Программа научных исследований НПКЦ ДиТ ДЗМ включала:

1. Аналитические исследования:

- мониторинг и анализ рынка;
- коммуникации с компаниями-разработчиками;
- коммуникации с врачебным сообществом;
- систематизация научных публикаций;
- теоретическая (эмпирическая) разработка методологий.

2. Научное исследование:

- тестирование диагностической точности алгоритмов ИИ;
- апробация и практическая валидация разработанных методологий.

Также использовались методы социологических опросов, экспертных интервью, оценки диагностической точности по совокупности метрик и т. д.

Особую сложность создавало *отсутствие в глобальной перспективе общепринятых инструментов научного анализа технологий искусственного интеллекта в медицине*. Поэтому ключевой задачей программы научных исследований стала разработка нужных методологий.

В ходе аналитической части исследования проводились постоянный мониторинг рынка, научной публикационной активности, а также дискуссии, как в формате бизнес-встреч, так и в ходе конференций, с разработчиками различного уровня (от инициативных стартапов до транснациональных производителей медицинской аппаратуры). Всего было проведено около 100 подобных встреч, в ходе которых были достигнуты соглашения о более углубленном научном исследовании 19 различных алгоритмов, предназначенных для использования в сфере лучевой диагностики (разработки компаний из Российской Федерации, Бельгии, Великобритании, Испании, Китая, Нидерландов, Объединенных Арабских Эмиратов, Соединенных Штатов Америки, Южной Кореи). Углубленное исследование

дование предполагало экспертные интервью и независимое тестирование точности и надежности работы алгоритмов на новых данных.

Для научного тестирования отобранных решений был сформирован ряд эталонных наборов подготовленных (размеченных) данных, так называемых «датасетов», 4 из которых получили официальное свидетельство о государственной регистрации базы данных. Для помощи многочисленным разработчикам в 2018 году в свободном доступе впервые в Российской Федерации был размещен ограниченный деперсонализированный набор размеченных компьютерных томограмм грудной клетки. Этот набор скачали несколько десятков раз и использовали для самотестирования и обучения несколько групп разработчиков и компании (подробнее см. главу 4).

Были осуществлены исследования 18 алгоритмов автоматизированного анализа диагностических изображений и 1 алгоритма для распознавания естественного языка и анализа медицинской документации (таблица 1).

Таблица 1 – Данные об изученных алгоритмах ИИ для лучевой диагностики

Компания	Страна	Модальность	Клиническая цель
Изображения			
Aidoc	США	КТ	Острая патология головного мозга
		КТ	Заболевания легких
Botkin Ai	Россия	КТ	Очаги в легких
Care mentor	Россия	РГ	Заболевания легких
Cogna	ОАЭ	КТ	Очаги в легких
		ММГ	ЗНО молочной железы
Icometrix	Бельгия	МРТ	Рассеянный склероз
HyHy	Китай	КТ	Очаги в легких
Kheironmed	Великобритания	ММГ	ЗНО молочной железы
Lunit	Южная Корея	РГ	Заболевания легких
		ММГ	ЗНО молочной железы
Philips	Нидерланды	НДКТ	Очаги в легких
Quibim	Испания	РГ	Заболевания легких
Qure.ai	Индия	РГ	Туберкулез, заболевания легких
Yrim	Россия	ММГ	ЗНО молочной железы
ЗМнение	Россия	РГ	Заболевания легких
Цельс	Россия	ММГ	ЗНО молочной железы
ФтизисБиоМед	Россия	РГ	Туберкулез
Документы			
IBM	США	—	Контроль качества медицинской документации
Примечание: КТ – компьютерная томография; РГ – рентгенография; ММГ – маммография; МРТ – магнитно-резонансная томография; НДКТ – низкодозная компьютерная томография; ЗНО – злокачественное новообразование.			

В ходе тестирования нами установлены типовые проблемы со стороны разработчиков:

1. Отсутствие или низкое качество клинически обоснованного целеполагания при разработке, непонимание контекста применения автоматизации в реальных производственных процессах (в связи с притоком в сферу медицины разработчиков из других отраслей, не имеющих представления о реальностях и особенностях практического здравоохранения).

2. Отсутствие или пренебрежение стандартами, методологические проблемы при формировании наборов данных для обучения (включая терминологические расхождения, игнорирование клинического контекста при разметке изображений).

3. Отсутствие методического понимания сути автоматизированного анализа медицинских изображений:

- некачественная сегментация анатомической области (например, поиск очага в легких по принципу «белое на черном» неоднократно приводил к ложному обнаружению «патологии» не только в брюшной полости, но и вне человеческого тела);

- отсутствие контроля качества изображения, получаемого на вход, с позиций его пригодности для диагностики (например, рентгенограмма органов грудной клетки с неполным захватом области, что чревато пропуском тяжелой патологии — рака в верхушке легкого или гемоторакса в плевральных синусах);

- отсутствие контроля целевой анатомической области, то есть контроля соответствия заявленной медицинской услуги фактическому изображению (в результате алгоритм старательно обнаруживает «очаги в легких» на рентгенограммах пяточной кости).

4. Создание дискретного программного обеспечения, не позволяющего проводить интеграцию с медицинскими (радиологическими) информационными системами (такой подход усложняет производственные процессы врача и резко ограничивает возможности данного продукта).

5. Игнорирование принципов объяснимости работы искусственного интеллекта (вывод результатов работы алгоритма без пояснений и детализации просто «ставит перед фактом» врача, что негативно сказывается на восприятии и используемости технологии).

6. Отсутствие независимой валидации алгоритмов на новых данных, в том числе в дизайне проспективных мультицентровых исследований в действующих медицинских организациях (такой

подход резко снижает воспроизводимость результатов работы алгоритмов, моментально дискредитируя разработки, оказавшиеся в реальных клинических условиях).

7. Незнание или игнорирование принципов доказательной медицины (это сразу блокирует восприятие разработки медицинским сообществом, в ответ же порождается миф о «костности и консерватизме врачей»).

8. Отсутствие внутренней системы менеджмента качества у компании-разработчика (типичный дефект — отсутствие преемственности в версионности, приводящее к нестабильной работе, падению точности алгоритмов после обновлений).

9. Низкая конверсия перспективных разработок алгоритмов в готовые продукты, сертифицированные в качестве медицинских изделий (что легко объясняется предыдущим перечнем проблем).

Полученные результаты наглядно продемонстрировали выраженную нехватку методологии разработки, валидации и применения технологий искусственного интеллекта в практическом здравоохранении, в частности — в лучевой диагностике. Вместе с тем многие научные тестирования были достаточно успешными, алгоритмы надежно и качественно справлялись с клинически вполне обоснованными задачами. Это вселяло оптимизм и убежденность в необходимости дальнейших исследований. Приведем несколько примеров.

Кейс №1. Автоматизированный анализ низкодозовых компьютерных томографий (НДКТ) органов грудной клетки для выявления признаков злокачественных новообразований¹

Цель: провести аналитическую валидацию (оценку диагностической точности) алгоритма выявления очагов на НДКТ.

Референс-тест: эталонный набор данных из 100 НДКТ, сформированный в рамках программы скрининга ЗНО легкого в Москве; каждое исследование прошло двухэтапную экспертную интерпретацию.

¹ Морозов С. П., Владимирский А. В., Гомболевский В. А. [и др.] Искусственный интеллект в скрининге рака легкого: оценка диагностической точности алгоритма для анализа низкодозовых компьютерных томографий // Туберкулез и болезни легких. 2020. Т. 98. №8. С. 24–31.

Индекс-тест: алгоритм ИИ Trajanovski S et al, 2019¹ для анализа НДКТ легких.

Результаты: в соответствии с разработанной методологией проведен этап аналитической валидации — доказательный тест (тестирование алгоритма на эталонном наборе данных, расчет стандартизированного набора метрик). Результаты представлены в таблице 2; проведено их сравнительное изучение с данными самих разработчиков и опубликованными результатами тестирования аналогичных алгоритмов. На рис. 1 представлены примеры работ изучаемого алгоритма ИИ.

Таблица 2 – Установленные характеристики диагностической точности индекс-теста при различных величинах порога активации

Показатель	Значение (среднее значение, 95 % ДИ)	
	Порог активации 0,166 (риск $\geq$ 16,6 %)	Порог активации 0,102 (риск $\geq$ 10,2 %)
Чувствительность	0,783 (0,658; 0,879)	0,817 (0,696; 0,905)
Специфичность	0,975 (0,868; 0,999)	0,925 (0,796; 0,984)
Точность (общая валидность)	0,860 (0,776; 0,921)	0,860 (0,776; 0,921)
Отношение правдоподобия положительного результата	31,3 (5,0; 879,0)	10,9 (3,4; 56,6)
Отношение правдоподобия отрицательного результата	0,22 (0,12; 0,39)	0,20 (0,10; 0,38)
Прогностическая ценность положительного результата	0,979 (0,889; 0,999)	0,942 (0,841; 0,988)
Прогностическая ценность отрицательного результата	0,750 (0,611; 0,860)	0,771 (0,627; 0,880)
Площадь под характеристической кривой	—	0,930 (0,880; 0,979)
Примечание – ДИ – доверительный интервал.		

¹ Trajanovski S., Mavroeidis D., Swisher C. L. et al. Towards radiologist-level cancer risk assessment in CT lung screening using deep learning // arXiv:1804.01901v2 [cs.CV] 11 Apr. 2019.

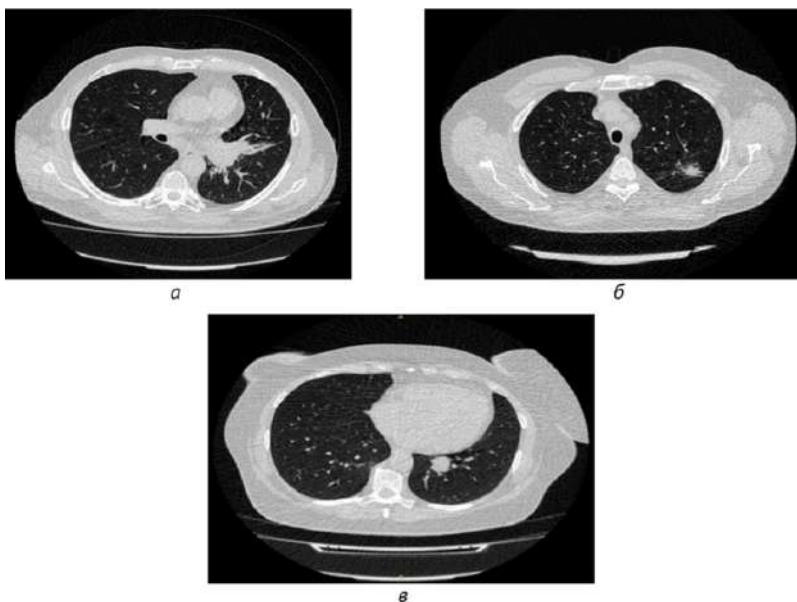


Рисунок 1 – Решения индекс-теста при анализе НДКТ органов грудной клетки: а – ложноотрицательный результат (при наличии ЗНО в корне левого легкого алгоритм предоставляет вероятность рака легкого 0,028 и степень сомнения (uncertainty) 0,0215); б – истинно положительный результат (при наличии ЗНО в верхней доле левого легкого алгоритм предоставляет вероятность рака легкого 0,91 и степень сомнения (uncertainty) 0,061); в – ложноположительный результат (при наличии гамартомы (плотность – 20 Нu) в нижней доле левого легкого алгоритм предоставляет вероятность рака легкого 0,48 и степень сомнения (uncertainty) 0,0288)

Выводы:

1. Установлены следующие показатели диагностической точности алгоритма ИИ для автоматизированного выявления очагов в легких на НДКТ: чувствительность – 0,817, специфичность – 0,925, точность – 0,860, площадь под характеристической кривой – 0,930.
2. Сопоставление полученных значений с итогами собственного тестирования разработчиками алгоритма ИИ свидетельствует о хо-

рошей воспроизводимости результатов работы ИИ на независимых данных, относящихся к популяции Москвы.

3. Изученный алгоритм ИИ может быть допущен к клинической апробации. В частности, он может применяться для первичного автоматизированного анализа результатов скрининговых обследований (с обязательным вторичным просмотром врачом-рентгенологом).

Кейс №2. Автоматизированный анализ рентгенограмм органов грудной клетки для выявления признаков туберкулеза¹

Цель: провести аналитическую валидацию (оценку диагностической точности) алгоритма выявления заболеваний органов грудной клетки для массовых профилактических осмотров.

Референс-тест:

— экспериментальная выборка №1 (n=140), соотношение «норма»: «патология» 50:50,

— экспериментальная выборка №2 (n=150), соотношение «норма»: «патология» 95:5.

Индекс-тест: система автоматизированного анализа флюорографических снимков на основе сверхточных нейронных сетей типа U-NET, модифицированных и обученных специальным образом.

Результаты: в соответствии с разработанной методологией проведен этап аналитической валидации — доказательный тест (тестирование алгоритма на эталонных наборах данных, расчет стандартизированного набора метрик). Результаты представлены в таблице 3; проведено их сравнительное изучение с данными самих разработчиков и опубликованными результатами тестирования аналогичных алгоритмов. На рис. 2 представлены примеры работ изучаемого алгоритма ИИ.

¹ Морозов С. П., Владимирский А. В., Ледихова Н. В. [и др.]. Оценка диагностической точности системы скрининга туберкулеза легких на основе искусственного интеллекта // Туберкулез и болезни легких. 2018. Т. 95. №8. С. 42–49.

Таблица 3 – Установленные характеристики диагностической точности индекс-теста

Параметры	Выборка №1	Выборка №2
Чувствительность	87,2% (95% ДИ 77,0–94,0%)	75,0% (95% ДИ 34,9–96,8%)
Специфичность	60,0% (95% ДИ 47,6–71,5%)	53,5% (95% ДИ 45,0–61,9%)
Отношение правдоподобия положительного результата	2,2 (95% ДИ 1,6–3,0)	1,62 (95% ДИ 1,04–2,50)
Отношение правдоподобия отрицательного результата	0,22 (95% ДИ 0,1–0,41)	0,47 (95% ДИ 0,14–1,6)
Прогностическая ценность положительного результата	68,6% (95% ДИ 61,7–74,7%)	8,3% (95% ДИ 5,6–12,4%)
Прогностическая ценность отрицательного результата	82,4% (95% ДИ 71,1–89,9%)	97,5% (95% ДИ 91,9–99,2%)

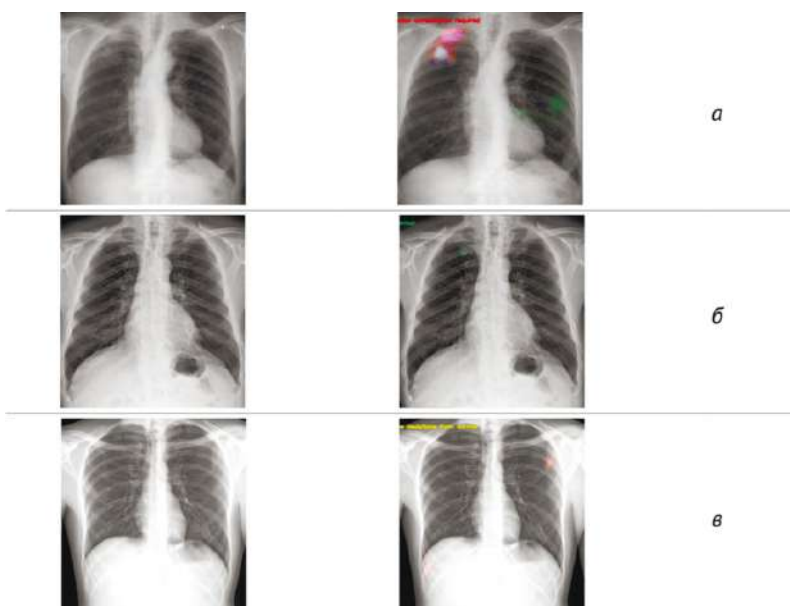


Рисунок 2 – Решения индекс-теста при анализе рентгенограмм грудной клетки: а – истинно положительный, б – ложноотрицательный, в – ложноположительный результаты

Выводы:

1. В результате ретроспективного исследования определены параметры диагностической точности системы скрининга туберкулеза легких на основе «искусственного интеллекта»: чувствительность 75,0–87,2%, специфичность 53,5–60,0%, площадь под кривой 0,64–0,74.

2. Система применима только для массовых профилактических осмотров в популяциях с низкой претестовой вероятностью наличия патологии, что подтверждается значением прогностической ценности отрицательного результата (97,5%).

3. Система может быть рекомендована для полуавтоматизированного формирования в процессе скрининга групп риска по туберкулезу легких для последующей верификации результатов компьютерного анализа флюорограмм врачом-рентгенологом.

Кейс №3. Автоматизированный анализ протоколов описаний результатов компьютерной томографии органов грудной клетки в целях контроля качества¹

Цель: провести аналитическую валидацию (оценку диагностической точности) алгоритма распознавания естественного языка.

Референс-тест: эталонный набор данных из протоколов описаний 5047 результатов НДКТ.

Индекс-тест: когнитивная система обработки естественного языка («IBM Watson's Natural Language Processing Algorithm»).

Результаты: проведен многосторонний анализ протоколов НДКТ органов грудной клетки с целью обнаружения в них изменений по заданным параметрам, а также для выявления неточностей. Установлено, что в 8,3% документов содержались расхождения между описанием и заключением. Суть расхождений состояла в том, что значимый элемент (например, наличие очагов в легких) был указан лишь в одном компоненте протокола. При математическом сопоставлении согласованности решений о расхождениях в наличии триггеров между компонентами протокола установлено,

¹ Морозов С. П., Владимирский А. В., Гомболевский В. А. [и др.]. Искусственный интеллект: автоматизированный анализ текста на естественном языке для аудита радиологических исследований // Вестник рентгенологии и радиологии. 2018. Т. 99. №5. С. 253–258.

что коэффициент каппа Козна составляет 0,95. Определено соответствие рекомендаций, содержащихся в заключении, типу выявленного очага по классификации LungRADS. Установлено, что для очагов LungRADS 3 рекомендованные принципы ведения пациентов использованы в 46% случаев, для LungRADS 4A — в 42%, а для LungRADS 4B — в 49%. Автоматическая классификация очагов и анализ заключений были верифицированы двумя независимыми врачами-радиологами; точность системы обработки естественного составляет 96%.

Выводы:

1. Алгоритмы коммерческой системы обработки естественного языка позволили провести автоматизированный анализ 5047 депersonализированных протоколов низкодозовых компьютерных томографий, выполненных в рамках масштабной программы скрининга рака легких. Согласованность решений (коэффициент каппа Козна) при использовании системы в рамках контроля качества лучевых исследований составляет 95–96%.

2. Показаны потенциальные возможности системы для выявления значимых слов и их сочетаний, характеризующих качество описаний и заключений, а также для анализа соответствия рекомендаций, формулируемых врачом-рентгенологом, методическим и иным требованиям.

3. Можно констатировать факт применимости системы обработки естественного языка в качестве инструмента для контроля качества лучевых исследований, то есть для сплошного автоматизированного анализа протоколов и иных текстовых документов.

Кейс №4. Автоматизированный анализ магнитно-резонансных томографий головного мозга для выявления признаков рассеянного склероза¹

Цель: провести аналитическую валидацию (оценку диагностической точности) алгоритма выявления признаков рассеянного склероза (проспективно, в условиях отделения лучевой диагности-

¹ Морозов С. П., Владимирский А. В., Черняева Г. Н. [и др.]. Валидация диагностической точности алгоритма «искусственного интеллекта» для выявления рассеянного склероза в условиях городской поликлиники // Лучевая диагностика и терапия. 2020. №2 (11). С. 58–65.

ки медицинской организации, оказывающей первичную (амбулаторно-поликлиническую) медицинскую помощь).

Референс-тест: поток результатов МР-исследований за период времени ($n=93$). Каждое исследование интерпретировано в 2 этапа. Первично — врачом-рентгенологом, непосредственно проводившим исследование; второе чтение проводилось двумя врачами-рентгенологами (заведующим отделением и независимым врачом-экспертом). При отсутствии расхождений во мнении исследование относили к одной из двух групп. При наличии расхождений исследование интерпретировалось повторно, коллегиально, с участием врача-рентгенолога с субспециализацией по нейрорадиологии, после чего исследование относили к одной из двух групп.

Индекс-тест: оригинальный алгоритм ИИ, Беляев М. с соавт., 2019.

Результаты: в соответствии с разработанной методологией проведен этап аналитической валидации — доказательный тест (проспективное тестирование алгоритма, расчет стандартизированного набора метрик). Результаты представлены в таблице 4; проведено их сравнительное изучение с данными самих разработчиков и опубликованными результатами тестирования аналогичных алгоритмов. На рис. 3 и 4 представлены примеры работ изучаемого алгоритма ИИ.

Таблица 4 – Установленные характеристики диагностической точности индекс-теста

Показатель	Значение (среднее значение, 95% ДИ)
Чувствительность	1,000 (0,398; 1,000)
Специфичность	0,753 (0,650; 0,838)
Точность (общая валидность)	0,763 (0,664; 0,845)
Отношение правдоподобия положительного результата	4,0 (1,1; 6,2)
Отношение правдоподобия отрицательного результата	0,00 (0,00; 0,93)
Прогностическая ценность положительного результата	0,154 (0,044; 0,349)
Прогностическая ценность отрицательного результата	1,000 (0,946; 1,000)
Площадь под характеристической кривой	0,861 (0,747; 0,975)



Рисунок 3 – Решения индекс-теста при анализе МРТ головного мозга. МР-томограммы головного мозга разных пациентов с ложноположительным заключением при использовании предложенного алгоритма: а, б – T2-ВИ FLAIR с подавлением сигнала от жира (FatSat) в аксиальной плоскости. Определяются многоочаговые поражения обеих гемисфер головного мозга, не характерные для рассеянного склероза (стрелки). Последствия вентрикулоперитонеального шунтирования (наконечник стрелки)

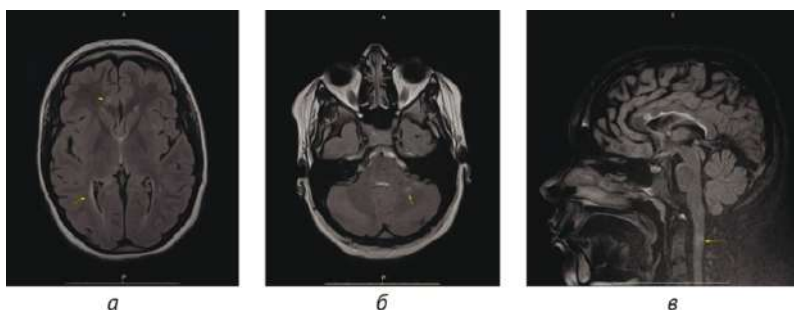


Рисунок 4 – Решения индекс-теста при анализе МРТ головного мозга. МР-томограммы головного мозга разных пациентов с истинно положительным заключением при использовании предложенного алгоритма: а, б – T2-ВИ FLAIR в аксиальной плоскости; в – T2-ВИ FLAIR с подавлением сигнала от жира (FatSat) в сагиттальной плоскости. Отмечаются характерные для рассеянного склероза очаговые изменения в головном мозге, в том числе с локализацией в юстакортикальных отделах (а – головка стрелки), перивентрикулярных отделах (а – стрелка), левой гемисфере мозжечка (б – стрелка), спинном мозге (в – стрелка)

Выводы:

1. Результатом работы стала разработка алгоритма ИИ для выявления рассеянного склероза на МРТ в условиях типовой городской поликлиники, который обеспечивает эффективную сортировку МР-исследований в условиях первичного звена здравоохранения с поддержанием оптимального уровня настороженности относительно рассеянного склероза.

2. Проведена валидация диагностической точности алгоритма ИИ, при этом чувствительность метода составила 100,0, специфичность – 75,3, точность – 76,3, площадь под характеристической кривой – 0,861. Уровень прогностической ценности отрицательного результата составил 100%, что свидетельствует о надежном «отсеивании» алгоритмом результатов исследований без признаков рассеянного склероза.

3. Сопоставление полученных значений с собственными предыдущими результатами и литературными данными свидетельствует о достаточном качестве и отличной воспроизводимости результа-

тов работы алгоритма на независимых данных. Установлено, что на снижение специфичности алгоритмов ИИ для выявления РС влияет возраст пациентов.

4. Рекомендуется расширенная клиническая апробация алгоритма в условиях медицинской организации первичного звена, оказывающей амбулаторно-поликлиническую медицинскую помощь.

В процессе исследования мы активно общались с профессиональной аудиторией — врачами-рентгенологами, радиологами, специалистами ультразвуковой диагностики, руководителями служб лучевой диагностики, организаторами здравоохранения. В ходе бесед и дискуссий легко было понять общее отношение: настороженное ожидание новых технологий на фоне недостатка объективных данных.

Тогда, помимо стандартных научных и популярных способов предоставления объективной информации, мы приняли решение провести специальное мероприятие в формате «поединка» алгоритмов и врачей, получившее название «ИИ-баттл: рентгенологи против искусственного интеллекта».

Мероприятие было запланировано как открытый конкурс разработчиков ИИ для лучевой диагностики. К участию допускались российские и зарубежные юридические лица, разработавшие или имеющие права на алгоритмы для автоматизированного анализа результатов компьютерной томографии или рентгенографии органов грудной клетки (выявление очагов и иных патологических состояний) или маммографии (скрининг злокачественных новообразований).

Участникам предлагалось продемонстрировать точность и качества своего алгоритма на эталонном наборе данных. Для конкурса были разработаны три полностью анонимизированных набора данных по 100 исследований в каждом (50% случаев с целевой патологией и 50% случаев нормы). Подготовка данных заключалась в разметке 3 врачами-экспертами с достижением консенсуса по каждому случаю. Отмечались исследования с патологией, координаты и диаметр патологических очагов.

Несмотря на широкое анонсирование мероприятия в открытом конкурсе — в открытой схватке с теми самыми рентгенологами, которых хотели заменить искусственным интеллектом — решились

принять участие только четыре компании. Это были разработчики из России («Медицинские скрининг-системы (Цельс)», «Третье мнение», «IRYM») и из Китая («HY Medical»).

Конкурс был проведен в рамках конференции МРО «Лучевая диагностика в онкологии» (Москва, Сколково, 13–14 декабря 2019 г.), на которой присутствовало более полутора тысяч участников из России, стран СНГ, Европы и Китая. Он состоял из заочной и очной фаз. За 12 часов до очной фазы участники получили доступ к анонимизированным наборам данных, провели их анализ и предоставили жюри результаты работы алгоритмов для расчета метрик диагностической точности. Очная фаза конкурса включала короткую самостоятельную презентацию каждым участником и, собственно, поединок в следующем формате. Из наборов данных были отобраны 5 клинических случаев по каждой модальности — как вполне тривиальных, так и диагностически сложных. Аудитории врачей-рентгенологов демонстрировали каждый клинический случай и предлагали вынести общий вердикт. Затем представлялось мнение искусственного интеллекта, а вердиктом служила согласованная позиция 3-х экспертов, проводивших разметку наборов данных.

В течение второго дня на конференции присутствовало большое количество участников, в аудитории проходило активное обсуждение: после абстрактных выступлений и маркетинговых выставок у врачей появилась возможность подробно ознакомиться с очередной революционной технологией. По модальности «рентгенография» с врачами состязался только один алгоритм, показавший неплохие результаты, которые в четырех случаях совпали с мнением врачей. Последний, пятый, снимок оказался фатальным: ИИ нашел патологию там, где ее не было; но при этом еще и пропустил перелом ребра довольно коварной локализации, однако вполне очевидный.

По модальности «маммография» врачам противостояли сразу два алгоритма, которые не смогли представить консолидированные результаты — их мнения постоянно различались. И если первый алгоритм почти все время совпадал в своих решениях с врачами, то второй часто ошибался. Общий исход раунда оказался нетривиальным. В трех случаях алгоритмы дали ложноположительный ответ, «увидев» онкологическую патологию. Проблема состояла в том, что и участники конференции разошлись во мнениях, од-

нако эксперты, размечавшие исследования, вынесли вердикт — на всех трех снимках присутствуют доброкачественные изменения.

Отличные результаты показал алгоритм по модальности «компьютерная томография»: полное совпадение мнение врача и ИИ было в четырех ситуациях. В последнем случае алгоритм сделал ошибку, хотя сомнения имелись и у врачебной аудитории. Для решения проблемы был созван консилиум экспертов с субспециализацией.

Эмоции присутствующих и атмосферу ИИ-баттла достоверно передал один из основателей ассоциации разработчиков и пользователей систем искусственного интеллекта в медицине «Национальная база медицинских знаний» Александр Гусев: «Долой скучные выставки и рекламный спам: живое сравнение систем с экспертами и сидящими в зале врачами создает совершенно другое восприятие».

Общая оценка участников проводилась по сумме четырех этапов: анкетирование (оценка системы качества компании), анализ набора данных из 100 клинических случаев (метрики диагностической точности), презентация компаний (оценка симпатий зрителей), соревнование в точности диагностики (голосование аудитории по итогам поединка). Для каждого этапа была сформирована система балльной оценки. Жюри в составе врачей, физиков и инженеров приняло окончательное решение. На торжественном закрытии конференции были вручены награды: приз зрительских симпатий получила компания «Третье мнение», в номинации «Помощник врача» лучшей стала «NY Medical», в номинации «Самый точный» — «IRYM», а гран-при заслужила компания «Медицинские скрининг-системы (Цельс)» (рис. 5).

Таким образом, надо отметить, что в течение 4-х лет велась системная работа по изучению феномена технологий искусственного интеллекта относительно применимости в сфере лучевой диагностики, созданию необходимых методологий и принципов, коммуникациям с индустриальным и медицинским сообществом.

В качестве итогов аналитического исследования и научного тестирования алгоритмов были опубликованы 14 научных работ, в том числе 2 монографии в издательствах «Springer-Verlag» и «ГЭОТАР-Медиа», 3 статьи в журналах, индексируемых Scopus и Web of Science (в том числе во входящем в первый квартиль цитируемости «Insights into Imaging»); разработаны 2 методических

рекомендаций; получено 6 документов о регистрации прав на интеллектуальную собственность для программного обеспечения и баз данных; сделано около 50 докладов на научно-практических конференциях. Также в целях популяризации тематики выпущено более 30 научно-популярных публикаций, организован первый в Российской Федерации «ИИ-баттл: рентгенологи против искусственного интеллекта».

Проведенная работа позволила получить широкое представление о текущем состоянии разработок, качестве, особенностях работы технологий искусственного интеллекта. Но прежде всего эта работа стала научно-практической основой для создания необходимой научной базы — в ее процессе нами формировалось понимание клинического контекста применения ИИ в лучевой диагностике, создавались и совершенствовались методология клинических испытаний и эталонные наборы данных, а также поэтапно развивались технологические условия для автоматизации лучевой диагностики. Подробнее об этом расскажем в следующих главах.

Глава 2. ТЕХНОЛОГИЧЕСКАЯ ОСНОВА ДЛЯ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВОЙ ДИАГНОСТИКЕ

Основное предназначение технологий искусственного интеллекта в лучевой диагностике — это автоматизация определенных участков производственных процессов, в которых задействованы руководители, врач и средний медицинский персонал, с целью повышения безопасности и качества, но прежде всего — с целью оптимизации (подробно это утверждение будет доказано в следующей главе). Поэтому использование технологий ИИ должно происходить в «фоновом» режиме, без необходимости увеличения количества трудовых операций, а конкретные программные решения должны быть встроены в информационную инфраструктуру отделений лучевой диагностики.

Наличие медицинской (радиологической) информационной системы с интегрированным диагностическим оборудованием и автоматизированными рабочими местами врачей-рентгенологов — это абсолютный минимум с точки зрения технической готовности. Для применения технологий ИИ в оптимальном масштабе и с высокой эффективностью требуется наличие **единого цифрового пространства службы лучевой диагностики** в рамках субъекта РФ: централизованного архива медицинских изображений с подключением до 100% диагностического оборудования и обеспечением доступа к результатам исследований до 100% врачей-рентгенологов.

Функциональные задачи единого цифрового пространства службы лучевой диагностики:

- обеспечить наличие хранилища данных медицинских изображений с возможностью «сквозной» идентификации результатов исследований данного пациента;
- предоставить защищенный доступ для уполномоченных сто-

рон (медицинских организаций, их сотрудников, пациентов и их законных представителей) ко всему массиву диагностических данных определенного пациента;

- выполнять централизованное накопление и анализ в режиме реального времени технологической информации, поступающей от диагностической аппаратуры (включая информацию о дозовой нагрузке, применяемых протоколах и проч.);

- реализовать поддержку принятия решений;

- обеспечить возможность проведения мероприятий по контролю качества процесса проведения и результатов диагностических исследований;

- формировать информационный базис для принятия управленческих решений в отрасли здравоохранения.

Единое цифровое пространство службы лучевой диагностики должно включать **технологические** и **методические компоненты**.

К технологическим компонентам относятся:

- парк цифровой диагностической аппаратуры, объединенный в сеть;

- централизованный вендор-нейтральный архив медицинских изображений;

- телекоммуникационное оборудование;

- рабочие станции;

- средства защиты персональных данных и обеспечения безопасности;

- системное и прикладное программное обеспечение;

- стандарты и протоколы обмена данными.

К методическим компонентам относятся:

- способ централизованного накопления и мониторинга динамики данных, характеризующих использование парка диагностической аппаратуры;

- метод дистанционного анализа и контроля качества результатов диагностических исследований с формированием индивидуальных стратегий повышения качества;

- методики описаний результатов лучевых исследований с применением телемедицинских технологий (первичные, экспертные, двойные и проч.).

В Москве технологической основой для применения искусственного интеллекта в лучевой диагностике стал Единый радиологический информационный сервис в составе Единой медицинской

информационно-аналитической системы города Москвы (ЕРИС ЕМИАС).

ЕРИС ЕМИАС — это информационная система в сфере здравоохранения, которая объединяет рабочие места рентгенолаборантов, врачей-рентгенологов и диагностическую аппаратуру, аккумулирует информацию о каждом исследовании или серии исследований, проведенных на подключенных к нему устройствах.

Функциональные возможности ЕРИС ЕМИАС включают:

- учет и регистрацию всех проводимых лучевых исследований (медицинские организации государственной системы здравоохранения г. Москвы амбулаторно-поликлинического и стационарного звеньев);

- хранение и доступ ко всему объему диагностических изображений;

- подсистему обмена информацией с любой внешней информационной системой в режиме реального времени;

- полную автоматизацию процесса логистики исследования, архивации и доступа к историческим данным пациента;

- визуализацию исследований, позволяющую провести рутинную диагностику без использования дополнительного программного обеспечения;

- шаблонизацию, структурирование и хранение медицинской информации в международных стандартах (включая структурированные шаблоны описаний);

- аналитический модуль, позволяющий оценивать ключевые показатели;

- модуль технического мониторинга доступности диагностического оборудования;

- подсистему запроса консультативной помощи в сложных клинических случаях;

- подсистему анализа и контроля качества.

Апробация ЕРИС проводилась в течение 2015 года, наращивание числа подключенных медицинских организаций амбулаторно-поликлинического звена интенсивно осуществлялось в 2016 году. В 2017 году ЕРИС вышел на рутинный порядок использования, в 2018 — проведена интеграция ЕРИС и ЕМИАС, в 2019 — начато подключение к ЕРИС ЕМИАС медицинских организаций стационарного звена, а в 2020 — завершено формирование единого цифрового пространства лучевой диагностики стои-

цы. Данные об интегрированном в ЕРИС ЕМИАС оборудовании приведены в таблице 5.

Таблица 5 – Распределение подключенного к ЕРИС ЕМИАС оборудования по типам медицинских организаций

Медицинские организации ДЗМ / Оборудование по модальностям	Амбулаторно- поликлиническое звено		Стационарное звено		Итого
	абс.	%	абс.	%	
Ангиография	-	-	52	100,0%	100,0%
Радионуклидная диагностика (гамма-камера)	-	-	10	100,0%	100,0%
Остеоденситометрия	27	90,0%	3	10,0%	100,0%
КТ	51	28,0%	131	72,0%	100,0%
ММГ	78	68,4%	36	31,6%	100,0%
МРТ	36	37,9%	59	62,1%	100,0%
ОФЭКТ/КТ	-	-	3	100,0%	100,0%
РГ	260	44,7%	322	55,3%	100,0%
Флюорография	160	78,0%	45	22,0%	100,0%
ПЭТ-КТ*	17	100,0%	-	-	100,0%
Всего	629	49,4%	661	51,9%	100,0%
* Интегрированы ПЭТ/КТ-устройства медицинских организаций всех форм собственности, участвующих в реализации территориальной программы государственных гарантий оказания гражданам бесплатной медицинской помощи в системе обязательного медицинского страхования.					

Также была обеспечена возможность работы для 100% врачей-рентгенологов и рентгенолаборантов, организован доступ к результатам исследований для лечащих врачей, направляющих пациентов на исследования. Путем интеграции с порталом государственных услуг г. Москвы (www.mos.ru) реализована возможность предоставления результатов исследований (как описания, так и непосредственно изображения) пациентам через личный кабинет.

В фазе апробации в ЕРИС ЕМИАС было накоплено около 95 тысяч изображений, в фазе рутинного применения к концу 2020 года их общее количество превысило 7 миллионов.

Управленческий компонент. Накапливаемая в ЕРИС информация о загруженности, режимах работы, протоколах исследований диагностической аппаратуры доступна в режиме реального времени посредством специального аналитического интерфейса или «дашборда» (рис. 5).

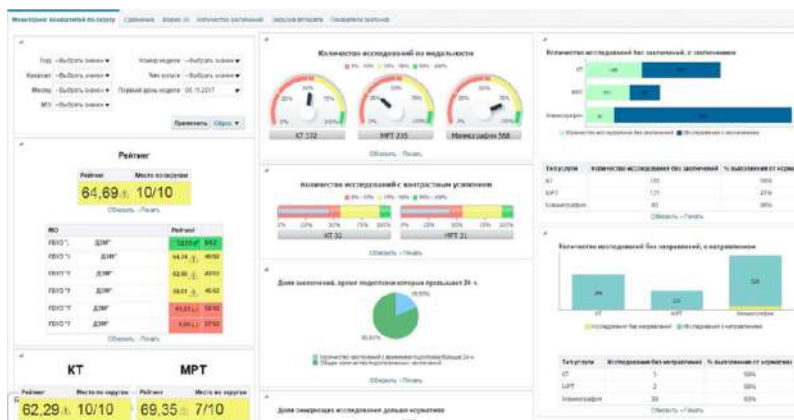


Рисунок 5 – Аналитический интерфейс («дашборд») ЕРИС ЕМИАС – инструмент для управления службой лучевой диагностики Москвы

Интерфейс – это уникальный аналитический инструмент, позволяющий в режиме реального времени мониторить работу всей службы лучевой диагностики г. Москвы, оперативно принимать необходимые управленческие решения. Данные, поступающие в систему с промежуток 2 секунды, накапливаются и доступны для ретроспективного анализа. Эта подсистема разработана по оригинальной авторской концепции и уникальному техническому заданию, включающему базу метрик и методики их расчета.

В частности, посредством интерфейса можно контролировать следующие основные параметры по каждому подключенному аппарату:

- количество отработанных дней;
- количество отработанных смен;
- количество проведенных исследований;

- количество исследований, проведенных по разным каналам финансирования;

- количество проведенных исследований с применением контрастного усиления;

- время и длительность проведения исследований.

Также ЕРИС позволяет учитывать некоторые параметры работы врача-рентгенолога, рентгенолаборанта и отделения лучевой диагностики в целом:

- количество заключений, подготовленных в системе;

- время подготовки заключения в системе (разница между временем окончания исследования и временем, когда заключение появляется в системе);

- время, которое пациенты ожидают проведения исследования (разница между датой выдачи направления на исследование и датой проведения исследования);

- количество пациентов, не имеющих направления на исследование.

Кроме того, в режиме реального времени можно оценивать техническое состояние оборудования и эффективность его загрузки. В пилотном режиме (в том числе с использованием специализированного программного обеспечения сторонних производителей) доступен контроль дозовой нагрузки.

Аналитический интерфейс может быть настроен для использования организаторами здравоохранения различного уровня, исходя из конкретных задач: отдельно для заведующего отделением, окружного внештатного специалиста и главного внештатного специалиста по лучевой диагностике столицы.

Благодаря ЕРИС ЕМИАС в автоматическом режиме налажен сбор данных по количеству выполненных исследований в разрезе модальностей оборудования, виду исследований, времени проведения исследования, режима работы медицинского персонала и эксплуатации оборудования. Проводится регулярный мониторинг работоспособности и загрузки оборудования, количества диагностических ошибок по модальностям, видам оборудования, отделениям лучевой диагностики, врачам и рентгенолаборантам. Актуализируется перечень малозагруженных аппаратов, потенциально пригодных для перемещения в медицинские организации, нуждающиеся в компьютерном или магнитно-резонансном томографе. Применяются рейтинги отделений для увеличения количества ис-

следований с внутривенным контрастированием и обеспечения их доступности. На основе данных из ЕРИС ЕМИАС готовятся управленческие решения, формируются нормативно-правовая и методическая документация.

Медицинский компонент. Наличие ЕРИС ЕМИАС позволило обеспечить цифровизацию производственных процессов медицинских организаций в части лучевой диагностики. Одновременно появилась возможность для проведения дистанционных интерпретаций и описаний результатов исследований. С применением телемедицинских технологий стали проводиться экспертные телемедицинские консультации (их среднее количество в год стабильно достигает 3–4 тысяч). Затем в 2018 году проведены успешные эксперименты по централизации описаний и перекрестным описаниям по субспециализациям.

В первом случае проведена реорганизация производственных процессов, кадрового и ресурсного обеспечения одной городской поликлиники с разветвленной филиальной сетью путем создания единого центра описаний, использующего в своей работе ЕРИС ЕМИАС и методики телемедицины. В результате достигнуты: экономия финансовых средств за счет снижения фонда оплаты труда на 59% (более 19 млн руб. в год), затрат на фотолабораторию на 40% (0,4 млн руб. в год), сокращения коммунальных и эксплуатационных расходов на 25% (более 6 млн руб. в год); рост производительности труда (число описаний, выполняемых одним врачом, увеличилось с 4,7 до 21 в день).

Во втором случае организованы описания результатов исследования по субспециализациям с учетом анатомического принципа. Для этого в ЕРИС ЕМИАС была налажена специальная схема маршрутизации. Такой подход обеспечил снижение удельного веса клинически значимых расхождений с 2,0% до 1,0%, незначимых – с 5,0% до 3,0%, замечаний общего характера – с 35,0% до 16,0% (в исследованиях по субспециализации «нейрорадиология»). Также благодаря «перекрестной» маршрутизации исследований для описаний полностью устранены простои и перебои, связанные с физическим отсутствием врача (например, по болезни или из-за отпуска).

Таким образом, были получены результаты, свидетельствующие об экономической эффективности, росте производительности труда, повышении качества и доступности медицинской помощи. Они

легли в основу концепции Референс-центра лучевой диагностики, созданного в Москве в 2019 году.

Особое значение имеет такой методический компонент ЕРИС ЕМИАС, как дистанционный анализ и контроль качества лучевых исследований. Пересмотр (от англ. «peer review») результатов диагностических исследований — глобально признанный подход, заключающийся в повторном анализе диагностических изображений и их описаний квалифицированным врачом-экспертом (чаще всего непосредственным руководителем) для выявления расхождений, ошибок, неточностей и их методического устранения.

Ограниченная методика пересмотра была развита нами в концепцию **дистанционного анализа и контроля качества**, которая подразумевает дистанционный, систематический, независимый и документируемый процесс оценки качества проведения и описания рентгенорадиологических исследований, выполненных в медицинской организации с целью определения степени их соответствия рекомендуемым стандартам. Оценка проводится постоянно, на регулярной основе, независимой группой экспертов; для анализа случайным образом формируется выборка исследований и их описаний, при этом она анонимизируется. Научно обоснована классификация расхождений, которые могут быть выявлены при анализе. Разработаны и автоматизированы в ЕРИС ЕМИАС процессы работы врачей-экспертов, обеспечивающие максимальную объективность, беспристрастность и точность выявления расхождений. Отличительная черта дистанционного анализа и контроля качества — направленность на непрерывное улучшение. По итогам анализа формируются рейтинги и индивидуальные стратегии повышения качества. Последние могут включать учебно-методические и образовательные мероприятия в разных форматах, организационные, технические и, крайне редко, административные мероприятия.

Для Московского референс-центра лучевой диагностики дистанционный анализ и контроль качества — один из стандартных производственных процессов. Научно доказана его эффективность: на материале 48520 исследований показано достоверное снижение удельного веса расхождений, в том числе клинически значимых — почти на 50% в показателях наглядности.

В 2020 году Московский референс-центр лучевой диагностики, как «ядро» единого цифрового пространства, работает в ру-

тинном режиме. Основными его процессами являются первичные описания результатов лучевых исследований, двойные просмотры результатов скринингов и результатов компьютерной томографии пациентов с подозрением на новую коронавирусную инфекцию, экспертные консультации, дистанционный контроль качества.

Таким образом, единое цифровое пространство службы лучевой диагностики города Москвы в виде ЕРИС ЕМИАС стало технологической, а в части дистанционного контроля качества — еще методической, основой для применения технологий искусственного интеллекта.

Глава 3. КЛИНИЧЕСКИЙ КОНТЕКСТ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Если исключить философский подход и применить к современным компьютерным наукам принципы механистического материализма, то технологии искусственного интеллекта окажутся всего лишь инструментом автоматизации тех или иных производственных операций в здравоохранении. Безусловно, автоматизация должна быть осознанной, резонной, обоснованной, наконец — эффективной. Иначе как не вспомнить афоризм психолога Абрахама Маслоу (1908–1970): «Я думаю, что если твоим единственным инструментом является молоток, то на что угодно хочется смотреть как на гвоздь»¹. Автоматизация — лишь один из подходов оптимизации и улучшения производственных процессов в здравоохранении, а искусственный интеллект — лишь один из инструментов такого подхода. Поэтому внедрение технологии ИИ в практическую медицину без понимания и должного описания соответствующего клинического контекста представляется бессмысленной тратой времени и ресурсов.

Под **клиническим контекстом** мы понимаем: цель, задачи (запросы), конкретные процессы и операции, нозологии, виды данных, функции, способы и формы представления результатов анализа, измеримые метрики качества — как единый комплекс специфической базовой информации для осознанного и эффективного применения конкретной технологии искусственного интеллекта в практическом здравоохранении.

Сразу хотим сказать о ключевом аспекте представления результатов анализа, то есть результатов работы искусственного ин-

¹ Maslow A.H. The Psychology of Science. New York: Harper&Row, 1966. 168 p.

теллекта. Алгоритмы работы нейронных сетей крайне сложны для интерпретации и, следовательно, результаты их работы могут быть подвергнуты сомнению и отменены человеком¹. Отсутствие понимания того, как искусственный интеллект достигает результатов, является одной из причин низкого уровня доверия медицинских работников к соответствующим технологиям. Недопустимо в итоге автоматизации просто ставить врача или медицинскую сестру перед фактом по принципу «есть болезнь» / «нет болезни» или «назначить медикамент Х». Решения и предложения ИИ должны быть объяснимы, содержать пояснения, статистические величины (шансы, риски, вероятности) в соответствии с канонами доказательной медицины. Конечно, в зависимости от клинической ситуации медицинский работник может воспользоваться предложением ИИ без детализации — например, искусственный интеллект анализирует результаты КТ головного мозга до их описания рентгенологом и сразу сообщает нейрохирургу о наличии внутричерепного кровотечения для оказания экстренной медицинской помощи. Однако все равно должна быть предусмотрена возможность «вернуться» к решениям алгоритма и просмотреть их доказательную базу.

Таким образом, особое условие для широкого применения технологий искусственного интеллекта в практическом здравоохранении — это скорейшее развитие направления **объяснимого ИИ** (от англ. «explainable AI»). Оно подразумевает функциональную возможность программного обеспечения на основе ИИ объяснить/интерпретировать свое решение и степень уверенности в этом решении человеку. Соответствующие положения, например, в форме утверждения «системы ИИ должны включать элемент объяснимого ИИ для повышения интерпретируемости ИИ решений», уже появляются в нормативно-правовых документах ряда стран. В России Национальная стратегия развития искусственного интеллекта на период до 2030 года² вводит такой основной принцип развития и использования ИИ, как **прозрачность**: объяснимость работы искусственного интеллекта и процесса достижения им результатов, недискриминационный доступ пользователей продуктов, которые созданы с использованием техноло-

¹ Указ Президента Российской Федерации от 10.10.2019 №490 «О развитии искусственного интеллекта в Российской Федерации».

² Там же.

гий искусственного интеллекта, к информации о применяемых в этих продуктах алгоритмах работы искусственного интеллекта. Следование этому принципу — строго обязательно для технологий искусственного интеллекта, разрабатываемых для здравоохранения. Целесообразность, объяснимость, верность принципам доказательной медицины, исходный высочайший уровень качества и точности алгоритмов — вот гарантия широкого принятия технологий искусственного интеллекта практическим здравоохранением.

Далее мы проведем обоснование клинического контекста для сферы лучевой диагностики и начнем с анализа основного производственного процесса — проведения и описания исследования.

§1. Основной производственный процесс службы лучевой диагностики

Цель основного производственного процесса службы лучевой диагностики (рис. 6) — предоставление качественного обоснованного ответа на клинический запрос лечащего врача путем организации, проведения, описания и контроля качества лучевого исследования определенной модальности.

В стандартном схематическом виде лечащий врач по результатам обследований иными способами и методами принимает решение о необходимости выполнения лучевого исследования, формирует соответствующее направление. Пациент осуществляет запись на исследование. Рентгенолаборант регистрирует исследование в информационной системе и непосредственно проводит его. Если исследование выполняется с диагностической целью, то один врач-рентгенолог интерпретирует его результаты и формирует описание. При необходимости или при наличии показаний (в том числе могут быть установлены нормативными документами) врач может обратиться за экспертной консультацией к более квалифицированному доктору, возможно — с субспециализацией. Обычно для таких консультаций применяются телемедицинские технологии. Если же цель исследования — профилактическая, скрининговая, то необходимо второе чтение — параллельная независимая интерпретация и описа-

ние другим рентгенологом. Финализированный протокол исследования направляют лечащему врачу, который формулирует диагноз и ведет дальнейшую работу с пациентом. В отношении результатов лучевого исследования осуществляется ретроспективный контроль качества. Обращает на себя внимание высокая степень информатизации данного процесса. Множество отдельных операций в его составе совершаются с применением тех или иных информационных технологий.

ЭКСПЕРТ (ВРАЧ-РЕНТГЕНОЛОГ)

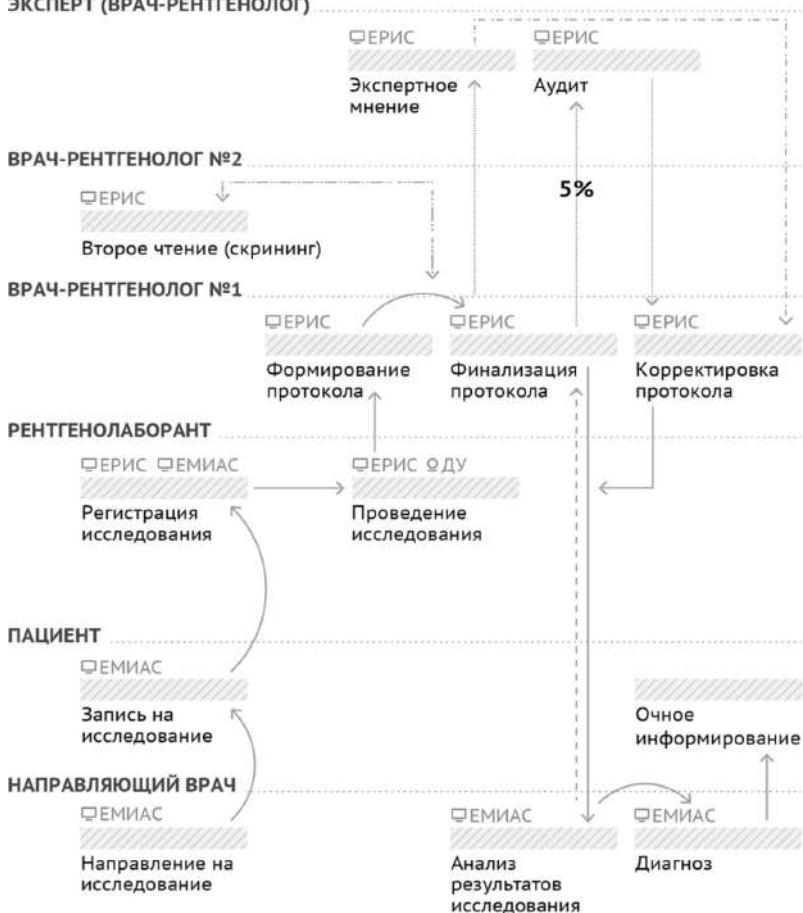
ВРАЧ-РЕНТГЕНОЛОГ №2

ВРАЧ-РЕНТГЕНОЛОГ №1

РЕНТГЕНОЛАБОРАНТ

ПАЦИЕНТ

НАПРАВЛЯЮЩИЙ ВРАЧ



□ЕРИС – Единый радиологический информационный сервис

□ЕМИАС – Единая медицинская информационно-аналитическая система

□ДУ – Диагностические устройства

Рисунок 6 – Схема основного производственного процесса службы лучевой диагностики

К сожалению, в таком варианте производственный процесс чреват дефектами, рисками, затрагивающими всех участников:

для пациента это пролонгация или ошибочный диагноз, негативные исходы; для медицинских работников — угроза возникновения юридической ответственности (рис. 7). На этапе назначения исследования есть риски некорректного информирования пациента о методике, показаниях и противопоказаниях к исследованию (например, назначение МРТ без учета наличия у пациента кардиостимулятора). Лечащий врач может быть недостаточно осведомлен об оптимальном виде и способе диагностики, что приводит к необоснованным назначениям, выбору устаревших методик. А вот уже после получения описания — в самом конце процесса — лечащему врачу может вновь не хватить компетенций с точки зрения интерпретации результатов лучевого исследования в конкретных клинических обстоятельствах. Ситуация самостоятельного выбора пациентом слотов, с одной стороны, комфортна, но, с другой — чревата дисбалансом в загруженности диагностических устройств. Рентгенолаборант может допустить механическую опечатку при регистрации исследования, совершить ошибку при укладке пациента, выбрать некорректный протокол сканирования. Самое плохое, что ошибку рентгенолаборанта выявят только на следующем этапе, когда изображения будет анализировать врач-рентгенолог. Это означает, что диагноз не будет определен вовремя, а пациенту придется снова проходить весь цикл: идти к лечащему врачу, получать направление, записываться, ждать... Вполне вероятно, что в итоге пациент напишет жалобу.

На своем этапе процесса врач-рентгенолог, конечно же, может пойти по ложному пути гипо- или гипердиагностики. Второе чтение для скрининговых исследований зачастую просто невозможно из-за выраженного кадрового дефицита, в итоге либо чрезвычайно затягиваются его сроки, либо оно совершается формально. Затянуть подготовку описания может и длительное ожидание экспертной консультации. Исходя из сложившейся практики известно, что на ретроспективный контроль качества попадает в среднем около 5% исследований. Очевидно, что подавляющее большинство дефектов при этом остается не выявленными. Подчеркнем, контроль проводится ретроспективно — через несколько недель, а то и месяцев после проведения исследования. Фактически он может обнаружить и исправить проблемы компетенций медицинского персонала. Но для конкретного пациента он уже бесполезен чаще всего.

Таким образом, есть значительные риски пролонгации диагностики, отказа в проведении исследования, необходимости повторного сканирования ионизирующим облучением, гипо- или гипердиагностики, накопления систематических ошибок персоналом. Все это в итоге ограничивает доступность и качество медицинской помощи, вызывает социальное недовольство, психоэмоциональные проблемы, а также самые негативные исходы болезни.

Потенциальную роль технологий искусственного интеллекта в рамках основного производственного процесса службы лучевой диагностики трудно переоценить (рис. 7). Фактически автоматизация может быть (должна быть!) применена на всех его этапах.

ЭКСПЕРТ (ВРАЧ-РЕНТГЕНОЛОГ)



ЕРИС – Единый радиологический информационный сервис

ЕМИАС – Единая медицинская информационно-аналитическая система

ДУ – Диагностические устройства

Рисунок 7 – Анализ проблем и рисков основного производственного процесса службы лучевой диагностики

Для лечащего врача необходимы системы поддержки принятия решений как при назначении исследования, так и при его интер-

претации, точнее — при дифференциальной диагностике по совокупности данных. Вклад автоматизации здесь составляет до 50%. Запись на исследование может быть полностью автоматизированной, учитывающей текущий статус загрузки и работоспособности парка оборудования, реальные сроки ожидания исследования, персональные данные и предпочтения пациента, возможности и специализацию отдельных медицинских организаций. Конечно, для удобства пациенту должны предлагаться несколько вариантов для записи, но не механически («все свободные»), а с учетом рациональности использования ресурсов системы здравоохранения и максимальной доступности медицинской помощи. Обеспечение идентичности персональных данных пациента на всех этапах процесса — это также функция для автоматизированного контроля. Здесь, пожалуй, стоит напомнить требование Национальной стратегии развития искусственного интеллекта на период до 2030 года¹ по обеспечению за счет технологий ИИ повышения качества предоставления государственных и муниципальных услуг, а также снижения затрат на их предоставление.

Значителен вклад технологий ИИ в работу рентгенолаборанта при непосредственном проведении исследования. Алгоритмы могут проконтролировать правильность укладки, предупредить о противопоказаниях, провести немедленный технический аудит полученного изображения (если будут обнаружены дефекты, то можно сразу провести повторное сканирование, не отправляя пациента «на второй круг»). В данном случае вклад технологий ИИ может достигать 75%.

При проведении первичного описания уровень автоматизации составляет не менее 50%: ИИ может приоритезировать исследования в рабочем списке автоматизированного рабочего места (ускорится описание случаев с признаками болезни); отметить на изображении патологические признаки и, возможно, измерить их (снизится риск пропусков патологии); помочь сопоставить текущие результаты для данного пациента с предыдущими исследованиями; сформировать шаблон описания. Таким образом, комплексно повысятся скорость и качество описаний. Второе чтение может быть полностью автоматизированным — все скрининговые

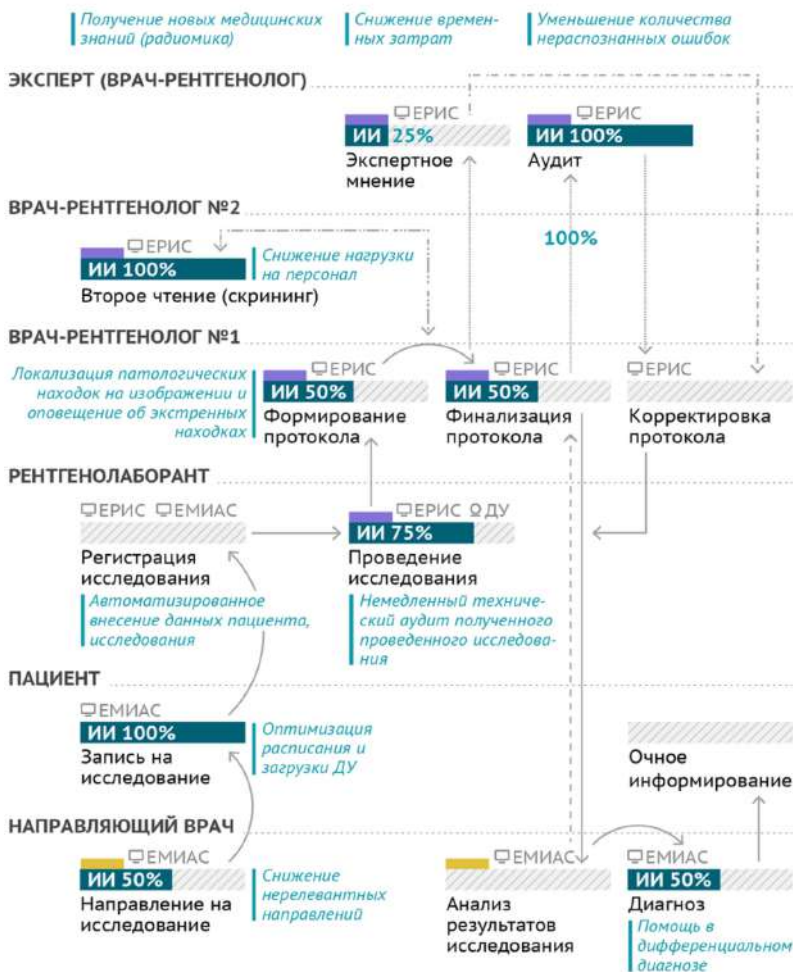
¹ Указ Президента Российской Федерации от 10.10.2019 №490 «О развитии искусственного интеллекта в Российской Федерации».

исследования могут на потоке анализироваться алгоритмом ИИ. Тогда полностью устраняется проблема кадрового дефицита и достигается максимальная точность профилактических осмотров. Немаловажное значение имеет и скорость — результаты скринингов с подозрением на наличие патологии не будут дожидаться своей очереди в огромном рабочем списке заваленного работой врача, а сразу «будут подняты наверх» с тревожной отметкой, быстрее будут описаны и направлены лечащему врачу. Следовательно, пациент быстрее получит нужное и иногда жизнеспасующее лечение. Системами поддержки принятия решений может пользоваться и врач-эксперт, в том числе с учетом особенностей субспециализации.

Благодаря комбинациям технологий компьютерного зрения и распознавания естественного языка можно добиться охвата контролем качества 100% исследований. Причем такой контроль будет происходить практически в реальном режиме времени (рис. 8). Дефекты можно будет устранить заблаговременно, то есть можно уже говорить о профилактике ошибок, а не борьбе с ними.

Таким образом, технологии искусственного интеллекта потенциально позволяют автоматизировать ряд операций в составе основного производственного процесса службы лучевой диагностики для снижения рисков дефектов и ошибок, повышения производительности труда, увеличения скорости постановки диагноза и начала специального лечения.

ИССЛЕДОВАТЕЛИ



ЕРИС – Единый радиологический информационный сервис

ЕМИАС – Единая медицинская информационно-аналитическая система

ДУ – Диагностические устройства

— Компьютерное зрение

— СПП

Рисунок 8 – Качественное улучшение и оптимизация основного производственного процесса службы лучевой диагностики

Далее мы проанализируем запросы системы здравоохранения, разберем ожидания врачебного сообщества и проанализируем потребность в автоматизации лучевой диагностики с позиции эпидемиологии и общественного здоровья.

§2. Запрос системы здравоохранения Российской Федерации

Перспективное становление технологий искусственного интеллекта должно быть согласовано с общим направлением развития отечественного здравоохранения. В России до 2024 г. выполняется Национальный проект «Здравоохранение», включающий ряд федеральных проектов¹. К наиболее релевантным нашей теме метрикам Национального проекта относятся:

- снижение уровней летальности и смертности от онкологических заболеваний;
- повышение выявляемости злокачественных новообразований на ранних стадиях;
- увеличение охвата населения профилактическими исследованиями;
- наличие в субъектах референс-центров для борьбы с онкологическими заболеваниями.

Представляется перспективным широкое внедрение технологий искусственного интеллекта для:

- работы с результатами массовых профилактических осмотров;
- выявления признаков онкологических заболеваний, особенно на ранних стадиях;
- поддержки решений по оптимальной, предписанной маршрутизации пациентов;
- оппортунистического поиска предикторов или явных проявлений иных особо значимых патологий;
- формирования проектов описаний с использованием установленных классификаций.

С учетом аспекта качества, работа решений на основе технологий искусственного интеллекта должна быть точной и сбалансированной. Категорически недопустимо создание дополнительной необоснованной нагрузки на систему здравоохранения: например, за счет избыточной генерации ложноположительных или клинически нецелесообразных результатов (к последним, в частности, относятся очаги в легких объемом менее 5 мм, которые

¹ URL: <https://minzdrav.gov.ru/poleznye-resursy/natsproektzdravooхранenie>.

невозможно верифицировать патогистологически или иным способом).

В соответствии с действующим законодательством программное обеспечение на основе технологий ИИ должно быть интегрировано с информационными системами в сферу здравоохранения субъектов РФ и/или медицинских информационных систем, в том числе референс-центров.

§3. Эпидемиологический и организационный аспекты

Онкологические заболевания по-прежнему представляют значительную угрозу и проблему для человечества. По официальным данным¹ в Российской Федерации в 2018 году показатель распространенности злокачественных новообразований составил 2 562,1 на 100 000 населения. В течение 10 лет отмечается рост этого параметра на 39,5%. С одной стороны, такая динамика обусловлена положительным развитием отечественного здравоохранения: значимо улучшается выявляемость и выживаемость онкологических больных. С другой стороны, имеется истинный рост заболеваемости, связанный с различными неблагоприятными факторами.

Об улучшении показателей выживаемости онкологических больных свидетельствует динамика индекса накопления контингента больных со злокачественными новообразованиями: за 10 лет данный показатель вырос с 5,7 до 6,9². Вместе с тем в структуре смертности населения России злокачественные ново-

¹ Злокачественные новообразования в России в 2018 году (заболеваемость и смертность) / под ред. А. Д. Каприна, В. В. Старинского, Г. В. Петровой. М.: МНИОИ им. П. А. Герцена — филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2019. 250 с.

² Состояние онкологической помощи населению России в 2018 году / под ред. А. Д. Каприна, В. В. Старинского, Г. В. Петровой. М.: МНИОИ им. П. А. Герцена — филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2019. 236 с.

образования занимают второе место (16,1%), уступая только болезням системы кровообращения (46,8%). В целом показатель запущенности (то есть выявление болезни на IV стадии) достигает 23,9%. На этом фоне уровень одногодичной летальности составляет 25,8%, что означает: почти четверть больных со злокачественными новообразованиями узнают о своем диагнозе только на практически безнадежной стадии, а в течение года после установления диагноза умирает каждый четвертый пациент. Невзирая на положительную динамику исходов и некоторое снижение смертности, проблема онкологических заболеваний еще явно далека от окончательного решения.

Высокий показатель запущенности свидетельствует о необходимости развития и активного совершенствования программ профилактических осмотров — скринингов, направленных на как можно более раннее выявление и своевременное лечение злокачественных новообразований. Методологии и технологии таких осмотров различны. Скрининг может быть селективным (исследования проводятся только в группах риска) или неселективным (проводятся исследования всего населения). Критично важно наличие эффективной системы маршрутизации лица с выявленным подозрением на опухоль для всестороннего дообследования и специализированного лечения. Так, в 2016–2020 гг. НПКЦ ДиТ ДЗМ инициировал и обеспечил проведение двух программ скрининга (ЗНО легких и молочной железы). В результате была сформирована методология, обеспечивающая значимый прирост выявляемости на ранних стадиях, а значит влияющая и на исходы. Таким образом, профилактические исследования могут внести значительный вклад в преодоление проблемы онкологических болезней, что полностью соответствует концепции и задачам развития здравоохранения Российской Федерации.

Какие же состояния несут наибольшую угрозу? Для обоих полов ведущими локализациями в общей структуре онкологической заболеваемости являются: кожа (12,6%, с меланомой — 14,4%), молочная железа (11,4%), трахея, бронхи, легкое (9,9%). У мужского населения первое место занимают опухоли трахеи, бронхов, легкого (16,9%); у женского — рак молочной железы (20,9%). Вполне закономерно, что обе нозологии — это одновременно и ведущие причины смерти. У мужчин более четверти (25,9%) смертей обусловлены раком трахеи, бронхов, легкого. У женщин злокачествен-

ные опухоли молочной железы (16,2%)¹. Далее следуют опухоли желудочно-кишечного тракта, кожи, внутренних мочеполовых органов. Примечательно, что именно лучевая диагностика помогает своевременно выявлять ЗНО легких и молочной железы (в то время как для других локализаций — используются иные инструментальные и лабораторные методы: эндоскопии, дерматоскопии, тестов на скрытую кровь и онкомаркеры).

Казалось бы, общая ситуация очевидна. Есть две формы злокачественных новообразований, лидирующие по распространенности и рискам смерти. У многих пациентов обе формы выявляются несвоевременно. Для решения этой проблемы есть программы скринингов с помощью лучевой диагностики с проверенной методологией. Почему же онкология до сих пор не побеждена?

Дело в том, что с позиций практической лучевой диагностики скрининги имеют ряд существенных проблем:

1. Нагрузка. Профилактические осмотры подразумевают проведение однотипных стандартизированных исследований у значительных контингентов лиц (от десятков тысяч до миллионов). Причем большинство этих людей здоровы; во всяком случае не имеют онкологической патологии, которую и пытаются обнаружить при скринингах. Для проведения исследования необходимо задействовать максимально большие аппаратные мощности, а также иные виды ресурсов системы здравоохранения.

2. Кадры. Для проведения такого колоссального объема исследований и интерпретации (описаний) их результатов требуется значительное, постоянно нарастающее количество среднего медицинского персонала и врачей. И это на фоне существующего дефицита квалифицированных кадров — типичной проблемы современного здравоохранения как на национальном, так и на общемировом уровнях.

3. Человеческий фактор. При описании результатов тысяч скрининговых исследований врач-рентгенолог в подавляющем большинстве случаев имеет дело с нормальной картиной, выискивая пациента с опухолью «как иголку в стоге сена». Поэтому при про-

¹ Злокачественные новообразования в России в 2018 году (заболеваемость и смертность) / под ред. А. Д. Каприна, В. В. Старинского, Г. В. Петровой. М.: МНИОИ им. П. А. Герцена — филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2019. 250 с.

смотре сотен однотипных нормальных исследований постоянно растет риск пропуска значимой находки, и та самая «иголка» может быть просто упущена. Для минимизации таких рисков применяют двойное чтение, когда результаты каждого исследования независимо интерпретируют два врача. Но в данном случае мы возвращаемся к ранее обозначенной проблеме — установленный объем дефицита специалистов удваивается.

4. Квалификация. Обнаружить рак сложно. Для качественного скрининга врач-рентгенолог должен иметь достаточный опыт, а иногда — как в случае с маммографией для выявления ЗНО молочной железы — еще и субспециализацию. Таким образом, возникает двойной дефицит не просто врачей, но специалистов наивысшего уровня.

Итак, общая ситуация крайне сложна: на фоне дефицита высококвалифицированных специалистов нужно проводить и описывать (качественно и быстро!) миллионы однотипных исследований с преимущественной нормой.

Дефицит кадров можно проиллюстрировать следующим образом: в соответствии с целями Национального проекта «Здравоохранение» должно быть достигнуто снижение смертности от новообразований до 185 случаев на 100 тыс. населения (на 7,8%). Для этого установлены целевые показатели ранней выявляемости. Если говорить о двух самых значимых формах ЗНО — опухолях легкого и молочной железы — то для достижения метрик Национального проекта потребуется выполнить около 1,8 миллиона низкодозовых компьютерных томографий и 14,8 миллиона маммографий. Для этого необходимо дополнительно задействовать порядка 5 400 врачей-рентгенологов, в то время как, по официальным данным, укомплектованность штатных должностей врачей-рентгенологов физическими лицами составляет около 59%¹. Физически увеличить количество врачей-рентгенологов, тем более необходимой специализации, не получится. Потребность всегда будет опережать реальное число должным образом подготовленных специалистов, что свидетельствует о том, что скрининги проводятся в недостаточном объеме или с неудовлетворительным качеством.

¹ Тюрин И. Е. Лучевая диагностика в Российской Федерации в 2016 г. // Вестник рентгенологии и радиологии. 2017. Т. 98. №4. С. 219–226.

Эта ситуация идеальна для автоматизации посредством технологий искусственного интеллекта. Должным образом обученный алгоритм может непрерывно анализировать результаты скрининговых исследований, выявлять подозрительные находки и направлять их для интерпретации квалифицированным врачом. Это и есть целевая модель скрининга будущего, которая может реально снизить смертность от злокачественных опухолей. Конечно, при условии очень качественного, можно сказать ответственного, обучения алгоритмов и масштабности внедрения.

§4. Ожидания врачебного сообщества

В ходе собственного аналитического и научного исследования проблемы искусственного интеллекта мы активно общались и вели дискуссию с профессиональным сообществом, также было организовано участие в масштабном опросе врачей-рентгенологов под эгидой Европейского общества радиологии. Всего были собраны и проанализированы ответы почти 675 респондентов из 39 стран Евразии и Северной Америки, в том числе из Российской Федерации¹. В течение проведения данных мероприятий стало известно, что примерно 20% врачей-рентгенологов и радиологов уже используют те или иные инструменты ИИ в своей работе.

В глобальной перспективе наиболее востребованной врачами будет являться автоматизированная обработка результатов исследований по таким модальностям (здесь и далее в порядке убывания), как:

- компьютерная томография;
- маммография или томосинтез молочной железы;
- магнитно-резонансная томография;
- позитронно-эмиссионная томография (в т.ч. в составе гибридных методик).

¹ Codari M., Melazzini L., Morozov S.P. et al. Impact of artificial intelligence on radiology: a EuroAIM survey among members of the European Society of Radiology // Insights Imaging. 2019. Oct 31. Vol. 10, №1. P. 105. DOI: 10.1186/s13244-019-0798-3.

Вместе с тем наиболее ожидаемы решения на основе ИИ для:

- анализа диагностических изображений молочной железы;
- поиска признаков онкологических заболеваний, прежде всего на изображениях органов грудной клетки;
- молекулярной визуализации;
- анализа диагностических изображений центральной нервной системы.

Примечательно, что с учетом клинического контекста ожидания врачебного сообщества относительно целей применения ИИ таковы:

- скрининг (массовые профилактические осмотры);
- стадирование новообразований;
- количественные измерения биомаркеров;
- постпроцессинг (в т.ч. для ускорения процессов получения изображений для анализа);
- генерация шаблонов описаний;
- поддержка в принятии решений при наличии симптомов.

Таким образом, профессиональным сообществом явным образом высказан запрос на автоматизацию массовых профилактических (скрининговых) исследований, в том числе исследований органов грудной клетки и молочной железы. Безусловно, скрининги не исчерпываются онкологической патологией. Специальные массовые осмотры методами лучевой диагностики проводятся для выявления туберкулеза, сердечно-сосудистых заболеваний, остеопороза, наконец, ЗНО иных локализаций. Это необходимо учитывать при разработке клинического контекста применения ИИ.

Проанализировав запрос профессиональной аудитории, мы, к сожалению, увидели вновь дисбаланс между мнениями разработчиков и врачей. По обобщенным данным среди создателей ИИ предпочтение отдается следующим модальностям: МРТ — 40% алгоритмов, КТ — 27%, УЗИ — 6%, маммография — 4%, рентгенография — 3%, ПЭТ/КТ — 1%. Распределение интереса разработчиков по анатомическим областям таково: центральная нервная система — 40%, опорно-двигательная система — 8%, сердечно-сосудистая система и молочная железа — по 7% каждая, мочевыводящая система, грудная полость и брюшная полость — по 6% каждая¹. Как видим, целевые модальности и анатомические локализации фигурируют всего лишь в 10–30% случаев.

Впрочем, преобладание мнения врачей очевидно, что иллюстрируется показателями регистрации технологий ИИ в качестве медицинских изделий. Например, в США среди модальностей лидируют КТ и МРТ, среди локализаций — центральная нервная система, органы грудной клетки, молочная железа². Следовательно, динамика регистраций алгоритмов в большей степени соответствует реальным потребностям здравоохранения и запросам врачебной аудитории. Вместе с тем мы не намерены преуменьшать роль автоматизированного анализа КТ и МРТ в нейровизуализации. Сложность диагностики заболеваний и повреждений центральной нервной системы, необходимость экстренного принятия решений во множестве ситуаций также представляет собой значительный потенциал для технологий искусственного интеллекта.

§5. Клинический контекст

В настоящее время технологии искусственного интеллекта находятся на первом, начальном, этапе системного внедрения в сферу лучевой диагностики. В данном случае **оптимальным клиническим контекстом** для их применения будет следующий:

1. Цель: достижение максимального уровня доступности и качества лучевых исследований. Уровень доступности определяется непрерывным выполнением описаний результатов исследований с опережением установленных законодательством норм времени. Уровень качества определяется удельным весом ложноположи-

¹ Pesapane F., Codari M., Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine // Eur Radiol Exp. 2018. Oct 24. Vol. 2, №1. P. 35. DOI: 10.1186/s41747-018-0061-6.

² Lu W., Hao W., Chen X. et al. Towards standardized premarket evaluation of computer aided diagnosis/detection products: insights from FDA-approved products // Expert Review of Medical Devices. 2020. Vol. 17, №9. P. 899–918. DOI:10.1080/17434440.2020.181356; ACR: URL: <https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms>.

тельных и ложноотрицательных результатов, обуславливающих клинически значимые расхождения.

2. Процесс: первичное или двойное описание врачом-рентгенологом результатов лучевого исследования, выполненного:

- в рамках массового профилактического осмотра (скрининга);
- при наличии высоких шансов выявления онкологической патологии (исходя из анатомической области);
- при оказании экстренной и неотложной медицинской помощи.

3. Функции:

- анализ соответствия номенклатуры заявленной услуги фактическому изображению;
- технический анализ диагностической ценности изображения;
- приоритизация исследований в автоматизированном рабочем месте врача для первоочередной интерпретации результатов с риском патологии;
- локализация патологических находок на диагностическом изображении с предоставлением информации о вероятности наличия патологии (а также иных данных в рамках концепции объяснимости решений ИИ);
- морфометрия патологических находок;
- сравнение результатов исследований пациента, выполненных в динамике;
- формирование проекта описания результатов исследований в виде структурированного документа и с использованием принятых классификаций;
- предоставление объяснений и подтверждений принятых решений, в том числе в соответствии с принципами доказательной медицины.

4. Приоритеты модальностей:

- компьютерная томография (в том числе низкодозная);
- маммография;
- рентгенография;
- магнитно-резонансная томография.

5. Приоритеты локализаций:

- органы грудной клетки;
- молочная железа;
- центральная нервная система.

6. Приоритеты нозологий:

- злокачественные новообразования;
- социально значимые инфекционные заболевания (туберкулез, новая коронавирусная инфекция);
- состояния, требующие экстренной или неотложной помощи;
- заболевания сердечно-сосудистой системы (в том числе их предикторы);
- иные социально значимые заболевания.

На текущем (начальном) этапе должны быть научно обоснованы методологии и сформирована доказательная база безопасности и результативности применения технологий искусственного интеллекта в лучевой диагностике. Это создаст основу для дальнейшего развития, появления новых этапов за счет расширения перечня процессов (в том числе включения участков работы рентгенолаборантов, направляющих лечащих врачей), модальностей, локализаций, нозологий.

Глава 4. КЛАССИФИКАЦИЯ И МЕТОДОЛОГИЯ ПОДГОТОВКИ ЭТАЛОННЫХ НАБОРОВ ДАННЫХ

Философы-футуристы обычно называют данные «топливом» для технологий искусственного интеллекта. Мы не согласны с этим утверждением и вот почему. Хаотичные массивы необработанных так называемых «сырых» данных непригодны для создания качественного ИИ в сфере медицины, так же как практически непригодна методология машинного обучения «без учителя» в силу колоссальных рисков и высочайшей цены ошибки, которые таит в себе сфера здравоохранения. В медицине данные требуют предпроцессинга, стандартизации, разметки — только тогда появляется возможность с их помощью разработать алгоритм для эффективного решения конкретной клинической задачи.

Сферы здравоохранения, медицины, здоровьесбережения содержат потенциально неисчислимые объемы информации. В качестве минимальной единицы измерения количества информации здесь можно использовать эксабайт. Однако наличие таких ресурсов совсем не гарантирует реализуемость востребованного искусственного интеллекта. Колоссальные массивы данных бесполезны без научно обоснованной постановки клинических задач и описания клинического контекста применения ИИ.

Принцип «Больше данных — больше интеллекта» в чем-то созвучен утверждению «Врач учится всю жизнь». Тем не менее с учетом создания конкретного продукта для здравоохранения необходимости в доступе к бесконечно пополняемому массиву данных нет. На этапе планирования разработки должны определяться целевые метрики качества и точности будущего алгоритма; соответственно определяется и объем данных, требуемый для их достижения. Таким образом, по нашему мнению, данные — это не «топливо», а «двигатель» искусственного интеллекта, требующий целеполагания, планирования, подготовки и управления.

Дополняя положение из Национальной стратегии развития искусственного интеллекта на период до 2030 года¹, можно сказать, что для поиска интеллектуальной вычислительной системой в сфере здравоохранения непредвзятого доказательного решения требуется ввести репрезентативный, релевантный и корректно размеченный набор биомедицинских данных. В соответствии с вышеназванным документом приняты следующие определения².

Набор данных — совокупность данных, прошедших предварительную подготовку (обработку) в соответствии с требованиями законодательства Российской Федерации об информации, информационных технологиях и о защите информации и необходимых для разработки программного обеспечения на основе искусственного интеллекта.

Разметка данных — этап обработки структурированных и неструктурированных данных, в процессе которого данным (в том числе текстовым документам, фото- и видеоизображениям) присваиваются идентификаторы, отражающие тип данных (классификация данных), и (или) осуществляется интерпретация данных для решения конкретной задачи, в том числе с использованием методов машинного обучения.

Медицинские данные — обширное понятие, включающее в себя разнообразные типы документов, возникающих в ходе медицинской деятельности. Несмотря на то, что понятие «медицинские данные» не имеет отдельного упоминания в Федеральном законе «Об основах охраны здоровья граждан в Российской Федерации», о нем косвенно сказано в ст. 13 упомянутого Федерального закона, а именно — в контексте соблюдения врачебной тайны. Под медицинскими данными в статье понимаются сведения о факте обращения гражданина за оказанием медицинской помощи, состоянии его здоровья и диагнозе и иные сведения, полученные при его медицинском обследовании и лечении³. В контексте своей работы под медицинскими данными, в первую очередь, мы понимаем результаты клинических, лабораторных и инструментальных исследова-

¹ Указ Президента Российской Федерации от 10.10.2019 №490 «О развитии искусственного интеллекта в Российской Федерации».

² Там же.

³ Ст. 17 Федерального закона от 21.11.2011 №323-ФЗ (ред. от 31.07.2020)

ний в виде текстовых, графических изображений, видеорядов и проч.

Во многих клинических и лабораторных исследованиях медицинские данные играли и играют большую роль с момента начала применения научного метода в медицине. Эта роль сводилась к пониманию наборов данных как вспомогательного материала для проведения аналитических расчетов в целях проверки или опровержения поставленной гипотезы⁴.

С приходом цифровых способов обработки больших массивов данных появилась потребность в правильном структурировании и подготовке данных, так называемой машиночитаемости. Это связано с тем, что исходные причины их возникновения не предполагали ретроспективной автоматической обработки: медицинские данные служат целью оказания медицинской помощи конкретному пациенту (от постановки правильного диагноза до назначения и проведения эффективного лечения). Однако первично цифровой характер многих результатов инструментальных исследований, а также активное внедрение современных систем диагностики и лечения, использующих такие данные, способствует пересмотру концепции медицинских данных как вспомогательных инструментов для оказания медицинской помощи⁵.

Для задач обучения нейронных сетей, входящих в понятие технологий искусственного интеллекта, хорошо структурированные данные являются ключевым фактором успеха, поскольку результаты работы алгоритмов при анализе новых, ранее не встречаемых медицинских данных, напрямую зависят от того, на каких наборах данных были обучены нейронные сети. На сегодняшний день алгоритмы искусственного интеллекта уже активно используются во многих аспектах бытовой жизни (например, персональные голосовые помощники) и профессиональной деятельности (в авиации, логистике и др.)⁶.

«Об основах охраны здоровья граждан в Российской Федерации» (с изм. и доп., вступ. в силу с 01.09.2020).

⁴ Dekker R. The importance of having data-sets // IATUL Annu. Conf. Proc. 2006.

⁵ Cobb A.N., Benjamin A. J., Huang E. S. et al. Big data: More than big data sets // Surgery. 2018. №164. P. 640–642.

Завершающим этапом разработки алгоритмов искусственного интеллекта до практического применения является их тестирование. Для этих целей обычно используется подвыборка из основного набора данных для обучения (например, 20%), которая изолируется от остальных 80% и не используется в процессе обучения. В случае независимого тестирования (например, для проведения сравнения результатов анализа между несколькими алгоритмами) может использоваться отдельный эталонный набор данных, в котором тестирующему заранее известны категории каждой из единиц данных (например, отношение единицы данных к категориям «норма» и «патология»). Для классической задачи бинарного классификатора часто используются сбалансированные «датасеты», в которых две категории представлены в равных частях⁷.

Другое назначение наборов данных — техническая проверка алгоритмов на правильность обработки исследований одной модальности, полученных на аппаратах различных производителей. В связи с особенностями различных моделей входные данные для алгоритма могут отличаться как по техническим характеристикам (например, размер матрицы изображения), так и по передаваемой вместе с ними метаданной (например, перечень DICOM-тегов).

Набор данных (синоним-жаргонизм «датасет» от англ. dataset) отличается от простого сбора медицинских данных тем, что он наделен особыми качествами, к которым относятся: унификация и структурированность данных; отсутствие грубых неточностей, ошибочных исследований; наличие дополнительной информации (категории и значения признаков или характеристик элементов данных); наличие сопроводительной документации.

Существует ряд трудностей, с которыми можно столкнуться при составлении медицинских наборов данных:

1. Вопросы этического характера: деперсонализация данных, «серая зона» законодательства о персональных данных применительно к медицинским данным, страх раскрытия врачебной тайны.

⁶ Mintz Y., Brodie R. Introduction to artificial intelligence in medicine // *Minim. Invasive Ther. Allied Technol.* 2019. Vol. 28, №2. P. 73–81.

⁷ Wei Q., Dunbrack R.L. The Role of Balanced Training and Testing Data Sets for Binary Classifiers in Bioinformatics // *PLoS One.* 2013. July 9.

2. Вопросы политического характера: вопрос трансграничной передачи данных.

3. Сложность получения самих данных: для получения медицинских данных требуется пройти весь процесс от начала регистрации пациента, проведения диагностики и исследования, где участвует медицинский персонал и используется медицинское оборудование, до составления заключения и передачи этих данных в единую систему хранения медицинских данных.

4. Сложнейшая структура организации данных с рядом специфических параметров, которые необходимо фильтровать для использования ИИ (в отличие от банковских транзакций, баз данных спамерских атак, распознавания номеров автомобилей и других наборов данных).

5. Разметка данных: может быть осуществлена экспертной группой врачей. Стоит отметить наличие данных, согласно которым в определенных случаях возможна разметка краудсорсингом (снижение вероятности ошибки путем усреднения).

6. Репрезентативность: сформированная выборка отражает общие характеристики популяции (генеральной совокупности). Репрезентативность также может относиться к сбалансированному соотношению элементов по каждой выбранной категории (используемое оборудование, классы разметки и т.п.). При корректно сформированной выборке выводы и заключения по исследованию будут верны и могут быть перенесены на уровень популяции¹.

Одной из основных сложностей, связанных с подготовкой наборов данных для машинного обучения и валидации алгоритмов искусственного интеллекта, является разнообразие характеристик, меняющихся от задачи к задаче. К **основным характеристикам наборов данных** относятся:

1. Формат данных. Может различаться во входных и выходных данных. Ко входным относятся данные, служащие материалом для составления набора. Примерами форматов входных данных лучевой диагностики служат изображения магнитно-резонансной томографии (МРТ), компьютерной томографии (КТ), рентгенографии (РГ) формата DICOM, клинические данные в текстовом формате, разметка изображений формата json и другие.

¹ Ramsey C.A., Hewitt A.D. A methodology for assessing sample representativeness // Environ. Forensics. 2005. Vol. 6, №1. P. 71–75.

Выходные данные — продукт готового набора. Формат выходных данных варьируется, может быть оптимизирован под задачу и заданные условия.

2. Объем набора данных, количество занимаемого пространства на диске.

3. Размер набора данных, показывающий количество единиц данных, содержащихся в наборе. Единица данных устанавливается в зависимости от задачи и принципа структурирования данных в наборе. В случае данных лучевой диагностики таким примером может стать одно исследование или одна серия изображений исследования, так как представляет собой минимальную самостоятельно информативную единицу данных.

4. «Сырые» и агрегированные данные. «Сырые» данные — это необработанные данные медицинских исследований в своей исходной форме (без обработки и предобработки, без применения фильтров и других методов анализа). Агрегированные данные — это сырые данные, прошедшие какую-либо форму обработки, анализа и прочих манипуляций.

5. Тип набора данных (его логическое представление). Набор может быть представлен в виде непрерывной последовательности записей, в виде отдаленных друг от друга членов и т. п.

6. Формат хранения набора данных.

7. Доступ к набору данных. Может быть организован как открытый, так и закрытый доступ.

8. Прочее. У каждого типа данных есть свой набор спецификаций и характеристик.

Вследствие широкого разнообразия характеристик данных, методов их сбора и способов организации в единый набор появляется большое множество наборов данных, каждый из которых может радикально отличаться от другого, что усложняет их поиск и применение. Особенно это касается случаев, где есть необходимость в большом количестве входных данных для обучения или, как еще один пример, необходимость в дополнительном наборе данных с той же спецификацией, но без повторения элементов набора, на которых алгоритм уже был обучен. Кроме того, обучение и развитие алгоритмов с искусственным интеллектом напрямую зависит от качества набора данных.

Под **качеством** набора принято понимать его структурированность, однородность, репрезентативность, сбалансированность

по классам, отсутствие выпадающих значений, наличие экспертной разметки (при необходимости), наличие описания модели данных и документации. Кроме того, в лучевой диагностике наблюдается спрос и нехватка открытых наборов данных для обучения, а также эталонных данных для валидации и калибровки (в ограниченном доступе) алгоритмов ИИ. Вышеперечисленные проблемы решаются путем стандартизации процесса сбора, обработки и проверки набора данных, чтобы упростить их использование и повысить эффективность для исследований в области машинного обучения.

Источниками для подготовки набора данных могут являться:

1. Открытые наборы данных (доступные для переработки).
2. Базы данных, к которым есть доступ.
3. Иные наборы данных, права использования которых позволяют создавать на их основе новые результаты интеллектуальной деятельности (например, полученные в рамках коммерческого договора на оказание услуг по разметке набора данных).

§1. Требования к эталонным наборам данных

Для обучения, внутренней и внешней валидации, клинико-технических и клинических испытаний технологий искусственного интеллекта применяют эталонные данные — наборы биомедицинских данных. Требования к таким наборам разработаны нами в ходе аналитических и научных исследований¹.

В лучевой диагностике набор эталонных данных — это упорядоченная совокупность:

¹ Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика) / сост. С. П. Морозов, А. В. Владзимирский, В. Г. Кляшторный [и др.] // Сер. «Лучшие практики лучевой и инструментальной диагностики». М., 2019. Вып. 57. 51 с. (Clinical acceptance of software based on artificial intelligence technologies (radiology) / S.P. Morozov, A.V. Vladzimirskyy, V.G. Klyashtornyy [et al.] // Ser. «Best practices in medical imaging». М., 2019. Issue 57. 45 p.

— диагностических изображений одной модальности и/или однотипных медицинских документов (например, протоколов описаний результатов исследований);

— сведений о наличии, характере и локализации патологических изменений на изображениях; для текстовых документов — библиотеки ключевых слов, словосочетаний и их критичных сочетаний;

— сведений о верификации диагноза (опционально).

Информация о наличии, характере и локализации патологических изменений (в том числе в соответствии с Международной классификацией болезней) может быть подтверждена объективно — в таком случае набор данных именуется **верифицированным**. Подробнее об этом будет сказано далее.

Размер эталонного набора данных (**математически — размер выборки**) для оценки диагностических характеристик алгоритма ИИ устанавливается с помощью методик определения размера выборки для пропорций. Для каждого конкретного случая (клинического сценария) должны быть заданы желаемые параметры точности оценки диагностических характеристик алгоритма ИИ; необходимо задать ожидаемую пропорцию и желаемую точность (ширину доверительного интервала) ее оценки. Надо учитывать, что чем выше желаемая точность оценки (т.е. чем уже доверительный интервал) и чем ближе пропорция к 50%, тем большее количество размеченных исследований понадобится включить в набор данных.

Эталонный набор данных должен быть проверен профильной медицинской научно-исследовательской организацией на предмет полноты и качества содержащейся в нем информации. Рекомендуется при проведении клинических испытаний применять эталонные наборы данных, имеющие государственную регистрацию в качестве базы данных.

Эталонный набор данных должен содержать такие *сведения* (описательного характера):

1. Номер свидетельства о государственной регистрации базы данных (рекомендательно).

2. Характеристика популяции (гендерно-возрастные показатели, этнический состав, регионы проживания и т.д.); сведения о деперсонализации; сведения о медицинских организациях, послуживших источниками для формирования набора данных.

3. Характеристика исследований: анатомическая область (и), модальность, проекции, типы медицинских изделий — диагностических приборов, виды и характеристики протоколов исследований.

4. Целевая патология в соответствии с Международной классификацией болезней.

5. Общее количество клинических случаев, исследований, изображений, документов и их распределение по диагностическим группам.

6. Соотношение случаев «норма»: «патология» (случаи «патология» могут быть разделены на несколько подклассов).

7. Сведения о верификации (патогистологическом или ином окончательном диагнозе).

8. Методология разметки (ранее представленной в виде научных публикаций, методических рекомендаций или патентов).

Эталонный набор данных должен отвечать таким *требованиям*:

1. Структура набора данных должна соответствовать поставленной цели его формирования (решаемой клинической задаче).

2. Включаемые данные должны быть получены из реальной клинической практики (не допускается получение синтезированных данных, применение фильтров и математических средств постобработки при выгрузке данных из медицинских информационных систем).

3. Соотношение «норма»: «патология» должно соответствовать распространенности целевой патологии в популяции.

4. При формировании эталонного набора должны использоваться данные из нескольких медицинских организаций.

5. Демографические, социально-экономические характеристики и основные показатели здоровья пациентов, чьи данные включаются в эталонный набор данных, должны соответствовать усредненным характеристикам популяции территории, на которой планируется использование ИИ.

6. Обязательно включение в набор исследований, выполненных с нарушением методики, содержащих дефекты, а также — исследований аналогичной модальности, но иной анатомической области.

7. Количество включенных элементов (клинических случаев, наблюдений, исследований) должно быть достаточным для достижения статистической значимости результата работы алгоритмов ИИ.

8. Планируемый размер эталонного набора данных должен быть обоснован в протоколе исследования, исходя из статистических соображений и желаемой точности оценки основных метрик, указанных выше.

9. Разметка должна быть проведена с использованием стандартизированной терминологии — т.н. тезауруса (кодированной библиотеки типовых формулировок, соответствующих нормативно-правовой документации, клиническим рекомендациям или рекомендациям профессиональных врачебных ассоциаций).

10. Подготовка и разметка должны быть проведены подготовленными техническими и медицинскими специалистами, имеющими соответствующие навыки и компетенции.

11. Эталонные наборы данных, используемые в процессе клинических испытаний для регистрации ПО в качестве медицинского изделия, не должны быть публично доступны (для исключения возможности обучения алгоритмов ИИ на эталонных наборах данных).

§2. Классификация наборов данных

Наборы данных для обучения и тестирования алгоритмов искусственного интеллекта можно классифицировать различными способами. Например, выделяют наборы со структурированными, частично структурированными и неструктурированными данными.

Мы рекомендуем разделять «датасеты» по источникам их формирования, условиям использования, типам биомедицинских и клинических данных, по временным характеристикам, файловой структуре, наконец — по видам задач, для решения которых наборы сформированы и т. д. Однако, изучив имеющиеся подходы с позиций их практического применения в лучевой диагностике, мы пришли к выводу об их несостоятельности. Предложенные классификации носили либо чересчур абстрактный характер, либо имели значение для математиков-разработчиков, но были неприменимы врачами. Некоторые систематики слишком специализированы, посвящены отдельным, относительно «узким» задачам.

В итоге было принято решение о создании совершенно иной классификации наборов данных для лучевой диагностики, которая должна была отличаться:

- клинической ориентированностью,
- универсальностью,
- простой применения.

Используя аналитические методы научной работы, мы предложили две классификации: по диагностической ценности и по функциональному назначению.

В таблице 6 приведена классификация наборов данных **по диагностической ценности**.

Таблица 6 – Классификация эталонных наборов данных в сфере лучевой диагностики по диагностической ценности

		А	В	С
		Пиксельная маска	Координаты области	Метаданные
1	Подтвержденный диагноз	Результаты подтверждающих исследований	Результаты подтверждающих исследований	Результаты подтверждающих исследований
2	Классификация находок	Стандартизованная шкала/ классификация	Стандартизованная шкала/ классификация	Стандартизованная шкала/ классификация
3	Наличие находок	Бинарная оценка	Бинарная оценка	Бинарная оценка

Классификация предполагает разделение наборов на три вида (1, 2, 3) и три класса (А, В, С).

Вид подразумевает типовой способ верификации:

1. Бинарная оценка факта наличия или отсутствия целевой патологической находки.

2. Классификация целевой патологической находки в соответствии со стандартизованными клинко-рентгенологическими классификациями, шкалами, системами.

3. Наличие данных о верификации природы целевой патологической находки.

Класс подразумевает типовой способ отображения патологической находки в результатах лучевого исследования:

1. Информация о наличии/отсутствии целевой патологической находке содержится в метаданных (то есть в аннотации – сопроводительных файлах), отсутствует на изображении.

2. Информация об имеющейся целевой патологической находке представлена в виде координат. Может помещаться в метаданные (аннотация, сводный табличный сопроводительный файл) и/или присутствовать на изображении в виде отметки области расположения простой геометрической фигурой.

3. Информация об имеющейся целевой патологической находке представлена на изображении в виде пиксельной маски (окоптуренной области изображения), дополнительно может содержаться в метаданных (в аннотации).

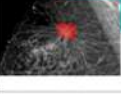
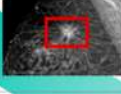
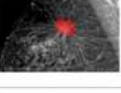
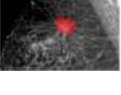
КЛАССЫ РАЗМЕТКИ	ПРОСПЕКТИВНАЯ		РЕТРОСПЕКТИВНАЯ
	А	В	С
	Пиксельная маска	Координаты области	Метаданные
1 Подтвержденный диагноз	 РМЖ (ДАННЫЕ ГИСТОЛОГИИ)	 РМЖ (ДАННЫЕ ГИСТОЛОГИИ)	 РМЖ (ДАННЫЕ ГИСТОЛОГИИ)
2 Классификация находок	 BI-RADS 2	 BI-RADS 2	 BI-RADS 2
3 Наличие находок	 'НЕТ ОЧАГОВ'/ 'ЕСТЬ ОЧАГИ'	 'НЕТ ОЧАГОВ'/ 'ЕСТЬ ОЧАГИ'	 'НЕТ ОЧАГОВ'/ 'ЕСТЬ ОЧАГИ'

Рисунок 9 – Классификация наборов данных для обучения и тестирования алгоритмов ИИ, предназначенных для скрининга ЗНО молочной железы

Классификация может применяться в отношении наборов данных для любых клинических задач в сфере лучевой диагностики. Она не зависит от типов (модальности) диагностических данных, но вместе с тем четко отображает взаимосвязь между собой: а) объемов и качества исходных данных; б) трудозатрат на подготовку; в) методик разметки и работы с первичными данными; г) диагностической ценности. На рис. 9 приведен пример использования

классификации в отношении наборов данных для обучения алгоритмов ИИ в целях скрининга злокачественных новообразований молочной железы.

В соответствии с классификацией минимальный вариант набора (C3) данных представляет собой результаты цифровой маммографии с информацией «Есть очаги» / «Нет очагов» в дополнительном файле (электронной таблице). Максимальный вариант (A1) — это результаты цифровой маммографии, на которых пиксельной маской отмечены патологические очаги, а в дополнительных файлах содержатся результаты патогистологического исследования материала, полученного при биопсии или хирургическом вмешательстве. Очевидно, что с учетом трудозатрат вариант C3 можно подготовить простой выгрузкой данных из медицинской информационной системы (т.н. ретроспективная разметка — подробнее об этом будет сказано далее). Однако этот же вариант наименее ценен с диагностической точки зрения; объяснимость решений алгоритма ИИ, разработанного на его основе, будет сомнительной. Вариант A1 требует длительного сбора данных, скорее всего из различных источников, информационных систем; также необходима непосредственная работа квалифицированных врачей-рентгенологов по нанесению пиксельной маски на изображения согласно определенной методике. Впрочем, именно вариант A1 позволит алгоритму не просто научиться отличать норму от болезни, но четко указывать локализацию и тип патологической находки (вплоть до отделения потенциально злокачественных находок от потенциально доброкачественных). «Золотая середина» для клинической задачи скрининга — это набор данных B2. В таком случае на изображении выделены области расположения патологических очагов, а само исследование классифицировано по стандартизированному инструменту BI-RADS. Следовательно, алгоритм ИИ сможет не просто выбрать исследования с патологией, но сразу сформировать предложения по дальнейшей маршрутизации пациентки.

Вторая классификация наборов данных — ***по функциональному назначению*** (таблица 7).

Таблица 7 – Классификация эталонных наборов данных в сфере лучевой диагностики по функциональному назначению

Название	Базовое количество исследований	Назначение
I. Общие наборы		
Самостоятельный тест (селф-тест)	10–20	Проверка на техническое соответствие модели
Клинический тест	100–200	Валидация и калибровка модели
Дообучение	1000	Дополнительное обучение готовой модели
Машинное обучение	>5000	Обучение новых моделей и решение новых клинических задач
II. Специализированные наборы		
Динамический	20–100	Оценка динамических эффектов
Специализированные технологические дефекты	10–30	Оценка дефектов

Медицинские данные подразделяются на несколько подмножеств, каждое из которых является важным компонентом в обучении, оценке качества алгоритмов ИИ и используется для других прикладных и фундаментальных задач в сфере искусственного интеллекта для здравоохранения. Каждый компонент (подмножество, набор) данных направлен на решение определенной задачи.

Обе предложенные классификации успешно применяются для подготовки эталонных наборов данных в службе лучевой диагностики города Москвы.

§3. Методология формирования наборов данных

Общие принципы. Формирование наборов данных состоит из двух взаимосвязанных процессов подготовки и разметки (аннотирования) данных.

Подготовка данных — процесс выгрузки структурированных и неструктурированных данных из медицинских информационных систем по заданным критериям (фильтрам).

Разметка (аннотация) данных — обработка структурированных и неструктурированных данных, формируемых ретро- или проспективно, в процессе которой данным (в том числе диагностическим изображениям) присваиваются идентификаторы, отражающие тип данных (классификация данных), и/или вносится информация об интерпретации данных для решения конкретной клинической задачи.

Следовательно, разметка — это процесс определения значения признаков или характеристик у элемента данных в наборе. В лучевой диагностике обычно таким «элементом» является диагностическое изображение, одно или в виде серии. На основе разметки становится возможным провести классификацию элементов, отнести их к той или иной группе. А под словами «определение значения» чаще всего понимают кластеризацию и классификацию исследований по группам (наличие или отсутствие рентгенологических признаков выбранного заболевания — целевой патологии), а также графическое обозначение области интереса, соответствующей искомым признакам (например, очаги демиелинизации при рассеянном склерозе на МР-изображениях головного мозга).

Для разметки может использоваться как уже имеющаяся на момент отбора исходных данных информация (ретроспективная подготовка данных), так и разметка, сделанная специалистом уже после этапа отбора (проспективная подготовка данных). Таким образом, подготовка наборов данных может производиться ретроспективно или проспективно.

Ретроспективная подготовка — это сбор элементов (результатов лучевых исследований) в соответствии с указанными метаданными, перечень которых выбирают в соответствии с поставленной клинической задачей. Ретроспективная подготовка обеспечивает формирование наборов данных класса С. Такую подготовку проводят путем выгрузки данных из медицинской информационной системы по предустановленным фильтрам. Этот подход не предполагает какой-либо обработки элементов, то есть аннотирования, анализа, нанесения меток и проч. на диагностические изображения. Для каждого элемента набора данных устанавливают

его соответствие медицинской информации (протоколу описания, диагнозу согласно принятой классификации, результатам лабораторных, инструментальных или иных лучевых исследований и т.п.). Эта подготовка не требует участия врача, а может быть выполнена техническим специалистом, который имеет опыт работы с наборами данных.

Пример ретроспективной подготовки: набор данных — результаты лучевых исследований пациентов с подтвержденной коронавирусной болезнью. Перечень метаданных в таком случае будет следующим: идентификационный номер, дата рождения, дата выполнения лучевого исследования, результаты иммунологических исследований и т. п. Значит, при ретроспективной подготовке суть разметки состоит в соотношении результатов лучевого исследования с некими данными из медицинской документации пациента без критического или экспертного анализа такого соответствия. В этом процессе могут быть использованы данные сопроводительных документов (например, тексты заключений для результатов инструментальных исследований), медицинских информационных систем, электронных медицинских карт, а также метаданные, генерируемые автоматически диагностическим устройством при проведении исследования и хранящиеся в исходных данных. Очевидное преимущество ретроспективного подхода заключается в значительно меньших временных затратах со стороны врачей, так как большую часть работы выполняет специалист по работе с данными.

Проспективная подготовка — это сбор элементов (результатов лучевых исследований) в соответствии с поставленной клинической задачей, а также проведение дополнительных манипуляций с элементами, то есть разметки, аннотирования. Проспективная подготовка обеспечивает формирование наборов данных классов А и В.

Пример разметки — непосредственно на изображение наносятся метки, аннотирования — определяются координаты и морфометрические характеристики объектов. Проспективная подготовка предполагает активное вовлечение обученного медицинского персонала в процесс аннотирования содержания данных или их частей набора. Этот процесс может происходить в графической или текстовой формах либо в их комбинации. Фактически он означает «насыщение» набора данных дополнительной информацией, позволяющей разделить его элементы на категории. Степень вовлечения медицинского персонала может быть более или менее затратной: в первом слу-

чае экспертам предлагается обвести контур области интереса, во втором — обозначить ее координаты простой геометрической фигурой.

Обязательный подготовительный этап перед проспективной подготовкой — выбор конкретной методики разметки (аннотирования), определение объемов, критериев, программного обеспечения для этой работы, подготовка инструкций для врачей-разметчиков, формирование и обучение команды. В процесс подготовки такой инструкции, по возможности, должна быть вовлечена та рабочая группа, которая определяла клиническую задачу на этапе планирования. Надо подчеркнуть, что в зависимости от клинической задачи и технических условий в одном наборе могут комбинироваться данные, подготовленные ретро- и проспективно.

Как видим, набор данных содержит результаты диагностических исследований, при этом наличие целевой патологии определяется извлечением информации из сопутствующей медицинской документации и/или интерпретацией изображений врачом-рентгенологом, осуществляющим разметку. В целом такую разметку можно назвать *экспертной*. Набор данных может быть дополнен информацией не только о наличии целевой патологии, но и об объективном подтверждении этого факта, то есть о верификации первичной диагностической гипотезы. Соответственно, такой набор называют **верифицированным**.

Верифицированный набор данных — это эталонный набор данных для обучения или тестирования алгоритмов искусственного интеллекта, подготовленный проспективно и содержащий данные из медицинской документации об окончательном и/или патологоанатомическом диагнозе.

Верификация осуществляется путем внесения в набор данных результатов (в порядке убывания диагностической ценности):

- патологоанатомических, патогистологических и цитологических исследований;
- лабораторных исследований (прежде всего, молекулярно-биологических, серологических, иммунологических);
- иных инструментальных исследований (например, эндоскопических);
- результатов динамического наблюдения (повторных исследований аналогичной модальности, выполненных через установленный интервал времени);

— открытого консенсуса не менее 3-х врачей-экспертов по каждому отдельному случаю;

— слепого анализа набора данных несколькими экспертами с достижением заданного уровня согласованности их решений.

Условно выражаясь, верификация обеспечивает степень «доверия» набору данных со стороны разработчиков или специалистов по оценке работы интеллектуальных систем.

В тех случаях, когда экспертное мнение является наиболее весомым фактором при определении значений признаков или характеристик данных, разумным решением будет проведение одновременного чтения исследования двумя независимыми экспертами. В случае несогласованности между двумя экспертами спорные исследования направляются третьему, более квалифицированному эксперту (по практическому опыту, наличию степени либо другим критериям). Исследования, оставшиеся спорными после трех экспертов, могут быть признаны «некачественными» и исключены из набора данных.

Оптимальный и наиболее ценный вариант — это верификация диагноза патологоанатомическими, клеточными и лабораторными методами. Консенсус и слепой анализ, как правило, приводят к исключению значительного количества выгруженных элементов (результатов исследований) из окончательного набора. По собственному опыту авторов до 25,0% исследований могут оказаться спорными после двух независимых чтений; до 4% могут остаться таковыми даже после чтения третьим, более квалифицированным экспертом (медицинский стаж которого более 5 лет).

Организация работы и контроль качества. Формирование набора данных должно быть тщательно спланировано, обеспечено ресурсами и эффективным управлением. Обязателен мониторинг процессов подготовки и разметки (аннотирования) для непрерывного улучшения качества.

Врачей-разметчиков рационально объединять в группы. В группе должны входить специалисты с достаточным опытом интерпретации и описаний результатов определенного вида исследования (обычно — от трех лет и выше); идеальный вариант — это наличие субспециализации. Группой должен руководить модератор, который не принимает непосредственного участия в разметке, но определяет и регулирует срочность, очередность и объем работы

между разметчиками. Также ключевая задача модератора — это контроль качества, в частности:

- проверка соблюдения требований к набору данных в соответствии с поставленной задачей и характеристиками эталонного набора;
- проверка отсутствия пропусков элементов в наборе данных;
- проверка отсутствия некорректных элементов для решения поставленных задач;
- проверка качества элементов и манипуляций с ними (разметки, аннотирования) в соответствии с установленными критериями.

Подготовка и разметка данных в лучевой диагностике. В лучевой диагностике стандартный набор данных состоит из диагностических изображений в формате DICOM, отобранных в соответствии с клинической задачей. Диагностическое изображение, включенное в набор, может содержать информацию о целевой патологической находке и иные сведения в метаданных или в сопроводительном файле — аннотации. Существуют такие виды аннотаций:

1. Полуструктурированное текстовое описание визуальных наблюдений с указанием содержащих их анатомических объектов и типов нарушений.

2. Структурированное описание с использованием стандартизированной терминологии (тезауруса).

3. Сводный табличный файл для всех изображений, включенных в набор, с координатами (x, y, z) и морфометрическими данными целевых патологических находок.

Конкретная информация о локализации целевых патологических находок может быть представлена с разным уровнем точности и детализации:

- приблизительное обозначение координат посредством задания ограничивающего параллелепипеда, эллипсоида или иной простой геометрической фигуры либо их сочетания;
- полная сегментация на основе маски минимальных элементов, обозначающей положение целевой патологической находки на фоне остальной части изображения.

Пример. Методика разметки результатов компьютерной томографии органов грудной клетки для клинической задачи: скрининг злокачественных новообразований легкого путем выявления очагов более 5 мм в диаметре¹.

1. Действия врача:

1.1. Просматривает срезы в аксиальной, фронтальной и сагитальной проекциях (допустимо использование режима Maximum intensity projection для более четкого различения очагов на фоне кровеносных сосудов).

1.2. В случае выявления очага на одной из проекций отмечает при помощи мыши сферу, ограничивающую все пораженные ткани.

2. Важные функциональные требования к программному обеспечению:

- нанесенная сфера должна синхронно отображаться на всех трех проекциях;

- должна быть возможность нанести метку на одной проекции, а затем внести в нее изменения (расширить или переместить) на другой;

- для каждого обнаруженного и отмеченного очага автоматически формируются пространственные координаты сферы и ее диаметр.

Примечание. В методике используется научно обоснованная кластерная модель определения локализации очаговых образований. Для выбора оптимальной геометрической фигуры был проведен численный эксперимент, в котором установлено соотношение между точностью и скоростью разметки при аппроксимации различными фигурами. Доказано, что разметка сферами обладает наименьшей сложностью и лишь незначительно уступает по точности разметке эллипсоидами. Для очагов простой формы (как правило, близкой к сферической) достаточно указать координаты центра очага и диаметр сферы, охватывающей весь очаг и часть прилегающих к нему здоровых тканей. Для объектов сложной формы (вытянутых вдоль какого-то направления или состоящих из конгломерата сферических очагов) формируется так называемый кластер.

Кластер — это покрытие очага несколькими сферами, каждая из которых захватывает какую-то часть данного объекта. Эти сферы, пересекаясь друг с другом, фиксируют локализацию всей опухоли. Пересекающиеся между собой сферы, объединенные в кластер, считаются относящимися к одному образованию. Условием

¹ Кульберг Н. С., Гусев М. А., Решетников Р. В. [и др.]. Методология и инструментарий создания обучающих выборок для систем искусственного интеллекта по распознаванию рака легкого на КТ-изображениях // Здоровоохранение Российской Федерации. 2020. Т. 64. №6. С. 343–350.

отнесения группы сфер к одному кластеру является расстояние между центрами, не превышающее диаметра сферы. Фактически кластер является массивом записей о сферических отметках, допускающим несколько вариантов представления.

3. Действия врача:

3.1. Вводит параметры обнаруженного и отмеченного очага:

А. Обозначение типа текстуры очага:

- со́лидный (очаг типичной структуры локального уплотнения округлой формы мягкотканой плотности с различными контурами);
- полусо́лидный (очаг имеет более плотный участок в центре и зоны низкой плотности по типу «матового стекла» по периферии);
- по типу «матового стекла» (очаг характеризуется незначительным повышением плотности легочной ткани, с сохранением видимости сосудов и бронхов в зоне патологического процесса).

В. Злокачественность очага:

- «да»,
- «нет» (для фиброзных уплотнений, кист и прочих доброкачественных образований).

4. Действия эксперта:

- просматривает ранее сделанные врачом-разметчиком отметки и отвечает на стандартные вопросы по каждой из них (три варианта ответа: «Согласен с отметкой», «Согласен частично», «Не согласен»);
- при выборе «Согласен частично» в поле описания добавляет информацию о причинах разногласий.

§4. Процесс создания наборов данных

На рис. 10 изображена схема жизненного цикла набора данных от независимых друг от друга исследований до конечного продукта.

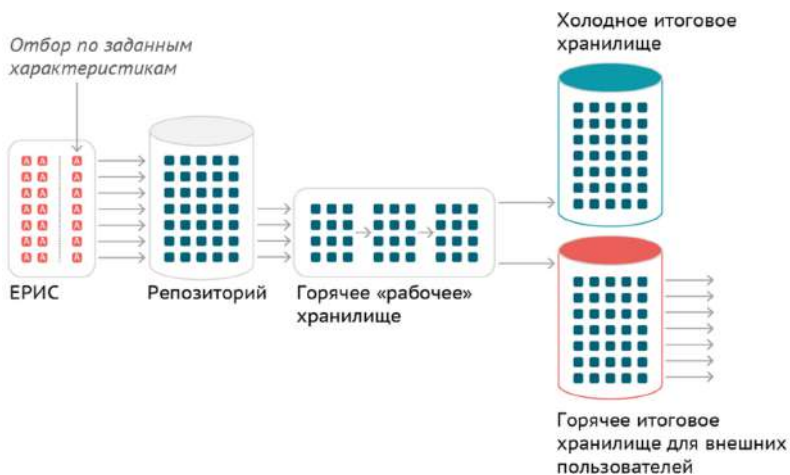


Рисунок 10 – Жизненный цикл набора данных (красным обозначены исследования, содержащие персональные данные, синим – анонимизированные; в качестве основного источника входных рентгенологических данных служит ЕРИС, для хранения данных используется стандарт DICOM)

Процесс подготовки набора данных состоит из 5 основных этапов:

- планирование;
- документирование;
- сбор данных;
- разметка и верификация;
- хранение и предоставление доступа.

Каждый этап, в свою очередь, состоит из совокупности процедур, выполнение которых позволяет получить качественный набор данных и в конечном итоге достигнуть цели обучения и тестирования алгоритма искусственного интеллекта.

I этап. Планирование. Основные процедуры этапа:

1. Постановка клинической задачи в сфере лучевой диагностики, которая потенциально решается путем автоматизации с помощью интеллектуальных систем. Составление перечня признаков и/или характеристик исходных данных, информацию о которых планируется получать от интеллектуальной системы в процессе решения поставленной задачи и по которым возможно оценить корректность принятого системой решения, то есть формирование требований к результатам работы технологий искусственного интеллекта по решению данной клинической задачи.

2. Определение цели применения набора данных: обучение, выполнение тестирования на этапе разработки или выполнение валидации (клинических испытаний) алгоритмов ИИ.

3. Оценка необходимости получения одобрения Комитета по биоэтике для сбора данных или использования деидентифицированных данных с целью подготовки набора данных. При выявленной необходимости — получение такого одобрения.

4. Выбор вида подготовки (ретроспективный или проспективный), методов разметки (аннотации), способа верификации значений выбранных признаков и/или характеристик у элементов формируемого набора данных.

5. Определение источников данных, критериев включения, исключения и исключения данных из набора. Определение значимых характеристик данных, необходимых для оценки не только точности, но и пределов надежности и масштабируемости интеллектуальной системы.

Постановка клинической задачи — одна из наиболее ответственных задач, стоящих перед создателем набора данных. Недостаточное внимание к ней приводит к внезапно всплывающим вопросам как в процессе подготовки набора данных, так и уже при внедрении диагностического алгоритма на основе ИИ в клиническую практику. В постановке задачи должна участвовать рабочая группа из профессионалов различных профилей: врачей-клиницистов, организаторов здравоохранения, специалистов по обработке медицинских данных, инженеров-исследователей (занимающихся машинным обучением или проводящих валидацию решений на основе ИИ), а также администраторов, обеспечивающих доступ и выгрузку исходных данных. В таком случае может быть поставлена задача, имеющая клиническое, организационно-управленческое

и социально-экономическое значение для практического здравоохранения, а также потенциально качественно решаемая именно интеллектуальными технологиями.

Клиническая задача должна позволить создателям набора данных ответить на следующие вопросы:

- Какие модальности, какие процедуры, какую клиническую, демографическую и подобного рода информацию следует принимать на вход алгоритму для ее решения, и что принимать за одну единицу данных?

- Какие признаки должны определяться с помощью технологий ИИ?

- Каким нозологиям или группам нозологий относятся искомые признаки?

- Как решение задачи помогает клиническому специалисту?

- Какое количество единиц данных необходимо и достаточно для цели применения создаваемого набора данных?

Из клинической задачи логическим образом вытекают критерии, по которым интеллектуальная система принимает решение, отнести ли то или иное исследование или найденную на изображении область к группе интереса. Речь идет о так называемых **базовых диагностических требованиях** к работе ИИ. Они включают формальное описание искомых признаков на исследовании, а также составляют перечень признаков и/или характеристик, на основе которых в дальнейшем будет проводиться разметка данных в наборе. Эта информация позволяет разработчикам точнее настроить решения для определения требуемых признаков, а специалистам по подготовке набора — составить инструкцию по разметке и верификации данных.

Критерии включения и невключения чаще определяются клинической и/или практической задачей, тогда как критерии исключения обычно дополняются в ходе работы с первичными данными, так как находятся те или иные критерии, негативно сказывающиеся на структурированности и унификации набора данных. Критерии могут носить как медицинский (например, возраст — от 18 до 99 лет, наличие сохранной структуры искомого органа и т. д.), так и технический характер (фильтр КТ — мягкотканый, ядро свертки — FC51 и т. д.). Унификация данных необходима для надежной работы инструментов оценки работы решения на основе ИИ.

Существуют два основных принципа отбора данных для формирования набора:

а) включение медицинских данных (патологических находок, заболеваний, состояний, артефактов, нормальной картины), отражающее максимальную вариативность, то есть и частые, и редкие случаи представлены в одинаковом объеме;

б) представление медицинских данных согласно их частоте встречаемости, претестовой вероятности, заболеваемости, распространенности в популяции.

II этап. Сбор данных. Основные процедуры этапа:

1. Получение доступа к источнику данных.
2. Отбор исходных (первичных, «сырых») данных.
3. Перенос данных в информационную систему для разметки (аннотирования), документирования набора данных.

Медицинская организация, участвующая и/или выполняющая подготовку набора, должна обеспечить доступ к необходимым данным в медицинской информационной системе. Кроме того, нужно обеспечить выполнение операций поиска, чтения, сбора и выгрузки данных, а в идеальной ситуации — еще и обезличивания данных средствами самой медицинской информационной системы. Процесс доступа следует оптимально изложить в виде документа (регламента, соглашения и проч.), прежде всего для обеспечения защиты данных, в том числе персональной информации согласно действующему законодательству.

Подход к получению (выгрузке) исходных данных зависит от источника и способа хранения данных. Общие принципы отбора первичных данных:

- выбор как можно большего спектра исследований интересующей модальности и процедуры;
- сохранение необходимого для решения клинической задачи количества сопроводительной информации (включая текстовые документы, описывающие результаты исследования, клинический диагноз пациента, которым закончился медицинский случай, и др.);
- обезличивание исследований в информационном контуре медицинской организации (медицинской информационной системе), в которой происходит отбор данных.

На этапе отбора применяются также критерии включения и невключения исследования в будущий набор. Эта операция может проводиться как непосредственно при отборе исследований

в медицинскую информационную систему (МИС), так и сразу после выгрузки (уже вне пределов информационного контура медицинской организации). Необходимо учитывать, что этот шаг может повлечь за собой уменьшение набора данных в несколько десятков раз.

Набор данных для обучения или тестирования алгоритмов ИИ не должен содержать какую-либо персональную информацию. В соответствии с действующим законодательством любая персональная информация должна быть удалена из исходных данных и метаданных. Стандартно проводится удаление номера полиса обязательного медицинского страхования застрахованного лица, наименования МО, фамилии, имени, отчества пациента, места проживания, даты исследования; иные идентификаторы, посредством которых потенциально можно установить личность пациента; проводится автоматическая замена сведений о дате рождения на точный возраст (годы, месяцы) на момент исследования.

Важно подчеркнуть, что в информированном добровольном согласии пациента, которое оформляется для проведения исследования, должно быть и согласие на обработку персональных данных, включая обезличивание.

Деидентификация метаданных изображений проводится в стандарте DICOM (E Attribute Confidentiality Profiles) в соответствии с ГОСТ Р ИСО 17432–2009.

III этап. Разметка и верификация. Основные процедуры этапа:

1. Формирование задания на разметку, обеспечение наличия команды разметчиков, требуемых методических материалов и программного обеспечения.

2. Проведение разметки (аннотирования).

3. Внесение данных о верификации.

Подробно перечисленные процедуры были описаны выше.

В целях эффективной разметки наборов данных в лучевой диагностике необходимо программное обеспечение, позволяющее «насытить» исходные медицинские данные дополнительной полезной информацией, полученной в ходе работы медицинских специалистов. Такое программное обеспечение должно соответствовать определенным критериям. Далее представлен список необходимых и желательных **функциональных критериев для программного продукта**, используемого для создания размеченных наборов данных.

1. Критерии пользователей:

а. Разметка с использованием изображений нескольких серий, модальностей.

б. Возможность оставлять комментарии при разметке, которые направляются от пользователя к модератору разметки и наоборот.

в. Отображение изображений в трех ортогональных плоскостях с указанием локализации отображаемого слоя.

г. Изменение фона во время разметки по выбору пользователя (однородный черный/серый/белый).

д. Возможность указания на клинический случай, который сложен для интерпретации врачом.

е. Указание времени, которое тратится на среднюю разметку одного кейса пользователем.

ж. Возможность пропустить представленный кейс.

2. Критерии разработчиков методологии разметки:

а. Поддержка совместимости с разными типами входных данных и данных, содержащих результаты разметки.

б. Создание собственных дополнительных модулей.

в. Возможность подключения Clara Nvidia segmentation models (ML based)¹ или других решений для помощи в разметке.

г. Программирование автоматического выполнения отдельных шагов пользователя.

д. Модификация пользовательского интерфейса под текущую задачу.

3. Требования к ИТ-реализации:

а. Автосохранение в процессе работы — обязательно.

б. Возможность подключения к PACS — обязательно.

в. Не требователен к ресурсам компьютера (графический процессор) при визуализации (обязательное требование для desktop-продуктов).

г. Open-source — желательно.

д. Веб-интерфейс — обязательно.

IV этап. Документирование. Основные процедуры этапа:

1. Оформление сопроводительной документации (в том числе в соответствии с рекомендованной базовой структурой readme-файла).

¹ NVIDIA Clara Imaging | NVIDIA Developer [Electronic resource]. URL: <https://developer.nvidia.com/clara-medical-imaging> (accessed: 30.11.2020).

2. Внесение информации о наборе данных во внутренний учетный реестр организации, проводящей его формирование.

3. Государственная регистрация набора в качестве базы данных (результата интеллектуальной деятельности) в соответствии с действующим законодательством.

В процессе подготовки набора данных неизбежно упускаются те или иные критерии, всплывающие при непосредственной работе с набором конечными пользователями. Внесение нужных корректировок должно быть прозрачным для всех участников процесса и пользователей. Ведение версионирования набора данных позволяет отслеживать подобные изменения.

Нами предложен следующий оригинальный подход как вариация семантического версионирования:

1. Мажорная версия (Major) — увеличивается при изменении значимых параметров набора данных, связанных с клинической задачей, целью, принципами разметки и верификации данных.

2. Минорная версия (Minor) — увеличивается при замене, добавлении или удалении единиц данных в наборе без изменения других значимых его параметров; при этом алгоритмы обучения или валидации могут использовать новую минорную версию без изменения кода. При выпуске новой мажорной версии минорная версия устанавливается равной 0.

3. Патч-версия (Patch) — увеличивается при внесении корректировок в сопроводительную документацию, исправлении опечаток и иных ошибок в файлах разметки; при этом не меняется количество и качество единиц данных в наборе. При выпуске новой мажорной и/или минорной версии патч-версия устанавливается равной 0.

В целях удобства использования набора данных в корневую директорию помещается файл с названием README.md в формате Markdown и сгенерированный на его основе README. pdf в формате Adobe PDF. Единый подход к структуре README файла (рис. 11) позволяет организовать удобный поиск и фильтрацию по всем опубликованным наборам.

Название датасета

Логотип организации

Аннотация
(1-2 предложения)

Disclaimer

- название датасета
- условия лицензии (основные)
- особые условия, запрещающие использование

Обозначение лицензии

Разрыв страницы

Общая информация

- Название
- Внутренний код
- Классы разметки
- Ключевые слова
- Язык
- Финансирование
- Версия
- Постоянная ссылка
- Дата публикации

Аффилиация и авторы

- Авторы
- Аффилиация

Структура набора данных

- Схема структуры директорий и файлов
- Описание файлов в корневой директории
- Описание принципа наименования директорий
- Описание принципа наименования файлов

Обзор данных

Параметр	Значение

- Особенности подготовки исследований
- Принципы разметки данных
- Принципы верификации разметки данных

Правила использования и распространения

- Лицензия
- Копирайт
- Рекомендованная форма цитирования
- Правила распространения

Рисунок 11 – Рекомендуемая базовая структура README-файла

С точки зрения удобства ведения отчетности, практическую ценность имеет единый реестр подготовленных наборов данных. Минимальный набор рекомендуемых полей реестра:

1. *Порядковый номер* записи в реестре.
2. *Внутренний код*, уникальный для набора данных в текущем реестре и/или учреждении.
3. *Назначение* набора данных, область применения.
4. *Модальность/процедура* (характеристики исследований, пригодные для их поиска и выбора в МИС).
5. *Искомые признаки и/или целевая патология* (при возможности — с указанием кода Международной классификации болезней).
6. *Определение единицы данных*.

7. *Количество единиц данных* (при возможности — с указанием выходного объема данных в Мб, Гб или Тб).

8. *Классы разметки* с указанием количества записей в каждом классе.

V этап. Хранение и предоставление доступа. Основные процедуры этапа:

1. Передача готового к использованию набора данных в локальное или внешнее хранилище-репозиторий (публикация набора данных).

2. Организация доступа к набору данных авторизованных пользователей.

3. Применение набора данных.

Завершающим этапом процесса подготовки набора данных является его публикация. Этот этап включает в себя формирование структурированного набора файлов и директорий, помещенного в отдельное подразделение файлового хранилища, имеющего доступ к сети «Интернет». Возможность передачи опубликованных наборов третьим лицам, а также получать наборы, сформированные третьими лицами, появляется после подписания соглашения о передаче данных в двустороннем порядке. Соглашение регламентирует цели, задачи, порядок, сроки и условия передачи данных, а также ответственности сторон за нарушение условий соглашения.

Публикация набора данных представляет собой организацию доступа пользователей к его первой мажорной версии (1.0.0), обязательно снабженной сопроводительной документацией.

Хранение набора данных может быть организовано на локальном сервере или с использованием облачного хранения (ГОСТ Р ИСО/МЭК 17826–2015). В первом случае обеспечивается более высокий уровень доступности и безопасности, а во втором — становятся возможными совместное использование данных и более надежное резервное копирование. Доступ пользователей к набору данных проводится в соответствии с установленными стандартными процедурами (для разработчиков, исследователей, сотрудников, проводящих клинические испытания и т.д.). Вопросы публикации и использования наборов данных должны быть урегулированы с медицинской организацией, на базе которой формируется сам набор.

Организационно могут использоваться два репозитория:

1. Холодное итоговое хранилище. Используется для хранения готовых наборов данных, доступ к которым имеет ограниченный круг лиц по внутренней сети учреждения. Не требует большой скорости, поскольку нет больших и частых объемов работы с ним.

2. Горячее итоговое хранилище. Горячее хранилище выделено отдельно для опубликованных наборов данных, к которым необходимо обеспечить доступ для внешних пользователей согласно принятому регламенту. Скорость должна быть по возможности высокой и стабильной для пользователей.

Авторский пример хранения наборов данных представлен на рис. 12.

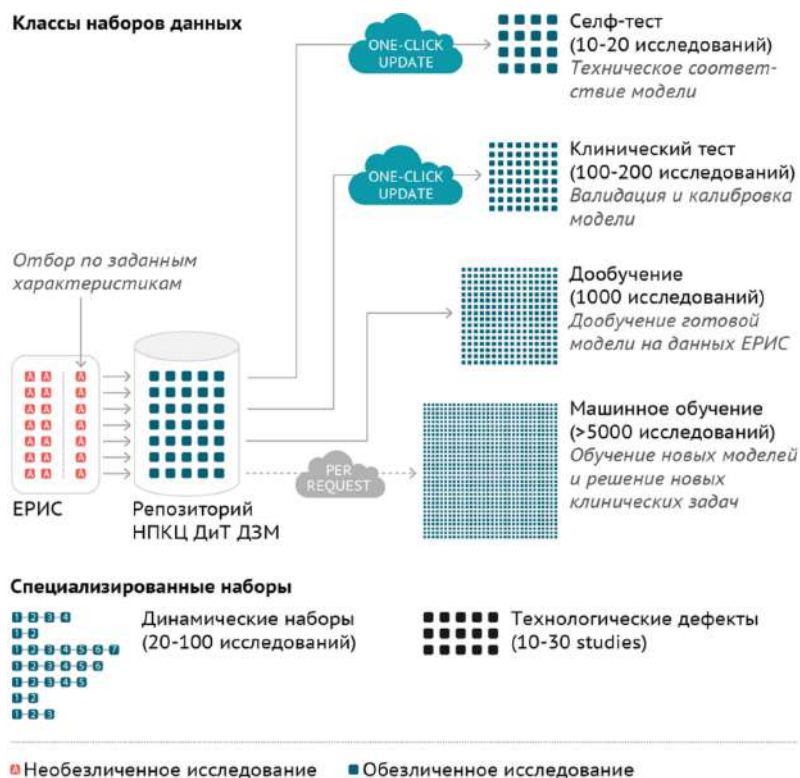


Рисунок 12 – Расположение наборов данных в ИТ-инфраструктуре (авторский пример)

К публикации допускаются наборы данных, соответствующие следующим условиям:

1. Набору данных присвоен внутренний код. Для составления внутреннего кода допустимо использование латинских букв, цифр и нижних подчеркиваний.

В качестве предложения сформулирован подход к названию набора данных (рис. 13), в соответствии с которым в наименование входят:

- классификация по диагностической ценности;
- общее количество исследований, вошедших в набор данных;
- область исследования;
- характеристика пациентов или изображения;
- клиническая задача;
- дополнительная информация о данных и/или разметке и/или верификации.

УНИФИЦИРОВАННЫЕ НАЗВАНИЯ НАБОРОВ ДАННЫХ



Примеры:

[1-A] 150 КОМПЬЮТЕРНЫХ ТОМОГРАММ ГОЛОВНОГО МОЗГА ПАЦИЕНТОВ, ПЕРЕНЕСШИХ ИНСУЛЬТ, ДЛЯ ОПРЕДЕЛЕНИЯ ЗОНЫ ПЕНУМБРЫ (С ВВЕДЕНИЕМ КОНТРАСТНОГО ВЕЩЕСТВА, С РАЗМЕТКОЙ ПО ТИПУ БИНАРНЫХ МАСОК)

[3-C] 500 ФЛЮОРОГРАММ ОРГАНОВ ГРУДНОЙ КЛЕТКИ С ПРИЗНАКАМИ И БЕЗ ПРИЗНАКОВ ТУБЕРКУЛЕЗА ДЛЯ ВЫЯВЛЕНИЯ СЛУЧАЕВ, ТРЕБУЮЩИХ ПОВЫШЕННОГО ВНИМАНИЯ

Рисунок 13 – Унифицированные названия наборов данных

2. Набор данных внесен в общий реестр наборов учреждения, которое осуществляет их разработку. Общий реестр представляет из себя таблицу, где ведется отчетность по готовности набора: ука-

зывается название и основная информация по нему (модальность, количество исследований, статус регистрации, анатомическая область, целевая задача, внутренний код, класс разметки, версия, путь к набору и ответственные лица).

3. Сохранены данные «цифровых двойников». «Цифровой двойник» представляет из себя цифровую копию информации об исследовании и пациенте, являющейся ключевой для подготовки наборов. Одни из наиболее значимых — теги с информацией по уникальным UID исследований и серий, производителей и модели оборудования, а также клиническая информация о пациенте (пол, возраст, история болезни или анамнез, если есть в доступе).

4. Набор сформирован в одной директории (по умолчанию — `datasets-collector/draft/NNN_INTERNALCODE/output/production-ready/`), где NNN — порядковый номер набора, INTERNALCODE — внутренний код.

5. В корневой директории набора содержится файл (ы):

- README_XX.md и README_XX. pdf (см. выше);
- LICENSE (если предусмотрено документацией README, в ином случае условие соблюдается автоматически).

6. Набор данных отвечает целям и задачам, указанным в документации README. Структура директорий, названия директорий и файлов соответствуют разделу Data Structure документации README.

7. Набор данных не содержит персональных данных субъектов медицинских данных (данные сохранены таким образом, чтобы исключить возможность явно или косвенно идентифицировать субъекта; произведена контрольная выборочная проверка данных).

Публикация готового набора данных может быть осуществлена в одном из трех режимов доступа:

1. **Закрытый.** Набор доступен ограниченному кругу лиц во внутренней сети (в пределах информационного контура).

2. **Ограниченный.** Набор доступен ограниченному кругу лиц, несущих ответственность за его дальнейшее распространение соглашения о передаче данных. Набор доступен для просмотра и загрузки из сети «Интернет» по инструкции и предоставляемым реквизитам пользователя. Набор данных должен иметь хотя бы одно из трех ограничений: срок действия, количество обращений или IP-адрес (а), с которого разрешен доступ.

3. Публичный. Набор данных доступен для просмотра и загрузки из сети «Интернет» по инструкции и реквизитам анонимными пользователями; информация о наборе распространяется через средства массовой информации, периодические издания, профильные веб-сайты и пр.

Каждое событие, связанное с обращением к набору данных, находящемуся в ограниченном или публичном режимах доступа, фиксируется в специальном журнале доступов со следующими параметрами:

- дата и время обращения;
- hash-маска пароля;
- IP адрес клиента;
- путь доступа;
- действие (чтение, загрузка).

Готовые наборы данных могут применяться для обучения, дообучения, внутреннего тестирования, калибровки, функционального тестирования, технологического мониторинга, самотестирования, технических, клинических испытаний и т. д.

§5. Практическая реализация подготовки наборов данных

Разработка и апробация методологии велись нашим коллективом с 2017 года. Для помощи многочисленным разработчикам в 2018 году в свободном доступе впервые в Российской Федерации нами был размещен ограниченный деперсонализированный набор размеченных компьютерных томограмм грудной клетки «Датасет Радиологии Москвы CT LungCa-500». Этот набор был скачан несколько десятков раз и использован для самотестирования и обучения несколькими группами разработчиков и компаниями.

Созданная нами методология подготовки эталонных наборов данных успешно апробирована в ходе выполнения научно-исследовательских работ. Был подготовлен ряд наборов данных, отвечающий обозначенным выше требованиям.

В качестве примера перечислим некоторые из них (таблица 8).

Таблица 8 – Пример наборов данных, подготовленных по авторской методологии

Название	Предназначение	Основные характеристики
Набор смоделированных диагностических изображений для технического тестирования разработок на основе компьютерного зрения/ «искусственно-го интеллекта»	Предназначен для тестирования технической готовности систем автоматического анализа диагностических изображений компьютерной томографии, маммографии и рентгенографии. Функциональные возможности базы данных обеспечивают возможность проверки готовности систем автоматического анализа к работе в Единой радиологической информационной системе г. Москвы (ЕРИС). База данных представляет собой структурированный набор деперсонализированных результатов диагностических исследований грудной клетки и молочных желез	Объем базы данных: 2,2 Гб. База данных состоит из 16 анонимизированных диагностических исследований в формате DICOM и включает в себя следующие модальности/направления: компьютерная томография/рак легких, маммография/рак молочной железы, рентгенография грудной клетки/туберкулез и рентгенография грудной клетки/патологии легких (не туберкулез). Исследования идентифицируются по трем признакам, содержащимся в DICOM-файлах: Accession Number, Study ID и Patient ID
MosMedData: Результаты маммографических исследований для калибровки сервисов на основе искусственного интеллекта	Оценка точности систем автоматического анализа диагностических изображений, полученных в результате проведения маммографии	Объем базы данных: 19 Гб. База данных состоит из результатов 100 маммографических (ММГ) анонимизированных диагностических исследований молочных желез в формате DICOM, а также из сопроводительной таблицы с разметкой в виде отнесения исследований к группе, содержащей признаки, коррелирующие с наличием рака молочной железы (код C50 по Международной классификации болезней 10-го пересмотра), либо к группе, в которой упомянутые признаки отсутствуют

Продолжение таблицы 8

Название	Предназначение	Основные характеристики
MosMedData: Результаты КТ-исследований легких для калибровки сервисов, работающих на основе искусственного интеллекта	Оценка точности систем автоматического анализа диагностических изображений, полученных в результате проведения компьютерной томографии	Объем базы данных: 22 Гб. База данных состоит из результатов 100 компьютерно-томографических (КТ) анонимизированных диагностических исследований органов грудной полости в формате DICOM, а также из сопроводительной таблицы с разметкой в виде отнесения исследований к группе, содержащей признаки, коррелирующие с наличием рака легкого (код C34 по Международной классификации болезней 10-го пересмотра), либо к группе, в которой упомянутые признаки отсутствуют
MosMedData: результаты исследований компьютерной томографии органов грудной клетки с признаками COVID-19	Предназначен для разработки, дообучения и тестирования программных продуктов (использующих в том числе методы компьютерного зрения), выявляющих признаки, характерные для коронавирусной инфекции (COVID-19)	Объем базы данных: 11 Гб. База данных представляет собой структурированный набор деперсонализированных результатов КТ-исследований органов грудной клетки с рентгенологическими признаками вирусной пневмонии (COVID-19), а также без признаков (норма). Структурирование по степени тяжести поражения легочной ткани проведено по классификации «КТ0-4»*. Для некоторых исследований представлена разметка областей интереса (зон уплотнений по типу «матового стекла» и консолидации) в виде бинарной пиксельной маски
* Морозов С. П., Гомблевский В. А., Чернина В. Ю. [и др.]. Прогнозирование летальных исходов при COVID-19 по данным компьютерной томографии органов грудной клетки // Туберкулез и болезни легких. 2020. Т. 98, №6. С. 7–14. URL: https://doi.org/10.21292/2075-1230-2020-98-6-7-14 .		

Дополнительно расскажем о наборе данных «MosMedData: результаты исследований компьютерной томографии органов грудной клетки с признаками COVID-19». Исследования для его формирования были собраны в отделениях лучевой диагностики лечебных учреждений города Москвы в период с 01.03.2020 по 25.04.2020. Уникальность набора данных заключается в представлении широкой распространенности находок в легких у пациентов с подозрением на заболевание новой коронавирусной инфекцией, выявляемых в первые месяцы пандемии, по данным КТ грудной клетки, выполненных в медицинских организациях, оказывающих первичную медико-санитарную помощь взрослому населению. На момент публикации набора в открытом доступе в литературе не обнаружилось более крупного общедоступного набора данных КТ органов грудной клетки с подозрением на поражение легких в результате инфицирования SARS CoV2.0.

Благодаря расположению в открытом доступе набор данных был использован независимыми разработчиками алгоритмов искусственного интеллекта. В частности, набор применен для:

1. Независимого тестирования точности алгоритмов двумя коллективами разработчиков технологий искусственного интеллекта из Китая — под руководством Pengyi Zhang (Пекинский технологический институт) и Cheng Jin (Пекинский национальный исследовательский центр информационных наук и технологий)¹; из США — под руководством Li Sun (Питтсбургский университет)²; из Германии — под руководством Keno K. Bressen (Университетская клиника Шарите, Берлин)³.

2. Разработки собственных алгоритмов ИИ командой из Южной

¹ Jin C., Chen W., Cao Y. et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis // Nat Commun. 2020 Vol. 11, №1. P. 5088. DOI:10.1038/s41467-020-18685-1. Zhang P., Zhong Y., Deng Y. et al. CoSinGAN: Learning COVID-19 Infection Segmentation from a Single Radiological Image // Diagnostics (Basel). 2020. Vol. 10, №11. P. 901. DOI:10.3390/diagnostics10110901.

² Sun L., Yu K., Batmanghelich K. Context Matters: Graph-based Self-supervised Representation Learning for Medical Images. 2020. arXiv:2012.06457v1.

Кореи под руководством Myeongkyun Kang (Институт науки и технологий Кенгбук)⁴ и доктором Sachin Sharma (Институт передовых исследований, Гандинагар)⁵.

3. Иных научных исследований в лучевой диагностике, не связанных с искусственным интеллектом⁶.

Таким образом, разработанная методология подготовки эталонных наборов данных была не только проверена в рамках собственной практической деятельности, но и широко валидирована на международном уровне.

³ Bressemer K.K., Niehues S.M. et al. 3D U-Net for segmentation of COVID-19 associated pulmonary infiltrates using transfer learning: State-of-the-art results on affordable hardware. 2021. arXiv:2101.09976v1.

⁴ Kang M., Hong K.S., Chikontwe P. et al. Quantitative Assessment of Chest CT Patterns in COVID-19 and Bacterial Pneumonia Patients: a Deep Learning Perspective // J Korean Med Sci. 2021. Vol. 36, №5. P. e46. Published 2021 Feb 1. DOI:10.3346/jkms.2021.36.e46.

⁵ Sharma S. Drawing insights from COVID-19-infected patients using CT scan images and machine learning techniques: a study on 200 patients // Environ Sci Pollut Res Int. 2020. Vol. 27, №29. P. 37155–37163. DOI:10.1007/s11356-020-10133-3.

⁶ Sushentsev N., Bura V., Kotnik M. et al. A head-to-head comparison of the intra- and interobserver agreement of COVID-RADS and CO-RADS grading systems in a population with high estimated prevalence of COVID-19 // BJR Open. 2020. Vol. 2, №1. P. 20200053. Published 2020 Dec 11. DOI:10.1259/bjro.20200053.

Глава 5. МЕТОДОЛОГИЯ КЛИНИЧЕСКИХ ИСПЫТАНИЙ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

§1. Разработка методологии

Формирование методологии клинических испытаний (оценки) технологий искусственного интеллекта включало в себя несколько этапов (рис. 14).



Рисунок 14 – Схема процесса разработки и апробации методологии клинических испытаний (оценки) технологий ИИ в лучевой диагностике

Первый этап: параллельно велись аналитические исследования и практический анализ. Обобщался мировой и национальный

опыт, изучалась нормативно-правовая база. В формате стандартного диагностического исследования были оценены 5 алгоритмов для автоматизированного анализа цифровых рентгенограмм, компьютерных томограмм грудной клетки и головы, а также для анализа текстовых документов. Второй этап: полученные знания, данные и опыт были систематизированы и обобщены в виде проекта методологии клинической оценки. Проект опубликован в виде препринта на русском и английском языках. Третий этап: препринт был доступен для широкого публичного обсуждения (в том числе в депозитарии ArXiv.org), в ходе которого был получен ряд отзывов, в том числе положительная рекомендация от Европейского общества информатики и медицинской визуализации (European Society of Medical Imaging Informatics). Параллельно методология была применена для анализа 14 новых алгоритмов ИИ (помимо вышеуказанных модальностей, исследованы средства автоматизированного анализа маммограмм, магнитно-резонансных томограмм). Четвертый этап: результаты открытого рецензирования и апробации систематизированы, в методологию внесены дополнения и уточнения, сформирован и опубликован финальный документ¹.

Далее мы подробно приводим разработанную методологию.

§2. Общие положения

Искусственный интеллект в лучевой диагностике представляет мультидисциплинарное (медицина, биология, математика и компьютерные науки) прикладное научное направление по постановке

¹ Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика) // сост. С. П. Морозов, А. В. Владзимирский, В. Г. Кляшторный [и др.] / Сер. «Лучшие практики лучевой и инструментальной диагностики». М., 2019. Вып. 57. 51 с. (Clinical acceptance of software based on artificial intelligence technologies (radiology) / S.P. Morozov, A.V. Vladzymyrskyy, V.G. Klyashtornyy [et al.] // Ser. «Best practices in medical imaging». М., 2019. Issue 57. 45 p.).

и решению задач моделирования отдельных компонентов аналитических процессов, являющихся частью интеллектуальной деятельности врача-рентгенолога, радиолога или специалиста ультразвуковой диагностики (таблица 9).

Таблица 9 – Типовые формы применения искусственного интеллекта в лучевой диагностике

Основные задачи (вопросы)	Действие ИИ	Типовая медицинская задача
Локализация объекта	Обнаружение (детекция)	Приоритизация исследований, поддержка принятия решений, контроль качества
1. Объект присутствует или нет. 2. Определение типа объекта	Классификация	Скрининг, массовые профилактические осмотры. Определение характера патологии
Определение размеров, характеристика объекта	Сегментация, определение радиомических особенностей	Определение анатомических границ. Медицинская морфометрия (мониторинг течения патологического процесса, автоматическое формирование описания диагностического изображения)
Распознавание речи и текста человека	Обработка естественного языка	Протоколирование, контроль качества
Оценка изменений объекта во времени	Обнаружение (детекция), классификация, сегментация	Оценка динамики патологических изменений

Любой алгоритм искусственного интеллекта, успешно прошедший клинические испытания, функционирует исключительно под контролем медицинских работников в качестве системы поддержки принятия врачебных решений. Генерируемые алгоритмом результаты могут использоваться или не использоваться медицинскими работниками. Применение искусственного интеллекта направлено на повышение эффективности системы здравоохранения и производительности труда персонала, снижение рисков и ошибок, стандартизацию представления результатов диагностических исследований.

В соответствии со статьей 38 Федерального закона от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан

в Российской Федерации» медицинское изделие — это «любые инструменты, аппараты, приборы, оборудование, материалы и прочие изделия, применяемые в медицинских целях отдельно или в сочетании между собой, а также вместе с другими принадлежностями, необходимыми для применения указанных изделий по назначению, включая специальное программное обеспечение, и предназначенные производителем для профилактики, диагностики, лечения и медицинской реабилитации заболеваний, мониторинга состояния организма человека, проведения медицинских исследований, восстановления, замещения, изменения анатомической структуры или физиологических функций организма, предотвращения или прерывания беременности, функциональное назначение которых не реализуется путем фармакологического, иммунологического, генетического или метаболитического воздействия на организм человека. Медицинские изделия могут признаваться взаимозаменяемыми, если они сравнимы по функциональному назначению, качественным и техническим характеристикам и способны заменить друг друга»¹.

Технологии искусственного интеллекта реализуются в виде программного обеспечения, используемого самостоятельно или в сочетании / в составе с иными устройствами. В любом случае для применения в сфере здравоохранения они должны получить статус медицинского изделия в соответствии с действующим законодательством.

В Российской Федерации программное обеспечение с применением технологий искусственного интеллекта относится к классу 3 — программное обеспечение с высокой степенью риска². Правила и требования к технической и эксплуатационной документации соответствующего программного обеспечения определяются действующим законодательством³.

Государственная регистрация медицинских изделий проводится на основании результатов технических испытаний, токсикологиче-

¹ Федеральный закон от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации». Ст. 38.

² Приказ Министерства здравоохранения Российской Федерации от 07.07.2020 №686н «О внесении изменений в приложения №1 и №2 к приказу Министерства здравоохранения Российской Федерации от 6 июня 2012 г. №4н „Об утверждении номенклатурной классификации медицинских изделий“».

ских исследований, клинических испытаний, представляющих собой формы оценки соответствия медицинских изделий с учетом классификации в зависимости от потенциального риска их применения, и экспертизы качества, эффективности и безопасности медицинских изделий с учетом классификации в зависимости от потенциального риска их применения, а также испытаний в целях утверждения типа средств измерений⁴.

Таким образом, на этапе разработки и подготовки к регистрации в качестве медицинского изделия программное обеспечение на основе технологий искусственного интеллекта проходит технические и клинические испытания. В процессе клинических испытаний должна быть установлена диагностическая эффективность, подтверждены безопасность использования и соответствие реальных характеристик ПО тем, которые были заявлены изготовителем.

Следовательно, применяемые в лучевой диагностике алгоритмы и программное обеспечение на основе искусственного интеллекта относятся к медицинским изделиям и должны проходить клинические испытания в установленном порядке. При этом программное решение может быть интегрировано в информационные системы в сфере здравоохранения, медицинские (радиологические) информационные системы⁵; либо быть составной частью иного медицинского изделия, например, диагностического устройства.

Методология клинических испытаний, в том числе в сокращенном варианте, может быть использована в рамках пилотных проек-

³ Приказ Министерства здравоохранения Российской Федерации от 20.11.2020 №1236н «О внесении изменений в требования к содержанию технической и эксплуатационной документации производителя (изготовителя) медицинского изделия, утвержденные приказом Министерства здравоохранения Российской Федерации от 19 января 2017 г. №11н».

⁴ Приказ Министерства здравоохранения Российской Федерации от 09.01.2014 №2н «Об утверждении Порядка проведения оценки соответствия медицинских изделий в форме технических испытаний, токсикологических исследований, клинических испытаний в целях государственной регистрации медицинских изделий».

⁵ Приказ Министерства здравоохранения Российской Федерации от 09.06.2020 №560н «Об утверждении Правил проведения рентгенологических исследований».

тов, научно-исследовательских, опытно-конструкторских работ и клинических апробаций (в соответствии со статьей 36.1 Федерального закона от 21.11.2011 №323-ФЗ¹), целью которых является развитие технологий искусственного интеллекта на принципах доказательной медицины.

В таких случаях следует использовать термин **«предварительные клинико-технические испытания (ПКТИ)»**.

Предварительные клинико-технические испытания — это разработанное и запланированное исследование, предпринятое для предварительной научно-практической оценки безопасности и эффективности программного обеспечения на основе искусственного интеллекта. ПКТИ проводятся по упрощенной и адаптированной методологии клинических испытаний с целью получения дополнительных знаний о возможностях и ограничениях ПО на основе ИИ. ПКТИ не являются заменой или аналогом клинических испытаний. ПКТИ могут включать этапы аналитической валидации и клинической апробации либо состоять только из аналитической валидации (см. далее). По результатам ПКТИ оформляется отчет (таблица 10).

ПКТИ применимы в контексте научно-исследовательских, опытно-конструкторских работ, а также в качестве инструмента оценки медико-технической готовности ПО на основе ИИ для углубленного изучения в реальных клинических условиях.

¹ Федеральный закон от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации». Ст. 36.1.

Таблица 10 – Рекомендованная форма отчета о предварительных клинико-технических испытаниях (ПКТИ) программного обеспечения на основе ИИ для лучевой диагностики

№	Пункт	Пояснение
1	Учреждение	Полное название учреждения, на базе которого проведены ПКТИ
2	Контактная информация	Контактная информация учреждения (адрес, телефон, сайт, электронная почта)
3	Даты проведения ПКТИ	Диапазон дат проведения ПКТИ
4	Резюме	Структурированное изложение дизайна исследования, материалов, методов, результатов и выводов
5	Цель, задачи и конечные точки ПКТИ	Цель, задачи и конечные точки ПКТИ
6	Референс-тест (эталонный набор данных)	Детальная характеристика использованных для ПКТИ эталонных наборов подготовленных данных
6.1	Вид данных	Вид(ы) данных (медицинская документация, результаты исследований и проч.), модальности для исследований, прочие характеристики
6.2	Количество включенных клинических случаев	Количество включенных клинических случаев (пациентов, результатов исследований и проч.)
6.3	Характеристика популяции	Характеристика популяции (расовые, гендерно-демографические, иные характеристики)
6.4	Характеристики набора данных и разметки	Где и когда сформирован эталонный набор данных с указанием ключевых характеристик (государственная регистрация базы данных, сведения о деперсонализации, о наличии информированного добровольного согласия пациентов, критерии включения/исключения, источники клинических случаев). Способ подготовки (разметки) набора данных
6.5	Характеристика патологии в наборе данных	Целевая патология и диагностические группы с распределением по ним. Способ верификации и наличие соответствующей информации в наборе данных
6.6	Способ формирования набора данных	Сформирован последовательно, случайно или другими способами. Обоснование размера выборки
6.7	Источники данных	Количество и локализация медицинских учреждений, послуживших источниками клинических случаев, включенных в ПКТИ

Продолжение таблицы 10

№	Пункт	Пояснение
6.8	Отметка о независимости ПКТИ	Набор данных, использованный для ПКТИ, не должен (как полностью, так и частично) использоваться для обучения или калибровки индекса-теста
7	Индекс-тест (ПО на основе ТИИ)	Информация об использовании, инсталляции, организации доступа и проч. индекс-теста (исследуемого ПО на основе ТИИ)
8	Процесс ПКТИ	Краткое общее описание процесса исследования; возможно использование схем, диаграммы CONSORT
9	Таблица результатов	Комбинированная таблица результатов индекс-теста и референс-теста
10	Порог активации	Точка отсечения («cut-off») для референс-теста и для индекс-теста, включая разделение исходов на «ожидаемые» и «непредвиденные»
11	Показатели диагностической точности	Метрики диагностической точности с 95% доверительным интервалом (чувствительность, специфичность, общая точность, площадь под характеристической кривой и т.д.)
12	Ограничения	Любые ограничения ПКТИ, включая источники систематических ошибок, статистических неточностей и обобщений. Любые значимые различия в методиках проведения индекс-теста и референс-теста
13	Выводы	Краткое обобщение результатов
14	Источники финансирования ПКТИ	Указание источника финансирования работ, связанных с проведением ПКТИ
15	Прочее	Произвольная дополнительная информация
16	Список исполнителей	Список сотрудников учреждения, проводивших ПКТИ (ФИО, должность, ученое звание/степень)
17	Дата подписания отчета	-
18	Подпись ответственного исполнителя	Личная подпись ответственного исполнителя ПКТИ, ФИО
19	Подпись руководителя учреждения	Личная подпись руководителя учреждения, ФИО
20	Печать учреждения	-

§3. Клинические испытания

Цель клинических испытаний программного обеспечения на основе технологий искусственного интеллекта: подтверждение эффективности, безопасности использования, а также соответствия его характеристик использованию по назначению, указанному изготовителем.

В соответствии с п.8 статьи 38 Федерального закона от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации»¹, в целях государственной регистрации медицинских изделий в порядке, установленном Росздравнадзором, проводится оценка соответствия в форме технических испытаний, клинических испытаний и экспертизы качества, эффективности и безопасности медицинских изделий.

Согласно п. 4 Правил государственной регистрации медицинских изделий², **технические испытания** проводятся с целью определения соответствия характеристик медицинского изделия требованиям нормативной документации, технической и эксплуатационной документации производителя и принятия последующего решения о возможности проведения клинических испытаний. Производитель сам определяет, каким нормативным документам соответствует его изделие.

Клинические испытания представляют собой разработанное и запланированное систематическое исследование, предпринятое для оценки безопасности и эффективности медицинского изделия. Провести клинические испытания можно только при условии успешного проведения технических испытаний.

Критерии допуска к клиническим испытаниям (предварительным клинико-техническим испытаниям, предварительной опытной эксплуатации) программного обеспечения на основе технологий искусственного интеллекта:

1. Цели:

¹ Федеральный закон от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации». Ст. 38.

² Постановление Правительства Российской Федерации от 27.12.2012 №1416 «Об утверждении Правил государственной регистрации медицинских изделий».

- ПО обеспечивает предварительный автоматический анализ диагностических изображений для повышения качества и скорости подготовки описания врачом;

- ПО обеспечивает приоритизацию исследований в рабочем списке (worklist) в соответствии с автоматически выявленной патологией;

- ПО обеспечивает предварительный сравнительный анализ исследований одного пациента в формате DICOM, выполненных в разное время (оценка динамики);

- ПО обеспечивает поддержку принятия врачебных решений;

- ПО обеспечивает автоматическую подготовку проекта описания исследования по результатам работы алгоритма ИИ.

2. Сертификация:

- пройдены технические испытания медицинского изделия в аккредитованной лаборатории;

- наличие у ПО на основе технологий искусственного интеллекта сертификации FDA и/или CE (класс II); наличие у юридического лица-разработчика фактического внедрения ПО в медицинские организации: не менее 2-х независимых учреждений; более 6 месяцев эксплуатации; не менее 1000 успешно выполненных анализов (подтвержденных врачом-диагностом) для каждой задачи (если ПО решает несколько задач);

- наличие научных статей (оригинальных исследований) в рецензируемых журналах, индексируемых базами данных Scopus и/или Web of Science и входящих в первый-второй квартиль согласно платформе International Scientific Journal & Country Ranking;

- должны быть доказаны диагностическая точность – $AUC > 0,81$ (классическая ROC-кривая) и повышение эффективности производственных процессов (сравнение работы специалистов с и без данного алгоритма ИИ, в том числе хронометраж).

3. Безопасность:

- соблюдение требований законодательства Российской Федерации в сфере персональных данных, информационной безопасности, охраны здоровья;

- наличие или готовность разместить на территории Российской Федерации серверные мощности, необходимые для работы ПО.

4. Доказательность:

- после окончания разработки оценка точности алгоритмов проводилась на независимых данных¹;

- диагностическая точность протестирована на популяции, характеристики которой аналогичны населению территории предполагаемого использования ИИ;

- ежегодное обновление информации о диагностической точности.

5. Стандартизация:

- автоматизированный анализ диагностических изображений в стандарте DICOM;

- поддержка стандарта HealthLevel7 (HL7) / FHIR (в частности, должен обеспечиваться обмен сообщениями о завершении автоматического анализа изображения, обнаружении и классификации патологии);

- при формировании проектов описаний используются рекомендованные классификации (RADS, MAGNIMS и т.д.).

6. Интеграция:

- наличие или готовность разработать средства для «бесшовной» интеграции с информационными системами в сфере здравоохранения данного субъекта РФ, медицинскими информационными системами;

- наличие средств для интеграции с PACS и РИС (DICOM «query and retrieve»);

- возможность «бесшовной» интеграции с PACS/РИС, обеспечиваемой только программными средствами.

7. Функциональность:

- возможность потоковой обработки с последующей отправкой в PACS серий, дополненных результатами анализа ИИ;

- возможность совмещения серий нативных изображений и серий, дополненных результатами анализа ИИ;

- выявление находок (нозологий, патологических признаков, отклонений от нормы) для данной модальности или рабочего процесса в соответствии с техническим заданием;

- классификация (определение типа, разновидности) находки, в том числе в соответствии с Международной классификацией болезней (версия 10) если предусмотрено техническим заданием;

¹ Набор медицинских данных для тестирования отличался от набора, использованного для обучения, калибровки и валидации алгоритма (то есть тестирование осуществлено на данных, с которыми ранее алгоритм не работал).

- предоставление информации о вероятности наличия целевой патологии;

- автоматическая подготовка и выгрузка в PACS проекта описаний находок (с соблюдением структуры шаблона описаний, предустановленной в PACS);

- результат автоматического анализа отображается стандартными средствами PACS, в том числе графическими возможностями DICOM;

- наличие шаблонов протоколов описаний с их автоматической генерацией и отправкой в РИС/МИС посредством сообщений HL7/FHIR;

- наличие встроенного инструмента оценки точности;

- длительность обработки одного исследования не превышает установленного временного периода (формируется индивидуально, исходя из клинического сценария, характеристик инфраструктуры и проч.; пример – 60 секунд, не учитывая время на пересылку данных; при оценке динамики (сравнении исследований) анализ может занимать более 60 секунд, но не более 60 секунд на одно исследование);

- функция контроля, обеспечивающая старт автоматического анализа только по факту получения из РИС сообщения HL7 «status complete by radiographer»;

- сообщения ПО в стандарте HL7 о завершении автоматического анализа и обнаружении патологии используются для приоритизации исследований в рабочем списке (worklist);

- возможность автоматизированного поиска в базе PACS/РИС аналогичных исследований для сравнения;

- для каждого проанализированного исследования ПО формирует декларацию, содержащую название ПО (алгоритма), список выявленных находок (нозологий, патологических признаков, отклонений от нормы) с указанием чувствительности и специфичности для каждого типа находок; декларация формируется в DICOM SR, Overlay и т. д. и передается в PACS как отдельная серия;

- результатом работы ПО является сформированная серия изображений в формате DICOM, которая:

- а) имеет количество срезов, аналогичное исходной серии для возможности синхронного просмотра врачом;

- б) обладает информацией на каждом срезе о названии ПО,

версии, диагностической точности с датой ее проверки и точным временем завершено анализа конкретного исследования;

с) допускается наличие дополнительных серий с информацией о результатах анализа (сводные таблицы анализа выявленных находок в динамике и/или отдельные изображения находок).

8. Контракт:

- наличие у юридического лица-разработчика системы менеджмента качества;

- наличие у юридического лица-разработчика политики подготовки и контроля версий;

- регулярные обновления системы, включая обновление информации о диагностической точности;

- обновления ПО включены в стоимость;

- все медицинские данные, сопутствующие материалы и результаты работы ПО являются собственностью заказчика.

Учет и оценку критериев проводят с использованием анкеты (таблица 11).

Весомость и значимость каждого критерия может определяться индивидуально для каждой ситуации. Критерии вносятся в техническое задание, описывающее требования к ПО на основе технологий искусственного интеллекта. Также в техническом задании обязательно должны быть указаны *цели и клинические сценарии применения* искусственного интеллекта по отдельным модальностям, видам исследований, производственным процессам.

Процесс клинических испытаний организуют в соответствии с действующей нормативно-правовой базой и рекомендованной методикой экспертизы качества, эффективности и безопасности медицинских изделий в части программного обеспечения. В ходе испытаний ПО на основе технологий искусственного интеллекта должна быть дана *«оценка клинической связи (научной обоснованности) степени адекватности подбора данных и системы их анализа (концепция, измерения, заключение) целевому назначению»* [2].

Клиническая оценка ПО, разработанного на основе искусственного интеллекта и предназначенного для применения в лучевой диагностике, осуществляется в 2 этапа (рис. 15).

1. Аналитическая валидация.
2. Клиническая апробация.
2. Клиническая апробация.

Таблица 11 – Анкета для допуска программного обеспечения на основе технологий искусственного интеллекта к предварительной опытной эксплуатации (предварительным клинико-техническим испытаниям)

Раздел	Метрики	Ответ (1 – да/ 0 – нет)	Комментарий
1. Цели	1.1. ПО обеспечивает предварительный автоматический анализ медицинских изображений в формате DICOM для повышения качества и скорости подготовки описания врачом. 1.2. ПО обеспечивает приоритизацию исследований в рабочем списке (worklist) в соответствии с автоматически выявленной патологией. 1.3. ПО обеспечивает автоматическую подготовку проекта описания исследования по результатам работы алгоритма ИИ. 1.4. ПО обеспечивает предварительный сравнительный анализ исследований одного пациента в формате DICOM, выполненных в разное время (оценка динамики)		
2. Сертификация	2.1. Наличие у ПО на основе технологий искусственного интеллекта сертификации FDA и/или CE (класс II). В случае несоответствия требованиям пункта 2.1 остальные показатели / документы должны отвечать требованиям пунктов 2.2 и 2.3. 2.2. Наличие у юридического лица-разработчика фактического внедрения ПО в медицинские организации: – не менее 2-х независимых учреждений; – более 6 месяцев эксплуатации; – не менее 1000 успешно выполненных анализов (подтвержденных врачом-диагностом) для каждой задачи (если ПО решает несколько задач).		

Продолжение таблицы 11

Раздел	Метрики	Ответ (1 – да/ 0 – нет)	Комментарий
	2.3. Наличие научных статей (оригинальных исследований) в рецензируемых журналах, индексируемых базами данных Scopus и/или Web of Science и входящих в первый-второй квартиль согласно платформе International Scientific Journal & Country Ranking; должны быть доказаны диагностическая точность – $AUC \geq 0,81$ (классическая ROC-кривая) и повышение эффективности производственных процессов (сравнение работы специалистов с и без данного алгоритма ИИ, в том числе хронометраж)		
3. Доказательность	3.1. После окончания разработки оценка точности алгоритмов проводилась на независимых данных (набор биомедицинских данных для тестирования отличался от набора, использованного для обучения, калибровки и валидации системы; то есть тестирование осуществлено на данных, с которыми ранее алгоритм не сталкивался). 3.2. Диагностическая точность протестирована на данных лиц, относящихся к популяции, характеристики которой аналогичны населению территории предполагаемого использования ИИ. 3.3. Ежегодное обновление информации о диагностической точности		
4. Функциональность	4.1. Наличие встроенного инструмента оценки точности. 4.2. Длительность обработки одного исследования не превышает установленного временного периода (формируется индивидуально, исходя из клинического сценария, характеристик инфраструктуры и проч.; пример – 60 секунд, не учитывая время		

Продолжение таблицы 11

Раздел	Метрики	Ответ (1 – да/ 0 – нет)	Комментарий
	на пересылку данных; при оценке динамики (сравнении исследований) анализ может занимать более 60 секунд, но не более 60 секунд на одно исследование). 4.3. Результатом работы ПО является сформированная серия изображений в формате DICOM, которая: – имеет количество срезов, аналогичное исходной серии для возможности синхронного просмотра врачом; – обладает информацией на каждом срезе о названии ПО, версии, диагностической точности с датой ее проверки и точным временем завершения анализа конкретного исследования; – допускается наличие дополнительных серий с информацией о результатах анализа (сводные таблицы анализа выявленных находок в динамике и/или отдельные изображения находок)		
5. Контракт	5.1. Наличие у юридического лица-разработчика системы менеджмента качества. 5.2. Регулярные обновления системы, включая обновление информации о диагностической точности. 5.3. Обновления ПО включены в стоимость. 5.4. Все медицинские данные, сопутствующие материалы и результаты работы ПО являются собственностью заказчика		



* С учетом действующих НПА

Рисунок 15 – Схема проведения клинической оценки программного обеспечения на основе технологий искусственного интеллекта (для сферы лучевой диагностики)

Таблица 12 – Этапы аналитической валидации ПО на основе технологий искусственного интеллекта для лучевой диагностики

Этап	Цель	Краткое описание	Результат этапа
1	2	3	4
I. Анкетирование	(1) сбор информации для рейтинга сервиса (цели, сертификация, доказательность, функциональность, контракт); (2) оценка клинической задачи сервиса и практической применимости в рабочем процессе отделений лучевой диагностики	Юридическому лицу- разработчику ИИ-сервиса предлагается заполнить анкету, затем ответы в анкете анализируются комиссией	Заполненная анкета
II. Самостоятельный тест	Проверка технической готовности сервиса (бесшовная интеграция в информационную систему в сфере здравоохранения, обработка данных со всех типов диагностических аппаратов, качественное предоставление результатов медицинским работникам)	Разработчику предоставляется типовый эталонный набор данных (обычно от 5 до 10 исследований) для самооценки технической готовности ПО	Файлы с результатами обработки набора данных ПО
III. Интервью	(1) детализация клинической задачи и сценария применения; (2) оценка юридического лица на соответствие корпоративной культуре превосходного качества и организации (КПКО)	С представителями компании проводится интервью согласно протоколу (таблицы 13, 14)	Ответы на вопросы в протоколе

Продолжение таблицы 12

Этап	Цель	Краткое описание	Результат этапа
IV. Онлайн-тест	Оценка работы сервиса в режиме реального времени на ограниченном наборе данных (включает выборку показательных случаев)	Анализ данных, интерпретация результатов проводится при непосредственном участии комиссии	Отчет о работе ИИ-алгоритма/сервиса по каждому исследованию
V. Доказательный тест	Ретроспективная проверка заявленной точности и остальных параметров работы алгоритмов на эталонном наборе данных	Происходит тестирование ПО на эталонном наборе данных, подготовленном в соответствии с клинической задачей, решаемой ПО	Оценка диагностической точности (аналитическая валидация – см. далее)
VI. Итоговая оценка	(1) Интегральная оценка работы ПО и готовности к проверке в реальных условиях ОЛД; (2) Интегральная оценка юридического лица-разработчика	Комиссией анализируются все прошедшие этапы тестирования и выставляются промежуточные оценки ПО и юридическому лицу	Решение о допуске ПО сервиса к интеграции с действующей информационной системой в сфере здравоохранения для оценки работы ПО в условиях рутинных производственных процессов отделений лучевой диагностики

Аналитическая валидация – оценка корректности обработки входных данных программным обеспечением для создания надежных выходных данных; оценивается с применением эталонных наборов размеченных данных. Аналитическая валидация (в рамках клинических испытаний) ПО на основе технологий ИИ включает 6 этапов (таблица 12).

Таблица 13 – Протокол интервью для оценки надежности и качества юридического лица-разработчика ПО на основе ИИ для лучевой диагностики

Общие требования	Наличие следующих составляющих: - рабочая станция врача-рентгенолога, поддерживающая визуализацию результатов работы алгоритмов ИИ-сервиса; - возможность во время очной встречи загрузить и обработать ИИ-сервисом исследования; - возможность хронометража обработки ИИ-сервисом от загрузки до предоставления результата; - информация об архитектуре используемых моделей; - доступ к коду; - присутствие разработчика, специалиста по анализу больших массивов данных, которые могут проконсультировать по коду и структуре алгоритмов, процессу обучения и валидации алгоритма(-ов); - присутствие врача, использующего ИИ-сервис в своей практике, либо документы, подтверждающие факт успешной клинической апробации с контактной информацией (опционально)
Оценка клинической значимости	1. Какую клиническую задачу решает ИИ-сервис? 2. Является ли применение ИИ-сервиса для выбранной клинической задачи обоснованным с точки зрения доказательной медицины? 3. Как влияет результат работы ИИ-сервиса на тактику ведения пациента? 4. Повышается ли качество работы врача? 5. Повышается ли скорость работы врача?
Аналитическая валидация	<i>А) Техническая готовность</i> 1. Самостоятельный тест. 2. Хронометраж, замер времени на обработку 1-го исследования. 3. Формат выходных данных. 4. Визуализация результатов работы ИИ-сервиса на рабочей станции врача-рентгенолога. 5. Формат входных данных (DICOM, JPEG, PNG, BMP, AVI (копия)). <i>Б) Детали создания и применения ИИ-модели</i> 1. Какие базы данных использовались при построении ИИ-моделей (из открытых ресурсов, от внешних партнеров, собственные, аугментированные)? 2. Баланс классов в базах данных. 3. Общее количество данных, использованных для построения ИИ-модели для каждой клинической задачи. 4. Архитектура ИИ-модели. 5. Библиотеки, ресурсы, использованные для построения модели. 6. Подходы к разработке модели (from scratch, transfer learning, retraining of existing architecture). 7. Тип машинного обучения (supervised / unsupervised / semi-supervised / weak supervision). 8. Валидация модели (independent validation set, cross-validation).

Продолжение таблицы 13

	9. Длительность создания ИИ-моделей. Имеющиеся вычислительные ресурсы. 10. Происходит ли дообучение ИИ-моделей? Как организована версионность? 11. ИИ-сервис работает с другими данными, кроме изображений? 12. ИИ-сервис способен оценить техническое качество проведения исследования? 13. Есть ли предобработка изображений? 14. Способен ли ИИ-сервис оценить качество интерпретации исследования (т.е. контроль работы врача-рентгенолога)? 15. Значения метрик оценки точности работы ИИ-сервиса для каждой клинической задачи (AUC, sens, spec, etc.). решения таких же клинических задач? 16. Сравнивались ли ИИ-модели с существующими аналогами для решения таких же клинических задач? <i>В) Детали использованных наборов данных и их разметки для разработки ИИ-моделей</i> 1. Были ли критерии для подборки базы данных для обучения или нет? 2. Были ли в базе для обучения разные исследования одного пациента (чтобы обучаться распознавать динамику)? 3. Что в каждом клиническом сценарии считалось «ground truth» в размеченной базе данных? 4. Были ли критерии контроля качества собранной базы данных перед разметкой? 5. Какой бизнес-процесс разметки? 6. Проходила ли подборка разметчиков? Критерии? 7. Был ли контроль качества входных данных и результатов разметки? 8. Сколько человек участвовали в разметке данных (сколько человек на 1 исследование; была ли выполнена разметка одного исследования одним специалистом дважды)? 9. Проходили ли разметчики обучение перед разметкой? 10. ПО для разметки свое или внешнее? Какое? 11. Подход в случае несогласованности разметчиков. 12. Бизнес-процесс работы эксперта, подтверждающего находку: – рабочий список; – доразметка; – цифровой след (серия комментариев от эксперта к разметчику). 13. Использовалась ли автоматизация процесса разметки (CAD для оконтуривания)? 14. Мониторинг процесса разметки
Клиническая валидация	<i>А) Практическое применение в клинической работе</i> 1. Происходила ли интеграция с PACS в медицинской организации? Если нет, то каким образом происходила интеграция в рабочий процесс?

Продолжение таблицы 13

	2. Какой бизнес-процесс работы врача-рентгенолога с ИИ-сервисом в практике? 3. Есть ли инструкции для врача по работе с ИИ-сервисом? 4. Вопросы к врачу: как он обучался и есть ли инструкции применения ИИ-сервиса и возможности для обратной связи? 5. Есть ли триаж исследований? 6. Есть ли локализация находки с помощью ИИ? 7. Есть ли классификация находки? 8. Есть ли сравнение исследований в динамике? 9. Есть ли формирование шаблона описания и/или заключения? 10. На каком этапе бизнес-процесса врач-рентгенолог получает результаты обработки ИИ-сервисом? 11. Есть ли обратная связь от врача-рентгенолога по использованию ИИ? Какая? Если есть, то она обязательная? 12. Что происходит, если врач-рентгенолог и ИИ имеют разные мнения (гипердиагностика и гиподиагностика)? Каким образом принимается окончательное решение по исследованию? 13. Есть ли метрики использования ИИ в клинической практике? Какие (например, сокращение количества ошибок врачей-рентгенологов и сокращение времени работы по интерпретации одного исследования)?
--	---

Клиническая апробация (оценка эффективности путем использования в рамках стандартного производственного процесса) состоит из двух компонентов (рис. 16).

– клинической корреляции (оценки наличия достоверной клинической связи между результатами и целевым клиническим состоянием ПО);

– клинической валидации (подтверждение достижения намеченной цели в целевой популяции в контексте лечебной работы благодаря использованию точных и надежных выходных данных).

На первом этапе определение клинического значения осуществляется путем:

а) оценки надежности юридического лица (разработчика ПО), в том числе посредством проверки на соответствие корпоративной культуре превосходного качества и организации;

б) генерации и анализа новых диагностических данных для конкретного предполагаемого использования ПО:

– формируется эталонный набор подготовленных биомедицинских данных;

Таблица 14 – Оценки составляющих корпоративной культуры юридического лица-разработчика ПО на основе ИИ (для лучевой диагностики) на соответствие корпоративной культуре превосходного качества и организации (КПКО)*

Пункт	Наименование	Баллы (20 – удовлетворительно, 15 – наличие некритических замечаний, 5 – наличие критических замечаний, 0 – нет результата)
А.	Безопасность пациентов. Демонстрация обязательств обеспечения безопасности пациента с выделением данного принципа в качестве основного при принятии любых решений	
Б.	Качество продукта. Демонстрация обязательств разрабатывать, тестировать и поддерживать медицинское программное обеспечение на самом высоком уровне качества	
В.	Клиническая ответственность. Демонстрация ответственности в проведении клинических оценок с должным вниманием к пациент-ориентированным вопросам, включая разметку и человеческий фактор	
Г.	Обеспечение кибербезопасности. Демонстрация обязательств по обеспечению кибербезопасности, с активным привлечением к данному рода проблемам партнеров и посредников	
Д.	Активная корпоративная культура. Демонстрация активного подхода к наблюдению и оценке потребностей пользователей, гибкости и непрерывного обучения	
Итоговый балл (максимум 100)		
* Developing a Software Precertification Program: A Working Model.V.1.0 – January 2019. URL: https://www.fda.gov/media/119722/download .		

- осуществляется автоматизированный анализ набора (как имитация диагностического процесса);
- результаты анализа сопоставляются с эталонным подготов-

ленным набором данных;

- проводится математический и статистический анализ (набор типовых метрик представлен в параграфе 4 данной главы).

На втором этапе испытываемое программное обеспечение применяется по целевому назначению в рамках типового производственного процесса. Процедуры, выполняемые с применением ПО на основе искусственного интеллекта, оцениваются с позиций производительности и качества. Соответственно, проводится сопоставительный хронометраж (при этом измерения должны проводиться многократно, с участием различных сотрудников-исполнителей, в разное время суток и периоды рабочих смен) и ретроспективный пересмотр результатов обследований. Организаторами испытания должны быть предусмотрены меры по минимизации рисков и обеспечению безопасности (безвредности для пациентов и их законных представителей) производственных процессов.

В программе проведения клинических испытаний должны быть указаны:

- детальные сведения о наборе данных (эталонных биомедицинских данных), включая номер свидетельства о государственной регистрации (желательно), количество и характер данных, сведения о разметке (см. далее);

- сведения о необходимости инсталляции ПО с применением технических средств испытательной медицинской организации;

- сведения об использовании вычислительных мощностей, предоставленных разработчиком ПО;

- сведения о деинсталляции ПО после завершения испытаний (в связи с тем, что ПО устанавливается и применяется на соответствующем оборудовании до получения регистрационного удостоверения).

В целом, формирование программы рекомендуется осуществлять в соответствии с чек-листом (таблица 15).

Таблица 15 – Чек-лист для детального описания результатов клинических испытаний программного обеспечения, разработанного для лучевой диагностики на основе технологий искусственного интеллекта (на основе STARD 2015 [7] с дополнениями)

Раздел и тема	№	Пункт
НАЗВАНИЕ		
	1	Акцент на оценку диагностической точности программного обеспечения на основе технологий ИИ, предназначенного для лучевой диагностики
РЕЗЮМЕ		
	2	Структурированное изложение дизайна исследования, материалов и методов, результатов и заключения
ВВЕДЕНИЕ		
	3	Клинические и научные сведения, включающие практическое применение результатов нового (исследуемого) метода – индекс-теста
	4	Гипотеза, цели и конечные точки исследования
МЕТОДЫ		
Дизайн исследования	5	Проспективный или ретроспективный характер исследования
Выборка	6	Критерии включения в исследование (включая целевую патологию)
	7	На основании каких критериев проведен отбор пациентов (симптомы, результаты обследований и т.д.)
	8	Где и когда сформирован набор биомедицинских данных с указанием ключевых характеристик (государственная регистрация, характеристика популяции, сведения о деперсонализации, целевая патология и диагностические группы с распределением, характеристика исследований, верификация, методология разметки)
	9	Запланированный размер выборки и способ его определения
	10а	Выборка сформирована последовательно, случайно или другими способами. Обоснование размера выборки

Продолжение таблицы 15

Раздел и тема	№	Пункт
Методы исследования	10b	Информация об индекс-тесте в достаточном для воспроизведения результатов объеме (включая сведения об использованных при разработке математических моделях, нейронных сетях, методах машинного обучения, выборках для обучения и калибровки)
	11	Информация о референс-тесте («золотом стандарте») в достаточном для воспроизведения результатов объеме (референс-тестом является размеченный набор данных (эталонный))
	12a	Обоснование выбора референс-теста (методологии разметки, всего набора данных)
	12b	Порог активации (точка отсечения, cut-off) для индекс-теста, включая разделение исходов на «ожидаемые» и «непредвиденные». Метод определения порога активации (cut-off): минимальное расстояние от верхнего левого угла до ROC-кривой, индекс Юдена, в зависимости от целей исследования
	13a	Порог активации (порог решающего правила, cut-off) для референс-теста, включая разделение исходов на «ожидаемые» и «непредвиденные». Метод определения порога активации: активации (cut-off): минимальное расстояние от верхнего левого угла до ROC-кривой, индекс Юдена, в зависимости от целей исследования
	13b	Была ли доступна участвующим в проведении/анализе индекс-теста критичная клинико-лабораторная информация, включая результаты применения референс-теста (было ли исследование слепым или нет, во втором случае – обоснование причины)
	14	Была ли доступна участвующим в разметке и анализе эталонных данных критичная клинико-лабораторная информация, включая результаты применения индекс-теста (было ли исследование слепым или нет, во втором случае – обоснование причины)
Анализ	15	Методы оценки и сравнения диагностической точности (метрики)
	16	Тактика при неоднозначных результатах индекс-теста или референс-теста
	17	Тактика при потере данных индекс-теста или референс-теста
	18	Анализ причин вариабельности диагностической точности с разделением на «ожидаемые» и «непредвиденные»

Продолжение таблицы 15

Раздел и тема	№	Пункт
РЕЗУЛЬТАТЫ		
Пациенты	19	Характеристика пациентов: диспозиция пациентов, определение популяции для анализа, число пациентов, включенных в исследование, число досрочно выбывших, число пропущенных или негодных данных. Рекомендуется в дополнение к текстовому описанию представить диаграмму CONSORT
	20	Демографические и клинические характеристики выборки на момент первичного обследования (всех субъектов, а также субъектов, у которых есть результаты обоих тестов)
	21a	Распределение по степени тяжести целевой патологии
	21b	Распределение прочих диагнозов (по Международной классификации болезней) у пациентов без целевой патологии
	22	Любые значимые различия в методиках проведения индекс-теста и референс-теста
Результаты	23	Комбинированная таблица результатов индекс-теста и референс-теста
	24	Подсчет метрик диагностической точности с определенным доверительным интервалом (например, 95%)
	25	Любые побочные эффекты вследствие индекс-теста или референс-теста
ОБСУЖДЕНИЕ		
	26	Ограничения исследования, включая источники систематических ошибок, статистических неточностей и обобщений
	27	Клиническая ценность, включая планируемое применение индекс-теста
ПРОЧЕЕ		
	28	Сведения о регистрации исследования (любого этапа)
	29	Где можно получить полные сведения о разработке (включая информацию о наборе биомедицинских данных для обучения и калибровки)
	30	Источники финансирования и прочее

§4. Метрики качества

Оценка диагностической точности. Аналитический этап осуществляется в дизайне диагностического исследования, при этом используется стандартный набор метрик.

Автоматизированный анализ медицинских данных посредством исследуемого ПО на основе ИИ считают индекс-тестом. Результаты разметки набора биомедицинских данных в соответствии с известной методологией – референс-тестом.

Для сопоставления результатов индекс- и референс-теста составляют четырехпольную таблицу и определяют абсолютное значение для каждого вида результатов (таблица 16). Далее производят выбор и расчет релевантных метрик (таблица 17) с определением достоверности в рамках 95% доверительного интервала.

Таблица 16 – Виды результатов применения ИИ в соответствии с четырехпольной таблицей (на примере ответа, представляющего собой бинарную переменную*)

Вид результата			
Истинно положительный	Истинно отрицательный	Ложноположительный	Ложноотрицательный
ИП	ИО	ЛП	ЛО
ПО выявило патологию при ее наличии	ПО не выявило патологию при ее отсутствии	ПО выявило патологию при ее отсутствии	ПО не выявило патологию при ее наличии
*Данный вид ответа соответствует большинству типовых задач; вместе с тем большинство иных задач практически всегда могут быть сведены к задаче бинарной классификации.			

Таблица 17 – Типовой выбор метрик для аналитического этапа исследования*

Действие ИИ	Базовые метрики
Обнаружение (детекция)	Стандартный набор метрик
Классификация	Стандартный набор метрик
Сегментация	Коэффициент сходства
Обработка естественного языка	Коэффициент согласованности классификаций
* Указанные метрики являются базовыми, обязательными для оценки; дополнительно могут быть использованы и другие метрики, применимые к конкретной задаче.	

В зависимости от решаемой задачи могут применяться различные сочетания метрик. Например, анализ характеристической кривой и определение коэффициента согласованности классификаций и т. д. Такие основные показатели, как чувствительность, специфичность, площадь под характеристической кривой, должны быть использованы для характеристики каждого типового действия искусственного интеллекта.

Стандартный набор метрик. Для сопоставления эффективности индекс-теста относительно референс-теста применяется классический набор показателей, характеризующих диагностическую результативность (таблица 18).

Таблица 18 – Стандартный набор метрик диагностической ценности

Показатель	Значение	Формула
Чувствительность	Вероятность того, что индекс-тест окажется позитивным при наличии признака	$ИП / (ИП + ЛО)$
Специфичность	Вероятность того, что индекс-тест окажется негативным при отсутствии признака	$ИО / (ИО + ЛП)$
Точность (общая валидность)	Уровень соответствия полученных результатов индекс-теста к действительным значениям референс-теста	$(ИП + ИО) / (ИП + ИО + ЛП + ЛО)$
Отношение правдоподобия положительного результата	Математически отображает, во сколько раз при положительном результате индекс-теста вероятность наличия целевой патологии превышает вероятность ее отсутствия	$Чувствительность / (1 - специфичность)$
Отношение правдоподобия отрицательного результата	Математически отображает, во сколько раз при отрицательном результате индекс-теста вероятность наличия целевой патологии превышает вероятность ее отсутствия	$(1 - чувствительность) / специфичность$
Прогностическая ценность положительного результата	Вероятность заболевания при положительном результате индекс-теста	$ИП / (ИП + ЛП)$
Прогностическая ценность отрицательного результата	Вероятность отсутствия заболевания при отрицательном (нормальном) результате индекс-теста	$ИО / (ИО + ЛО)$
Частота ложных срабатываний	Вероятность ложного срабатывания	$1 - специфичность$

Все метрики, кроме отношения правдоподобия, **оцениваются** в диапазоне 0–1 или в процентах от 0 до 100%:

- <0,6 – непригодно;
- 0,61 – 0,8 – требуется доработка;
- >= 0,81 – может быть допущено к клинической валидации.

Отношение правдоподобия положительного результата должно быть как можно выше, тогда как отношение правдоподобия отрицательного результата — как можно ниже.

Характеристическая кривая (ROC-кривая). Характеристическая кривая — отображение соотношения между долей объектов от общего количества носителей признака, верно классифицированных как несущих признак (то есть чувствительностью), и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущих признак (то есть специфичностью) при варьировании порога активации. Метрика диагностической ценности: площадь под кривой (AUC) — площадь, ограниченная ROC-кривой и абсциссой.

Классическая ROC-кривая: график зависимости чувствительности от 1-специфичности (по оси абсцисс — 1-специфичность, по оси ординат — чувствительность). При анализе ROC-кривой необходимо провести определение оптимальных значений порога активации (так называемого cut-off). Для этого используются различные методики. В качестве примера (но не обязательного требования) можно привести следующие:

1) минимальное расстояние от верхнего левого угла до ROC-кривой (минимум d);

2) индекс Юдена (Youden index), который отображает максимальное расстояние от диагональной линии до ROC-кривой (рис. 16).

Приоритет той или иной методике определения cut-off должен отдаваться в зависимости от целей исследования.



Рисунок 16 — Методика определения оптимальных значений порога активации (cut-off)

Площадь под кривой (AUC) **оценивается** в диапазоне 0–1:

- <0,6 – непригодно;
- 0,61–0,8 – требуется доработка;
- >=0,81 – может быть допущено к клинической валидации.

Коэффициент согласованности классификаций. Согласованность (конкордантность) – математическое отображение согласованности классификации двух экспертов относительно одного явления.

Формула расчета (1) по методу Козна (каппа Козна) (Cohen J., 1960):

$$K = (P_0 - P_e) / (1 - P_e), (1)$$

где P_0 – доля случаев, когда измерения совпадали, и P_e – ожидаемая доля эпизодов случайного совпадения. Порядок расчета (2):

Эксперт №2	Эксперт №1			
		1	2	Всего
	1	P11	P12	P1_
	2	P21	P22	P2_
	Всего	P_1	P_2	1

$$P_0 = P_{11} + P_{22}$$

$$P_e = P_{_1}P_{1_} + P_{_2}P_{2_}, (2)$$

Оценивается в диапазоне 0–1 или в процентах от 0 до 100%:

- <0,6 – непригодно;
- 0,61–0,8 – требуется доработка;
- >=0,81 – может быть допущено к клинической валидации.

Коэффициент сходства. Бинарный показатель для количественного определения степени сходства биологических объектов. Формула расчета (3) по методу Дайса-Серенсена (Dice L., 1945, Sørensen T., 1948):

$$DSC = 2|A \cap B| / (|A| + |B|), (3)$$

где в числителе — удвоенное количество совпавших оценок, в знаменателе — общее количество оценок.

Для множеств с одинаковой мощностью **оценивается** в диапазоне 0–1:

- $<0,6$ — непригодно;
- $0,61-0,8$ — требуется доработка;
- $\geq 0,81$ — может быть допущено к клинической валидации.

Клиническая апробация. Клиническая апробация подразумевает оценку влияния исследуемого ПО на показатели деятельности медицинского персонала, отдельных подразделений и медицинских организаций в целом.

Хронометраж — метод оценки результативности применения ПО, фиксация и замеры продолжительности выполняемых действий (процедур, операций) в рамках производственных процессов. Например, продолжительность описания цифровой рентгенографии легких, выполненной в профилактических целях с и без искусственного интеллекта. Базовая методика выполнения строится на способе определения норм времени¹.

Аудит результатов исследований — ретроспективный анализ качества интерпретации медицинских данных; например, проведения и описания лучевых исследований, выполненных с применением технологий искусственного интеллекта, с целью определения степени соответствия рекомендуемым стандартам. Аудит позволяет оценить качество оказания медицинской помощи, включая проведение и интерпретацию результатов диагностических исследований с и без использования ПО на основе технологий искусственного интеллекта; достоверность различий или их отсутствия определяется статистически. Базовая методика выполнения строится на способе организации и осуществления анализа рентгенорадиологических изображений². Могут применяться иные методики (например, социологический опрос медицинских работников) и показатели, в том числе относительные величины,

¹ Методика разработки норм времени и нагрузки медицинского персонала. М.: ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, 2013. 25 с.

² Оценка качества рентгенорадиологических исследований / сост. Н. Н. Ветшева, Н. В. Ледихова // Сер. «Лучшие практики лучевой и ин-

рассчитываемые в соответствии с правилами медицинской статистики.

§5. Оформление результатов

При оценке результатов ПКТИ и клинических испытаний необходимо учитывать следующее:

- ПО проверяется с применением биоматериалов и всего необходимого оборудования (в том числе других медицинских изделий), предназначенного для обеспечения надлежащей работы исследуемого ПО;

- в процессе проведения клинических испытаний должны быть задействованы все модули ПО (при модульной архитектуре), опробованы все функции, указанные производителем в эксплуатационной документации;

- должны быть оценены все позиции доказательных баз ПО (клинические данные, анализ взаимозаменяемых медицинских изделий и т.п.), представленных производителем (изготовителем) в испытательную медицинскую организацию;

- испытательная медицинская организация в обязательном порядке оценивает документацию на программный продукт, включая техническую и эксплуатационную, делает вывод и о том, что содержание документа доступно и понятно специалистам соответствующего профиля, и о том, что его содержание обеспечивает надлежащее эффективное и безопасное медицинское применение.

Результаты клинических испытаний ПО (как и любого другого медицинского изделия) должны быть оформлены в соответствии с приложениями к приказу Министерства здравоохранения Российской Федерации от 09.01.2014 №2н «Об утверждении Порядка проведения оценки соответствия медицинских изделий в форме технических испытаний, токсикологических исследований, клинических испытаний в целях государственной регистрации медицинских изделий³».

струментальной диагностики». Вып. 22. М., 2019. 50 с.

При подготовке документации необходимо формировать подробные результаты оценки и анализа данных в соответствии с дополненным чек-листом STARD 2015 (таблица 15), которые рекомендуется публиковать в формате научных статей в рецензируемых научных журналах.

* * *

В разделе «Введение» мы уже рассказали, что к 2019 году в глобальной перспективе итоги работы по формированию методологических возможностей медицинского ИИ были неутешительными. К этому времени путем многолетних аналитических и научных исследований мы создали ключевые компоненты методологии применения технологий искусственного интеллекта в лучевой диагностике.

Собственные результаты, опыт разработки методологий и скрупулезный анализ международной ситуации, представленные в предыдущих главах, позволили нам сформировать стратегию дальнейшего развития — ***необходимо проведение проспективного мультицентрового комплексного клинического исследования неограниченного числа алгоритмов искусственного интеллекта в условиях реального лечебно-диагностического процесса.***

³ Приказ Министерства здравоохранения Российской Федерации от 09.01.2014 №2н «Об утверждении Порядка проведения оценки соответствия медицинских изделий в форме технических испытаний, токсикологических исследований, клинических испытаний в целях государственной регистрации медицинских изделий».

Часть вторая. ВНЕДРЕНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВУЮ ДИАГНОСТИКУ

Стратегическая важность развития технологий искусственного интеллекта обусловила принятие Правительством Москвы решения о проведении проспективного мультицентрового комплексного клинического исследования неограниченного числа алгоритмов искусственного интеллекта в условиях реального лечебно-диагностического процесса. С учетом накопленного опыта, технологической и методологической готовности предметной областью исследования стала лучевая диагностика. В соответствии с Постановлением Правительства Москвы от 21 ноября 2019 г. №1543-ПП в системе здравоохранения города Москвы организован эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы (далее — Эксперимент)¹.

В Эксперименте должны принять участие юридические лица — резиденты Российской Федерации, разработавшие или имеющие права на предоставление сервисов на базе технологий компьютерного зрения для анализа медицинских изображений (они именуются ИИ-сервисами). Организационная, научная и методическая работа в Эксперименте должна осуществляться Департаментом здравоохранения города Москвы и уполномоченным ГБУЗ г. Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий ДЗМ». Технологическая составляющая Эксперимента обеспечивалась Департаментом информационных технологий города Москвы. Для поддержки и мотивации юридических лиц, представляющих ИИ-сервисы, Правительство Москвы предусмотрело выделение грантов на исследовательские проекты.

Также постановлением Правительства Москвы от 21 ноября 2019 г. №1543-ПП установлены в общем виде порядки предоставления и рассмотрения заявок на получение грантов, собственно механизмы расчета объемов и условия соответствующих выплат. Детализированные порядки и условия проведения Эксперимента,

¹ Постановление Правительства Москвы от 21.11.2019 №1543-ПП «О проведении эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы».

а также состав и положение о комиссии, рассматривающей организационно-финансовые вопросы, регламентированы приказом Департамента здравоохранения города Москвы от 19 февраля 2020 года №142. Этот нормативный документ устанавливает цели и задачи Эксперимента, регламентирует процессы, применяемые методологии, определяет зоны ответственности, права и обязанности всех участвующих сторон, а также — форму заявки на участие в Эксперименте и соответствующее соглашение.

С момента вступления приказа в силу началась практическая работа по подготовке и осуществлению Эксперимента. В ходе его выполнения были получены новые знания, потребовавшие пересмотра некоторых положений приказа, их конкретизации. Основной целью пересмотров стало обеспечение должного уровня качества и точности ИИ-сервисов, допускаемых к использованию в условиях реального здравоохранения, а также обеспечение максимальной объективности, «прозрачности» и наукоемкости Эксперимента.

В целом, порядок и правила проведения Эксперимента устанавливались следующими нормативно-правовыми документами:

- Постановление Правительства Москвы №1543-ПП «О проведении эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы»;

- Постановление Правительства Москвы от 28.04.2020 №464-ПП «О внесении изменений в постановление Правительства Москвы от 21.11.2019 №1543-ПП»;

- Приказы Департамента здравоохранения города Москвы от 19.02.2020 №142 и от 13.10.2020 №1199 «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».

В дальнейших главах подробно описан комплекс работ по организации и проведению Эксперимента, а также результаты этой деятельности.

Глава 6. ОРГАНИЗАЦИОННАЯ, ТЕХНИЧЕСКАЯ И МЕТОДОЛОГИЧЕСКАЯ ПОДГОТОВКА ЭКСПЕРИМЕНТА

Организацию проведения и сопровождение Эксперимента осуществляло государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы» (далее – ГБУЗ «НПКЦ ДиТ ДЗМ»). Организационно-методическая подготовка и проведению Эксперимента включала следующие компоненты:

- 1) разработку нормативно-правового обеспечения;
- 2) обеспечение технологической инфраструктуры;
- 3) организацию и проведение комплекса работ по технологической интеграции ИИ-сервисов в ЕРИС ЕМИАС;
- 4) организацию и проведение процедур отбора, оценки, включения, мониторинга участников Эксперимента, а также выплаты предусмотренных грантов;
- 5) организацию информирования и вовлечения компаний-работчиков ИИ-сервисов в Эксперимент;
- 6) организацию информирования, обучения и вовлечения врачей-рентгенологов;
- 7) разработку специальных методологий и процедур (как для операционных, так и для научных задач Эксперимента);
- 8) обеспечение эффективной системы коммуникаций со всеми участниками Эксперимента;
- 9) информирование профессиональной аудитории и широкой общественности о ходе Эксперимента.

Далее приведено описание основных результатов данных мероприятий, имеющих значение с научной и методической точек зрения.

§1. Основные организационные аспекты Эксперимента

В IV квартале 2019 года коллективом ГБУЗ «НПКЦ ДиТ ДЗМ» под патронатом Департамента здравоохранения г. Москвы и при участии сотрудников Департамента информационных технологий г. Москвы разработаны:

- проект приказа Департамента здравоохранения Москвы с детальным описанием всех аспектов Эксперимента;
- основные производственные процессы;
- методологии обеспечивающих процедур.

В 2020 году на базе ГБУЗ «НПКЦ ДиТ ДЗМ» была организована бесперебойная работа по проведению процедур привлечения, информирования, отбора, оценки, включения, мониторинга, поддержки, научного анализа участников Эксперимента; организационно-технически обеспечена *работа комиссии* Департамента здравоохранения Москвы (ДЗМ); ведение соответствующего документооборота, в том числе для выплаты предусмотренных грантов. Данная комиссия создана для рассмотрения заявок на участие и выплату грантов, результатов регулярного контроля работы участников Эксперимента и решения иных организационных вопросов.

Организован процесс получения и обработки *заявок от потенциальных участников* Эксперимента для последующего рассмотрения комиссией ДЗМ в установленном порядке. Всего в 2020 году было получено 48 заявок от 24-х компаний, принято к участию в Эксперименте 81,0% (21) заявок, отклонено – 11,0% (5) из-за несоответствия установленным нормативным требованиям. По состоянию на 25.12.2020 8,0% (4) заявок оставались на рассмотрении в силу ожидания от заявителя запрошенной дополнительной информации.

Для участия в Эксперименте от юридического лица требовалось предоставить следующие обязательные документы:

- заявку по установленной форме;
- заверенные копии уставных документов;
- документы, подтверждающие право на предоставление предлагаемого ИИ-сервиса и описывающие функции, техническую архитектуру и характеристики, предусмотренные сценарии исполь-

зования сервиса, требования к программно-аппаратным средствам для работы сервиса;

- отчет о предварительных клинико-технических испытаниях, выполненных в соответствии с методическими рекомендациями¹.

Дополнительно рекомендовалось предоставить:

- действующий сертификат соответствия системы менеджмента качества стандарту ISO;

- заверенные копии государственной сертификации ИИ-сервиса в качестве медицинского изделия;

- заверенные копии сертификатов соответствия качества предлагаемого участником сервиса региональным стандартам FDA, CE и их аналогам в других странах;

- копии научных статей об оценке точности и эффективности предлагаемого ИИ-сервиса;

- оригиналы или копии отчетов о клиническом внедрении или применении ИИ-сервиса в медицинских организациях Российской Федерации или других стран.

Определены критерии для принятия решения комиссией ДЗМ о включении ИИ-сервиса в Эксперимент (таблица 19). С учетом совершенной новизны всего процесса первоначально критерии были сформулированы общими, базовыми, создающими максимальные возможности для вовлечения компаний-разработчиков.

¹ Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика) / сост. С. П. Морозов, А. В. Владзимирский, В. Г. Кляшторный [и др.] // Сер. «Лучшие практики лучевой и инструментальной диагностики». М., 2019. Вып. 57. 51 с.

Таблица 19 – Критерии оценки ИИ-сервисов

Формальные (есть в наличии)	Переменные	
	Количественные	Качественные
Документы, подтверждающие его право на предоставление предлагаемого сервиса на основе компьютерного зрения для анализа медицинских изображений	Длительность апробации предлагаемого сервиса в медицинских организациях Российской Федерации или других стран	Качество и полнота предоставленной документации, описывающей функции, техническую архитектуру и характеристики, предусмотренные сценарии использования сервиса, требования к программно-аппаратным средствам для работы сервиса
Заверенные копии сертификатов соответствия качества предлагаемого участником сервиса региональным стандартам FDA, CE и их аналогам в других странах или государственной регистрации в качестве медицинского изделия	Количество медицинских учреждений, в которых выполнялась апробация предлагаемого сервиса	Наличие и качество реализации в предлагаемом сервисе функциональности автоматического анализа медицинских изображений с указанием локализации выявленных сервисом патологических находок и уведомлением о результатах
Копии научных статей об оценке точности и эффективности предлагаемого сервиса в рецензируемых журналах, индексируемых Scopus и/или Web of Science	Длительность анализа одного исследования предлагаемым сервисом	Наличие и качество реализации в предлагаемом сервисе возможности формирования в результате анализа исследования – серии в формате DICOM, которая содержит информацию о наличии или отсутствии патологий, результаты анализа с визуализацией находок на исходных изображениях
Оригиналы или копии отчетов о клиническом внедрении или апробации предлагаемого сервиса в медицинских организациях Российской Федерации или других стран	Критерии диагностической точности предлагаемого сервиса, в том числе: классическая ROC-кривая, площадь под кривой (AUC); удельный вес ложноотрицательных результатов; удельный вес ложноположительных результатов	Возможность поддержки предлагаемым сервисом стандарта HealthLevel7 (HL7)/FHIR

Продолжение таблицы 19

Формальные (есть в наличии)	Переменные	
	Количественные	Качественные
Отчет о предварительных клинико-технических испытаниях в соответствии с утвержденной методологией	Иные критерии диагностической точности	Наличие и качество реализации в предлагаемом сервисе возможности автоматически предоставлять информацию для приоритизации исследований в рабочем списке врача-рентгенолога
Наличие результатов тестирования диагностической точности предлагаемого сервиса на деперсонализированных медицинских снимках граждан Российской Федерации и/или лиц европеоидной и монголоидной рас		Наличие и качество реализации в предлагаемом сервисе функциональности сравнительного анализа исследований одного пациента, выполненных в разное время (оценка динамики патологических процессов)
		Наличие и качество реализации в предлагаемом сервисе функциональности автоматической подготовки проекта описания рентгенологического исследования

Сотрудниками ГБУЗ «НПКЦ ДиТ ДЗМ» налажена эффективная система коммуникаций со всеми сторонами, вовлеченными в Эксперимент, включая списки рассылок, выделенные телефонные номера, тематические чаты и группы в интернет-мессенджерах, также разработан и запущен *официальный веб-сайт Эксперимента* www.mosmed.ai и создан электронный почтовый ящик ai@npsmr.ru.

Интенсивность коммуникаций наглядно иллюстрирует следующий факт: ежедневно на официальный электронный адрес Эксперимента поступало от 30 до 50 обращений. В результате анализа коммуникации с компаниями было выявлено 120 типовых вопросов, на которые были подготовлены стандартные справочные ответы и опубликованы на сайте www.mosmed.ai.

Для информирования профессиональной аудитории и широкой общественности о ходе Эксперимента осуществлялся план публикаций в средствах массовой информации и социальных медиа. Всего в 2020 году было выпущено 2742 тематических материала новостного, научно-популярного и публицистического характера об Эксперименте, что составило 57,0% от общего объема публикаций в масс-медиа по проблематике ИИ и смежным темам. Ряд публикаций и сообщений были сделаны ТАСС и РИА Новости. Общий охват аудитории превысил 110 миллионов.

Особым направлением работы стало широкое информирование и *вовлечение компаний-разработчиков* в Эксперимент. В IV квартале 2019 года был проведен масштабный поиск информации о разработчиках из России и из-за рубежа; составлен общий список потенциальных компаний (n=176), посредством разных каналов коммуникаций направлены повторяющиеся приглашения. Из-за несоответствия предлагаемых ИИ-сервисов условиям Эксперимента (трем предусмотренным видам исследований) 21 (12,0%) компанию исключили из рассылки, таким образом, приглашения получили 155 разработчиков. Из этого числа 20 (13,0%) компаний являлись российскими, 135 (87,0%) — зарубежными, соответственно был обеспечен 100,0% охват рынка в РФ и 65,0% на международном уровне¹.

В итоге в Эксперименте приняла участие 21 компания, в том числе 8 российских представителей зарубежных разработчиков ИИ-сервисов (согласно нормативно-правовому обеспечению Эксперимента). В соответствии с условиями Эксперимента каждое юридическое лицо могло представить для участия несколько отдельных сервисов по разным видам исследований. Поэтому общее число участвовавших в Эксперименте ИИ-сервисов составило 39. Таким образом, Эксперимент охватил 14,0% рынка разработчиков технологий искусственного интеллекта для лучевой диагностики.

В силу разного уровня технологической готовности и сроков начала интеграции ИИ-сервисы достигли разного прогресса (таблица 20). Участниками активной фазы Эксперимента (реальной ра-

¹ Tadavarthi Y., Vey B., Krupinski E. et al. The State of Radiology AI: Considerations for Purchase Decisions and Current Market Offerings. Radiology: Artificial Intelligence. 2020. Vol. 2, №. 6. P. e200004.

боты с потоком результатов лучевых исследований) стали 18 ИИ-сервисов (46,0% из общего числа начавших интеграцию).

Таблица 20 – Прогресс ИИ-сервисов в Эксперименте

Количество ИИ-сервисов / Модальность	КТ/НДКТ РЛ	КТ Covid	ММГ	РГ/ФЛГ	Итого
Общее количество	5	15	6	13	39
Техническая интеграция	1	2	0	2	5
Интегрированы в ТПАК	2	6	4	4	16
Интегрированы в ППАК	2	7	2	7	18

Для мониторинга общего процесса и основных метрик Эксперимента был разработан интерактивный интуитивный инструмент – **дашборд** (онлайн-табло). Основными средствами разработки выступили язык программирования Python и открытая библиотека Plotly. В качестве входных данных использовались выгрузка из ЕРИС ЕМИАС об обработанных исследованиях по специально разработанной форме и ряд других таблиц.

Ключевым функционалом дашборда является возможность доступа к информации с любого устройства компьютера, планшета или мобильного телефона с выходом в интернет для авторизованных пользователей и возможность установки фильтров по датам, ИИ-сервисам и их модальностям. Кроме этой базовой статистики, для пользователя доступна также следующая информация:

- количество подключенных медицинских организаций и диагностических устройств к каждому ИИ-сервису;
- количество проанализированных исследований в день по ИИ-сервисам и по их модальностям;
- количество отправленных на анализ и полученных результатов от ИИ-сервисов (каждым сервисом в неделю);
- медианное время ожидания результата от ИИ-сервиса по каждому ИИ-сервису за каждый день (от момента проведения исследования до получения результата);
- медианное время анализа исследования по каждому ИИ-сервису;

- статистика по количеству проанализированных исследований на карте Москвы;
- общий анализ обратной связи от врачей;
- анализ обратной связи от врачей с детализацией по медицинским организациям;
- количество обработанных исследований по каждой медицинской организации с детализацией по модальностям и ИИ-сервисам;
- статистика по отправленным/полученным исследованиям по каждому диагностическому устройству;
- количество исследований, подходящих для анализа ИИ-сервисами и отправленных ИИ-сервисам.

Дашборд размещен в интернете по адресу: www.dash.mos-med.ai; однако доступ к нему ограничен и предоставляется по логину и паролю. Для данного инструмента было получено свидетельство о государственной регистрации программы для ЭВМ¹.

§2. Технологические аспекты Эксперимента

Технологической средой для проведения Эксперимента стал Единый радиологический информационный сервис в составе Единой медицинской информационно-аналитической системы города Москвы (ЕРИС ЕМИАС). Подробно эта информационная система, образующая единый контур и общее цифровое пространство лучевой диагностики города Москвы, описана в главе 2.

Сервисы на основе искусственного интеллекта, включенные в Эксперимент, проходили процедуру интеграции с ЕРИС ЕМИАС для обеспечения бесшовных производственных процессов отделений лучевой диагностики. Процессы интеграции и работы ИИ-сервисов в ЕРИС ЕМИАС реализованы сотрудниками Департамента

¹ Свидетельство о государственной регистрации программы для ЭВМ №2020660360. Дашборд для мониторинга работы сервисов, основанных на использовании искусственного интеллекта, в здравоохранении. Заявка №2020619516. Дата поступления: 25.08.2020. Дата госрегистрации: 02.09.2020.

информационных технологий города Москвы в соответствии с функциональным заданием ГБУЗ «НПКЦ ДиТ ДЗМ» г. Москвы.

Концептуальная схема интеграции представлена на рис. 17. Для проведения интеграции в составе ЕРИС была специально разработана подсистема «Продукт управления моделями (ПУМ)» — «точка входа» для решений на основе искусственного интеллекта. В ЕМИАС применялась подсистема «Единая система уведомлений для внешних взаимодействий (ЕСУВВ)», построенная на продукте Apache Kafka (распределенный программный брокер сообщений). Также в составе ЕРИС была сформирована тестовая среда — так называемый тестовый программно-аппаратный комплекс (ТПАК). Это «точка» первоначального подключения ИИ-сервисов для выполнения инженерно-технических работ, настройки, функционального и калибровочного тестирования.

Легко понять, что при работе с ИИ-сервисами в рутинно работающей радиологической информационной системе могут возникать технические ошибки, приводящие к абсолютно недопустимым результатам — затруднению или остановке работы врача. С целью уменьшения рисков необходимо предварительное тестирование корректности работы ИИ-сервиса в данной системе, а в случае нашего Эксперимента — в ЕРИС. Для этого и был организован обособленный (тестовый) контур — ТПАК. На этапе интеграции ИИ-сервиса в ТПАК выполнялась проверка корректности работы и предоставляемых результатов в условиях возможной вариативности входных данных (различные диагностические устройства, наименования процедур и т.д.). В ТПАК исследователи и представители технической поддержки ЕРИС ЕМИАС тестировали функциональные возможности каждого ИИ-сервиса. Фактически в тестовом контуре сервис «адаптировался» к работе в ЕРИС ЕМИАС в условиях «бесшовной» интеграции.

После успешного завершения тестовых процедур ИИ-сервис «переключался» в рабочую среду ЕРИС — так называемый продуктивный программно-аппаратный комплекс (ППАК). Здесь уже и велась непосредственная работа с потоком реальных исследований.

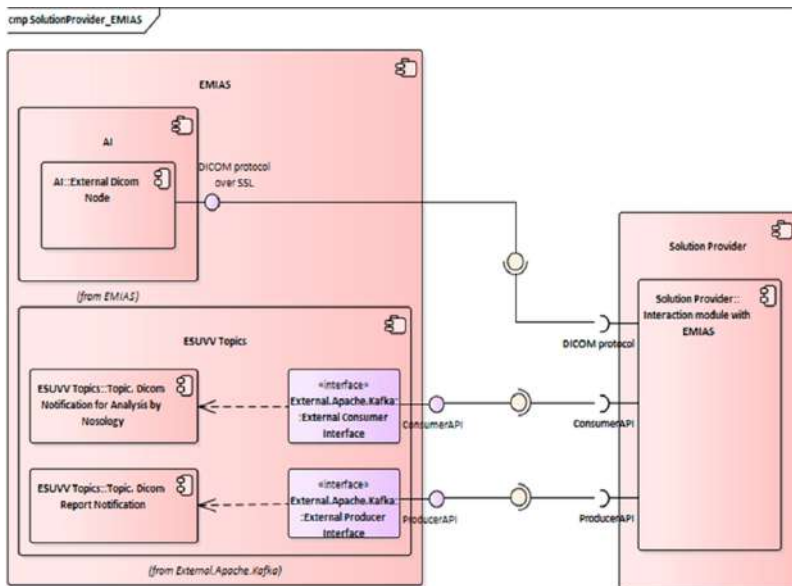


Рисунок 17 – Концептуальная схема интеграции ИИ-сервисов в ЕРИС ЕМИАС

Также были определены минимальные базовые требования к разработчикам ИИ-сервисов, в том числе:

- наличие в инфраструктуре компании временного хранилища DICOM-файлов для обмена исследованиями и результатами с ЕРИС ЕМИАС;
- наличие статического внешнего «белого» IP-адреса;
- наличие компетенций разработки приложений для работы с протоколом DICOM;
- наличие компетенций разработки приложений с взаимодействием с Apache Kafka.

Кроме того, разработчик ИИ-сервиса должен обеспечить:

1. Взаимодействие с ЕРИС ЕМИАС через:

а. ЕСУВВ (Apache Kafka API):

- в качестве подписчика – для получения уведомления о наличии исследования для анализа;
- в качестве продюсера – для отправки уведомления о выпол-

ненном анализе исследования и информации по отчету.

б. Узел DICOM (реализация протокола DICOM версии 3.0 — работает в роли SCU; необходима реализация TCP over SSL):

- получение исследований;
- публикация отчетов.

2. Анализ исследований:

а. Формирование отчетности по исследованию.

Были сформированы и представлены в виде справочников:

— база документации, расположенная в справочной системе ЕМИАС (wiki.emias.mos.ru);

- логика работы;
- требования и процедуры получения сертификатов;
- требования к численным и текстовым данным;
- рекомендации по тегам DICOM SR с примерами;
- требования к визуальному представлению (формат результата DICOM SR, формат данных DICOM GSPS и CSPS, перечни и последовательности данных в случае обнаружения/отсутствия находок);

— сопроводительная, статистическая, сигнальная и иная информация.

Интеграция и работа ИИ-сервисов в ЕРИС ЕМИАС осуществлялись в строгом соответствии с действующим законодательством в сфере государственных информационных систем, информационной безопасности, защиты персональных данных. Данные, передаваемые для обработки ИИ-сервисами, были обезличены, т. е. по ним нельзя было идентифицировать физическое лицо, которому соответствуют те или иные результаты исследования:

А. Удалены и/или искажены данные, однозначно характеризующие физическое лицо, включая, но не ограничиваясь:

- фамилию, имя или имена, отчество (при наличии) пациента;
- дату рождения пациента и его возраст на момент проведения исследования;
- идентификатор (ы) исследования.

В. Удалены и/или искажены данные, позволяющие косвенно определить физическое лицо по совокупности входных данных, включая, но не ограничиваясь:

- дату и время проведения исследования;
- место проведения исследования;

- наименование учреждения, в котором было проведено исследование, и сопутствующую информацию;
- фамилии, имена, отчества (при наличии) медицинских работников, например, направляющих врачей, рентгенолаборантов;
- информацию об идентификационном номере диагностического оборудования;
- техническую информацию о проведенном исследовании.

ИИ-сервис не допускался к дальнейшему внедрению и испытаниям по соображениям безопасности в случае, если для его функционирования необходимы были любые данные из пункта А.

Если для функционирования ИИ-сервиса необходимы были данные из пункта В, то их предоставляли в искаженном виде либо не предоставляли совсем (в таком последнем случае ИИ-сервис не допускался к дальнейшему внедрению и испытаниям по соображениям безопасности).

Организацию и проведение комплекса работ по интеграции ИИ-сервисов, технической поддержке и сопровождению разработчиков, обеспечению информационной безопасности на высоком профессиональном уровне выполнил коллектив Департамента информационных технологий города Москвы.

§3. Обеспечение компетенций

Постоянное развитие системы знаний и компетенций — это необходимый элемент цифровизации здравоохранения. Ранее была предложена трехэтапная схема обучения медицинского персонала при внедрении телемедицинских технологий, включающая¹:

I этап — общее информирование (изучение истории, теории, базовых методологий), проводится перед непосредственным развертыванием технологического решения;

II этап — практическое обучение (освоение данного технологического решения, процедур, производственных процессов), проводится во время развертывания технологического решения;

¹ Владимирский А. В., Лебедев Г. С. Телемедицина. М.: ГЭОТАР-Медиа, 2018. 576 с.

III этап — разбор проблем (обсуждение дефектов, проблем и ошибок в работе, разбор сложных ситуаций, повторное изучение отдельных навыков), проводится через 4–6 месяцев после развертывания технологического решения.

Данная схема была успешно использована в процессе организации и проведения Эксперимента по компьютерному зрению. Акцент сделан на дистанционные и электронные формы преподавания. Был подготовлен лекционный материал, инструкции и руководства по эксплуатации ИИ-сервисов (как в общем виде, так и для каждого сервиса по отдельности), организован канал технической поддержки пользователей (чат в интернет-мессенджере), предусмотрен сбор обратной связи о качестве обучения.

На первом этапе в формате вебинаров проведены шесть дистанционных лекций общей длительностью 10 академических часов. На втором этапе врачам предоставлен свободный доступ к 21 оригинальному видеоуроку (на момент подготовки книги количество оригинальных просмотров этих материалов превышало 7500)¹. Содержание видеоуроков неоднократно актуализировалось и дополнялось. Также для врачей была подготовлена короткая инструкция (таблица 21).

Тематический план видеоуроков:

1. Искусственный интеллект: применение в лучевой диагностике.
2. Что является результатами работы ИИ-сервисов?
3. Работа с результатами работы ИИ-сервисов.
4. Как оценить качество работы ИИ-сервиса?
5. Оформление результатов исследования, содержащего результаты работы ИИ-сервиса.
6. Обучение работе с конкретным ИИ-сервисом (темы 6–21).

На третьем этапе проведены 2 очные рабочие встречи при участии наиболее активных и заинтересованных врачей.

¹ URL:<https://www.youtube.com/playlist?list=PLtdh74emfl1YEJP0mzueKB1X7GOo4SK2n>.

Таблица 21 – Базовая инструкция для врачей-рентгенологов по работе в условиях Эксперимента

№	Действия
1	Открыть рабочий список исследований. Исследования, проанализированные ИИ-сервисом, имеют маркировку. Рабочий список сортирован по приоритету анализа исследований (исследования, обозначенные цветом или с буквенной маркировкой, имеют максимальный приоритет – в них, по результатам обработки ИИ-сервиса, были обнаружены целевые патологические находки)
2	Выбрать и начать просмотр исследования. В случае отсутствия дополнительных указаний необходимо начинать с исследований с максимальным приоритетом
3	Проанализировать необходимые данные в привычном для врача режиме
4	При желании ознакомиться с результатами анализа ИИ-сервиса необходимо выбрать и просмотреть дополнительную(ые) серию(и) с изображениями исследования, обработанными ИИ-сервисом. Подпись к дополнительной серии соответствует названию ИИ-сервиса, создавшего данную серию. При первом знакомстве с результатами работы текущего ИИ-сервиса рекомендуем обратиться к краткому руководству, присутствующему в DICOM SR, или к учебным материалам по данному сервису
5	Посмотреть заключение и детализацию находок ИИ-сервиса по данному исследованию можно в разделе с текстовой информацией (DICOM SR). Важно выбрать нужную вкладку DICOM SR с наименованием сервиса, результаты от которого были оценены врачом на предыдущем шаге
6	Заполнить протокол исследования в привычном режиме
7	В случае обращения к результатам анализа ИИ-сервиса рекомендуем оставить обратную связь по данному исследованию. Для этого необходимо выбрать сервис, по которому оценивалось заключение, выбрать «Согласен» или «Не согласен», в случае несогласия указать причину из выпадающего списка или оставить текстовый комментарий

Анкета обратной связи для оценки качества обучения размещена в интернете¹. На момент подготовки книги получены 378 отзывов от 339 уникальных пользователей-врачей. Среди прошедших обучение врачей-рентгенологов 43,9% работали в стационарных медицинских организациях, 56,1% – соответственно в амбулаторно-поликлинических.

Более половины опрошенных были в возрасте 30–50 лет, 26,0% относились к группе старше 50 лет, лишь 18,0% являлись относительно молодыми специалистами. В плане стажа работы раз-

¹ URL: <https://forms.yandex.ru/u/5efb444e968c8c24d17bf7ae/>

деление было практически равным (рис.18). Результаты опроса представлены в таблице 22.

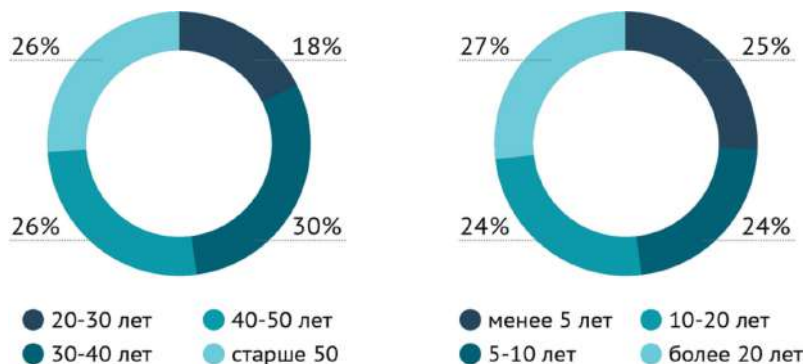


Рисунок 18 – Характеристика возраста и стажа работы врачей-рентгенологов, прошедших обучение

В 8 и более баллов работу лектора оценили 73,8% опрошенных, качество материала – 76,7%, доступность информации – 65,6%, практическую значимость – 58,7%. Оценки в пять и менее баллов поставили 10,9%, 11,1%, 11,1% и 23,6% врачей соответственно.

В целом большинство врачей положительно оценили качество обучения. Обращает на себя внимание более низкая оценка практической значимости представленных материалов, что вполне характерно для второго этапа схемы развития компетенций. В данный период пользователи имеют минимальный или нулевой собственный опыт применения технологии создает некоторый «разрыв» между теорией и практикой, но минимизируется в дальнейшем.

На основании накопленных материалов научного, учебно-методического и практического характера в конце 2020 года была сформирована программа дополнительного профессионального образования «Технологии искусственного интеллекта в лучевой диагностике», рассчитанная на 18 академических часов.

Таблица 22 – Результаты оценки врачами-рентгенологами качества обучения теоретическим и практическим вопросам применения технологий искусственного интеллекта

Вопрос / Показатели	Ваша оценка работы лектора по содержанию материала (развернутость, обоснованность, ответы на вопросы) (0 баллов – плохо, 10 баллов – прекрасно)	Ваша оценка работы лектора по оформлению презентации, доступности материала (0 баллов – плохо, 10 баллов – прекрасно)	Оцените полноту и доступность предоставляемой информации (0 баллов – осталось множество вопросов, я не понимаю, как буду работать с ИИ-сервисами, 10 баллов – вопросов нет, все ясно)	Как Вы оцениваете практическую значимость обучения (0 – плохо, 10 баллов – прекрасно)
Минимальное значение	0	0	0	0
Максимальное значение	10	10	10	10
Медиана	9	9	8	8
Мода	10	10	10	10
Среднее значение	8,3+2,1	8,4+2,1	7,8+2,4	7,5+2,7

§4. Специальные методологии и процедуры эксперимента

В соответствии с нормативно-правовым обеспечением Эксперимента и его стратегией существовала необходимость разработки ряда оригинальных, специфических методологий и процедур, предназначенных для решения как операционных, так и научных задач.

К таковым относились:

1. Стандартизированные функциональные и диагностические требования к ИИ-сервисам.

2. Процедуры тестирования ИИ-сервисов на этапах интеграции в ЕРИС ЕМИАС для определения качества, точности и воз-

возможности работы в условиях радиологической информационной системы:

- функциональное тестирование;
- калибровочное тестирование;
- контрольно-техническое тестирование.

3. Процедура технологического мониторинга работы ИИ-сервисов в процессе работы в условиях реального потока результатов лучевых исследований в ЕРИС ЕМИАС.

Функциональное и контрольно-техническое тестирования решали операционные задачи Эксперимента, а калибровочное и технологический мониторинг — операционные и научные. Вместе с тем все разработанные методики имеют методическую значимость при внедрении технологий искусственного интеллекта в практическое здравоохранение.

4.1. Стандартизированные требования к ИИ-сервисам

Базовые функциональные требования. С целью стандартизации представления результатов работы ИИ-Сервисов был сформирован перечень базовых функциональных требований к ним:

1. Приоритизация исследований в рабочем списке врача (триаж).
2. Буквенная маркировка исследований, обработанных ИИ-сервисом (AS, ASMT, BI). Обозначение исследований, обработанных ИИ, например, информацией о наличии дополнительных серий (AS, ASMT, BI).
3. Вероятность патологии в исследовании в целом (0–100%), в случае маммографии — по каждой железе.
4. Оригинальная серия не изменяется.
5. Дополнительная серия (должна всегда присутствовать независимо от результатов анализа):
 - а. Название дополнительной серии соответствует названию ИИ-сервиса.
 - б. В обязательном порядке должно присутствовать предупреждение в виде текста «Только для исследовательских целей», вшитого в неотключаемое изображение (не overlay). Также в дополнительной серии должны быть отражены название ИИ-сервиса, его версия, дата и время обработки исследования.
 - в. В случае отсутствия патологических изменений в дополнительной серии необходимо указать «Целевая патология не выявлена».

d. При наличии патологии, настройки яркости и контрастности изображения (окно) должны соответствовать принятым нормам в рентгенологии (например, результаты изменений в легких должны отражаться в легочном окне).

e. При наличии патологии для исследований КТ и НДКТ дополнительная серия должна содержать количество изображений, аналогичное количеству оригинальной серии. Также необходимо обеспечить функционирование режима синхронизации серий. Серии с патологическими находками должны быть промаркированы на ScrollBars (полосе прокрутки изображений в серии).

f. Патологические находки должны быть локализованы (обозначены), оптимальным решением маркировки патологических образований является оконтуривание. Например, для ММГ допустима только контрастная маркировка, хорошо визуализируемая на монохромных мониторах и различная для разных типов находок. В случае оценки соотношения объемных показателей (% поражения легочной ткани) необходима четкая визуализация как пораженной ткани, так и здоровой. Для этих целей оптимальный способ визуализации — цветовая карта.

g. В случае определения находок разного типа необходимо обеспечить цифровую идентификацию каждого типа находки. Перечень цифровых кодов должен быть отражен в кратком руководстве пользователя.

6. Текстовое описание (DICOM SR).

a. Структура должна быть следующая: название ИИ-сервиса/заключение/детализация находок/краткое руководство пользователя.

b. Детализация находок должна содержать изображение находки или срез с находкой, при этом следует определить класс патологической находки и, в случае установления очага, указать размер. При оценке признаков COVID необходимости в детализации находок нет, однако допустимо в сводной информации о пораженной ткани (цифровой или графической) указать % изменений легочной ткани каждого легкого.

c. Заключение должно содержать:

- для КТ РЛ — вероятность ЗНО в данном исследовании, вероятность иных патологий, определяемых ИИ-сервисом, за исключением изменений, характерных для COVID;

- для КТ ОГК COVID — процент вовлечения легочной ткани

в патологический процесс для каждого легкого, степень выявленных изменений в виде КТ 0–4;

- для РГ/ФЛГ — выявленные патологические процессы с указанием их вероятности;

- для ММГ — оценка по шкале BI-RADS 0–2, где к категории 0 относятся признаки, характерные для BI-RADS 3–5 с указанием вероятности злокачественности.

Базовые диагностические требования. Базовые диагностические требования к ИИ-сервисам формируются в зависимости от решаемой клинической задачи с учетом аналитики технической возможности реализации. В настоящее время большинство представленных на рынке решений для автоматизации диагностического процесса основаны исключительно на анализе изображений и не предназначены для постановки диагноза. Наименование клинической задачи должно быть привязано к шифру заболевания по МКБ-10. Для формулирования диагностических требований к ИИ-сервисам необходимо проведение предварительной аналитики возможностей автоматизации диагностического процесса по данным лучевых методов исследования.

ИИ-сервисы в обязательном порядке проходят проверку соответствия диагностическим требованиям в несколько этапов. Подготовительный этап состоит из ретроспективного исследования на калибровочном наборе изображений с верифицированными целевыми находками. Основной этап осуществляется проспективно в продуктивном программно-аппаратном контуре.

При анализе исследований ИИ-сервисам выдвигается требования обнаружения хотя бы одной находки из перечня целевых, указанных для каждой модальности и клинической задачи. Требования к содержанию ответа ИИ-сервиса в обязательном порядке включают указание процента вероятности наличия патологических изменений в исследовании и локализацию выявленных находок.

Требования к формату представления ответа предполагают указание вероятности наличия патологических изменений во всем исследовании в виде целого или дробного числа; обязательное указание патологических находок предполагает использование тепловой карты, цветовой маски или контура на всем протяжении находки применительно к оригинальному изображению, обработанному ИИ-сервисом. В зависимости от целевой

группы пользователей и основных клинических задач, требования могут быть пересмотрены и адаптированы.

Требованием к форме ответа ИИ-сервиса является создание Message/Сообщения и DICOM-файла с последующей передачей через систему Apache Kafka.

Первоначально в Эксперимент включались ИИ-сервисы, осуществляющие автоматизированный анализ результатов трех видов лучевых исследований:

1. Компьютерная томография и низкодозная компьютерная томография (КТ/НДКТ) грудной клетки с целью выявления рака легких.

2. Маммография (ММГ) с целью выявления рака молочных желез.

3. Рентгенография (РГ) легких с целью определения патологии. Наравне с рентгенографией рассматривались и результаты профилактической флюорографии (ФЛГ).

Однако внешние обстоятельства внесли существенные изменения в первоначальные планы – 11 марта 2020 года Всемирная организация здравоохранения объявила вспышку нового типа коронавируса 2019-nCoV (COVID-19) пандемией. В связи с этим резко снизилось количество профилактических лучевых исследований, временно заморозились программы скрининга, а в Эксперимент были включены ИИ-сервисы для выявления признаков коронавирусной инфекции (COVID 19) на КТ/НДКТ и рентгенограммах¹. Подробные клинико-диагностические требования к результатам работы ИИ-сервисов представлены в таблицах 23–26.

¹ В соответствии с приказом Департамента здравоохранения города Москвы от 15.05.2020 №522 «О внесении изменений в приказ Департамента здравоохранения города Москвы от 19 февраля 2020 г. №142».

Таблица 23 – Диагностические требования к ИИ-сервисам, анализирующим результаты рентгенографии или флюорографии органов грудной клетки

Наименование	Клиническая задача, решаемая ИИ-сервисом	Подготовительный этап (ретроспективное исследование) – признаки исследований калибровочного набора данных	Основной этап (проспективное исследование) – признаки, для которых ожидается положительный и отрицательный результаты работы ИИ-сервиса	Содержание ответа ИИ-сервиса	Формат ответа ИИ-сервиса	Форма предоставления ответа ИИ-сервиса
Рентгенография или флюорография органов грудной клетки	Определение наличия и локализация рентгенографических признаков (минимум одного), коррелирующих с минимум одной приоритетной нозологией (из перечня ниже): 1. Туберкулез (A15–A16, A19). 2. Пневмония, гнойные и некротические состояния (J10–J18, J80–J86). 3. Гидроторакс (J94, R09.1). 4. Пневмоторакс (S27.0). 5. Ателектаз (J98.1). 6. Объемные образования (D38.1–D38.4, C34–C39). 7. Перелом ребра (ребер), грудины и грудного отдела позвоночника (S22)	Есть признаки патологии: А. Наличие хотя бы одного рентгенологического признака* из перечня: 1. Плевральный выпот. 2. Пневмоторакс (S27.0). 3. Ателектаз 4. Очаг затемнения. 5. Инфильтрация/консолидация. 6. Диссеминация (>20 очаговых изменений). 7. Полость с распадом. 8. Полость с уровнем жидкости. 9. Кальцинат/кальцинированная тень в легких. 10. Нарушение целостности кортикального слоя		Обязательно – вероятность наличия во всем исследовании хотя бы одного рентгенологического признака из перечня А	Дробное или целое число	Apache Kafka Message + DICOM
		Б. Положительные результаты верификации хотя бы одной из приоритетных нозологий				
				Нет признаков патологии: ни одного рентгенологического признака из перечня А		Опционально – наименование одного или нескольких рентгенологических признаков

Продолжение таблицы 23

Наименование	Клиническая задача, решаемая ИИ-сервисом	Подготовительный этап (ретроспективное исследование) – признаки исследований калибровочного набора данных	Основной этап (проспективное исследование) – признаки, для которых ожидается положительный и отрицательный результаты работы ИИ-сервиса	Содержание ответа ИИ-сервиса	Формат ответа ИИ-сервиса	Форма представления ответа ИИ-сервиса
Рентгенография или флюорография органов грудной клетки	Определение наличия и локализация рентгенографических признаков, коррелирующих с наличием коронавирусной инфекции (COVID-19)	Есть признаки патологии: А. Инфильтрация (затемнение) легочной ткани с обеих сторон и преимущественным поражением периферических и базальных отделов легких		Обязательно – триаж исследований в рабочем списке (отсутствует или присутствует признак А); ИЛИ Обязательно – вероятность наличия во всем исследовании признака А	Дробное или целое число	Apache Kafka Message + DICOM
		Б. Положительные результаты верификации вирусной пневмонии посредством КТ ОГК и коронавирусной инфекции (COVID-19) посредством ОТ-ПЦР				
		Нет признаков патологии: отсутствие рентгенологического признака А		Обязательно – локализация найденных признаков (если применимо)	Тепловая карта/ Контур/ Маска/ др.	DICOM

Таблица 24 – Характеристика рентгенологических признаков, включенных в диагностические требования к ИИ-сервисам для анализа результатов рентгенографии или флюорографии органов грудной клетки

Наименование рентгенологического признака	Клинически значимые характеристики признаков
1. Плевральный выпот	Однородное затемнение в нижних отделах легких с практически горизонтальным уровнем, не определяется анатомическая структура наружно-нижних отделов легких – реберно-диафрагмальный синус
2. Пневмоторакс (S27.0)	Однородное просветление в периферических отделах верхних долей легких, соответствующее распределению воздуха в замкнутом пространстве, в зоне интереса не определяется (смещен) легочный рисунок
3. Ателектаз	Снижение объема легкого вследствие коллапса. Субсегментарный, сегментарный, долевого, тотальный. Однородное затемнение структурной единицы легкого со смещением анатомических структур в сторону поражения на фоне уменьшения объема
4. Очаг затемнения	Локальная зона затемнения легочной ткани с нарушением дифференцировки легочного рисунка, не соответствующая анатомическому перибронхиальному распределению, возможен значительный разброс локализации/размеров/контуров/формы
5. Инфильтрация/ консолидация	Локальная зона затемнения легочной ткани с полным/неполным нарушением дифференцировки легочного рисунка, соответствующая анатомическому перибронхиальному/ сегментарному/ долевого распределению
6. Диссеминация	Множественные однотипные субсантиметровые локальные зоны затемнения легочной ткани, соответствующие анатомическому перибронхиальному распределению
7. Полость с распадом	Локальная зона затемнения легочной ткани с нарушением дифференцировки легочного рисунка, центральным просветлением, не соответствующая анатомическому перибронхиальному распределению, возможен значительный разброс локализации/размеров/контуров/формы

Продолжение таблицы 24

Наименование рентгенологического признака	Клинически значимые характеристики признаков
8. Полость с уровнем жидкости	Локальная зона затемнения легочной ткани с нарушением дифференцировки легочного рисунка, горизонтальным уровнем раздела содержимого «газ-жидкость», не соответствующая анатомическому перибронхиальному распределению, возможен значительный разброс локализации/размеров/контуров/формы
9. Кальцинат/кальцинированная тень в легких	Локальное однородное высокоинтенсивное (высокоплотное) затемнение с четкими контурами
10. Нарушение целостности кортикального слоя	Локальное прерывание линии наружного контура кости, возможно с наличием смещения/расхождения костных отломков

Таблица 25 – Диагностические требования к ИИ-сервисам, анализирующим результаты компьютерной томографии или низкодозной компьютерной томографии легких

Наименование	Клиническая задача, решаемая ИИ-сервисом	Подготовительный этап (ретроспективное исследование) – признаки исследований калибровочного набора данных	Основной этап (проспективное исследование) – признаки, для которых ожидается положительный и отрицательный результаты работы ИИ-сервиса	Содержание ответа ИИ-сервиса	Формат ответа ИИ-сервиса	Форма предоставления ответа ИИ-сервиса
Компьютерная томография легких/ Низкодозная компьютерная томография легких для скрининга рака легкого	Выявление компьютерно-томографических признаков, коррелирующих с наличием рака легкого (C34)	Есть признаки патологии: А. Хотя бы один солидный или субсолидный узел объемом >100 куб. мм		Обязательно – вероятность наличия во всем исследовании признаков рака легкого (C34): перечень А	Дробное или целое число	Apache Kafka Message + DICOM
		Б. Результаты патоморфологического исследования: злокачественное новообразование		Обязательно – локализация найденных очагов (находок) Опционально – наличие скроллинга с указанием положения текущего среза и указанием положения срезов с патологическими находками	Тепловая карта/ Контур/ Маска/ др.	DICOM
		Нет признаков патологии: ни одного очага, подпадающего под указанные условия			-	DICOM
				Опционально – для каждого найденного очага вероятность отнесения к признаку рака легкого (C34)	Дробное или целое число	DICOM

Продолжение таблицы 25

Наименование	Клиническая задача, решаемая ИИ-сервисом	Подготовительный этап (ретроспективное исследование) – признаки исследований калибровочного набора данных	Основной этап (проспективное исследование) – признаки, для которых ожидается положительный и отрицательный результаты работы ИИ-сервиса	Содержание ответа ИИ-сервиса	Формат ответа ИИ-сервиса	Форма представления ответа ИИ-сервиса
Компьютерная томография легких	Выявление компьютерно-томографических признаков, коррелирующих с наличием коронавирусной инфекции (COVID-19)	Есть признаки патологии: А. 1. Инфильтрация легочной паренхимы по типу «матовых стекол» с обеих сторон, преимущественно периферической локализации с или без инфильтрации легочной паренхимы по типу консолидации с положительным признаком воздушной бронхограммы. 2. Инфильтрация легочной паренхимы по типу «булжной мостовой» (утолщение междолькового интерстиция на фоне «матового стекла») с обеих сторон, преимущественно периферической локализации, с или без инфильтрации легочной паренхимы по типу консолидации с положительным признаком воздушной бронхограммы	Обязательно – триаж исследований в рабочем списке (отсутствует или присутствует хотя бы один из признаков из перечня А); ИЛИ Обязательно – вероятность наличия во всем исследовании хотя бы одного признака из перечня А	Дробное или целое число	Apache Kafka Message + DICOM	

Продолжение таблицы 25

Наименование	Клиническая задача, решаемая ИИ-сервисом	Подготовительный этап (ретроспективное исследование) – признаки исследований калибровочного набора данных	Основной этап (проспективное исследование) – признаки, для которых ожидается положительный и отрицательный результаты работы ИИ-сервиса	Содержание ответа ИИ-сервиса	Формат ответа ИИ-сервиса	Форма представления ответа ИИ-сервиса
		Б. Положительные результаты верификации коронавирусной инфекции (COVID-19) посредством ОТ-ПЦР		Обязательно – локализация найденных очагов (находок)	Тепловая карта/ Контур/ Маска/ др.	DICOM
				Опционально – наличие скроллинга с указанием положения текущего среза и указанием положения срезов с патологическими находками	-	DICOM
				Опционально – классификация степени изменений (КТ 0-4) на основе значения относительного объема вовлечения паренхимы легкого	Текст	DICOM

Таблица 26 – Диагностические требования к ИИ-сервисам, анализирующим результаты профилактической маммографии

Наименование	Задача, решаемая ИИ-сервисом	Подготовительный этап (ретроспективное исследование) – признаки исследований калибровочного набора данных	Основной этап (проспективное исследование) – признаки, для которых ожидается положительный и отрицательный результаты работы ИИ-сервиса	Содержание ответа ИИ-сервиса	Формат ответа ИИ-сервиса	Форма представления ответа ИИ-сервиса
Диагностическая / скрининговая маммография молочной железы	Определение признаков наличия и локализации находок, характерных для рака молочной железы (C50)	Есть признаки патологии: A. BI-RADS 3–5 / 0*		Обязательно – вероятность наличия во всем исследовании признаков рака молочной железы (C50): перечень A	Дробное или целое число	Apache Kafka Message + DICOM
		Б. Результаты патоморфологического исследования: злокачественное новообразование		Обязательно – локализация находок, свидетельствующая о наличии во всем исследовании признаков рака молочной железы (C50): перечень A	Тепловая карта/ Контур/ Маска/ др.	DICOM
			Нет признаков патологии: B. BI-RADS 1–2	Опционально – для каждой находки наиболее вероятная категория из приведенного ниже перечня и вероятность отнесения находки к этой категории в процентах: • злокачественное новообразование; • доброкачественное образование; • подозрительные кальцинаты; • патологически измененные лимфоузлы; • утолщение кожи	Дробное или целое число/ Текст/ др.	DICOM

4.2. Функциональное тестирование ИИ-сервисов

Цель функционального тестирования (ФТ): проверка технологического качества интеграции ИИ-сервиса в Единый радиологический информационный сервис.

В ходе функционального тестирования осуществляется проверка наличия и работоспособности функций ИИ-сервиса, заявленных компанией-поставщиком и соответствие базовым функциональным требованиям (рис. 19).



Рисунок 19 – Процесс функционального тестирования

Функциональное тестирование проводится как с технической, так и медицинской точек зрения. Техническая часть отвечает за корректность отображения всех требуемых полей в АРМ врача-рентгенолога, включая корректность результатов работы ИИ-сервиса. Медицинская часть проводится врачами-экспертами с точки зрения оценки грубых нарушений в работе ИИ-сервиса (например, пропуск явных патологических находок).

По готовности ИИ-сервиса к проведению функционального тестирования на сервис маршрутизируются 5 исследований из специального набора данных для функционального тестирования, содержащих нормальные и патологические исследования. После обработки сервисом данного набора специалистами, проводящими ФТ, осуществляется оценка функционала Сервиса и информирование компании-поставщика ИИ-сервиса о резуль-

татах; в случае отсутствия значимых замечаний к работе ИИ-сервиса последний может быть допущен к проведению калибровочного тестирования.

Оценка функционала ИИ-сервиса включает в себя такие процедуры:

- заполнение шаблона протокола ФТ;
- проведение ФТ;
- анализ дефектов ИИ-сервиса;
- составление протокола ФТ;
- привлечение клинического эксперта;
- анализ дефектов ИИ-сервиса с клинической точки зрения.

В ходе проведения проверки наличия и работоспособности функций ИИ-сервиса заполняется протокол функционального тестирования (приложение 1). В указанном документе отражается информация об ИИ-сервисе, происходит проверка заявленных параметров компанией-поставщиком (рабочий список, заключение DICOM SR, дополнительная серия изображения от ИИ-сервиса) с занесением соответствующих результатов, затем происходит оценка основных критериев.

Проверка осуществляется по следующим параметрам:

а) Рабочий список:

- триаж — определение приоритета исследования в зависимости от результатов работы ИИ-сервиса и присвоение ему одной из трех категорий, обозначаемых различными цветами (красный, оранжевый, бесцветный);

- графическое обозначение — наличие отдельного столбца в рабочем списке с отметкой о завершении анализа исследования ИИ-сервисом. В поле Modality в общем рабочем списке будет уведомление через определенный тег (AS, ASMT, BI);

б) Заключение DICOM SR:

- DICOM SR 1 — назначение сервиса: модальность, анатомическая область, возрастная группа, показания к применению (перечень выявляемых патологий, требования к изображениям и т.д.);

- DICOM SR 2 — служебная информация: название сервиса, версия ПО, дата анализа и т. д.

- DICOM SR 3 — краткое руководство пользователя;

- DICOM SR 4 — заключение по выполненному анализу исследования;

- DICOM SR 5 – детализация выполненного анализа по находкам;

- шаблон протокола – наличие предзаполненного заключения (наличие предзаполненного описания – опционально). Источником заполнения может быть либо предустановленный универсальный шаблон, либо шаблон по форме конкретного ИИ-сервиса.

- в) Дополнительная серия изображений от ИИ-сервиса:

- наличие дополнительной серии от ИИ-сервиса – дополнительная серия с результатами работы ИИ-сервиса, содержащая маркировку и/или температурную карту, локализирующую находку. В случае, если визуализация находок на изображениях исследования не требуется или отсутствует, информация о результатах анализа должна быть предоставлена в текстовом виде в этой доп. серии;

- название дополнительной серии – однозначная идентификация серий, сгенерированных ИИ-сервисом;

- синхронизация серий – возможность параллельного просмотра исходных серий и серий, сгенерированных ИИ-сервисом;

- информация о сервисе – наличие на изображениях дополнительных серий служебной информации об ИИ-сервисе: название сервиса, версия ПО, дата анализа, предупреждения «Только в исследовательских целях»;

- вероятность (-и) находки (-ок) – наличие на изображениях дополнительных серий информации о значении вероятности наличия патологических изменений в целом для всего исследования и/или для каждой находки (в %);

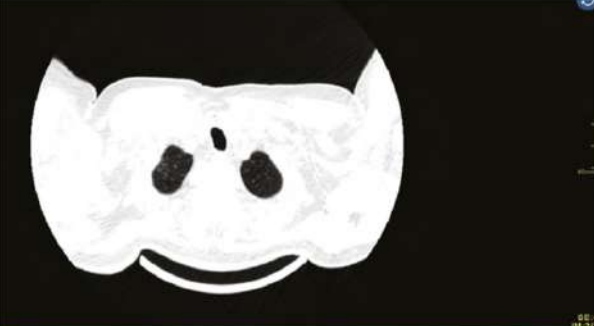
- категории находок – в случае, если ИИ-сервис определяет несколько категорий находок, каждая из них должна быть однозначно идентифицируема в дополнительной серии;

- отключение маркировки – возможность отключения отображения на дополнительной серии любой информации от ИИ-сервиса, в т.ч. маркировки находки (-ок).

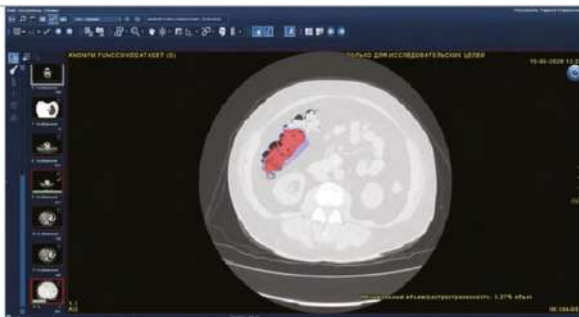
По результатам проверки и оценки формируются выводы и заключение.

Результатом функционального тестирования является протокол с оценкой полноты реализованного функционала ИИ-сервиса для его использования врачами-рентгенологами (приложение 1). Протокол может содержать значимые замечания и замечания рекомендательного характера.

Таблица 27– Виды замечаний при проведении функционального тестирования ИИ-сервиса

Критические замечания	Влияют на безопасность пациента и на работу врача. Примеры: отсутствие реализации заявленного функционала от компании-поставщика; замечания, оказывающие влияние на врача-рентгенолога или затрудняющие его деятельность; отклонения функционала ИИ-сервиса от базовых функциональных требований
Примеры	 Краткое руководство пользователя автоматическая система, выявляющая области на рентгенограммах грудной клетки, в которых предполагается наличие патологий. Области локализуются красными контурами на снимке. Вероятность наличия <i>Несоответствие информации в дополнительной серии и в DICOM SR</i>  Краткое руководство пользователя Очаги поражения отмечены красным контуром Заключение Результат обработки Испытания: КТ1 (Процент поражения легких = 0); "Подозрение на COVID-19" Степень выявленных <i>В качестве результата представлены только два изображения, по которым невозможно принять клиническое решение</i>

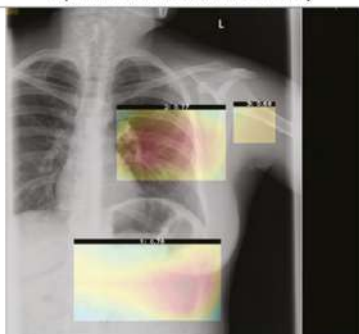
Продолжение таблицы 27



Находка вне области целевого органа

[illegible]

*Заклучения по одному и тому же исследованию,
противоположные по смыслу*



Присутствуют находки вне целевого органа

Продолжение таблицы 27

Значимые замечания	Не влияют на безопасность пациента, но влияют на работу врача. Недоработки функционала, которые отклоняются от общепринятых норм в деятельности врача-рентгенолога
Примеры	<i>Некорректное отображение подписей в дополнительной серии</i> <i>Некорректная приоритизация исследования (обнаружены патологические изменения, но приоритет исследования не повышен)</i>
Незначимые замечания	Не влияют на безопасность пациента, не влияют на работу врача. Несущественные недоработки, которые необходимо устранить для более удобной, интуитивно понятной и оперативной работы врача-рентгенолога
Примеры	<i>Излишнее количество информации в DICOM SR</i> Требования к изображениям: файлы DICOM рентгеновского снимка грудной клетки или переднего обзора <i>Неточные формулировки в DICOM SR</i>

Все **замечания к функционалу** можно условно разделить на три группы (таблица 27):

1. Влияют на безопасность пациента и на работу врача — отсутствие реализации заявленного функционала от компании-поставщика; замечания, оказывающие влияние на врача-рентгенолога или затрудняющие его деятельность; отклонения функционала ИИ-сервиса от базовых функциональных требований.

2. Не влияют на безопасность пациента, но влияют на работу врача — недоработки функционала, которые отклоняются от общепринятых норм в деятельности врача-рентгенолога.

3. Не влияют на безопасность пациента, не влияют на работу врача — несущественные недоработки, которые необходимо устранить для более удобной, интуитивно понятной и оперативной работы врача-рентгенолога.

Если реализованные функциональные элементы не соответствуют перечню базовых функциональных требований, то указываются рекомендации по доработке ИИ-сервиса. В случае, если реализованные функциональные элементы в полном объеме соответствуют перечню базовых функциональных требований, то ИИ-сервис допускается к проведению калибровочного тестирования.

4.3. Калибровочное тестирование ИИ-сервисов

Цель калибровочного тестирования: проверка диагностической точности ИИ-сервиса в условиях работы в Едином радиологическом информационном сервисе.

В процессе калибровочного тестирования фактически происходит проверка соответствия заявленных разработчиком и реальных значений метрик диагностической точности, достигаемых на данных из ЕРИС (рис. 20).

После успешного прохождения функционального тестирования происходит калибровка ИИ-сервиса с вычислением оптимального порога срабатывания (т.н. cut-off), то есть такой вероятности наличия целевой патологии, при которой, по мнению ИИ-сервиса, присутствует патология, а также расчет показателей точности работы ИИ-сервиса согласно методологии.



Рисунок 20– Процесс калибровочного тестирования (здесь КТ – калибровочное тестирование)

В ходе этого процесса используется специальный калибровочный набор данных, содержащий не менее 100 исследований требуемого вида. В этом наборе часть исследований соответствует категории «без патологии» и другая часть – категории «с патологией» в равных долях. Отнесение исследования к одной из категорий происходит в соответствии с клинической задачей ИИ-сервиса и базовыми требованиями. Исследования из калибровочного набора данных направляются ИИ-сервису, анализируются, результаты возвращаются. Результаты обработки тестовых исследований в виде таблицы с указанием временных интервалов, а также со значениями вероятности наличия целевой патологии в каждом исследовании направляются сотруднику, проводящим калибровку.

В отчете формата .csv (.xls) для каждого исследования указаны:

- StudyInstanceUID;
- дата и время отправки в сервис (downloadStartDT) -сервис;
- дата и время получения сервисом (downloadEndDT) -сервис;
- дата и время начала анализа сервисом (processStartDT) -сервис;
- дата и время окончания анализа сервисом (processEndDT) -сервис;
- дата и время отправки в сервис-ПУМ;
- дата и время получения результата от сервиса-ПУМ;

- вероятность наличия во всем исследовании признаков патологии;

- успешно обработано (1 — да/0 — нет);

- длительность обработки исследования ИИ-сервисом.

На основе полученных вероятностей целевой патологии проводится расчет параметров диагностической точности, пороговых значений и времени обработки исследований. Для автоматизации этой процедуры может применяться специальное программное обеспечение. В частности, авторами разработан программный модуль, написанный на Python 3 и R. Пример отчета, формируемого этим ПО, показан на рис. 21. На основании данного отчета заполняется протокол калибровочного тестирования (приложение 2).

В протоколе отражается информация об ИИ-сервисе и его тестировании, происходит проверка статистических параметров, полученных в ходе калибровки с занесением и сравнением с параметрами, предоставленными разработчиком; формируются выводы и рекомендации.

Важно подчеркнуть, что полученные значения оптимального порога срабатывания (cut-off) ИИ-сервиса вносятся в специальный справочник ЕРИС ЕМИАС. В протокол обязательно включаются значение площади под ROC-кривой (AUC), точность, чувствительность, специфичность, удельный вес ложноотрицательных результатов, удельный вес ложноположительных результатов, оптимальный порог, которые:

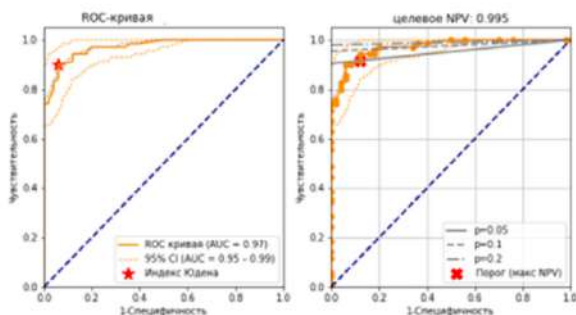
- заявлены компанией-поставщиком;

- получены на эталонном наборе данных (с использованием в расчетах индекса Юдена);

- получены на эталонном наборе данных (с использованием в расчетах максимального значения прогностической ценности отрицательного результата).

Отчет по калибровочному тестированию

Всего исследований: 120
 Отправлено и обработано: 120
 Отправлено и не обработано: 0
 Не отправлено: 0



Оптимальный порог (индекс Юдена): 0.22

Четырехпольная таблица:

ИП: 63 ЛО: 7 ЛП: 3 ИО: 47

Точность: 0.92 (0.85 – 0.96)

Чувствительность: 0.9 (0.8 – 0.96)

Специфичность: 0.94 (0.83 – 0.99)

Оптимальный порог (макс NPV): 0.19

претестовая вероятность: 0.05

Четырехпольная таблица:

ИП: 64 ЛО: 6 ЛП: 6 ИО: 44

Точность: 0.9 (0.83 – 0.95)

Чувствительность: 0.91 (0.82 – 0.97)

Специфичность: 0.88 (0.76 – 0.95)

Среднее время анализа (стандартное отклонение, медианное значение), мин: 2 (0, 2)

Минимальное время анализа, мин: 1

Максимальное время анализа, мин: 4

Среднее время ПУМ (стандартное отклонение, медианное значение), мин: 3 (1, 3)

Минимальное время ПУМ, мин: 1

Максимальное время ПУМ, мин: 6

Рисунок 21 – Пример автоматизированного отчета, формируемого по результатам калибровочного тестирования

Также указываются время обработки одного исследования и время, затрачиваемое на принятие исследования, его обработку

и передачу результатов анализа по данным ЕРИС ЕМИАС.

Для калибровочного тестирования нами приняты оптимальные значения параметров работы ИИ-сервиса, установленные экспертным и опытным путем (таблица 28). Под термином «Относительная разница» подразумевается разница между двумя значениями параметра: заявленным разработчиком, установленным при калибровке. Причем разница выражена в процентах.

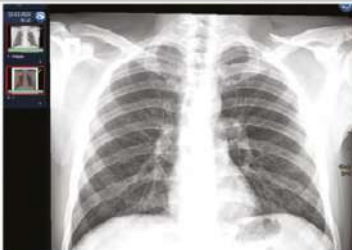
Таблица 28 – Оптимальные значения параметров работы ИИ-сервиса при калибровочном тестировании

Параметр	Значение
Площадь под ROC-кривой (AUC)	Более 0,81%
Относительная разница значения AUC	Не менее минус 10%
Относительная разница значения точности	Не менее минус 10%
Относительная разница значения чувствительности	Не менее минус 10%
Относительная разница значения специфичности	Не менее минус 10%
Время, затрачиваемое на принятие исследования, его обработку и передачу результатов анализа по данным ЕРИС ЕМИАС	Не более 10 минут
Удельный вес успешно обработанных исследований	Не менее 90%

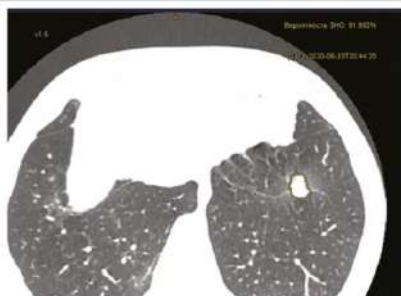
В ходе калибровочного тестирования также проводится дополнительная ручная проверка функционала ИИ-сервиса: оценка ложноотрицательных и ложноположительных срабатываний ИИ-сервиса, анализ корректности локализации патологических изменений в исследованиях с истинно положительным срабатыванием.

В таблицах 28 и 29 приведены примеры ложноотрицательных и ложноположительных срабатываний ИИ-сервиса по каждой модальности, примеры истинно положительных срабатываний.

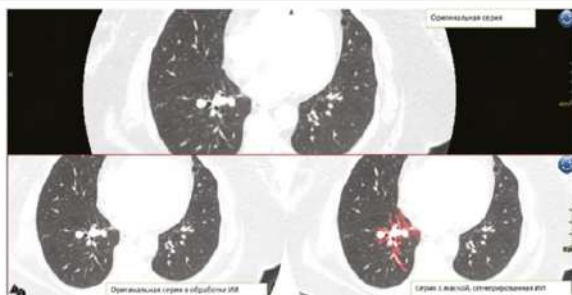
Таблица 28 – Примеры ложноотрицательных и ложноположительных срабатываний ИИ-сервисов (по результатам калибровочного тестирования)

Рентгенография органов грудной клетки	
	
<i>Пример ложноположительного срабатывания. Патологические изменения в легких и ребрах отсутствуют</i>	
	
<i>Пример ложноотрицательного срабатывания. ИИ-сервис не локализовал отчетливую округлую тень в средней или нижней доле левого легкого. Мелкий очаг с кальцинацией справа</i>	
Компьютерная томография органов грудной клетки	
	
<i>Пример ложноотрицательного срабатывания. ИИ-сервис локализовал маленькую находку, но пропустил большую</i>	

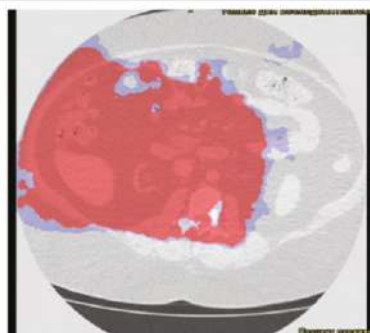
Продолжение таблицы 28



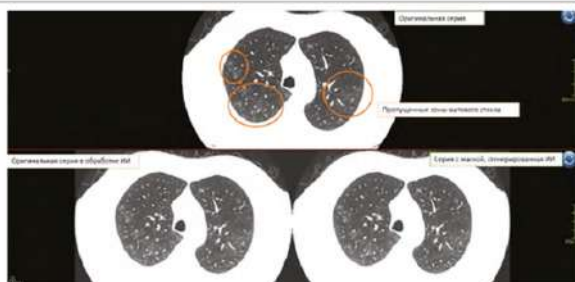
Пример ложноположительного срабатывания. ИИ-сервис локализовал купол диафрагмы



Пример ложноположительного срабатывания. ИИ-сервис (COVID-19) локализовал нормальные сосудистые структуры корня



Пример ложноположительного срабатывания. ИИ-сервис (COVID-19) локализовал находки в брюшной полости

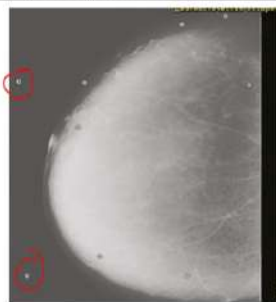


Пример ложноотрицательного срабатывания. ИИ-сервис (COVID-19) не отметил распространенные уплотнения по типу «матового стекла»

Маммография

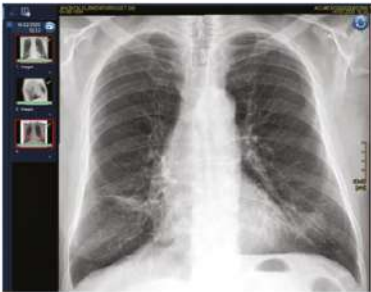
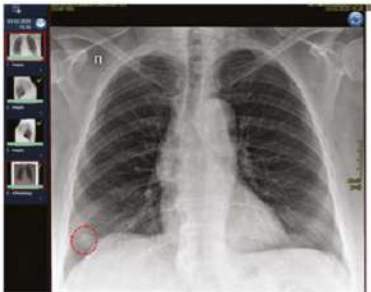


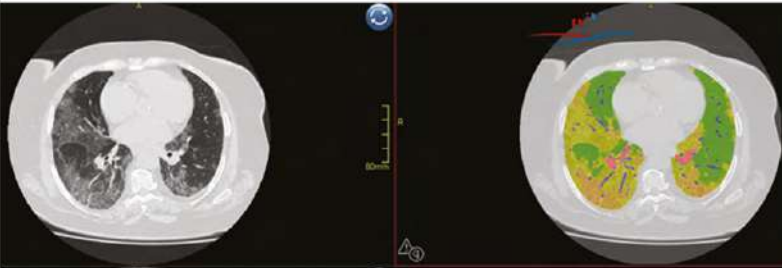
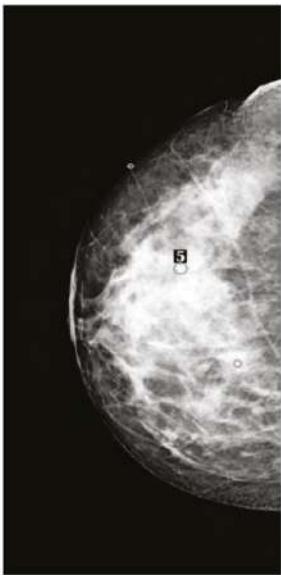
Пример ложноотрицательного срабатывания. ИИ-сервис не локализовал новообразование, подозрительное на злокачественное



Пример ложноположительного срабатывания. ИИ-сервис локализовал находки вне области интереса, а также за пределами изображения

Таблица 29 – Примеры истинно положительного срабатывания ИИ-сервисов (по результатам калибровочного тестирования)

Рентгенография органов грудной клетки	
	
Изменений в левом легком нет. В правом – ИИ-сервисом локализовано верно	
	
Непосредственно патологический очаг в правом легком локализован ИИ-сервисом частично верно, но обозначен менее ярко, чем нормальные ткани вокруг	
Компьютерная томография органов грудной клетки	
	
ИИ-сервис локализовал периферическое образование с тяжистыми контурами	


<i>ИИ-сервис локализовал крупные сливные участки уплотнения паренхимы типа «матового стекла» и консолидации, субплеврального и перибронхиального расположения</i>
Маммография

<i>ИИ-сервис верно локализовал новообразование, подозрительное на злокачественное</i>

Результатом калибровочного тестирования является протокол (приложение 2) с оценкой реализованного функционала ИИ-серви-

са на эталонном наборе данных.

Протокол может содержать:

- критические несоответствия (существенные отклонения от нормативных требований и заявленных параметров компаний-поставщиком);

- несоответствия рекомендательного характера (пограничные значения времени, затрачиваемого на принятие исследования, его обработку и передачу результатов анализа по данным ЕРИС ЕМИАС).

В выводах отражается информация о соответствии/несоответствии ИИ-сервиса технологическим требованиям прохождения стадии тестирования, описанной в нормативно-правовых документах¹.

Протокол калибровочного тестирования направляется компании-поставщику ИИ-сервиса для ознакомления. В случае соответствия ИИ-сервиса технологическим требованиям прохождения стадии тестирования последний получает возможность перейти в производственный контур ЕРИС ЕМИАС и начать проспективный анализ исследований.

4.4. Контрольно-техническое тестирование ИИ-сервисов

Рутинная эксплуатация ИИ-сервисов происходит в промышленном контуре ЕРИС ЕМИАС. После интеграции, при условии успешного прохождения функционального и калибровочного тестирования, ИИ-сервис начинает работу с потоком исследований от подключенных диагностических устройств в указанном контуре. При этом происходит настройка маршрутизации определенных видов исследования с назначенных диагностических устройств медицинских организаций. После начала работы проводится контрольно-техническое тестирование (КТТ).

Цель контрольно-технического тестирования – проверка корректности маршрутизации и обмена данными между каждым диагностическим устройством и ИИ-сервисом.

¹ На момент подготовки рукописи монографии согласно приказу ДЗМ от 13.10.2020 №1199 «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».

КТТ проводится путем проспективной (реже — ретроспективной) проверки двух исследований от каждого диагностического устройства, проанализированных ИИ-сервисом; имеются в виду устройства, от которых настроена маршрутизация исследований на данный сервис. КТТ проводится однократно или при смене версии для каждого ИИ-сервиса.

Проверка осуществляется по тем же параметрам, по которым проводится функциональное тестирование (см. выше); также осуществляется контроль среднего времени, затрачиваемого на обработку одного исследования и по условиям Эксперимента не превышающего 10 минут.

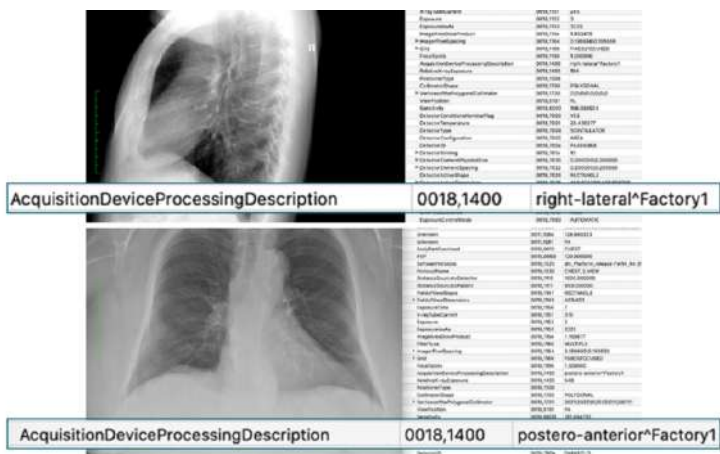
По итогам КТТ составляется отчет контрольно-технического тестирования (приложение 3). В указанном документе отражается информация об ИИ-сервисе и тестировании, происходит проверка заявленных параметров компанией-поставщиком с занесением соответствующих результатов. Формируются выводы и заключение, где предоставляются следующие списки диагностических устройств (ДУ):

1. Исследования с ДУ не содержат технологических дефектов и могут направляться на анализ ИИ-сервису в полном объеме.

2. Исследования с ДУ частично содержат технологические дефекты (в одном из двух исследований от одного ДУ содержится дефект либо содержатся дефекты рекомендательного характера), которые необходимо устранить. Исследования от данных ДУ могут направляться на анализ ИИ-сервису в полном объеме.

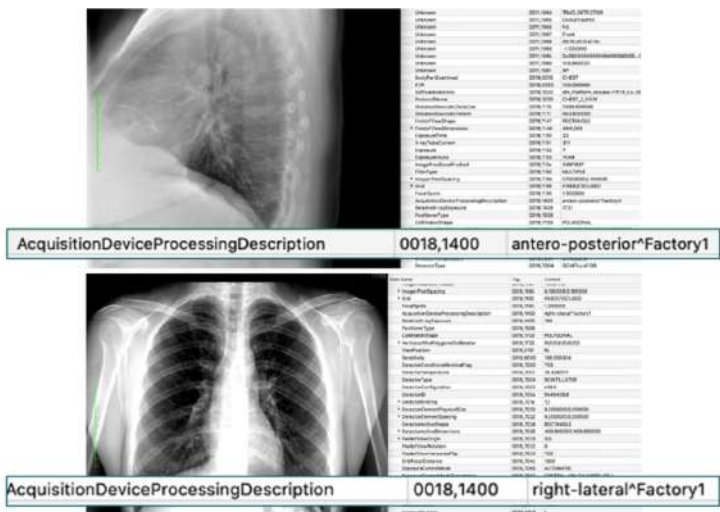
3. Исследования с ДУ содержат технологические дефекты. Исследования от данных ДУ не могут направляться на анализ ИИ-сервису до выяснения причин и устранения выявленных дефектов. В текущих настройках данные ДУ из маршрутизации необходимо исключить.

Считаем целесообразным привести сведения о типичных проблемах, выявляемых на данной стадии. Эта информация подчеркивает важность этапа КТТ — на фоне эйфории от успешной интеграции в промышленный контур радиологической информационной системы могут быть упущены критичные дефекты. Итак, в ходе проведения КТТ можно своевременно обнаружить дефект, связанный с неправильным заполнением тегов DICOM (рис. 22).



a

Рисунок 22 а – Примеры заполнения тегов DICOM: пример правильного тега



б

Рисунок 22 б – Примеры заполнения тегов DICOM: пример неправильного тега

Эти ошибки имеют критическое значение при работе отдельных сервисов. Так, при обработке результатов рентгенографии многие ИИ-сервисы определяют проекцию снимка исключительно по DICOM-тегам. С учетом систематического обнаружения некорректно заполненных соответствующих тегов нами был составлен алгоритм определения проекции грудной клетки (таблица 30). Фактически это последовательность просмотра определенных тегов с выполнением различных условий «если..., то...».

Таблица 30 – Алгоритм определения проекции рентгенограммы органов грудной клетки

1. Смотрим DICOM-тер (0018, 1400) Acquisition Device Processing Descr	
	а) Проекция прямая, если значение равно значению из списка: – ГРУДНАЯ КЛЕТКА, Пр стоя – ГРУДНАЯ КЛЕТКА, ПЕРЕДНЯЯ Р->А – postero-anterior^Factory1
	б) Проекция боковая, если значение равно значению из списка: – ГРУДНАЯ КЛЕТКА, БОК стоя – ГРУДНАЯ КЛЕТКА, БОКОВАЯ – right-lateral^Factory1
2. Если нет DICOM-тера из пункта 1 или нет значения из списка, то смотрим DICOM-тер (0018, 5101) View Position	
	а. Проекция прямая, если значение: – AP
3. Если нет DICOM-тера из пунктов 1 и 2 или нет значения из списков, то смотрим DICOM-тер (0054, 0220) View Code Sequence вкладку (0008, 0104) Code Meaning	
	а. Проекция прямая, если значение: – postero-anterior б. Проекция боковая, если значение: – lateral
4. Если нет этих тегов, то проекцию определить невозможно	

При работе с результатами компьютерной томографии органов грудной клетки также были выявлены технические сложности при выборе подходящей серии КТ (таблица 31).

Таблица 31 – Выбор оптимальной толщины среза в зависимости от производителя КТ-сканера

Название производителя (варианты в теге 0008,0070)	TOSHIBA	SIEMENS	PHILIPS	GE	Другие
Выбор наименьшей толщины среза (варианты в теге 0018,0050, от наиболее приоритетного к наименее приоритетному)	1,5 1,0 0,5 прочие	1,5 1,0 0,6 прочие	1,5 1,0 0,6 прочие	1,5 1,25 1,0 0,6 прочие	в порядке убывания приоритета 1,5 1,25 1,0 0,6 0,5 прочие
Выбор фильтра (варианты в теге 0018,1210, от наиболее приоритетного к наименее приоритетному)	FC06 FC07 FC08 FC09 FC10 FC11 FC12 FC13 FC14 FC15 FC16 FC17 FC18 FC19 FC50 FC51 FC52 FC53 FC54 FC55	I70 B70 B75 B80	SOFT STANDARD LUNG Y-Detail DETAIL BONE	SOFT STANDARD LUNG BONE	в порядке убывания приоритета SOFT STANDARD LUNG BONE прочие

Специально был разработан алгоритм по выбору серии исследования компьютерной томографии органов грудной клетки для отправки в ИИ-сервис при подозрении на COVID-19, который представлен ниже:

1. Определение производителя томографа (DICOM tag 0008,0070).
2. Определение минимальной толщины среза (DICOM tag

0018,0050).

3. Определение фильтра (DICOM tag 0018,1210).

Сформированный отчет о КТТ направляется компании-поставщику ИИ-сервиса для возможных последующих доработок ИИ-сервиса.

Главный результат контрольно-технических испытаний — это своевременное внесение корректирующих действий в маршрутизацию в случае некорректной или конфликтной работы ИИ-сервиса с отдельными диагностическими устройствами.

На основе полученных данных запускается маршрутизация исследований, выполненных на ДУ в группах «Без технологических дефектов» и «Частично содержащие технологические дефекты». Контроль корректности работы ИИ-сервиса с ДУ из последней группы осуществляется в ходе дальнейших регулярных мониторингов работы ИИ-сервиса.

4.5. Мониторинг работы ИИ-сервисов

Цель регулярного мониторинга работы ИИ-сервисов в промышленном контуре радиологической информационной системы — контроль качества для своевременного выявления и устранения технологических дефектов, а также оценки стабильности работы сервиса.

Суть мониторинга состоит в проверке результатов работы ИИ-сервиса на наличие технологических дефектов, которые разделены на несколько категорий (таблица 32, рис. 23–26). Проверка осуществляется в ручном и автоматическом режимах путем пересмотра случайно сформированной выборки исследований (минимум $n=20$, обычно $n=50$), проведенных за период времени. Регулярность проверок для данного ИИ-сервиса составляет 1 раз в месяц.

Для автоматизации формирования выборки, выявления исследований с дефектами (в том числе типов А, Б), устранения дублирования идентификационных номеров разработаны оригинальный алгоритм и программное обеспечение на его основе (на языке Python). Посредством этой же разработки автоматизированы сбор и анализ результатов мониторинга, формирование отчетов (рис. 27).

Таблица 32 – Виды технологических дефектов при работе ИИ-сервисов

Вид дефекта	Тип	Описание
Превышение установленной длительности анализа исследования	А	Среднее время, затрачиваемое на анализ (включая принятие, его обработку и передачу результатов обработки в ЕРИС ЕМИАС) исследования, превышает установленные нормы на 50%. Установленной нормой считается 6,5 минут*
Отсутствие результатов работы ИИ-сервиса в исследовании	Б	В исследовании, находящемся в ЕРИС ЕМИАС и ранее маршрутизированном на данный ИИ-сервис, отсутствуют результаты анализа/работы ИИ-сервиса
Изображения, содержащиеся в результатах работы сервиса, не соответствуют изображениям нативного исследования	В	Изображения в результатах работы ИИ-сервиса искажены, присутствуют не полностью, содержат данные другого пациента и т.д. В1 – изображения обрезаны; В2 – изменена яркость/контрастность; В3 – проанализированы не все необходимые изображения
Некорректная работа ИИ-сервиса, затрудняющая работу врача-рентгенолога или делающая ее выполнение невозможным с надлежащим качеством	Г	Г1 – полное отсутствие результатов работы ИИ-сервиса; Г2 – отсутствие дополнительной серии; Г3 – отсутствие DICOM SR; Г4 – наличие 2-х и более DICOM SR; Г5 – отсутствие названия ИИ-сервиса; Г6 – отсутствие сведений о версии ИИ-сервиса; Г7 – отсутствие предупреждающей надписи: «Только для использования в исследовательских/научных целях»; Г8 – отсутствие разметки при наличии патологии
Иные нарушения целостности и содержимого файлов с результатами исследований, обуславливающих невозможность их диагностической интерпретации	Д	Д1 – противоречие информации DICOM SR и в дополнительной серии; Д2 – разметка за пределами целевого органа; Д3 – проанализирована некорректная анатомическая область, проекция или серия
Изменение оригинальной серии исследования	Е	ИИ-сервис внес изменения, дополнения и проч. в оригинальные серии изображений исследования
* На момент подготовки рукописи монографии согласно приказу ДЗМ от 13.10.2020 № 1199 «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».		

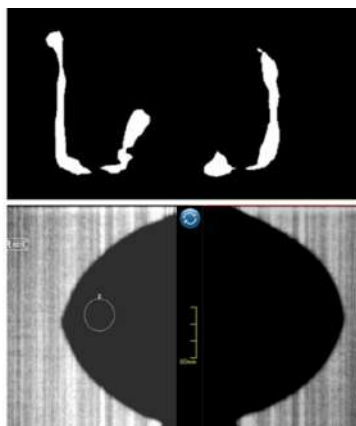


Рисунок 23 – Пример дефектов типа «В». Изображения, содержащиеся в результатах работы ИИ-сервиса, не соответствуют изображениям оригинального исследования



Рисунок 24 – Пример дефектов типа «Г». Отсутствие результатов работы ИИ-сервиса в направленном ранее на анализ исследовании: нет дополнительной серии, DICOM SR, предупреждающей надписи, разметки, названия сервиса

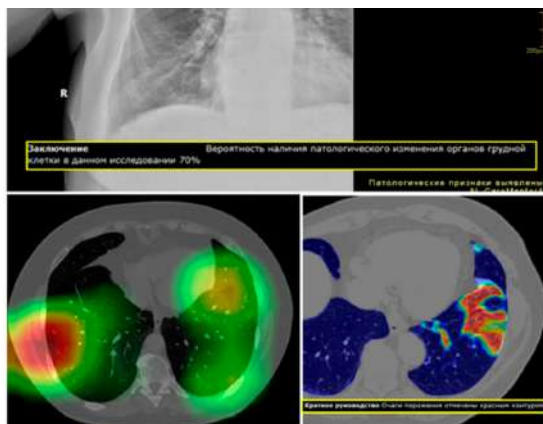


Рисунок 25 – Пример дефектов типа «Д». Разметка за пределами целевого органа; выводимая информация противоречит руководству пользователя сервиса: отсутствие разметки при указании на патологию в DICOM SR, противоречие данных в дополнительной серии и в DICOM SR



Рисунок 26 – Пример дефектов типа «Е». В результате работы ИИ-сервиса изменена оригинальная серия изображений. Некорректная работа сервиса, оказывающая непосредственное влияние на безопасность пациента, затрудняющая работу врача-рентгенолога или делающая ее выполнение невозможным

ОТЧЕТ ПО МОНИТОРИНГУ ТЕХНОЛОГИЧЕСКИХ ПАРАМЕТРОВ РАБОТЫ СЕРВИСА

1. Идентификатор Сервиса в Эксперименте	
2. Отчетный период	09.11.2020 - 15.11.2020
3. Вид исследований	ММГ
4. Клиническая цель:	ЗНО
Дата оформления отчёта:	2020-11-24
5. Общее количество исследований:	
5.1 Согласно отчёту от компании	
«НПКЦ_2 вкладки_Исследования_ММГ_09.11.2020_15.11.2020.xlsx», шт.*	611
5.2 Направленных на анализ Сервису за отчетный период согласно выгрузке из ЕРИС ЕМИАС, шт.	613
5.3 Согласно отчёту от компании, уникальных и подтвержденных выгрузкой из ЕРИС ЕМИАС, шт.	611
6. Количество исследований, прошедших контроль, шт.	611
6.1 Прошедших ручной контроль, шт.	20
7. Количество исследований, содержащих дефекты:	
7.1 Содержащих технологический дефект «а», шт., Приложение 1	3
7.2 Содержащих технологический дефект «б», шт., Приложение 2	140
7.3 Содержащих технологический дефект «в»-«е», шт., Приложение 3	0
8. Удельный вес исследований, содержащих тех. дефек. «а» относительно 471 исследований, %	1
9. Удельный вес исследований, содержащих тех. дефек. «б» относительно 611 исследований, %	23
10. Удельный вес исследований, содержащих тех. дефек. «в»-«е» относительно 20 исследований, %	0
11. Количество исследований без дефектов, шт.	468
12. Решение:	
участнику Эксперимента предлагается внести изменения в работу его Сервиса	
13. Примечания: Уд. вес дефек. «б» > 10 %.	
14. ФИО ответственного лица: отчёт сформирован автоматически	

* В отчете от компании:

Рисунок 27 — Пример отчета по результатам периодического технологического мониторинга, сформированного в автоматическом режиме

В условиях научного Эксперимента установлено, что доля исследований с технологическими дефектами не должна превышать 10% от всех исследований, обработанных ИИ-сервисом за отчетный период. Превышение заданного уровня может привести к прекращению участия сервиса в Эксперименте или послужить причиной приостановки работы до устранения дефектов.

В перспективе методика организации и проведения регулярного мониторинга технологических дефектов может стать основой для пострегистрационного контроля программного обеспечения, разработанного на основе технологий искусственного интеллекта и зарегистрированного в качестве медицинского изделия.

§5. Подготовка наборов данных в рамках Эксперимента

Разработанная ранее методология (см. главу 4) позволила нам подготовить для задач Эксперимента целый ряд наборов данных, в том числе:

1. Набор для самостоятельного тестирования разработчиками ИИ-сервисов своих программных продуктов. Содержит примеры результатов лучевых исследований рентгенографии органов грудной клетки, флюорографии, компьютерной томографии ОГК, низкодозной компьютерной томографии ОГК и маммографии, выполненных на различных моделях диагностического оборудования, установленного в отделениях лучевой диагностики МО ДЗМ и подключенного к ЕРИС ЕМИАС (таблица 33). Всего в наборе результаты 41 лучевого исследования по 5 модальностям. Набор размещен в свободном доступе на официальном сайте Эксперимента (www.mosmed.ai).

2. Наборы данных для проведения процедуры функционального тестирования: 54 исследования в 12 наборах (таблица 34). Предназначены для проверки соответствия ИИ-сервиса базовым функциональным требованиям и оценки качества работы заявленного функционала сервиса.

3. Наборы данных для проведения процедуры калибровочного тестирования: 1425 исследований в 13 наборах (таблица 35). Предназначены для определения порога и метрик качества работы ИИ-сервиса.

Таблица 33 — Примеры моделей диагностических устройств, результаты работы которых включены в эталонный набор данных

Модальность	Бренд	Модель
Компьютерная томография	Siemens	Somatom Emotion 16
		Somatom Sensation 40
	Toshiba Medical Systems Corporation	Aquilion 64
		Alexion TSX-032A
		Aquilion 32
	GE Medical Systems, LLC	Optima CT 660
		Light Speed VCT
Рентгенография/ Флюорография	Электрон	ФЦ-01 «Электрон»
	ЗАО «РЕНТГЕНПРОМ»	Про граф РП 4000
		ПроСкан-2000
	С.П. ГЕЛПИК	РенексФлюоро
		ФЦМ Барс-«Ренекс»
		ФМЦ
		Ренекс ФЛЦ
Маммография	Planmed	Clarity
	Fujifilm	Amulet
	ЗАО «Медицинские технологии Лтд» (МТЛ)	Маммо-4МТ
		Маммо-3МТ

Таблица 34 – Эталонные наборы данных, разработанные для функционального тестирования

№	Модальность	Диагностическое направление	Количество исследований в наборе данных, абс.
1	КТ	Поиск COVID-19	5
2	КТ*	Поиск COVID-19	5
3	КТ РЛ	Поиск ЗНО легкого	5
4	НДКТ РЛ	Поиск ЗНО легкого	5
5	ММГ	Поиск ЗНО молочной железы	5
6	ММГ*	Поиск ЗНО молочной железы	4
7	РГ	Поиск COVID-19	4
8	РГ*	Поиск COVID-19	4
9	РГ	Поиск патологий по списку	5
10	РГ*	Поиск патологий по списку	5
11	ФЛГ	Поиск патологий по списку	4
12	ФЛГ*	Поиск патологий по списку	4

* Отдельные оригинальные наборы данных для повторного проведения функционального тестирования.

Таблица 35 – Эталонные наборы данных, разработанные для калибровочного тестирования

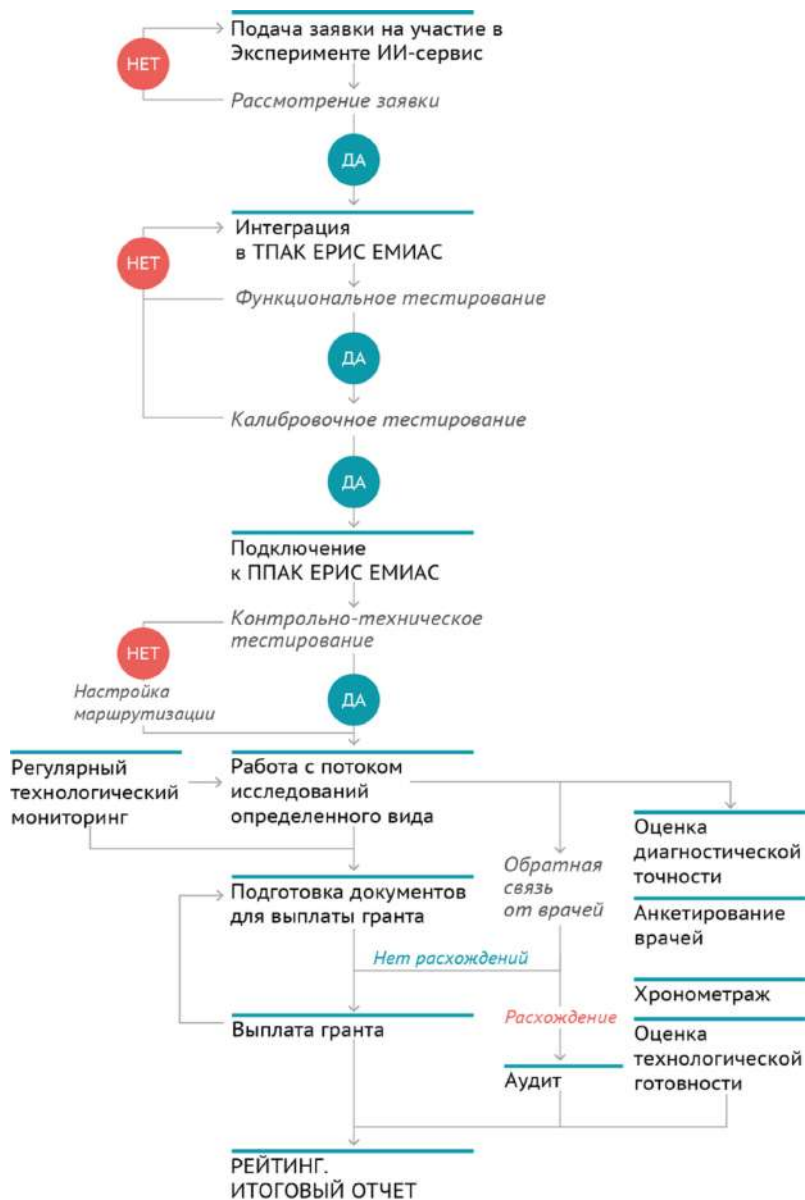
№	Модальность	Диагностическое направление	Количество исследований в наборе данных, абс.
1	КТ	Поиск COVID-19	100
2	КТ *	Поиск COVID-19	100
3	КТ	Поиск COVID-19 с классификацией КТ0-4	125
4	КТ	COVID-19 с классификацией КТ0-4	200
5	КТ РЛ	Поиск ЗНО легкого	100
6	НДКТ РЛ	Поиск ЗНО легкого	100
7	ММГ	Поиск ЗНО молочной железы	100
8	ММГ*	Поиск ЗНО молочной железы	100
9	РГ	Поиск COVID-19	100
10	РГ	Поиск патологий по списку	100
11	РГ*	Поиск патологий по списку	100
12	ФЛГ	Поиск патологий по списку	100
13	ФЛГ*	Поиск патологий по списку	100

* Отдельные оригинальные наборы данных для повторного проведения калибровочного тестирования.

Таким образом, разработанная нами методология подготовки эталонных наборов данных ранее прошла успешную апробацию в отдельных научно-практических исследованиях, затем была валидирована в международных публикациях независимыми авторами. В условиях Эксперимента методология была также успешно внедрена для массовой, фактически конвейерной подготовки эталонных наборов данных.

§6. Общий ход Эксперимента

Мультицентровое проспективное научное исследование «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы» включало ряд организационных, методических, образовательных, обеспечивающих, а также непосредственно научных мероприятий. Общий ход Эксперимента представлен на рисунках 28, 29.



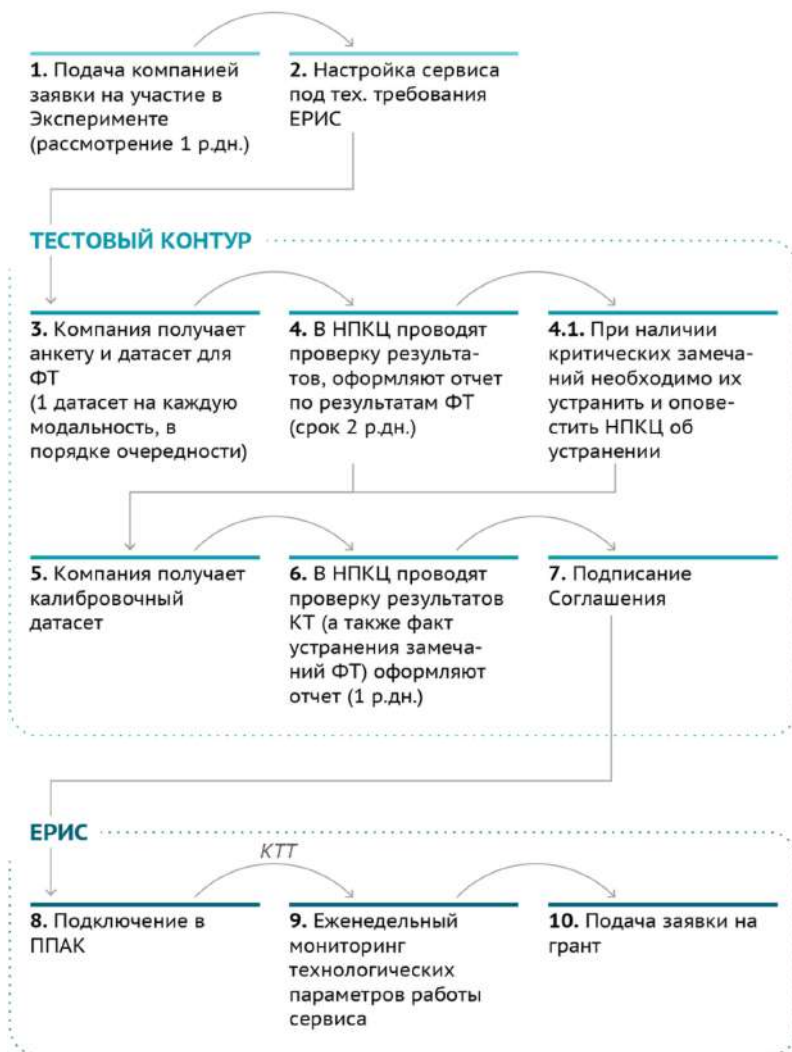


Рисунок 29 – Схема основных организационных этапов Эксперимента

Как было сказано выше, участие в Эксперименте осуществляется на основании заявки юридического лица, разработавшего или

имеющего права на предоставление ИИ-Сервиса. Претендент на участие подает заявку в Департамент здравоохранения города Москвы, приложив к ней нужные документы. В течение 1-го рабочего дня после получения заявки в Департаменте здравоохранения города Москвы она передается в ГБУЗ «НПКЦ ДиТ ДЗМ», который проверяет ее на соответствие нормативным требованиям. В случае, если заявка не отвечает требованиям, она возвращается по указанному в ней адресу электронной почты. При наличии полного комплекта документов ГБУЗ «НПКЦ ДиТ ДЗМ» направляет ИИ-сервис на процедуру технической интеграции с ЕРИС ЕМИАС посредством адресации соответствующего письма в Департамент информационных технологий города Москвы.

Техническая интеграция сервиса выполняется в тестовом программно-аппаратном комплексе ЕРИС ЕМИАС (далее – ТПАК ЕРИС ЕМИАС), срок ее проведения – не более 7 рабочих дней с момента поступления заявки претендента на участие в Эксперименте в ГБУЗ «НПКЦ ДиТ ДЗМ». Срок интеграции может быть увеличен по запросу претендента на участие в Эксперименте.

Техническая интеграция считается осуществленной при условии прохождения каждой стадии тестирования (таблица 36):

- 1) функционального,
- 2) калибровочного,
- 3) контрольно-технического.

Таблица 36 – Основные характеристики тестирований на этапах технологической интеграции

Тестирование	Цель	Критерии прохождения
Функциональное	Оценить корректность и полноту реализации заявленного функционала ИИ-сервиса, соответствие базовым требованиям	Отсутствие значимых замечаний к работе ИИ-сервиса
Калибровочное	Оценить метрики качества работы ИИ-сервиса на эталонных данных; определить порог срабатывания ИИ-сервиса	Метрики качества работы (AUC, чувствительность, специфичность, точность, время обработки) не отличаются более чем на 10% от заявленных в худшую сторону; AUC > 0,81
Контрольно-техническое (при необходимости)	Оценка корректности обработки ИИ-сервисом данных с каждого подключенного диагностического устройства	Данные с диагностического устройства обрабатываются без технологических дефектов или содержат их частично

Функциональное тестирование (методология проведения приведена в главе 6). Всего проведено 40 процедур (33 первичных и 7 повторных) для 85,0% ИИ-сервисов, заявки которых были приняты к участию. Скорость устранения замечаний в среднем составила 26 дней (медианное значение –18 [3 – 118]) (рисунок 30).

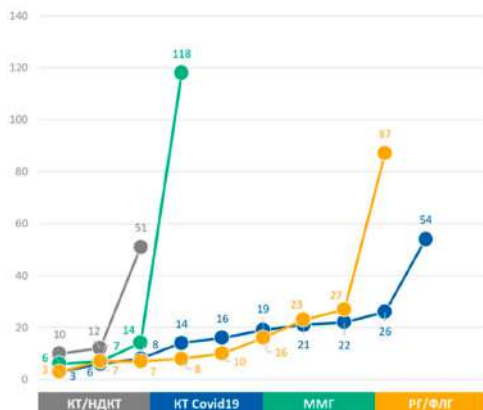


Рисунок 30 – Динамика устранения замечаний по итогам функционального тестирования

В ходе функционального тестирования среднее количество замечаний значимого характера составило 2 (максимальное – 6, минимальное – 0), незначимого характера – 1 (максимальное – 8, минимальное – 0). Наиболее часто встречающиеся замечания по функциональному тестированию:

1. Отсутствует предупреждение об исследовательских целях, информация об ИИ-сервисе и его работе.
2. Отсутствие подписей к названию серии от ИИ-Сервиса.
3. Отсутствуют результаты работы ИИ-Сервиса.
4. Дефекты при работе с дополнительной серией – изменение яркости, контрастности, искажение.
5. Дублирование DICOM SR и/или дополнительной серии.

Калибровочное тестирование (методология проведения приведена в подпараграфе 4.3 данной главы). Всего проведено 38 процедур (23 первичных и 15 повторных) для 70,0% ИИ-сервисов, успешно прошедших функциональное тестирование. Скорость доработки ИИ-сервиса при неуспешном первичном прохождении тестирования в среднем составила 52 дня [11 – 163] (рис. 31). Не смогли успешно завершить этап 6 сервисов – 15,0% сервисов, участвующих в Эксперименте. Результаты калибровочных тестирований как научного исследования (второго этапа диагностической точности) представлены в главе 8.



Рисунок 31 – Динамика устранения замечаний по итогам калибровочного тестирования

Контрольно-техническое тестирование (методология проведения приведена в подпараграфе 4.4 данной главы). Всего проведено 20 процедур с охватом 295 диагностических устройств (из них КТ – 32,0%, РГ/ФЛГ – 53,0%, ММГ – 15,0%) (таблица 37).

Таблица 37 – Результаты контрольно-технического тестирования (по состоянию на 25.12.2020)

Качество результатов работы ИИ-сервисов	Количество диагностических устройств, шт.
Без технологических дефектов	156
Частично содержат технологические дефекты	80
Содержат технологические дефекты	59

Выявленные технологические дефекты обусловлены как особенностями конкретных диагностических устройств (корректность выбора настроек рентгенолаборантом, корректность DICOM-тегов, стабильность подключения к ЕРИС ЕМИАС), так и особенностями ИИ-сервиса (отсутствие алгоритмов проверки входных данных, настройка DICOM-тегов, алгоритма выбора проекций и серий и т.д.).

В результате аналитики работы ИИ-сервисов с диагностическими устройствами выявлено отсутствие результатов работы ИИ-сервисов в исследованиях с данными ДУ: для КТ в 18,0–72,0% исследований, для РГ/ФЛГ – в 19,0–28,0%, для ММГ – в 26,0%.

В ходе Эксперимента в 2020 году не было полного охвата процедурой контрольно-технического тестирования всех диагностических устройств МО ДЗМ, подключенных к ЕРИС ЕМИАС, что могло сказаться на качестве части исследований. Для решения этой проблемы необходимо проводить самотестирование ИИ-сервиса на данных со всего парка оборудования.

По итогам контрольно-технического тестирования вносились уточнения в схему маршрутизации исследований в Эксперименте. Среднее время от подачи заявки до успешного завершения интеграции (прохождения всех этапов тестирования и допуска к работе с реальным потоком исследований) составило 65 дней±34 дня (таблица 38). На рис. 32 приведен график, отражающий время интеграции ИИ-сервисов в зависимости от модальности.



Рисунок 32 – Длительность технологической интеграции ИИ-сервисов

Таблица 38 – Диапазоны длительности технологической интеграции ИИ-сервисов

Модальность	Время до успешного завершения технологической интеграции (сут.)
КТ / НДКТ РЛ	39 – 128
КТ Covid	16 – 110
ММГ	59 – 154
РГ / ФЛГ	54 – 106

Тестирование также могло проводиться после интеграции в промышленный программно-аппаратный комплекс ЕРИС ЕМИАС (ППАК ЕРИС ЕМИАС). При прохождении технической интеграции работа Сервиса переносится в ППАК ЕРИС ЕМИАС. По письменному запросу претендента на участие в Эксперименте информация о прохождении технической интеграции и переносе сервиса в ППАК ЕРИС ЕМИАС может быть подтверждена оператором ЕРИС ЕМИАС. С момента технической интеграции в ППАК ЕРИС ЕМИАС претендент на участие в Эксперименте становится участником Эксперимента.

О неудачном прохождении технической интеграции претендент на участие в Эксперименте уведомляется письмом, направляемым по адресу электронной почты, указанной им в заявке, а также по месту нахождения юридического лица. Риск неполучения корреспонденции по адресам, указанным в заявке, несет претендент на участие в Эксперименте.

В случае положительного решения по итогу рассмотрения предоставленного комплекта документов, а также после прохождения интеграции участник Эксперимента вправе заключить с ГБУЗ «НПКЦ ДИТ ДЗМ» соглашение об участии в Эксперименте по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы.

Общий сценарий работы ИИ-сервисов в ППАК ЕРИС ЕМИАС. Эксперимент проводился по четырем типам лучевых исследований:

1. Компьютерная томография и низкодозная компьютерная томография грудной клетки с целью выявления злокачественных новообразований (рака легких) (далее – КТ РЛ / НДКТ РЛ).

2. Маммография с целью выявления злокачественных новообразований молочных желез (далее – ММГ).

3. Рентгенография / флюорография легких с целью определения патологии легких (далее – РГ / ФЛГ).

4. Компьютерная томография грудной клетки с целью выявления новой коронавирусной инфекции COVID-19 (далее – КТ COVID).

Как было сказано выше, первоначально предполагалось проведение Эксперимента только по 3 типам, однако интенсивное развитие пандемии новой коронавирусной инфекции обусловило необходимость оперативного расширения.

Диагностические исследования поступают от подключенных диагностических устройств в ППАК ЕРИС ЕМИАС. Техническими средствами ЕРИС ЕМИАС осуществляется обезличивание исследований (изображений), их отправка на обработку ИИ-сервисом, возвращение исследований (изображений) в ЕРИС ЕМИАС и деобезличивание их результатов. ИИ-сервис проводит анализ обезличенных исследований (изображений), представляет результаты анализа для обратной отправки в ЕРИС ЕМИАС.

В ходе работы в ЕРИС ЕМИАС ИИ-сервисам предлагалось решать последовательно несколько задач для врача-рентгенолога:

- предоставлять информацию о повышении приоритета исследования в рабочем списке в случае выявления патологической находки на изображениях исследования;

- локализовать патологические находки на изображениях исследования в виде маркировок на дополнительной серии изображений, эквивалентных оригинальным изображениям исследования;

- классифицировать выявленные находки с указанием вероятности патологии;

- автоматически формировать отчет о результате работы ИИ-сервиса с предлагаемым шаблоном описания исследования.

Отметим, что ИИ-сервисы предлагали именно вероятность наличия той или иной патологии, а не однозначный ответ. Соответственно, финальное решение при интерпретации и описании результатов лучевого исследования всегда принимает врач-рентгенолог.

Результаты автоматизированного анализа становились доступными на рабочих местах врачей-рентгенологов ЕРИС ЕМИАС в виде дополнительной серии, содержащей маркировку находок на нативных изображениях исследования, текстовую информацию, при этом исследования с выявленными патологиями автоматически приоритизируются в рабочих списках врачей.

В процессе работы ИИ-сервисов в ППАК ЕРИС ЕМИАС осуществлялся регулярный **технологический мониторинг** (методология проведения приведена в подпараграфе 4.5 данной главы).

Всего проведено 158 процедур. Результаты использованы для научного анализа целесообразности применения ИИ-сервисов в лучевой диагностике (подробно представлены в главе 9).

Результаты технологического мониторинга регулярно представлялись на заседаниях комиссии ДЗМ, на их основе принимались

решения о продолжении, приостановлении или прекращении участия данного ИИ-сервиса в Эксперименте. Также результаты технологического мониторинга учитывались при определении объемов грантов, выплачиваемых участникам Эксперимента в соответствии с порядком его проведения.

На всех этапах Эксперимента осуществлялись сбор и анализ информации для **научных целей**; результаты этой работы подробно представлены в следующих главах.

Итоговые данные. Всего в Эксперименте приняли участие 39 ИИ-сервисов от 21 компании-разработчика; 18 ИИ-сервисов успешно интегрировались в ППАК ЕРИС ЕМИАС и работали в условиях реального потока результатов лучевых исследований.

Всего на анализ было направлены результаты 1762949 исследований; из них проанализированы ИИ-сервисами 83,3% (1468872), в том числе: результаты КТ и НДКТ — 56,0% (818296), ММГ — 4,0% (61497), РГ — 18,0% (270965), ФЛГ — 22,0% (318114).

На анализ направлялись результаты исследований по четырем модальностям, проведенных на 935 диагностических устройствах в 296 медицинских организациях Департамента здравоохранения Москвы, оказывающих медицинскую помощь взрослому населению (включая филиалы). Результаты работы ИИ-сервисов были доступны для использования и оценки на автоматизированных рабочих местах в ЕРИС ЕМИАС 538 врачей-рентгенологов (физических лиц) (рис. 33).

По состоянию на 25.12.2020 компаниям-участникам Эксперимента выплачен 41 грант на общую сумму 213 391 240 рублей.

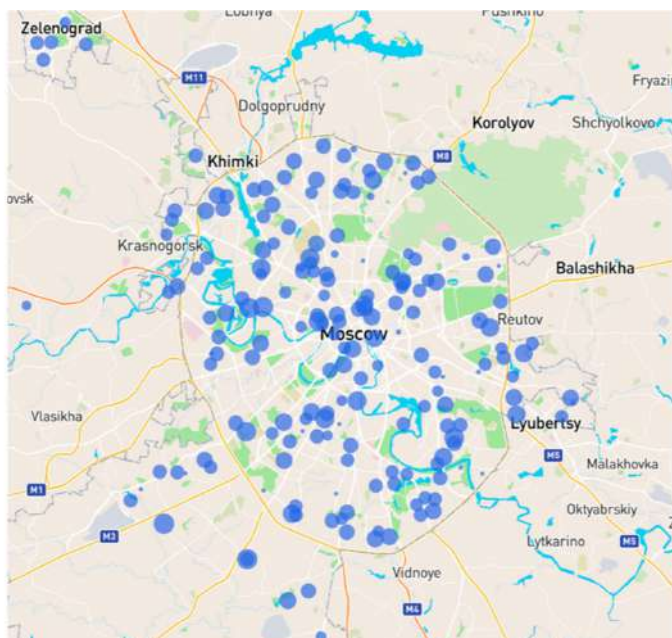


Рисунок 33 – Схема распределения и интенсивности работы медицинских организаций, участвующих в Эксперименте

В декабре 2020 года по инициативе Департамента информационных технологий города Москвы проведен сбор обратной связи от компаний-участников Эксперимента. Как показали результаты опроса, участие в Эксперименте позволило большинству респондентов развить свой продукт (рис. 34). Самым ценным в участии в Эксперименте компании назвали практический опыт, разностороннюю экспертную оценку продукта и гранты (рис.35).

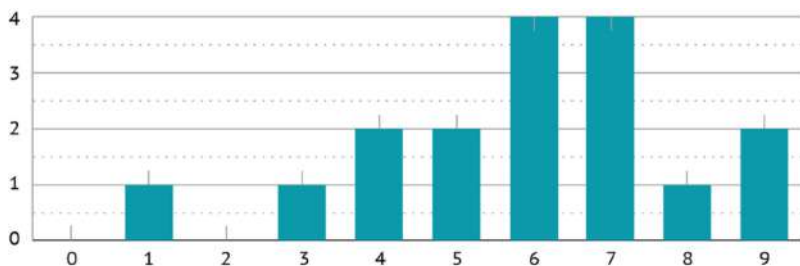


Рисунок 34 – Эволюция ИИ-сервисов по оценке компаний-разработчиков в баллах от 0 до 9



Рисунок 35 – Определение ценностей участия в Эксперименте (каждая компания могла выбрать 3 ответа)

Примечательно, что все разработчики однозначно рекомендовали участие другим компаниям (индекс лояльности составил 100). При этом абсолютное большинство компаний хотели бы улучшить способ получения обратной связи и коммуникации с организаторами Эксперимента.

Глава 7. НАУЧНАЯ МЕТОДОЛОГИЯ ЭКСПЕРИМЕНТА

В этой главе подробно представлена методология проведения научного исследования «Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы».

Цель исследования: оценить возможность использования в системе здравоохранения города Москвы технологий компьютерного зрения, их качество и влияние на работу службы лучевой диагностики.

Задачи исследования:

1. Обосновать клинический контекст применения технологий компьютерного зрения в лучевой диагностике.
2. Разработать методологическую основу для оценки качества и применимости технологий компьютерного зрения в лучевой диагностике.
3. Разработать и реализовать организационно-методические мероприятия по внедрению технологий компьютерного зрения в работу отделений лучевой диагностики.
4. Оценить диагностическую точность сервисов на основе технологий компьютерного зрения в условиях проспективной работы в радиологической информационной системе г. Москвы.
5. Изучить целесообразность применения технологий компьютерного зрения, в том числе с позиций технологического качества и устойчивости сервисов, их влияния на удовлетворенность и производительность труда врачей-рентгенологов.

Гипотезы исследования:

№1. H_0 : {Чувствительность $< 0,81$ и/или Специфичность $< 0,81$ и/или AUC $< 0,81$ }.

№2. H_0 : {Уровень дефектов работы ИИ-сервисов типа «а» и/или «б» и/или «в» и/или «г» и/или «д» и/или «е» $> 10\%$ }.

№3. H_0 : {Снижение длительности описания результатов лучевых

исследований при работе ИИ-сервисов <30%}.

Дизайн: исследование со смешанными методами (конкурентный дизайн с полным смешиванием и равными статусами¹; вариант интеграции: «дополнительное покрытие»²).

Научное исследование «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы» (далее — Эксперимент) проведено в период с 01.01.2020 по 31.12.2020 в медицинских организациях Департамента здравоохранения города Москвы, оказывающих первичную медико-санитарную помощь и специализированную взрослому населению в амбулаторных или стационарных условиях.

Критерии включения:

1. Возраст старше 18 лет.
2. Пол мужской или женский.
3. Наличие в карте пациента направления на проведение и результатов проведения одного из видов лучевых исследований.

4. Результаты в стандарте DICOM следующих видов лучевых исследований:

- компьютерной томографии органов грудной клетки,
- низкодозной компьютерной томографии органов грудной клетки,
- маммографии профилактической,
- рентгенографии органов грудной клетки легких,
- флюорографии органов грудной клетки легких.

Критерии исключения:

1. Лучевые исследования другого вида (в т.ч. иная модальность, анатомическая область).
2. Отсутствие подписанного информированного согласия на участие в Эксперименте.

Экспериментальное вмешательство.

¹ Schoonenboom J., Johnson R.B. How to Construct a Mixed Methods Research Design. // Kolner Z Soz Sozpsychol. 2017. Vol. 69. Suppl 2. P. 107–131. DOI:10.1007/s11577-017-0454-1.

² O’Cathain A., Murphy E., Nicholl J. The quality of mixed methods studies in health services research // J Health Serv Res Policy. 2008 Apr. Vol. 13, №2. P. 92–98. DOI: 10.1258/jhsrp.2007.007074.

ИИ-сервисы использовались врачами-рентгенологами при описании результатов лучевых исследований в ЕРИС ЕМИАС. Перед внедрением ИИ-сервисов проводилось обучение пользователей (врачей-рентгенологов) (см. параграф 3 главы 6).

При обращении пациента в медицинскую организацию ДЗМ осуществлялось стандартное обследование, в процессе которого лечащий врач устанавливал наличие показаний для лучевой диагностики и формировал соответствующее направление на исследование, которое проводилось в установленном порядке, а результаты автоматически сохранялись в ЕРИС ЕМИАС. Для видов исследований, включенных в Эксперимент (см. критерии включения), настраивалась автоматическая маршрутизация на ИИ-сервис. Перечень ИИ-сервисов — участников Эксперимента представлен в таблице 41. Схема маршрутизации устанавливалась в рамках организационно-методических мероприятий по подготовке и проведению Эксперимента в соответствии с его нормативным обеспечением. От данного диагностического устройства результаты исследований конкретного вида маршрутизировались только на один ИИ-сервис в данный период времени и автоматически обрабатывались им. Результаты такой обработки отправлялись в ЕРИС ЕМИАС в формате DICOM SR, где становились доступными в АРМе врача-рентгенолога в виде дополнительной серии. Все изображения в этой серии маркировались сообщением «Только в исследовательских целях».

Функциональные задачи ИИ-сервисов в ЕРИС ЕМИАС:

- предоставлять информацию о повышении приоритета исследования в рабочем списке в случае выявления патологической находки на изображениях исследования;
- локализовать патологические находки на изображениях исследования в виде маркировок на дополнительной серии изображений, эквивалентных оригинальным изображениям исследования;
- классифицировать выявленные находки с указанием вероятности патологии;
- автоматически формировать отчет о результате работы ИИ-сервиса с предлагаемым шаблоном описания исследования.

ИИ-сервисы предлагают именно вероятность наличия той или иной патологии, а не однозначный ответ. Соответственно, финальное решение при интерпретации и описании результатов лучевого исследования всегда принимает врач-рентгенолог.

Таблица 39 – Список участников Эксперимента (ИИ-сервисов)

№ п/п	Юридическое лицо	Наименование ИИ-сервиса	Модальность
1	АНО ВО «Университет Иннополис»	AI RADIOLOGY CXR	РГ
2	АО «Р-Фарм»	Lunit INSIGHT CXR	РГ
3		Lunit INSIGHT MMG	ММГ
4	ООО «АЙРА Лабс»	COVID-19 Intelligent Radiology Assistant	КТ Covid
5	ООО «Айрим»	IRYM COVIDetect	КТ Covid
6	ООО «Гаммамед-Софт»	Гамма Мультивокс Ковирус	КТ Covid
7	ООО «Интеллоджик»	Botkin.AI	КТ РЛ
8	ООО «КэреМенторЭйАй»	CarementorAI	РГ
9		CarementorAI	КТ Covid
10	ООО «Медицинские скрининг системы»	Цельс	ФЛГ
11		Цельс	ММГ
12	ООО «МИП «Биномикс-Рэй»	Virtual Resident (Radlogics)	КТ Covid
13	ООО «СиВижинЛаб»	CVL	КТ Covid
14	ООО «Платформа Третье Мнение»	ТретьеМнение	РГ
15		Третье Мнение. КТ-COVID-19	КТ Covid
16	ООО «Филипс Инновационные Лаборатории РУС»	Philips Research Lung Cancer Screening	КТ РЛ (НДКТ)
17	ООО «ФтизисБиоМед»	FBM	РГ
18		FBM	ФЛГ
19	АНО ВО «Университет Иннополис»	AI RADIOLOGY CXR	ФЛГ
20	ООО «Ай.Эс.Эм.Ай»	JLD-02K	РГ
21		JBD-01K	ММГ
22		JLD-01K	КТ РЛ
23	ООО «Айрим»	IRYM mammography	ММГ

Продолжение таблицы 39

№ п/п	Юридическое лицо	Наименование ИИ-сервиса	Модальность
24	ООО «Генетические технологии»	G42 Cogna AI	KT Covid
25	ООО «Интеллоджик»	Botkin.AI	KT Covid
26		Botkin.AI	РГ
27	ООО «Крайон»	CT Image Analytics for COVID-19	KT Covid
28	ООО «КэреМенторЭйАй»	CarementorAI	ММГ
29	ООО «Лаборатория Сенсмейкинг»	Aspect	KT Covid
30	ООО «Медицинские скрининг системы»	Цельс	КТ РЛ
31		Цельс	КТ Covid
32	ООО «МИП «Биномикс-Рэй»	Virtual Resident (Radlogics)	КТ РЛ
33		Radlogics (РГ)	РГ
34	ООО «ОТР – безопасность информационных технологий»	HY Covid-19	KT Covid
35	ООО «СберМедИИ»	SAI Pneumonia	KT Covid
36	ООО «Студия Данных»	DS_lung_diseases_CT	KT Covid
37		DS_lung_diseases	РГ
38		Mamm	ММГ
39	ООО «Честная медицина»	qXR V	РГ
Примечание: ИИ-сервисы №1–18 успешно интегрировались в ТПАК ЕРИС ЕМИАС; ИИ-сервисы №19–39 на дату завершения Эксперимента (25.12.2020) находились на этапе технической интеграции в ППАК ЕРИС ЕМИАС.			

Врач-рентгенолог, проводящий описание исследования в ЕРИС ЕМИАС, мог открыть дополнительную серию, изучить результаты работы ИИ-сервиса, далее – использовать, не использовать или частично использовать эти результаты при формировании описания. Обращение к дополнительной серии было сугубо добровольным, зависело от индивидуального решения врача. По результатам работы

с дополнительной серией врач-рентгенолог мог добровольно заполнить форму обратной связи. Детальное описание требований к результатам работы ИИ-сервисов приведено в подпараграфе 4.1 главы 6. В процессе участия ИИ-сервисов в Эксперименте проводились регулярный технологический мониторинг (см. подпараграф 4.5 главы 6) отбор исследований для аудита (см. далее). С учетом ограниченного периода наблюдения и отсутствия полноценного объема данных в исследование были **включены только ИИ-сервисы №1–18** (таблица 39).

Первичная точка: обеспечение значений основных показателей диагностической точности ИИ-сервисов на уровне не менее 81%.

Вторичные точки:

1. Обеспечение уровня технологических дефектов всех типов в работе ИИ-сервисов не более 10%.
2. Увеличение не менее чем на 30% количества описываемых врачом-рентгенологом исследований за смену.
3. Сохранение или улучшение показателей удовлетворенности врачей-рентгенологов от взаимодействия с ЕРИС ЕМИС при использовании ИИ-сервисов.

Схемы исследования и потока данных в исследовании представлены на рис. 36, 37.

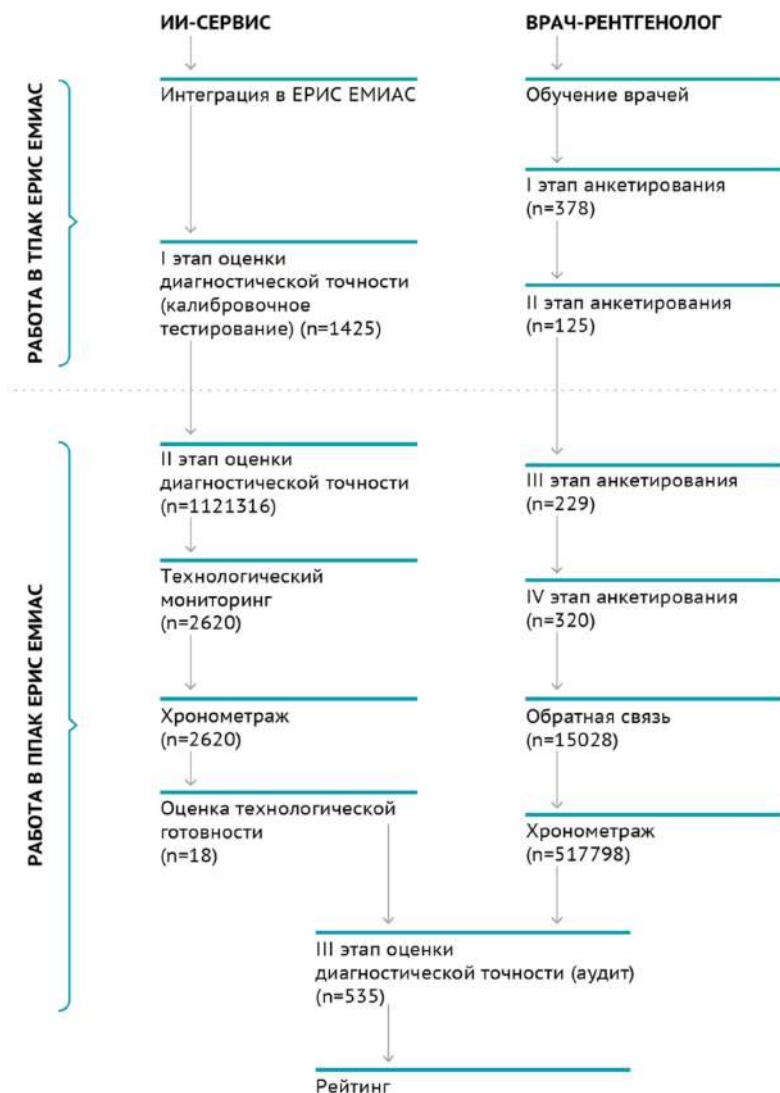


Рисунок 36 – Общий ход научного исследования

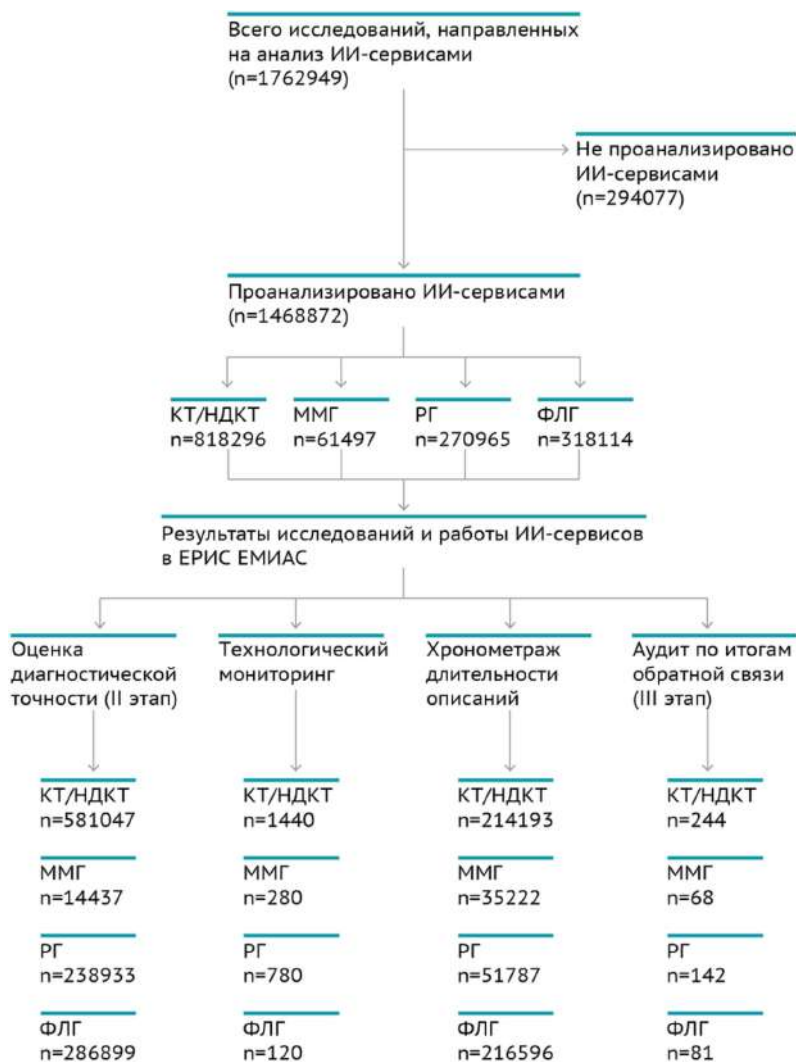


Рисунок 37 – Поток данных в научном исследовании (без учета данных контрольных групп)

Первый компонент исследования (количественный): мульти-центровое, проспективное диагностическое исследование ИИ-сер-

висов (diagnostic study). Формирование гипотезы для диагностического компонента исследования осуществлено по методике Korevaar et al., 2019¹. В соответствии с методикой для индекс-теста (ИИ-сервисов) установлены минимально приемлемые уровни критериев, оправдывающие предполагаемое клиническое применение исследуемого индекс-теста:

- чувствительность – 0,81 (с учетом 95% ДИ),
- специфичность – 0,81 (с учетом 95% ДИ), – площадь под кривой (AUC) – 0,81 (с учетом 95% ДИ).
- площадь под кривой (AUC) – 0,81 (с учетом 95% ДИ).

Нулевая гипотеза для каждого из этапов первого компонента исследования: $H_0: \{\text{Чувствительность} < 0,81 \text{ и/или Специфичность} < 0,81 \text{ и/или AUC} < 0,81\}$.

Первый компонент содержал три этапа оценки для каждого, включенного ИИ-сервиса:

1. Первый этап оценки:

- индекс-тест: ИИ-сервис, интегрированный в ТПАК ЕРИС;
- референс-тест: эталонные наборы данных с разметкой.

Подготовка и разметка эталонных наборов данных выполнена по оригинальной методологии, описанной в главе 4. Объем выборки составил 1425 исследований, сгруппированных в 13 наборов (таблица 40). В рамках Эксперимента указанные наборы использованы для процедуры калибровочного тестирования, которая методически представляла собой оценку диагностической точности ИИ-сервисов.

¹ Korevaar D.A., Gopalakrishna G., Cohen J.F. et al. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses // *Diagn Progn.* 2019. Res 3, 22. URL: <https://doi.org/10.1186/s41512-019-0069-2>.

Таблица 40 – Эталонные наборы данных с разметкой, использованные для калибровочного тестирования и оценки диагностической точности на первом этапе

№	Модальность	Диагностическое направление	Количество исследований в наборе данных, абс.
1	КТ	Поиск COVID-19	100
2	КТ*	Поиск COVID-19	100
3	КТ	Поиск COVID-19 с классификацией КТ0-4	125
4	КТ	COVID-19 с классификацией КТ0-4	200
5	КТ РЛ	Поиск ЗНО легкого	100
6	НДКТ РЛ	Поиск ЗНО легкого	100
7	ММГ	Поиск ЗНО молочной железы	100
8	ММГ*	Поиск ЗНО молочной железы	100
9	РГ	Поиск COVID-19	100
10	РГ	Поиск патологий по списку	100
11	РГ*	Поиск патологий по списку	100
12	ФЛГ	Поиск патологий по списку	100
13	ФЛГ*	Поиск патологий по списку	100
* Отдельные оригинальные наборы данных для повторного проведения калибровочного тестирования.			

2. Второй этап оценки:

– индекс-тест: ИИ-сервис, интегрированный в ППАК ЕРИС, успешно прошедший функциональное и калибровочное тестирование;

– референс-тест: заключение врача-рентгенолога, непосредственно проводившего исследование в МО ДЗМ.

Выборка сформирована последовательно, проспективно включались 100,0% исследований каждого вида, проанализированных ИИ-сервисами в период Эксперимента. Общий объем выборки составил 1 468 872, в том числе результаты КТ и НДКТ – 818296,

ММГ — 61497, РГ — 270965, ФЛГ — 318114. Информация по выборкам для каждого сервиса представлена в таблице 41.

Таблица 41 – Выборки для второго этапа оценки диагностической точности каждого ИИ-сервиса

Название ИИ-сервиса	Объем выборки
AI RADIOLOGY CXR (РГ)	17006
Lunit INSIGHT CXR (РГ)	8010
Lunit INSIGHT ММГ (ММГ)	3123
COVID-19 Intelligent Radiology Assistant (КТ Covid)	110773
IRYM COVIdetect (КТ Covid)	27859
Гамма Мультивокс Ковирус (КТ Covid)	139143
Botkin.AI (КТ РЛ)	46713
CarementorAI (РГ)	96517
CarementorAI (КТ Covid)	40975
Цельс (ФЛГ)	265366
Цельс (ММГ)	11314
Virtual Resident (Radlogics) (КТ Covid)	142538
CVL (КТ Covid)	9290
ТретьеМнение (РГ)	53309
Третье Мнение. КТ-COVID-19 (КТ Covid)	60363
Philips Research Lung Cancer Screening (НДКТ РЛ)	3393
FBM (РГ)	64091
FBM (ФЛГ)	21533

Заключения с описанием результатов лучевых исследований, использованных в качестве референс-теста, находятся в базе данных ЕРИС ЕМИАС. Для обработки такого массива информации нами разработано прикладное программное обеспечение MedLabel — автоматизированный анализ медицинских протоколов¹.

Основной функционал этого продукта — извлечение из текстовых данных (описания и заключения лучевых исследований) признаков описания патологических изменений путем выполнения следующих этапов:

- Предварительная обработка, на которой текстовые и неструктурированные рентгенологические отчеты исследований преобразовываются и стандартизируются с помощью различных методов обработки естественного языка.

- Текстовое представление. На данном этапе неструктурированные входные данные преобразуются в подходящую модель представления, которая позволяет проводить эффективно анализ на последующих этапах, например, так называемый «мешок слов» (англ. «Bag of Words») и др.

- Стадия обнаружения, на которой полезная информация извлекается из коллекций текстовых данных посредством применения определенных методов и приемов, например, классификация, кластеризация методами машинного обучения, применение алгоритма на основе ключевых слов, признаков.

Инструмент анализа врачебных протоколов Medlabel использован нами для проспективного исследования диагностической точности ИИ-сервисов по видам исследований.

3. Третий этап оценки:

- индекс-тест: ИИ-сервис, интегрированный в ППАК ЕРИС, успешно прошедший функциональное и калибровочное тестирование;

- референс-тест: заключение эксперта-врача-рентгенолога Референс-центра лучевой диагностики.

На данном этапе проводился так называемый аудит результатов лучевых исследований в соответствии с ранее разработанной и успешно валидированной авторской методикой². Аудит выполнялся высококвалифицированными экспертами (врачами-рентгенологами с подготовкой по субспециализациям) — сотрудниками

¹ Свидетельство о государственной регистрации программы для ЭВМ «MedLabel — автоматизированный анализ медицинских протоколов» №2020664321, дата заявки: 27.10.2020, дата регистрации: 11.11.2020.

² Оценка качества рентгенорадиологических исследований / сост. С. П. Морозов, Н. Н. Ветшева, Н. В. Ледихова // Сер. «Лучшие практики лучевой и инструментальной диагностики». М., 2019. Вып. 22.

ГБУЗ «НПКЦ ДиТ ДЗМ» г. Москвы, с 05.08.2020 – Референс-центра лучевой диагностики на базе указанной медицинской организации³.

Выборка исследований на аудит происходила автоматически, согласно заданным информационной системой параметрам (таблица 42, рис. 38). На аудит исследование отправляется даже в том случае, если оно параллельно обработано другим ИИ-сервисом, подключенным ранее.

В рамках Эксперимента аудит ИИ-сервиса – это оценка результатов как врача-рентгенолога, так и ИИ-сервиса. Согласно базовой методике автоматически отобранное исследование направляется на первый этап аудита. Врач-эксперт №1 проводит анализ результатов работы врача и ИИ-сервиса, заполняет поля чек-листа (таблица 43).

Таблица 42 – Критерии формирования выборки для аудита

Критерий	Максимальный объем исследований для одного ИИ-сервиса
Врач-рентгенолог отклонил результаты работы ИИ-сервиса (в форме обратной связи выбрано значение «Не согласен» и/или вынесены дефекты в поле «Комментарий»)	Не более 10 в квартал
ИИ-сервис выходит из Эксперимента (прекращает работу)	20

50 с.

³ Приказ Департамента здравоохранения города Москвы от 05.08.2020 №764 «Об организации описания данных, полученных при проведении лучевых методов исследований, в ЕРИС ЕМИАС».

Сводный отчет о результатах обоих этапов аудита формируется в информационной системе на основе заполненных чек-листов по каждому исследованию, прошедшему аудит. Информация по выборкам для каждого сервиса представлена в таблице 44.

Таблица 44 – Выборки для третьего этапа оценки диагностической точности каждого ИИ-сервиса

Название ИИ-сервиса	Объем выборки
AI RADIOLOGY CXR (РГ)	18
Lunit INSIGHT CXR (РГ)	19
Lunit INSIGHT MMG (ММГ)	4
COVID-19 Intelligent Radiology Assistant (КТ Covid)	54
IRYM COVIdetect (КТ Covid)	12
Гамма Мультивокс Ковирус (КТ Covid)	75
Botkin.AI (КТ РЛ)	32
CarementorAI (РГ)	53
CarementorAI (КТ Covid)	19
Цельс (ФЛГ)	51
Цельс (ММГ)	64
Virtual Resident (Radlogics) (КТ Covid)	
CVL (КТ Covid)	20
ТретьеМнение (РГ)	-
Третье Мнение. КТ-COVID-19 (КТ Covid)	30
Philips Research Lung Cancer Screening (НДКТ РЛ)	2
FBM (РГ)	52
FBM (ФЛГ)	30

При выполнении первого компонента исследования использованы методы оценки и сравнения диагностической точности путем определения и сравнения следующих метрик: чувствительность, специфичность, площадь под характеристической кривой (т.н. «AUC» от англ. area under curve), коэффициент согласованности классификаций (каппа Козна). При расчетах диагностических метрик установлен порог активации, рассчитанный по индексу Юдена.

В целом для проведения первого компонента использована авторская методология клинических испытаний программного обеспечения на основе интеллектуальных технологий в сфере лучевой диагностики, описанная в главе 5.

Второй компонент исследования (качественный): оценка целесообразности применения ИИ-сервисов (feasibility study). В рамках данного компонента изучены приемлемость, спрос, реализация, осуществимость, интеграция и способность к адаптации в соответствии с типовыми аспектами и результатами исследований целесообразности (таблица 45)¹.

¹ Bowen D.J., Kreuter M., Spring B. et al. How we design feasibility studies // Am J Prev Med. 2009. Vol. 36, №5. P. 452–457. DOI:10.1016/j.amepre.2009.02.002.

Таблица 45 – Аспекты оценки целесообразности применения ИИ-сервисов

Аспект	Гипотезы	Метрики
Приемлемость	Внедрение ИИ-сервисов не вызывает интерес пользователей (врачей-рентгенологов). Использование ИИ-сервисов не изменяет уровень удовлетворенности пользователей (врачей-рентгенологов)	– Удовлетворенность пользователей. – Вовлеченность пользователей. – Предполагаемая целесообразность использования – Предполагаемое положительное или отрицательное воздействие на производительность и качество работы
Спрос	Реальный спрос на постоянное использование ИИ-сервисов у пользователей (врачей-рентгенологов) отсутствует	– Фактическое использование. – Намерение продолжать использование
Реализация	В условиях интеграции в радиологическую информационную систему субъекта РФ ИИ-сервисы не обеспечивают качественный анализ результатов >90% исследований в установленные временные рамки Нулевая гипотеза: H_0 : {Уровень дефектов работы ИИ-сервисов типа «а» и/или «б» и/или «в» и/или «г» и/или «д» и/или «е» >10%}	– Степень реализации (объем проанализированных исследований по видам). – Уровень и структура технологических дефектов. – Время, затрачиваемое на обработку одного исследования
Осуществимость	В условиях реального производственного процесса отделений лучевой диагностики ИИ-сервисы не повлияют на скорость подготовки описаний результатов исследований врачами-рентгенологами. Нулевая гипотеза: H_0 : {Снижение длительности описания результатов лучевых исследований при работе ИИ-сервисов <30%}	– Положительное/отрицательное воздействие на производительность труда врачей-рентгенологов

Продолжение таблицы 45

Аспект	Гипотезы	Метрики
Интеграция	На данном уровне технического развития ИИ-сервисы не могут быть интегрированы в существующие радиологические информационные системы с достаточным уровнем качества	– Соответствие инфраструктуре и стандартам – Скорость интеграции – Устойчивость в работе
Способность к адаптации	ИИ-сервисы утрачивают производительность при изменении клинического контекста применения	– Сроки готовности к испытаниям обновленных алгоритмов. – Показатели производительности и стабильности работы при изменении клинического контекста. – Первоначально планировалось экспериментальное изменение клинических сценариев и подключение дополнительных популяций пациентов. Однако из-за развития пандемии COVID-19 появилась возможность оценить адаптируемость ИИ-сервисов в реальных условиях

В рамках данного компонента использованы следующие методы:

1. Социологический опрос пользователей (врачей-рентгенологов) ИИ-сервисов.

Социологическое исследование проводилось с целью сбора и систематизации мнений и оценок практикующих врачей-рентгенологов вопросов использования технологий искусственного интеллекта в их профессиональной деятельности.

Были поставлены следующие задачи:

1. Оценить готовность врача-рентгенолога к практическому использованию технологий искусственного интеллекта/компьютерного зрения по результатам проведенного обучения.

2. Оценить удовлетворенность и приверженность врачей-рентгенологов использованию технологий искусственного интеллекта/компьютерного зрения на практике, в том числе провести оценку:

— первичных ожиданий врачей-рентгенологов от внедрения технологии ИИ;

— динамики отношения врачей-рентгенологов к технологиям ИИ в процессе работы с ними.

3. Оценить эффективность работы ИИ-сервисов и сравнить их по параметрам практического использования.

4. Оценить качество работы службы технической поддержки при использовании ИИ-технологий.

Опрос врачей-рентгенологов проводится путем онлайн-анкетирования. Анкетирование проходят все врачи-рентгенологи амбулаторно-поликлинического звена, подтвердившие успешное прохождение обучения по результатам тестирования, а также давшие свое согласие на участие в данном компоненте Эксперимента и обработку персональных данных. В рамках исследования предполагается использование технологии ИИ данными специалистами в их повседневной работе.

Идентификация респондентов. При регистрации на обучающий вебинар (см. параграф 3 главы 6) врач-рентгенолог указывает свою электронную почту, Ф.И.О. и специализацию. В дальнейшем каждый раз проходя анкетирование, респондент также указывает электронную почту в онлайн-форме опросника, посредством которой идентифицируется.

Структура опросника. Все вопросы разбиты на 5 блоков:

I. Оценка качества обучения (n=378).

II. Ожидания от внедрения технологий ИИ в работу врача-рентгенолога (n=125).

III. Отношение к технологиям ИИ в работе врача-рентгенолога (n=229).

IV. Оценка конкретного (тестируемого) ИИ-сервиса (n=320).

V. Оценка качества работы службы технической поддержки.

План-график анкетирования. Анкетирование проводится в несколько этапов (рис. 39):

Этап 1. Проводится сразу по окончании обучения. Включает в себя I и II блоки вопросов.

Этап 2. Проводится через 3 месяца работы с первым ИИ-сервисом. Включает в себя III и IV блоки вопросов.

Этап 3. Проводится после 1-го месяца работы со следующим (не первым) ИИ-сервисом. Включает в себя IV блок вопросов. Данный этап может проводиться несколько раз, в зависимости от того,

сколько новых (не первых) ИИ-сервисов будет тестировать врач-рентгенолог.

Этап 4. Проводится через 6–8 месяцев использования ИИ-сервисов (декабрь 2020 — январь 2021 гг.). Включает в себя III и V блоки вопросов.

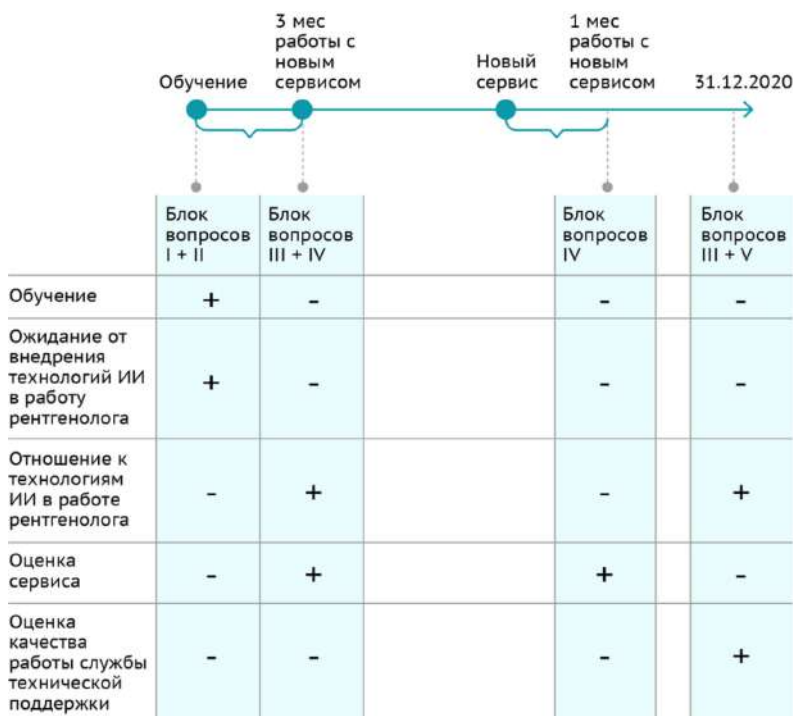


Рисунок 39 – План-график социологического опроса врачей-рентгенологов в рамках Эксперимента

Методика опроса. Применяется выборочный опосредованный социологический опрос. Анкеты содержат открытые вопросы как форму обратной связи; вопросы, в которых предложено выбрать один или несколько вариантов ответа; а также вопросы, оцениваемые по шкале от 0 до 10. Для определения приверженности врачей-рентгенологов ИИ-технологиям и ИИ-сервисам (индекс го-

товности рекомендовать) используется индекс потребительской лояльности (NPS от англ. Net Promoter Score).

Расчет индекса NPS включает в себя несколько шагов.

1. Респондентам предлагается ответить на несколько вопросов, оцениваемых по 11-балльной шкале, где 0 соответствует «Ни в коем случае не буду рекомендовать», а 10 – «Обязательно порекомендую».

2. На основе полученных оценок все потребители разделяются на 3 группы: 9–10 баллов – сторонники, 7–8 баллов – нейтральные, 0–6 баллов – критики.

3. Непосредственно расчет индекса NPS: $NPS = \% \text{ сторонников} - \% \text{ критиков}$.

1. Процедура технологического мониторинга (описана в параграфе 2 главы 9).

2. Сбор и анализ обратной связи по конкретным клиническим случаям от пользователей (врачей-рентгенологов).

Для получения информации об оценке пользователями (врачами-рентгенологами) качества работы ИИ-сервиса в каждом конкретном клиническом случае, а также для оценки вовлеченности врачей разработана специальная форма обратной связи. Форма интегрирована в интерфейс автоматизированного рабочего места врача-рентгенолога в ЕРИС ЕМИАС (рис. 40, 41).

Поле обратной связи

Выберите сервис

Выберите «Согласен»

Рисунок 40 – Заполнение формы обратной связи в ЕРИС ЕМИАС при согласии врача с ИИ-сервисом

Поле обратной связи

Выберите сервис

Выберите «Не согласен»

Заполните Комментарий

Рисунок 41 – Заполнение формы обратной связи в ЕРИС ЕМИАС при несогласии врача с ИИ-сервисом

В том случае, если рентгенолог уже работает с результатами ИИ-сервиса, оформление результатов исследования, обработанного ИИ, подразумевает особые правила:

- если врач согласен с результатами ИИ-сервиса, то не рекомендуется изменять основную часть предзаполненного шаблона;
- взаимодействие с ИИ-сервисом, как при согласии, так и несогласии с его работой, включает в себя заполнение формы обратной связи в установленном порядке.

В форме обратной связи, интегрированной в ЕРИС ЕМИАС, основными критериями оценки в случае несогласия врача является перечень технологических дефектов и диагностических неточностей (таблица 46), которые могут быть выявлены только специалистами лучевой диагностики.

Таблица 46 – Расшифровка пунктов в меню «Комментарии» поля обратной связи, заполняемой врачом при работе с результатами ИИ-сервиса в ЕРИС ЕМИАС

Наименование типа расхождения	Описание и примеры расхождения
01. Дефект (технологический)	Наличие технологического дефекта, который не позволяет использовать результаты ИИ-сервиса при описании результатов исследования: пустая/неполная серия с ИИ; в серии от ИИ-сервиса данные другого исследования; результаты представлены для серии, которая не подлежит обработке ИИ (толщина среза, проекция и т.п.); DICOM SR не соответствует серии ИИ в окне работы с изображениями; присутствуют несколько противоречащих друг другу результатов ИИ-сервиса
02. Грубое расхождение в работе сервиса	При отсутствии технологических дефектов: находки вне области интереса; находки, не соответствующие целям ИИ-сервиса
03. Неверный диагноз	Исходя из общего заключения сервиса по исследованию
04. Неверная локализация внутри целевого органа	Ошибка локализации (например, отмечен очаг в правом легком, а очаг в левом легком не размечен)
05. Неверная классификация находки	Ошибка классификации. Неверные результаты анализа (например, неправильный размер очагов или их характеристика)
06. Прочее (указать)	

Врач-рентгенолог оставляет обратную связь по работе ИИ-сервиса для отдельного исследования, выбирая «Согласен» или «Не согласен» с результатами работы ИИ-сервиса. В случае несогласия указывается причина («дефект», «грубая ошибка», «неверный диагноз», «неверная локализация внутри целого органа», «неверная классификация находки»).

2. Оценка технологической готовности.

Оценка технологической готовности компаний-разработчиков ИИ-сервисов, участвовавших в Эксперименте, проводилась сотрудниками Департамента информационных технологий города Москвы. Для проведения оценки использовалась шкала (таблица 47).

Таблица 47 – Шкала оценки технологической готовности

Вопрос	1. Оперативность реакции на замечания ДИТ по технологической интеграции	2. Результативность реакции на замечания ДИТ по технологической интеграции	3. Проактивность компании (предложения по оптимизации процессов схемы интеграции)	4. Оценка технической грамотности инженеров компании: DICOM, Apache Kafka	5. Проактивность компании по определению дефектов при работе в ППАК ЕРИС ЕМИАС
Шкала	0–2	0–2	0–2	0–2	0–2
«0» соответствует	Отсутствие реакции (нет ответа)	Замечания не устранены	Отсутствует	Требовались консультации по всем 3-м инструментам	Отсутствие обратной связи
«1» соответствует	Ответ в течение недели или дольше	Устранены с несколькими итерациями (после дополнительных замечаний или консультаций)	Низкая ценность полученных предложений	Требовались дополнительные консультации минимум по 1-му инструменту	Заявленные дефекты находились на стороне компании
«2» соответствует	Ответ в течение суток (до трех суток)	Устранены в полном объеме с 1-го раза	Высокая ценность полученных предложений	В достаточном объеме (по всем 3-м инструментам, освоены самостоятельно из технической литературы)	Заявленные дефекты относились к работе ЕРИС ЕМИАС

3. Хронометраж.

Исследование временных показателей в Эксперименте выполнено по двум задачам:

1. Оценить длительность обработки ИИ-сервисами результатов одного исследования.

Нормативным временем анализа одного исследования считается время, прошедшее с момента публикации сообщения о доступности исследования для скачивания и анализа сервисом в Единой системе уведомлений для внешних взаимодействий (далее – ЕСУ-ВВ) до момента публикации сервисом в ЕСУВВ сообщения о доступности результатов анализа исследований в ЕРИС ЕМИАС. По условиям Эксперимента 2020 года нормативное время

не должно превышать 10 минут. В качестве источника данных использована база данных ЕРИС ЕМИАС. Произведен расчет медианных значений нормативного времени анализа для каждого ИИ-сервиса и совокупно по модальностям.

2. Изучить влияние ИИ-сервисов на длительность описаний результатов лучевых исследований.

Определение длительности описаний результатов лучевых исследований выполнено на принципах аналитического (поэлементного) метода нормирования труда¹. Исследование длительности описаний врачами-рентгенологами результатов лучевых исследований выполнено в 2-х группах медицинских организаций (таблица 48):

1. Группа 1 – медицинские организации до внедрения ИИ-сервисов, в период 10.09.2019 – 13.07.2020, в том числе:

- группа 1.1 – амбулаторные МО,
- группа 1.2 – стационарные МО.

2. Группа 2 – медицинские организации после внедрения ИИ-сервисов, в период 01.05.2020 – 25.12.2020, в том числе:

- группа 2.1 – амбулаторные МО,
- группа 2.2 – стационарные МО.

Медицинские организации групп 1 и 2 были идентичны. Пересечение временных диапазонов связано с поэтапным внедрением ИИ-сервисов.

¹ Методика разработки норм времени и нагрузки медицинского персонала. М.: ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения», 2013. 25 с.

Таблица 48 – Структура групп для исследования длительности описаний результатов лучевых исследований

Группа	Модальность					Всего
	КТ РЛ	КТ Covid	ММГ	РГ	ФЛГ	
1	90859	43938	7372	28982	71896	243047
1.1	75138	41777	7049	26088	53801	203853
1.2	15721	2161	323	2894	18095	39194
2	39308	174885	35222	51787	216596	517798
2.1	26817	157690	34432	42492	201568	462999
2.2	12491	17195	790	9295	15028	54799
Всего	130167	218823	42594	80769	288492	760845

Способ получения информации для разработки норм времени: сбор информации по данным отчетности в информационной системе в сфере здравоохранения субъекта Российской Федерации. В качестве источника данных использована база данных ЕРИС ЕМИАС. Для хронометража применялась арифметическая разница между временем начала работы над заключением и временем первого подписания заключения. Если указанная разница превышала 40 минут, то данный случай из исследований исключали (подобные задержки связаны с выполнением описаний особо сложных клинических случаев, получением экспертных консультации внешних специалистов).

Рейтинг. Итоговая оценка ИИ-сервисов: с целью составления рейтинга ИИ-сервисов по совокупности данных разработана специальная методика.

Итоговый балл D складывается из параметров A , B , C с учетом индивидуального вклада, формула (4): A – характеризует соответствие минимальным функциональным требованиям (вклад в суммарный балл 10%); B – характеризует скорость обработки исследований (вклад в суммарный балл 5%); C – характеризует качество работы сервиса (вклад в суммарный балл 85%):

$$D = A * 0,1 + B * 0,05 + C * 0,85, (4)$$

А — соответствие минимальным функциональным требованиям (для текущей версии ИИ-сервиса):

— соответствует (реализована визуализация и описательная часть в полном объеме) = 100 баллов;

— частично соответствует (реализована в полном объеме только визуализация или только описательная часть, другая часть реализована частично) = 50 баллов;

— грубое несоответствие (визуализация и описательная часть не реализованы в должном объеме) = 0 баллов.

В — время ожидания результатов исследования (t , скорость обработки исследований): соответствует времени с момента отправки исследования ИИ-сервису до момента возвращения результата в ЕРИС ЕМИАС (считается как медианное время за последнюю неделю работы). Норматив — 10 минут. При расчете итогового рейтинга учитывалось среднее значение B в баллах за последний месяц работы ИИ-сервиса.

Перевод в баллы осуществляется по формуле (5):

$$(10-t) * 10, (5)$$

где 0 баллов при $t \geq 10$ минут, 100 баллов при $t \rightarrow 0$.

С — интегральная оценка диагностической ценности и качества, формула (6):

$$C = 0,3 * (C1 + C2) + 0,15 * (C3 + C4) + 0,1 * C5, (6)$$

где $C1$ — диагностическая точность (AUROC) по результатам калибровочных тестирований согласно базовым диагностическим требованиям (средняя, с коррекцией на дов. интервал последнего тестирования), рассчитывается как (7):

$$AUROC_{\text{средний}} - ДИ_AUROC/2, (7)$$

где $C2$ — диагностическая точность (AUROC) по результатам проспективного сравнения с текстами протоколов согласно базовым диагностическим требованиям (с коррекцией на дов. интервал), рассчитывается как (8):

$$(AUROC - ДИ_AUROC/2), (8)$$

C3 – удельный вес согласия врачей с локализацией находок.

C4 – удельный вес исследований без дефектов «в-е», рассчитывается как (9):

$(100 - \text{уд. вес дефектов}), (9)$

C5 – оценка пользователей-врачей (NPS), нормированная на максимальный NPS среди сервисов, рассчитывается как (10):

$(100 * (NPS + 100) / (NPS_{max} + 100)), (10)$

Методология Эксперимента как научного исследования

Методологически научное исследование выполнено *на принципах системного подхода*.

Для сбора данных использованы база данных ЕРИС ЕМИАС и специально разработанные процедуры.

В исследовании использованы следующие **методы научного познания**:

- аналитические (анализ, синтез)¹;
- оценки диагностической точности²;
- анализа результатов рентгенорадиологических исследований³;
- социологический опрос (опосредованный, выборочный, повторный, методом анкетирования)⁴;
- метод аналитического (позлементного) нормирования труда⁵;

¹ Долгушина Н. В., Воронов Д. А., Грачев С. В. Методология научных исследований в клинической медицине. М.: ГЭОТАР-Медиа, 2016. 112 с.

² Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика) / сост. С. П. Морозов, А. В. Владимировский, В. Г. Кляшторный / Сер. «Лучшие практики лучевой и инструментальной диагностики». М., 2019. Вып. 57. 51 с.

³ Оценка качества рентгенорадиологических исследований / сост. С. П. Морозов, Н. Н. Ветшева, Н. В. Ледихова [и др.] / Сер. «Лучшие практики лучевой и инструментальной диагностики». М., 2019. Вып. 22. 50 с.

⁴ Яковлева Н. Ф. Социологическое исследование [Электронный ресурс]: учеб. пособие. 2-е изд., стер. М.: ФЛИНТА, 2014. 250с.

- метод функционального тестирования ИИ-сервисов;
- метод калибровочного тестирования ИИ-сервисов;
- метод технологического мониторинга ИИ-сервисов;
- статистического анализа (описательная статистика, критерии Стьюдента, Манна-Уитни, Уилкоксона, коэффициент корреляции Пирсона, критерий Шапиро-Уилка (проверка гипотезы о нормальном распределении)); статистическая достоверность различий: вероятности случайного характера различий «р» — различия достоверны при $p < 0,05$; 95%-й доверительный интервал (ДИ) — результаты значимы только в пределах рассчитанного интервала.

При описании результатов исследования мы руководствовались методологиями **CONSORT-AI**, **STARD 2015**⁷ и **StaRI**⁸.

Исследование одобрено Независимым этическим комитетом МРО РОРР (протокол 2/2020 от 20.02.2020) (приложение 4). Методика проведения Эксперимента соответствует Совместному заявлению Европейского и Североамериканского обществ радиологов и др. об этике искусственного интеллекта в радиологии⁹. Для включения в исследование пациенты подписывали специальную форму информированного добровольного согласия (приложение

⁵ Методика разработки норм времени и нагрузки медицинского персонала. М.: ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения», 2013. 25 с.

⁶ Liu X., Rivera S.C., Moher D. et al. on behalf of the SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension // Lancet Digital Health. 2020. Vol. 2, №10. P. e537–e548.

⁷ Bossuyt P.M., Reitsma J.B., Bruns D.E. et al. For the STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies // Radiology. 2015. Vol. 277, №3. P. 826–832.

⁸ Pinnock H., Barwick M., Carpenter C. et al. Taylor SIC for the StaRI Group. Standards for Reporting Implementation Studies (StaRI) statement // BMJ. 2017. 356:i6795.

⁹ Geis J.R., Brady A., Wu C.C. et al. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement // Insights Imaging. 2019. Vol. 70, №4. P. 329–334. URL: <https://doi.org/10.1186/s13244-019-0785-8>.

5), для дополнительного информирования использовалась специально разработанная брошюра (приложение 6).

Для представления результатов Эксперимента на международном уровне (включая публикации в высокорейтинговых научных изданиях) данное научное исследование зарегистрировано в базе данных Clinical Trials с присвоением идентификационного номера ID Clinical Trials NCT04489992¹.

¹ URL: <https://www.clinicaltrials.gov/ct2/show/NCT04489992>.

Глава 8. ОЦЕНКА ДИАГНОСТИЧЕСКОЙ ТОЧНОСТИ СЕРВИСОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ

§1. Первый этап оценки

На данном этапе каждый ИИ-сервис, интегрированный в ТПАК ЕРИС ЕМИАС, проходил процедуру калибровочного тестирования на эталонных наборах данных. С точки зрения организации Эксперимента, установленные диагностические метрики сравнивались с заявленными (указанными в заявке на участие в Эксперименте): если отклонение превышало 10,0%, то ИИ-сервис не допускался к работе в ППАК ЕРИС ЕМИАС. Также в результате процедуры тестирования определялся индивидуальный порог срабатывания для каждого конкретного ИИ-сервиса. Данные о количестве и результатах процедуры как организационной составляющей Эксперимента представлены в главе 6.

С научной точки зрения, результаты этапа позволяли оценить точность работы конкретного ИИ-сервиса на реальных данных ЕРИС ЕМИАС, относящихся к московской популяции (приложение 7). Результаты оценки диагностической точности ИИ-сервисов представлены в таблицах 49, 50, 51.

Таблица 49 – Метрики диагностической точности ИИ-сервисов, установленные на первом этапе оценки

ИИ-сервис	AUC	Чувствительность	Специфичность
ИИ-сервис-1 (РГ)*	0,74	0,4	0,96
	0,86	0,83	0,88
ИИ-сервис-2 (РГ)	0,91	0,86	0,85
ИИ-сервис-3 (ММГ)	0,96	0,84	0,94
ИИ-сервис-4 (КТ Covid)	0,98	0,94	0,94
ИИ-сервис-5 (КТ Covid)	0,82	0,84	0,78
ИИ-сервис-6 (КТ Covid)	0,9	0,88	0,84
ИИ-сервис-7 (КТ РЛ)	0,91	0,84	0,92
ИИ-сервис-8 (РГ)	0,91	0,81	0,92
ИИ-сервис-9 (КТ Covid)	0,82	0,82	0,82
ИИ-сервис-10 (ФЛГ)	0,82	0,72	0,9
ИИ-сервис-10 (ММГ)	0,81	0,68	0,84
ИИ-сервис-11 (КТ Covid)	0,91	0,91	0,92
ИИ-сервис-12 (КТ Covid)	0,94	0,94	0,9
ИИ-сервис-13 (РГ) *	0,6	0,98	0,23
	0,92	0,86	0,86
ИИ-сервис-14 (КТ Covid) *	0,86	0,78	0,86
	0,97	0,9	0,94
ИИ-сервис-15 (НДКТ РЛ)	0,92	0,87	0,88
ИИ-сервис-16 (РГ)	0,8	0,9	0,69
ИИ-сервис-17 (ФЛГ)*	0,77	0,74	0,76
	0,9	0,8	0,92
* Представлены результаты первичного и повторного тестирований			

Таблица 50 – Средние значения метрик диагностической точности, установленных на первом этапе оценки (для всех ИИ-сервисов)

Параметр	Мин	Макс	Среднее значение	Медиана
AUC	0,80	0,98	0,89+0,06	0,91
Чувствительность	0,68	0,94	0,85+0,07	0,85
Специфичность	0,69	0,94	0,88+0,07	0,89

Таблица 51 – Средние значения метрик диагностической точности, установленных на первом этапе оценки (по модальностям), M+m

Модальность	AUC	Чувствительность	Специфичность
КТ/НДКТ РЛ	0,92+0,01	0,86+0,02	0,9+0,03
КТ Covid	0,91+0,07	0,89+0,05	0,88+0,06
ММГ	0,89+0,10	0,76+0,11	0,89+0,07
РГ	0,88+0,05	0,85+0,04	0,84+0,09
ФЛГ	0,86+0,06	0,76+0,06	0,91+0,02

С организационной точки зрения, полученные результаты соответствовали установленным требованиям точности, что позволило допустить ИИ-сервисы к работе в условиях реального потока исследований в ППАК ЕРИС ЕМИАС. С научной точки зрения, наиболее важен следующий факт. ИИ-сервисы отличаются хорошей воспроизводимостью результатов своей работы на новых эталонных данных. Это подтверждается тем, что отклонения значений основных диагностических метрик, полученные на первом этапе, от заявленных не превышали 10,0%. Совокупные средние значения метрик диагностической точности всех ИИ-сервисов колебались около 0,9.

В разрезе модальностей наиболее высокие и одновременно сбалансированные значения метрик были присущи ИИ-сервисам, работающим с результатами компьютерной томографии. С учетом специфики клинической задачи ИИ-сервисы для анализа результатов флюорографии отличались низкими значениями AUC и чув-

ствительности, но наиболее высокими значениями специфичности, что заслуживает положительной оценки.

52. Второй этап оценки

На данном этапе каждый ИИ-сервис работал с потоком исследований в ППАК ЕРИС ЕМИАС. Результаты анализа автоматически сопоставлялись с заключениями врачей-рентгенологов, непосредственно проводивших лучевые исследования в медицинских организациях. Результаты оценки диагностической точности ИИ-сервисов представлены в таблицах 52, 53, 54 и на рис. 42.

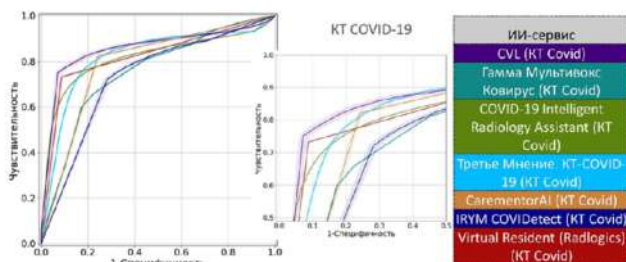


Рисунок 42 а – Характеристические кривые ИИ-сервисов (второй этап оценки диагностической точности)

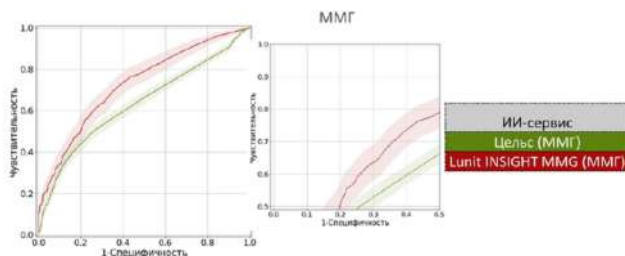


Рисунок 42 б – Характеристические кривые ИИ-сервисов (второй этап оценки диагностической точности)

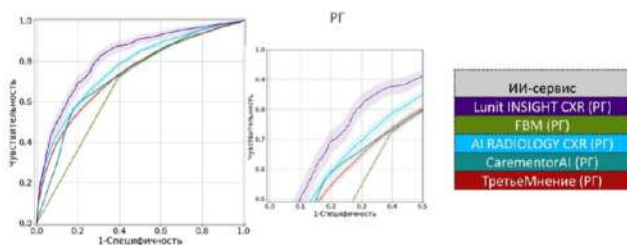


Рисунок 42 в – Характеристические кривые ИИ-сервисов (второй этап оценки диагностической точности)

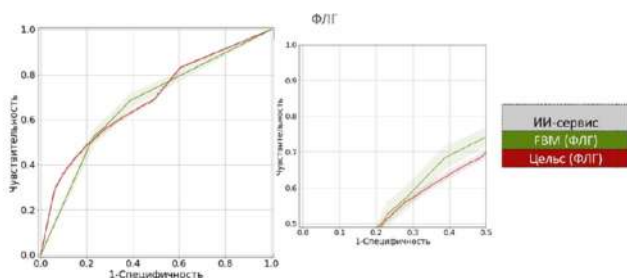


Рисунок 42 г – Характеристические кривые ИИ-сервисов (второй этап оценки диагностической точности)

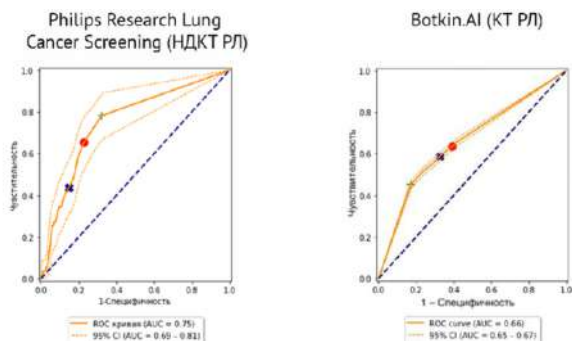


Рисунок 42 д– Характеристические кривые ИИ-сервисов (второй этап оценки диагностической точности)

Таблица 52 – Метрики диагностической точности ИИ-сервисов, установленные на втором этапе оценки

ИИ-сервис	AUC	Чувствительность	Специфичность
ИИ-сервис-1 (РГ)	0,77	0,56	0,83
ИИ-сервис-2 (РГ)	0,82	0,82	0,70
ИИ-сервис-3 (ММГ)	0,72	0,62	0,72
ИИ-сервис-4 (КТ Covid)	0,81	0,71	0,86
ИИ-сервис-5 (КТ Covid)	0,73	0,72	0,71
ИИ-сервис-6 (КТ Covid)	0,75	0,66	0,77
ИИ-сервис-7 (КТ РЛ)	0,66	0,45	0,83
ИИ-сервис-8 (РГ)	0,73	0,57	0,81
ИИ-сервис-9 (КТ Covid)	0,79	0,82	0,76
ИИ-сервис-10 (ФЛГ)	0,69	0,53	0,73
ИИ-сервис-11 (ММГ)	0,63	0,50	0,74
ИИ-сервис-12 (КТ Covid)	0,82	0,73	0,91
ИИ-сервис-13 (КТ Covid)	0,85	0,75	0,93
ИИ-сервис-14 (РГ)	0,74	0,60	0,76
ИИ-сервис-14 (КТ Covid)	0,82	0,78	0,80
ИИ-сервис-15 (НДКТ РЛ)	0,75	0,78	0,68
ИИ-сервис-16 (РГ)	0,68	0,70	0,62
ИИ-сервис-17 (ФЛГ)	0,67	0,63	0,62

Таблица 53 – Средние значения метрик диагностической точности, установленных на втором этапе оценки (для всех ИИ-сервисов)

Параметр	Мин	Макс	Среднее значение	Медиана
AUC	0,63	0,85	0,74±0,07	0,745
Чувствительность	0,45	0,82	0,66±0,11	0,68
Специфичность	0,62	0,93	0,76±0,09	0,76

Таблица 54 – Средние значения метрик диагностической точности, установленных на втором этапе оценки (по модальностям), M+m

Модальность	AUC	Чувствительность	Специфичность
КТ/НДКТ РЛ	0,71±0,06	0,61±0,23	0,75±0,11
КТ Covid	0,80±0,05	0,74±0,01	0,82±0,11
ММГ	0,68±0,06	0,56±0,08	0,73±0,01
РГ	0,75±0,02	0,65±0,11	0,74±0,09
ФЛГ	0,68±0,04	0,58±0,07	0,67±0,08

Оценка диагностической точности на данном этапе представляет наибольшее научное значение, так как позволяет получить реальные характеристики ИИ-сервисов в условиях работы с потоком результатов лучевых исследований в режиме реального времени.

Специфическая ситуация 2020 года (пандемия новой коронавирусной инфекции, временная приостановка массовых профилактических исследований) сказалась на структуре потока исследований, направляемых для анализа ИИ-сервисами. Тем не менее были получены первичные данные, вполне достаточные для научного анализа.

При применении в реальных клинических условиях у всех ИИ-сервисов отмечается выраженное снижение значений метрик диагностической точности. Только 27,8% (5) ИИ-сервисов сохранили достаточный уровень воспроизводимости результатов своей работы ($AUC > 0,81$ в соответствии с нормативными требованиями Эксперимента). Большая часть ИИ-сервисов – 44,4% (8) – «удержали» площадь под характеристической кривой в пределах 0,7–0,8. Остальные «провалились» ниже 0,69.

В разрезе модальностей наиболее стабильные и относительно высокие значения метрик показали ИИ-сервисы для анализа результатов компьютерной томографии и рентгенографии. К критично низким значениям приблизились ИИ-сервисы для маммографии и флюорографии.

В целом, полученные результаты свидетельствовали о необходимости дальнейшей доработки ИИ-сервисов, возможном потенциальном дообучении на наборах данных московской популяции.

§3. Сравнение результатов первого и второго этапов оценки

Представляет интерес сравнение полученных на первом и втором этапах данных, в том числе с заявленными разработчиками (таблица 55, 56, 57). Фактически Эксперимент представляет собой масштабную независимую валидацию ИИ-сервисов на новых данных, не использованных для обучения алгоритмов. Наличие независимой валидации (тем более в формате мультицентрового клинического испытания) является обязательным для успешного развития технологий искусственного интеллекта¹.

На первом этапе для 33,3% (6) ИИ-сервисов зафиксировано незначительное, но все же снижение значений АУС относительно заявленных. Вместе с тем подавляющее большинство ИИ-сервисов 66,7% (12) — продемонстрировало некоторое повышение; при этом 5 ИИ-сервисов улучшили свой показатель более, чем на 0,1.

На втором этапе только 11,1% (2) сервисов сохранили уровень АУС на заявленном уровне.

Большинство ИИ-сервисов — 88,9% (16) показали выраженное ухудшение, причем у 66,7% (12) это отклонение было значимым и составило 0,1 или более.

В целом различия между значениями АУС (заявленными и полученными на I и II этапах) были статистически достоверны. В обоих случаях U-критерий Манна-Уитни при заданной численности сравниваемых групп составлял 99 ($p < 0,05$).

Также отрицательная динамика зафиксирована при сравнении результатов первого и второго этапов — 100,0% ИИ-сервисов продемонстрировали снижение площади под характеристической кривой. Менее 0,1 отклонения зафиксированы у 33,3% (6) ИИ-сервисов. Максимальным ($> 0,2$) было отклонение у 3 ИИ-сервисов по ФЛГ, ММГ и КТ РЛ; различия были статистически достоверны (U-критерий Манна-Уитни = 99, $p < 0,05$).

¹ Baldwin D.R., Gustafson J., Pickup L., Arteta C. et al. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules // Thorax. 2020. Vol. 75, №4. P. 306–312. DOI: 10.1136/thoraxjnl-2019-214104.

Таблица 55 – Сравнение средних значений параметра «площадь под характеристической кривой (AUC)» на этапах оценки диагностической точности (по ИИ-сервисам)

ИИ-сервис	По данным разработчика	Первый этап	Второй этап
ИИ-сервис-1 (РГ)	0,907	0,86	0,77
ИИ-сервис-2 (РГ)	0,92	0,91	0,82
ИИ-сервис-3 (ММГ)	0,90	0,96	0,72
ИИ-сервис-4 (КТ Covid)	0,94	0,98	0,81
ИИ-сервис-5 (КТ Covid)	0,87	0,82	0,73
ИИ-сервис-6 (КТ Covid)	>0,81	0,9	0,75
ИИ-сервис-7 (КТ РЛ)	0,94	0,91	0,66
ИИ-сервис-8 (РГ)	>0,81	0,91	0,73
ИИ-сервис-9 (КТ Covid)	0,85	0,82	0,79
ИИ-сервис-10 (ФЛГ)	0,87	0,82	0,69
ИИ-сервис-11 (ММГ)	>0,81	0,81	0,63
ИИ-сервис-12 (КТ Covid)	>0,81	0,91	0,82
ИИ-сервис-13 (КТ Covid)	0,82	0,94	0,85
ИИ-сервис-14 (РГ)	0,78	0,92	0,74
ИИ-сервис-15 (КТ Covid)	0,95	0,97	0,82
ИИ-сервис-16 (НДКТ РЛ)	0,85	0,92	0,75
ИИ-сервис-17 (РГ)	0,79	0,8	0,68
ИИ-сервис-18 (ФЛГ)	0,79	0,9	0,67

Таблица 56 – Различия (дельта) средних значений параметра «площадь под характеристической кривой (AUC)» на этапах оценки диагностической точности (по ИИ-сервисам)

ИИ-сервис	Заявленные	Первый этап	Второй этап
ИИ-сервис-1 (РГ)	-0,047	-0,137	-0,09
ИИ-сервис-2 (РГ)	-0,01	-0,1	-0,09
ИИ-сервис-3 (ММГ)	0,06	-0,18	-0,24
ИИ-сервис-4 (КТ Covid)	0,04	-0,13	-0,17
ИИ-сервис-5 (КТ Covid)	-0,05	-0,14	-0,09
ИИ-сервис-6 (КТ Covid)	0,09	-0,06	-0,15
ИИ-сервис-7 (КТ РЛ)	-0,03	-0,28	-0,25
ИИ-сервис-8 (РГ)	0,1	-0,08	-0,18
ИИ-сервис-9 (КТ Covid)	-0,03	-0,06	-0,03
ИИ-сервис-10 (ФЛГ)	-0,05	-0,18	-0,13
ИИ-сервис-11 (ММГ)	0	-0,18	-0,18
ИИ-сервис-12 (КТ Covid)	0,1	0,01	-0,09
ИИ-сервис-13 (КТ Covid)	0,12	0,03	-0,09
ИИ-сервис-14 (РГ)	0,14	-0,04	-0,18
ИИ-сервис-15 (КТ Covid)	0,02	-0,13	-0,15
ИИ-сервис-16 (НДКТ РЛ)	0,07	-0,1	-0,17
ИИ-сервис-17 (РГ)	0,01	-0,11	-0,12
ИИ-сервис-18 (ФЛГ)	0,11	-0,12	-0,23

Таблица 57 – Сравнение средних значений параметра «площадь под характеристической кривой (AUC)» на этапах оценки диагностической точности (по модальностям)

Модальность	Первый этап	Второй этап
КТ/НДКТ РЛ	0,92±0,01	0,71±0,06
КТ Covid	0,91±0,07	0,80±0,05
ММГ	0,89±0,10	0,68±0,06
РГ	0,88±0,05	0,75±0,02
ФЛГ	0,86±0,06	0,68±0,04

Вместе с тем, необходимо четко обозначить **ограничения** данного этапа исследования:

1. Наборы данных для первого этапа — нами применялись наборы с искусственным балансом классов, что может вызвать дискуссию с точки зрения значимости влияния на результаты оценки точности.

2. Дизайн второго этапа — в качестве «ground truth» использовалось только одно мнение врача, непосредственно описывающего результаты исследования, не верифицированное иными способами (экспертным консенсусом, аудитом и т.д.); не было механизмов анализа содержания протоколов для исключения описаний, содержащих предубеждения и дефекты.

3. Длительность проспективного участия ИИ-сервисов в Эксперименте была различна, что сказывается на сравнительном анализе.

Исходя из сказанного, полученные результаты надо интерпретировать как предварительные, требующие дальнейших скрупулезных исследований.

54. Третий этап оценки

На данном этапе часть исследований, обработанных каждым ИИ-сервисом, направлялась на аудит врачами-экспертами по установленным правилам. Референсным считалось решение по результатам аудита, включая консенсус двух экспертов при необходимости. С учетом существенных различий в объемах исследований, направляемых на аудит от разных ИИ-сервисов, расчет метрик AUC, чувствительности и специфичности невозможен. Поэтому на данном этапе определялся коэффициент согласованности классификаций решений в парах: «врач-рентгенолог» — «врач-эксперт» и «ИИ-сервис» — «врач-эксперт». Результаты оценки представлены в таблицах 58, 59, 60.

Полученные результаты свидетельствуют о высоком уровне согласованности решений экспертов (аудиторов) и врачей-рентгенологов, непосредственно проводивших и делавших описания исследований; значения коэффициента каппа Коэна стремятся к 1,0.

Вместе с тем наблюдается крайне низкая конкордантность решений аудиторов и ИИ-сервисов; значения коэффициента колеблются от отрицательных до минимально положительных. Единственное исключение составляет один ИИ-сервис по ФЛГ, для которого коэффициент согласованности достигает 0,7 (95% ДИ 0,5; 0,91).

В разрезе модальностей слабый уровень согласованности решений эксперта и ИИ-сервисов отмечен только для флюорографии — $0,45 \pm 0,09$.

Основной причиной направления результатов исследований на аудит было расхождение мнения врача-рентгенолога с ИИ-сервисом. Низкая согласованность решений экспертов с решениями сервисов подтверждает правоту врачей-рентгенологов и ошибочность суждений ИИ в указанных случаях.

Таблица 58 – Метрики диагностической точности ИИ-сервисов, установленные на третьем этапе оценки

ИИ-Сервис	Согласованность для ИИ-сервиса		Согласованность для врача-рентгенолога	
	Каппа Козна	95% ДИ	Каппа Козна	95% ДИ
ИИ-сервис-1 (РГ)	0*	(0; 0)	-	-
ИИ-сервис-2 (РГ)	-0,19	(-0,7; 0,32)	1,0	(1,0; 1,0)
ИИ-сервис-3 (ММГ)	-0,8	(0,8; 0,8)	1,0	(1,0; 1,0)
ИИ-сервис-4 (КТ Covid)	-0,22	(-0,48; 0,03)	0,77	(0,6; 0,94)
ИИ-сервис-5 (КТ Covid)	0,27	(-0,25; 0,79)	1,0	(1,0; 1,0)
ИИ-сервис-6 (КТ Covid)	0,01	(-0,18; 0,21)	1,0	(1,0; 1,0)
ИИ-сервис-7 (КТ РЛ)	0,17	(-0,27; 0,6)	1,0	(1,0; 1,0)
ИИ-сервис-8 (РГ)	0,02	(-0,22; 0,27)	1,0	(1,0; 1,0)
ИИ-сервис-9 (КТ Covid)	-0,06	(-0,56; 0,44)	1,0	(1,0; 1,0)
ИИ-сервис-10 (ФЛГ)	0,7	(0,5; 0,91)	1,0	(1,0; 1,0)
ИИ-сервис-11 (ММГ)	-0,08	(-0,28; 0,12)	0,9	(0,79; 1,0)
ИИ-сервис-12 (КТ Covid)	-	-	-	-
ИИ-сервис-13 (КТ Covid)	-0,4	(-0,89; 0,08)	1,0	(0,79; 1,0)
ИИ-сервис-14 (РГ)	-	-	-	-
ИИ-сервис-15 (КТ Covid)	-0,59	(-0,86; 0,31)	1,0	(1,0; 1,0)
ИИ-сервис-16 (НДКТ РЛ)	0	-	-	-
ИИ-сервис-17 (РГ)	-0,22	(-0,36; -0,08)	1,0	(1,0; 1,0)
ИИ-сервис-18 (ФЛГ)	0**	-	-	-
* ИИ-сервис отнес все исследования к патологии, врач и врач-эксперт (аудитор) указывают на то, что все 18 исследований содержат норму.				
** Во всех случаях норма				

Таблица 59 – Средние значения метрик диагностической точности, установленных на третьем этапе оценки (для всех ИИ-сервисов)

Параметр	Мин	Макс	Среднее значение	Медиана
Согласованность для ИИ-сервиса	-0,8	0,7	-0,095±0,367	-0,04
Согласованность для врача-рентгенолога	0,77	1	0,97±0,07	1

Таблица 60 – Средние значения метрик диагностической точности, установленных на третьем этапе оценки (по модальностям), каппа Козна, M+m

Модальность	Согласованность для ИИ-сервиса	Согласованность для врача-рентгенолога
КТ/НДКТ РЛ	-0,44±0,51	0,95±0,07
КТ Covid	-0,17±0,06	0,94±0,02
ММГ	0,08±0,12	1,0
РГ	-0,1±0,06	1,0
ФЛГ	0,45±0,09	1,0

Важно отметить, что полученные данные подтверждают результаты второго этапа оценки диагностической точности в качестве референсного метода, на котором использованы именно решения врачей-рентгенологов.

Результаты третьего этапа позволяют сделать выводы методологического характера:

1. Результаты работы технологий искусственного интеллекта в лучевой диагностике должны в обязательном порядке верифицироваться врачом (специалистом по лучевой диагностике).

2. В перспективе – при масштабном внедрении и рутинном применении технологий искусственного интеллекта необходимо предусмотреть функциональные возможности медицинских (радиологических) информационных систем для автоматизированной отправки результатов исследований на независимый контроль качества (аудит) при наличии расхождений решений врача и ИИ.

Таким образом, по результатам трехэтапной оценки диагностической точности ИИ-сервисов нулевая гипотеза H_0 : {Чувствительность $< 0,81$ и/или Специфичность $< 0,81$ и/или AUC $< 0,81$ } для условий:

- калибровочного тестирования — отклонена,
- работы с потоком исследований отклонена для модальности KT Covid, для остальных — принята.

Глава 9. ОЦЕНКА ЦЕЛЕСООБРАЗНОСТИ ПРИМЕНЕНИЯ ИИ- СЕРВИСОВ В ЛУЧЕВОЙ ДИАГНОСТИКЕ

В соответствии с существующей методологией¹ целесообразность применения ИИ-сервисов в лучевой диагностике изучена по нескольким основным аспектам: приемлемость, спрос, реализация, осуществимость, интеграция, способность к адаптации. Для этого параллельно проводились несколько научных исследований с различными методиками.

§1. Социологический опрос пользователей (врачей-рентгенологов) ИИ-сервисов

Первый этап анкетирования. Результаты оценки качества обучения врачей-рентгенологов вопросам применения ИИ-сервисов в рамках Эксперимента приведены в главе 6.

Второй и третий этапы анкетирования. На втором этапе получено 125 заполненных анкет, на третьем — 229. Этапность опроса позволила оценить субъективные ожидания врачей от использования ИИ-сервисов и объективную оценку работы с ними. Всего было получено 359 заполненных анкет (что составило 67,0% от числа врачей, работающих с ИИ-сервисами). До начала активной фазы Эксперимента на все ключевые вопросы, касающиеся

¹ Bowen D.J., Kreuter M., Spring B. et al. How we design feasibility studies // Am J Prev Med. 2009. Vol. 36, №5. P. 452–457. DOI:10.1016/j.amepre.2009.02.002.

ожиданий от применения технологий искусственного интеллекта, мнения респондентов почти всегда разделялись на примерно равные доли.

На вопрос, поможет ли ИИ сократить количество пропущенных клинически значимых патологий, 46,0% московских врачей ответили положительно, 36,0% — нейтрально, 18,0% — отрицательно. Позволит ли ИИ больше сосредоточиться на проблемных областях изображений: 43,0%, 38,0%, 19,0% соответственно.

Решит ли ИИ проблему увеличивающейся рабочей нагрузки: 35,0% положительных ответов, 38,0% нейтральных, 26,0% отрицательных. На вопрос «Ожидаете ли Вы, что ИИ-сервис сократит время на написание заключений» утвердительно («сократит») ответили 33,0% респондентов, «ничего не изменится» — 35,0%, «увеличит» — 32,0%. Примечательно, что, по данным масштабного международного опроса врачей лучевой диагностики, наблюдается немного иная картина: 50,8% опрошенных ожидают снижения длительности, а 49,2% — наоборот увеличения длительности описаний¹. Следовательно, ожидания московских врачей более нейтральны, в то время как ожидания почти половины международной аудитории менее оптимистичны.

Интересны опасения, возникающие у профессиональной аудитории в связи с внедрением технологий ИИ. Применение искусственного интеллекта может увеличить количество ложноположительных находок — так считают 32,0% врачей; увеличится время работы врача — 23,0%, есть риск пропуска значимой патологии — 19,0%, будут трудности по использованию ИИ в рутинной работе — 13,0%. Все же 13,0% врачей считают, что профессия рентгенолога находится под угрозой в будущем.

Полученные ответы можно интерпретировать как недостаточную информированность врачей о реальных возможностях и ограничениях ИИ, а также недостаток собственного практического опыта. Впрочем, это полностью объяснимо — технологии искусственного интеллекта в медицине находятся в самом начале пути своего развития. На международном уровне наблюдается та-

¹ Pesapane F., Codari M., Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine // Eur Radiol Exp. 2018. Oct. 24. Vol. 2, №1. P. 35. DOI: 10.1186/s41747-018-0061-6.

кая же картина: 68,6% врачей хотели бы пройти дополнительное углубленное обучение по вопросам возможностей и ограничений ИИ, а 58,1% — по практическим аспектам его применения. Примечательно, что 11,1% опрошенных хотели бы узнать «как выжить во время ИИ-революции», лишь 0,9% международной аудитории хотели бы полностью избежать применения ИИ¹.

В целом, после второго этапа анкетирования отношение к технологиям ИИ московских врачей-рентгенологов можно охарактеризовать как выжидательное, умеренно-оптимистичное. После нескольких месяцев активного использования ИИ-сервисов мнения и оценки врачей-рентгенологов изменились (рис. 43).

Возможности технологий ИИ сократить количество пропущенных клинически значимых патологий «плохо» охарактеризовали 41,0% опрошенных, «хорошо» — 25,0%. Здесь отмечается отрицательная динамика (на предыдущем этапе аналогичные значения составляли 18,0% и 46,0% соответственно). Оценки способности ИИ помочь врачу больше сосредоточиться на проблемных областях изображений практически не изменились (небольшое увеличение доли нейтральных оценок). Зато снизились отрицательные оценки возможностей ИИ-сервисов по решению проблемы постоянно увеличивающейся трудовой нагрузки на врачей. Теперь оценку «плохо» дали 16,0% респондентов (на 10% меньше, чем на предыдущем этапе). Большая часть ответов здесь «сместилась» в нейтральную сторону. Реальное снижение длительности подготовки описаний благодаря ИИ-сервисам указали только 20,0% респондентов, здесь произошло снижение на 12,0%. Снизилось количество отрицательных оценок, ощутимо возросло число нейтральных: с 35,0% до 61,0%. Примечательно, что существенно изменилась структура опасений, возникающих у врачей из-за внедрения технологий ИИ. Полностью исчезло мнение о наличии риска пропуска значимой патологии (на предыдущем этапе — 19,0%). На 8,0% возросло число утверждающих, что ИИ создает риски увеличения числа ложноположительных находок (с 32,0% до 40,0%). Больше число врачей-рент-

¹ Pesapane F., Codari M., Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine // Eur Radiol Exp. 2018. Oct. 24. Vol. 2, №1. P. 35. DOI: 10.1186/s41747-018-0061-6.

генологов стали утверждать, что их профессия находится под угрозой: 20,0% против 13,0% на предыдущем этапе.



КАКИЕ ОПАСЕНИЯ ОТ ВНЕДРЕНИЯ ИИ У ВАС ВОЗНИКАЮТ?



ПЕРВЫЙ ЭТАП ОПРОСА



- ИИ решает проблему увеличивающейся нагрузки на врачей
- ИИ позволяет больше сосредоточиться на значимых областях изображений исследований
- ИИ позволяет сократить количество пропусков клинически значимых находок

ВТОРОЙ ЭТАП ОПРОСА



- ИИ решает проблему увеличивающейся нагрузки на врачей
- ИИ позволяет больше сосредоточиться на значимых областях изображений исследований
- ИИ позволяет сократить количество пропусков клинически значимых находок

Рисунок 43 – Результаты анкетирования врачей-рентгенологов для оценки ожиданий и отношения к технологиям искусственного интеллекта (периоды опросов: 10.2020 и 01.2021)

Важно отметить, что все отличия в ответах на втором и третьем этапах тестирования не имели статистической значимости, за исключением одного. Отрицательная динамика ответов на утверждение «ИИ помогает уменьшить количество клинически значимых расхождений» имеет статистическую значимость: критерий Манна-Уитни 12194,0 $p=0,0204$.

Таким образом, в динамике субъективных оценок отмечены следующие изменения:

- не оправдались ожидания снижения длительности описаний результатов лучевых исследований;
- вместе с тем не оправдались и опасения относительно увеличения такого времени;
- увеличились опасения генерации искусственным интеллектом значительного числа ложноположительных находок;
- возросло число врачей-рентгенологов, опасющихся за свою профессию в будущем.

По субъективным оценкам врачей, ИИ-сервисы не показали эффективности относительно создания условий для более сосредоточенного анализа важных находок. Также сервисы не помогли намного уменьшить количество клинически значимых расхождений.

Однако в целом заметно вырос уровень нейтральных оценок, причем за счет отрицательных. Можно констатировать, что за время активной фазы Эксперимента доля негативно настроенных к технологиям ИИ врачей-рентгенологов стала меньше, а положительно и нейтрально настроенных — больше.

По результатам этого этапа социологического опроса отношение к технологиям ИИ врачей-рентгенологов можно охарактеризовать как выжидательное, умеренное. Конечно, данную динамику можно расценивать как отрицательную, но сугубо арифметический подход мало применим в этой ситуации.

Многomesячный практический опыт использования ИИ-сервисов в повседневной работе позволил врачам взглянуть на проблему более прагматично. В их ответах мы видим запрос на научные данные о работе искусственного интеллекта, подготовленные на принципах доказательной медицины. Ряд суждений респондентов обусловлен нестабильностью работы ИИ-сервисов, в том числе недостаточным уровнем диагностической точности. Оценка производительности труда носила субъективный характер, соответственно требуется ее проверка объективными методами.

Четвертый этап анкетирования. На данном этапе изучены оценки врачей-рентгенологов по тем ИИ-сервисам, с которыми они работали непосредственно. Всего было получено 320 заполненных анкет (что составило 67,0% от числа врачей, работающих с ИИ-сервисами).

Для обобщенного представления подсчитаны средние оценки (см. таблицу 61). Обращает на себя внимание нейтральный подход врачей к оценкам: по всем пунктам выбраны средние оценки. На основе заполненных анкет для каждого ИИ-сервиса рассчитан индекс потребительской лояльности (NPS) (рис. 44).

Пятый этап анкетирования. Результаты не представляют научной ценности и использованы для улучшения производственных процессов и организационно-методического обеспечения Эксперимента. Таким образом, результаты социологического исследования врачей-рентгенологов позволяют сформулировать следующие утверждения.

1. Общее отношение врачей к технологиям ИИ выжидательное, умеренное, нейтральное.

2. Требуется развитие доказательного научного подхода к внедрению технологий искусственного интеллекта в лучевой диагностике, систематизация накапливаемого практического опыта с активными научными публикациями.

3. Актуальна системная постоянная работа по дополнительному профессиональному обучению медицинских работников вопросам цифровизации, информатизации и автоматизации лучевой диагностики.

4. Необходимо дальнейшее изучение вопросов целеполагания, стандартизации, качества технологий компьютерного зрения, а также — обязательного внедрения процедур технических и клинических испытаний технологий искусственного интеллекта как этапа получения статуса медицинского изделия.

Таблица 61 – Обобщенные результаты анкетирования для оценки качества используемых ИИ-сервисов (M+m)

Критерий	Средняя оценка
ИИ решает/способен решить проблему растущей информационной и трудовой нагрузки в Вашей работе (0 – совершенно не способен, 10 – решит существенно)	5+2,6
ИИ позволяет Вам больше сосредоточиться на значимых областях изображений исследований (0 – ничего не меняется, 10 – дает возможность сосредоточиться)	5+2,7
ИИ помогает уменьшить количество клинически значимых расхождений (0 – нет, количество клинически значимых расхождений меньше не стало; 10 – да, количество клинически значимых расхождений существенно сокращается)	4,2+2,8
Оцените удобство использования ИИ-сервиса (0 – совершенно неудобно, 10 – очень удобно)	5,6+2,4
Оцените уровень диагностической точности ИИ-сервиса (0 – абсолютно неприемлемое качество, 10 – отличное качество)	5,2+2,3
ИИ-сервис привел к увеличению времени подготовки заключения (0 – существенно увеличилось, 10 – несущественно увеличилось)	4,1+2,1
Оцените полезность приоритизации исследований в рабочем списке (0 – крайне низкая, 10 – очень высокая)	5,1+2,6
Насколько вам нравится отображение локализации находок на изображениях (0 – абсолютно не нравится, 10 – очень нравится)	5,6+2,5
Оцените точность локализации находок (0 – крайне низкая точность, 10 – точность очень высокая)	5,2+2,4
Оцените качество описания, созданного ИИ-сервисом (0 – очень плохое качество, 10 – прекрасное качество)	5,1+2,3
Если ИИ-сервис оценивает динамику, то оцените удобство реализации функциональности, связанной с оценкой в динамике (0 – абсолютно неудобно, 10 – очень удобно)	4,0+2,7
С какой вероятностью Вы бы порекомендовали данный ИИ-сервис своим коллегам, врачам-рентгенологам? (0 – ни в коем случае не буду рекомендовать, 10 – обязательно буду рекомендовать)	4,9+2,5

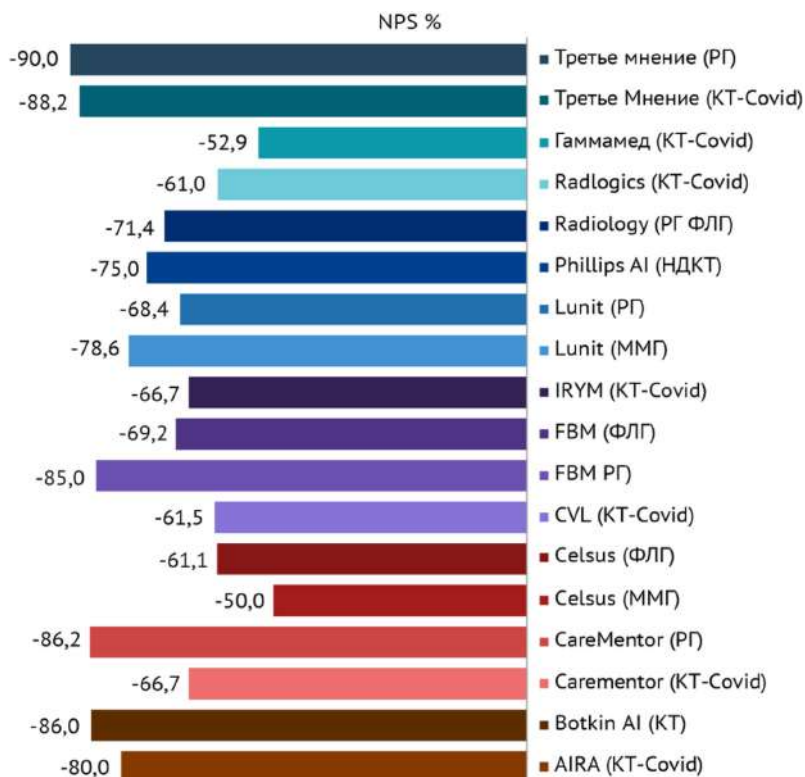


Рисунок 44 – Индекс потребительской лояльности для ИИ-сервисов

§2. Результаты технологического мониторинга

Технологический мониторинг проводился по оригинальной процедуре (см. главу 6) как обязательная организационно-методическая процедура Эксперимента, предусмотренная нормативно-правовой документацией. По результатам мониторинга принимались решения о продолжении участия в Эксперименте каждого конкретного ИИ-сервиса, об объемах выплат грантовых средств. С научной точки зрения, процедура технологического мониторинга послужила методикой оценки качества и надежности ИИ-сервисов

как информационной технологии – прикладного программного продукта.

За весь период Эксперимента процедура технологического мониторинга проводилась 161 раз. На разных этапах у ИИ-сервисов фиксировались принципиально разные уровни дефектов каждой из групп (таблица 62). Если опираться на средние значения, то наиболее частой проблемой было превышение предусмотренного нормативного времени обработки одного исследования – 21,3%. Указанный норматив, по условиям Эксперимента в 2020 г., составлял 10 минут. Достаточно часто представленные результаты работы ИИ-сервиса искаженными, ошибочными, неполными данными, целостность их была нарушена вплоть до негативного влияния на безопасность пациента и работу врача-рентгенолога. Средний уровень таких дефектов составлял 17,1%. В среднем в 5,0% ситуаций исследования были переданы для обработки в ИИ-сервис, но не вернулись в ЕРИС ЕМИАС с результатами обработки.

Таблица 62 – Удельный вес технологических дефектов ИИ-сервисов

Группа дефектов	Удельный вес (%)		
	Среднее значение	95% ДИ	Диапазон значений
Группа «а» – превышение норматива времени обработки исследования	21,3	(17,2; 25,5)	0,0 – 93,0
Группа «б» – отсутствие результатов работы ИИ-сервиса	5,0	(3,4; 6,6)	0,0 – 89,2
Группа «б*– г» / «в–е» – дефекты, связанные с представлением результатов исследования	17,1	(13,2; 20,1)	0,0 – 100,0
* С 13.10.2020 (приказ ДЗМ от 13.10.2020 г. №1199).			

Удельный вес технологических дефектов значительно колебался в ходе Эксперимента. Подавляющее большинство компаний-участников занимало активную позицию, оперативно и скрупулезно работало над постоянным улучшением своих программных решений. Поэтому представляет интерес оценка каждого ИИ-сервиса в динамике.

На рис. 45 (а-в) представлена динамика технологических дефектов ИИ-сервисов, осуществляющих выявление признаков новой коронавирусной инфекции на результатах компьютерной томографии органов грудной клетки.

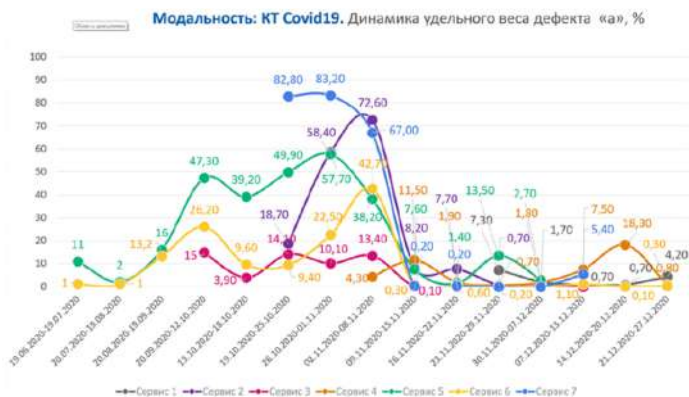


Рисунок 45 а – Динамика технологических дефектов ИИ-сервисов модальности КТ Covid-19. Знак * означает, что дефект соответствует группе дефектов «б-г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)



Рисунок 45 б – Динамика технологических дефектов ИИ-сервисов модальности КТ Covid-19. Знак * означает, что дефект соответствует группе дефектов «б-г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)

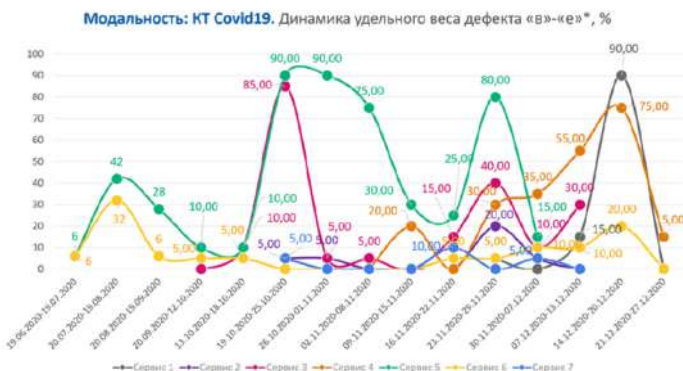


Рисунок 45 в – Динамика технологических дефектов ИИ-сервисов модальности КТ Covid-19. Знак * означает, что дефект соответствует группе дефектов «б–г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)

Наиболее показательна ситуация с дефектом «а» – превышением нормативного времени обработки исследования. В начальном периоде отмечаются значительные колебания, ИИ-сервисы работают нестабильно. Затем, по мере стандартизации принципов работы, внедрения требований по клиническому контексту их применения (в частности, классификации тяжести поражения легочной ткани «КТ0–4»), ситуация стабилизируется. Уровень дефектов снижается, в большинстве случаев укладывается в допустимые условиями Эксперимента пределы. Аналогична ситуация для дефектов группы «б»: совершенствование собственных программных решений обеспечило стабилизацию работы ИИ-сервисов. Частота отсутствия результатов обработанных исследований в ППАК ЕРИС ЕМИ-АС минимизировалась. К сожалению, для дефектов групп «в–е», связанных с искажением и нарушением целостности данных, ситуация в основном была негативной. В целом отмечались значительные колебания.

По уровню технологических дефектов показательна ситуация с первым ИИ-сервисом для выявления на КТ признаков новой коронавирусной инфекции, принявшим участие в Эксперименте. В ходе работы он продемонстрировал положительную динамику по наращиванию скорости обработки исследований, почти что из-

бежал «пропусков» исследований. Однако относительно часто искажал и нарушал целостность данных, причем здесь отмечалась явная нестабильность работы программного обеспечения – выраженные колебания уровня дефектов групп «в–е».

Другой сервис продемонстрировал положительную динамику в достижении установленного временного норматива обработки одного исследования; был стабилен по минимальному уровню дефектов групп «в–е». Однако заметны периодические проблемы с потерей данных – выраженные «всплески» отсутствия результатов обработанных исследований в ППАК ЕРИС ЕМИАС.

На рис. 46 (а-в) представлена динамика технологических дефектов ИИ-сервисов, также работавших с результатами КТ или низкодозной КТ органов грудной клетки, но для выявления признаков злокачественных новообразований.



Рисунок 46 а – Динамика технологических дефектов ИИ-сервисов модальностей КТ РЛ и НДКТ РЛ. Знак * означает, что дефект соответствует группе дефектов «б–г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)



Рисунок 46 б – Динамика технологических дефектов ИИ-сервисов модальностей КТ РЛ и НДКТ РЛ. Знак * означает, что дефект соответствует группе дефектов «б-г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)



Рисунок 46 в – Динамика технологических дефектов ИИ-сервисов модальностей КТ РЛ и НДКТ РЛ. Знак * означает, что дефект соответствует группе дефектов «б-г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)

По данной модальности работали только два ИИ-сервиса. Оба

продемонстрировали достаточно стабильную работу с позиций отсутствия потерь и искажений данных (дефекты групп «б–е»), отклонения от установленных допустимых значений встречались спорадически. Однако оба сервиса отличались неприемлемо высоким уровнем превышений нормативного времени обработки исследования (дефект группы «а»). Отчасти это можно объяснить недостаточными вычислительными мощностями ИИ-сервиса. Такие значительные задержки в предоставлении результатов автоматизированного анализа в реальной практике делают ИИ-сервис полностью непригодным для врача-рентгенолога. Технология значительно «опаздывает», в итоге врач формирует заключение, не дожидаясь результатов работы искусственного интеллекта.

На рис. 47 (а-в) представлена динамика технологических дефектов ИИ-сервисов, работавших с результатами профилактической маммографии.



Рисунок 47 а – Динамика технологических дефектов ИИ-сервисов модальности ММГ. Знак * означает, что дефект соответствует группе дефектов «б–г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)



Рисунок 47 б – Динамика технологических дефектов ИИ-сервисов модальности ММГ. Знак * означает, что дефект соответствует группе дефектов «б–г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)



Рисунок 47 в – Динамика технологических дефектов ИИ-сервисов модальности ММГ. Знак * означает, что дефект соответствует группе дефектов «б–г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)

В работе одного из ИИ-сервисов по ММГ отмечена несколько

парадоксальная картина. В начале работы в Эксперименте сервис не справляется с установленной длительностью обработки исследований, стремительно нарастает число потерь данных. Однако на этом фоне искажения и нарушения целостности данных минимальны или даже отсутствуют.

Примерно за 2 месяца ситуация по дефектам «а» и «б» значительно улучшается — налицо положительная динамика. Тем не менее скачкообразно вырастает уровень дефектов групп «в–е». Более того, в двух случаях удельный вес проверенных исследований с искажениями и нарушениями целостности достигает 100,0%. Затем происходит небольшое снижение, но в целом уровень дефектов этого типа остается совершенно неприемлемым.

Второй ИИ-сервис по ММГ работает стабильно с точки зрения отсутствия пропусков, потерь и искажений данных. Однако за весь период участия в Эксперименте так и не смог снизить до приемлемого уровня удельный вес дефекта «а». Длительность обработки исследований была очень высокой, полностью обесценивавшей результаты применения ИИ для врача-рентгенолога. Определенная положительная динамика наблюдалась, однако указанная технологическая проблема так и не была решена окончательно. На рис. 48 (а-в) представлена динамика технологических дефектов ИИ-сервисов, работавших с результатами рентгенографии и флюорографии.



Рисунок 48 а –Динамика технологических дефектов ИИ-сервисов модальностей РГ и ФЛГ. Знак * означает, что дефект соответствует группе дефектов «б–г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)

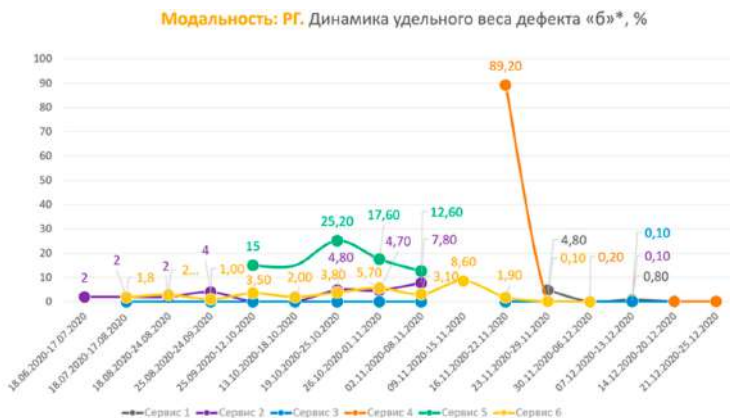


Рисунок 48 б –Динамика технологических дефектов ИИ-сервисов модальностей РГ и ФЛГ. Знак * означает, что дефект соответствует группе дефектов «б–г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)



Рисунок 48 в –Динамика технологических дефектов ИИ-сервисов модальностей РГ и ФЛГ. Знак * означает, что дефект соответствует группе дефектов «б–г» редакции приказа ДЗМ от 19.02.2020 г. №142 (в ред. от 15.05.2020 №552)

Три ИИ-сервиса продемонстрировали хорошую тенденцию к стабильности соблюдения временных норм обработки исследования и низкий удельный вес потерь данных. Однако все они отличались неприемлемо высоким уровнем искажений и нарушений целостности данных; динамика решения этой проблемы здесь часто не просто отсутствовала, но была отрицательной.

Для одного ИИ-сервиса отмечается нестабильная работа со значительным превышением допустимого уровня технологических дефектов. Только один ИИ-сервис (ФЛГ/РГ) продемонстрировал уверенность и эффективность в улучшении качества своей работы: нестабильность и высокий уровень дефектов были устранены. Почти того же самого добился еще один ИИ-сервис по РГ, однако, улучшив качество, этот сервис неприемлемо увеличил длительность обработки исследований.

Таким образом, при анализе результатов технологического мониторинга однозначно определяется тренд на снижение удельного веса дефектов, что отражает эволюцию ИИ-сервисов в Эксперименте. Программное обеспечение, аппаратные ресурсы компаний дорабатываются и развиваются с целью соответствия выставляемым к ним требованиям. Наиболее значимыми проблемами были:

1. Неприемлемо высокая длительность обработки одного исследования, обесценивающая результаты работы ИИ для врача-рентгенолога. В среднем этот дефект встречался в 22,0% ситуаций.

2. Дисбаланс работы ИИ-сервисов: улучшение качества обработки за счет длительности и наоборот.

Выявленный средний уровень дефектов всех групп свидетельствует о необходимости дальнейшего совершенствования аппаратно-программного обеспечения, а также математических моделей ИИ-сервисов. Вместе с тем 65,0% ИИ-сервисов успешно справились с задачей проведения анализа исследований в установленные временные нормативы.

§3. Сбор и анализ обратной связи по конкретным клиническим случаям от пользователей (врачей-рентгенологов)

В начале активной фазы Эксперимента (с 18.06.2020) результаты работы сервисов в виде приоритизации в рабочем списке и дополнительной серии изображений были доступны на рабочих местах в ЕРИС ЕМИАС 538 врачей-рентгенологов. В течение нескольких месяцев работы количество подключенных врачей неуклонно возрастало и составило 1054 к октябрю 2020 года.

Среди подключенных врачей-рентгенологов были выделены 2 категории (рис. 40):

1. Активные — использующие в той или иной мере при интерпретации и описании результатов исследований результаты работы ИИ-сервисов.

2. Дающие обратную связь — вносящие данные в форму обратной связи в ЕРИС ЕМИАС.

Количественные данные первой категории мы трактовали как используемость, второй — как вовлеченность.

При оценке в динамике (рис. 50) отмечается положительная тенденция используемости: плавное нарастание количества врачей-рентгенологов, применяющих в той или иной мере ИИ-сервисы в течение 5 месяцев, затем — признаки выхода на плато в течение 2-х месяцев.

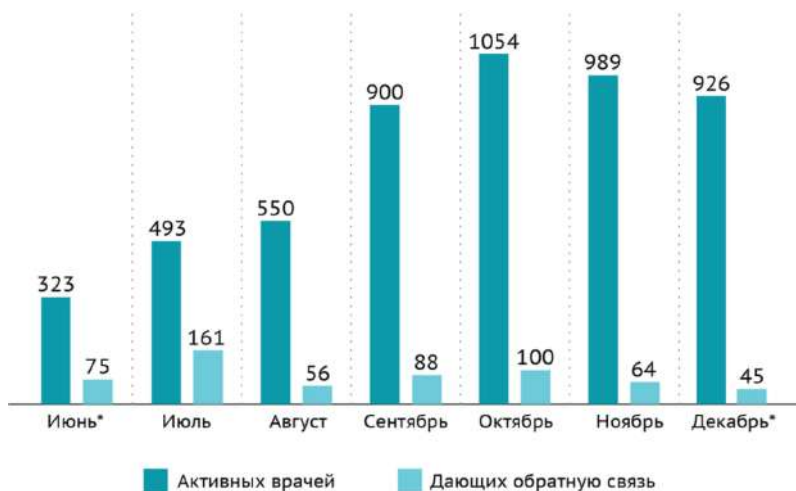


Рисунок 49 – Количественные данные об использовании ИИ-сервисов и вовлеченности врачей-рентгенологов (знаком * отмечены неполные месяцы: в июне работа велась с 18.06.2020 по 30.06.2020, в декабре с 01.12.2020 по 25.12.2020)



Рисунок 50 – Динамика использования ИИ-сервисов и вовлеченности врачей-рентгенологов

Для вовлеченности отмечается отрицательная динамика: небольшой подъем с 23,2% до 32,7% во втором сменяется плавным снижением до 4,9% в последнем. Подобная динамика в целом характерна для этапа внедрения новых информационных технологий, особенно с учетом добровольности предоставления обратной связи¹. Вместе с тем сформировалось «ядро» из специалистов с высокой вовлеченностью, постоянно сотрудничающих по вопросам обратной связи в Эксперименте.

За весь период наблюдения активными пользователями ИИ-сервисов стали 1338 врачей-рентгенологов, обратную связь по качеству работы сервисов предоставляли 300 из них, что составило 22,4%.

В целом вовлеченность в 22,4% не только соответствует опубликованным ранее данным, но и немного превышает их. Так, по итогам международного опроса 675 членов Европейского общества радиологов из 39 стран, установлено, что технологиями искусственного интеллекта в своей практической работе пользуется 20% врачей². Региональные опросы демонстрируют меньшие значения. Например, в Саудовской Аравии определен уровень использования ИИ врачами лучевой диагностики в 18%³.

Далее охарактеризуем и проанализируем структуру обратной связи от врачей-рентгенологов. За весь период Эксперимента получено 15028 результатов обратной связи, примерно в равных долях по каждой модальности (рис. 51). Суммарные данные об оценках качества работы ИИ-сервисов врачами-рентгенологами представлены в таблице 63.

¹ Christodoulakis C., Asgarian A., Easterbrook S. et al. Barriers to adoption of information technology in health care. Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering. Armonk, NY: IBM Corp, 2017.

² Pesapane F., Codari M., Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine // Eur Radiol Exp. 2018. Oct. 24. Vol. 2, №1. P. 35. DOI: 10.1186/s41747-018-0061-6.

³ Tajaldeen A., Alghamdi S. Evaluation of radiologist's knowledge about the Artificial Intelligence in diagnostic radiology: a survey-based study // Acta Radiol Open. 2020. Jul 31. Vol. 9, №7. DOI: 10.1177/2058460120945320.

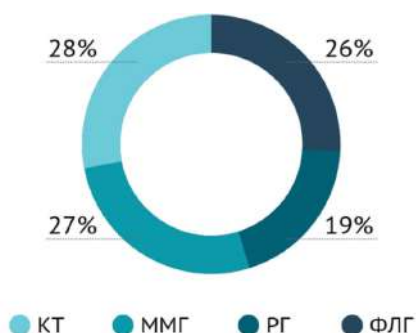


Рисунок 51 – Удельный вес результатов обратной связи от врачей-рентгенологов по разным модальностям

Таблица 63 – Результаты обратной связи от врачей-рентгенологов по качеству работы ИИ-сервисов

Категория отзыва	Удельный вес исследований, %*
Полное соответствие (согласен)	62,0
Не согласен	38,0
Значимое расхождение	8,0
Дефект	3,0
Грубая ошибка	5,0
Незначимое расхождение	26,0
Неверный диагноз	14,0
Неверная локализация	4,0
Неверная классификация	7,0
Прочее	4,0

*Данные удельного веса не суммируются, так как врач-рентгенолог одновременно может указать несколько вариантов в форме обратной связи (например, «неверный диагноз» и «неверная локализация»).

Врачи-рентгенологи определили полное соответствие результатов работы сервисов компьютерного зрения реальной рентгеноло-

гической картине в 62,0% случаев. В 38,0% зафиксированы несоответствия; из этого числа значимые расхождения имели место в 8,0% случаев, незначимые – в 26,0%, грубые ошибки отмечены в 5,0% исследований.

В таблице 64 приведены результаты обратной связи по каждому отдельному ИИ-сервису; в таблице 65 – комплексная оценка отзывов по модальностям.

Таблица 64 – Результаты обратной связи от врачей-рентгенологов по качеству работы каждого ИИ-сервиса

ИИ-сервис	Общее количество отзывов	Полное соответствие (согласен)	
		Абс.	Относит.
ИИ-сервис-1 (РГ)	250	204	82,0
ИИ-сервис-2 (КТ РЛ)	434	305	70,0
ИИ-сервис-3 (КТ Covid)	82	43	52,0
ИИ-сервис-4 (РГ)	2295	1117	49,0
ИИ-сервис-5 (КТ Covid)	35	22	63,0
ИИ-сервис-6 (РГ), (ФЛГ)	1117	497	44,0
ИИ-сервис-7 (КТ Covid)	499	242	48,0
ИИ-сервис-8 (КТ Covid)	61	26	43,0
ИИ-сервис-9 (РГ)	114	102	89,0
ИИ-сервис-10 (ММГ)	8	1	13,0
ИИ-сервис-11 (НДКТ РЛ)	1	1	100,0
ИИ-сервис-12 (Radlogics) (КТ Covid)	1962	1386	71,0
ИИ-сервис-13 (КТ Covid)	545	334	61,0
ИИ-сервис-14 (КТ Covid)	65	32	49,0
ИИ-сервис-15 (РГ)	741	360	49,0
ИИ-сервис-16 (ММГ)	2797	1540	55,0
ИИ-сервис-17 (ФЛГ)	4022	3084	77,0

Таблица 65 – Результаты обратной связи от врачей-рентгенологов по качеству работы ИИ-сервисов по модальностям

ИИ-сервис	Общее количество отзывов	Полное соответствие (согласен)	
		Абс.	Относит.
КТ ЗНО ОГК	434	305	70,0
КТ COVID-19	3249	2085	64,0
ММГ	2805	1541	55,0
НДКТ	1	1	100,0
РГ	4205	2103	50,0
ФЛГ	4334	3261	75,0

Наибольший процент согласия с заключением по результатам обратной связи был получен для модальностей «Флюорография» – 75,0% и «Компьютерная томография» (для выявления ЗНО легкого – 70,0%, COVID-19 – 64,0%). Примечательным фактом является низкое согласие врачей с результатами работы ИИ-сервисов по модальности «Рентгенография» – 50,0%.

Качество и объем полученных данных не позволяют использовать результаты обратной связи для объективной оценки и тем более – сравнения ИИ-сервисов.

Таким образом, в ходе Эксперимента фиксируется интенсивный рост с последующей стабилизацией используемости ИИ-сервисов врачами-рентгенологами. Общая вовлеченность врачей составила 22,4%, что несколько превышает общемировые показатели. Отрицательная динамика показателя обусловлена психологическими факторами и требует дальнейших научных исследований. За весь период работы предоставлено 15028 отзывов через форму обратной связи в ЕРИС ЕМИАС. Врачи-рентгенологи определили полное соответствие результатов работы сервисов компьютерного зрения реальной рентгенологической картине в 62,0% случаев. В 38,0% ситуаций зафиксированы несоответствия (в том числе значимые – 8,0% случаев, незначимые – 26,0%, грубые ошибки – 5,0%).

54. Оценка технологической готовности

Оценка технологической готовности компаний-разработчиков ИИ-сервисов проведена сотрудниками Департамента информационных технологий (ДИТ) города Москвы. Для этого использовался специальный опросник. Усредненные результаты оценки представлены в таблице 66.

Таблица 66 – Усредненные результаты оценки технологической готовности компаний-разработчиков ИИ-сервисов

Критерий	Результат
Оперативность ответа на замечания по технологической интеграции	72,0% – среднее значение 1–3 дня 28,0% – более 7 дней
Результативность ответа на замечания по технологической интеграции	64,0% – меры в полном объеме с 1 раза 36,0% – требовалась дополнительная консультация
Оценка технической грамотности инженеров компании	78,0% компаний имеют персонал с высокой квалификацией
Проактивность компании	50,0%
Проактивность компании по определению дефектов при работе в ППАК ЕРИС ЕМИАС	15,0%

Высокий уровень технологической готовности имели 46,0% (6) компаний-разработчиков ИИ-сервисов (оценки 8–10 баллов). В частности, это подтверждается быстрым и эффективным реагированием на замечания и требования, высказываемые со стороны Департамента информационных технологий города Москвы в процессе технологической интеграции ИИ-сервисов в ЕРИС ЕМИАС: 72,0% компаний вносили необходимые изменения в течение 1–3 рабочих дней, при этом 64,0% разработчиков не требовались повторные консультации или разъяснения. В целом, в 78,0% компаний инженерный персонал имеет высокий уровень квалификации. Предложения по оптимизации процессов схемы технологической интеграции активно вносили представители половины компаний. Значимые дефекты и недостатки в работе ППАК ЕРИС ЕМИАС вы-

явили 15,0% разработчиков и активно сообщили об этом. Это позволило сотрудникам ДИТ города Москвы, теперь уже со своей стороны, оперативно внести изменения и повысить качество работы. По результатам оценки составлен рейтинг компаний по показателям технической готовности, использованный для внутренних процессов организаторов Эксперимента.

В целом, отметим, что максимально возможное количество баллов удалось набрать двум юридическим лицам, что указывает на изначально высокий уровень готовности инженерных групп данных компаний к интеграции их сервисов в ЕРИС ЕМИАС. Однако при сопоставлении с количеством дней, спустя которые ИИ-сервис был готов провести функциональное тестирование (эквивалентно интеграции), явной закономерности обнаружено не было: коэффициент корреляции Пирсона равен — 0,1057. Таким образом, высокая техническая готовность ИИ-сервиса не означает быструю интеграцию: например, один ИИ-сервис, получивший средний балл оценки 5, интегрировался за 22 дня.

Таким образом, полученные данные свидетельствуют о наличии достаточно высокого уровня технологической готовности компаний-разработчиков ИИ-сервисов к интеграции и работе в условиях реальных информационных систем в сфере здравоохранения на уровне субъекта РФ.

§5. Хронометраж

5.1. Длительность обработки ИИ-сервисами результатов одного исследования

Данные о длительности анализа одного исследования каждым ИИ-сервисом получены путем выгрузки файлов регистрации из ЕРИС ЕМИАС. Далее проведен расчет средних значений (таблицы 67, 68).

Таблица 67 – Медианное время обработки исследований ИИ-сервисами по модальностям

Модальность	Среднее время, сек.	95% ДИ
КТ РЛ	817	(803, 832)
НДКТ РЛ	656	(512, 921)
КТ Covid	276	(274, 277)
ММГ	144	(140, 149)
РГ	23	(23, 24)
ФЛГ	12	(12, 12)

Минимальным было медианное время обработки результатов рентгенографий и флюорографий – 23 и 12 секунд соответственно. Дольше всего выполнялся анализ КТ – 13 минут и 37 секунд. Для компьютерной томографии длительность обработки результатов колебалась в очень большом диапазоне, в отдельных случаях превышая 6 часов. Это обусловлено несоответствием аппаратного обеспечения компаний-разработчиков ИИ-сервисов.

Таблица 68 – Медианное время обработки исследований ИИ-сервисами

Наименование ИИ	Средняя длительность, сек.	95% ДИ	Отклонение от нормативного значения, %
ИИ-сервис-1 (РГ)	50	(49, 52)	0
ИИ-сервис-2 (КТ РЛ)	817	(803, 832)	+36,2%
ИИ-сервис-3 (КТ Covid)	206	(164, 284)	0
ИИ-сервис-4 (РГ)	12	(11, 12)	0
ИИ-сервис-5 (КТ Covid)	504	(475, 532)	0
ИИ-сервис-6 (РГ/ФЛГ)	22	(21, 23)	0
ИИ-сервис-7 (КТ Covid)	140	(137, 144)	0
ИИ-сервис-8 (КТ Covid)	79	(74, 84)	0
ИИ-сервис-9 (РГ)	1167	(70, 4728)	+94,5%
ИИ-сервис-10 (ММГ)	2486	(1127, 4271)	+314,3%
ИИ-сервис-11 (НДКТ РЛ)	656	(512, 921)	+9,5%
ИИ-сервис-12 (КТ Covid)	416	(411, 422)	0
ИИ-сервис-13 (КТ Covid)	191	(189, 194)	0
ИИ-сервис-14 (КТ Covid)	1592	(1166, 2090)	+165,3%
ИИ-сервис-15 (РГ)	36	(29, 45)	0
ИИ-сервис-16 (ММГ)	133	(129, 137)	0
ИИ-сервис-17 (ФЛГ)	13	(13, 13)	0

При анализе работы каждого ИИ-сервиса установлено, что среднее время анализа одного исследования за весь период работы в Эксперименте не соответствовало установленному нормативному значению в 600 секунд (10 минут) у 27,8% (5) ИИ-сервисов.

Исходя из полученных данных и сведений о средней длительности подготовки описаний результатов лучевых исследований врачами-рентгенологами, утверждаем следующее. Отклонение от нормативного времени анализа одного исследования является критичной проблемой ИИ-сервисов. Предоставление результатов

автоматизированного анализа с критичным опозданием делает бессмысленным применение технологий искусственного интеллекта. В предусмотренных Экспериментом клинических сценариях врач-рентгенолог просто успевает подготовить протокол раньше, чем ИИ-сервис направляет свои результаты в ЕРИС ЕМИАС. Выше нами показан высокий удельный вес выявляемых при технологическом мониторинге дефектов группы «а»: превышение нормативного времени анализа одного исследования в среднем встречается в 22,0% случаев.

Таким образом, среднее время обработки исследований ИИ-сервисами по модальностям соответствовало нормативным значениям, однако у 27,8% ИИ-сервисов зафиксированы критичные отклонения от установленного времени анализа одного исследования. Данная проблема является значимой. Для обеспечения клинической приемлемости применения ИИ-сервисов считаем обязательным нормативное установление требований к максимальному времени анализа одного исследования (с учетом клинических сценариев).

5.2. Влияние ИИ-сервисов на длительность описаний результатов лучевых исследований

Изучено влияние ИИ-сервисов на длительность описаний результатов лучевых исследований по модальностям. Проведен статистический анализ различий длительности описаний в медицинских организациях до и после внедрения ИИ-сервисов (группы 1 и 2). Расчеты проводились для трех вариантов развития ситуации: ИИ-сервис уменьшает, увеличивает или не влияет на длительность описаний.

При совокупном анализе по модальностям единственным статистически достоверным эффектом было увеличение медианной длительности описаний результатов КТ/НДКТ, выполненных для выявления злокачественных новообразований, на 42,0% (1,18 минуты) (критерий Уилкоксона = больше («greater»), $p=0,04$). При анализе с учетом условий оказания медицинской помощи установлены следующие факты: указанное выше увеличение медианной длительности описаний результатов КТ/НДКТ характерно только для амбулаторного звена (увеличение на 37,0%, $p=0,01$).

При углубленном анализе установлено, что применение, по крайней мере, одного ИИ-сервиса достоверно снижает длительность описаний результатов профилактической маммографии в амбулаторном звене на 15,0% или 1,08 минуты (критерий Уилкоксона = меньше («less»), $p = 0,03$). Аналогично зафиксировано достоверное снижение длительности подготовки описаний результатов маммографии в стационарном звене на 2 минуты или 50,0% (критерий Уилкоксона = меньше («less»), $p = 0,05$).

Отмечается факт противоположного воздействия на длительность описаний при использовании различных ИИ-сервисов в рамках одной модальности. Например, при описаниях КТ Covid одного ИИ-сервиса привело к сокращению медианных временных затрат на 1 минуту или 22,0% (критерий Уилкоксона = меньше («less»), $p = 0,14$), а другого — к увеличению на 1 минуту или 27,0% (критерий Уилкоксона = больше («greater»), $p = 0,01$). Разнонаправленный характер воздействий, вероятно, зависит от функционального качества ИИ-сервисов (включая особенности интерфейса и отображения результатов работы), от диагностической точности (уровень ложноположительных находок), а также от степени получения врачами практических навыков по применению технологий компьютерного зрения. Данный феномен требует дальнейшего углубленного изучения.

Таким образом, применение ИИ-сервисов может положительно влиять на производительность труда врачей-рентгенологов, однако выраженность и достоверность этого влияния зависит от вида клинического сценария (производственного процесса), модальности и условий оказания медицинской помощи.

Выявленное увеличение длительности описаний для КТ/НДКТ РЛ может быть истолковать двояко:

А. Отрицательно — некачественная работа ИИ-сервисов и их негативное влияние.

В. Положительно — более вдумчивая, скрупулезная работа врача благодаря поддержке принятия решений со стороны ИИ-сервиса.

В рамках данного исследования получить убедительные доказательства в пользу одного из вариантов не представляется возможным. Парадоксальные результаты свидетельствуют о необходимости дальнейших углубленных исследований.

Нулевая гипотеза H_0 : {Снижение длительности описания результатов лучевых исследований при работе ИИ-сервисов <30%}

отклонена для описания результатов профилактической маммографии. Для остальных ситуаций – принята.

§6. Обобщение результатов оценки целесообразности применения ИИ-сервисов

В соответствии с существующей методологией¹ целесообразность применения ИИ-сервисов в лучевой диагностике изучена по нескольким основным аспектам:

1. Приемлемость.

Общая вовлеченность врачей-рентгенологов в Эксперимент составила 22,4%. Внедрение и активное использование ИИ-сервисов изменили отношение врачей к данному виду технологий с позиций их влияния на производительность и качество работы: имеет место переход от крайних оценок к нейтральным, умеренным. Изменилась структура опасений. Это свидетельствует о появившемся запросе со стороны врачебного сообщества на предоставление большего объема научной доказательной информации о технологиях ИИ. ИИ-сервисы представляют собой приемлемую, но по-прежнему недостаточно изученную с точки зрения реальной эффективности технологию.

2. Спрос.

Количество активных пользователей ИИ-сервисов колебалось от 323 до 1054. За период Эксперимента наблюдалось плавное нарастание их количества в течение 5 месяцев, затем появились признаки выхода на плато в течение 2 месяцев. Для 100,0% ИИ-сервисов зафиксирован отрицательный индекс потребительской лояльности. Требуется развитие методической базы для применения ИИ в решении более конкретных клинических задач и формирование системы финансирования (тарифов в системе обязательного медицинского страхования).

¹ Bowen D.J., Kreuter M., Spring B. et al. How we design feasibility studies // Am J Prev Med. 2009. Vol. 36, №5. P. 452–457. DOI:10.1016/j.amepre.2009.02.002.

3. Реализация.

Интегрированные в ЕРИС ЕМИАС ИИ-сервисы обработали 83,3% направленных на анализ исследований (1 468 872 из 1 762 949) по 4 разным модальностям. Средний удельный вес технологических дефектов имеет достаточно высокий уровень, многие ИИ-сервисы работают нестабильно. Вместе с тем уровень и структура технологических дефектов ИИ-сервисов значительно различаются. В условиях интеграции в радиологическую информационную систему субъекта РФ 65,0% ИИ-сервисов обеспечивают анализ результатов лучевых исследований в соответствии с установленными временным нормативами. Исходя из анализа результатов технологического мониторинга нулевая гипотеза H_0 : {Уровень дефектов типа «а» и/или «б» и/или «в» и/или «г» и/или «д» и/или «е» > 10%} принята для групп дефектов «в–е», отклонена для групп дефектов «а» и «б».

4. Осуществимость.

По результатам объективного исследования (хронометража) длительности описаний результатов лучевых исследований врачами-рентгенологами нулевая гипотеза H_0 : {Снижение длительности описания результатов лучевых исследований при работе ИИ-Сервисов <30%} отклонена для описания результатов профилактической маммографии; для остальных ситуаций – принята. Вместе с тем использование ИИ-сервисов может достоверно удлинять длительность описаний некоторых видов исследований, в частности – результатов КТ/НДКТ, проведенных для выявления ЗНО легкого. Парадоксальное влияние требует дальнейшего исследования.

5. Интеграция.

Заявки на участие в Эксперименте в установленном порядке подало 21 юридическое лицо (39 ИИ-сервисов). По состоянию на 25.12.2020 успешно интегрировались в ППАК ЕРИС ЕМИАС и приступили к работе в условиях реальных производственных процессов службы лучевой диагностики 46,0% (18) ИИ-сервисов. Остальные находятся на этапе интеграции в тестовый контур. Отчасти это вызвано временным ограничением – фактический средний срок технологической интеграции составляет 65 ± 34 дней. Не менее 78,0% инженерного персонала компаний-разработчиков имеет достаточный уровень квалификации. Большая часть компаний работает интенсивно и проактивно, быстро адаптируя собственные разработки к установленным стандартам и требованиям. Таким об-

разом, достигнутый уровень технологического развития и корпоративной культуры позволяет не менее 46,0% компаний-разработчиков успешно и в актуальные сроки интегрировать ИИ-сервисы в функционирующие радиологические информационные системы (в том числе на уровне субъектов РФ).

6. Способность к адаптации.

В условиях начавшейся в 2020 году пандемии COVID-19 заявки на участие в Эксперименте были дополнительно поданы для 15 ИИ-сервисов по выявлению COVID-19. Из них 47,0% успешно интегрировались в ППАК ЕРИС ЕМИАС, прошли все предусмотренные процедуры и приступили к работе с потоком исследований. Примечательно, что 4 ИИ-сервиса были заявлены компаниями, уже участвующими в Эксперименте по другим модальностям (половина из них интегрирована в ППАК ЕРИС ЕМИАС). При технологическом мониторинге именно для ИИ-сервисов, работающих по модальности КТ Covid, наиболее часто отмечены положительные тенденции оперативной доработки и снижения уровня дефектов. Отмечается высокая способность к адаптации и развитию разработчиков ИИ-сервисов, в том числе в условиях чрезвычайных ситуаций.

Таким образом, применение ИИ-сервисов в лучевой диагностике целесообразно для повышения производительности труда врачей-рентгенологов (в том числе в условиях чрезвычайных ситуаций). Отмечается высокий уровень технологической готовности к развитию технологий искусственного интеллекта. Необходимо введение обоснованных нормативных требований к показателям работы ИИ-сервисов, включая временные нормативы. Требуется развитие научных и методических основ применения ИИ в лучевой диагностике (в том числе для решения конкретных клинических задач) на принципах доказательной медицины. Перспективным представляется разработка системы медицинских услуг, оказываемых с применением ИИ-сервисов и финансируемых за счет средств системы обязательного медицинского страхования.

ЗАКЛЮЧЕНИЕ

В 2020 году в городе Москве в результате реализации комплекса организационно-методических, научных, образовательных и технологических мероприятий успешно проведено мультицентровое проспективное научное исследование «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы».

Разработаны и успешно апробированы клинические сценарии применения технологий искусственного интеллекта в лучевой диагностике в рамках массовых профилактических исследований, а также в условиях пандемии новой коронавирусной инфекции. Согласно данным сценариям сформированы технические и диагностические требования к программному обеспечению на основе технологий компьютерного зрения, включая специфические задачи для каждой модальности.

Для оценки качества и применимости технологий компьютерного зрения в лучевой диагностике разработаны методики:

- этапного тестирования (функционального, калибровочного, контрольно-технического);
- технологического мониторинга.

Методики апробированы и валидированы путем проведения:

- 1) 40 процедур функционального тестирования (33 первичных и 7 повторных) для 85,0% ИИ-сервисов, заявки которых были приняты к участию;
- 2) 38 процедур калибровочного тестирования (23 первичных и 15 повторных) для 70,0% ИИ-сервисов, успешно прошедших функциональное тестирование;
- 3) 20 процедур контрольно-технического тестирования с охватом 295 диагностических устройств;
- 4) 161 процедура технологического мониторинга.

На основе указанных методик разработана обобщенная методология клинических испытаний технологий искусственного интеллекта для сферы лучевой диагностики (в том числе процедура технологического мониторинга рассматривается как способ пост-регистрационных испытаний).

Научно обоснованы классификации, методики формирования, схема жизненного цикла эталонных наборов данных для обучения и тестирования технологий искусственного интеллекта. На основе авторской методологии для целей Эксперимента подготовлено 26 эталонных наборов данных с общим числом подготовленных результатов 1520 лучевых исследований по 4 модальностям. Авторская методология формирования наборов данных валидирована на международном уровне независимыми исследователями.

Диагностическая точность и воспроизводимость результатов работы ИИ-сервисов на эталонных данных — достаточная, на данных в условиях реального производственного процесса — недостаточная. При применении в реальных клинических условиях у всех ИИ-сервисов отмечается выраженное статистически достоверное (U-критерий Манна-Уитни = 99, $p < 0,05$) снижение значений метрик диагностической точности ниже нормативных значений. Более стабильные и относительно высокие значения метрик показали ИИ-сервисы для анализа результатов компьютерной томографии и рентгенографии; остальные приблизились к критично низким значениям.

По результатам экспертного аудита случаев расхождения мнений врача-рентгенолога и ИИ-сервисов установлен высокий уровень согласованности решений экспертов и врачей-рентгенологов (каппа Козна стремится к 1,0) и крайне низкий для решений экспертов и ИИ-сервисов (каппа Козна от -0,59 до 0,7).

Результаты работы ИИ-сервисов были доступны для использования и оценки на автоматизированных рабочих местах в ЕРИС ЕМИАС 538 врачей-рентгенологов (физических лиц). Общая вовлеченность врачей-рентгенологов в Эксперимент составила 22,4%. Внедрение и активное использование ИИ-сервисов изменили отношение врачей к данному виду технологий: отмечен переход от крайних оценок к нейтральным, умеренным. Крайне низкий индекс потребительской лояльности (-50,0 — -90,0) указывает на то, что у врачебного сообщества сформирован запрос на предоставление научной доказательной информации о технологиях искусственного интеллекта.

Благодаря применению ИИ-сервисов достоверно снижается длительность подготовки описаний результатов профилактической маммографии в амбулаторном звене на 15,0% ($p=0,03$), в стационарном звене — на 50,0% ($p=0,05$). Длительность описа-

ний результатов КТ/НДКТ для выявления ЗНО легких достоверно увеличивается на 42,0% ($p=0,04$). Разнонаправленный и парадоксальный характер влияния ИИ-сервисов на производительность труда врачей-рентгенологов требует дальнейшего углубленного изучения с учетом разных клинических сценариев, форм, видов и условий оказания медицинской помощи.

До 46,0% ИИ-сервисов имеют высокий уровень технологической готовности к работе в условиях практического здравоохранения и способны интегрироваться в реально функционирующую радиологическую информационную систему субъекта РФ в течение 65 ± 34 дней. Отсутствует статистически значимая связь между уровнем технологической готовности и скоростью интеграции (коэффициент корреляции Пирсона равен $-0,1057$).

Из числа ИИ-сервисов, успешно интегрированных в радиологическую информационную систему субъекта РФ, 65,0% обеспечивают анализ результатов лучевых исследований в соответствии с установленными временным нормативами. Зафиксировано объективное отсутствие стабильности в работе ИИ-сервисов, что требует обязательного применения процедур пострегистрационного контроля при внедрении таких технологий в практическое здравоохранение.

Таким образом, в результате проведенного Эксперимента:

1. Нулевая гипотеза H_0 : {Уровень дефектов типа «а» и/или «б» и/или «в» и/или «г» и/или «д» и/или «е» > 10%} принята для групп дефектов «в–е», отклонена для групп дефектов «а» и «б».

2. Нулевая гипотеза H_0 : {Снижение длительности описания результатов лучевых исследований при работе ИИ-сервисов <30%} отклонена для описания результатов профилактической маммографии.

3. Нулевая гипотеза H_0 : {Чувствительность <0,81 и/или Специфичность <0,81 и/или AUC <0,81} для условий: калибровочного тестирования – отклонена; работы с потоком исследований отклонена для модальности КТ Covid, для остальных – принята.

С методической и нормативной точек зрения, результаты работы технологий искусственного интеллекта в лучевой диагностике должны в обязательном порядке верифицироваться врачом (специалистом по лучевой диагностике).

При масштабном внедрении и рутинном применении технологий искусственного интеллекта необходимо предусмотреть функ-

циональные возможности медицинских (радиологических) информационных систем для автоматизированной отправки результатов исследований на независимый контроль качества (аудит) при наличии расхождений решений врача и ИИ. Участвовавшие в Эксперименте ИИ-сервисы нуждаются в технологической доработке, дополнительном обучении на наборах данных московской популяции, проведении технических и клинических испытаний (в том числе с целью получения статуса медицинского изделия).

Технологии искусственного интеллекта (компьютерного зрения) целесообразно применять в системе здравоохранения города Москвы при проведении массовых профилактических исследований методами лучевой диагностики. Дальнейшие научные исследования должны быть посвящены вопросам клинической и экономической эффективности применения технологий искусственного интеллекта в рамках конкретных производственных процессов и медицинских услуг.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Блохин И. А., Морозов С. П., Чернина В. Ю. [и др.]. Использование искусственного интеллекта в здравоохранении: опыт валидации алгоритма искусственного интеллекта в медицинских организациях в условиях пандемии COVID-19 // Мониторинг общественного мнения: экономические и социальные перемены. 2021. №1. С. 271–282.
2. Владимирский А. В., Лебедев Г. С. Телемедицина. М.: ГЭОТАР-Медиа, 2018. 576 с.
3. Гончар А. П., Елизаров А. Б., Кульберг Н. С. [и др.]. Автоматическое определение плотности печени по данным компьютерной томографии и ультранизкодозной компьютерной томографии // Новости хирургии. 2020. Т. 28. №6. С. 636–647.
4. 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. №1 (43). С. 36–45.
5. Гусев А. В., Плисс М. А. Основные рекомендации к созданию и развитию информационных систем в здравоохранении на базе искусственного интеллекта // Врач и информационные технологии. 2018. №3. С. 45–60.
6. Гусев А. В., Добридень С. Л. Искусственный интеллект в медицине и здравоохранении // Информационное общество. 2017. №4–5. С. 78–93.
7. Долгушина Н. В., Воронов Д. А., Грачев С. В. Методология научных исследований в клинической медицине. М.: ГЭОТАР-Медиа, 2016. 112 с.
8. Злокачественные новообразования в России в 2018 году (заболеваемость и смертность) / под ред. А. Д. Каприна, В. В. Старинского, Г. В. Петровой. М.: МНИОИ им. П. А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2019. 250 с.
9. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика) /

сост. С. П. Морозов, А. В. Владзимирский, В. Г. Кляшторный [и др.] // Сер. «Лучшие практики лучевой и инструментальной диагностики». М., 2019. Вып. 57. 51 с.

10. Корсаков С. Н. Начертание нового способа исследования при помощи машин, сравнивающих идеи / пер. с франц. под ред. А. С. Михайлова. М.: МИФИ, 2009. 44 с.

11. Кульберг Н. С., Гусев М. А., Решетников Р. В. [и др.]. Методология и инструментарий создания обучающих выборок для систем искусственного интеллекта по распознаванию рака легкого на КТ-изображениях // Здоровоохранение Российской Федерации. 2020. Т. 64. №6. С. 343–350.

12. Кульберг Н. С., Елизаров А. Б., Новик В. П. [и др.]. Автоматическое пакетное определение рентгеновской плотности печени для выявления субклинических заболеваний печени // Радиология – практика. 2020. №3 (81). С. 50–61.

13. Кудрявцев Н. Д., Сергунова К. А., Иванова Г. В. [и др.]. Оценка эффективности внедрения технологии распознавания речи для подготовки протоколов рентгенологических исследований // Врач и информационные технологии. 2020. № S1. С. 58–64.

14. Лебедев Г. С., Маслюков А. П., Шадеркин И. А., Шадеркина А. И. Глубокое машинное обучение (искусственный интеллект) в ультразвуковой диагностике // Журнал телемедицины и электронного здравоохранения. 2020. Т. 6. №2. С. 22–29.

15. Лебедев Г. С., Фомина И. В., Шадеркин И. А. [и др.]. Основные направления развития интернет-технологий в здравоохранении (систематический обзор) // Социальные аспекты здоровья населения. 2017. №5 (57). С. 10.

16. Ли Кай-фу. Сверхдержавы искусственного интеллекта. Китай, кремниевая долина и новый мировой порядок. М.: Манн, Иванов и Фербер, 2019. – 240 с.

17. Логунова Т. А., Арзамасов К. М., Кирпичев Ю. С. [и др.]. Методика опроса пациентов и врачей при внедрении интеллектуальных технологий в практическое здравоохранение / В сб.: Социология здоровья: здравоохранение, которому доверяют. Труды II Форума с международным участием. М., 2020. С. 56–60.

18. Мелдо А. А., Уткин Л. В., Трофимова Т. Н. Искусственный интеллект в медицине: современное состояние и основные направления развития интеллектуальной диагностики // Лучевая диагностика и терапия. 2020. №1 (11). С. 9–17.

19. Методика разработки норм времени и нагрузки медицинского персонала. М.: ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения», 2013. 25 с.

20. Методические рекомендации по порядку проведения экспертизы качества, эффективности и безопасности медицинских изделий (в части программного обеспечения) для государственной регистрации в рамках национальной системы. М.: ФГБУ «ВНИИ-ИМТ» Росздравнадзора, 2018. 31 с.

21. Морозов С. П., Владимирский А. В., Ледихова Н. В. [и др.]. Московский эксперимент по применению компьютерного зрения в лучевой диагностике: вовлеченность врачей-рентгенологов // Врач и информационные технологии. 2020. №4. С. 14–23.

22. Морозов С. П. Решения на базе искусственного интеллекта – новый стандарт безопасности в лучевой диагностике // Московская медицина. 2020. №2 (36). С. 24–26.

23. Морозов С. П., Владимирский А. В., Гомболевский В. А. [и др.]. Искусственный интеллект в скрининге рака легкого: оценка диагностической точности алгоритма для анализа низкодозовых компьютерных томографий // Туберкулез и болезни легких. 2020. Т. 98. №8. С. 24–31.

24. Морозов С. П., Владимирский А. В., Ледихова Н. В. [и др.]. Оценка диагностической точности системы скрининга туберкулеза легких на основе искусственного интеллекта // Туберкулез и болезни легких. 2018. №96 (8). С. 42–49. DOI:10.21292/2075-1230-2018-96-8-42–49.

25. Морозов С. П., Владимирский А. В., Гомболевский В. А. [и др.]. Искусственный интеллект: автоматизированный анализ текста на естественном языке для аудита радиологических исследований // Вестник рентгенологии и радиологии. 2018. Т. 99. №5. С. 253–258.

26. Морозов С. П., Владимирский А. В., Черняева Г. Н. [и др.]. Валидация диагностической точности алгоритма «искусственного интеллекта» для выявления рассеянного склероза в условиях городской поликлиники // Лучевая диагностика и терапия. 2020. №2 (11). С. 58–65.

27. Морозов С. П., Гомболевский В. А., Чернина В. Ю. [и др.]. Прогнозирование летальных исходов при COVID-19 по данным компьютерной томографии органов грудной клетки // Туберкулез

и болезни легких. 2020. Т. 98, №6. С. 7–14. URL: <https://doi.org/10.21292/2075-1230-2020-98-6-7-14>.

28. Морозов С. П., Владзимирский А. В., Сафронов Д. С. Бенчмаркинг для оценки качества цифровизации отделений лучевой диагностики: разработка методологии // Врач и информационные технологии. 2019. №1. С. 40–45.

29. Морозов С. П., Владзимирский А. В., Ветшева Н. Н. [и др.]. Референс-центр лучевой диагностики: обоснование и концепция // Менеджер здравоохранения. 2019. №8. С. 25–34.

30. Морозов С. П., Владзимирский А. В., Ледихова Н. В. Телерадиология в Российской Федерации: достигнутый уровень // Врач и информационные технологии. 2019. №2. С. 67–73.

31. Морозов С. П., Гомболевский В. А., Владзимирский А. В. [и др.]. Результаты первого года скрининга рака легкого с помощью низкодозной компьютерной томографии в Москве // Вопросы онкологии. 2019. Т. 65. №2. С. 224–233.

32. Морозов С. П., Чернина В. Ю., Андрейченко А. Е., Владзимирский А. В. [и др.]. Как искусственный интеллект влияет на оценку поражения легких при COVID-19 по данным КТ грудной клетки? // Digital Diagnostics. 2021. Т. 2. №1. С. 27–38.

33. Морозов С. П., Говорухина В. Г., Диденко В. В. [и др.]. Перспективы использования технологий искусственного интеллекта (ИИ) в скрининге рака молочной железы // Вопросы онкологии. 2020. Т. 66. №6. С. 603–608.

34. Морозов С. П., Андрейченко А. Е., Блохин И. А. [и др.]. MOSMEDDATA: Датасет 1110 компьютерных томографий органов грудной клетки, выполненных во время эпидемии COVID-19 // Digital Diagnostics. 2020. Т. 1. №1. С. 49–59.

35. Морозов С. П., Кузьмина Е. С., Ледихова Н. В. [и др.]. Мобилизация научно-практического потенциала службы лучевой диагностики г. Москвы в пандемию COVID-19 // Digital Diagnostics. 2020. Т. 1. №1. С. 5–12.

36. Мураев А. А., Гусейнов Н. А., Цай П. А. [и др.]. Искусственные нейронные сети в лучевой диагностике, в стоматологии и в челюстно-лицевой хирургии (обзор литературы) // Клиническая стоматология. 2020. №3 (95). С. 72–80.

37. Николаев А. Е., Чернина В. Ю., Блохин И. А. [и др.]. Перспективы использования комплексной компьютер-ассистированной ди-

агностики в оценке структур грудной клетки // Хирургия. Журнал им. Н. И. Пирогова. 2019. №12. С. 91–99.

38. Николенко С., Кадуринов А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018. 480 с.

39. Основы менеджмента медицинской визуализации / С. П. Морозов [и др.]; под ред. С. П. Морозова. М.: ГЭОТАР-Медиа, 2020. 432 с.: ил. DOI:10.33029/9704-5247-9-men-2020-1-424.

40. Оценка качества рентгенорадиологических исследований / сост. С. П. Морозов, Н. Н. Ветшева, Н. В. Ледихова [и др.] // Сер. «Лучшие практики лучевой и инструментальной диагностики». М., 2019. Вып. 22. 50 с.

41. Петряйкин А. В., Белая Ж. Е., Гомболевский В. А. [и др.]. Оценка деформации тел позвонков в диагностике компрессионных переломов с помощью методов искусственного интеллекта // Остеопороз и остеопатии. 2020. Т. 23. №2. С. 118–119.

42. Петряйкин А. В., Белая Ж. Е., Киселева А. Н. [и др.]. Технология искусственного интеллекта для распознавания компрессионных переломов позвонков с помощью модели морфометрического анализа, основанной на сверточных нейронных сетях // Проблемы эндокринологии. 2020. Т. 66. №5. С. 48–60.

43. Состояние онкологической помощи населению России в 2018 году / под ред. А. Д. Каприна, В. В. Старинского, Г. В. Петровой. М.: МНИОИ им. П. А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2019. 236 с.

44. Тюрин И. Е. Лучевая диагностика в Российской Федерации в 2016 г. // Вестник рентгенологии и радиологии. 2017. Т. 98. №4. С. 219–226.

45. Чернина В. Ю., Писов М. Е., Соловьев А. В. [и др.]. Оценка объемов эпикардальной и перикардальной жировой ткани с применением алгоритмов искусственного интеллекта по данным скрининговой ультранизкодозной компьютерной томографии грудной клетки: первые результаты // Трансляционная медицина. 2019. Т. 6, № S1. С. 383–384.

46. Чернина В. Ю., Писов М. Е., Беляев М. Г. [и др.]. Волюметрия эпикардальной жировой ткани: сравнение полуавтоматического измерения и алгоритма машинного обучения // Кардиология. 2020. Т. 60, №9. С. 46–54.

47. Шапиро Л., Стокман Дж. Компьютерное зрение / пер. с англ. 3-е изд. 763 с. М.: БИНОМ; Лаборатория знаний, 2015

48. Яковлева Н. Ф. Социологическое исследование [Электронный ресурс]: учеб. пособие. 2-е изд., стер. М.: ФЛИНТА, 2014. 250с.
49. Allen B., Dreyer K. The Role of the ACR Data Science Institute in Advancing Health Equity in Radiology // J Am Coll Radiol. 2019. Vol. 16, №14. P. 644–648.
50. Artificial Intelligence in Medical Imaging. Opportunities, Applications and Risks / Erik R. Ranschaert, Sergey Morozov, Paul R. Algra (Eds.). Springer, 2019.
51. Ayer A.J. Language, Truth, and Logic. 2nd edition. London: Gollancz, 1946. 160 p.
52. Baldwin D.R., Gustafson J., Pickup L., Arteta C. et al. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules // Thorax. 2020. Vol. 75, №4. P. 306–312. DOI: 10.1136/thoraxjnl-2019-214104.
53. Bossuyt P.M., Reitsma J.B., Bruns D.E. et al. For the STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies // Radiology. 2015. Vol. 277, №3. P. 826–832.
54. Bowen D.J., Kreuter M., Spring B. et al. How we design feasibility studies // Am J Prev Med. 2009. Vol. 36, №5. P. 452–457. DOI:10.1016/j.amepre.2009.02.002.
55. Bressen K.K., Niehues S.M. et al. 3D U-Net for segmentation of COVID-19 associated pulmonary infiltrates using transfer learning: State-of-the-art results on affordable hardware. 2021. arXiv:2101.09976v1.
56. Cobb A.N., Benjamin A. J., Huang E. S. et al. Big data: More than big data sets // Surgery. 2018. №164. P. 640–642.
57. Christodoulakis C., Asgarian A., Easterbrook S. et al. Barriers to adoption of information technology in health care. Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering. Armonk, NY: IBM Corp, 2017.
58. Codari M., Melazzini L., Morozov S.P. et al. Impact of artificial intelligence on radiology: a EuroAIM survey among members of the European Society of Radiology // Insights Imaging. 2019. Oct 31. Vol. 10, №1. P. 105. DOI: 10.1186/s13244-019-0798-3.
59. Dekker R. The importance of having data-sets // IATUL Annu. Conf. Proc. 2006.

60. Duggan G.E., Reicher J.J., Liu Y. et al. Improving reference standards for validation of AI-based radiography // *Br J Radiol.* 2021. Jul 1. №94 (1123). P. 20210435. DOI: 10.1259/bjr.20210435.
61. Geis J.R., Brady A., Wu C.C. et al. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement // *Insights Imaging.* 2019. Vol. 70, №4. P. 329–334. URL: <https://DOI.org/10.1186/s13244-019-0785-8>.
62. Goncharov M., Pisov M., Shevtsov A. et al. CT-Based COVID-19 triage: Deep multitask learning improves joint identification and severity quantification // *Med Image Anal.* 2021. №71:102054. DOI: 10.1016/j.media.2021.102054.
63. Jin C., Chen W., Cao Y. et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis // *Nat Commun.* 2020. Vol. 11, №1. P. 5088. DOI:10.1038/s41467-020-18685-1.
64. Kang M., Hong K.S., Chikontwe P. et al. Quantitative Assessment of Chest CT Patterns in COVID-19 and Bacterial Pneumonia Patients: a Deep Learning Perspective // *J Korean Med Sci.* 2021. Vol. 36, №5. P. e46. Published 2021 Feb 1. DOI:10.3346/jkms.2021.36.e46.
65. Kim D.W., Jang H.Y., Kim K.W., Shin Y. et al. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers // *Korean J Radiol.* 2019. Mar. Vol. 20, №3. P. 405–410. URL: <https://DOI.org/10.3348/kjr.2019.0025>.
66. Klassen V.I., Safin A.A., Maltsev A.V. et al. AI-based screening of pulmonary tuberculosis: diagnostic accuracy // *Journal of eHealth Technology and Application.* 2018. T. 16. № S1. C. 28–32.
67. Korevaar D.A., Gopalakrishna G., Cohen J.F. et al. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses // *Diagn Progn.* 2019. Res 3, 22. URL: <https://DOI.org/10.1186/s41512-019-0069-2>.
68. Kulberg N.S., Elizarov A.B., Novik V.P. et al. Automatic segmentation and determining radiodensity of the liver in a large-scale ct database. 31 Dec 2019. arXiv: 1912.13290.
69. Kwee T.C., Kwee R.M. Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence // *Insights Imaging.* 2021. Jun 29. Vol. 12, №1. 88. DOI: 10.1186/s13244-021-01031-4.

70. Lebedev G. The main trends in the development of digital health and their teaching in a medical university // Russian Journal of Cardiology. 2018. Vol. 23. №10. P. 170a–170b.

71. Liu X., Rivera S.C., Moher D. et al. on behalf of the SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension // Lancet Digital Health. 2020. Vol. 2, №10. P. e537–e548.

72. Lu W., Hao W., Chen X. et al. Towards standardized premarket evaluation of computer aided diagnosis/detection products: insights from FDA-approved products // Expert Review of Medical Devices. 2020. Vol. 17, №9. P. 899–918. DOI:10.1080/17434440.2020.181356; ACR: URL: <https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms>.

73. Maslow A.H. The Psychology of Science. New York: Harper&Row, 1966. 168 p.

74. Mintz Y., Brodie R. Introduction to artificial intelligence in medicine // Minim. Invasive Ther. Allied Technol. 2019. Vol. 28, №2. P. 73–81.

75. Morozov S.P., Ranschaert E., Parker W. et al. Is radiology ready for the implementation of artificial intelligence? // An international survey of 1,041 radiologists and residents. Insights into imaging. 2020. №11 (Suppl 2). P. 25. DOI: 10.1186/s13244-020-00849-8

76. Morozov S., Guseva E.B., Ledikhova N.V. et al. Telemedicine-based system for quality management and peer review in radiology // Insights into Imaging. 2018. Vol. 9, №3. P. 337–341.

77. Morozov S.P., Gombolevskiy V.A., Elizarov A.B. et al. A simplified cluster model and a tool adapted for collaborative labeling of lung cancer CT scans // Comput Methods Programs Biomed. 2021. Jul. 206:106111. DOI: 10.1016/j.cmpb.2021.106111.

78. Morozov S.P., Gombolevsky V.A., Vladzimirsky A.V. et al. Incorporating Task-Specific Structural Knowledge into CNNs for Brain Midline Shift Detection. In: Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support (Lecture Notes in Computer Science) – Springer International Publishing, 2019. P.30–38. DOI: 10.1007/978-3-030-33850-3_4.

79. O’Cathain A., Murphy E., Nicholl J. The quality of mixed methods studies in health services research // J Health Serv Res

Policy. 2008. Apr. Vol. 13, №2. P. 92–98. DOI: 10.1258/jhsrp.2007.007074.

80. Pesapane F., Codari M., Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine // Eur Radiol Exp. 2018. Oct 24. Vol. 2, №1. P. 35. DOI: 10.1186/s41747-018-0061-6.

81. Pinnock H., Barwick M., Carpenter C. et al. Taylor SJC for the StaRI Group. Standards for Reporting Implementation Studies (StaRI) statement // BMJ. 2017. №356:i6795. P. 1–9.

82. Pisov M., Kondratenko V., Belyaev M. et al. Keypoints localization for joint vertebra detection and fracture severity quantification. Lecture Notes in Computer Science (см. в книгах). 2020. Vol. 12266 LNCS. P. 723–732.

83. Ramsey C.A., Hewitt A.D. A methodology for assessing sample representativeness // Environ. Forensics. 2005. Vol. 6, №1. P. 71–75.

84. RCR position statement on artificial intelligence // The Royal College of Radiologists. 2018. July. URL: <https://www.rcr.ac.uk/posts/rcr-position-statement-artificial-intelligence>.

85. Seah J., Brady Z., Ewert K., Law M. Artificial intelligence in medical imaging: implications for patient radiation safety // Br J Radiol. 2021. Jun. 23:20210406. DOI: 10.1259/bjr.20210406.

86. Schoonenboom J., Johnson R.B. How to Construct a Mixed Methods Research Design // Kolner Z Soz Sozpsychol. 2017. Vol. 69. Suppl 2. P. 107–131. DOI:10.1007/s11577-017-0454-1.

87. Sharma S. Drawing insights from COVID-19-infected patients using CT scan images and machine learning techniques: a study on 200 patients // Environ Sci Pollut Res Int. 2020. Vol. 27, №29. P. 37155–37163. DOI:10.1007/s11356-020-10133-3.

88. Sounderajah V., Ashrafian H., Golub R.M. et al. STARD-AI Steering Committee. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol // BMJ Open. 2021. Jun. 28:047709. DOI: 10.1136/bmjopen-2020-047709.

89. Strand F., Patel B.K., Allen B. A Call for Controlled Validation Data Sets: Promoting the Safe Introduction of Artificial Intelligence in Breast Imaging. // J Am Coll Radiol. 2021. Jun. 17:S1546–1440 (21) 00463–4. DOI: 10.1016/j.jacr.2021.06.001.

90. Sun L., Yu K., Batmanghelich K. Context Matters: Graph-based Self-supervised Representation Learning for Medical Images. 2020. arXiv:2012.06457v1.
91. Sushentsev N., Bura V., Kotnik M. et al. A head-to-head comparison of the intra- and interobserver agreement of COVID-RADS and CO-RADS grading systems in a population with high estimated prevalence of COVID-19 // *BJR Open*. 2020. Vol. 2, №1. P. 20200053. Published 2020 Dec 11. DOI:10.1259/bjro.20200053.
92. Tadavarthi Y., Vey B., Krupinski E. et al. The State of Radiology AI: Considerations for Purchase Decisions and Current Market Offerings // *Radiology: Artificial Intelligence*. 2020. Vol. 2, №. 6. P. e200004.
93. Tang A., Tam R., Cadrin-Chênevert A. et al. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology // *Can Assoc Radiol J*. 2018. Vol. 69, №2. P. 120–135.
94. Tajaldeen A., Alghamdi S. Evaluation of radiologist's knowledge about the Artificial Intelligence in diagnostic radiology: a survey-based study // *Acta Radiol Open*. 2020. Jul 31. Vol. 9, №7. DOI: 10.1177/2058460120945320.
95. Trajanovski S., Mavroeidis D., Swisher C. L. et al. Towards radiologist-level cancer risk assessment in CT lung screening using deep learning // arXiv:1804.01901v2 [cs.CV] 11 Apr. 2019.
96. Turing A. Computing Machinery and Intelligence // *Mind*. 1950. Vol. 59, №236. P. 433–460.
97. van Leeuwen K.G., de Rooij M., Schalekamp S. et al. How does artificial intelligence in radiology improve efficiency and health outcomes? // *Pediatr Radiol*. 2021. Jun 12. DOI: 10.1007/s00247-021-05114-8.
98. Wei Q., Dunbrack R.L. The Role of Balanced Training and Testing Data Sets for Binary Classifiers in Bioinformatics // *PLoS One*. 2013. July 9.
99. Zhang P., Zhong Y., Deng Y. et al. CoSinGAN: Learning COVID-19 Infection Segmentation from a Single Radiological Image // *Diagnostics (Basel)*. 2020. Vol. 10, №11. P. 901. DOI:10.3390/diagnostics10110901.

ПРИЛОЖЕНИЯ

Приложение 1

ПРОТОКОЛ ФУНКЦИОНАЛЬНОГО ТЕСТИРОВАНИЯ ИИ-СЕРВИСА

№ ____ от ____ . ____ .20 ____

1. Компания-поставщик ИИ-сервиса: _____
2. Наименование ИИ-сервиса: _____
3. Компания-производитель ИИ-сервиса: _____
4. Модальность: _____
Анатомическая область: _____
Клиническая цель: _____
5. Дата получения сведений об успешной интеграции ИИ-Сервиса: _____
Дата проведения испытаний: _____
6. Инструменты тестирования: тестовый программно-аппаратный комплекс (ТПАК), набор данных из ____ исследований
7. Нормативные документы: _____
8. Исходные сведения, предоставленные разработчиком

8.1. Предоставленные данные для тестирования

№	Наименование параметра	Оценочные данные	Комментарий
1	Анализируемая модальность		
2	Определяемые нозологии		
3	Функциональность		
3.1	Возможность приоритезации (триаж)		
3.2	Наличие графического обозначения		
3.3	Наличие дополнительной серии от ИИ-сервиса		
3.4	Название дополнительной серии		
3.5	Возможность синхронизации серий		
3.6	Отображение информации об ИИ-сервисе		
3.7	Отображение вероятности(-ей) находки(-ок)		
3.8	Указание категории находок		
3.9	Возможность отключения маркировки		
3.10	Наличие DICOM SR 1)		

3.11	Наличие DICOM SR 2)		
3.12	Наличие DICOM SR 3)		
3.13	Наличие DICOM SR 4)		
3.14	Наличие DICOM SR 5)		
3.15	Создание шаблона протокола		
3.16	Возможность сравнения исследований в динамике		
3.17	Прочее (указать)		

8.2. Перечень StudyInstanceUID исследований для функционального тестирования

№	Номер исследования	Примечание
1		
2		
3		
4		
5		

9. Результаты проверки заявленных параметров в каждом исследовании по п. 7.2

№	Наименование параметра	Оценочные данные (предоставленные разработчиком)	Результат в соответствии с порядковым номером исследования (согласно п. 7.2)					Комментарий
			1	2	3	4	5	
1	Возможность приоритизации (триаж)							
2	Наличие графического обозначения							
3	Наличие дополнительной серии от ИИ-Сервиса							
4	Название дополнительной серии							
5	Возможность синхронизации серий							
6	Отображение информации об ИИ-сервисе							
7	Отображение вероятности(-ей) находки(-ок)							
8	Указание категории находок							
9	Возможность отключения маркировки							

10	Наличие DICOM SR 1)							
11	Наличие DICOM SR 2)							
12	Наличие DICOM SR 3)							
13	Наличие DICOM SR 4)							
14	Наличие DICOM SR 5)							
15	Создание шаблона протокола							
16	Возможность сравнения исследований в динамике							
17	Прочее (указать)							

Примечание: в случае наличия оцениваемой функции отображается «да», в случае отсутствия функции выставляется «нет», в случае неприменимости данной функции «тире». Исследования маркируются «#» и «N» в зависимости от того, выявлены или нет патологические изменения соответственно в текущем исследовании по данным ИИ-сервиса.

10. Оценка основных критериев

№	Критерии*	Оценка	Комментарии
1	ИИ-сервис в ТПАК выполняет функцию триаж (повышение приоритета исследования с патологической находкой в рабочем списке)		
2	ИИ-сервис в ТПАК не повышает приоритет исследования, если ИИ-сервис не выявил патологию или не смог проанализировать исследование		
3	ИИ-сервис информирует о наличии неполного охвата области интереса на оригинальном изображении		

4	Количество переданных разработчику ИИ-сервиса исследований соответствует количеству проанализированных исследований в ТПАК		
5	Для каждого тестового исследования добавлена по крайней мере одна дополнительная серия, проанализированная ИИ-сервисом, синхронизированная с исходной серией исследования		
6	Серии с результатами ИИ-сервиса содержат информацию о названии, разработке, версии и уведомление «Только в исследовательских целях»		
7	Серии с результатами ИИ-сервиса содержат информацию о показаниях и противопоказаниях к применению результатов работы ИИ-сервиса		
8	Серии с результатами ИИ-сервиса включают в свой состав измененное изображение оригинала (например, температурная карта или локализация патологической находки)		
9	Исследования содержат результат анализа ИИ-Сервиса даже в случае отсутствия патологических находок с информацией об этом на изображении		
10	Серии с результатами ИИ-сервиса включают минимальное необходимое количество релевантных изображений, соответствующих оригинальным		
11	Изображение в серии с результатами ИИ-Сервиса включает в себя все необходимые для интерпретации исследования области интереса из оригинального исследования		

12	В серии с результатами ИИ-сервиса, присутствует информация о дате и времени завершения анализа		
13	Отсутствие грубых ошибок ИИ-сервиса (например, в результатах ИИ-сервиса локализация патологии находится вне области интереса)		
14	ИИ-сервис не локализует явные патологические находки (например, ИИ-сервис не должен пропустить легочный узел, подозрительный на рак легкого при анализе КТ)		
15	Другие критические несоответствия к работе ИИ-сервис не выявлены		

Примечание: * критерии 1-3 являются – некритическими, 4-15 – критическими

11. Вывод

По результатам промежуточного тестирования функционала ИИ-сервиса выявлены несоответствия:

11.1. Критические: _____

11.2. Некритические: _____

12. Рекомендации: _____

13. Заключение: _____

Тестирование провели:

Должность специалиста

ФИО специалиста

ПРИЛОЖЕНИЕ К ПРОТОКОЛУ ФУНКЦИОНАЛЬНОГО ТЕСТИРОВАНИЯ ИИ-СЕРВИСА

Приложение 2

ПРОТОКОЛ КАЛИБРОВОЧНОГО ТЕСТИРОВАНИЯ ИИ-СЕРВИСА

№ ____ от _____.20__ г.

1. Компания-поставщик ИИ-сервиса: _____
2. Наименование ИИ-сервиса: _____
3. Компания-производитель ИИ-сервиса: _____
4. Модальность: _____
Анатомическая область: _____
Клиническая цель: _____
5. Дата получения результатов калибровки ИИ-сервиса: _____
Дата анализа: _____
6. Инструменты калибровки: _____

тестовый программно-аппаратный комплекс (ТПАК), набор данных из ____ исследований (согласно базовым требованиям к результатам работы ИИ-сервисов) для _____ (вид исследования).

7. Нормативные документы: _____
8. Контролируемые параметры и характеристики: _____

8.1. Метрики точности и время анализа исследований

	Наименование	Численное значение	Ученое значение на эталонном наборе данных		Примечание
			Метод определения оптимального порога		
			индекс Юдена (YI)	максимизации прогностической ценности отрицательного результата (max NPV)	
1	Значение площади под ROC-кривой (AUC)				
2	точность				

3	чувствительность				
4	специфичность				
5	Удельный вес ложноотрицательных результатов (%)				
6	Удельный вес ложноположительных результатов (%)				
7	Оптимальный порог				
8	Время анализа одного исследования, с				
9	Время, затрачиваемое на принятие исследования, его анализ и передачу результатов анализа по данным ЕРИС, с				

8.1.1. Оценка результатов

№	Наименование	Нормативное значение	Метод определения оптимального порога		Результат
			индекс Юдена (YI)	максимизации прогностической ценности отрицательного результата (max NPV)	
1	Значение площади под ROC-кривой (AUC), %	>0,81			
2	Δ AUC, %	≥ -10			
3	Δ точности, %	≥ -10			
4	Δ чувствительности, %	≥ -10			
5	Δ специфичности, %	≥ -10			
6	Время, затрачиваемое на принятие исследования, его анализ и передачу результатов анализа по данным ЕРИС, минут	≤ 10			
7	Удельный вес успешно проанализированных исследований, %	≥ 90			

Δ – Относительная разница значения.

9. Выводы и рекомендации.

Калибровочное тестирование провели:

Должность специалиста

ФИО специалиста

Приложение 3

ПРИЛОЖЕНИЕ К ПРОТОКОЛУ КАЛИБРОВОЧНОГО ТЕСТИРОВАНИЯ ИИ-СЕРВИСА

ОТЧЕТ О КОНТРОЛЬНО-ТЕХНИЧЕСКОМ ТЕСТИРОВАНИИ

№ ____ от ____ г.

1. Наименование организации, проводившей тестирование: ГБУЗ «НПКЦ ДиТ ДЗМ»

2. Компания-поставщик ИИ-сервиса: _____

3. Наименование ИИ-сервиса: _____

4. Компания-производитель ИИ-сервиса

5. Модальность: _____

Анатомическая область: _____

Клиническая цель: _____

6. Дата получения результатов анализа набора исследований в количестве 2 единиц с каждого подключенного к ИИ-сервису диагностического устройства:

Дата проведения тестирования:

7. Инструменты тестирования: продуктивный программно-аппаратный комплекс (ППАК), _____ исследований

8. Нормативные документы: Постановление Правительства Москвы от _____, Приказ ДЗМ от _____.

9. Основание для проведения КТТ: _____

10. Перечень ae-title ДУ (т.е. уникальных идентификаторов диагностических устройств в ЕРИС.ЕМИАС) и Study_UID исследований (т.е. уникальных идентификаторов исследований в ЕРИС.ЕМИАС) для тестирования.

Таблица 1. Список ДУ и исследований, по которым проводятся тестирования

№	ae-title ДУ	study_uid	Медицинская организация
1	2	3	4
1.			

Детализация проверенного функционала ИИИ-сервиса по каждому исследованию приведена в Таблице 2.

Таблица 2. Результаты проверки наличия функционала ИИИ-сервиса в ППАК по каждому тестируемому исследованию

№	Наименование параметра	Оценочные данные (предоставленные разработчиком)	Результат в соответствии с уникальным идентификатором исследования (согласно п. 9)																Комментарий
1	Возможность привязки и (триажа)																		
2	Наличие графического обозначения																		
3	Наличие дополнительной серии ИИИ																		
4	Наличие дополнительной серии																		

12. Приложение

Приложение 1. Список ДУ, в исследовании от которых отсутствуют технологические дефекты

№	ae-title ДУ
1	2
1.	

Приложение 2. Список ДУ, в части исследований от которых присутствуют технологические дефекты

№	ae-title ДУ	study_uid	Комментарии
1	2	3	4
1			

Приложение 3. Список ДУ, во всех исследованиях от которых присутствуют технологические дефекты

№	ae-title ДУ	study_uid	Комментарии
1	2	3	4
1			

Приложение 4

**Независимый этический Комитет
Московского регионального отделения
Российского общества рентгенологов и радиологов**

г. Москва, ул. Средняя Калитниковская, д. 28, стр. 1; e-mail: ethics@npcmr.ru

**Выписка из протокола №2 (1-II-2020)
НЭК МРО РОРР от 20 февраля 2020г.**

Присутствовали: Агафонова О. А., Бахтеева Е. Г., Лайпан А. Ш., Комолов И. С., Мокиенко О. А., Наделяева И. И., Низовцова Л. А., Омелянская О. В., Петрайкин А. В.

Слушали: Экспертиза материалов планируемого научного исследования «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы»

Представленные документы:

1. Аннотация исследования «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы»;
2. Сопроводительное письмо на имя председателя НЭК МРО РОРР для проведения этической экспертизы исследования «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы»;
3. CV главного исследователя
4. Информационный листок для пациента и информированное согласие,
5. Информация о проекте.

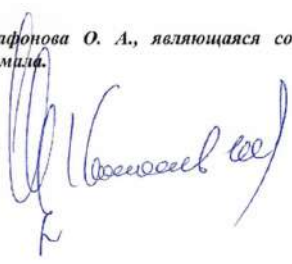
Постановили: одобрить планируемое научное исследование «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы»

В связи с конфликтом интересов Агафонова О. А., являющаяся соисполнителем проекта, участие в голосовании не принимала.

Выписка выдана:

Ответственный секретарь НЭК МРО РОРР:

**Независимый
этический комитет
МРО РОРР**
ethics@npcmr.ru
+7(495)2760436



Приложение 5

ИНФОРМАЦИОННЫЙ ЛИСТОК ДЛЯ УЧАСТНИКА НАУЧНОГО ИССЛЕДОВАНИЯ

*«Эксперимент по использованию инновационных технологий в области
компьютерного зрения для анализа медицинских изображений и дальнейшего
применения в системе здравоохранения города Москвы»*

Уважаемый пациент!

Предлагаем Вам ознакомиться с информацией о научном исследовании по изучению возможности использования искусственного интеллекта (ИИ) как вспомогательного инструмента при проведении лучевых исследований.

Департамент здравоохранения города Москвы (ДЗМ) в рамках научного исследования будет проводить оценку возможности использования сервисов на основе технологий компьютерного зрения и искусственного интеллекта для обработки лучевых изображений, получаемых при проведении диагностических и профилактических процедур в отделениях лучевой диагностики медицинских организаций ДЗМ г. Москвы.

Пожалуйста:

1. Внимательно прочтите данную информацию.
2. Задайте столько вопросов врачу, сколько Вам необходимо для понимания того, что включает в себя Ваше участие в исследовании.

Если Вы согласитесь принять участие в данном научном исследовании, Вам необходимо подписать два экземпляра информированного согласия.

Для чего проводится это исследование?

Данное научное исследование будет проводиться в нескольких медицинских учреждениях ДЗМ г. Москвы в целях оценки возможности использования различных сервисов на основе компьютерного зрения и искусственного интеллекта для анализа изображений, полученных при проведении диагностических и профилактических процедур по компьютерной томографии (КТ), рентгенографии, маммографии, флюорографии.

Врач-рентгенолог медицинской организации будет проводить описание полученных изображений/снимков участников, а также оценку результатов анализа этих изображений (предварительно обезличенных) сервисами на основе компьютерного зрения и искусственного интеллекта, которые потом будут сравниваться и перепроверяться экспертами Государственного бюджетного учреждения здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы» (ГБУЗ «НПКЦ ДиТ ДЗМ»), уполномоченной специализированной организацией (<http://medradiology.moscow/>).

Процедуры и мероприятия

Все результаты полученных исследований, в случае выявления подозрительных явлений, будут проходить процедуру экспертной оценки («второе мнение») экспертами ГБУЗ «НПКЦ ДиТ ДЗМ» для подтверждения качества проведённой оценки и выбора наиболее эффективных сервисов на основе компьютерного зрения и искусственного интеллекта. Для пациента процедура экспертной оценки является бесплатной.

Всем участникам научного эксперимента, ожидаемая численность которых составит $\approx 500\,000$ человек, в зависимости от результатов исследования будет предложен индивидуальный план дальнейшего сопровождения: регулярные профилактические осмотры или наблюдение и лечение у профильного специалиста в связи с возможным установленным диагнозом, согласно стандартной тактике ведения пациентов.

Что произойдет с информацией, полученной в результате проведения данного исследования?

Информация, полученная в результате проведения данного исследования, будет использоваться в целях научных исследований при разработке методов диагностики и лечения пациентов. Ваши обезличенные данные будут храниться как в печатном, так и в электронном виде. При опубликовании какой-либо информации о результатах проведенного исследования, Ваше имя нигде упоминаться не будет. Подписывая данный документ, Вы даете согласие на такое использование информации, собранной в ходе Вашего участия в исследовании.

Контролирующие органы, члены комитетов по этике и сотрудники экспертных организаций могут проверять информацию, собранную в ходе Вашего исследования, и сверять ее с Вашей медицинской документацией. Сведения в Вашей медицинской документации не подлежат разглашению, кроме случаев, предусмотренных действующим законодательством. Подписывая данный документ, Вы понимаете, что разрешаете уполномоченным лицам прямой доступ к своей медицинской документации.

Обработка Ваших персональных данных будет осуществляться с соблюдением требований законодательства Российской Федерации в области персональных данных и с соблюдением врачебной тайны.

Что будет, если Вы передумаете?

Ваше участие в научном исследовании является осознанным и добровольным. Вам необязательно участвовать в этом исследовании. Вы можете принять в нем участие, а потом передумать. Ваше решение об участии не повлияет на качество оказываемой Вам медицинской помощи и не приведет к нарушению Ваших прав.

Какая информация медицинского характера будет использована в исследовании?

- Сведения из медицинской документации, результаты медицинских осмотров и других процедур, которые необходимы для проведения настоящего исследования.
- Информация, полученная или зарегистрированная в ходе Вашего участия в настоящем исследовании, включая результаты анализов, процедур и обследований, предусмотренных исключительно для настоящего исследования.
- Вы имеете право получить в отношении себя всю собранную информацию в результате данного исследования

Для получения дополнительной информации Вы можете обратиться за информацией на сайт <https://mosmed.ai/> или задать вопрос главному рентгенологу г. Москвы <http://medradiology.moscow/>; телефон: +7 (495) 276-04-36.

**Форма информированного согласия для участника научного исследования
«Эксперимент по использованию инновационных технологий в области
компьютерного зрения для анализа медицинских изображений и
дальнейшего применения в системе здравоохранения города Москвы»**

- Я прочитал(-а) и понял(-а) информацию, предоставленную в данном документе о настоящем медицинском научном исследовании. У меня было достаточно времени для принятия решения. Врач и/или персонал, участвующий в данном исследовании, объяснили мне неясную информацию и полностью ответили на все мои вопросы.
- Настоящим подтверждаю, что я подробно проинформирован(-а) о моём участии в исследовании по использованию методов компьютерного зрения и искусственного интеллекта в описании медицинских изображений
- Я добровольно согласен(-на) принять участие в настоящем исследовании
- Я разрешаю врачу - исследователю и/или команде исследователей обработку моих медицинских данных в целях настоящего исследования, а также передачу этих данных в обезличенном виде.
- Я согласен(-на) предоставить доступ к моей медицинской документации официальным уполномоченным лицам при условии соблюдения правил конфиденциальности персональных данных.
- Мне предоставили подписанный и датированный экземпляр информационного листка пациента и формы информированного согласия на участие в исследовании. Я понимаю, что запись о моем согласии на участие в исследовании будет храниться в моей медицинской карте.

Фамилия, Имя, Отчество пациента (напишите полностью и разборчиво)

Подпись

Дата

Фамилия, Имя, Отчество лица, получившего согласие (напишите полностью и разборчиво)

Подпись

Дата

Приложение 6



Москва внедряет новые технологии в медицине:

Пройдите комплексное обследование на заболевание легких, онкологию и COVID-19 с использованием «компьютерного зрения»

На помощь врачам в диагностике заболеваний приходят **«искусственный интеллект»** и **«компьютерное зрение»**.

Благодаря новым технологиям появилась возможность значительно сократить время на диагностику и улучшить ее качество.

Внесите свой вклад в развитие инновационных технологий в медицине!



Для этого достаточно пройти назначенное врачом обследование с применением технологий «искусственного интеллекта»*



Диагноз мне поставит компьютер?
Окончательный диагноз может поставить только врач. Но использование технологий «компьютерного зрения» и «искусственного интеллекта» помогает врачу сделать диагностику более качественной.

*Алгоритмы «искусственного интеллекта» применяются для обработки и анализа медицинских изображений в рамках научно-практического эксперимента.

Чем мне это полезно?

Все участники получают индивидуальные планы дальнейшего наблюдения. Это профилактические осмотры или наблюдение и лечение у профильного специалиста.

Для проведения обследования необходимо получить направление у вашего лечащего врача.

Что дает этот эксперимент врачам?

Вы можете помочь московским врачам в изучении возможностей технологий «искусственного интеллекта» с целью:

- | | | |
|--|---|---|
| ⁰¹ повышения
точности работы
врача-рентгенолога | ⁰² повышения качества
всех видов лучевой
диагностики | ⁰³ сокращения
сроков постановки
диагноза |
|--|---|---|

Где это применяется?

-  ⁰¹ Низкодозная компьютерная томография органов грудной клетки (НДКТ) для диагностики рака легкого и КТ для диагностики COVID-19.
-  ⁰² Рентгенография/флюорография органов грудной клетки (РГ/ФГ) для выявления заболеваний легких.
-  ⁰³ Маммография (ММГ) для раннего выявления рака молочной железы.

Если у вас остались вопросы, задайте их на сайте:
<http://medradiology.moscow> или mosmed.ai

Опрос:

Пожалуйста, пройдите анонимный опрос.
Нам важно понять, что пациенты знают
об «искусственном интеллекте»
и как относятся к его применению в медицине.



ЦЕНТР ДИАГНОСТИКИ
И ТЕЛЕМЕДИЦИНЫ

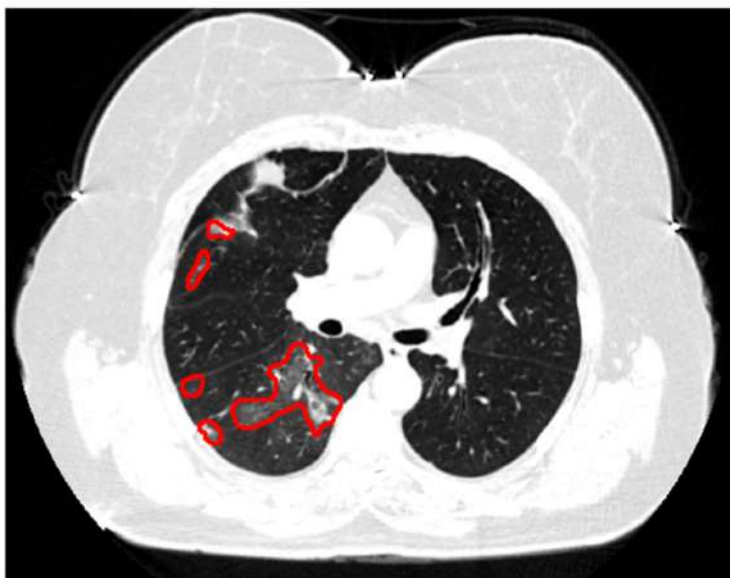


КОМПЛЕКС
СОЦИАЛЬНОГО
РАЗВИТИЯ
МОСКВЫ

Приложение 7

1. Virtual Resident (Radlogics) (КТ Covid)

ИП: Уплотнения по типу «матового стекла» в нижней и средней долях правого легкого.



ЛП: купол диафрагмы неверно размечен как патологическое уплотнение

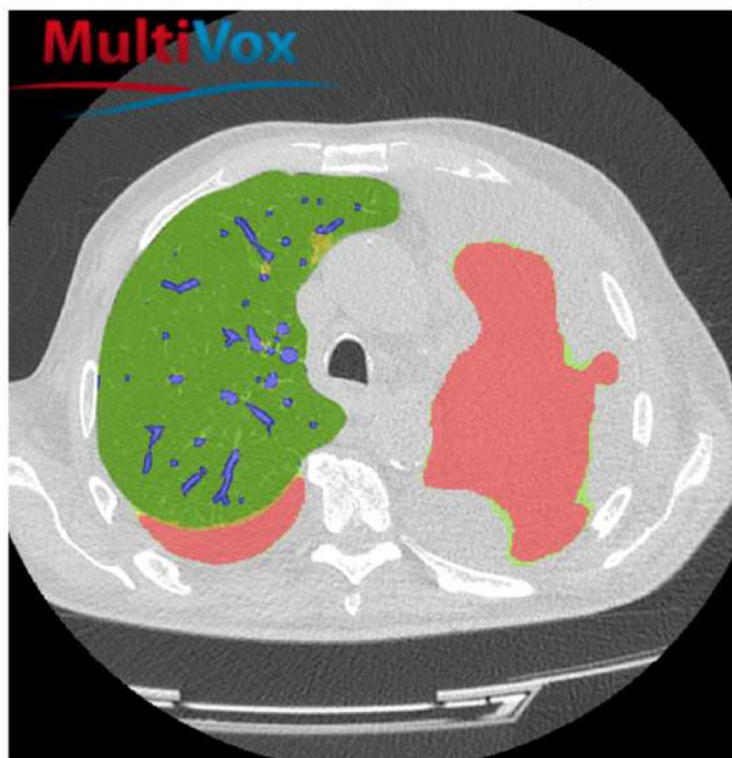


2. Гамма Мультивокс Ковирус (КТ Covid)

ИП: цветовая маркировка уплотнений по типу "матового стекла" с элементами консолидации в нижних долях легких.

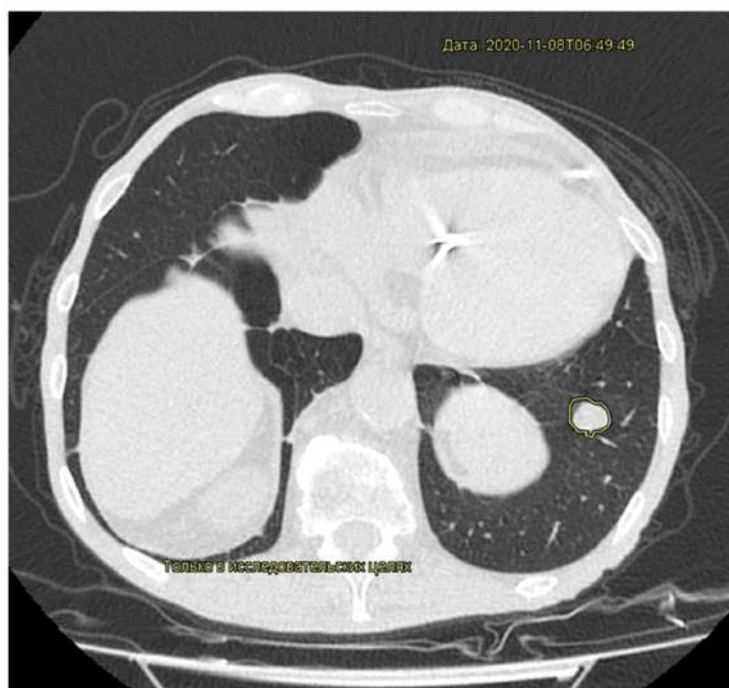


ЛП: разметка в брюшной полости – за пределами целевого органа

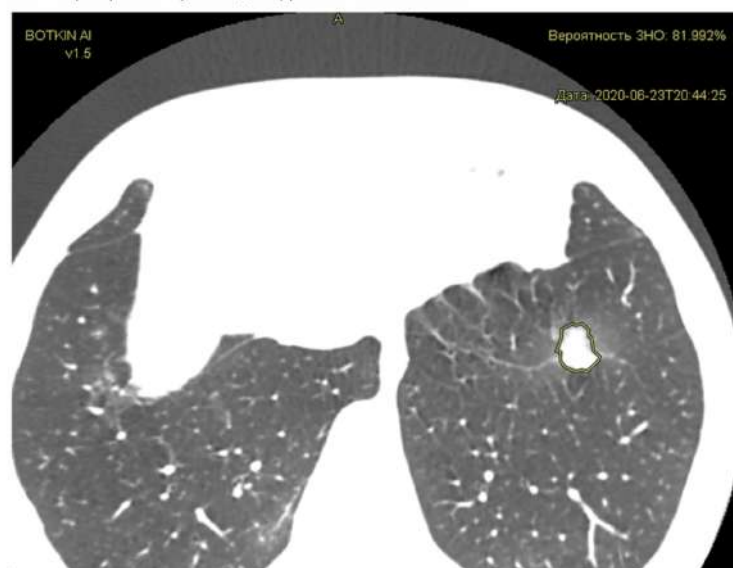


3. Botkin.AI (КТ РЛ)

ИП: узловое образование нижней доли левого легкого

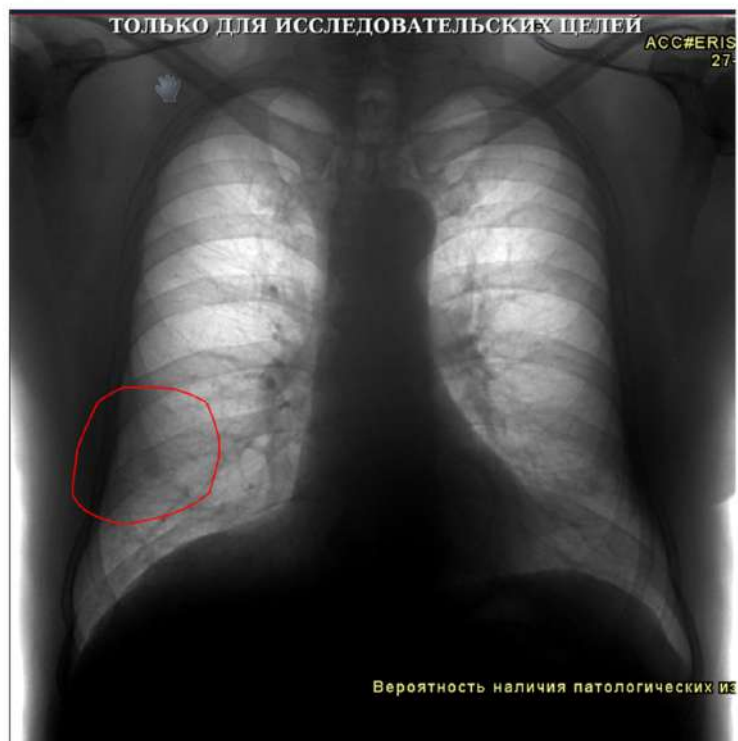


ЛП: маркировка купола диафрагмы со спайками

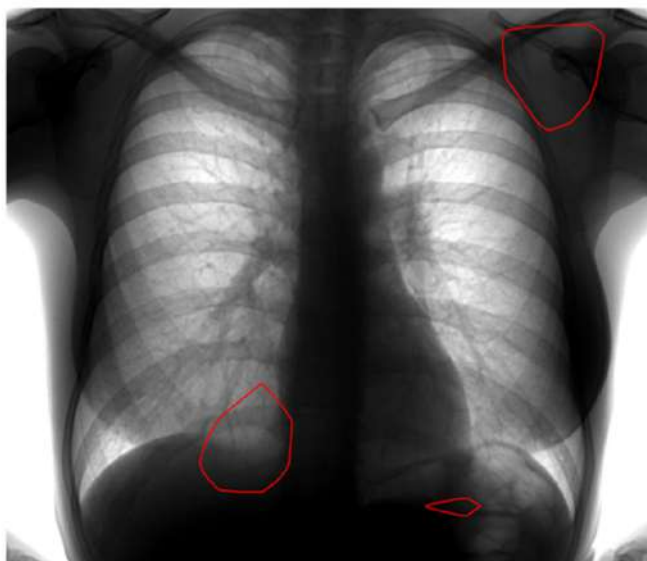


4. Цельс (ФЛГ)

ИП: участок инфильтрации на границе нижней и средней долей правого легкого

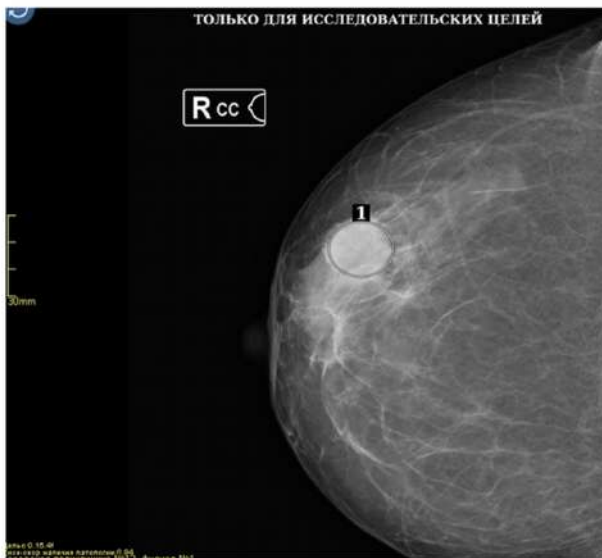


ЛП: маркировка за пределами целевого органа (плечевой пояс, газовый пузырь желудка, мягкие ткани грудной стенки). Область маркировки в прикорневых отделах правого легкого не содержит патологических затенений.

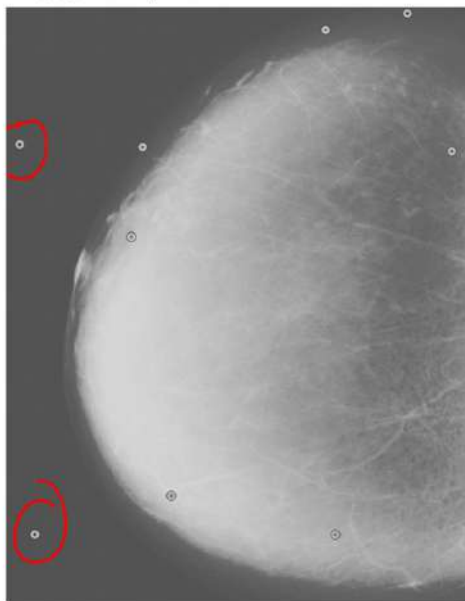


4. Цельс (ММГ)

ИП: образование ретроареолярной области правой молочной железы

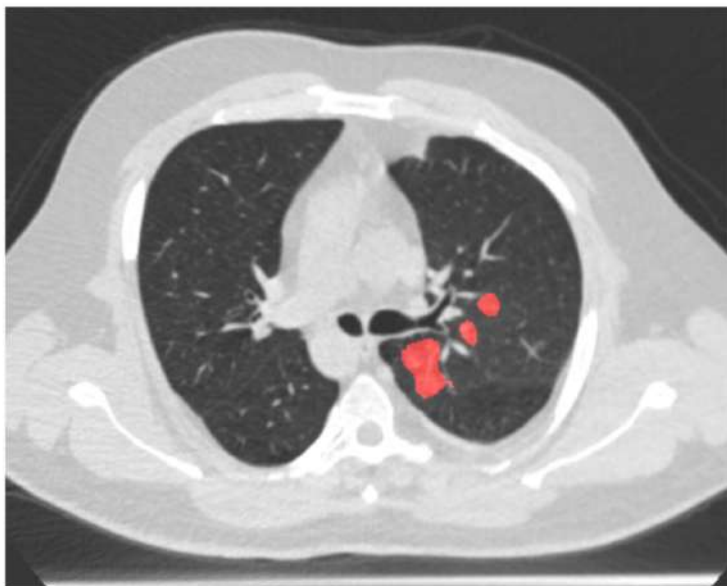


ЛП: разметка за пределами органа

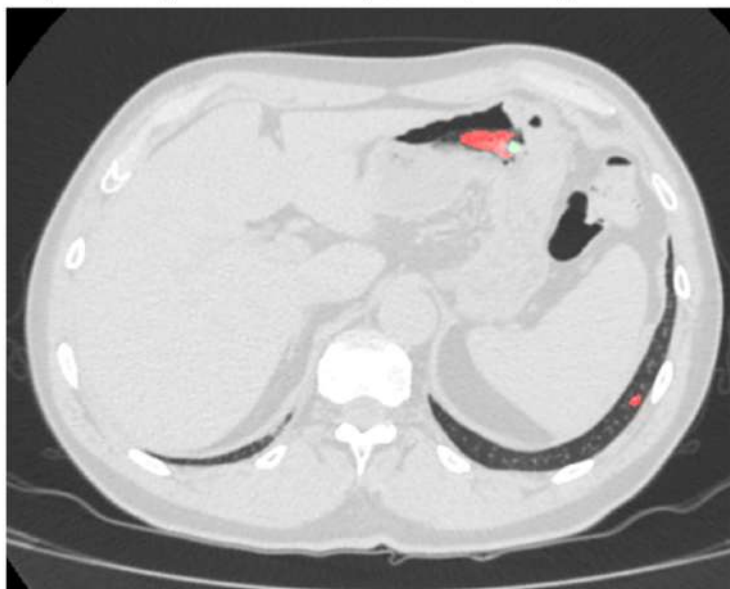


6. CarementorAI (КТ Covid)

ИП: мелкие участки инфильтрации по типу «матового стекла»

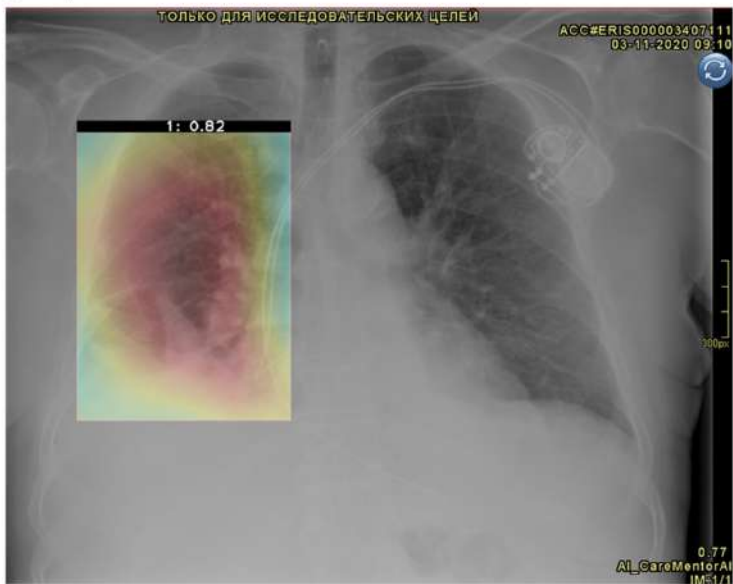


ЛП: разметка в просвете кишки – за пределами целевого органа

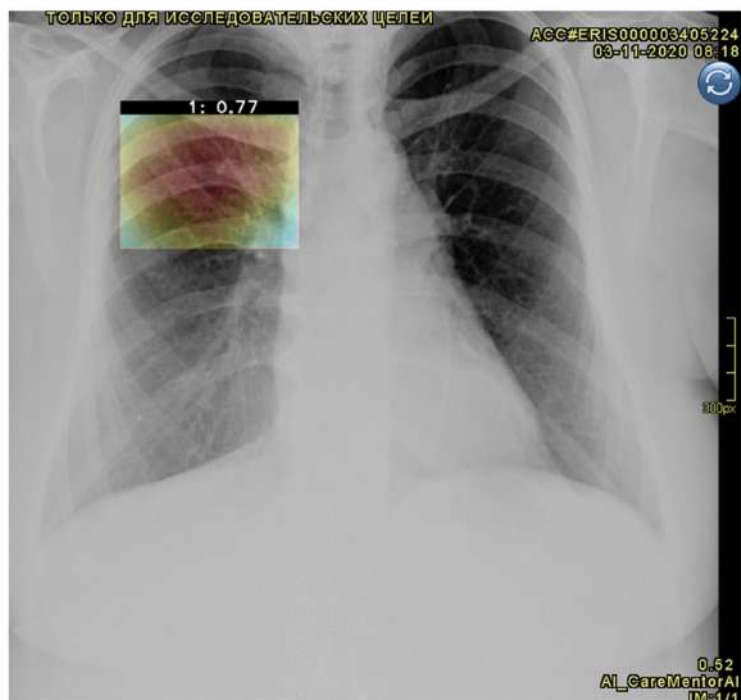


7. CarementorAI (РГ)

ИП: правосторонний гидроторакс и патологическая инфильтрация в нижней доле правого легкого

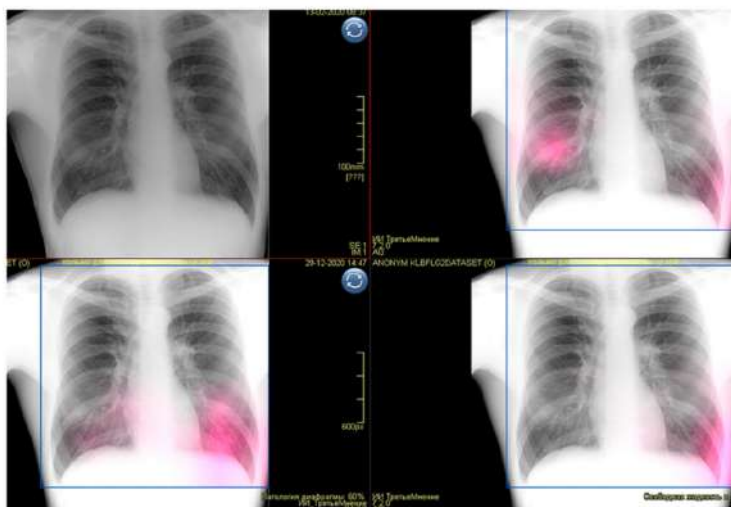


ЛП: Область маркировки в прикорневых отделах правого легкого не содержит патологических затенений.

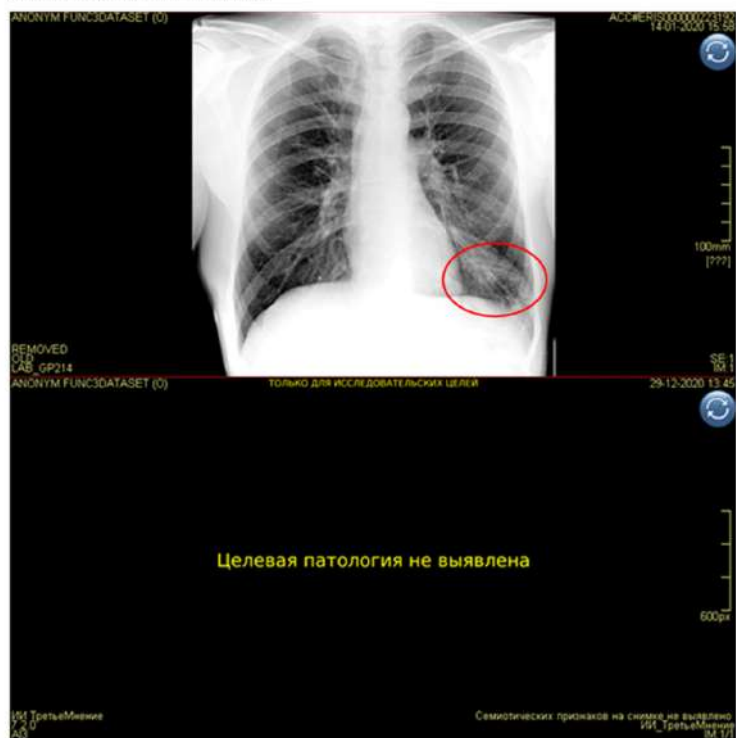


8. Третье Мнение (РГ)

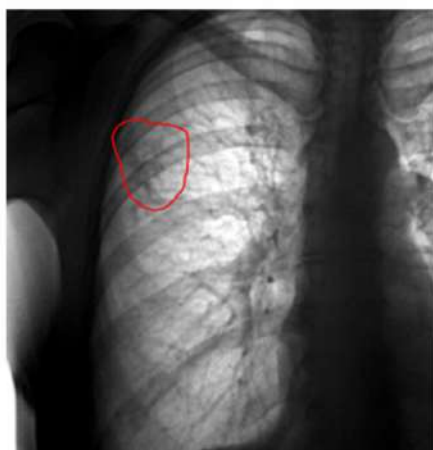
ЛП: Область маркировки не содержит патологических затенений: контур лопатки и контур ребра размечены как патологическое затенение (перелом и патологический процесс диафрагмы), изменений в плевральной полости нет



ЛО: затенение левого легкого, которое может соответствовать частичному ателектазу или пневмонии.

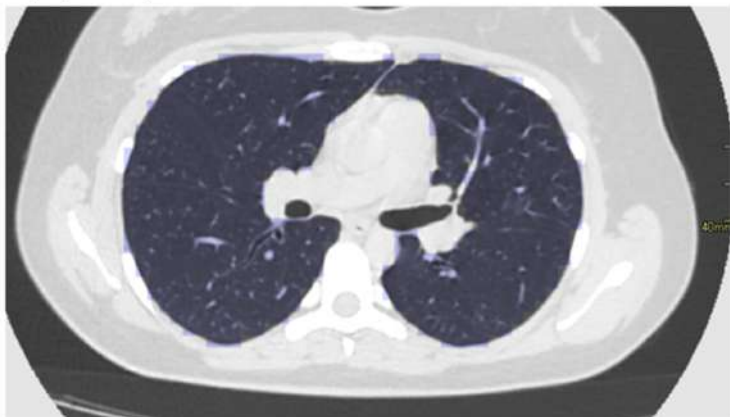


ИП: Очаг или инфильтрация в субплевральных отделах верхней доли правого легкого – верная локализация

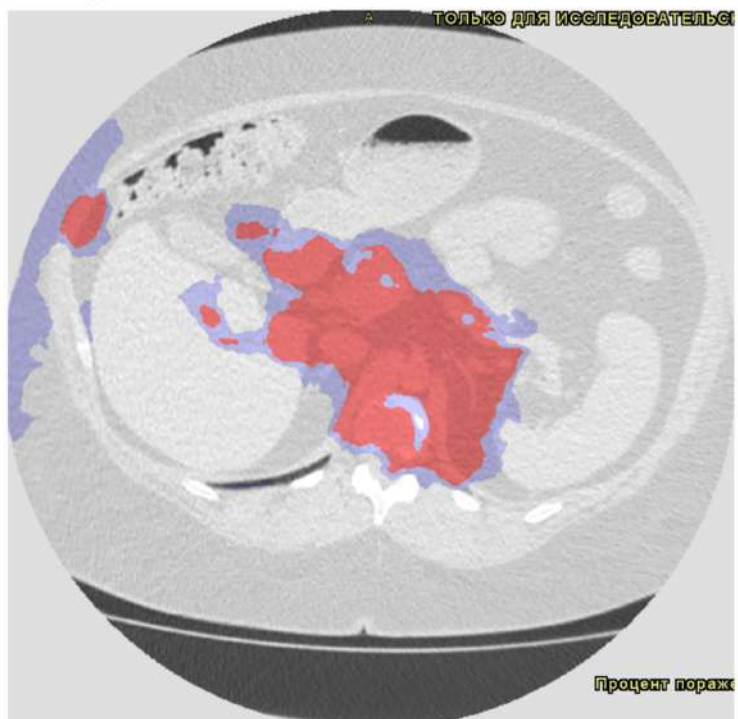


9. Третье Мнение. КТ-COVID-19 (КТ Covid)

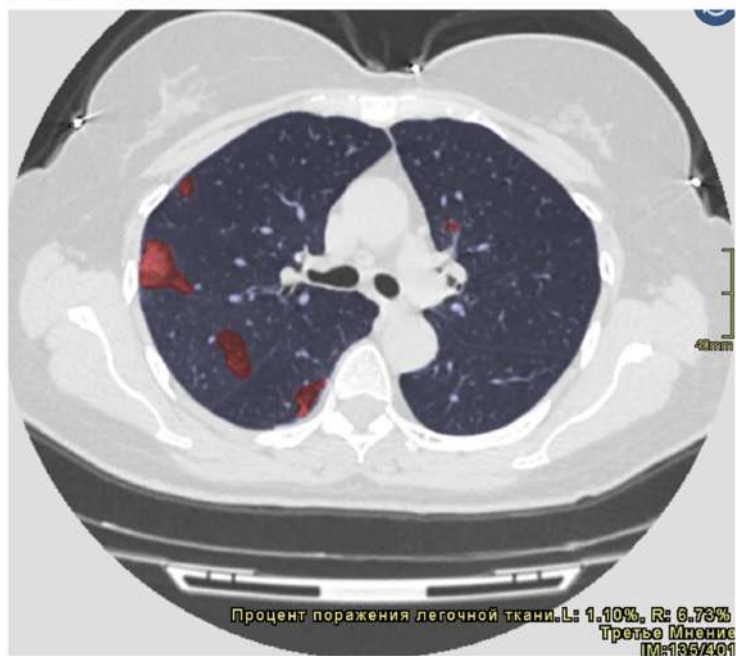
ЛО: Слабоинтенсивное затенение в верхней доле левого легкого не маркировано при анализе ИИ



ЛП: маркировка вне органов грудной клетки, захватывающая области в подкожной клетчатке, позвоночник, внутренние органы (область малого сальника)

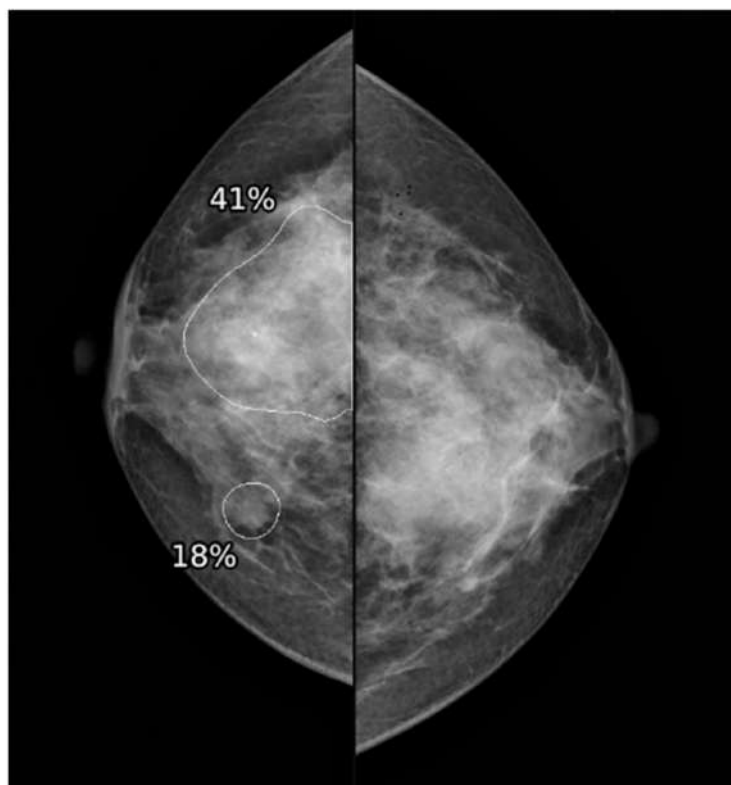


ИП: патологические уплотнения по типу «матового стекла» в субплевральных отделах правого легкого

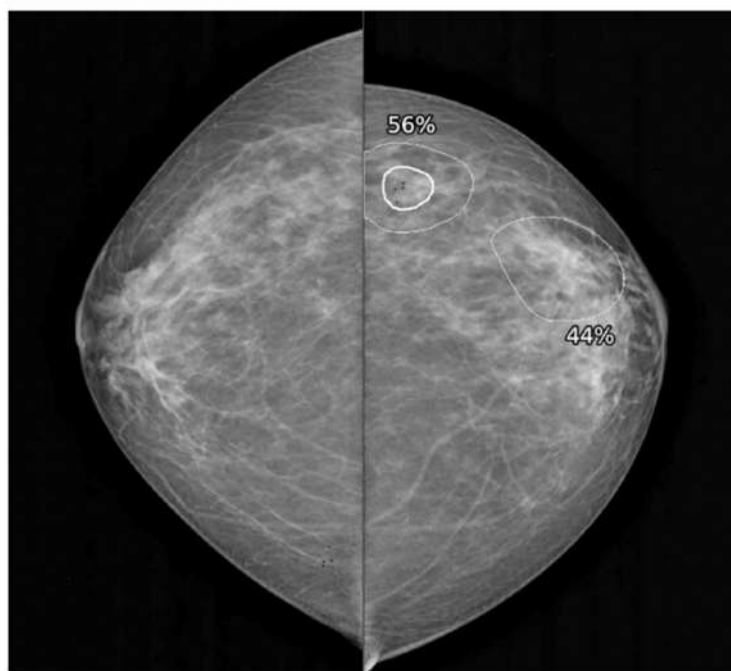


10. Lunit INSIGHT MMG (ММГ)

ИП: массивное уплотнение в правой молочной железе



ЛП: участки фиброзно-железистой стромы неверно классифицированы как патологический процесс

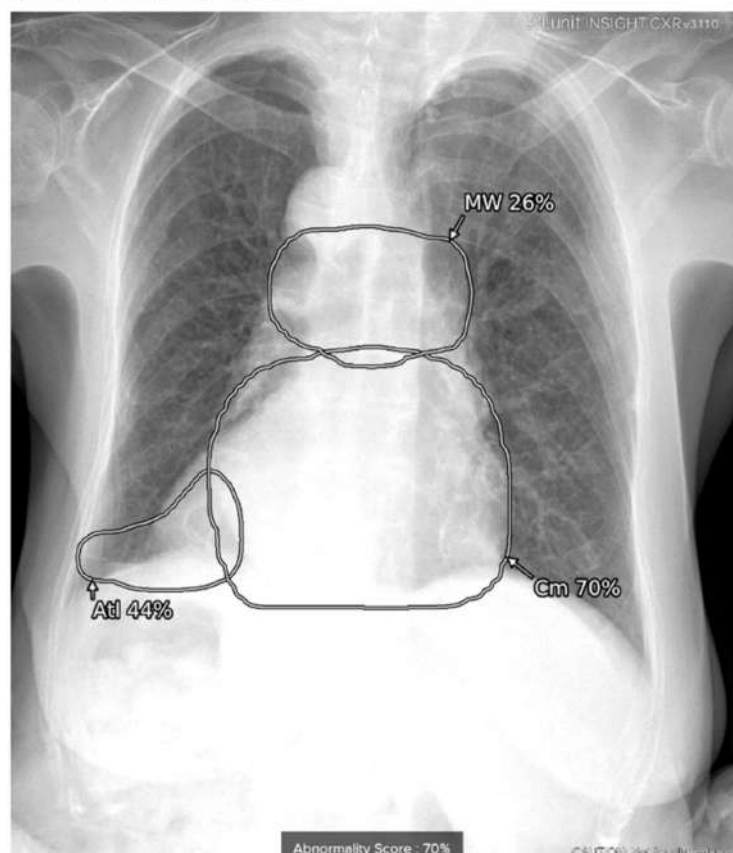


11. Lunit INSIGHT CXR (PT)

ЛП: подозрение на правосторонний пневмоторакс не подтвердилось, паренхима легких излишне прозрачна, но сосудистый рисунок в зоне маркировки прослеживается

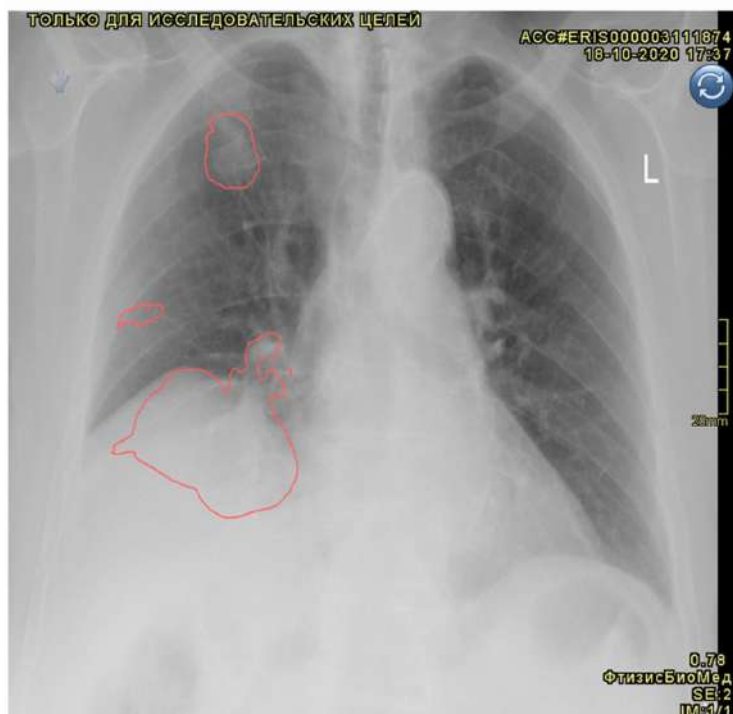


ИП: расширение тени средостения, расширение тени сердца, ателектаз средней доли правого легкого

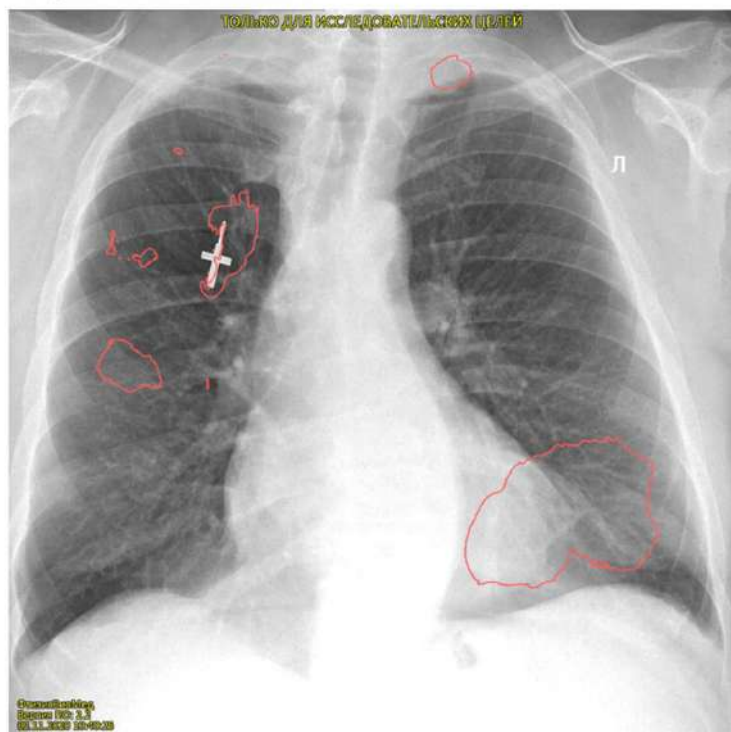


12. FBM (РГ)

ИП: образование верхней доли правого легкого выявлено корректно, высокое стояние правого купола диафрагмы размечено в рамках патологического затенения



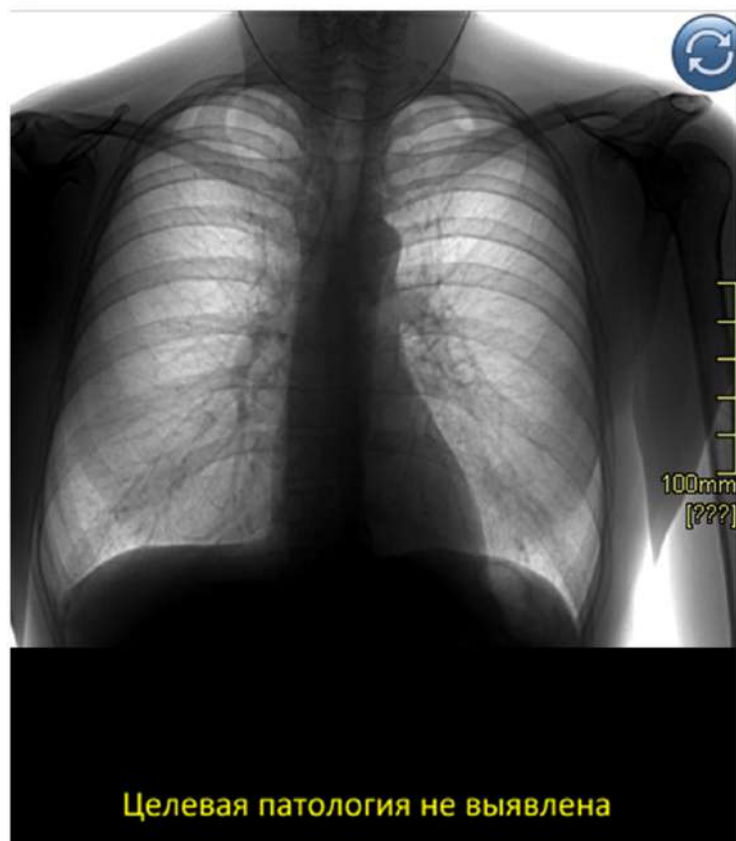
ЛП: неверная разметка ювелирного изделия, заднего отрезка ребра на уровне вершины левого легкого; патологических теней на фоне тени сердца и в правом легком не выявлено



ЛО: Очаг в базальных отделах нижней доли левого легкого не маркирован

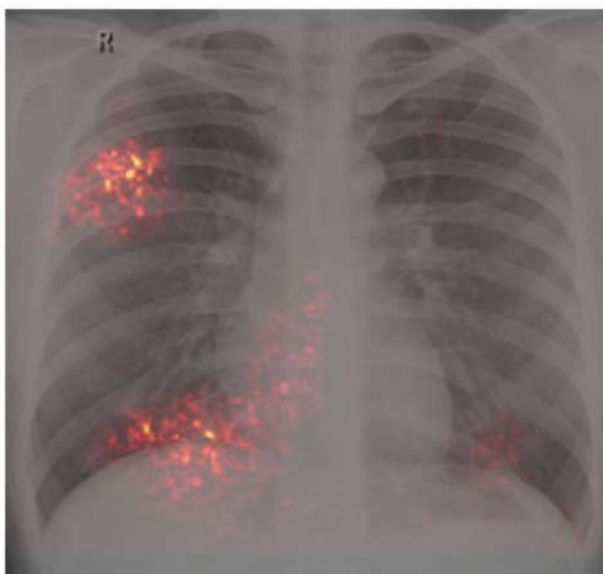


ИП:

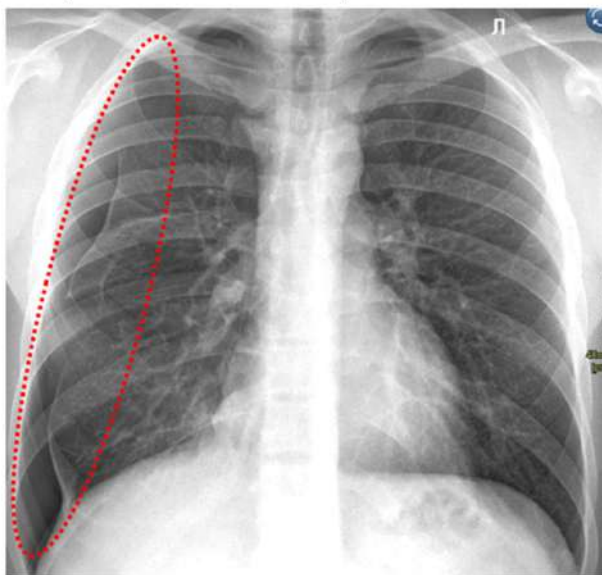


14. AI RADIOLOGY CXR (PI)

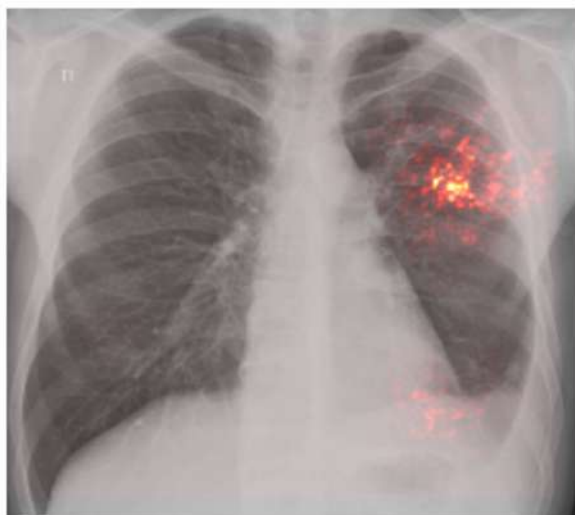
ЛП



ЛО: правосторонний массивный пневмоторакс не выявлен

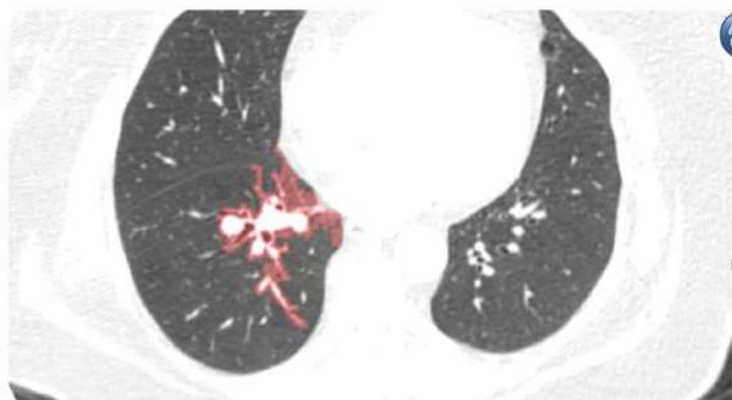
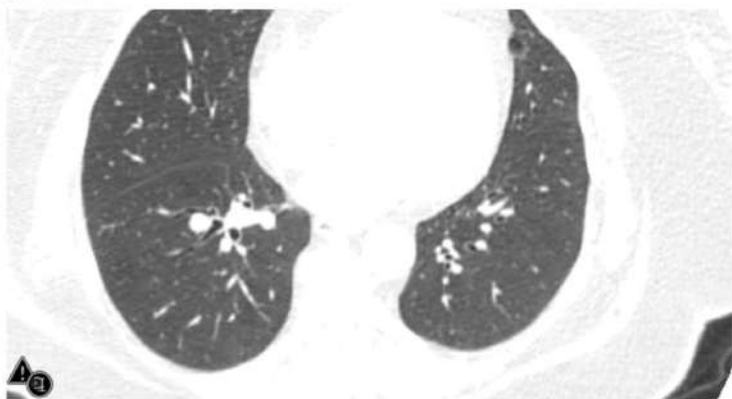


ИП: тядистые затенения в верхней доле левого легкого выявлены корректно, подозрительны на туберкулезный процесс; подозрение на левосторонний гидроторакс

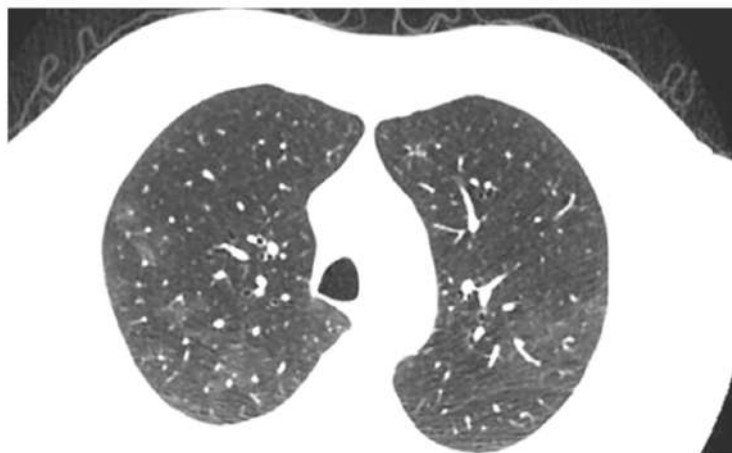
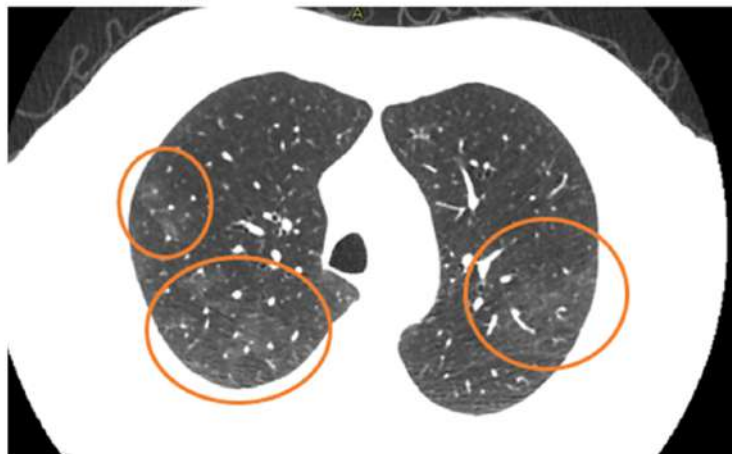


15. IRYM COVIDetect (KT Covid)

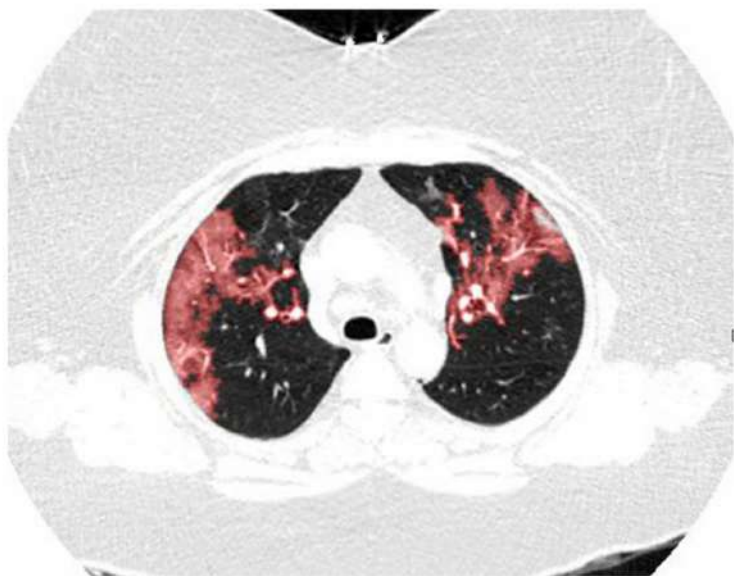
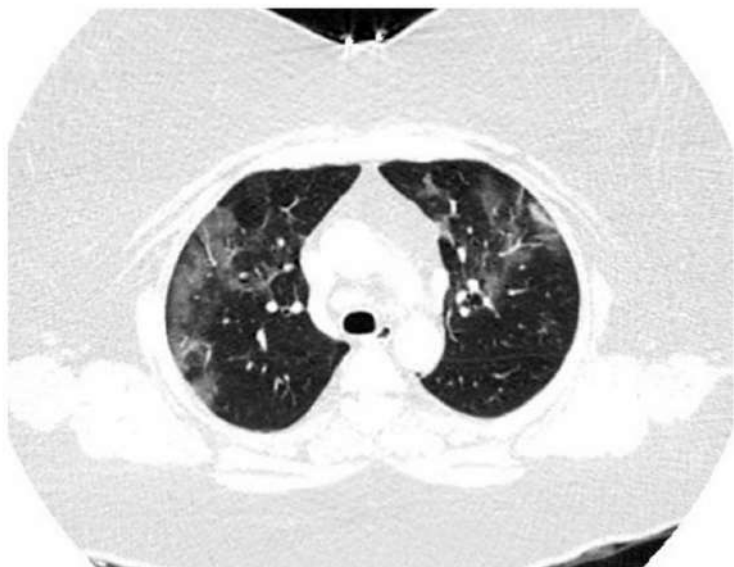
ЛП: неверная разметка сосудов легкого в качестве патологического уплотнения



ЛО: не выявлены крупные зоны «матового стекла» малой интенсивности

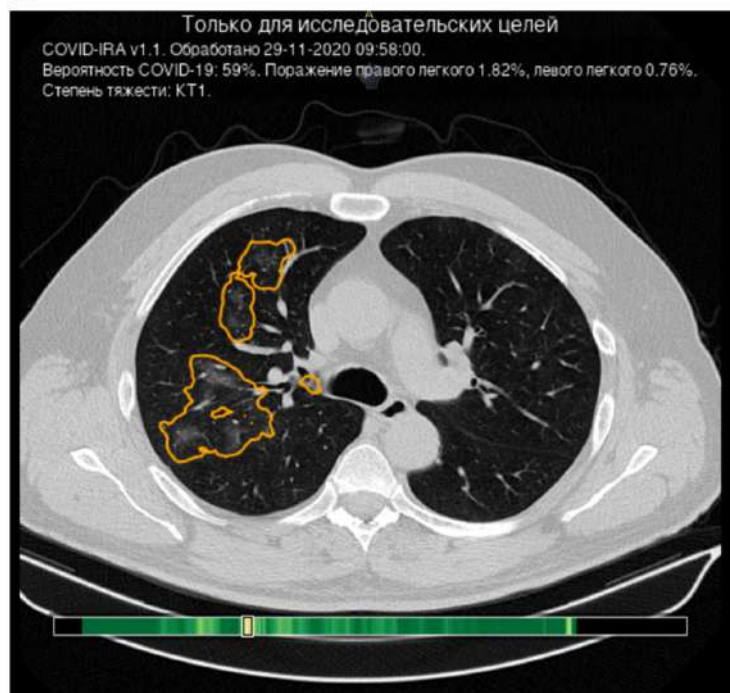


ИП

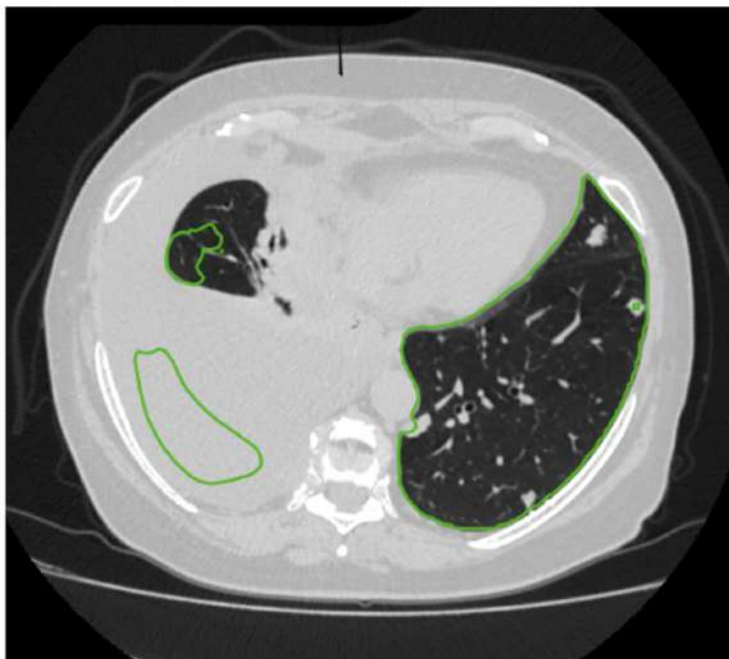


16. COVID-19 Intelligent Radiology Assistant (KT Covid)

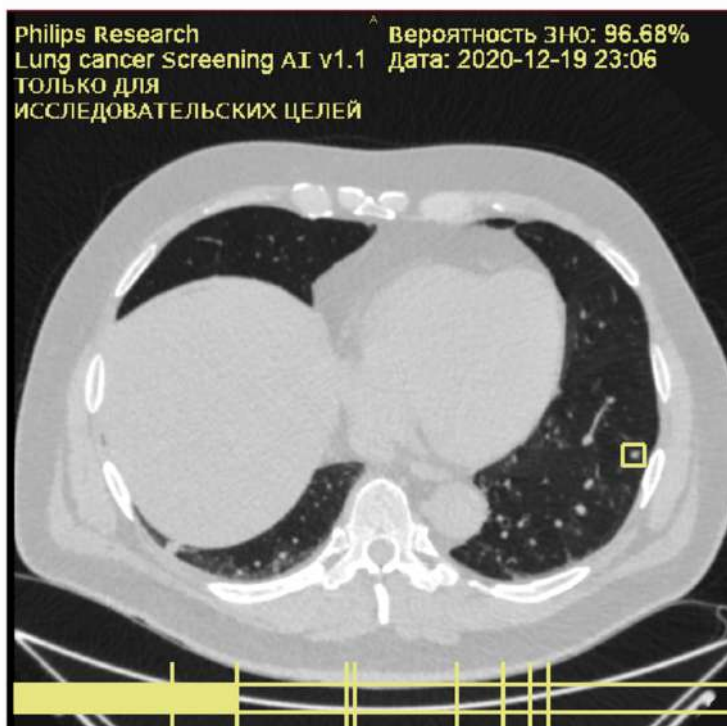
ИП



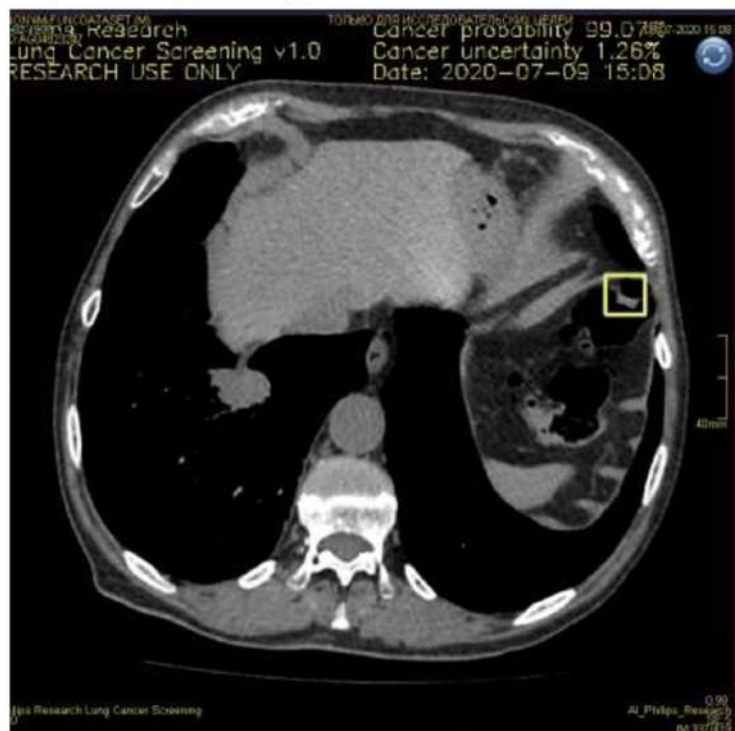
ЛП: разметка за пределами легкого – в паренхиме печени



17. Philips Research Lung Cancer Screening (НДКТ РЛ)
ИП

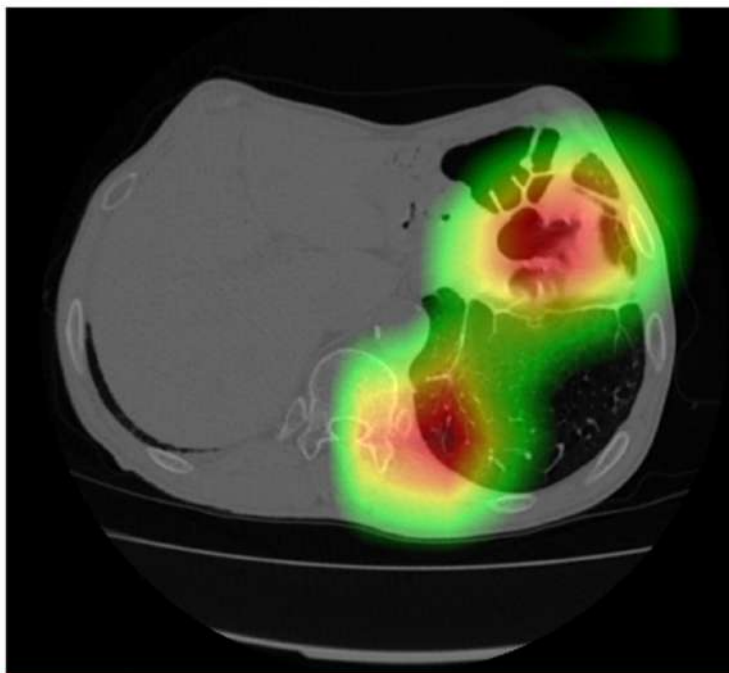


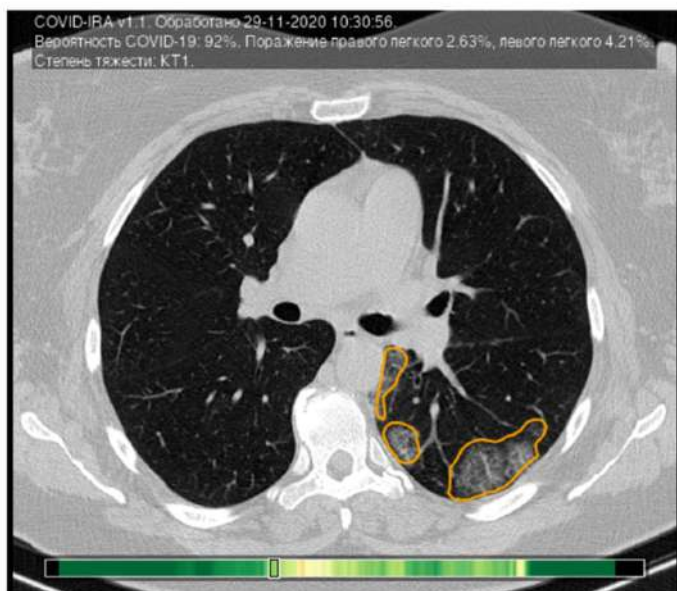
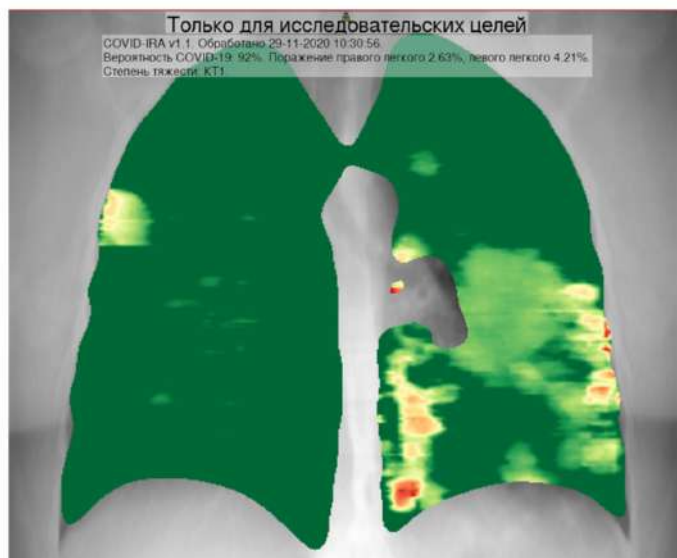
ЛП: разметка за пределами грудной клетки, в стенках кишки



18. CVL (KT Covid)

ЛП: разметка за пределами грудной клетки, в стенках кишки и на уровне реберно-позвоночного сочленения





Ю. А. Васильев
А. В. Владимирский
К. М. Арзамасов
А. Е. Андрейченко
В. А. Гомболевский
Н. С. Кульберг
О. В. Омелянская
Н. А. Павлов
Р. В. Решетников
К. А. Сергунова
Д. Е. Шарова
И. М. Шулькин

Компьютерное зрение в лучевой диагностике: первый
этап Московского эксперимента
Монография. 2-е издание, переработанное и дополненное

Технический редактор А. И. Овчарова
Компьютерная верстка Е. Д. Бугаенко

Ю. А. Васильев, А. В. Владзимирский, К. М. Арзамасов,
А. Е. Андрейченко, В. А. Гомболевский, Н. С. Кульберг,
О. В. Омелянская, Н. А. Павлов, Р. В. Решетников, К. А. Сергунова,
Д. Е. Шарова, И. М. Шулькин.

Систематизированы результаты первого этапа (2019–2020 гг.) крупнейшего в мире научного «Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы». Представлены методологии испытаний и применения ИИ в лучевой диагностике. Для организаторов здравоохранения, научных и медицинских работников, специалистов по разработке ИИ, сотрудников и учащихся учебных заведений.

ISBN 978-5-0059-3043-9



9 785005 930439 >