

ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ЗДРАВООХРАНЕНИЯ
ГОРОДА МОСКВЫ
«НАУЧНО-ПРАКТИЧЕСКИЙ КЛИНИЧЕСКИЙ ЦЕНТР ДИАГНОСТИКИ
И ТЕЛЕМЕДИЦИНСКИХ ТЕХНОЛОГИЙ
ДЕПАРТАМЕНТА ЗДРАВООХРАНЕНИЯ ГОРОДА МОСКВЫ»

На правах рукописи

АРЗАМАСОВ КИРИЛЛ МИХАЙЛОВИЧ

**ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
ПРИ МАССОВЫХ ПРОФИЛАКТИЧЕСКИХ И ДИАГНОСТИЧЕСКИХ
ЛУЧЕВЫХ ИССЛЕДОВАНИЯХ**

3.3.9. – Медицинская информатика (медицинские науки)

3.2.3. – Общественное здоровье, организация и социология здравоохранения,
медико-социальная экспертиза (медицинские науки)

Диссертация на соискание ученой степени
доктора медицинских наук

Научные консультанты:
профессор РАН,
доктор медицинских наук, профессор
Колсанов Александр Владимирович;
доктор медицинских наук
Владзимирский Антон Вячеславович

Москва – 2024

ОГЛАВЛЕНИЕ

Введение.....	5
Глава 1. ИНФОРМАТИЗАЦИЯ И ЦИФРОВИЗАЦИЯ ЛУЧЕВОЙ ДИАГНОСТИКИ В КОНТЕКСТЕ ОРГАНИЗАЦИИ, УПРАВЛЕНИЯ И ПРОВЕДЕНИЯ МАССОВЫХ ПРОФИЛАКТИЧЕСКИХ ИССЛЕДОВАНИЙ (ОБЗОР ЛИТЕРАТУРЫ)	19
1.1. Массовые профилактические исследования методами лучевой диагностики	19
1.2. Достигнутый уровень информатизации и цифровизации лучевой диагностики	23
1.3. Возможности автоматизации процессов массовых лучевых исследований на основе технологий искусственного интеллекта.....	36
Глава 2. МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ	54
2.1. Общий ход и этапы исследования.....	54
2.2. Информационная инфраструктура.....	58
2.3. Сопоставления показателей диагностической точности врача- рентгенолога и программного обеспечения на основе технологий искусственного интеллекта.....	64
2.4. Программное обеспечение на основе технологий искусственного интеллекта (ИИ-сервисы).....	71
2.5. База исследования	75
2.6. Методическое обеспечение комплексной системы контроля качества программного обеспечения на основе ТИИ	77
Глава 3. ОБОСНОВАНИЕ ЦЕЛЕСООБРАЗНОСТИ И ПРИМЕНИМОСТИ АВТОМАТИЗАЦИИ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ МАССОВЫХ ПРОФИЛАКТИЧЕСКИХ ИССЛЕДОВАНИЙ, ОСУЩЕСТВЛЯЕМЫХ С ПРИМЕНЕНИЕМ МЕТОДОВ ЛУЧЕВОЙ ДИАГНОСТИКИ	90
3.1. Проблемы управления массовыми профилактическими исследованиями, осуществляемыми с применением методов лучевой диагностики.....	90

3.2. Оценка применимости технологий искусственного интеллекта для профилактических исследований	104
3.3. Разработка, внедрение и оценка результативности организационной технологии автоматизации рабочего процесса врача-рентгенолога при проведении профилактических исследований.....	110
Глава 4. СОЗДАНИЕ И ОЦЕНКА РЕЗУЛЬТАТИВНОСТИ СИСТЕМЫ КОНТРОЛЯ КАЧЕСТВА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОФИЛАКТИЧЕСКИХ ИССЛЕДОВАНИЙ.....	136
4.1. Методика многоступенчатой системы контроля качества программного обеспечения на основе технологий искусственного интеллекта	136
4.2. Разработка инструментов для осуществления контроля качества работы программного обеспечения на основе технологий искусственного интеллекта.....	157
4.3. Апробация системы контроля качества при проведении профилактических исследований с применением технологий искусственного интеллекта	181
4.4. Инструмент поддержки принятия управленческих решений о выборе программного обеспечения на основе технологий искусственного интеллекта.....	194
Глава 5. МОДЕЛЬ ОРГАНИЗАЦИИ И ПРОВЕДЕНИЯ МАССОВЫХ ПРОФИЛАКТИЧЕСКИХ ИССЛЕДОВАНИЙ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА.....	212
5.1. Особенности применения программного обеспечения на основе технологий искусственного интеллекта для маммографии	212
5.2. Оценка эффективности автоматизации процессов организации и проведения маммографии	218

5.3. Особенности применения программного обеспечения на основе технологий искусственного интеллекта для рентгенографии/флюорографии органов грудной клетки.....	225
5.4. Оценка эффективности автоматизации процессов организации и проведения флюорографии	231
5.5. Оценка влияния применения программного обеспечения на основе технологий искусственного интеллекта на длительность описания результатов профилактических исследований	239
Заключение	243
Выводы	269
Практические рекомендации	271
СПИСОК СОКРАЩЕНИЙ, АББРЕВИАТУР И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	272
СПИСОК ЛИТЕРАТУРЫ.....	276
ПРИЛОЖЕНИЯ.....	310
Приложение А	Ошибка! Закладка не определена.
Приложение Б.....	Ошибка! Закладка не определена.
Приложение В.....	Ошибка! Закладка не определена.

ВВЕДЕНИЕ

Актуальность темы исследования. Профилактика как комплекс направленных на сохранение и укрепление здоровья населения мероприятий является одним из приоритетных направлений деятельности и развития системы здравоохранения. В Российской Федерации (РФ) основополагающее звено вторичной профилактики – это массовые профилактические осмотры, порядки и правила проведения которых регламентированы нормативно-правовыми документами органов исполнительной власти в сфере здравоохранения.

Несмотря на многочисленные научные исследования, практические комплексные меры по развитию системы массовых профилактических осмотров (Еганян Р. А. с соавт., 2019; Концевая А. В. с соавт., 2019; Красникова Н. В. с соавт., 2019; Лобеев А. В. с соавт., 2020; Максикова Т. М. с соавт., 2019; Aceves B. et al., 2020; Lane M. M. et al., 2021; Tapia-Conyer R. et al., 2017) [1–8], проблема своевременного выявления хронических неинфекционных заболеваний и иных социально значимых заболеваний далека от решения. Это подтверждается, в частности, статистическими данными о выявлении онкологической патологии. В России удельный вес злокачественных новообразований (ЗНО), впервые выявленных на III–IV стадиях, остается недопустимо высоким: в 2021 г. он составил 37,7%. В динамике за 10 лет отмечается некоторое снижение этого показателя (в 2011 г. – 43,1%), что свидетельствует об успешности многих научно-практических подходов (Каприн А. Д. с соавт., 2021) [9]. Однако без принципиального повышения активной выявляемости ЗНО на I–II стадиях невозможно добиться действительного снижения соответствующей летальности.

Типичной проблемой организации и проведения массовых профилактических осмотров (как в России, так и в мире) является обеспечение доступности, качества, максимального охвата населения, эффективной маршрутизации лиц с выявленными нарушениями (Драпкина О. М. с соавт., 2022; Кобякова О. С. с соавт., 2019; Buse K. et al, 2017; Du M. et al., 2023; Williams J.

et al., 2018) [10–14]. Для решения указанных проблем применяются различные подходы, однако в целом актуальность данного вопроса остается очень высокой. В подавляющем большинстве случаев, даже на фоне развитой экономики, ресурсные возможности систем здравоохранения оказываются ограниченными с точки зрения обеспечения нужного охвата населения профилактикой.

Еще большую угрозу для реализации массовых профилактических осмотров создают чрезвычайные ситуации – особенно наглядно это стало в ходе пандемии новой коронавирусной инфекции. В силу необходимости задействования максимальных ресурсов системы здравоохранения для борьбы с COVID-19 профилактические программы были остановлены (Дроздова Л. Ю. с соавт., 2020; Марцевич С. Ю. с соавт., 2020; Фролова И. А. с соавт., 2022; Azarpazhooh M. R. et al., 2020; Basu S., 2020; Palmer K. et al., 2020) [15–20]. Такое решение является, безусловно, обоснованным. Однако оно показало актуальность разработки принципиально новых подходов к обеспечению стабильности и безотказности работы системы профилактической медицины в будущем.

Массовые профилактические осмотры предполагают поэтапное осуществление различных видов обследований – от анкетирования до инструментальных и лабораторных исследований. Особое место в этой системе занимает лучевая диагностика. Рентгенологические исследования органов грудной клетки (флюорография) и молочной железы (маммография) представляют собой фундаментальные методики массовых профилактических осмотров, направленных на раннее выявление туберкулеза и ЗНО, необходимость их проведения зафиксирована в Приказах Минздрава России № 404н¹ и №29н². При этом в структуре смертности женщин наибольший удельный вес имеют ЗНО

¹ Приказ Минздрава России от 27.04.2021 N 404н "Об утверждении Порядка проведения профилактического медицинского осмотра и диспансеризации определенных групп взрослого населения».

² Приказ Минздрава России от 28.01.2021 N 29н "Об утверждении Порядка проведения обязательных предварительных и периодических медицинских осмотров работников, предусмотренных частью четвертой статьи 213 трудового кодекса Российской Федерации, перечня медицинских противопоказаний к осуществлению работ с вредными и (или) опасными производственными факторами, а также работам, при выполнении которых проводятся обязательные предварительные и периодические медицинские осмотры».

молочной железы – 15,8%, а мужчин – ЗНО трахеи, бронхов, легкого – 25,0% (Каприн А. Д. с соавт., 2021) [9].

В последние годы в России и других наиболее экономически развитых странах мира наблюдается значительное увеличение парка цифрового диагностического оборудования для лучевых исследований. Также стремительно нарастает спрос на соответствующие диагностические исследования, в том числе с целью скрининга и профилактики (Голубев Н.А. с соавт., 2021; Тюрин И. Е. с соавт., 2018; Шелехов П. В. с соавт., 2017, 2019) [21–24]. Вместе с тем дефицит кадровых ресурсов становится все более явным. Фактически на современном уровне оснащенности уже физически невозможно обеспечить наличие врача-рентгенолога возле каждого диагностического устройства. Усугубляет ситуацию необходимость выполнения двойных независимых описаний маммографии. По общепринятым мировым данным именно такой подход обеспечивает максимальное качество, своевременную выявляемость ЗНО молочной железы на ранних стадиях (Chen Y. et al., 2022; Euler-Chelpin M. et al., 2018; Posso M. et al., 2017; Schünemann H. J. et al., 2020; Рожкова Н. И. с соавт., 2020) [25–29]. Необходимость постоянного наращивания охвата населения профилактическими исследованиями, обеспечения устойчивости работы на фоне неустраняемого кадрового дефицита и значительного роста спроса на лучевые исследования является серьезным вызовом для организаторов здравоохранения. Очевидной становится высокая перспективность создания и внедрения новых цифровых технологий.

Степень разработанности темы исследования. Цифровые технологии активно применяются в контексте проблемы профилактической медицины. В лучевой диагностике успешно реализуется модель централизации. Благодаря созданию референс-центров устраняется кадровый дефицит, повышается доступность лучевых исследований (Глушкова И. В. с соавт., 2018;

Кошкаров А. А. с соавт., 2017, 2020; Пшекарский Д. С. с соавт., 2020; Bashshur R. L. et al., 2016; Seo H. S. et al., 2021; Weikert T. et al., 2019) [30–36].

В последние годы значительное внимание уделяется новому поколению средств автоматизации производственных процессов в сфере здравоохранения – так называемым технологиям искусственного интеллекта (ТИИ). Более того, Национальной стратегией развития искусственного интеллекта на период до 2030 года³ (Указ Президента РФ от 10.10.2019 №490) предполагается использование ТИИ в социальной сфере для повышения качества услуг в области здравоохранения, включая, прежде всего, профилактические обследования и диагностику, основанную на анализе изображений.

Представляется актуальным создание организационных подходов, основанных на ТИИ и направленных на совершенствование массовых профилактических мероприятий. Вместе с тем научные знания о применимости и качестве искусственного интеллекта (ИИ) в медицине остаются достаточно разрозненными, неполными; многие исследования отличаются крайне слабой доказательностью, нерепрезентативным статистическим анализом (Гусев А. В. с соавт., 2022; Collins G. S. et al., 2021; Hardy M. et al., 2020; Milne-Ives M. et al., 2020; Shahzad R. et al., 2022; Stanfill M. et al., 2019) [37–42]. Не разработаны принципы организации системного внедрения ТИИ для автоматизации производственных процессов в сфере здравоохранения. Отсутствует научно обоснованная система обеспечения качества медицинской помощи (в частности, профилактических исследований), оказываемой с применением технологий искусственного интеллекта.

Цель исследования – совершенствование организации и проведения профилактических и диагностических лучевых исследований с использованием технологий искусственного интеллекта.

³ Указ Президента Российской Федерации от 10 октября 2019 г. №490 «О развитии искусственного интеллекта в Российской Федерации»

Задачи исследования:

1. Изучить проблемы управления профилактическими и диагностическими исследованиями, осуществляемыми с применением методов лучевой диагностики.
2. Оценить применимость технологий искусственного интеллекта в лучевой диагностике путем сравнительного изучения точности диагностических решений, принимаемых врачом самостоятельно и формируемых автоматически.
3. Разработать методику внедрения систем поддержки принятия врачебных решений для автоматизации рабочего процесса врача-рентгенолога при проведении профилактических и диагностических исследований.
4. Обосновать организационно-методические основы контроля качества программного обеспечения на основе технологий искусственного интеллекта для профилактических и диагностических лучевых исследований.
5. Обосновать и внедрить модель организации и проведения профилактических и диагностических исследований с применением технологий искусственного интеллекта.
6. Оценить результативность применения систем поддержки принятия врачебных решений для автоматизации процессов организации и проведения профилактических и диагностических исследований.

Научная новизна исследования. Получены новые теоретические знания о возможностях цифровых технологий для управления массовыми профилактическими исследованиями, направленными на выявление социально значимой патологии.

Разработано программное обеспечение для подготовки наборов медицинских данных (Свидетельства о государственной регистрации программы для ЭВМ № 2023617333, № 2023619686, № 2022617275), на основе которого созданы наборы данных для оценки диагностической точности программного обеспечения на основе технологий искусственного интеллекта, анализирующего

результаты лучевых исследований (Свидетельства о государственной регистрации базы данных № 2023621204, № 2023621234, № 2023621305, № 2023622421, № 2023624880).

Впервые доказана достижимость высокого сходства диагностической точности врача-рентгенолога и алгоритма искусственного интеллекта при стандартизированной классификации патологических изменений на результатах профилактических лучевых исследований (диагностическая точность программного обеспечения (ПО) на основе ТИИ достигает 0,96 для маммографии; 0,95 – для флюорографии; 0,91 – для рентгенографии; диагностическая точность врача – 0,93 для маммографии; 0,96–0,97 – для рентгенографии/ флюорографии).

Впервые разработана масштабируемая методика бесшовного внедрения технологий искусственного интеллекта в производственные процессы организации и проведения диагностических исследований (50 решений были протестированы и внедрены в деятельность 153 медицинских организаций города Москвы; 7 решений внедрены в 12 медицинских организаций Ямало-Ненецкого автономного округа).

Впервые разработаны инструмент оценки зрелости и система оценки и контроля технических и клинических показателей (эксплуатационных параметров) программного обеспечения на основе технологий искусственного интеллекта на этапах жизненного цикла (Свидетельства о государственной регистрации программы для ЭВМ № 2023611181, № 2023665713, № 2022617275), в том числе с позиций применения в системе обязательного медицинского страхования (разработанная система контроля качества обеспечила рост диагностической точности на 41% для флюорографии, на 25% – для маммографии, на 19% – для рентгенографии органов грудной клетки).

Впервые научно обоснованы сценарии автоматизации описания результатов профилактических лучевых исследований (сценарий двойного описания с применением технологий искусственного интеллекта позволяет сократить число пропусков патологии для маммографии на 3%; для

рентгенографии/флюорографии предложенный подход позволяет максимально, на 77%, сократить нагрузку на врача-рентгенолога).

Теоретическая и практическая значимость. Разработан новый методический подход к организации и управлению массовыми профилактическими лучевыми исследованиями с применением технологий искусственного интеллекта.

Разработан и внедрен комплекс инструментов (программного обеспечения), обеспечивающий контроль качества применения технологий искусственного интеллекта в здравоохранении и включающий: инструмент расчета показателей диагностической точности, инструмент мониторинга технологических и клинических показателей работы, инструмент тестирования при смене версионности, инструмент сопоставления с заключением врача-рентгенолога, инструмент оценки диагностической точности врача-рентгенолога.

Создан ряд наборов данных для оценки врачебной точности при описании профилактических лучевых исследований и для оценки диагностической точности программного обеспечения на основе технологий искусственного интеллекта.

Разработан инструмент оценки зрелости программного обеспечения на основе технологий искусственного интеллекта для применения в сфере здравоохранения. Установлены показатели минимальной диагностической точности для допуска программного обеспечения на основе технологий искусственного интеллекта к практической эксплуатации: площадь под характеристической кривой (AUC) для профилактической маммографии – 0,88, для рентгенографии/флюорографии – 0,93–0,94.

Впервые в государственную систему здравоохранения внедрена медицинская услуга, оказываемая с применением технологий искусственного интеллекта и финансируемая за счет средств обязательного медицинского страхования (001601 «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта»).

Методология и методы исследования. Базой исследования послужили медицинские организации, подведомственные Департаменту здравоохранения г. Москвы, и ГБУЗ г. Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы». Диссертационное исследование выполнено на основе принципов системного подхода.

Дизайн: исследование со смешанными методами.

Методы исследования: эпидемиологические; аналитические (анализ, синтез, индукция, дедукция); статистические, в т. ч. метод оценки диагностической точности (построение и анализ характеристической кривой).

Объект исследования: результаты обработки лучевых исследований ПО на основе ТИИ; протоколы врачебных описаний профилактических и диагностических лучевых исследований.

Предмет исследования: диагностическая точность «среднего» врача-рентгенолога и ПО на основе ТИИ; внедрение ПО на основе ТИИ в рабочий процесс врача-рентгенолога; организация проведения диагностических и профилактических исследований с применением ТИИ.

Единица наблюдения: лучевое исследование с результатами работы ПО на основе ТИИ; лучевое исследование, описанное врачом-рентгенологом.

Основные положения, выносимые на защиту:

1. Для совершенствования организации и управления массовыми профилактическими лучевыми исследованиями целесообразна автоматизация отдельных процессов, которая может быть выполнена за счет применения технологий искусственного интеллекта.

2. При анализе результатов профилактических и диагностических лучевых исследований точность программного обеспечения на основе технологий искусственного интеллекта соответствует точности диагностических решений врача-рентгенолога.

3. Программное обеспечение на основе технологий искусственного интеллекта для анализа и интерпретации результатов профилактических и диагностических лучевых исследований целесообразно интегрировать в информационную систему в сфере здравоохранения субъекта Российской Федерации. Результаты работы программного обеспечения на основе технологий искусственного интеллекта рекомендуется предоставлять в виде двух дополнительных серий: текстовой (DICOM SR) с описанием и заключением и графической (DICOM SC) с визуальным отображением выявленных патологических изменений.

4. Контроль качества, реализованный на всех этапах жизненного цикла, позволяет дорабатывать программное обеспечение на основе технологий искусственного интеллекта для анализа профилактических и диагностических лучевых исследований, увеличивая его диагностическую точность.

5. Автоматизация процессов организации и проведения профилактических и диагностических лучевых исследований позволяет сократить число клинически значимых пропусков патологии и существенно повысить производительность труда врача-рентгенолога.

Степень достоверности исследования. Научные результаты и практические рекомендации диссертации, сформулированные автором лично, базируются на анализе достаточного объема первичных данных (6 млн исследований). В исследовании применены адекватные методы научного познания, полностью соответствующие поставленным задачам. Анализ данных проводился с использованием современных подходов и методов, в том числе машинного обучения. Статистическая обработка данных диссертации включала описательную статистику, оценку распределения, проведение ROC-анализа, расчет показателей диагностической точности, коэффициента согласованности и расчет 95% доверительного интервала для всех показателей. Статистический анализ проведен с использованием языка программирования R и Python.

Внедрение результатов исследования. Результаты диссертационного исследования послужили основой для нормативно-правовых актов – приказов органа исполнительной власти в сфере здравоохранения субъекта Российской Федерации (Департамента здравоохранения города Москвы): № 142 от 19.02.2020 «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»; № 51 от 26.01.2021 «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 2021 году»; № 160 от 24.02.2022 «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 2022 году»; № 134 от 16.02.2023 «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» .

Результаты исследования внедрены в Территориальную программу государственных гарантий бесплатного оказания гражданам медицинской помощи в городе Москвы (приложение № 6 к Тарифному соглашению на 2023 год от 30.12.2022), в деятельность медицинских организаций Москвы, Всеволожска, ЯНАО, организующих проведение диагностических исследований, а также Клиник Самарского государственного медицинского университета.

Полученные результаты также были внедрены в практическое здравоохранение в виде методических рекомендаций: «Использование сервисов на основе технологии искусственного интеллекта при проведении описаний рентгенологических снимков», 2020 год; «Базовые рекомендации к работе

сервисов искусственного интеллекта для лучевой диагностики», 2022 год «Клинические испытания систем искусственного интеллекта (лучевая диагностика)», 2023 год.

Результаты работы внедрены в педагогический процесс ФГАОУ ВО «Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет)», ФГБОУ ВО «МИРЭА – Российский технологический университет» и ФГБОУ ВО СамГМУ Минздрава России. Результаты диссертационного исследования стали основой национальных стандартов Российской Федерации: ГОСТ Р 59921.1-2022 «Системы искусственного интеллекта в клинической медицине. Часть 1. Клиническая оценка» и ГОСТ Р 59921.4-2021 «Системы искусственного интеллекта в клинической медицине. Часть 4. Оценка и контроль эксплуатационных параметров».

Личный вклад автора. Автор самостоятельно установил актуальность и сформулировал тему диссертационной работы, определил цель и задачи, подготовил дизайн и методологию исследования, а также установил перечень необходимых методов. Соискатель лично провел накопление и анализ первичных данных, разработал организационные технологии и методику, осуществил научно-организационную деятельность по разработке инструментов, участвовал во внедрении результатов диссертационного исследования (в том числе осуществляя научно-организационное руководство), лично провел оценку результативности внедрения, систематизировал результаты диссертации в виде единой модели, заключения, выводов, практических рекомендаций.

Апробация результатов исследования. Основные положения и результаты работы были доложены и обсуждены на следующих конференциях международного, всероссийского и регионального уровней: Международный конгресс «Суперкомпьютерные дни в России», 21–22.09.2020, Москва; вебинар «Практический вебинар с разбором клинических случаев использования ИИ-

сервисов», 06.11.2020, Москва; Открытое совещание с разработчиками и пользователями ИИ-сервисов, 19.11.2020, Москва; Форум Московского регионального отделения Российского общества рентгенологов и радиологов, секция «ИИ в медицинской диагностике: что нужно знать врачу», 17.12.2020, Москва; Открытая конференция по искусственному интеллекту OpenTalks.AI, 04.02.2021, Москва; ECR 2021 – European Congress of Radiology, 03.03–07.03.2021, Вена; XIII Всероссийский научно-образовательный форум с международным участием «Медицинская Диагностика – 2021», 25.05–27.05.2021, Красногорск; Первый Российский диагностический саммит, 19.10–21.10.2021, Красногорск; Открытая конференция по искусственному интеллекту OpenTalks.AI, 16.02–18.02.2022, Москва; V Всероссийский конгресс «Онкорadiология, лучевая диагностика и терапия», 16.02–18.02.2022, Москва; ECR 2022 – European Congress of Radiology, 02–06.2022, Вена; VI форум «Онлайн-диагностика 3.0», 13.04–16.04.2022, Москва; XVI Всероссийский национальный конгресс лучевых диагностов и терапевтов «Радиология – 2022», 24.05–26.05.2022, Красногорск; Всероссийская конференция «Обеспечение потребностей национального здравоохранения. Лето-2022», 08.06–09.06.2022, Москва; Всероссийская научно-практическая конференция с международным участием «Медицина катастроф – 2022», 15.09–16.09.2022, Москва; V Юбилейный международный форум онкологии и радиотерапии For Life («Ради жизни»), 19.09–23.09.2022, Москва; XXIII Всероссийский научно-образовательный форум «Мать и Дитя, 28.09–30.09.2022, Москва; II Междисциплинарный форум «Медицина молодая», 07.12.2022, Москва; VII Всероссийская научно-практическая конференция по искусственному интеллекту в здравоохранении и системам поддержки принятия врачебных решений ITM-AI-2023, 09.02–10.02.2023 (онлайн); VI Всероссийский конгресс «Онкорadiология, лучевая диагностика и терапия», 10.02–12.02.2023, Москва; XIV Международный конгресс «Невский радиологический форум-2023», 07–08.04.2023, Санкт-Петербург; VI Инновационный Петербургский медицинский форум «Медицина 2023 – наука, инновации и практика», 18.05–

20.05.2023, Санкт-Петербург; V Международный конгресс Health Age, 25.05.2023, Москва; XVII Всероссийский национальный конгресс лучевых диагностов и терапевтов «Радиология – 2023», 30.05–01.06.2023, Красногорск; 22-й Всероссийский конгресс с международным участием «Скорая медицинская помощь – 2023», 08.06–09.06.2023, Санкт-Петербург.

Публикации по теме диссертации. По материалам диссертационного исследования опубликована 31 печатная работа в отечественных и зарубежных изданиях, из них 22 – в изданиях, рекомендованных ВАК при Минобрнауки России, в том числе 16 – в изданиях, входящих в международные базы данных Web of Science и Scopus. Получено 5 свидетельств о регистрации программ для ЭВМ и 40 свидетельств о регистрации баз данных. Также опубликованы: 1 монография, 5 методических рекомендаций и 1 учебное пособие.

Соответствие паспорту научной специальности. Диссертационное исследование соответствует п. 11 «Интеллектуальные системы для клинической практики. Системы поддержки принятия клинических решений. Системы в области здоровьесбережения» и п. 17 «Методы и алгоритмы интеллектуальной поддержки принятия управленческих решений в здравоохранении. Ситуационный анализ и управление в системе охраны здоровья населения» паспорта научной специальности 3.3.9. «Медицинская информатика». Диссертационное исследование также соответствует п. 14 «Исследование проблем управления здравоохранением, разработка цифровых технологий управления лечебно-профилактическими учреждениями, службами и здравоохранением в целом с целью совершенствования организационных форм и методов работы органов управления здравоохранением и медицинскими организациями, оценки эффективности их деятельности» и п. 18 «Разработка теоретических и методологических основ обеспечения для населения доступности, качества и безопасности медицинской помощи» паспорта научной специальности 3.2.3.

«Общественное здоровье, организация и социология здравоохранения, медико-социальная экспертиза».

Структура и объем работы. Текст диссертации изложен на 326 страницах, состоит из введения, обзора литературы, главы о материалах и методах исследования, 3 глав об оригинальных исследованиях, заключения, выводов, практических рекомендаций и списка цитируемой литературы (300 источников). Диссертация включает 28 таблиц, 76 рисунков, 3 приложения.

Глава 1. ИНФОРМАТИЗАЦИЯ И ЦИФРОВИЗАЦИЯ ЛУЧЕВОЙ ДИАГНОСТИКИ В КОНТЕКСТЕ ОРГАНИЗАЦИИ, УПРАВЛЕНИЯ И ПРОВЕДЕНИЯ МАССОВЫХ ПРОФИЛАКТИЧЕСКИХ ИССЛЕДОВАНИЙ (ОБЗОР ЛИТЕРАТУРЫ)

1.1. Массовые профилактические исследования методами лучевой диагностики

Переход стратегии системы здравоохранения от лечения пациента с клиническими проявлениями заболевания к предупреждению возникновения заболевания на ранних стадиях позволил существенно улучшить ситуацию со многими болезнями. Эта стратегия оказалась одинаково эффективной при борьбе с онкологическими заболеваниями [43,44], туберкулезом [45] и другой патологией [46,47]. В XXI веке скрининг стал основным направлением политики здравоохранения для многих стран. Термин «скрининг» нашел широкое применение в отечественной научной литературе, однако в официальных документах его не используют, заменяя термином «профилактические исследования», закрепленным в приказе Министерства здравоохранения Российской Федерации от 27.04.2021 № 404н «Об утверждении Порядка проведения профилактического медицинского осмотра и диспансеризации определенных групп взрослого населения» [48]. Профилактика как комплекс направленных на сохранение и укрепление здоровья населения мероприятий является одним из приоритетных направлений деятельности и развития системы здравоохранения. В Российской Федерации (РФ) основополагающее звено вторичной профилактики – это массовые профилактические осмотры, порядки и правила проведения которых регламентированы нормативно-правовыми документами органов исполнительной власти в сфере здравоохранения.

Активному распространению массовых профилактических исследований способствовало развитие в первую очередь инструментальной диагностики,

позволившее выполнять исследования с достаточной чувствительностью при относительно низких затратах. Именно сочетание высокой чувствительности и низкой стоимости таких методов лучевых исследований, как флюорография, рентгенография и маммография, обеспечивает охват целевой группы населения. На примере маммографии можно увидеть, как при сравнительно низкой стоимости исследования были получены высокие показатели экономической эффективности раннего выявления рака молочной железы [49].

Согласно данным Всемирной организация здравоохранения (ВОЗ) рекомендуется проводить скрининговые программы только на тех территориях, где уровень распространенности заболеваний является достаточно высоким, чтобы оправдать подобные меры. Программа скрининга может быть внедрена, если доказана эффективность диагностических методов, ее составляющих, а также имеются достаточные ресурсы (как людские, так и технические) для охвата значительной доли целевой группы и обеспечения последующей медицинской помощи [50].

Ввиду технического прогресса и открытия новых методов визуализации многие медицинские работники все чаще стали отдавать предпочтение компьютерной и магнитно-резонансной томографии. Однако в профилактической медицине высокие технологии не способны полностью вытеснить «классическую рентгенологию» по ряду причин: высокая стоимость аппаратуры и самого исследования, малая оснащенность медицинских учреждений, необходимость обоснования показаний к исследованию, высокая лучевая нагрузка, наличие множества противопоказаний, сложность интерпретации результатов и недостаточная компетенция рентгенологов [51].

Высокая лучевая нагрузка может быть оценена в виде коллективной дозы. По этому показателю по состоянию на 2019 год лидирует компьютерная томография (56,2%), при довольно низком количестве процедур (4,5%). Рентгеновские исследования составляют большую часть всех процедур (64,4%), однако их вклад в общую дозу облучения составляет только 19,2%. В 2019 году

флюорография была на втором месте по количеству процедур (29,5%), принося 6,2% в коллективную дозу [52].

В отношении дозовой нагрузки в качестве профилактических исследований флюорография и рентгенография органов грудной клетки выглядят более привлекательно. Однако в целях практического применения метод для массовых профилактических обследований должен быть не только дешевым и безопасным, но, прежде всего, чувствительным.

Ранее проведенные обзорные исследования указывают на большую по сравнению с рентгенографией органов грудной клетки эффективность низкодозовой компьютерной томографии, однако при этом указывают на отсутствие экономической эффективности, в том числе из-за низкого уровня специфичности метода [53–55]. Рентгенографию органов грудной клетки современные исследователи рассматривают как метод, обладающий относительно высокой чувствительностью при выявлении рака легкого [56], но при этом выявление данного заболевания благодаря рентгенографии не приводит к снижению смертности от него [57]. Это отчасти обусловлено обнаружением заболевания на более поздних стадиях по сравнению с компьютерной томографией и низкодозной компьютерной томографии органов грудной клетки. Тем не менее необходимо четко понять причину пропуска узлов, подозрительных на ЗНО легкого на рентгенограммах: это техническое ограничение или человеческий фактор?

С одной стороны, рентгенология развивается и совершенствуется, появляются аппараты с высокой разрешающей способностью, с другой стороны, работы пятилетней давности указывают на сложность в определении легочных узлов на рентгенограмме с линейным размером менее 2 см [58]. Интересным представляются зарубежные исследования, в которых указывается на пропуски легочных узлов в 90% рентгенограмм, фигурировавших в судебных делах [59–61]. Этот факт позволяет заподозрить, что в большинстве случаев имеет место человеческий фактор (о чем ранее уже сообщали отдельные исследователи [62]),

влияние которого может быть потенциально снижено при организации двойного чтения, как было реализовано для маммографии [69, 70]. Однако в условиях ограниченности ресурсов двойное чтение может быть недоступно для большинства медицинских организаций. Кроме этого, низкая частота встречаемости патологии при проведении скрининговых исследований означает большой риск ее пропуска [65]. В этом случае необходимо обратить внимание на использование современных систем поддержки принятия врачебных решений.

В Москве ежедневно осуществляется свыше 45 тысяч диагностических исследований, результаты которых анализируются врачами-рентгенологами посредством Единого радиологического информационного сервиса Единой медицинской информационно-аналитической системы г. Москвы (ЕРИС ЕМИАС). Такая высокая производительность обусловлена цифровизацией в сфере лучевой диагностики, позволяющей централизовать процесс описания рентгеновских исследований с использованием телемедицинских технологий [66]. Большая часть проведенных исследований приходится на рутинные профилактические проверки, включая флюорографию, рентген органов грудной клетки и маммографию.

Профилактические исследования часто выявляют малый процент патологий. В России при проведении флюорографии обнаружение разнообразных заболеваний грудной клетки варьируется от 0,07 до 18,4% [67–71]. При этом высокие показатели обнаружения (10,1–18,4%) связаны с диагностированием различных заболеваний, таких как туберкулез, онкологические и другие состояния, в то время как низкие показатели (0,07–0,76%) относятся к выявлению туберкулеза.

Представляется целесообразным создание организационных технологий, позволивших бы использовать ресурсы системы здравоохранения более рационально; в частности – «разгрузить» врача-рентгенолога, позволив сфокусировать его внимание на клинически сложных случаях.

Активное развитие радиологических информационных систем (на примере вышерассмотренного ЕРИС ЕМИАС) позволило интегрировать системы поддержки принятия врачебных решений в привычную для врача-рентгенолога инфраструктуру [72,73]. Однако представленные работы по внедрению технологий на основе искусственного интеллекта являются скорее исключением, чем настоящим научным трендом [74]. Отдельные исследователи описывают применение технологий на основе ИИ в будущей рентгенодиагностике [75,76], но только в теории, без апробации на практике. В основной массе научные статьи представляют собой исследовательские работы в масштабах лаборатории или одной медицинской организацией. Вместе с тем даже ограниченные исследования последних лет демонстрируют достаточно высокую эффективность программного обеспечения на основе технологий искусственного интеллекта (ПО на основе ТИИ).

Показатели диагностической точности некоторых алгоритмов не только сравнивались, но и по ряду исследовательских работ превосходили таковые для врача-рентгенолога. Однако при переносе этих технологий на практику мы видим снижение показателей диагностической точности ПО на основе ТИИ [77]. Возможность осуществления тонкой настройки алгоритмов дает потенциал внедрению этого инструмента в массовые профилактические исследования. Таким образом, вопросы масштабного внедрения ПО на основе ТИИ в клиническую практику для массовых профилактических рентгеновских исследований остаются открытыми.

1.2. Достигнутый уровень информатизации и цифровизации лучевой диагностики

В России, с момента выполнения первого рентгеновского снимка в 1896 г., рентгенология на протяжении века была аналоговой: результаты исследования фиксировались сначала на экранах (рентгеноскопия), а после – на фотопленке.

С учетом необходимости применения серебра при изготовлении рентгеночувствительной пленки рентгеновское исследование нельзя было назвать бюджетным. Так, например, в России за год в среднем проводилось до 200 миллионов рентгеновских снимков, при этом суммарный расход серебра составлял не менее 40 тонн [78]. Несмотря на высокую стоимость выполненного рентгенологического исследования, в 1930-х годах С. А. Рейнберг пропагандировал «рентгенологическую диспансеризацию» здоровых и больных людей [79]. В 30–40-х годах прошлого столетия в России и в мире была эпидемия туберкулеза, который в основном поражал легкие. Именно с целью раннего выявления поражений легких и использовалась рентгенография органов грудной клетки. Классический рентген обладал умеренной лучевой нагрузкой не только на пациента, но и на медицинский персонал, а также требовал больших затрат. В этот момент вспомнили открытие Дж. Блейера, который в 1896 году продемонстрировал первый флюороскоп.

Флюорографические исследования обладают меньшей лучевой нагрузкой на медицинских работников, большой пропускной способностью и имеют низкую стоимость. В России первые массовые флюорографические обследования начались в 1947 году. Обеспечив широкий охват населения, флюорография позволила оперативно выявлять пациентов с туберкулезом, изолировать их и эффективно лечить за счет установления заболевания на ранних стадиях. Пленочная флюорография органов грудной клетки обладала рядом недостатков: относительно низкая чувствительность с экспозиционной дозой 4–5 мР в сравнении с 1–2 мР для стандартной рентгенографии, затратный и длительный процесс обработки изображений, а также высокая вероятность получения некачественных снимков, достигающая 13%. Но даже несмотря на эти недостатки, в конце XX века пленочная флюорография оставалась крайне востребованным методом. При помощи устаревших пленочных флюорографов в 1997 году были выявлены 1360 пациентов с активной формой туберкулеза и 832 пациента с раком легкого, а также другая патология легких. Эти показатели

подтверждают эффективность скрининговой флюорографии, что свидетельствует о высокой эффективности таких обследований, но также и о необходимости переоснащения [80–82].

Очевидно, что устаревшая технология (более 100 лет) имеет свои недостатки, которые могут быть устранены с применением современных технологий. Так, цифровые технологии, внедряющиеся повсеместно, не обошли и рентгенологию. Первый коммерческий цифровой рентгеновский аппарат появился в начале 80-х годов прошлого века [83]. Первые цифровые рентгеновские сканеры обладали низким разрешением (64×40 точки) [84]. Однако очень быстро разрешение цифровой рентгенограммы удалось довести до уровня стандартной рентгенографической кассеты [85,86].

Отечественная рентгенология также развивалась в направлении цифровых рентгенодиагностических систем, которые были реализованы к концу 1980-х годов [87–89]. В основном отечественные системы были направлены на оцифровку рентгеновских пленок, их обработку и дальнейшее хранение.

В конце XX века было уже очевидно, что будущее рентгенологии именно за цифровыми технологиями построения и обработки изображений, а век аналоговой рентгенографии завершается. В этом ключе развитие рентгеновского оборудования основывается на необходимости перехода на цифровые технологии, а диагностическое устройство в этой концепции следует рассматривать как единую систему преобразования информации [90]. Цифровые рентгеновские аппараты экономически более выгодные за счет быстрой окупаемости и экономии на расходниках. Положительно сказывается применение цифровых рентгеновских аппаратов на работе персонала (рентгенолаборант и врач-рентгенолог): снижается лучевая нагрузка, повышается производительность труда [91–93]. По этой причине мы наблюдаем постепенный переход от аналоговой пленочной рентгенографии к цифровой [92,94].

При внедрении многих прорывных технологий возникают силы, противостоящие их активному продвижению. Рентгенодиагностика не являлась

исключением. Цифровые рентгеновские аппараты имели существенно более высокую первоначальную стоимость по сравнению с аналоговыми. Также необходимо было подготовить рентгенолаборантов и врачей-рентгенологов к работе с цифровой аппаратурой [94]. Парадоксально, но технология, направленная на сокращение лучевой нагрузки и снижение времени на проведение исследования в начале, продемонстрировала диаметрально противоположный эффект. Изучение новых методов и приемов работы на цифровых рентгеновских системах привело к значительному увеличению времени проведения обследования по сравнению с традиционным аналоговым методом. В результате пациенты подвергались более высоким дозам облучения.

Цифровые рентгеновские аппараты позволяли дистанцировать пациента и медперсонал, сократив таким образом лучевую нагрузку на рентгенолаборантов и врачей-рентгенологов. Однако это также приводило к необходимости выполнять дополнительные сеансы сканирования из-за боязни пропустить патологию. В результате накапливалось большое количество цифровых копий исследования, переполнявших внутреннюю память аппаратов. Нужны были методические рекомендации по применению новых аппаратов, а также опытные специалисты, способные научить рационально использовать цифровое оборудование, но ничего из перечисленного не было [95].

Активная цифровизация, наблюдаемая в рентгенологии с начала 2000 годов, в итоге привела к практически полному отказу от пленочных технологий. Новые технологии получения изображения реализуются ведущими производителями аппаратуры для лучевой диагностики. К ним можно отнести малошумящие ПЗС-матрицы, а также наиболее перспективных полномасштабных полупроводниковых детекторов. Наибольшую перспективу открывают устройства, использующие технологию преобразования рентгеновского изображения с экрана в видеосигнал. Этот метод выделяется своей адаптивностью и обширными опциями для настройки ключевых параметров, таких как разрешение, динамический диапазон, размер поля обзора, чувствительность и скорость реакции [94, 96]. Интересен

подход, основанный на включении усилителя в каждый пиксель; он позволяет существенно повысить чувствительность системы [97, 98].

Цифровые детекторы в разы превосходят пленочные по показателям квантовой эффективности (0,5–0,8) и динамическому диапазону (>1000). По этой причине все недостатки, связанные с неправильным экспонированием, могут быть устранены последующей обработкой изображения. В результате из одного цифрового снимка можно извлечь информацию о плотностях разных структур без необходимости выполнения нескольких снимков, как это требовалось бы при аналоговом исследовании [99].

Однако сам переход на полноразмерные цифровые детекторы был связан с необходимостью оптимизации параметров «сигнал-шум». Высокие значения усиления сигнала сопровождались появлением собственного электронного шума. Такой шум становился заметным при самой низкой экспозиции, при которой может быть использован детектор. Поэтому переход на цифровые детекторы был не одномоментным и связан с жесткой конкуренцией производителей рентгеновских детекторов [100]. В настоящее время существуют несколько технологических систем цифровой рентгенографии [101], каждая из которых имеет свои плюсы и минусы. Однако благодаря возможности осуществления предварительной обработки изображения получаемое на экране монитора изображение фактически идентично для любой из систем.

Отдельного внимания заслуживает цифровая маммография. Рентгеновская маммография – последняя область рентгенологии, которая трансформировалась из аналогового изображения в цифровое. Ранее были описаны преимущества цифровой рентгенографии по сравнению с аналоговой, однако для маммографии все оказалось не так очевидно в части эффективности выявления рака молочной железы. В 2011 году В. Иареди и соавторы (W. Iaredi et al.) опубликовали систематический обзор с метаанализом, посвященный сравнению аналоговой и цифровой маммографии [102]. Приведенные результаты указывают на то, что в работах, опубликованных до 2005 г., не было достоверных различий в частоте

выявления рака молочной железы между двумя типами маммографических аппаратов. Напротив, исследования, проведенные после 2007 года, показали более высокую выявляемость рака молочной железы именно на цифровой маммограмме. Это может быть связано с постепенным развитием цифровых технологий и повышением квалификации рентгенологов при работе с цифровыми изображениями [103]. Разрешение цифрового маммографического исследования для выявления потенциально злокачественных образований должно быть выше, чем для рентгенографии органов грудной клетки. Для обнаружения объемных образований и микрокальцинатов цифровые ММГ системы должны обладать высоким разрешением и контрастностью. Способность цифровой ММГ выявлять мельчайшие структуры ограничена размером пикселя ее детектора. [104].

Цифровая маммография объединяет различные технологические подходы в зависимости от типа системы захвата изображений, включая:

- системы с CCD-матрицами, которые уже не выпускаются из-за ограниченного динамического диапазона и высоких уровней шума;
- компьютерную радиографию (CR), использующую кассеты с фосфорными пластинами и сканер для цифрового отображения на экране;
- полноформатную цифровую маммографию с двумя типами детекторов для прямого и непрямого преобразования сигнала в цифровую форму.

Современные исследования указывают на превосходство качества изображений от детекторов с прямым преобразованием по сравнению с CR-системами [105].

Маммография требует выявления объектов с размерами от 50 микрон (микрокальцинатов). Вследствие этого на всю техническую часть цифрового маммографа накладываются жесткие ограничения: высокая пространственная разрешающая способность, большая площадь сканирования, высокая квантовая эффективность и соотношение «сигнал-шум». Полученные таким образом снимки занимают существенно больший объем в хранилище по сравнению с рентгенографией органов грудной клетки. Таким образом, накладываемые

на цифровые маммографические системы требования к качеству должны быть выше [106].

С целью ускорения перехода на цифровую диагностику были реализованы различные проекты, направленные на переоснащение медицинских организаций. Среди основных двигателей цифровизации стоит отметить: национальный проект «Здоровье», программы высокотехнологичной медицины, осуществление региональных и муниципальных проектов в здравоохранении, а также развитие частной медицины [107].

Государственная программа развития здравоохранения на 2013–2020 гг. включает в себя положения о совершенствовании технологий и методов сохранения и укрепления здоровья населения, что в свою очередь невозможно без развития диагностических методов [108,109]. Чтобы повысить эффективность (результативность) всей отрасли, внедряются инновационные методы и проводится модернизация системы [110].

Несмотря на активную поддержку со стороны государства, к настоящему моменту переход на цифровые рентгеновские аппараты еще не закончен. Так, в Московской области трехлетняя программа по замене аналогового медицинского оборудования запущена была только в 2022 году [111]. Отдельно хотелось бы отметить развитие отечественной цифровой рентгеновской техники: в России цифровые рентгеновские сканеры разрабатывают и производят не более десятка отечественных компаний [96,112]. Ее удельный вес в оснащении отделений лучевой диагностики превышает 70% [113].

В начале цифровизации рентгенологии врачам-рентгенологам очень сложно было перестроиться описывать исследование не по твердой копии, а по изображению на мониторе. Наиболее объективной причиной приверженности твердой копии являлось низкое качество диагностических мониторов [114]. В начале 2000-х годов мониторы не могли воспроизводить необходимый динамический диапазон для отображения всех оттенков серого на цифровых рентгеновских снимках. Чтобы соответствовать возможностям

человеческого зрения, требуемый диапазон монитора должен был составлять не менее 1000. Однако мониторы того времени достигали значения чуть более 100. В результате при оценке цифровой рентгенограммы врач-рентгенолог был вынужден последовательно менять «окна», что являлось непривычным и более затратным по времени, если сравнить с оценкой изображения на негатоскопе.

Необходимо отметить, что технологии представления информации врачу-рентгенологу существенно развились с начала века, и уже сейчас мы видим принципиально новые диагностические мониторы, обладающие более высокими техническими показателями. Изучаются новые технологии конструкции матрицы мониторов: жидкокристаллические и светодиодные [115]. Перспективным в настоящее время является способ построения диагностического монитора, основанный на использовании последовательно расположенных жидкокристаллических матриц. Такой способ, по мнению Ц. Марчессоух и соавторов (C. Marchessoux et al.), позволяет получить изображение с высоким динамическим диапазоном, превосходящее по качеству обычные (с низким динамическим диапазоном) диагностические мониторы для рентгенологии [116]. При работе с цифровой маммографией требуются диагностические мониторы еще большего качества. Минимальное требование к матрице маммографического монитора составляет 5 мегапикселей. Этот критерий является ключевым для выявления ЗНО молочной железы. Использование мониторов с недостаточно высокими характеристиками приводит к потере информации, зафиксированной детектором, что может в свою очередь привести к пропускам патологии [104].

Еще одним важным вопросом при переходе с аналогового на цифровое рентгенологическое исследование стал вопрос об оригинале исследования, то есть если ранее врач-рентгенолог мог передать снимок коллеге для консультации или лечащему врачу, то с переходом на цифру этот вопрос снова стал актуальным. Изначально исследования хранились исключительно на рабочей станции, но, т.к. место для хранения не безгранично, приходилось удалять старые исследования, в результате они утрачивались. Автор диссертации даже стал

свидетелем того, как рентгенолаборанты распечатывали цифровые флюорограммы на лазерном принтере на обычной бумаге формата А4. Очевидно, что диагностическая ценность таких распечатанных изображений стремится к нулю. Был и альтернативный способ – записать изображение на компакт-диск или цифровой многоцелевой диск, но далеко не у каждого пациента с собой был диск, а в отделении лучевой диагностики этот расходный материал не всегда есть в наличии.

Переход на полностью цифровые изображения существенно расширяет возможности врача-рентгенолога. Цифровое изображение может быть обработано с применением разных цифровых фильтров и представлено в виде цветного изображения, что позволяет по-новому взглянуть на рентгеновское исследование, повышая чувствительность метода. Цифровая рентгенограмма с цветовым контрастом оптимальной насыщенности дает значительное улучшение в передаче информации по сравнению с обычным черно-белым изображением. Так, цветное изображение позволяет визуализировать диагностические данные, которые могут оставаться скрытыми или едва заметными из-за ограниченной способности человека различать оттенки серого. Исследования указывают на увеличение диагностической эффективности такого метода до 13% по сравнению с анализом классических черно-белых снимков [117].

Ключевым же преимуществом цифровых рентгенологических исследований является возможность сохранять их в едином формате хранения цифровых медицинских изображений. Рентгенодиагностические устройства генерируют большой объем данных, который необходимо хранить и обрабатывать. Для взаимодействия медицинского оборудования разных производителей необходим единый протокол. Таким протоколом выступает DICOM (Digital Imaging and COmmunications in Medicine) – стандарт создания, хранения, передачи и визуализации медицинских изображений и документов [118,119]. Данный стандарт был разработан для обеспечения наилучшей совместимости оборудования разных производителей. Каждый файл DICOM можно рассматривать как пакет, включающий медицинское изображение вместе с

дополнительными данными (техническая и клиническая информация), называемыми «метаданные». Благодаря стандартизации формата хранения медицинских изображений, а также перечня и значений для метаданных, стандарт DICOM объединил разные виды исследований, позволив реализовать PACS (систему архивирования и передачи изображений). Это позволяет проводить объективную оценку исследования в динамике, а также открывает возможность централизованного хранения и обработки исследования.

Итак, цифровые рентгеновские исследования обладают рядом преимуществ перед аналоговыми, но для раскрытия их в полном объеме недостаточно просто их визуализировать на диагностическом мониторе или сохранить на рабочей станции. Разработка информационной системы для обработки и хранения медицинских изображений началась с интеграции лучевой диагностики в единую систему архивации, передачи и анализа изображений. В 1981 году была создана первая экспериментальная версия такой системы, известная как PACS (Picture Archiving and Communication System), предназначенная для архивации и коммуникации медицинских изображений и связанных с ними данных [120]. Концепция и ключевые функции PACS были представлены Р. Бауманом в 1996 году [121]. С тех пор ее основной функционал остался неизменным: обеспечение накопления и хранения цифровых изображений; обеспечение передачи медицинских изображений; информационное обеспечение деятельности врачей-рентгенологов и врачей клинических специальностей, среднего медицинского и вспомогательного персонала. Развитие информационных систем вида PACS в России и в мире протекало неравномерно: возникало большое количество технических и экономических проблем. Технические заключались в том, что не все цифровое диагностическое оборудование поддерживает стандарт коммуникации DICOM или в силу разных факторов не может передавать диагностическое изображение на PACS, отличный от рекомендуемого производителем. Экономические проблемы чаще всего были обусловлены необходимостью организации или совершенствования имеющейся сетевой

инфраструктуры. Вторым этапом развития информатизации лучевой диагностики стало создание систем для автоматизации производственных процессов – так называемых «радиологических информационных систем (РИС)» (от англ. «Radiology Information System (RIS)») [122]. Интеграция PACS и РИС создает единое цифровое пространство лучевой диагностики как на уровне отдельной медицинской организации, так и на уровне административно-территориальной единицы (например, как компонент информационной системы в сфере здравоохранения субъекта РФ, включающий централизованный архив медицинских изображений). В настоящее время рынок PACS и РИС в Российской Федерации отличается хорошим развитием и общей активностью [123,124] [125], [126], [127], [128], [129], что свидетельствует о масштабной цифровизации и централизации в лучевой диагностике. Определенную альтернативу многочисленным разработкам специализированных ИТ-компаний представляют решения на основе программного обеспечения с открытым кодом [130], [131]. Соответствующая концепция имеет ряд значительных преимуществ, однако крайне редко применяется в государственной системе здравоохранения в силу высоких рисков для безопасности персональных данных.

Системы PACS позволяют подключать различные модули обработки цифрового медицинского изображения. Многие исследователи указывают на то, что постобработка цифровых диагностических изображений представляет собой ценный инструмент [132–134]. Однако все ли рентгенологи умеют эффективно использовать этот инструмент? Для полноценного использования возможностей PACS необходимо обучение специалистов работе с функциями постобработки изображений. Этот процесс может значительно повысить диагностическую точность благодаря улучшению качества визуализации на мониторе [135]. Отсутствие четких стандартов постобработки, вариативность инструментов у производителей приводят к необходимости постоянного обучения навыкам работы с ними, а также указывает на необходимость разработки соответствующих практических рекомендаций по их применению.

При внедрении РИС и PACS в медицинской организации важно понимать возможные трудности, с которыми придется столкнуться. Одним из важных ограничивающих факторов является низкий уровень компьютерной грамотности медицинских работников. Всего чуть более десяти лет назад один компьютер приходился на дюжину врачей [136]. За прошедшее время оснащенность возросла, но при этом исследователи отмечают недостаточный уровень обеспечения квалифицированным персоналом для работы с новым диагностическим оборудованием [137].

Второй значимой проблемой является вопрос хранилища данных. Даже рассматриваемые в рамках настоящей работы проекционные исследования быстро заполняют свободное место на дисках PACS. Так, одно маммографическое исследование, включающее в среднем 4–6 снимков, занимает около 120 МБ, а стандартная серия из 2–3 цифровых рентгеновских снимков – до 50 МБ. С учетом потока пациентов на профилактические исследования получается большой объем. Если мы говорим о централизации, то с такой PACS-системой будут работать несколько врачей-рентгенологов, а следовательно, необходимо обеспечить высокую прописную способность каналов связи, что не везде возможно [136]. Главным преимуществом внедрения единого PACS/РИС является сокращение до 80% затрат на весь рабочий процесс [138].

Отдельным аспектом является интеграция РИС/PACS с медицинскими информационными системами, представляющие собой системы автоматизации документооборота для медучреждений, включающей электронные карты пациентов, систему поддержки принятия врачебных решений и т.п. [139][139].

Взаимосвязь всех указанных информационных систем особенно нужна в контексте телерадиологии, когда врач-рентгенолог оказывается удаленным от пациента [137].

Телерадиология, представляющая применение телемедицинских технологий в сфере радиологии, активно используется в России; особенно ценна она стала для улучшения доступа к диагностическим услугам и повышения их качества в малонаселенных регионах. Технологические аспекты телерадиологии были

подробно изучены, в том числе в контексте работы многопрофильных медицинских учреждений. Особый рост интереса к телерадиологии наблюдался после принятия в 2017 году Федерального закона № 242-ФЗ, способствовавшего интеграции информационных технологий в здравоохранение [140]. Телерадиология в отделениях лучевой диагностики служит для непрерывного анализа результатов исследований, рационального распределения ресурсов и контроля качества.

Развитие телерадиологии позволило реализовать проекты по организации референс-центров. Одна из моделей телерадиологии, а именно модель «централизации описаний», стала основой для организации референс-центров. Технически все было готово с момента внедрения в клиническую практику PACS-систем, однако юридическая основа для появления референс-центров появилась только после 2018 года благодаря приказам Минздрава России и постановлениям Правительства РФ [141–143].

В крупных городах России, включая Москву и Санкт-Петербург, созданы консультационные центры. В 2020 году в Москве начал работу крупнейший референс-центр, где врачи-рентгенологи централизованно описывают лучевые исследования из медицинских организаций, подведомственных Департаменту здравоохранения города Москвы. С 1 января 2021 года по приказу Минздрава России № 560н «Об утверждении Правил проведения рентгеновских исследований» закреплена возможность дистанционного анализа рентгеновских изображений [144], что изменило подход к работе рентгеновских отделений и утвердило дистанционное описание изображений врачами-рентгенологами, в том числе в референс-центрах [145].

Таким образом, общий переход на цифровой формат рентгенологических исследований открывает новые возможности для врача-рентгенолога и повышает качество и безопасность проводимого исследования для пациента. Надо отметить, что описанные в данном параграфе проблемы являются общими при внедрении любых современных технологий и поэтому при внедрении новых методов крайне

важно акцентировать внимание на наличии следующих материалов: научно обоснованные методические рекомендации по практическому применению новой технологии, инструкции пользователя (для устройства или программного обеспечения), а также возможности и ограничения метода. Цифровизация лучевой диагностики создает новые потенциальные возможности для совершенствования процессов организации и управления системой здравоохранения в целом и процессами массовых профилактических исследований в частности. Вместе с тем практическая реализация таких возможностей требует научной основы, которая во многом еще не сформирована.

1.3. Возможности автоматизации процессов массовых лучевых исследований на основе технологий искусственного интеллекта

Первые попытки применения систем поддержки принятия врачебных решений в рентгенодиагностике датируются 50-ми годами XX века [146]. С того времени было апробировано множество различных решений, но в клиническую практику они смогли внедриться только в эру цифровой рентгенографии и PACS-систем.

Первые перспективные результаты были продемонстрированы экспертными системами, которые играли роль электронного ассистента врача-диагноста – CAD (computer aided diagnostic). Первая попытка создать автоматизированную систему CAD была предпринята в 1960-х годах [147]. Это исследование показало повышение точности обнаружения болезней органов грудной клетки при использовании врачом-рентгенологом в качестве помощника системы CAD. Системы CAD используют предобработку изображений: повышение контрастности, определение границ легких, уменьшение теней костных структур. Такая предобработка позволяет сделать патологические объекты более заметными [148,149].

Эпоха расцвета CAD-систем, анализирующих рентгенограммы органов грудной клетки с целью выявления изменений, характерных для туберкулеза, пришлась на период 1996–2013 гг. [150]. Особенность таких систем заключалась в высокой специфичности к определенному типу рентгенологических признаков туберкулеза. CAD-системы позволяли с большой точностью определять специфические признаки туберкулеза, например, как в работе Т. Ху и соавторов (Т. Xu et al.). Разработанный ими алгоритм безошибочно выявлял на рентгенограммах полости распада, которые зачастую свидетельствуют о бактериовыделении у пациента и опасности возникновения тяжелых осложнений [151]. Применение на практике системы, позволяющей определять один-единственный рентген-признак патологии, имеет сомнительную ценность, учитывая, что полость распада не является сложным для выявления рентгеновским признаком. CAD-системы эволюционировали по пути добавления новых рентгенологических признаков. Так, С. Джагер и соавторы (S. Jaeger et al.) использовали для разработки CAD-системы наборы рентгенограмм с различными признаками туберкулеза, включая полости, инфильтративные изменения, плевральный выпот. Полученная CAD-система обладала невысоким значением площади под характеристической кривой – AUC (0,83) [149], позднее коллективу исследователей удалось повысить точность, достигнув AUC 0,87–0,90. Но и эти значения нельзя было назвать достаточными, т.к. они все еще уступали показателям точности врачей-рентгенологов [152]. Тем не менее CAD-системы вышли за пределы лабораторий и стали активно применяться для диагностики туберкулеза в регионах, имеющих недостаточное обеспечение квалифицированными рентгенодиagnстами. Ярким примером такой системы является CAD4TB [153,154]. Обогащение дополнительными клиническими данными (такими как температура, потливость и кровохарканье) позволило увеличить эффективность диагностической CAD4TB и достичь AUC 0,84, чувствительности – 0,95, при этом специфичность была 0,49 [155].

В 2000-х годах активно стали развиваться CAD-системы для анализа рентгенографии органов грудной клетки с целью выявления легочных узлов. Для этих систем использовались различные комбинации алгоритмов обработки изображений [156]. Зачастую использовался подход построения разностных изображений, при котором нормальный фон удалялся, а легочные узлы оставались. Далее осуществлялась дифференцировка легочных узлов по группам, в соответствии с ранее установленными пороговыми значениями [157]. Результаты работы таких CAD-систем на тот момент были многообещающими.

Первый CAD-алгоритм, получивший сертификат соответствия FDA (Food and Drug Administration), был RapidScreen Digital. Примечательно, но этот алгоритм обладал далеко не 100% точностью, однако в ходе испытаний он позволил выделить 36 случаев (41%) из 88, в которых врач-рентгенолог пропустил образование, подозрительное на ЗНО легкого [157]. По этой причине данную систему рекомендовали как второе мнение, позволяющее минимизировать пропуск злокачественных новообразований в легких [158]. Примечательно, что в вопросе использования CAD-систем для диагностики злокачественных новообразований легких на рентгенограммах после 2010 года нет значимых продвижений. Все дело в том, что в основу CAD закладывался принцип анализа всего изображений или его части на основе функций, причем эти функции подбираются разработчиками вручную. Ожидалось, что такой подход будет применяться и далее, однако на практике невозможно было описать все возможные варианты признаков. По этой причине точность CAD для диагностики злокачественных новообразований легких не могла достигнуть таковой, как для врача-рентгенолога экспертного уровня. Отчасти именно поэтому известные CAD-системы фокусировались на двух основных патологиях на рентгенограмме органов грудной клетки: туберкулезе и раке легкого, в то время как признаки другой патологии оставались без должного внимания [159].

Процесс принятия решения о том, какие функции являются оптимальными для решения конкретной задачи, очень сложен. Как правило, невозможно доказать, что набор признаков является оптимальным; выбор набора функций – это в некотором смысле больше искусство, чем наука [160]. Однако благодаря научно-техническому прогрессу после 2010 года задача оптимизации относительно просто стала решаться при помощи машинного обучения. Может показаться парадоксальным, но машина гораздо точнее выделяет нужные признаки из большого объема и вычисляет решающую функцию. Поэтому следующая ветка развития CAD-систем уходит в гибридное направление CAD + ИИ.

Стоит отметить, что в отличие от рентгенографии органов грудной клетки CAD-системы для маммографии оказались более точными и продолжают успешно применяться и по сей день. Принцип работы CAD-систем с цифровой маммограммой принципиально не отличается от работы с рентгенографией органов грудной клетки. На вход системе подаются изображения (до четырех), при этом обработка происходит зачастую независимо. При помощи CAD выявляются аномалии в ткани молочной железы: скопления микрокальцинатов и уплотнений, подозрительных на новообразования. Затем включается следующий алгоритм CAD, который осуществляет классификацию этих областей, используя количественные инструменты оценки таких характеристик, как пространственная плотность и регулярность кластера. Вероятность малигнизации оценивается при помощи третьего модуля – автоматизированного обучаемого классификатора. В результате врач-рентгенолог видит на изображении области, отмеченные алгоритмом как подозрительные на злокачественный процесс [161].

Первый одобренный FDA алгоритм CAD для маммографии был зарегистрирован раньше, чем для рентгенографии органов грудной клетки – в 1998 г. [162]. Эффективность применения CAD для выявления рака молочной железы была настолько многообещающая, что уже к 2010 году примерно 74% всех скрининговых и диагностических маммографических исследований

описывались с применением CAD [163]. Стоит отметить, что CAD в маммографии стали использовать в качестве альтернативы второму чтению. Первые исследования демонстрировали прирост эффективности в обнаружении рака молочной железы по разным исследованиям от 2 до 10%, что ускоряло его раннее выявление на 2 месяца. Единственное, что ограничивало от повсеместного внедрения – экономическая эффективность [164–166]. Однако спустя некоторое время стали появляться работы, ставящие под сомнение эффективность CAD-систем [167,168], а последующие исследования вообще указывали на то, что врачи-рентгенологи не меняли своих решений, что не позволило повысить точность исследования [169].

Все это может быть свидетельством того, что врачи обучались, и в конечном счете, когда их способности сравнивались по чувствительности с CAD-системами, они перестали быть актуальными в качестве второго чтения. По описанной ранее логике, CAD-системы вышли на предел своих возможностей, став альтернативой глазам рентгенологов, но превзойти их по точности не смогли. А. Кохли и С. Йха (A. Kohli, S. Jha) в своем исследовании «Почему CAD потерпел неудачу в маммографии?» (Why CAD Failed in Mammography?) высказывались о недостаточности вычислительных мощностей и об отсутствии контролируемого дообучения CAD-систем [170]. Авторы этого исследования указывали на то, что для востребованности CAD-система должна «уметь видеть то, что мы не можем». Таким образом, в маммографии также заканчивается эра простых CAD-систем и начинается эра применения машинного обучения для анализа исследований.

Основополагающим отличием CAD-систем последнего поколения, базирующихся на технологиях ИИ [171], от предыдущих поколений является принцип их создания: это больше не запрограммированный свод эмпирических правил и обобщение статистической информации, а сложная система связей, натренированная на больших массивах данных с минимальной вовлеченностью оператора. В связи с этим ключевым фактором успешного развития таких

направлений являются качественные наборы данных, собранные и структурированные из больших массивов медицинских данных [172]. Появление их в открытом доступе [173–176], а также доступность больших вычислительных мощностей привели к экспоненциальному росту исследований в области применения ИИ в лучевой диагностике (рисунок 1) [177].

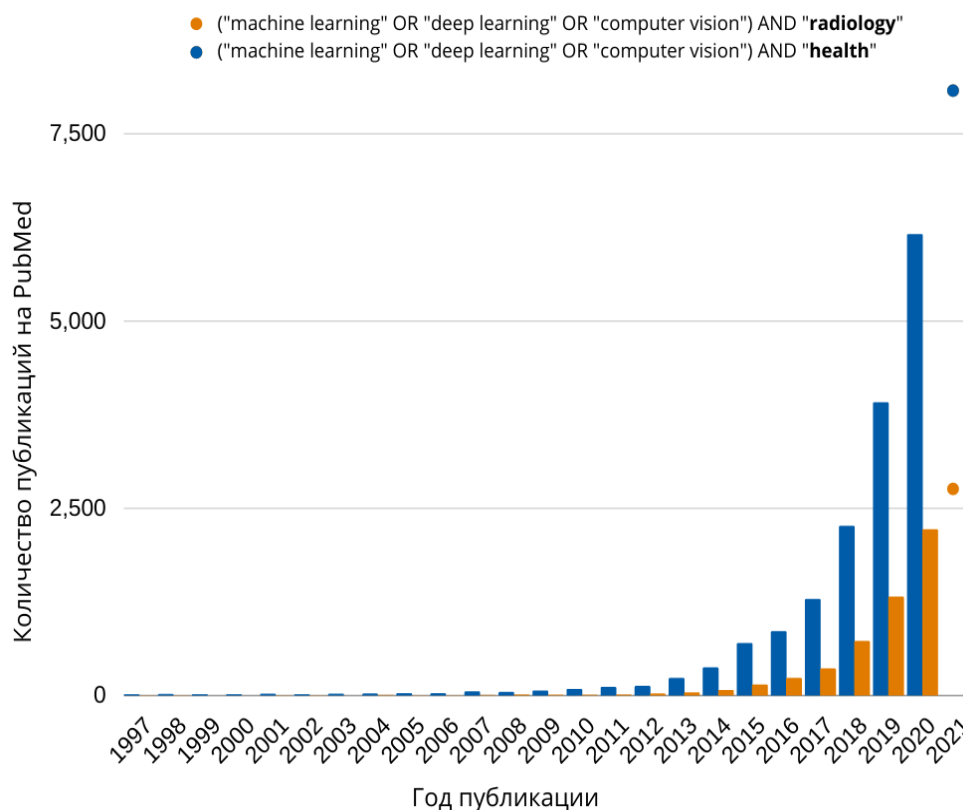


Рисунок 1 – Количество публикаций, посвященных теме машинного обучения в здравоохранении (синий) и радиологии (оранжевый) по данным PubMed

Логичным продолжением научной публикации является внедрение алгоритма в клиническую практику, а это возможно только после регистрации медицинского изделия. На рисунке 2 приведена динамика роста количества зарегистрированных в FDA сервисов на основе технологии ИИ.

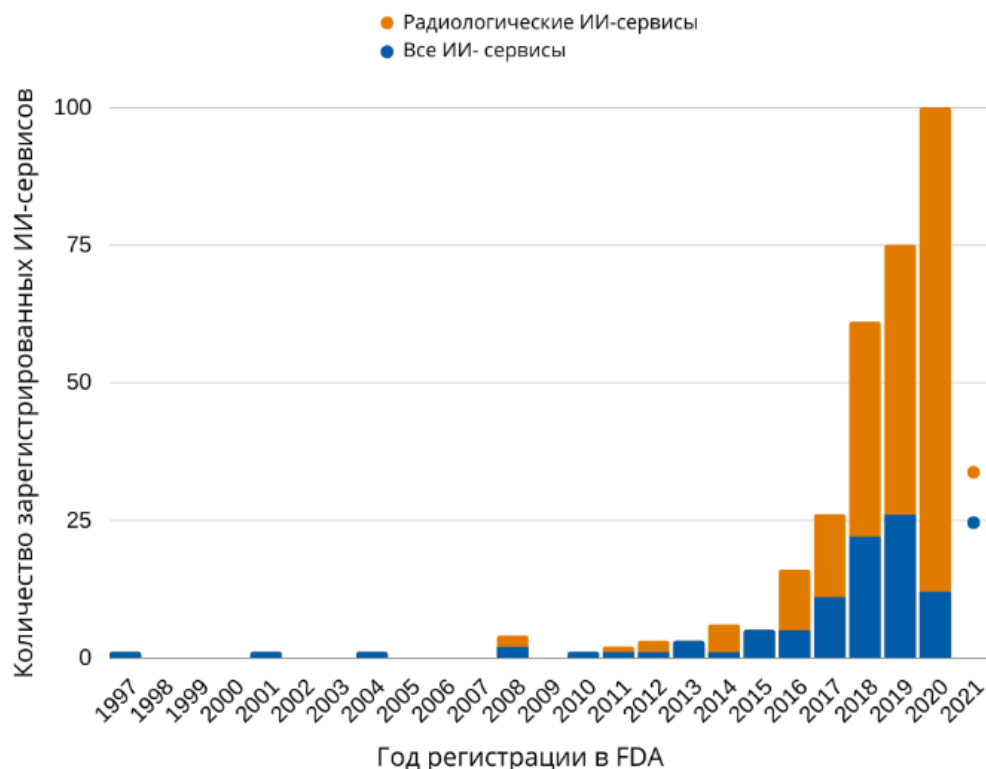


Рисунок 2 – Количество зарегистрированных медицинских изделий в США на основе ТИИ (ИИ-сервисов) в лучевой диагностике (оранжевый), в других направления медицины (синий)

Далее в тексте диагностические алгоритмы и системы, использующие искусственный интеллект, будут определяться термином «программное обеспечение на основе технологий искусственного интеллекта». Принципиальное отличие ПО на основе ТИИ от ранее рассмотренных экспертных систем заключается в том, что оно способно обучаться и накапливать опыт [178].

Ключевые области применения искусственного интеллекта включают: 1) технологии обработки естественного языка (NLP) для распознавания текста и речи; 2) компьютерное зрение (CV) для идентификации, классификации и анализа признаков на диагностических изображениях; 3) аналитику больших данных (Data Science) для обнаружения закономерностей в данных, их структурирования, извлечения новых знаний и прогнозирования результатов [179]. В данной главе будут рассмотрены только ПО на основе ТИИ для компьютерного зрения с целью классификации исследования, а также локализации патологических областей.

ПО на основе ТИИ используют два основных подхода: машинное обучение и глубокое обучение. Машинное обучение – тип анализа, при котором важность показателей определяется машиной [180,181]; глубокое обучение имеет больше параметров, моделируя архитектуру мозга [181]. Для практического применения ПО на основе ТИИ необходимы доказательства качества работы такого ПО [181].

Для диагностики туберкулеза имеются два большие систематические обзора с метаанализом, позволяющие оценить тенденции в этой области касательно перспектив ПО на основе ТИИ: первый, опубликованный в 2016 году [182], и второй – в 2019 [183]. Подробнее остановимся на результатах наиболее актуального исследования. Авторы разделили все публикации на посвященные разработки и более редкие – клинические. Результаты диагностической точности первых существенно превосходят клинические исследования (AUC 0,88 [0,82–0,90]) против 0,75 [0,66–0,87]. Авторы исследования указывают на перспективность разработки ПО на основе ТИИ для диагностики туберкулеза, однако клинической оценке таких алгоритмов уделяется недостаточно внимания [183]. Технологии ИИ стремительно развиваются, поэтому результаты, указанные в этих обзорах на данный момент, уже полностью устарели. Актуальный обзор, выполненный К. Сантош и соавторами (K. C. Santosh et al.) и опубликованный в октябре 2022 года, содержит существенно более высокие показатели диагностической точности. Так, AUC в актуальных работах последних 3 лет превысил порог в 0,90, а, по данным отдельных исследователей, достиг значения 0,99 [184,185].

Пандемия COVID-19 продиктовала необходимость разработки алгоритмов, позволяющих диагностировать признаки вирусной пневмонии по рентгенограмме органов грудной клетки. Как показал один из актуальных на сегодняшнее время систематических обзоров ПО на основе ТИИ для диагностики COVID-19, подавляющее большинство актуальных архитектур позволяют достигнуть значения AUC более 0,95, самые точные имели AUC более 0,99 [186–189].

При анализе литературы на предмет алгоритмов, выявляющих легочные узлы, подозрительные на злокачественные образования, удалось обнаружить ограниченное количество исследований. Так, за последние 3 года был выявлен только один систематический обзор с метаанализом, который на фоне различной патологии органов грудной клетки также упоминал ПО на основе ТИИ для диагностики рака легкого по рентгенограмме. По данным авторов, AUC для выявления легочных узлов составил 0,884 (0,842–0,925). При этом чувствительность равняется 0,75 (0,634–0,866), специфичность – 0,944 (0,912–0,976). Выявление признаков злокачественных новообразований легких по данным цифровой рентгенограммы продемонстрировало AUC 0,864 (95% ДИ 0,827–0,901) для рентгенографии, при этом чувствительность составила 0,801 (0,683–0,919), специфичность не указана. [185]. Как следует из приведенных данных, ПО на основе ТИИ не позволяет выявлять отдельную патологию (за исключением туберкулеза и COVID-19) с чувствительностью, близкой к 100%.

Особенность рентгенографии органов грудной клетки заключается в том, что ПО на основе ТИИ должно выявлять более 1 целевой патологии. Существуют алгоритмы, позволяющие проводить бинарную классификацию: норма/патология. Они прекрасно подходят для триажа (приоритизации исследований в рабочем списке врача-рентгенолога). В среднем диагностическая точность (AUC), которую демонстрирует ПО на основе ТИИ для выявления патологии в исследовании в целом составляет 0,917 (0,869–0,966), чувствительность – 0,873 (0,762–0,985), специфичность – 0,894 (0,860–0,929) [185].

Разработчики алгоритмов зачастую демонстрируют впечатляюще высокие показатели точности ПО на основе ТИИ [183,184,190]. Примечательно, что при разработке такого программного обеспечения исследователи по-прежнему уделяют мало внимания сопоставлению полученной диагностической точности с врачом-рентгенологом. Например, из 39 исследовательских работ в области ПО на основе ТИИ, анализирующих рентгенограммы органов грудной клетки, только в 13 сопоставляли полученные результаты с результатами врачей,

что позволило бы более объективно взглянуть на полученные результаты. Наибольшую ценность представляет сопоставление показателей диагностической точности, полученных ПО на основе ТИИ, и врачом-рентгенологом – на одном и том же наборе данных. К сожалению, при анализе литературы удалось обнаружить только один обзор, датированный 2019 годом и систематизировавший такие исследования [191]. Однако уже по состоянию на 2019 год авторы обзора с осторожностью заявляли, что точность ПО на основе ТИИ эквивалентна точности медицинских работников.

Необходимо обратить внимание и на этические вопросы: из научных статей мы понимаем, что ПО на основе ТИИ уже не уступает точности врача-рентгенолога, а в отдельных исследованиях конкретно указывается на его превосходство. Казалось бы, логичным начать использовать такое ПО, однако, как было показано выше, точность машины лишь для отдельных патологий достигает 99%, т.е. на практике ошибки неизбежны. Получается, что ПО на основе ТИИ на данный момент обладает точностью, недостаточной для замены собой врача-рентгенолога [192]. В этой связи возникает парадокс – применить ПО на основе ТИИ изолированно мы не можем, но при этом врач, как следует из ранее опубликованных работ, ошибается чаще. На этом фоне наблюдается активное обсуждение (в том числе в профессиональном сообществе) целесообразности применения ПО на основе ТИИ в лучевой диагностике [193–197], продолжающееся до настоящего момента [198].

Выход из сложившейся ситуации нашли исследователи, предлагающие использование врача-рентгенолога, результат работы которого будет усилен таким ПО. Такой тандем приводит к повышению диагностической точности по сравнению с результатами работы одного врача-рентгенолога [190].

Эффективность решений на основе ТИИ в области лучевой диагностики подтверждена многочисленными научными работами. ПО на основе ТИИ по точности определения отдельных рентгенологических признаков находится на уровне среднестатистического врача-рентгенолога. Автоматизированный анализ

рентгенологических исследований демонстрирует чувствительность, специфичность и площадь под ROC-кривой в 75,4%, 90,6% и 0,89 соответственно, сопоставимые с показателями врача-рентгенолога, которые составляют 73,0%, 88,6% и 0,85, при отсутствии статистически значимых различий [199]. Согласно метаанализу, автоматизированный анализ маммографии (ММГ) показывает чувствительность, специфичность и площадь под кривой в 91,4%, 91,6% и 0,945 соответственно [200]. При решении задачи сегментации патологических областей коэффициенты сходства IoU для алгоритма на основе сверточных нейронных сетей и для врача-рентгенолога составляют 0,86 и 0,96 соответственно [201].

Нет оснований отрицать необходимость применения ПО на основе ТИИ в клинической практике врача-рентгенолога. Однако, для того чтобы такое ПО вышло за пределы лабораторных условий, необходимо осуществить его регистрацию как медицинского изделия. К концу декабря 2022 года 29 ПО на основе ТИИ получили европейскую сертификацию CE MDR/MDD как медицинские изделия, из которых 11 также сертифицированы в США [202]. Особое внимание заслуживает программный продукт «ChestLink®», который имеет разрешение на использование в качестве медицинского изделия для полностью автономного анализа рентгенограмм органов грудной клетки [203]. «ChestLink®» проводит сортировку исследований, выявляет рентгенограммы органов грудной клетки без патологии, формирует полный протокол описания, не требующий валидации со стороны врача-рентгенолога; такой подход снижает нагрузку на врача, позволяя сосредоточиться на случаях с патологиями [204]. На наш взгляд, это очень важный шаг, требующий детальной оценки и проработки – есть ли тот допустимый минимальный уровень погрешности (ошибки) ПО на основе ТИИ?

Выше был рассмотрен вид исследования «цифровая рентгенография органов грудной клетки», по которой достаточно масштабно ведется разработка ПО на основе ТИИ: за период с 2020 по 2022 гг. было опубликовано 2248 работ [205]. При этом в части цифровой флюорографии количество международных

исследований, изучающих возможности применения ПО на основе ТИИ, исчисляется единицами. Так, удалось найти упоминание о том, что в 2018 был представлен алгоритм для анализа цифровых флюорограмм органов грудной клетки от компании Google [206]. Однако никаких подробностей в части значений диагностической точности не раскрывается. Необходимо отметить, что флюорография в настоящее время как метод активно используется в Российской Федерации и странах СНГ. Поэтому работы по применению ПО на основе ТИИ для флюорографии выполняются в основном нашими соотечественниками [207–210]. Например, исследователям удалось достичь уровня диагностической точности при выявлении туберкулеза – 91,31% [208], ~0,945 [209].

В настоящее время в научной литературе можно найти немного работ, целью которых является оценка усредненной диагностической точности ПО на основе ИИ для анализа профилактических исследований в лучевой диагностике. При этом, общее количество работ, посвященных анализу диагностической точности отдельных решений на основе ИИ в лучевой диагностике, напротив, стремительно растет [211–214]. Необходимо отметить, что при принятии решения о масштабном применении ТИИ в лучевую диагностику для профилактических исследований важно понимать показатели диагностической точности не только самого ПО на основе ТИИ, но и врача-рентгенолога, что в настоящее время изучено недостаточно.

В последние пять лет в мире развиваются отдельные решения на основе ТИИ, способные анализировать цифровые рентгенологические маммографические исследования. Значительное число публикаций, посвященных выявлению признаков злокачественных новообразований молочной железы [190,192,215,216], обусловлено высокой нагрузкой на врачей-рентгенологов при скрининге рака молочной железы.

Имеются исследования, подтверждающие, что чувствительность некоторых ИИ-алгоритмов находится на одном уровне с врачами-рентгенологами [217]. В то же время отдельные научные работы заявляют, что ПО, разработанное

на основе ТИИ, может демонстрировать более высокую диагностическую точность по сравнению с врачом [215]. Также имеются противоречивые мнения об использовании ПО на основе ТИИ при сравнении с врачами-рентгенологами, имеющими разный опыт. С одной стороны, исследования продемонстрировали способность ПО на основе ТИИ достигать уровня производительности врачей-рентгенологов [192]. С другой – точность интерпретации результатов маммографии опытными рентгенологами сильно различается [218], а ранее рассмотренные CAD-системы не позволили значительно улучшить ее [219].

Обобщив вышесказанное, можно с уверенностью констатировать, что в условиях быстро растущего количества рентгенологических исследований применение ПО на основе ТИИ в качестве первого/второго чтения становится целесообразным. Это снижает нагрузку на врачей-рентгенологов и повышает точность диагностики [212,220]. Недавние исследования показали, что диагностическая точность моделей ИИ для медицинской визуализации приближается к показателям медицинских экспертов и даже превосходит их в некоторых областях. Это привело к быстрому росту количества зарегистрированных и коммерчески доступных моделей искусственного интеллекта для анализа маммографии и рентгенографии органов грудной клетки [221]. Таким образом, модели ИИ активно интегрируются в рабочий процесс радиологии. Однако, хотя их показатели точности, заявленные разработчиками, довольно высоки, их реальные характеристики необходимо тщательно оценивать и сравнивать с показателями рентгенологов, чтобы обеспечить клиническую ценность и безопасность [190,192,215,222].

Ранее уже была показана целесообразность применения ПО на основе ТИИ для скрининга злокачественных новообразований молочной железы [223]. В ряде исследований авторы доказали эффективность применения ИИ в качестве одного из этапов чтения исследования [192,224]. Было доказано, что комбинация врача-рентгенолога и ПО на основе ТИИ имеет более высокую чувствительность и меньшую специфичность по сравнению с одним врачом и статистически

незначимые отклонения показателей диагностической точности по сравнению классическим двойным чтением маммографического исследования [225]. Последующее исследование показало увеличение показателей диагностической точности при использовании следующей схемы: на первом этапе ММГ-исследование анализируется с помощью ИИ-алгоритма, далее второе чтение проводит врач-рентгенолог. Такая схема позволяет повысить чувствительность метода за счет алгоритма, а специфичность – за счет врача [226].

Ожидания от внедрения ИИ в рентгенографию также остаются неоднозначными: даже многообещающие результаты ПО на основе ТИИ с показателями высокой точности связаны с ограниченной специфичностью классификации конкретных патологий. Таким образом, на данный момент ПО на основе ТИИ едва ли может заменить врача-рентгенолога [214,227,228]. Поэтому ВОЗ предупреждает, что, несмотря на очевидные преимущества, внедрение ИИ в медицину сопряжено с рисками, которые следует минимизировать [229].

Вышеописанные исследования демонстрируют потенциальные возможности применения ПО на основе ТИИ для ММГ, рентгенографии и флюорографии органов грудной клетки. Определение пороговых значений для диагностической точности алгоритмов ИИ является ключевым для оценки их пригодности к практическому использованию. В научной литературе упоминаются рекомендации относительно минимального значения площади под ROC-кривой, приемлемого для клинического применения [226]. Порог в 0,81 для AUC фактически указывает, что до 19% решений, сделанных ПО на основе ТИИ, могут быть неверными. Однако специфические пороговые значения для диагностической точности ПО, анализирующего профилактические исследования, в настоящее время отсутствуют.

В целом разнообразие представленных на рынке ПО для применения в медицине позволяет пользователям выбрать наиболее подходящие решения в зависимости от специфики как самих задач, так и медицинских организаций [230]. В рекомендациях Recht и соавторов были озвучены многие проблемы

внедрения и применения ПО на основе ТИИ в лучевой диагностике. При этом обращает на себя внимание детальный анализ этических проблем и отсутствие конкретных требований к технической интеграции таких решений [231].

В настоящее время существуют единичные успешные случаи внедрения ПО на основе ТИИ в работу отдельных медицинских организаций, чаще всего на базе академических медицинских центров. Однако существует колоссальная разница между внедрением технологий ИИ в масштабах одной локальной организации и в масштабах целой системы медицинских организаций. Кроме самих технических проблем, следует выделить проблемы, связанные с применением результатов работы ПО на основе ТИИ. Так, в академических центрах пользователями результатов работы ПО являются в основной массе эксперты и научные сотрудники, в то время как внедрение этих технологий в масштабе Департамента здравоохранения города Москвы затронет сотни врачей-рентгенологов, различающихся по опыту и квалификации, которые на первых порах будут воспринимать такое ПО как своего конкурента, поэтому и оценка работы ПО будет предвзятой. С технической точки зрения работа ПО на основе ТИИ в масштабах одной медицинской организации и в масштабах мегаполиса будет различаться объемом исследований, направляемых на анализ, что потребует увеличения вычислительных мощностей и пропускной способности каналов связи для обеспечения требуемой скорости обработки исследований.

Критерии, которые устанавливают медицинские организации и врачи-рентгенологи для оценки результатов рентгенологических исследований, нуждаются в стандартизации. Следующим шагом должна быть разработка на основе стандартизированных требований к такому ПО методики отбора и мониторинга качества ПО на основе ТИИ. Только в этом случае будет происходить развитие и совершенствование ПО, а конечный пользователь сможет выбрать качественные ИИ-решения, способные удовлетворить специфическим требованиям, предъявляемым к профилактическим или диагностическим исследованиям.

В любом случае, краеугольным камнем внедрения ТИИ являются научная и клиническая обоснованность метода, актуальность и удобство целевого применения. Что же касается практической эффективности методов обработки рентгеновских исследований с помощью ПО на основе ТИИ, то она может быть достигнута в конечном счете благодаря автоматизации процессов подготовки протокола исследования. Это позволит сократить трудозатраты врача и повысить его эффективность. Таким образом можно будет достичь цели автоматизации процессов с помощью интеллектуальных технологий.

В последние годы активно ведутся разработки отечественного ПО на основе ТИИ по всем направлениям профилактических исследований. Так, в Российской Федерации по состоянию на декабрь 2022 года статус медицинского изделия получен двумя программными продуктами на основе технологий ИИ для автоматизированного анализа результатов флюорографии (ФЛГ) и рентгенографии (РГ) компании ООО «Медицинские скрининг системы» (РЗН 2021/14449) и ООО «ФтизисБиоМед (РЗН 2022/17406), а также двумя программными продуктами для автоматизированного анализа ММГ (компания ООО «Медицинские скрининг системы» (РЗН 2021/14449) и Экосистемный интегратор «СберМедИИ»). Это указывает на потенциальную возможность применения ПО на основе ТИИ в медицинской практике.

В 2020 году нами было проведено социологическое исследование [232], направленное на выяснение отношения пациентов к ТИИ в медицине. Опрос охватил более 600 респондентов, большинство из которых относятся положительно или нейтрально к применению ТИИ, а 67% считают целесообразным применение ТИИ в здравоохранении, 57% считают, что качество медицинской помощи станет лучше при применении ТИИ. Таким образом, большая часть пациентов положительно оценивают применение ТИИ в медицине и ожидают, что итоговое заключение об исследовании будет сформировано врачом при поддержке ИИ.

Таким образом, переход к цифровым беспленочным технологиям в области рентгенологии привел к появлению инновационных решений, таких как PACS и RIS, что в свою очередь способствовало внедрению ТИИ. Эти технологии направлены на повышение качества лучевой диагностики и оптимизацию рабочих процессов отделений лучевой диагностики. Вследствие того, что эти инновации начали активно внедряться относительно недавно не только в России, но и во всем мире, имеется ряд задач и проблем, решение которых актуально и социально значимо [233].

Заключение. Обобщая результаты анализа обзора литературы, в порядке заключения следует отметить основные актуальные проблемы и задачи, которые требуют решения. Профилактические рентгенологические исследования переходят на цифровые технологии, что открывает безграничный потенциал не только для научных исследований, но и для повышения качества и безопасности самих исследований. Цифровизация и централизация, ставшие возможными благодаря появлению PACS и РИС, позволили повысить качество описания рентгенологических исследований, а также стали основой внедрения технологий постобработки, а в последующем и технологий на основе искусственного интеллекта.

Отдельные результаты применения ПО на основе ТИИ для массовых профилактических исследований указывают на перспективность этих технологий. Появляется возможность повысить показатели диагностической точности, что в конечном счете снижает количество пропусков патологии и повышает качество оказываемой медицинской помощи. На текущий момент времени рано говорить о замене врача-рентгенолога технологиями на основе искусственного интеллекта, однако уже стали появляться автономные алгоритмы, позволяющие без дополнительного контроля со стороны врача классифицировать рентгенологические исследования на «без патологии» и «с патологией».

Основная проблема заключается в отсутствии четкого регламента по применению ПО на основе ТИИ для профилактических рентгеновских исследований, а также критериев для допуска этого ПО к клиническому применению. В отличие от врача-рентгенолога ПО может быть настроено для оптимальной работы в конкретных условиях.

Большинство рассмотренных исследований базировалось на ограниченном наборе данных, на котором проводилась оценка результатов работы ПО на основе ТИИ, при этом единичные исследования касались клинической валидации, большая часть из которых заключалась в ретроспективном пересмотре исследований с целью выявления пропуска патологии.

Готовность общества к повсеместному внедрению инноваций, активное развитие отечественного ПО на основе ТИИ с его последующей регистрацией в качестве медицинского изделия, а также введенные в действие нормативные документы о внедрении и развитии цифровых технологий и искусственного интеллекта в частности, являются основой для массового внедрения технологий на основе искусственного интеллекта в практическое здравоохранение Российской Федерации.

Таким образом, очевидна необходимость создания организационных подходов, основанных на ТИИ и направленных на совершенствование массовых профилактических мероприятий. Требуют разработки принципы организации системного внедрения ТИИ для автоматизации производственных процессов в сфере здравоохранения. Должна быть научно обоснована система обеспечения безопасности и качества медицинской помощи (в частности, профилактических исследований), оказываемой с применением технологий искусственного интеллекта.

Глава 2. МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Общий ход и этапы исследования

Настоящая работа состоит из последовательных связанных шести этапов, соответствующих задачам проведенного исследования. Результаты работы, полученные на текущем этапе, являются входными данными для последующего этапа. Программа исследования представлена в таблице 1.

Таблица 1 – Программа исследования

Цель исследования – совершенствование организации и проведения профилактических и диагностических лучевых исследований с использованием технологий искусственного интеллекта.		
Задача	Источники и объем информации	Методы анализа
1. Изучить проблемы управления профилактическими и диагностическими исследованиями, осуществляемыми с применением методов лучевой диагностики.	Дашборд Business intelligence (BI) ЕРИС ЕМИАС	эпидемиологический, статистический
2. Оценить применимость технологий искусственного интеллекта в лучевой диагностике путем сравнительного изучения точности диагностических решений, принимаемых врачом самостоятельно и формируемых автоматически.	База данных (MONGO Express) платформы оценки врачебной точности, врачей-рентгенологов: 113 (ММГ), 185 (РГ) и 69 (ФЛГ). Тестовый контур ЕРИС – выгрузка в формате xls, а также DICOM – файлы с результатами анализа лучевых исследований ПО на основе ТИИ Наборы данных лучевых исследований из ЕРИС ЕМИАС: 269 (ММГ), 140 (РГ), 184 (ФЛГ).	Статистические (ROC-анализ, проверка гипотез)
3. Разработать, внедрить и оценить результативность организационной технологии автоматизации рабочего процесса врача-рентгенолога при проведении профилактических и диагностических исследований.	Выгрузки формата 3.1, предусмотренная Приказом ДЗМ от 19.02.2020 № 142, объединенные в базу данных в формате pandas dataframe (более 6 млн. исследований). Тестовый контур ЕРИС ЕРИС ЕМИАС	Аналитические, статистические

Продолжение таблицы 1

Задача	Источники и объем информации	Методы анализа
4. Обосновать организационно-методические основы контроля безопасности и качества программного обеспечения на основе технологий искусственного интеллекта для профилактических и диагностических лучевых исследований.	Выгрузки формата 3.1, предусмотренная Приказом ДЗМ от 19.02.2020 № 142, объединенные в базу данных в формате pandas dataframe (более 6 млн. исследований). Тестовый контур ЕРИС ЕРИС ЕМИАС Методические рекомендации о проведении клинических испытаний программного обеспечения на основе интеллектуальных технологий (лучевая диагностика)	Аналитические, статистические
5. Обосновать и внедрить модель организации и проведения профилактических и диагностических исследований с применением технологий искусственного интеллекта.	База данных (MONGO Express) платформы оценки врачебной точности, врачей-рентгенологов: 113 (ММГ), 69 (ФЛГ). Тестовый контур ЕРИС – выгрузка в формате xls, а также DICOM – файлы с результатами анализа лучевых исследований ПО на основе ТИИ Наборы данных лучевых исследований из ЕРИС ЕМИАС: 100 (ММГ), 100 (ФЛГ).	Аналитические, статистические
6. Оценить результативность автоматизации процессов организации и проведения профилактических и диагностических исследований	Выгрузки формата 3.1, предусмотренная Приказом ДЗМ от 19.02.2020 № 142, объединенные в базу данных в формате pandas dataframe (10 тыс. ММГ, 25 тыс. ФЛГ исследований). ЕРИС ЕМИАС	Аналитические, статистические

На **первом** этапе выполнялся анализ организации и проведения массовых профилактических исследований, осуществляемых с применением методов лучевой диагностики. В качестве целевого региона была выбрана Москва, целевых медицинских организаций – амбулаторные медицинские организации, подведомственные Департаменту здравоохранения города Москвы. Изучались оснащенность диагностическим оборудованием для проведения лучевых

профилактических исследований, среднее количество проводимых исследований и их динамика за последние 3 года. Также оценке подлежало количество выявленных патологических случаев для каждого типа профилактических исследований. В качестве источника данных использовался инструментарий ВІ ЕРИС ЕМИАС. Этот этап являлся необходимым для формирования клинических (диагностических) требований к решениям на основе ТИИ в целях профилактических лучевых исследований.

На **втором** этапе была проведена оценка диагностической точности «среднего» врача-рентгенолога, что позволило определить критерии применимости ТИИ для профилактических исследований. Были протестированы решения на основе ТИИ для профилактических исследований, а полученные значения диагностической точности сопоставлены с таковыми для врача-рентгенолога. Таким образом были определены минимальные пороговые значения для принятия решения о целесообразности внедрения указанного ПО. Этот этап проводился на эталонных наборах данных с применением специально подготовленной платформы для проведения тестирования врачей-рентгенологов. Целевые направления лучевых исследований: рентгенография/флюорография органов грудной клетки и маммография. Необходимость второго этапа заключалась в формировании методов и критериев отбора решений на основе ТИИ.

На **третьем** этапе была создана методика внедрения ТИИ в рабочий процесс врача-рентгенолога. В качестве ПО на основе ТИИ выступали разработки на базе технологии «компьютерное зрение», именуемые ИИ-сервисами (детальное описание представлено в параграфе 2.4. Данные решения разворачивались с применением облачных технологий, подключались к городской PACS-системе (ЕРИС ЕМИАС) и должны были предоставлять результаты работы в привычной рабочей среде врача-рентгенолога. На этом этапе формировалась максимально удобная для конечного пользователя, «бесшовная» интеграция ПО на основе ТИИ. В рамках третьего этапа были разработаны технические требования к ПО на основе ТИИ.

Наконец, на **шестом** этапе были проанализированы результаты ранее реализованной модели применения ТИИ для массовых профилактических исследований и оценена целесообразность дальнейшего масштабирования.

```
graph TD; A[Анализ организации и проведения массовых профилактических исследований с применением методов лучевой диагностики (задача 1)] --> B[Определение критериев для оценки целесообразности внедрения в клиническую практику]; B --> C[Сопоставление показателей диагностической точности врача (n>100) и ПО на основе ТИИ для РГ/ФЛГ, ММГ на одинаковых наборах данных (задача 2)]; C --> B; B --> D[ПО на основе ТИИ для лучевых профилактических исследований: ММГ (3) и РГ/ФЛГ (6)]; B --> E[Внедрение в рабочий процесс врача-рентгенолога (задача 3)]; B --> F[Работа ПО на основе ТИИ на потоке]; B --> G[Концепция массовых профилактических исследований с применением ТИИ (задача 5)]; B --> H[Анализ эффективности применения ПО на основе ТИИ для профилактических исследований (задача 6)]; D --> E; E --> F; F --> G; G --> H; I[Разработка комплексной системы контроля качества] --> J[Разработка инструментария контроля качества]; J --> K[Инструменты контроля качества (задача 4)]; I --> E; J --> F; K --> G; K --> H;
```

Анализ организации и проведения массовых профилактических исследований с применением методов лучевой диагностики *(задача 1)*

Определение критериев для оценки целесообразности внедрения в клиническую практику

Сопоставление показателей диагностической точности врача ($n > 100$) и ПО на основе ТИИ для РГ/ФЛГ, ММГ на одинаковых наборах данных *(задача 2)*

ПО на основе ТИИ для лучевых профилактических исследований: ММГ (3) и РГ/ФЛГ (6)

Внедрение в рабочий процесс врача-рентгенолога *(задача 3)*

Работа ПО на основе ТИИ на потоке

Концепция массовых профилактических исследований с применением ТИИ *(задача 5)*

Анализ эффективности применения ПО на основе ТИИ для профилактических исследований *(задача 6)*

Разработка комплексной системы контроля качества

Разработка инструментария контроля качества

Инструменты контроля качества *(задача 4)*

Рисунок 3 – Схема диссертационного исследования

Результаты первого, второго и третьего этапов приведены в главе 3, результаты четвертого этапа описываются в главе 4, результаты пятого и шестого этапов отражены в главе 5. Далее будут детально описаны материалы и методы, необходимые для реализации каждого из представленных этапов.

2.2. Информационная инфраструктура

Для реализации первого этапа необходимо было изучить текущую организацию работы рентгеновских отделений с массовыми профилактическими исследованиями. Это предполагает получение и анализ данных в первую очередь о работе диагностического оборудования, его количестве и загруженности. Эту информацию можно было бы собирать отдельно по медицинским организациям, но всё это оборудование объединено в единую систему. Основой системы, на которой была реализована настоящая работа, является Единый радиологический информационный сервис (ЕРИС) Единой медицинской информационно-аналитической системы (ЕМИАС).

ЕМИАС представляет собой государственную информационную систему города Москвы, обеспечивающую автоматизацию процессов организации и оказания медицинской помощи населению в городе Москве. ЕМИАС обрабатывает и сохраняет данные, получаемые в результате оказания медицинских услуг населению города Москвы. Одной из составных частей (сервис) ЕМИАС является Единый радиологический информационный сервис (ЕРИС), который был создан в рамках проекта информатизации здравоохранения Москвы.

ЕРИС ЕМИАС представляет собой информационную систему, состоящую из трех компонентов: диагностические устройства, PACS для хранения цифровых изображений и АРМ врачей-рентгенологов.

ЕРИС ЕМИАС выполняет следующие задачи:

1. Регистрация данных (о пациенте, об исследовании и оборудовании, о медицинской организации и медицинском персонале)
2. Передача данных.
3. Создание заключений о проведенном исследовании.
4. Хранение данных.
5. Анализ качества выполнения и интерпретации диагностических исследований.
6. Защита данных.

Система ЕРИС ЕМИАС была выбрана в качестве основы для оценки возможности применения ТИИ в диагностике, что обуславливалось, в первую очередь, большим потоком данных в системе. Однако описываемый далее подход может быть применен к любой МИС, осуществляющей работу с цифровыми форматами диагностических данных.

Для обеспечения безопасности и контроля качества ПО на основе ТИИ до этапа работы на потоке осуществляется тестирование в так называемом тестовом программно-аппаратном контуре – ТПАК, являющемся копией ЕРИС, имеющей идентичный функционал, только с тем отличием, что к нему не подключены диагностические устройства (ДУ) и не работают врачи-рентгенологи, а все находящиеся в нем исследования полностью анонимизированы. В ТПАК содержится ограниченное количество исследований, которых достаточно для проверки функционала и контроля показателей диагностической точности. При этом взаимодействие с ПО на основе ТИИ осуществляется аналогично целевой модели работы в ЕРИС. Для работы с ПО на основе ТИИ в ТПАК загружаются наборы данных – исследования из ЕРИС ЕМИАС, предварительно анонимизированные с учетом требований и стандартов анонимизации [234].

Аналитика потока данных ЕРИС ЕМИАС. ЕРИС ЕМИАС – уникальная система, объединяющая все рентгеновские отделения из 162 лечебных учреждений Москвы в единую сеть. К системе подключено около двух тысяч

единиц диагностической аппаратуры. Ежедневно в ЕРИС поступает почти 20 тысяч исследований. Потоки данных в ЕРИС в рамках настоящей работы анализировались по данным дашборда на основании Business intelligence (BI) ЕРИС ЕМИАС.

По результатам аналитики потока исследований в ЕРИС за 2019 год, массовые профилактические исследования составляют: ММГ – 548 781 исследований в год, ФЛГ – 6 579 646 исследований в год. При этом по состоянию на декабрь 2019 года в поликлиниках города Москвы к ЕРИС подключено 245 рентгенологических аппаратов, 155 флюорографических и 66 маммографических аппаратов. По состоянию на 2020 год был определен перечень медицинских организаций, к которым предполагалось подключение ПО на основе ТИИ.

К этой работе предполагалось подключить 291 медицинскую организацию, итоговые данные содержат исследования из 315 медицинских организаций ДЗМ, 856 уникальных диагностических устройств и 1410 врачей-рентгенологов. Ниже приведены количественные показатели по каждому направлению в отдельности, при этом сумма значений для каждого из них превышает общее значение в силу того, что на одном диагностическом устройстве может выполняться два вида исследований, аналогично врачи-рентгенологи могут описывать 2–3 вида исследований.

По направлению «маммография» результаты поступали из 125 медицинских организаций и 138 уникальных диагностических устройств и от 345 врачей-рентгенологов.

По направлению «флюорография» результаты поступали из 167 медицинских организаций и 176 уникальных диагностических устройств и от 703 врачей-рентгенологов.

По направлению «рентгенография» ОГК результаты поступали из 274 медицинских организаций и 542 уникальных диагностических устройств и от 1232 врачей-рентгенологов.

Типичный рабочий процесс рентгеновского отделения. Пациент обращается в отделение лучевой диагностики по направлению на исследование, которое выдал врач-клиницист. Рентгенолаборант в ЕРИС осуществляет поиск пациента и заполняет необходимые сведения о предстоящем исследовании в случае, если такой пациент найден в базе. После заполнения всех необходимых для исследования данных (выбор вида исследования, оборудования, на котором будет осуществляться исследование) задача на проведение исследования автоматически передается на консоль диагностического устройства (worklist). После выбора на консоли ДУ протокола исследования рентгенолаборант проводит исследование. Если связи с ЕРИС нет и рентгенолаборант не может сформировать направление в ЕРИС, он должен создать задание на исследование непосредственно на консоли ДУ. В результате проведения исследования формируется DICOM-файл с изображениями и отправляется в ЕРИС при наличии связи. Если связи с ЕРИС нет, DICOM-файл сохраняется на аппарате и передается в информационный сервис после восстановления связи. Врач-рентгенолог на рабочей панели с изображениями в программе-клиенте ЕРИС просматривает и анализирует изображения исследования и формирует протокол (описание и медицинское заключение) исследования в ЕРИС. Этот протокол хранится вместе с изображениями DICOM – они могут быть доступны при наличии доступа к ЕРИС.

Источники событий: направления на исследования лучевой диагностики формируют врачи в ЕМИАС. Запись на исследование осуществляется посредством записи и самозаписи в ЕМИАС. Рентгенолаборанту в ЕМИАС в плане приема доступна следующая информация о предстоящем исследовании: Ф.И.О. пациента, дата рождения, пол; номер полиса обязательного медицинского страхования (ОМС); название процедуры; предположительный диагноз по Международной классификации болезней 10-го пересмотра (МКБ-10); цель исследования; время записи на исследование в расписании кабинета отделения лучевой диагностики; наименование медицинской организации (МО), из которой

был направлен пациент; сведения о направляющем враче. После идентификации пациента и автоматического создания направления в ЕРИС рентгенолаборанту доступен активный рабочий список заданий на получение изображений (направлений на исследования) со статусами («Назначено», «Выполняется», «Приостановлено», «Выполнено», «Отменено»). Каждое задание содержит следующие сведения:

- сведения о пациенте: Ф.И.О. пациента, дата рождения, пол, номер полиса ОМС;
- сведения о направлении на исследование: клиническая информация (цель исследования), приоритет направления, плательщик по направлению, дата создания направления в ЕРИС;
- сведения о процедуре: наименование, код, вид исследования, диагностические коды (МКБ-10), часть тела, диагностический кабинет, в котором будет проводиться исследование;
- сведения об аллергических реакциях, угрозах безопасности: если исследование проводится с контрастированием, рентгенолаборант в соответствующем разделе задания указывает продукты контрастирования, объем, способ введения и скорость введения.

После проведения исследования врачу-рентгенологу в ЕРИС доступен рабочий список активных заданий на создание протокола исследования со статусами («Назначено», «Выполняется», «Завершено», «Приостановлено»), а также начатые задания.

При обращении к открытому заданию доступна следующая информация: сведения о направлении и пациенте; изображения исследований в рабочей панели с изображениями; статус заключений («Конфликтует», «Предварительный», «Утверждено», «Частично заверено»).

Статус подтверждения исследования: «Подтверждено», «Без подтверждения», «Ошибка проверки»; протокол исследования (заключение), сформированный врачом-рентгенологом: время проведения исследования, вид

исследования, дата исследования, дата создания задания, часть тела, инвентарный номер исследования, инвентарный номер (уровень направления), тип тома (хранение данных), название исполняющего отделения, медицинское учреждение, название процедуры, Ф.И.О. описывающего врача, уникальный идентификатор исследования, ключевые слова к исследованию, сведения о предыдущих исследованиях.

Таким образом, для реализации первого этапа настоящей работы, были проанализированы данные по массовым профилактическим исследованиям в ЕРИС за 2020 – 2022 года. Основные статистические показатели: среднее количество, среднеквадратичное отклонение, t-критерий. Для предварительного анализа частоты встречаемости патологии, были проанализированы журналы выявленной патологии, которые заполняют врачи-рентгенологи в случае выявления социально-значимой патологии.

С целью оценки качества проводимых массовых профилактических исследований были проанализированы данные аудита и второго мнения: эксперты ГБУЗ «НПКЦ ДиТ ДЗМ» проводят консультации (предоставление «второго мнения») и аудит исследований (случайно выбранные исследования с финализированными протоколами обезличиваются и проверяются врачом-рентгенологом). Методикой ретроспективных пересмотров предусмотрены следующие их результаты:

- полное соответствие (технически и клинически корректно проведенное и интерпретированное исследование);
- замечания общего характера (оформление протокола);
- клинически незначимое расхождение (не приводят к изменению тактики лечения, не угрожают жизни и здоровью пациента);
- клинически значимое расхождение (потенциально приводят к нарушениям тактики лечения и/или угрожают жизни и здоровью пациента);
- дефект (связан с работой рентгенолаборанта, укладкой пациента, выполнением исследования).

Описанная методика позволяет стандартизировать оценку врача-рентгенолога, а также направлена на повышение качества работы врача рентгенолога. Однако, эта методика не позволяет оценить показатели диагностической точности «среднего» врача-рентгенолога, что имеет решающее значение при принятии решения о внедрении ТИИ в практику.

2.3. Сопоставления показателей диагностической точности врача-рентгенолога и программного обеспечения на основе технологий искусственного интеллекта

В рамках второго этапа предполагалось провести сопоставление показателей диагностической точности ПО на основе ТИИ и «среднего» врача рентгенолога. Для этого необходимо было определить перечень показателей диагностической точности. Этот перечень был подготовлен на основании данных анализа литературы и на основании методических рекомендаций по проведению клинических испытаний программного обеспечения на основе интеллектуальных технологий [235, 236].

С учетом того, что в основной массе профилактические исследования направлены на бинарную классификацию результатов диагностической процедуры, в настоящей работе будет использоваться бинарная оценка: «норма» – исследование, не содержащее целевой патологии, и «не норма», т.е. исследование содержит признаки целевой патологии. При этом под целевой патологией понимается конкретные рентгенологические признаки. Например, для маммографии наличие признаков злокачественного образования будет являться маркером патологии. Для рентгенологического исследования ОГК исследование будет классифицировано как «не норма» при наличии минимум одного рентгеновского признака патологии. Для проведения оценки качества работы ПО на основе ТИИ используются бинарный ответ от ПО и бинарная классификация протокола врача-рентгенолога. Исходя из стандартной четырехпольной таблицы,

можно классифицировать результаты анализа, выполненного ПО на основе ТИИ, следующим образом:

- 1) истинно положительные результаты (ИП) – когда ПО правильно идентифицирует наличие патологии;
- 2) истинно отрицательные результаты (ИО) – когда ПО корректно определяет отсутствие патологии;
- 3) ложноположительные результаты (ЛП) – когда ПО указывает на наличие патологии, которой на самом деле нет;
- 4) ложноотрицательные результаты (ЛО) – когда ПО не обнаруживает патологию, хотя она присутствует.

При проведении статистической обработки результатов использовались следующие показатели: точность, чувствительность, специфичность. Все показатели рассчитывались с 95% доверительным интервалом.

Статистический анализ проведен с использованием языка программирования R и Python (библиотеки: Statsmodels, Pandas, NumPy, SciPy, Matplotlib, Seaborn, Scikit Learn, Plotly).

Для всех ПО на основе ТИИ ответ подразумевает вероятность патологии. С целью предоставления ответа ПО в понятном для конечного пользователя виде («норма» и «не норма») необходимо определить оптимальное пороговое значение (cut-off), выше которого исследование считается с патологией, ниже – трактуется как норма. Это также необходимо для расчета всех вышеприведенных значений. Для расчета порогового значения использовался метод ROC-анализа. Для ROC-анализа включались исследования, имеющие описание врача-рентгенолога и результат работы сервиса. Исследования без описания или с пропущенными значениями вероятности наличия патологии исключались.

Для построения ROC-кривой необходимо рассчитать чувствительность и специфичность со всеми возможными пороговыми точками. ROC-кривая представляет собой зависимость чувствительности от 1-специфичности с перечнем пороговых точек. Площадь под этой кривой, или AUC, является

ключевым показателем, который будет активно использоваться на всех этапах анализа результатов. ROC-кривая может выглядеть как «лестница» при ограниченном количестве исследований. Поэтому будет использоваться подход, основанный на интерполяции и сглаживании характеристической кривой [237].

Определение оптимальных пороговых точек имеет разные подходы и в основном зависит от поставленных целей и задачи, но есть и стандартные, такие как максимальный индекс Юдена [236,238]. Этот показатель можно визуально интерпретировать как наибольшую вертикальную линию от случайной линии до ROC-кривой [235,236].

Также важно иметь 95% доверительный интервал для всех показателей как лучший способ оценить значимость различий в эффективности теста по сравнению с другими [239]. Для выполнения ROC-анализа существуют различные инструменты [240–242], но ни один из них не обладал необходимым функционалом для проведения настоящего исследования или имел ограничения [243]. «Сравнительный анализ AUC проводился с помощью перестановочного теста [244]. Проверялась нулевая гипотеза H_0 об отсутствии различий между ROC-кривыми против альтернативной H_1 о существовании различий. Уровень значимости $p < 0,05$ » [245, стр. 59].

В основу исследования по оценке диагностической точности врача-рентгенолога легла методика исследования с несколькими читателями (Multi-reader study). Чтобы оценить диагностическую точность врачей-рентгенологов при описании ММГ, применялись упрощенные критерии оценки наличия или отсутствия патологических изменений, не углубляясь в детальную классификацию или локализацию обнаруженных изменений. Врачам-рентгенологам предлагалось выбрать для каждого случая один из пяти вариантов ответа, оценивая вероятность наличия патологии на шкале от «определенно без патологии» до «определенно с патологией». Категориальная оценка переводилась в цифровые значения от 0 до 1 с шагом в 0,25. Количество анализируемых исследований варьировалось: 20, 50 или 80 на одного врача.

В исследовании участвовали волонтеры – врачи-рентгенологи из разных медицинских организаций в России. Результаты врачей, не завершивших анализ заданного числа исследований, в анализ не включались. В итоге в исследование были включены результаты 113 врачей-рентгенологов. Из них 16 врачей имели опыт до 1 года, 32 – от 1 до 5 лет, 28 – от 5 до 10 лет, и 37 врачей работали более 10 лет. Персональные данные врачей – участников исследования не собирались.

Для решения задачи использовался один из ранее подготовленных калибровочных наборов данных. Исследование считалось патологическим, если врач-рентгенолог поставил диагноз по шкале BI-RADS как 3, 4a, 4b, 4c, или BI-RADS 0 для скрининга. Для классификации исследования как патологического необходимо было выполнение двух условий: консенсус двух экспертов с опытом работы в маммографии более трех лет и подтверждение злокачественности новообразований гистологически. При несогласии двух экспертов привлекался третий. Соглашение двух экспертов определяло включение в исследование. Исследования без консенсуса исключались. В итоге был сформирован набор данных из 269 исследований, из которых 167 – с патологией (см. таблицу 2).

Для решения задачи оценки точности врачей при описании РГ/ФЛГ-исследований также была использована вышеописанная методика выставления степени уверенности врача-рентгенолога.

В исследование были включены данные от 185 врачей-рентгенологов, анализировавших рентгеновские снимки, и 69 врачей, работавших с флюорографиями. Среди оценивавших рентгенограммы большинство составляли врачи со стажем от 1 до 5 лет (60 врачей), по 36 специалистов имели опыт от 0 до 1 и от 5 до 10 лет, в то время как 53 врача имели стаж более десяти лет. Среди радиологов, оценивавших флюорограммы, распределение по стажу было следующим: 6 с опытом до года, 8 – от 1 до 5 лет, 5 – от 5 до 10 лет и 9 – более 10 лет. В этом исследовании анализировалось 140 цифровых рентгеновских снимков, из которых 47 содержали патологические изменения, и 184 цифровые флюорографии с 84 случаями патологии. Классификация на категории

«с патологией» и «без патологии» основывалась на консенсусе двух опытных врачей-рентгенологов. Процент и типы обнаруженных патологий соответствовали таковым в клинической практике (см. таблицу 2).

«Коллективная оценка врачей выполнялась отдельно для рентгенографии, отдельно для флюорографии. Все полученные ответы были деперсонализированы. Далее выполнялось разделение ответов по видам исследований (рентгенография и флюорография). Осуществлялись балльная оценка и вычисление средней оценки среди ответов всех врачей по данному исследованию» [245, стр. 59].

Подготовка наборов данных для тестирования программного обеспечения на основе технологий искусственного интеллекта

Для оценки диагностической точности врачей-рентгенологов, а также для оценки ПО на основе ТИИ и последующего внедрения требовались референтные наборы. С этой целью были подготовлены наборы данных: калибровочные (для оценки показателей диагностической точности) и функциональные (для оценки полноты и качества реализованного функционала ИИ-Сервисов). С учетом технических ограничений по обработке ФЛГ отдельными решениями на основе ТИИ эти виды исследований были также разделены при подготовке наборов данных.

Методика подготовки наборов данных

При подготовке наборов данных для решения задач тестирования ПО на основе ТИИ необходимо исходить из базовых диагностических требований (БДТ) к ПО на основе ТИИ. Так, согласно разработанным БДТ, в первую очередь необходима оценка наличия патологии в исследовании, т.е. классификация исследований должна осуществляться путем отнесения к категориям «норма» и «патология», причем в последней категории указываются рентгенологические признаки заболевания. Критическую роль в подготовке итоговой базы данных

играет соблюдение баланса между классами для тестов калибровки, обеспечивая равное количество случаев без патологии и с ней.

В подготовке наборов данных участвовали следующие специалисты:

- 1) аналитик, осуществлявший предварительный отбор исследований, используя поиск по ключевым словам;
- 2) два разметчика (врачи-рентгенологи со стажем работы более 5 лет) для проведения оценки исследования. Каждый разметчик независимо оценивал каждое исследование, указывая, есть в исследовании патология или нет, а если есть – то какая;
- 3) один эксперт, проверявший корректность разметки исследований, а также участвовавший в обсуждении случаев с расхождением оценок от разметчиков;
- 4) технический специалист, осуществлявший сборку, выгрузку, анонимизацию данных и подготовку сопроводительной документации.

Набор данных для флюорографии

ФЛГ-исследования отбирались из ЕРИС ЕМИАС по ключевым словам, проходили анонимизацию и размечались экспертами. Исследование классифицировалось как патологическое в том случае, если был достигнут консенсус экспертов по одному патологическому признаку как минимум. Перечень патологических признаков формировался на основании БДТ в 2020 году. В итоговый набор данных не включались исследования, по которым не был достигнут консенсус, а также содержащие технические дефекты или признаки проведенных операций на целевом органе. Общий набор данных был аналогичный тому, на котором проводилась оценка врачебной точности (таблица 2). Для оценки диагностической точности из этого набора данных была сформирована подвыборка из 100 исследований с балансом классов: 50 – с целевой патологией, 50 – без целевой патологии.

Набор данных для рентгенографии органов грудной клетки

РГ-исследования по ключевым словам отбирались из ЕРИС ЕМИАС, проходили анонимизацию и размечались экспертами. Исследование

классифицировалось как патологическое в том случае, если был достигнут консенсус экспертов по одному патологическому признаку как минимум. Перечень патологических признаков формировался на основании БДТ в 2020 году аналогично описанному выше набору данных для ФЛГ. Также, по аналогии с набором данных для ФЛГ из общего набора для оценки врачебной точности (таблица 2) была сформирована подвыборка из 100 РГ-исследований, дополненные исследованиями с патологией для обеспечения баланса классов: 50 – с целевой патологией, 50 – без целевой патологии.

Набор данных для маммографии

Для оценки диагностической точности из ранее сформированного набора данных для оценки врачебной точности была сформирована выборка из 100 ММГ-исследований: 50 – с целевой патологией, 50 – без целевой патологии.

Таблица 2 – Характеристики набора данных для проведения оценки диагностической точности

Параметр	РГ ОГК	ФЛГ	ММГ
Всего число исследований (с патологией) *	140 (47)	184 (84)	269 (167)
Разметка	Два эксперта с опытом работы более 5 лет		Трепан-биопсия для патологии
Мужчины/женщины/нет данных	59/77/4	94/113/4	0/269/0
Возраст	49±18 [15, 89]	53±19 [19, 93]	63±6 [34, 80]
Рентгенологические признаки	Пневмоторакс; ателектаз; узлы или массы; инфильтрация или консолидация; диссеминация; полость; кальцинат; нарушение целостности кортикального слоя.		BiRADS 0
Количество диагностических устройств	61	69	11
Производители	1. GE Medical Systems, LLC 2. Fujifilm 3. Toshiba Medical Systems, Inc 4. RENinMED, LLC	S.P. Gelpik, LLC	Fujifilm
*Исследования с патологией содержали минимум 1 патологический признак			

2.4. Программное обеспечение на основе технологий искусственного интеллекта (ИИ-сервисы)

Для решения задач в рамках третьего этапа необходимо было оценить текущее состояние и возможности алгоритмов на основе ИИ. В настоящее исследование было включено 16 ИИ-сервисов от 10 компаний-разработчиков. Все показатели диагностической точности ПО на основе ТИИ, используемые в данной работе и заявленные разработчиками, представлены в таблице 3.

Таблица 3 – Показатели диагностической точности ПО на основе ТИИ

ПО на основе ТИИ	Определяемая патология	Площадь под ROC-кривой (AUC)	Чувствительность	Специфичность	Точность	Длительность анализа 1 исследования, мин
Цельс РГ	Различные патологии легких	0,91	0,8	0,92	0,86	0,4
Цельс ФЛГ	Различные патологии легких	0,91	0,8	0,92	0,86	0,4
Цельс ММГ	ЗНО молочной железы	0,886	0,82	0,82	0,82	0,55
«Программа автоматизированного анализа цифровых рентгенограмм органов грудной клетки»	Различные патологии легких	0,965	0,92 (0,87–0,97)	0,94 (0,89–0,99)	0,93 (0,89–0,97)	0,15±0,05
«Программа автоматизированного анализа цифровых флюорограмм»	Различные патологии легких	0,965	0,92 (0,87–0,97)	0,94 (0,89–0,99)	0,93 (0,89–0,97)	0,15±0,05
Третье Мнение, рентгенограммы	Различные патологии легких	0,97	0,87	0,96	0,9	1
Третье Мнение, флюорограммы	Различные патологии легких	0,92	0,9	0,81	0,85	0,05

Продолжение таблицы 3

ПО на основе ТИИ	Определяемая патология	Площадь под ROC-кривой (AUC)	Чувствительность	Специфичность	Точность	Длительность анализа 1 исследования, мин
Care Mentor AI	Различные патологии легких	0,93	0,86	0,92	0,91	0,17
AI RADIOLOGY CXR	Различные патологии легких	0,9	0,92	0,63	0,73	0,5
qXR	Различные патологии легких	0,92	0,9	0,82	0,85	1±0,5
Lunit INSIGHT CXR	Различные патологии легких	0,92	0,79	0,95	–	0,33
Lunit INSIGHT MMG	ЗНО молочной железы	0,9	0,83	0,95	0,96	0,5
Трио ДМ	ЗНО молочной железы	0,88	0,83	0,83	0,83	0,5
MIDA	ЗНО молочной железы	0,88	0,94	0,72	0,91	0,5
ISMI	ЗНО молочной железы	0,93	0,88	0,89	0,89	0,17
ISMI	Различные патологии легких	0,98	0,96	0,93	0,94	0,17

Несмотря на то, что отличия современной цифровой флюорографии и РГ ОГК минимальны, в данном параграфе и далее в результатах эти два вида исследований будут рассматриваться отдельно, т.к. не все разработчики заявляли в технической документации возможность обработки ФЛГ. Далее на основе информации, представленной разработчиками, приведено описание каждого решения в отдельности:

1. «Цельс» (ООО «Медицинские скрининг системы»). Сервис состоит из трех компонентов. Собственно, «Цельс РГ» – предназначен для автоматического анализа рентгеновских снимков органов грудной клетки с целью выявления признаков различных заболеваний; «Цельс ФЛГ» – для автоматического анализа флюорограмм с целью выявления признаков различных заболеваний [246];

«Цельс ЗНО молочной железы» – детектирует и выделяет на изображении злокачественные и доброкачественные новообразования, кальцинаты, лимфоузлы, фиброзно-кистозную мастопатию, определяет плотность ткани молочной железы по ACR [247,248]. Лицензионные права принадлежат ООО «Медицинские скрининг системы». ПО имеет регистрацию как медицинское изделие: РЗН 2021/14449. Показатели диагностической точности представлены в таблице 1.

2. «Программа автоматизированного анализа цифровых рентгенограмм органов грудной клетки» и «Программа автоматизированного анализа цифровых флюорограмм» (ООО «ФтизисБиоМед»). Соответствующие ИИ-сервисы автоматически оценивают вероятность наличия патологий на рентгенограммах органов грудной клетки и флюорограммах; осуществляет поиск и локализацию патологических признаков [249–251]. Лицензионные права принадлежат ООО «ФтизисБиоМед». ПО имеет регистрацию как медицинское изделие: РЗН 2022/17406. Показатели диагностической точности, заявляемые разработчиком, представлены в таблице 3.

3. «Третье Мнение. Рентгенограммы» и «Третье Мнение. Флюорограммы» (ООО «Платформа Третье Мнение»). Соответствующие ИИ-сервисы осуществляют автоматический поиск семиотических признаков на рентгенограммах органов грудной клетки и флюорограммах, выполненных в прямой проекции; для найденных семиотических признаков сервис указывает их локализацию и вероятность; на основании найденных изменений сервис формирует интегральную оценку наличия признаков отклонения от нормы в исследовании для проведения процедуры триажа потока исследований. Авторские права принадлежат ООО "ПТМ". Медицинское изделие РЗН 2021/14506 получено на единый программный модуль для анализа флюорограмм и рентгенограмм грудной клетки человека. Показатели диагностической точности представлены в таблице 3.

4. «Care Mentor AI» для анализа рентгеновских изображений (ООО «КэреМенторЭйАй»). Данный ИИ-сервис предназначен для проведения

рентгеновского скрининга органов грудной клетки в прямой проекции. Функционал ИИ-сервиса: детектирование, классификация и выделение на РГ ОГК патологических областей [195]. Лицензионные права принадлежат компании ООО «КэреМенторЭйАй». Программное обеспечение зарегистрировано как медицинское изделие (РЗН 2020/11137). Показатели диагностической точности представлены в таблице 3.

5. «AI RADIOLOGY CXR» (АНО ВО «Университет Иннополис»). ИИ-сервис распознает признаки патологий легких на рентгенограммах в прямой проекции и флюорограммах, а также осуществляет подготовку протокола исследования. [252]. Лицензионные права принадлежат АНО ВО «Университет Иннополис». Данные о регистрации данного ПО как медицинского изделия на момент подготовки рукописи отсутствовали. Показатели диагностической точности имеются в таблице 3.

6. «qXR» (индийской командой «Qure.ai»). Назначение ИИ-сервиса: обнаружение и классификация патологических изменений на рентгенограмме органов грудной клетки [253–256]. qXR является медицинским изделием, сертифицированным CE-MDD, класса 2а, данных о регистрации в базе данных отечественных медицинских изделий на момент подготовки рукописи не получено. В России представлен компанией ООО «Честная Медицина». Показатели диагностической точности представлены в таблице 3.

7. «Lunit» (корейская компания «Lunit.Inc»). Сервис состоит из двух компонентов: «Lunit INSIGHT CXR» – ИИ-сервис для анализа рентгенограмм органов грудной клетки в прямой проекции с целью выявления патологий легких и «Lunit INSIGHT MMG» – ИИ-сервис для анализа цифровых маммографических исследований для определения признаков ЗНО молочной железы. Показатели диагностической точности представлены в таблице 3. Сведений о регистрации в базе данных отечественных медицинских изделий на момент подготовки рукописи не получено. В России данные ИИ-сервисы представлены компанией АО «Р-Фарм».

8. «Трио ДМ» (АО «МТЛ»). Данный ИИ-сервис предназначен для анализа маммографических исследований: определяет признаки ЗНО, и визуализирует подозрительные области на снимке, а также указывает категорию BI-RADS. Сведений о регистрации в базе данных медицинских изделий на момент подготовки рукописи не получено. Лицензионные права принадлежат компании АО «МТЛ». Показатели диагностической точности представлены в таблице 3.

9. «MIDA» (ООО «АЙРИМ»). Назначение сервиса: обнаружение образований, подозрительных на ЗНО молочных желез и их визуальное выделение. Сведений о регистрации в базе данных медицинских изделий на момент подготовки рукописи не получено. Лицензионные права принадлежат компании ООО «АЙРИМ». Показатели диагностической точности представлены в таблице 3.

10. «ISMI» (разработчик – JLK Inc) состоит из двух модулей: JBD-01K для анализа маммографических исследований с целью выявления признаков ЗНО молочных желез и JLD-02K для выявления признаков патологии легких на рентгенографических исследованиях органов грудной клетки. Лицензионные права в России принадлежат компании ООО «Ай.Эс.Эм.Ай. Инновационные решения для медицины». Показатели диагностической точности указаны в таблице 3.

Далее по тексту все торговые наименования ПО на основе ТИИ будут анонимизированы как «ИИ».

2.5. База исследования

База исследования – медицинские организации, подведомственные Департаменту здравоохранения г. Москвы: ГБУЗ г. Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения Москвы» (ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ»);

медицинские организации, оказывающие медицинскую помощь в амбулаторных условиях и в условиях стационара (в том числе дневного стационара).

Исследование проводилось в 2020–2022 гг. (36 месяцев). Всего было проанализировано 918 115 ММГ исследований (преобладала профилактическая ММГ – 743 693 (81,01%, 174422 (18,99%) - диагностическая). Общее количество исследований с результатами работы по РГ 2 206 123 рентгенограмм и 2 868 544 флюорограмм. С учетом особенности многих ПО на основе ТИИ, не позволяющих одинаково эффективно оценивать показатели диагностической точности для РГ и ФЛГ, выборки анализировались независимо. Также рамках работы были исключены отдельные исследования, в т. ч. не имевшие заключения. Возраст исследуемых пациентов составил от 18 до 102 лет (средний возраст для рентгенограмм – $53,93 \pm 18,28$ лет, для флюорограмм – $49,82 \pm 17,66$ лет).

Для рентгенологических исследований органов грудной клетки распределение по полу составило: 41,25% мужчин, 58,16% женщин. В исследовании участвовали 264 медицинских учреждения и 1301 врач-рентгенолог. В рамках настоящей работы применялось 7 ПО на основе ТИИ, перечисленных выше. Среднее время работы ПО на основе ТИИ равнялось 55,24 секунды.

Для флюорографических исследований распределение по полу составило: 34,95% мужчин и 64,89% женщин. Исследования проводились в 165 медицинских учреждениях, описания исследований были выполнены 920 врачами-рентгенологами. Флюорографические исследования выполнялись в абсолютном большинстве (93,2%) по приказу Минздрава России № 404н, по приказу № 29н – 4,9%, прочее (платно, по добровольному медицинскому страхованию) – 1,9%.

Отдельно были сформированы репрезентативные выборки из 10 тыс. профилактических ММГ исследований и 25 тыс. флюорографических исследований, выполненных в 1 квартале 2023 года для контроля моделей применения ПО на основе ТИИ для профилактических и диагностических исследований.

2.6. Методическое обеспечение комплексной системы контроля качества программного обеспечения на основе ТИИ

В рамках четвертого этапа разрабатывается комплекс инструментов и методов контроля качества работы ПО на основе ТИИ, а также алгоритмов принятия управленческих решений. Для этого требуется прежде всего методологически обоснованный объем выборки, инструментарий (в рамках настоящей работы был разработан комплекс инструментов, созданных на языке Python) и прогностические модели.

Обоснование объема выборки

При проведении любых исследований необходимо провести обоснование объема выборки. Объем выборки в рамках настоящего исследования должен быть обоснован для двух задач контроля: 1) контроль качества работы ПО на основе ТИИ и 2) контроль показателей диагностической точности ПО на основе ТИИ.

В рамках настоящей задачи решался вопрос о минимально необходимом и достаточном количестве исследований для проведения контроля качества ПО на основе ТИИ. Для этого были проанализированы два подхода, в основе которых лежала проверка статистических гипотез.

«Подход, основанный на проверке статистических гипотез (вариант 1). Этот подход предполагает проверку основной статистической гипотезы H_0 , согласно которой качество продукции отвечает требованиям, и альтернативной гипотезы H_1 , указывающей на несоответствие качества продукции. Если количество дефектных единиц в выборке (m) не превышает заранее установленный предел (c), партия считается приемлемой. В противном случае она отвергается. Для определения объема выборки применяется формула (1):

$$p(q) = p_n(m \leq c) = \sum_{m=0}^c p_n(m), \quad (1)$$

где m – число дефектных единиц продукции в выборке n ; $p_n(m)$ – вероятность появления дефектных единиц продукции m в выборке n ; c – приемочное число.

В связи с тем, что в рамках настоящей работы объем всей партии превышал объем выборки более чем на 10% [257], оперативные характеристики определяли по формуле (2):

$$P(X = m) = C_n^m q^m (1 - q)^{n-m} \quad (2)$$

где C_n^m – количество сочетаний появления дефектных единиц продукции m в выборке n (3):

$$C_n^m = \frac{n!}{m! (n - m)!} \quad (3)$$

В рамках настоящей работы было использовано приемочное число, равное двум единицам продукции, произведены расчеты и построены кривые для выборок в 30, 50, 80, 138 единиц продукции (см. рисунок 4)» [258, стр. 22].

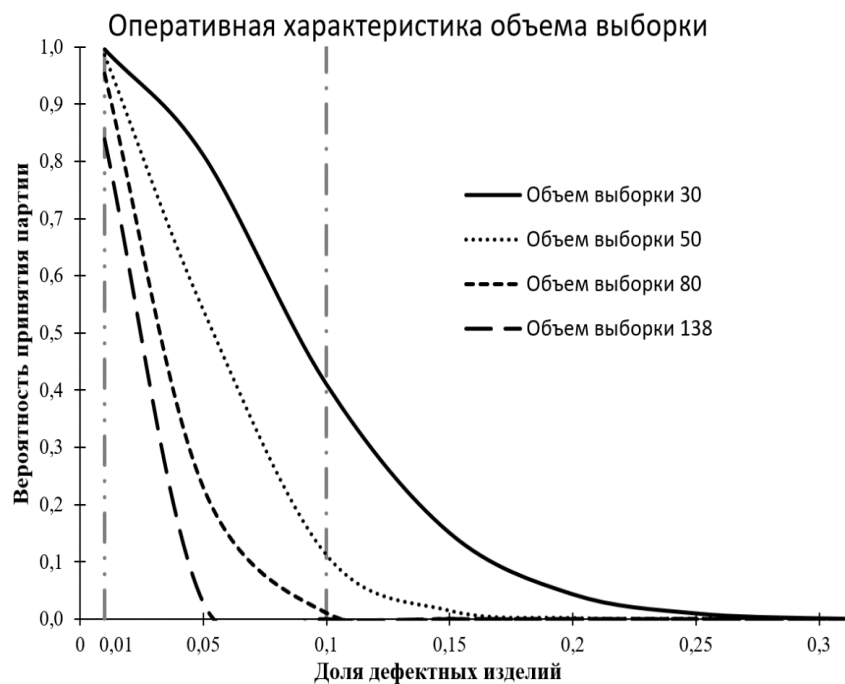


Рисунок 4 – Оперативная характеристика для различных объемов выборки: вертикальная штрихпунктирная линия с двумя точками – риск поставщика; вертикальная штрихпунктирная линия с одной точкой – риск потребителя

На рисунке 4 представлены риски потребителей и поставщиков. Риск для поставщика заключается в вероятности неправильного отклонения качественной партии (когда доля брака в общем количестве продукции меньше 10%). С учетом процента бракованных товаров, указанных поставщиком, этот риск оценивается как 1%. Риск для потребителя означает вероятность принятия некачественной партии, и при учете обнаруженного потребителем брака этот риск оценивается в 10%. Исходя из анализа и установленных рисков в 10% для потребителя и 5% для поставщика (таблица 4), было определено, что выборка в 80 единиц продукции соответствует требованиям обеих сторон.

Таблица 4 – Зависимость рисков от объема выборки

Риск	Объем выборки, шт.			
	30	50	80	138
Потребитель	41%	11%	1%	0%
Поставщик	0%	1%	5%	16%

«Подход, основанный на проверке статистических гипотез (вариант 2). Этот подход основан на вероятностной проверке нулевой гипотезы H_0 , учитывая риски как для поставщика, так и для потребителя. Нулевая гипотеза предполагает, что партия считается дефектной, если содержит более 10% бракованных продуктов. Альтернативная гипотеза H_1 предполагает менее 10% дефектных изделий. Шанс отклонения нулевой гипотезы составляет как минимум 80%. Анализ проведен для выборок объемом 30, 50, 80, 120 единиц с допустимым количеством дефектов до четырех (таблица 5), при этом риски для потребителей и поставщиков не превышали 10 и 5% соответственно. С учетом доли дефектных единиц продукции при приемочных числах больше нуля наиболее подходящие объемы выборок равнялись 80 или 120 единицам» [258, стр. 23].

Таблица 5 – Зависимость рисков от объема выборки и приемочного числа

Размер выборки	Приемочное число/доля дефектов, %	Приемочное число/доля дефектов, %	Риск поставщика	Риск потребителя
30	0 (0,0%)	100,0%	0,0%	8%
30	1 (3,3%)	36,2%	63,8%	37%
50	0 (0,0%)	100,0%	0,0%	1%
50	1 (2,0%)	36,4%	63,6%	7%
80	0 (0,0%)	100,0%	0,0%	<1%
80	1 (1,3%)	92,1%	7,9%	<1%
80	2 (2,5%)	67,7%	32,3%	2%
120	0 (0,0%)	100,0%	0,0%	<1%
120	1 (0,8%)	100,0%	0,0%	<1%
120	2 (1,7%)	98,4%	1,6%	<1%
120	3 (2,5%)	91,9%	8,1%	<1%
120	4 (3,3%)	78,8%	21,2%	2%

Таким образом, два варианта, основанные на проверке статистических гипотез, позволили получить минимальный объем выборки в количестве 80 исследований. Объем выборки в 80 исследований позволяет осуществлять выборочный контроль качества работы ПО на основе ТИИ, однако для задач оценки диагностической точности требуется немного другой подход.

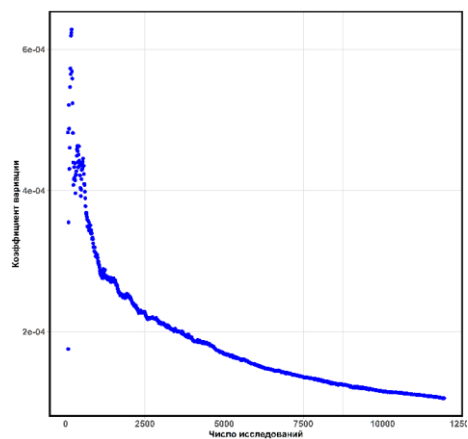
Применим ли описанный выше подход к оценки диагностической точности ИИ-алгоритмов? Для проведения калибровочного тестирования мы используем объем выборки, составляющий 100 исследований с соблюдением баланса классов. Это количество было обусловлено путем анализа эмпирических значений коэффициента вариации.

На рисунке 5 представлены результаты вычисления зависимости коэффициента вариации распределения значений AUC ROC с учетом

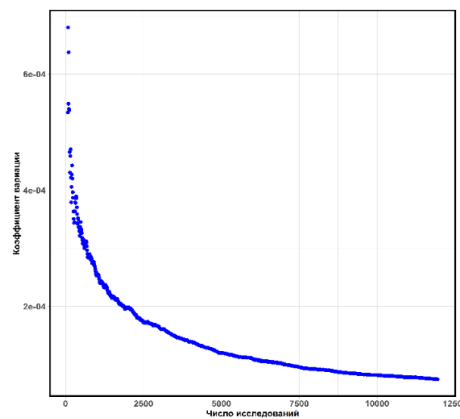
числа исследований для пяти долей «не нормы» в балансе классов «норма» /«не норма»: *a* – для доли «не нормы» 10%; *б* – для доли «не нормы» 20%; *в* – для доли «не нормы» 30%; *г* – для доли «не нормы» 40%; *д* – для доли «не норма» 50%; *е* – обобщенное представление для всех долей «не нормы» баланса классов «норма»/«не норма».

Максимальное значение коэффициента вариации значений AUC ROC для доли «не норма» 10% достигается при числе исследований равном 190, для доли 20% – равно 80 исследований, для доли 30% – 120 исследований, для доли 40% – 110 исследований и для доли 50% – 70 исследований.

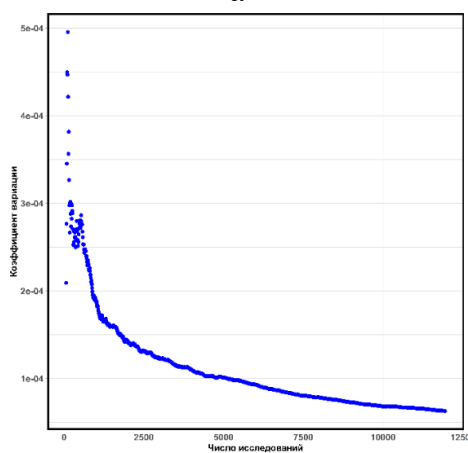
Таким образом, максимальное отклонение значений AUC ROC от общего среднего наблюдается у баланса классов «норма»/«не норма» с долей «не нормы» 10% и при количестве исследований, равном 190. В нашем случае объем исследований на потоке составлял десятки и сотни тысяч, что безусловно, достаточно для проведения достоверной оценки показателей диагностической точности.



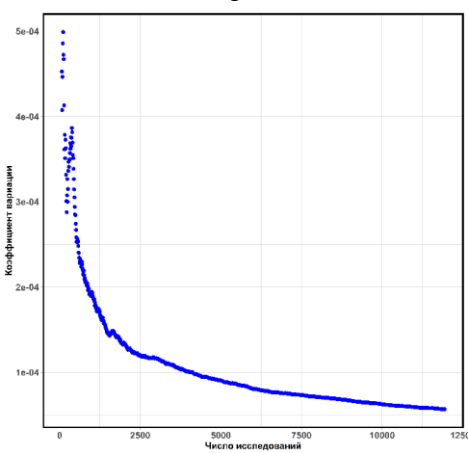
а



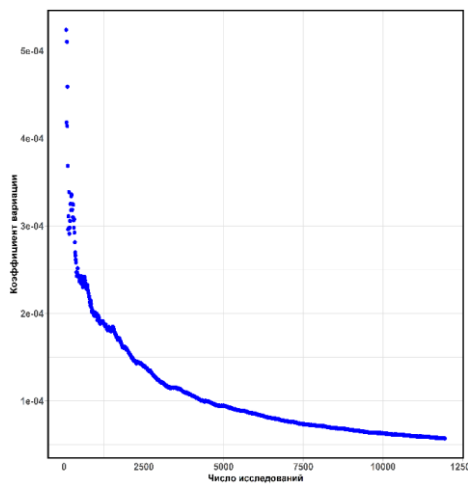
б



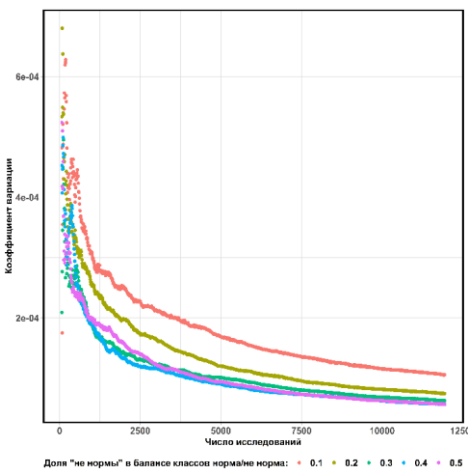
в



г



д



е

Рисунок 5 – Коэффициент вариации значений AUC ROC в зависимости от числа исследований

Оценка результативности работы программного обеспечения на основе технологий искусственного интеллекта

В настоящем подпараграфе будут описаны материалы (выборки исследований), на которых была произведена оценка результативности ПО на основе ТИИ. Для этого были сформированы две выборки: для оценки качества работы алгоритмов ИИ в составе новой медицинской услуги с кодом 1601 «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта» и для оценки качества работы ПО при анализе РГ/ФЛГ исследований.

Маммография, в рамках услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта»

«Выборка содержала 10 417 цифровых маммографических исследований, которые были выполнены пациенткам в медицинских организациях Департамента здравоохранения города Москвы в рамках территориальной программы государственных гарантий бесплатного оказания гражданам медицинской помощи в городе Москве в течение первых двух недель февраля 2023 года» [259, стр. 55,56].

Процедура проведения исследования

Выполнялась цифровая маммография обеих молочных желез. Описанием каждого исследования занимался врач-рентгенолог, описывающий маммографию на потоке не менее 5 лет. Исследования классифицировались по системе BI-RADS (Breast Imaging Reporting and Data System), причем выставлялась максимальная категория BI-RADS из двух молочных желез. Используемое в работе ПО зарегистрировано как медицинское изделие с показателями AUC,

чувствительности, специфичности и точности 0,91, 0,8, 0,92 и 0,86 соответственно.

Статистический анализ результатов

Из первоначальной выборки в 10 417 исследований были исключены те, по которым отсутствовали результаты ПО или была присвоена категория BI-RADS 0 врачом. В результате дальнейшему анализу подверглись 9684 исследования, среди которых для детального сравнения было отобрано 45 случаев с противоположными категориями BI-RADS 1 и 4–5, выставленными врачом и ИИ. Врач-эксперт провел дополнительный анализ этих случаев: разногласия между мнениями врача и ИИ были замечены в 37 случаях, тогда как в 8 случаях оба метода сходились в категории 0 (норма) по бинарной шкале 1 (рисунок 6).

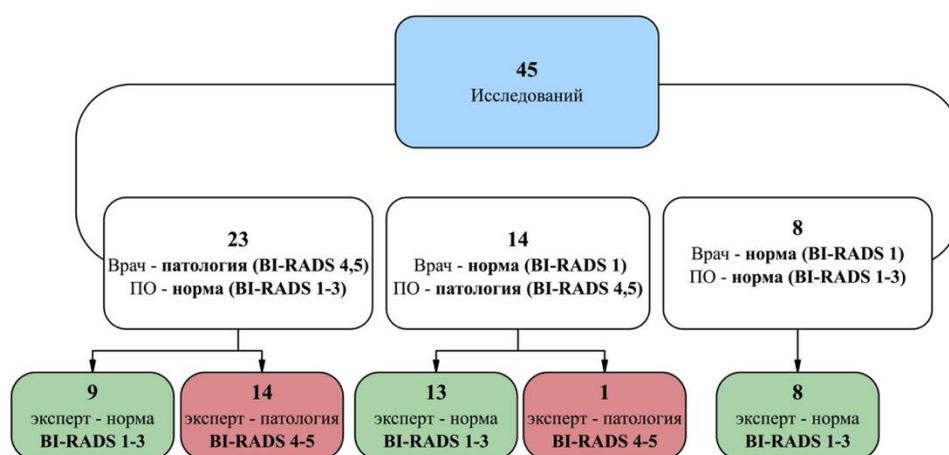


Рисунок 6 – Визуализация формирования выборки, а также распределения результатов оценки маммографии врачом, ИИ-сервисом и экспертом

«Сравнение результатов ПО на основе ТИИ проводили:

- относительно врача (для 9684 исследований);
- врач относительно эксперта (45 исследований);
- ПО на основе ТИИ относительно эксперта (45 исследований).

В первом случае проводили анализ всех данных (9684 исследований), во втором и третьем – анализировали только те исследования, которые были

направлены на пересмотр (45 исследований для бинарной шкалы BI-RADS 1, 29 исследований для бинарной шкалы BI-RADS 2)» [259, стр. 55, 56] (рисунок 7).

Оценку результатов проводили для следующих шкал: диагностическая шкала BI-RADS 1–6, бинарная шкала 1, полученная из диагностической (BI-RADS 1–3 = 0, отсутствие признаков ЗНО; BI-RADS 4–6 = 1, наличие признаков ЗНО) и бинарная шкала 2 (BI-RADS 1, 2 = 0, отсутствие признаков ЗНО; BI-RADS 4–6 = 1, наличие признаков ЗНО).

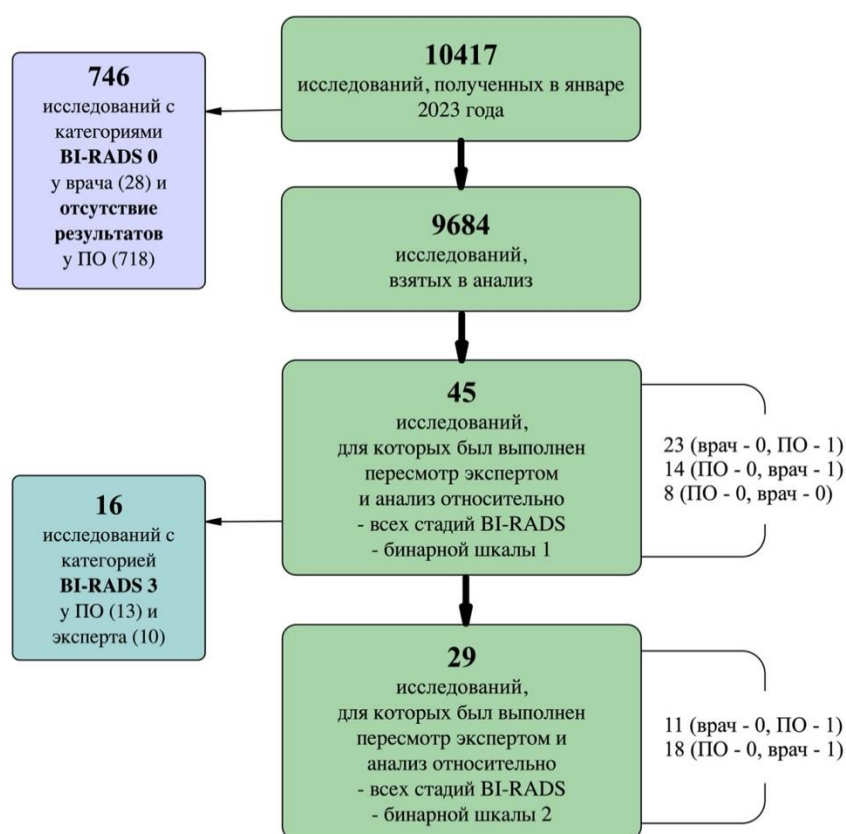


Рисунок 7 – Блок-схема количественного описания исследований, включенных в анализ

«Результаты оценивали относительно шкалы BI-RADS, за исключением 6-й категории, которая оценивалась как 5-я, так как ПО не имело данных о гистологическом подтверждении. Также оценку проводили относительно

бинарных шкал (рисунок 8). При исключении категории BI-RADS 3 из бинарной шкалы 2 были удалены все строки с результатом BI-RADS 3 у врача, ПО на основе ТИИ и эксперта» [259, стр. 57].

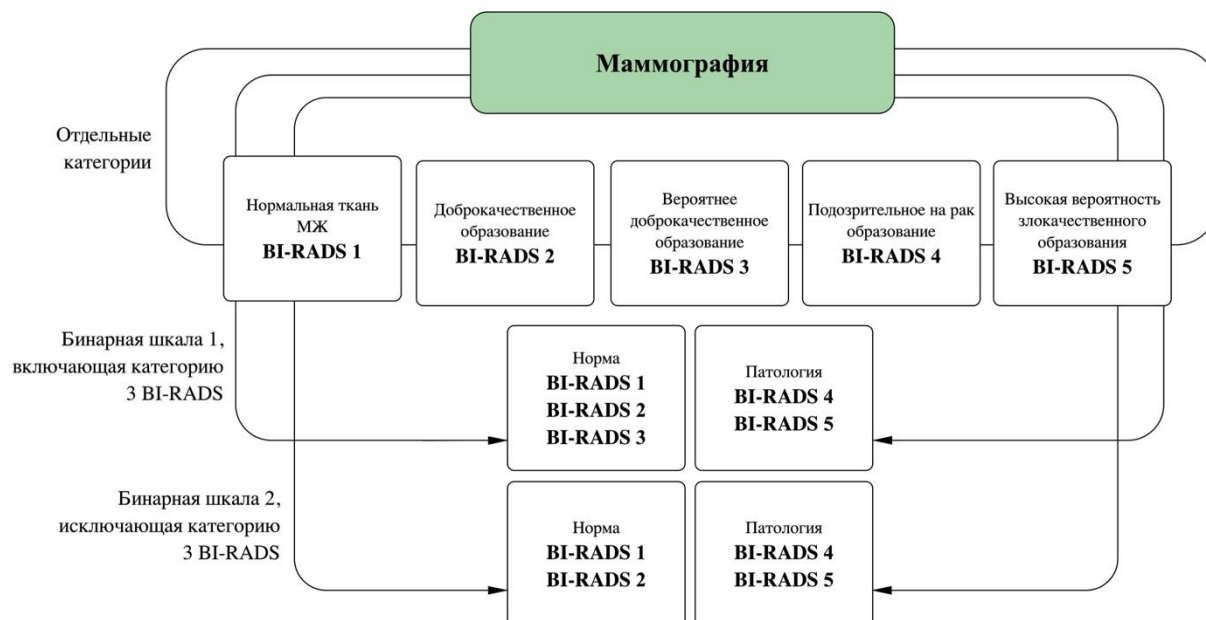


Рисунок 8 – Значения категорий шкалы BI-RADS, а также бинарные шкалы, на основании которых проводился анализ цифровой маммографии

«Согласованность результатов, выраженная в количественной оценке вероятности совпадений и расхождений, а также коэффициент межэкспертного согласия (коэффициент Каппа Козна) рассчитывали в парах ПО/врач, врач/эксперт и ПО/эксперт относительно отдельных категорий BI-RADS и бинарных шкал 1 и 2.

Оценку точности ПО на основе ТИИ проводили относительно врача (9684 исследований) и эксперта (45 исследований для шкалы BI-RADS 1 и 29 исследований для шкалы BI-RADS 2) путем расчета следующих показателей: AUC ROC (micro), точности, чувствительности и специфичности» [259, стр. 57, 58].

Флюорографические исследования в модели автономного описания

Анализируемые данные

Все исследования выполнены в медицинских организациях, подведомственных ДЗМ, в период с 1 января по 31 марта 2023 года, общее количество исследований – 88 048 цифровых флюорограмм. Классификация исследований на категории «норма» и «не норма» проводилась на основе заключений врачей-рентгенологов. В качестве целевых патологий учитывались: плевральный выпот, пневмоторакс, ателектаз, очаги затемнения, инфильтрации или консолидации, диссеминация, полости с распадом, кальцинаты, нарушения целостности кортикального слоя.

Методика построения прогностической модели возникновения технологических дефектов

Для построения модели были вычислены средневзвешенные значения дефектов для каждого сервиса по месяцам (4):

$$\widehat{y}_s = \frac{\sum_{i=1}^n p_i^s y_i^s}{\sum_{i=1}^n p_i^s} \quad (4)$$

где y_i^s – количество дефектов в i -м месяце на сервисе; p_i^s – вероятность возникновения дефекта в i -м месяце на сервисе ПО на основе ТИИ.

Средневзвешенное значение количества дефектов по сервисам рассчитывается по формуле (5):

$$\widehat{D}_s = \frac{\sum_{i=1}^n \omega_i D_i^s}{\sum_{i=1}^n \omega_i} \quad (5)$$

где ω_i – вероятность возникновения дефекта на i -м сервисе; D_i^s – общее количество дефектов на i -м сервисе.

Средневзвешенное значение количества дефектов по видам исследований (6):

$$\widehat{M} = \frac{\sum_{i=1}^n m_i M_i}{\sum_{i=1}^n m_i} \quad (6)$$

где m_i – вероятность возникновения дефектов на i -ом виде исследований; M_i – число дефектов на i -ом виде исследования.

Количество дефектов в прогнозируемом месяце вычислялось по уравнению (7):

$$\widehat{y}_{i+1} = (1 - \alpha) * \widehat{y}_i^s + (1 - \delta) * \widehat{D}_i^s + (1 - \mu) * \widehat{M}_i, \quad (7)$$

где α – вероятность возникновения дефекта в i -месяце на сервисе; δ – вероятность возникновения дефекта на i -сервисе; μ – вероятность возникновения дефекта на i -ом виде исследования.

Вероятность возникновения дефекта для каждого члена уравнения вычислялась исходя из анализа типа распределения дефектов по месяцам, сервисам или видам исследований. На основании критерия Акаике и Бейса из десяти теоретических типов распределений выбирался наиболее близкий к фактическому тип.

В рамках представленной работы для всех вероятностей возникновения дефектов использовалось допущение сохранности типа распределения на последующий период эксплуатации сервиса. Конечное значение вероятности возникновения дефекта вычислялось как (8):

$$\begin{aligned}
\alpha &= \max (P(p_i)) \\
\delta &= \max (P(\omega_i)) \\
\mu &= \max (P(m_i)), \\
\alpha &= \max (P(p_i)) \delta = \max (P(\omega_i)) \mu = \max (P(m_i)),
\end{aligned} \tag{8}$$

где $P(p_i)$ – максимальная вероятность возникновения дефектов в i -м месяце из 1000 случайно сгенерированных значений, соответствующих наиболее близкому теоретическому распределению; $P(\omega_i)$ – вероятность возникновения дефекта на i -м сервисе из 1000 случайно сгенерированных значений, соответствующих наиболее близкому теоретическому распределению; $P(m_i)$ – вероятность возникновения дефекта на i -ом виде исследования из 1000 случайно сгенерированных значений, соответствующих наиболее близкому теоретическому распределению.

Вновь полученные значения количества дефектов в $(i+1)$ -м месяце использовались для подсчета новых значений средневзвешенных величин и вероятностей возникновения дефектов.

Оценка точности разработанной модели осуществлялась с помощью среднеквадратичной ошибки («невязки»), выраженной в процентах (9):

$$\Delta = \frac{1}{N} * \sqrt{\sum_{i=1}^N (P(y^F)_i^F - P(\hat{y}^T)_i^T)^2} * 100 \%, \tag{9}$$

где N – количество месяцев, на которые строится прогноз; $P(y^F)_i^F$ – фактическая вероятность возникновения дефектов в i -м месяце на i -м сервисе; $P(\hat{y}^T)_i^T$ – теоретическая вероятность возникновения дефектов в i -м месяце на i -м сервисе.

Глава 3. ОБОСНОВАНИЕ ЦЕЛЕСООБРАЗНОСТИ И ПРИМЕНИМОСТИ АВТОМАТИЗАЦИИ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ МАССОВЫХ ПРОФИЛАКТИЧЕСКИХ ИССЛЕДОВАНИЙ, ОСУЩЕСТВЛЯЕМЫХ С ПРИМЕНЕНИЕМ МЕТОДОВ ЛУЧЕВОЙ ДИАГНОСТИКИ

3.1. Проблемы управления массовыми профилактическими исследованиями, осуществляемыми с применением методов лучевой диагностики

Для решения обозначенной задачи необходимо изучить действующую систему организации массовых профилактических исследований с применением методов лучевой диагностики. Далее будут выявлены основные проблемы управления и предложены способы их решения с применением ТИИ.

Анализ организации и проведения массовых профилактических исследований, осуществляемых с применением методов лучевой диагностики (на примере г. Москвы)

Массовые профилактические осмотры населения осуществляются в соответствии с приказом Минздрава России от 27.04.2021 № 404н «Об утверждении Порядка проведения профилактического медицинского осмотра и диспансеризации определенных групп взрослого населения» и иными нормативно-правовыми актами.

Согласно действующему законодательству методы лучевой диагностики применяются и на первом, и на втором этапах диспансеризации. На первом этапе мужскому и женскому населению (от 18 до 64 лет включительно и старше 65 лет) в рамках профилактического медицинского осмотра с целью скрининга туберкулеза, злокачественных новообразований и иных социально значимых заболеваний проводится флюорография или рентгенография легких с частотой 1 раз в 2 года. С целью скрининга злокачественных новообразований молочных

желез женщинам в возрасте от 40 до 75 лет включительно выполняется рентгеновская маммография обеих молочных желез в двух проекциях с двойным прочтением рентгенограмм 1 раз в 2 года. Необходимо обратить внимание на наличие требований в указанном выше приказе о проведении именно двойных прочтений результатов профилактической маммографии. На втором этапе диспансеризации предусмотрено выполнение рентгенографии или компьютерной томографии легких с целью дополнительного обследования и уточнения диагноза заболевания. Исследование осуществляется при наличии медицинских показаний в соответствии с клиническими рекомендациями по назначению врача клинической специальности.

Очевидно, что массовый характер носит именно выполнение флюорографии (реже рентгенографии в качестве альтернативы) и маммографии на первом этапе диспансеризации, причем (важно с организационной точки зрения) существует нормативное требование по реализации двойного прочтения результатов маммографии. Такой подход обеспечивает максимальную диагностическую ценность скрининга, что убедительно доказано в многочисленных отечественных и зарубежных публикациях; однако создает значительную нагрузку на систему здравоохранения.

С целью объективизации состояния и результативности массовых профилактических осмотров с применением методов лучевой диагностики проведен анализ соответствующей деятельности в медицинских организациях государственной системы здравоохранения Москвы в период 2019–2022 гг. Исследованы процессы, связанные с выполнением и интерпретацией результатов лучевых исследований первого этапа диспансеризации – флюорографии (ФЛГ) и маммографии (ММГ).

Первый этап диспансеризации в г. Москве осуществляется в медицинских организациях, оказывающих первичную медико-санитарную помощь в амбулаторных условиях (далее – городских поликлиниках). Особенностью системы здравоохранения г. Москвы является то, что часть медицинских

организаций второго и третьего уровней медицинской помощи (оказывающих специализированную помощь в условиях стационара; далее – стационары) имеют в своей структуре отдельные подразделения, также оказывающие первичную медицинскую помощь в амбулаторных условиях. В таких подразделениях тоже проводятся медицинские профилактические осмотры в рамках первого и второго этапов диспансеризации. Соответственно, тоже выполняются профилактические лучевые исследования – флюорография и маммография.

Предусмотренные нормативно-правовой базой лучевые исследования назначаются врачами и выполняются в структурных подразделениях лучевой диагностики медицинских организаций. Результаты проведенных исследований сохраняются в централизованном архиве медицинских изображений в государственной информационной системе в сфере здравоохранения субъекта РФ – Едином радиологическом информационном сервисе Единой медицинской информационно-аналитической системы г. Москвы (ЕРИС ЕМИАС).

В период 2019–2020 гг. интерпретация и описания результатов профилактических лучевых исследований выполнялись врачами-рентгенологами медицинских организаций. В 2020 г. на базе ГБУЗ «НПКЦ ДиТ ДЗМ» создан Московский референс-центр лучевой диагностики (в соответствии с приказом Минздрава России от 09.06.2020 № 560н «Об утверждении Правил проведения рентгенологических исследований» и приказом Департамента здравоохранения города Москвы от 01.04.2020 г. № 323 «О создании дистанционного референс-центра по лучевой диагностике»). Реализация концепции централизации лучевой диагностики обеспечила недопущение кадрового дефицита, оптимальную загрузку оборудования и высокую степень доступности лучевых исследований для населения, однако детальное изучение эффективности референс-центра выходит за рамки нашего исследования.

В период 2020–2021 гг. интерпретация и описания результатов флюорографий и маммографий, выполненных в городских поликлиниках, осуществлялась врачами-рентгенологами Московского референс-центра лучевой

диагностики. Описания профилактических маммографий, выполненных в стационарах – врачами-рентгенологами соответствующих медицинских организаций.

Парк диагностического оборудования. Городские поликлиники и стационары оснащены оборудованием для профилактических лучевых исследований согласно приказам Департамента здравоохранения города Москвы от 15.09.2020 № 1044 «Об утверждении модели расчета потребности в оснащении рентгенографическими аппаратами медицинских организаций государственной системы здравоохранения города Москвы» и от 10.02.2021 № 99 «Об утверждении моделей расчета потребности в оснащении медицинских организаций государственной системы здравоохранения города Москвы оборудованием лучевой диагностики». Сводные данные представлены в таблице 6. Обращает на себя внимание снижение количества диагностических устройств, что обусловлено полным списанием аналоговых аппаратов, плановыми мероприятиями по обновлению парка, программой модернизации и капитальных ремонтов медицинских организаций.

С точки зрения организации здравоохранения, важным вопросом является обеспечение требуемого объема и качества профилактических лучевых исследований на основе имеющегося парка оборудования. Для решения такого вопроса требуются специальные организационные подходы, управленческие решения, в том числе на основе автоматизации и цифровой трансформации. Подробнее эти аспекты изучены в дальнейших главах.

Таблица 6 – Оснащенность медицинских организаций г. Москвы оборудованием для профилактических лучевых исследований

Тип аппарата	2019	2020	2021	2022
Маммограф	183	173	161	156
Флюорограф	294	285	253	161
Итого	477	458	414	317

На рисунке 9 представлено соотношение распределения диагностических устройств между городскими поликлиниками и стационарами. Очевидно преобладание количества диагностических устройств, применяемых для профилактических исследований, именно в городских поликлиниках.

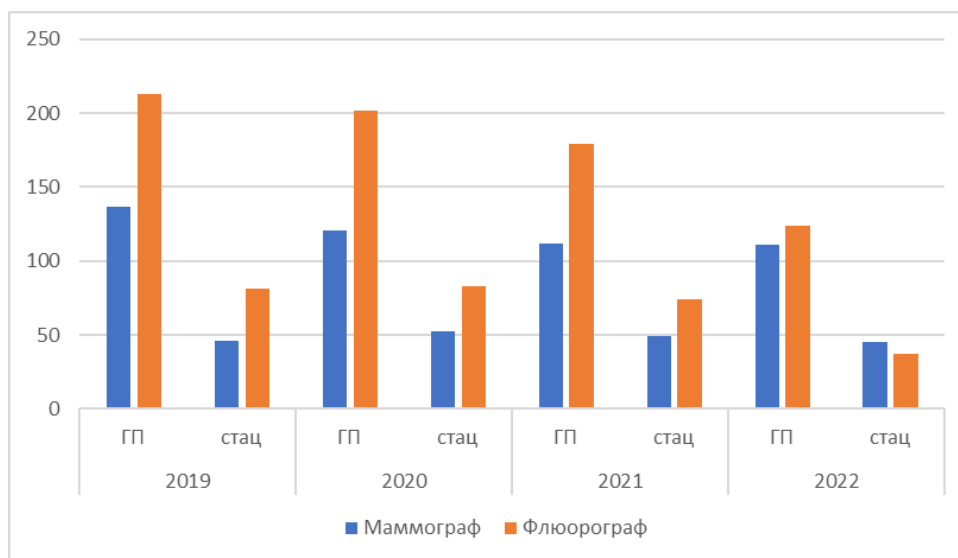


Рисунок 9 – Распределение оборудования для профилактических лучевых исследований в медицинских организациях г. Москвы

Профилактические лучевые исследования. Сводные данные о профилактических лучевых исследованиях в г. Москве за период 2019–2022 гг. представлены на рисунке 10.

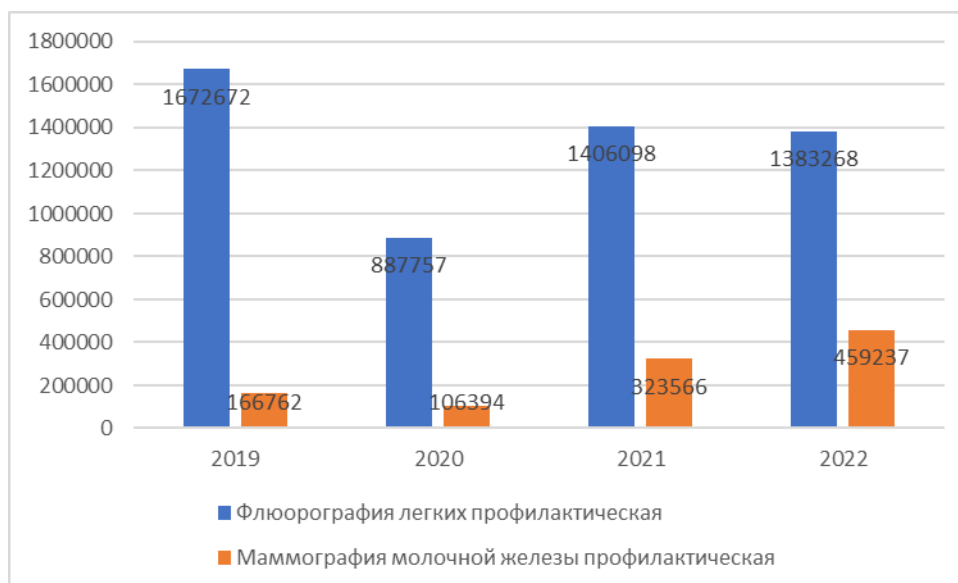


Рисунок 10 – Количество и динамика профилактических лучевых исследований в г. Москве (2019–2022 гг.)

Количество флюорографий в 2019 г. составило 1 672 672, в следующем году оно снизилось на 47,0%, до 887 757, что обусловлено временной остановкой профилактических исследований и общими ограничениями в период пика пандемии COVID-19. В 2021 и 2022 гг. отмечается постепенное возобновление объемов диспансеризации населения (таблица 7): количество флюорографий составило 1 406 098 и 1 383 268 соответственно. При сравнении среднего количества ФЛГ в месяц наблюдается статистически значимая разница между 2020 г. и 2021, 2022 гг. Вместе с тем различия между 2021 и 2022 гг. не имеют значимости (t -критерий = 0,183, $p=0,8567$), что свидетельствует о достижении адекватного уровня профилактических исследований органов грудной клетки и общем восстановлении профилактического направления деятельности отделений лучевой диагностики городских поликлиник после пандемии.

Таблица 7 – Динамика среднего ежемесячного количества флюорографий органов грудной клетки

Флюорография легких	2020	2021	2022
Среднее количество в месяц	73 979,75	117 174,8	115 272,3
Среднеквадратичное отклонение	37 916,21	26 340,88	24 645,47
t -критерий	-	3,241	-3,163
p	-	0,0038	0,0045

Обращает на себя внимание тот факт, что после пандемии количество флюорографий возросло, но осталось меньше уровня 2019 г. примерно на 20,0%. Это обусловлено переоснащением парка оборудования лучевой диагностики: заменой флюорографов на полноценные рентген-аппараты (так называемые U-дуги). Соответственно, с целью скрининга примерно 20,0–25,0% обследуемых выполняется цифровая рентгенография органов грудной клетки (отметим, что флюорография и рентгенография эквивалентны согласно действующему законодательству в сфере диспансеризации). Вместе с тем текущая номенклатура инструментальных исследований в ЕРИС ЕМИАС создает определенные

ограничения для учета профилактических рентгенографий. Тем не менее в рамках данной диссертации основное внимание будет сосредоточено именно на флюорографии как основной вид профилактических исследований.

Схожая тенденция в динамике общего количества исследований в период 2019–2022 гг. наблюдается для профилактической маммографии. До пандемии, в 2019 г., выполнено 166 762 таких исследований. В следующем, 2020 г. отмечено снижение на 36,0%. Зато в 2021 и 2022 гг. наблюдается стремительное увеличение числа таких исследований до уровня, значительно превышающего «допандемийный» в 2–2,5 раза (323 566 и 459 237). Констатируем, что имеет место увеличение количества профилактических маммографий в 2022 г., по сравнению с 2019 г., на 175%. Такое явление обусловлено повторным запуском программы «Московский скрининг рака молочной железы» и разворачиванием системы «Павильонов здоровья», то есть масштабными городскими проектами по повышению вовлеченности и охвата населения профилактическими мероприятиями.

Различные подходы к сбору данных (помесячный мониторинг в информационной системе введен для маммографии с 2021 г.) не позволяют провести статистический анализ различий по аналогии с флюорографией. В данном случае ограничиваемся лишь описательным изучением.

Качество лучевых профилактических исследований. В медицинских организациях г. Москвы проводятся системные мероприятия по обеспечению безопасности и качества лучевой диагностики, основанные на регулярном ретроспективном пересмотре результатов проведенных исследований. Пересмотр осуществляется случайным или целевым образом, в анонимизированном виде специально сформированной группой экспертов, размещенной на базе ГБУЗ «НПКЦ ДиТ ДЗМ» и подчиненной главному внештатному специалисту по лучевой и инструментальной диагностике Департамента здравоохранения Москвы. Методика пересмотра и соответствующая деятельность регулируются специальными распоряжениями и методическими рекомендациями. В контексте

нашего исследования представляет особый интерес качество профилактических лучевых исследований, которое и представляется возможным оценить путем анализа результатов ретроспективных экспертных пересмотров.

Для анализа использованы данные за 2020 и 2021 гг. Кроме того, проведено сравнение количества и структуры дефектов при описании профилактических маммографий врачами-рентгенологами Московского референс-центра лучевой диагностики (МРЦЛД) и стационаров.

В 2020 г. проведены ретроспективные пересмотры (рисунок 11) 1014 профилактических ММГ, в 2021 г. – 4314, в 2022 г. – 4496. Полное соответствие установлено в первый год в 47,2% случаев, в два последующих – фактически в 61,0% исследований. Среди несоответствий в каждом периоде лидируют замечания общего характера, клинически незначимые расхождения. Дефекты, связанные с работой рентгенолаборанта, отмечены в 1,0–2,0% случаев. Следует подчеркнуть, что в 2020 г. клинически значимые расхождения выявлены в 4,8% случаев, в 2021 г. – в 1,7%, а в 2022 г. – только в 0,8%.

Сравнительно изучено качество профилактических ММГ, описание которых выполнено в разных учреждениях. В 2022 г. выполнены ретроспективные пересмотры 4496 профилактических маммографий, из них 46,9% (677), описанных в МРЦЛД, и 53,1% (68) – в стационарах. Клинически незначимые расхождения встречались чаще при описаниях результатов ММГ в стационарах – 22,7% (489), чем в МРЦЛД – 16,4% (384). Также в описаниях, сделанных в стационарах, выявлено 2 случая (0,1%) с клинически значимыми расхождениями; в МРЦЛД таких случаев не было. При этом важно подчеркнуть отсутствие субъективизма: врачам-экспертам недоступна информация о враче (фамилия, место работы) и медицинской организации, выполнившим исследование.

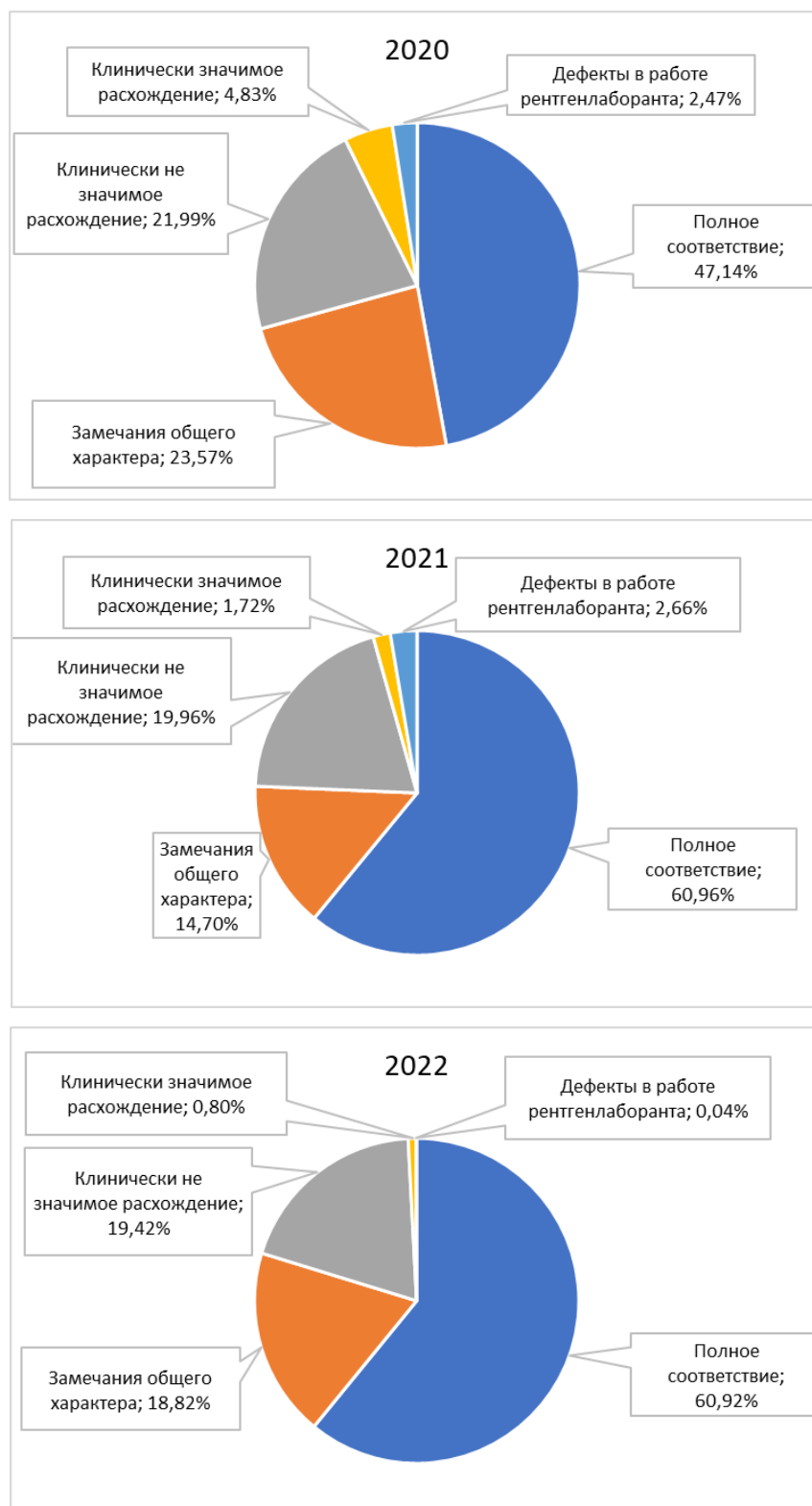


Рисунок 11 – Структура результатов ретроспективных пересмотров профилактических маммографий, выполненных в 2022 г.

Обращает на себя внимание выраженная тенденция снижения количества профилактических ММГ с клинически значимыми расхождениями, в том числе описания которых выполняются в МРЦЛД. Это является результатом комплексных мероприятий, включающих развитие централизации лучевой диагностики (активизация и улучшение деятельности МРЦЛД), системную методическую и образовательную работу с врачами-рентгенологами, а также внедрение систем поддержки принятия врачебных решений на основе технологий искусственного интеллекта, что более подробно будет изучаться в последующих параграфах.

Ретроспективные пересмотры результатов флюорографии органов грудной клетки проводятся в крайне малом объеме. В 2020 г. проверены только 4 исследования данного вида исследований (значимых расхождений нет), в 2021 г. – 144, из них 10 (0,7%) с клинически значимыми расхождениями. Такая ситуация объясняется тем, что организационно ресурсы врачей-экспертов фокусируют на диагностически сложных видах исследований лучевой диагностики (компьютерная и магнитно-резонансная томография). Вместе с тем, если интерполировать результаты пересмотров ФЛГ 2021 г., то клинически значимые расхождения могут потенциально встречаться в 9800 случаев (из расчета общего количества ФЛГ в год 1,4 млн). Это подчеркивает актуальность вопроса обеспечения качества флюорографии.

Выявляемость. Постоянное совершенствование ЕРИС ЕМИАС и общая цифровая трансформация московского здравоохранения позволили с 2022 г. начать получать объективные, достоверные данные о выявляемости значимой патологии в результате лучевых исследований, прежде всего – профилактических.

В 2022 г. признаки злокачественных новообразований выявлены в результатах 8175 профилактических маммографий из 459 237, что составило 1,78%. За первые четыре месяца 2023 г. проведено 185 130 исследований. Признаки ЗНО выявлены в 2,17% (4016).

Таким образом, усредненная выявляемость злокачественных новообразований молочной железы по результатам профилактической маммографии в Москве составляет 1,98%.

В 2022 г. признаки злокачественных новообразований выявлены в результатах 534 флюорографий из 1 383 268, что составило 0,038%; признаки туберкулеза – в 1529, что составило 0,11%. За первые четыре месяца 2023 г. проведено 443 462 исследований. Признаки ЗНО выявлены в 166 случаях (также 0,037%), туберкулеза – в 359 (0,08%).

Таким образом, по результатам флюорографии органов грудной клетки в Москве усредненная выявляемость злокачественных новообразований составляет 0,04%, туберкулеза – 0,1%.

Следовательно, профилактическое направление является одним из приоритетных в деятельности системы здравоохранения г. Москвы. Лучевые методы исследований составляют значительный компонент массовых профилактических осмотров населения. На фоне интенсивной цифровизации парка диагностического оборудования создан централизованный архив медицинских изображений (ЕРИС ЕМИАС), реализована централизация лучевой диагностики. Достигнут прогресс в обеспечении оптимальной загруженности оборудования, доступности и качестве лучевых исследований.

На пике пандемии отмечалось резкое снижение количества профилактических исследований, что подчеркивает актуальность создания мер по обеспечению бесперебойности и надежности функционирования профилактического направления в здравоохранении в любых, в том числе чрезвычайных ситуациях. В 2021–2022 гг. фиксировался стремительный рост количества профилактических исследований, соответственно расширился охват населения. Значительный рост количества профилактических мероприятий (прежде всего маммографий) создает дополнительную нагрузку на подразделения лучевой диагностики, еще более значимую в силу необходимости выполнения двойных прочтений. Анализ организации и проведения массовых

профилактических исследований, осуществляемых с применением методов лучевой диагностики в г. Москве, подтверждает актуальность цели и задач диссертационного исследования.

Способы решения проблем управления массовыми профилактическими исследованиями с применением технологий на основе искусственного интеллекта. При проведении массовых профилактических исследований с применением методов лучевой диагностики возникает ряд проблем, требующих эффективного решения. Ниже будут последовательно рассмотрены основные проблемы и предложены способы их решения на примере медицинских организаций, подведомственных Департаменту здравоохранения города Москвы.

Обеспечение функциональной инфраструктуры. Одной из основных проблем управления массовыми профилактическими исследованиями является потребность в надежной инфраструктуре. Для эффективного выполнения крупномасштабных программ скрининга необходимо иметь соответствующие средства, включая рентгенологическое оборудование, обученный персонал и информационные системы. Обеспечение финансирования в целях развития и обслуживания инфраструктуры имеет жизненно важное значение для поддержки текущих усилий по проверке. Данную проблему ДЗМ эффективно решил несколько лет назад, когда была внедрена единая радиологическая система ЕРИС ЕМИАС, что позволило не только централизовать хранение результатов лучевых исследований, но и решить кадровые вопросы за счет реализации референс-центра. Теперь врачу-рентгенологу нет необходимости находиться рядом с рентген-аппаратом, напротив, он может проводить описание исследований, выполненных в разных медицинских организациях.

Обеспечение доступности массовых профилактических исследований для населения. Организация референс-центра позволила частично решить кадровый вопрос и минимизировать время на подготовку протокола

профилактического исследования, что лишь частично решило проблему. Исследования, предполагающие двойное прочтение, не могут быть описаны сразу после того, как пациент оделся и вышел из кабинета. Также необходимо отметить разную квалификацию врачей, описывающих исследования, что в теории (данный вопрос детально не исследовался в отечественной научной литературе) может приводить к неточностям в описании и в конечном счете к разному качеству выполнения медицинской услуги. Решить этот вопрос могло бы ПО на основе ТИИ, которое, как было показано в обзоре литературы, способно выступать в качестве первого «читателя» исследования, минимизируя вероятность пропуска патологии и решая вопрос кадрового дефицита при проведении двойных описаний.

Обеспечение гарантии качества массовых профилактических исследований. Данная проблема является логичным продолжением вышеописанной ситуации и может быть частично устранена путем внедрения стандартизированных протоколов, которые уже реализованы в ДЗМ. Важными элементами контроля качества являются непрерывное обучение медицинского персонала и внедрение строгих мер контроля качества, включающие регулярные аудиты и экспертные оценки. Сейчас в ГБУЗ «НПКЦ ДиТ ДЗМ» действует отдел врачебного аудита, который занимается выборочным контролем качества выполненных врачом-рентгенологом описаний исследований. Выборочный контроль позволяет охватить лишь ограниченный процент исследований. При применении ТИИ существует возможность организации тотального контроля качества массовых профилактических исследований за счет автоматизированного сопоставления заключений врача и ПО на основе ТИИ. Случаи расхождения могут быть направлены на дополнительный контроль врачам-экспертам. Кроме этого, необходимо разработать многоступенчатую систему контроля качества работы ПО на основе ТИИ для обеспечения отбора и применения на практике

систем с максимальными показателями диагностической точности и технической стабильности.

Управление данными и информационными системами. В ходе проведения массовых профилактических исследований генерируются огромные объемы данных, которыми необходимо эффективно управлять для обеспечения точного и своевременного анализа. Эта задача также эффективно решается в настоящее время за счет реализации ЕРИС и ее интеграции в ЕМИАС. Кроме этого, подключение алгоритмов на основе ТИИ позволяет эффективно управлять потоком данных. Одно из перспективных решений – реализация триажа (сортировки) исследований по результатам обработки ПО. Так, исследования с патологией имеют максимальный приоритет, в то время как исследования с «нормой» могут быть описаны в последнюю очередь. Применение ТИИ на этапе первого чтения позволяет решать важную задачу управления потоком исследований, особенно в случае дефицита кадров, уменьшая количество времени на получение заключения у тех пациентов, для которых важно быстрое принятие решения о дальнейшей маршрутизации. Напротив, пациенты, у которых исследование не показало наличие патологии, за пределами медицинской организации могут воспользоваться электронной медицинской картой для просмотра результатов исследования (за счет интеграции ЕРИС с ЕМИАС).

Таким образом, применение ТИИ для анализа лучевых исследований в теории является действенным способом снижения рисков и оптимизации процессов управления массовыми профилактическими исследованиями. В последующих параграфах оценим применимость ТИИ для профилактических исследований.

3.2. Оценка применимости технологий искусственного интеллекта для профилактических исследований

Чтобы оценить применимость ТИИ для профилактических исследований, необходимо определить показатели диагностической точности таких решений и сопоставить с показателями диагностической точности врача-рентгенолога при описании профилактических исследований. Ниже представлены результаты двух исследований по оценке диагностической точности врача и ТИИ при описании маммографического исследования и при описании рентгенографии/флюорографии ОГК.

Показатели диагностической точности врача-рентгенолога в сравнении с программным обеспечением на основе технологий искусственного интеллекта для маммографии. Для каждого ПО на основе ТИИ и «среднего» врача-рентгенолога построены ROC-кривые (рисунок 12). На характеристических кривых отмечены пороговые значения, определенные методом максимизации индекса Юдена. Значения показателей диагностической точности представлены в таблице 8.

«Средний» врач-рентгенолог обладал следующими минимальными показателями диагностической точности (нижняя граница ДИ): AUC – 0,883, чувствительность – 0,677, специфичность – 0,874, точность – 0,800» [260, стр. 120]. Из таблицы 8 следует, что наивысшей чувствительностью при выбранном пороге, определенном методом максимизации индекса Юдена, отличались решения ИИ-4 и ИИ-5, превосходя по этому показателю «среднего» врача-рентгенолога. Самые высокие показатели специфичности были достигнуты для ИИ-2 и ИИ-5 при том же пороговом значении. Используя показатель точности для оценки баланса между чувствительностью и специфичностью, ИИ-5 превзошел «среднего» врача-рентгенолога на 3%, хотя это превосходство статистически было не значимым.

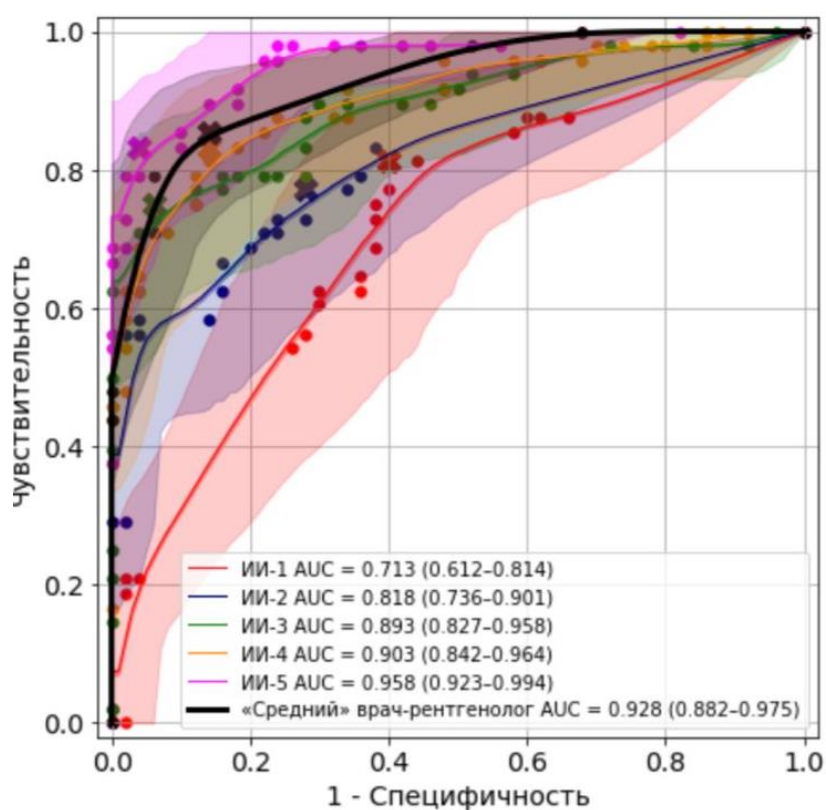


Рисунок 12 – Результаты ROC-анализа «среднего» врача-рентгенолога и ИИ-алгоритмов при анализе ММГ-исследований для групп «с патологией» и «без патологии». Увеличенным маркером отмечена точка с оптимальным значением показателей (с максимизацией индекса Юдена)

Таблица 8 – Показатели диагностической точности ПО на основе ТИИ и «среднего» врача-рентгенолога

	AUC (95% ДИ)	Чувствительность* (95% ДИ)	Специфичность* (95% ДИ)	Точность* (95% ДИ)
ИИ-1	0,713 (0,613–0,811)	0,812 (0,702–0,923)	0,600 (0,464–0,736)	0,704 (0,614–0,794)
ИИ-2	0,818 (0,736–0,897)	0,583 (0,444–0,723)	0,960 (0,906–1,000)	0,776 (0,693–0,858)
ИИ-3	0,893 (0,826–0,958)	0,750 (0,628–0,872)	0,940 (0,874–1,000)	0,847 (0,776–0,918)
ИИ-4	0,903 (0,841–0,963)	0,833 (0,728–0,939)	0,860 (0,764–0,956)	0,847 (0,776–0,918)

Продолжение таблицы 8

	AUC (95% ДИ)	Чувствительность* (95% ДИ)	Специфичность* (95% ДИ)	Точность* (95% ДИ)
«Средний» врач- рентгенолог	0,928 (0,883–0,976)	0,792 (0,677–0,907)	0,940 (0,874–1,000)	0,867 (0,800–0,935)
ИИ-5	0,958 (0,923–0,994)	0,833 (0,728–0,939)	0,960 (0,906–1,000)	0,898 (0,838–0,958)
* Показатели рассчитаны для порогового значения, определенного по максимуму индекса Юдена				

Рисунок 12 и таблица 8 приведены из ранее опубликованной работы [[260], стр 120, цветная вклейка]. Следует подчеркнуть, что ранее проведенное сравнение показателей не учитывало взаимное пересечение доверительных интервалов у ИИ-алгоритмов. Для объективного сравнения этих ПО между собой и со «средним» врачом-рентгенологом используется показатель AUC, который представляет собой интегральную оценку диагностической точности, не зависящую от выбранного порогового значения. Также применяется перестановочный тест для определения статистической значимости различий. «Полученные значения приведены в таблице 9.

Таблица 9 – Сравнение AUC-алгоритмов ИИ

j \ i	ИИ-1	ИИ-2	ИИ-3	ИИ-4	«Средний» врач-рентгенолог	ИИ-5
ИИ-1	0,71	-0,11	-0,18	-0,19	-0,22	-0,25
ИИ-2	0,11	0,82	-0,07	-0,08	-0,11	-0,14
ИИ-3	0,18	0,07	0,89	-0,01	-0,04	-0,07
ИИ-4	0,19	0,08	0,01	0,90	-0,03	-0,06
«Средний» врач-рентгенолог	0,22	0,11	0,04	0,03	0,93	-0,03
ИИ-5	0,25	0,14	0,07	0,06	0,03	0,96

На главной диагонали в таблице 9 указаны значения AUC-алгоритмов, в остальных ячейках – разница AUC $i-j$. Цветом и жирным шрифтом выделены статистически значимые различия ($p<0,05$). Данные, приведенные в таблице 9, указывают на статистически значимое превосходство «среднего» врача-рентгенолога над ИИ-1 и ИИ-2 и отсутствие статистически значимой разницы с ИИ-3, ИИ-4 и ИИ-5. В тройку с лучшими показателями по AUC, помимо «среднего» врача-рентгенолога, входят ИИ-4 и ИИ-5, между которыми не было получено статистически значимой разницы» [260, стр. 120].

Обобщив результаты исследования в настоящем подпараграфе, можно сделать вывод о том, что в качестве нижнего порога для показателя диагностической точности «площадь под характеристической кривой» рекомендуем использовать значение 0,883.

Показатели диагностической точности врача-рентгенолога в сравнении с программным обеспечением на основе технологий искусственного интеллекта для рентгенографии/флюорографии

Чувствительность, специфичность, точность и AUC на тестовом наборе данных приведены в таблице 10, характеристические кривые алгоритмов ИИ и «среднего» врача-рентгенолога представлены на рисунке 13.

Таблица 10 – Показатели диагностической точности

	AUC (95% ДИ)	Сравнение AUC ИИ и врачи, р	Чувствительность (95% ДИ)	Специфичность (95% ДИ)	Точность (95% ДИ)
	Рентгенография				
Врачи	0,96 (0,93–1,0)		0,90 (0,82–0,99)	0,96 (0,9–1,0)	0,93 (0,88–0,98)
ИИ -1	0,91 (0,85–0,97)	0,019	0,86 (0,75–0,96)	0,89 (0,80–0,98)	0,88 (0,81–0,94)
ИИ -2	0,91 (0,84–0,98)	0,123	0,76 (0,63–0,89)	0,98 (0,94–1,0)	0,88 (0,81–0,94)
	Флюорография				
Врачи	0,97 (0,94–1,0)		1,0 (1,0–1,0)	0,91 (0,83–0,99)	0,96 (0,92–1,0)
ИИ -1	0,91 (0,86–0,97)	0,049	1,0 (1,0–1,0)	0,72 (0,60–0,85)	0,86 (0,79–0,93)
ИИ -2	0,95 (0,90–0,99)	0,459	0,89 (0,80–0,98)	0,96 (0,90–1,0)	0,92 (0,87–0,98)

Из представленных данных следует, что показатели диагностической точности «среднего» врача-рентгенолога выше, чем у ИИ-1 и ИИ-2, но различия статистически не значимы из-за пересекающихся доверительных интервалов. Применение перестановочного теста показало статистически значимое ($p < 0,05$) снижение площади под характеристической кривой для ИИ-1 по сравнению с врачами. Несмотря на схожие средние значения AUC ИИ-1 и ИИ-2 для рентгенографии и выше для ИИ-2 в флюорографии, отсутствие статистически значимых различий между ИИ-алгоритмами подтверждается перестановочным тестом: при сравнении ИИ-алгоритмов между собой $p = 0,990$ для рентгенографии и $p = 0,411$ для флюорографии.

В ходе сопоставления значений специфичности максимальные значения были получены для ИИ-2, однако значения специфичности врачей и ИИ оказались в границах пересекающихся доверительных интервалов. Чувствительность при оптимальном пороге для ИИ-2 оказалась значительно ниже, чем у ИИ-1 и врачей. Средняя точность оценок врачей при выявлении патологий в органах грудной клетки превысила точность ИИ-алгоритмов, вне зависимости от типа исследования – рентгенография или флюорография.

Настоящее исследование показало, что диагностическая точность (AUC) «среднего» врача-рентгенолога и ПО на основе искусственного интеллекта сопоставима в пределах доверительных интервалов.

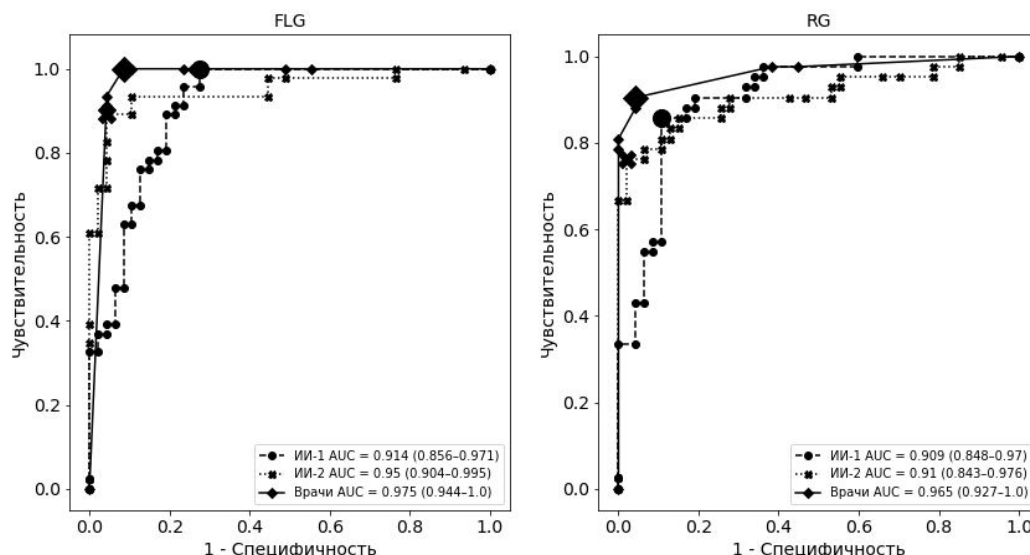


Рисунок 13 – Результаты ROC-анализа средней оценки врачей и ИИ-алгоритмов при анализе флюорографии (FLG) и рентгенографии (RG) органов грудной клетки для групп «с патологией» и «без патологии». Увеличенным маркером отмечена точка с оптимальным значением показателей (с максимизацией индекса Юдена)

Оценка, выполненная профессиональным сообществом, способствовала получению обобщенной и усредненной оценки точности врачей-рентгенологов, основанной на данных большого количества специалистов разного уровня квалификации и опыта. Применение бинарных критериев «с патологией» и «без патологии» обеспечило унификацию и корректное сравнение решений врачей и ИИ, хотя и привело к ограничению в сравнении точности выявления конкретных патологий.

Сравнивая AUC алгоритмов ИИ между собой, статистически значимых различий не обнаружено, однако детальное сопоставление показателей чувствительности и специфичности выявляет тенденцию к преобладанию одного из ИИ-алгоритмов при равных условиях (определении порога методом максимизации индекса Юдена). Например, ИИ-1 достигает 100% чувствительности за счет снижения специфичности. Это указывает на то, что оценка диагностической точности ИИ не должна ограничиваться только AUC; важно также рассматривать чувствительность и специфичность в контексте конкретной клинической задачи, требующей тонкой настройки алгоритма.

При более глубоком сравнительном анализе были обнаружены статистически значимые отличия между результатами работы врачей и одним из ИИ-алгоритмов, тогда как сравнение ИИ-алгоритмов между собой значимых различий не показало. Это различие не ограничивается только средними значениями AUC, но также включает их разброс, что может быть полезно для контроля качества ИИ-алгоритмов.

Таким образом, в качестве результата изложения данного подпараграфа можно зафиксировать минимальные значения показателей диагностической точности для принятия решения о внедрении ПО на основе ТИИ в клиническую практику: минимальный AUC 0,93–0,94. Однако такие значения оказываются недостижимыми для ПО на основе ТИИ, но при этом тонкая настройка позволяет настроить ПО на основе ТИИ на максимальную чувствительность.

3.3. Разработка, внедрение и оценка результативности организационной технологии автоматизации рабочего процесса врача-рентгенолога при проведении профилактических исследований

В контексте настоящего параграфа под термином «организационные технологии автоматизации рабочего процесса» будут пониматься и рассматриваться решения на основе ТИИ. Для эффективной работы врача-рентгенолога с результатами ПО на основе ТИИ необходима интеграция этого инструмента в привычный рабочий процесс врача. ПО на основе ТИИ интегрируются в ЕРИС по бесшовной технологии. Суть бесшовной технологии заключается в автоматической обработке исследований ПО на основе ТИИ и предоставлении результатов работы ПО на основе ТИИ в ЕРИС, что достигается путем добавления серий с результатами работы ПО на основе ТИИ в исходное исследование. Врач-рентгенолог может просматривать как оригинальные серии исследования, так и серии от ПО на основе ТИИ в едином программном интерфейсе. Для осуществления данного функционала (т.к. ПО на основе ТИИ

расположены на сторонних программно-аппаратных комплексах) используется продукт управления моделями (ПУМ).

Перечень необходимого функционала для осуществления производственных процессов с применением ПО на основе ТИИ и контроля качества:

1. Представление пользователю результатов работы ПО на основе ТИИ в трех областях работы пользователя в ЕРИС (рабочий список, область работы с изображениями и область работы с протоколом исследования).
2. Интегрированный в ЕРИС сбор обратной связи о результатах работы ПО на основе ТИИ от пользователей.
3. Автоматическое сравнение результатов работы пользователей и ПО на основе ТИИ.
4. Аудит результатов работы ПО на основе ТИИ.
5. Маршрутизация исследований к (от) ПО на основе ТИИ.
6. Обезличивание исследований и деобезличивание поступивших от ПО на основе ТИИ результатов обработки исследований.
7. Обеспечение формирования периодической отчетности о работе ПО на основе ТИИ и о работе врачей-рентгенологов с результатами работы ПО на основе ТИИ.
8. Обеспечение предварительного тестирования ПО на основе ТИИ в ТПАК до их эксплуатации в ППАК.
9. Шаблон протоколов исследований.

Представленный перечень необходимого функционала был разработан с учетом требований конечного пользователя, а также с учетом технических и инфраструктурных возможностей системы ЕРИС.

Для реализации эффективного применения ПО на основе ТИИ предложена дополнительная функциональная единица – ПУМ, которая осуществляет перенаправление исследования из DICOM-хранилища в ПО на основе ТИИ. После обработки исследования ПО оно направляется в PACS также посредством

ПУМ. С рабочей станции происходит обращение к PACS с последующей выгрузкой исследования уже с результатами работы ПО на основе ТИИ. Схематично рабочий процесс применения ПО на основе ТИИ представлен на рисунке 14.

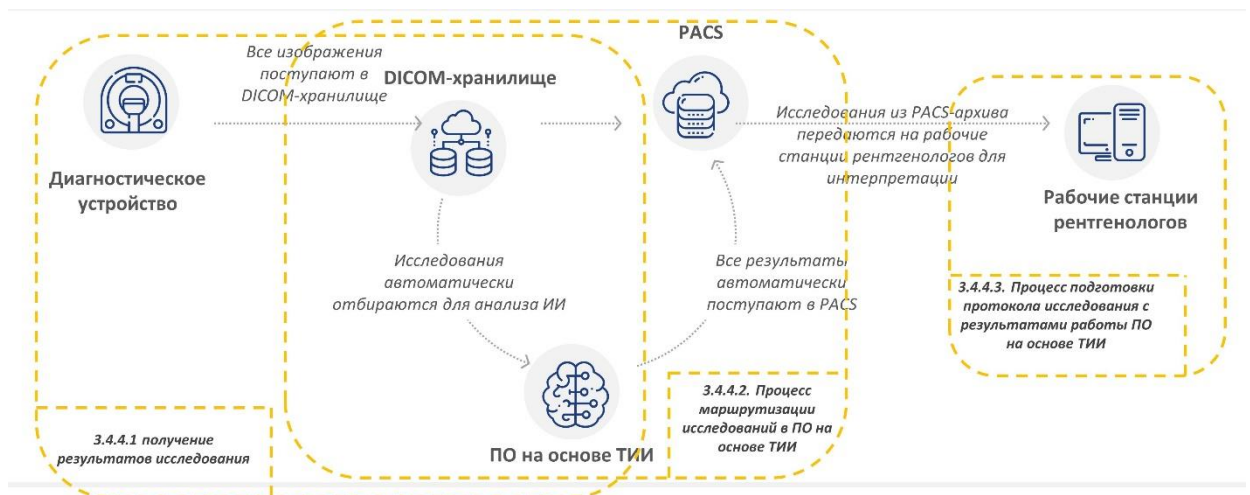


Рисунок 14 – Рабочий процесс применения ПО на основе ТИИ

Основными пользователями объединенной системы, состоящей из ЕРИС и ПО на основе ТИИ, и результатов работы данной системы являются:

1. Врачи-рентгенологи в ОЛД, которые работают в ЕРИС. Данная категория пользователей получает дополнительный инструмент.
2. Врачи-эксперты клинического аудита со стороны ГБУЗ «НПКЦ ДиТ ДЗМ», которые выполняют аудит и мониторинг работы ПО на основе ТИИ.
3. Исследователи со стороны ГБУЗ «НПКЦ ДиТ ДЗМ», участвующие в настройке и мониторинге производственных процессов в ЕРИС, связанных с применением ПО на основе ТИИ. Данная категория пользователей обеспечивает допуск ПО на основе ТИИ к работе в ППАК; контрольно-техническое испытание и технологический мониторинг объединенной системы, состоящей из ЕРИС и ПО на основе ТИИ; анализирует отчеты о работе ПО на основе ТИИ; настраивает маршрутизацию потоков исследований на определенные ПО на основе ТИИ.

Эмпирическим путем разработаны технические и функциональные требования к результатам работы ПО на основе ТИИ. При интеграции в АРМ

врача-рентгенолога такое программное обеспечение должно обеспечить следующие функции: сортировку, маркировку патологических находок на изображении, подготовку проекта текстового описания результатов лучевого исследования.

1. **Сортировка.** Процесс определения приоритета исследования в зависимости от результатов работы ПО на основе ТИИ и присвоение ему одной из трех категорий, обозначаемых различными цветами (красный, оранжевый, бесцветный). Пример реализации триажа продемонстрирован на рисунке 15.

Modality	Assigned to	Patient name and Patient ID	Patient date of birth	Procedure name
AS, ASMT, CT	_RUS Радиологи, _R...	♂ ТЕСТИИ1 ТЕСТИИ1 ТЕСТИИ1 AG00125...	01-01-1956	Rg-скопия брюшной...
AS, CT	_RUS Радиологи, _R...	♂ ANONYM PATIENT01	26-05-1945	Компьютерная томог...
AS, CT	_RUS Радиологи, _R...	♂ ANONYM PATIENT01	08-05-1960	Компьютерная томог...
AS, CT	_RUS Радиологи, _R...	♀ ANONYM PATIENT01	17-12-1987	Компьютерная томог...
AS, CT	_RUS Радиологи, _R...	♀ ANONYM PATIENT01	21-06-1940	Компьютерная томог...
AS, CT	nil lu	♂ ANONYM FUNCDATASET	20-02-1937	Компьютерная томог...
AS, CT	_RUS Радиологи, _R...	♀ ANONYM FUNCDATASET	06-06-1972	Компьютерная томог...
AS, CT	_RUS Радиологи, _R...	♀ ANONYM KLBDATASET	26-08-1943	Компьютерная томог...
ASMT, CT	_RUS Радиологи, _R...	♀ ANONYM PATIENT01	17-12-1987	Компьютерная томог...
ASMT	_RUS Радиологи, _R...	♂ CTTESTSR CTTESTSR AG00098555	01-01-1956	Компьютерная томог...
ASMT	_RUS Радиологи, _R...	♂ TESTXRAY1 TESTXRAY1 TESTXRAY1 AG...	01-01-1956	Rg-графия диафрагмы
ASMT	_RUS Лаборанты, _R...	♂ TESTXRAY1 TESTXRAY1 TESTXRAY1 AG...	01-01-1956	ASMT_RG-GRAFIYA DI...
ASMT	_RUS Лаборанты, _R...	♂ TESTXRAY1 TESTXRAY1 TESTXRAY1 AG...	01-01-1956	ASMT_RG-GRAFIYA DI...
ASMT	_RUS Лаборанты, _R...	♂ GHOLSTON WRIGHT	16-12-1993	ASMT_RG-SKOPIYA B...
ASMT, CR	_RUS Лаборанты, _R...	♂ GASKEY MARLOW R	13-01-2013	CR_RG-SKOPIYA BRY...
ASMT, CR...	_RUS Лаборанты, _R...	♂ BRAUBURGER COYE STEVE	14-08-1966	CR_RG-SKOPIYA BRY...
ASMT, CT	_RUS Радиологи, _R...	♀ ANONYM PATIENT01	04-11-1966	Компьютерная томог...
ASMT, CT	_RUS Лаборанты	♀ ANONYM PATIENT01	17-12-1987	Компьютерная томог...
AS, ASMT, ...	_RUS Лаборанты	♂ ТЕСТИИ1 ТЕСТИИ1 ТЕСТИИ1 AG00125...	01-01-1956	Rg-скопия брюшной...
ASMT	_RUS Лаборанты	♂ CTTESTSR CTTESTSR AG00098555	01-01-1956	Компьютерная томог...
ASMT	_RUS Лаборанты	♂ TESTXRAY1 TESTXRAY1 TESTXRAY1 AG...	01-01-1956	Rg-графия диафрагмы

Рисунок 15 – Пример реализации сортировки для исследований, обработанных ПО на основе ТИИ. Цветом выделены исследования с максимальным приоритетом (классифицированные ПО как «с патологией»)

Сортировка предназначена для оптимизации рабочего процесса врача-рентгенолога: в первую очередь описываются исследования с подозрением на патологию.

2. **Дополнительная серия (DICOM secondary capture (SC)).** Дополнительная серия с результатами работы ПО, содержащая маркировку и/или температурную карту, локализирующую находку. Пример DICOM SC представлен на рисунке 16.

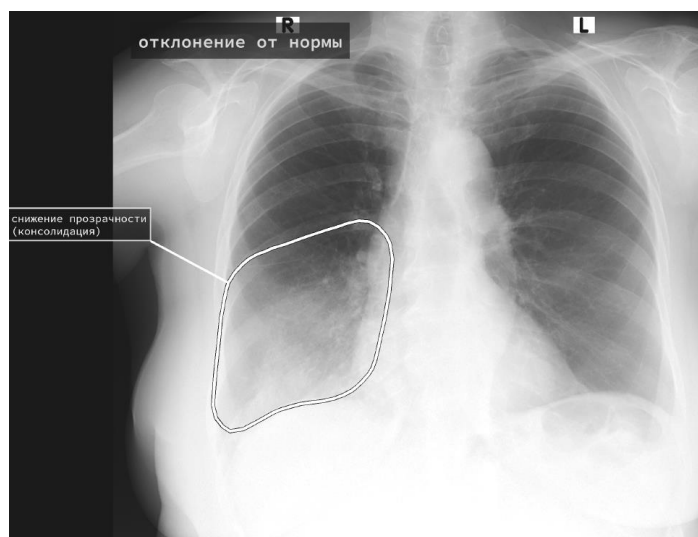


Рисунок 16 – Пример реализации DICOM SC

В случае если визуализация находок на изображениях исследования не требуется или отсутствует, информация о результатах анализа должна быть предоставлена в текстовом виде в этой дополнительной серии. Предназначена для визуализации результатов работы ПО на основе ТИИ в привычном врачу-рентгенологу формате. Для однозначной идентификации пользователем (врачом-рентгенологом) ПО на основе ТИИ название дополнительной серии должно соответствовать формату «**Наименование ПО**». Кроме того, на дополнительной серии должна присутствовать служебная информация: название сервиса, версия ПО, дата анализа и т.д. В случае, если ПО на основе ТИИ определяет несколько категорий находок, каждая из них должна быть однозначно идентифицируема. Отсутствие выявляемой патологии также требует формирования отчета от ПО на основе ТИИ для информирования врача-рентгенолога.

ПО на основе ТИИ, вне зависимости от количества определяемых целевых патологий, должно представить результат обработки исследования в виде одной дополнительной серии изображений. Одно и то же исследование может быть обработано несколькими ПО на основе ТИИ, и от каждого отдельного ПО на основе ТИИ должна быть предоставлена только одна дополнительная серия в формате SOPClass 1.2.840.10008.5.1.4.1.1.7.

Для исследований, содержащих несколько оригинальных серий, ПО на основе ТИИ должно предоставить результаты обработки наиболее подходящих серий или проекций.

Название дополнительной серии соответствует названию ПО на основе ТИИ и прописывается в теге Series Description.

Настройки яркости и контрастности изображения, а также разрешение изображения дополнительной серии не должны отличаться от исходных или, в случае представления искажения изображения в DICOM SR, должно быть указание на то, что эта дополнительная серия не диагностического качества.

В случае отсутствия патологических изменений в дополнительной серии необходимо указать «Целевая патология не выявлена». При этом дополнительная серия, подготовленная ПО на основе ТИИ, может быть представлена одним изображением.

3. Текстовое описание – DICOM structure report (SR) – содержит текстовую информацию о ПО на основе ТИИ и результатах его работы. Оптимальная для отображения структура DICOM SR: название ПО на основе ТИИ, заключение, детализация находок, краткое руководство пользователя. Данная структура оптимальна, но может быть дополнена иной информацией. Информирование врача-рентгенолога:

- 1) о принадлежности результатов конкретному ПО на основе ТИИ;
- 2) об общем выводе ПО на основе ТИИ по всему исследованию в целом, который врач-рентгенолог может использовать при составлении заключения;
- 3) о деталях результатов анализа, которые врач-рентгенолог может использовать при составлении описания;
- 4) о показаниях и ограничениях использования ПО на основе ТИИ (в том числе при проверке на технологические дефекты);
- 5) о корректном использовании результатов работы ПО на основе ТИИ.

Пример и структура DICOM SR представлен на рисунке 17.

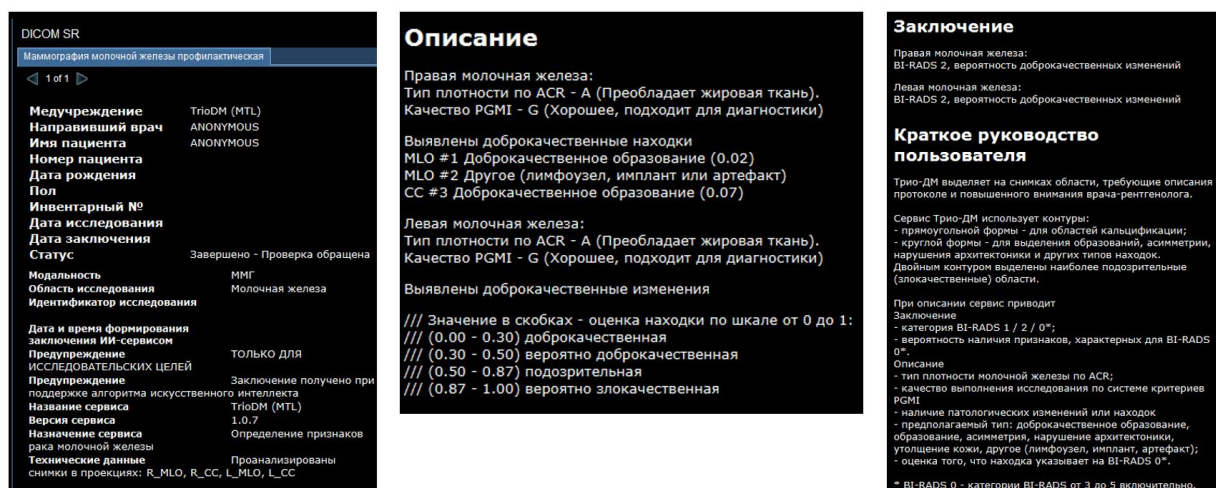


Рисунок 17 – Пример DICOM SR

DICOM SR должен обязательно присутствовать вне зависимости от выявления ПО на основе ТИИ наличия или отсутствия признаков патологии. Структура DICOM SR представлена ниже:

- Вид исследования.
- Область исследования: анатомическая область исследования.
- Идентификатор исследования: Study UID исследования в ЕРИС ЕМИАС.
- Дата и время формирования заключения ПО на основе ТИИ: дата формирования заключения ПО на основе ТИИ.
- Предупреждение:
 - а) заключение получено при поддержке алгоритма искусственного интеллекта;
 - б) только для исследовательских целей.
- Наименование сервиса: наименование ПО на основе ТИИ.
- Версия сервиса: номер версии ПО на основе ТИИ.
- Назначение сервиса: клинические задачи, которые решает сервис.
- Технические данные: количество и наименование обработанных проекций или толщина и количество обработанных ПО срезов в исследовании.
- Описание: текст описания.

- Заключение: текст заключения.
- Детализация находок: изображения находки.
- Руководство пользователя.

Раздел «Описание» должен воспроизводить информацию в соответствии с врачебным заключением по исследуемой патологии, чтобы на его основе можно было создать электронную медицинскую запись в системе. Содержание описания, предоставляемого ПО на основе ТИИ, варьируется в зависимости от типа исследования и должно соответствовать определенным критериям.

Раздел «Заключение» – аналогично врачебному, в целях формирования на его основе электронной медицинской записи в информационной системе. В зависимости от вида исследования заключение ПО на основе ТИИ должно содержать данные, соответствующие требованиям.

Раздел «Краткое руководство» должен содержать перечень целевой патологии, обнаруживаемой ПО на основе ТИИ, описание маркировки патологических изменений.

Для ПО на основе ТИИ, одновременно обрабатывающих исследование по нескольким целевым патологиям, необходимо представление результата в виде одного DICOM SR.

Взаимодействие ПО на основе ТИИ с ПУМ осуществляется посредством JSON, так называемого Kafka-сообщения. Общая вероятность патологии в исследовании в целом, передаваемая в системном сообщении Kafka – в диапазоне 0–100. В случае работы ПО на основе ТИИ одновременно по нескольким целевым патологиям в одной анатомической области необходимо направление в сообщении Kafka о вероятности наличия каждой из определяемых патологий.

Количественные характеристики патологии должны отражаться в следующем формате: результаты измерений, стадия/степень по принятой классификации.

**Основные процессы при работе с программным обеспечением
на основе технологий искусственного интеллекта**

С целью обеспечения стабильной работы врача-рентгенолога с ПО на основе ТИИ были разработаны и осуществлены следующие рабочие процессы.

Процесс: Получение результатов исследования

Иницилирующее событие: создано задание на исследование (статус **«Назначено»**).

Завершающее событие: сформировано задание на создание заключения, DICOM-файл передан для обработки ПО на основе ТИИ.

Участники процесса: рентгенолаборант, ЕРИС.

После сохранения направления на исследование рентгенолаборантом в ЕРИС автоматически создается задание на исследование и передается на ДУ аппарата (worklist). Статус задания на исследование: **«Назначено»**.

Рентгенолаборант запускает процесс исследования (сканирования исследуемой области) через консоль аппарата. Статус задания на исследование: **«Выполняется»**.

После проведения исследования формируется DICOM-файл, содержащий данные о результатах исследования (изображения исследования), и передается в ЕРИС. В ЕРИС при получении DICOM-файла выполняется проверка финализации и полноты исследования (установлен статус **«Исследование завершено»**).

После получения с аппарата DICOM-файла со статусом **«Исследование завершено»** должны выполняться следующие действия:

1. Заданию на исследование устанавливается статус **«Выполнено»**.
2. DICOM-файл передан во внешний модуль для обработки изображений исследования ПО на основе ТИИ, если исследования данного МО/ДУ/кода процедуры предназначены для анализа ПО на основе ТИИ.

3. Заданию на создание заключения устанавливается статус **«Назначено, приоритет „Норма”»**. Задание на создание заключения должно формироваться не позднее чем через 1 минуту после получения DICOM-файла с изображениями.

Если нет связи ЕРИС с ДУ, рентгенолаборант должен вручную внести данные пациента на ДУ аппарата и провести исследование. После проведения исследования DICOM-файл с изображениями исследований сохраняется в памяти аппарата и передается в ЕРИС после восстановления связи. Результаты проведенного исследования должны быть доступны рентгенолаборанту для проведения контроля качества. Рентгенолаборант должен создать направление на исследование, при помощи функции «Контроль качества» связать направление на исследование с полученными результатами (DICOM-файл) и завершить исследование, чтобы задание на исследование стало недоступно на ДУ аппарата. После проведения контроля качества рентгенолаборантом для врача-рентгенолога должно быть сформировано задание на создание заключения, и результаты исследования должны быть доступны для расшифровки.

Процесс маршрутизации исследований в программном обеспечении на основе технологий искусственного интеллекта

Исследования, направляемые на обработку к ПО на основе ТИИ, должны быть предварительно обезличены, а возвращаемые результаты деобезличены. Исследования, направляемые на ИИ-аудит, должны быть предварительно обезличены.

Иницилирующее событие: поступление исследования в ЕРИС.

Завершающее событие: от ПО на основе ТИИ результаты получены в ЕРИС.

Участники процесса: ЕРИС, ПУМ, ПО на основе ТИИ, ГБУЗ «НПКЦ ДиТ ДЗМ» (формирует правило маршрутизации).

Типы маршрутизации: постоянная (на протяжении всего периода маршрутизации ДУ не меняются), динамическая (ДУ меняются через определенный промежуток времени, например, «шахматная» маршрутизация).

Бизнес-процесс маршрутизации:

1. ГБУЗ «НПКЦ ДиТ ДЗМ» формирует задание на маршрутизацию.
2. ЕРИС настраивает маршрутизацию.
3. Исследование направляется – в ЕРИС.
4. Проверяется условие соответствия маршрутизации.
5. Исследование направляется в ПУМ.
6. Исследование забирается ПО на основе ТИИ, если соответствует правилу маршрутизации.
7. Возврат результата работы ПО на основе ТИИ в ЕРИС.

Алгоритм маршрутизации:

1. Проверка по AE-title: подлежат ли исследования с данного ДУ анализу каким-либо ПО на основе ТИИ. Если нет, то дальнейшие действия отсутствуют, если да, то переходим к п. 2 (далее проверяется соответствие привязанному ПО на основе ТИИ).

Если по AE-title обнаружена привязка к нескольким ПО на основе ТИИ, то дальнейшие действия, согласно пунктам, выполняются параллельно для всех привязанных ПО на основе ТИИ.

2. Проверяем соответствие кода процедуры. Если нет, то дальнейшие действия отсутствуют, если да, то переходим к п. 3.

3. Проверяем соответствие анатомической области. Если нет, то дальнейшие действия отсутствуют, если да, то переходим к п. 4.

4. Проверяем соответствие по возрастному ограничению. Если нет, то дальнейшие действия отсутствуют, если да, то переходим к п. 5.

5. Счетчик количества исследований текущего ПО на основе ТИИ +1. Сравниваем с ограничением по количеству исследований на данном ПО на основе ТИИ. Если меньше, то исследование направляется ПО на основе ТИИ, если

больше, то дальнейших действий с исследованием не происходит. Формируется сообщение в логе о превышении величины ограничения числа исследований данного сервиса.

Завершающие событие: от ПО на основе ТИИ результаты получены в ЕРИС или завершение по причине несоответствия критериям на любом этапе алгоритма.

Количество исследований от ДУ к ПО на основе ТИИ предварительно рассчитывается исходя из загруженности ДУ/ПО на основе ТИИ и потребности в таких исследованиях от МО. После успешного прохождения ПО на основе ТИИ этапа функционального тестирования, а также при успешной интеграции ПО на основе ТИИ в ППАК вводится в действие предварительно разработанный алгоритм маршрутизации. Данный алгоритм содержит данные о МО, а также ДУ в этих МО, исследования из которых перенаправляются к интегрированному ПО на основе ТИИ. Маршрутизация исследований осуществляется по «Заданию на подключение», содержащему следующие параметры: наименование ПО на основе ТИИ; количество исследований для маршрутизации; ae-title диагностических устройств; тип оборудования; код диагностической процедуры; наименование диагностической процедуры; анатомическая область; возрастные ограничения. Назначение исследования на обработку ПО на основе ТИИ происходит в ПУМе, где располагаются исследования после деперсонализации. Необходимо разработать правило отбора отдельных серий исследования для маршрутизации к ПО на основе ТИИ в целях оптимизации трафика. Количество исследований к выбранному ПО на основе ТИИ может быть изменено при отклонении от установленных целевых значений с целью оптимизации трафика и загрузки отдельных ПО на основе ТИИ.

Проверку корректности работы ПО на основе ТИИ и выдаваемых ими результатов в условиях возможной вариативности входных данных (различные ДУ, различные наименования процедуры) необходимо выполнить на этапе интеграции ПО на основе ТИИ в ТПАК. На этапе перехода в ППАК необходимо

осуществить проверку корректности маршрутизации от ДУ к ПО на основе ТИИ и получаемых результатов от ПО на основе ТИИ для исследований от каждого ДУ.

Источником данных для интерфейса маршрутизации является ГБУЗ «НПКЦ ДиТ ДЗМ». Затем эти данные передаются в техподдержку ЕРИС для подключения либо отключения требуемых ДУ.

Отключение ПО на основе ТИИ осуществляется удалением строки с ним. Отключение оборудования осуществляется удалением ДУ из таблицы.

Требования по приоритизации исследований в зависимости от обнаруженных патологий: в случае выставления вероятности патологии, превышающей порог срабатывания (cut-off) для данного ПО на основе ТИИ, данное исследование должно приоритизироваться в рабочем списке врача-рентгенолога. Это позволяет врачам-рентгенологам описывать в первую очередь исследования с патологией.

В случае, если в результате обработки исследования ПО на основе ТИИ была выявлена патология, данному исследованию должен устанавливаться высший приоритет. Такие задания должны быть подсвечены для врача-рентгенолога в списке активных заданий и иметь соответствующий индикатор. Необходимо также реализовать ранжирование заданий на создание заключений по приоритету (сначала все с приоритетом «**НЕМЕДЛЕННО!**»), затем с приоритетом «Срочно», и далее все задания с приоритетом «Норма») и по дате создания (сначала в списке отображаются исследования, проведенные раньше по времени).

Статус выявленной патологии определяется путем сравнения порога срабатывания данного ПО на основе ТИИ и полученной вероятности патологии в текущем исследовании от ПО на основе ТИИ в сообщении Kafka. Порог срабатывания задается поставщиком ПО на основе ТИИ, в дальнейшем может быть скорректирован в ходе процедуры калибровки (см. параграф 4.1).

После получения результатов обработки изображений ПО на основе ТИИ и добавления их в исходное исследование в рабочем списке врача-рентгенолога

в строке с соответствующим исследованием должна появиться маркировка, обозначающая, что в исследовании хранятся завершённые результаты работы ПО на основе ТИИ. Кроме этого, исследования с завершёнными результатами обработки хотя бы одного ПО на основе ТИИ также отображаются в отдельном списке (ИИ-список). Правило ранжирования исследований в ИИ списке аналогично общему рабочему списку.

В случае обработки одного исследования несколькими ПО на основе ТИИ приоритет исследования выставляется в соответствии с максимальным из определенных ПО на основе ТИИ (например, если одно исследование обрабатывается тремя ПО на основе ТИИ, каждое из которых вернуло разные приоритеты: «НЕМЕДЛЕННО!», «Срочно», «Норма», то приоритет для «Задания на создание заключения» должен быть выставлен «НЕМЕДЛЕННО!»).

Регламент подготовки протокола исследования с результатами работы программного обеспечения на основе технологий искусственного интеллекта

Иницилирующее событие: сформировано задание на создание заключения (статус «*Назначено*», приоритет «*Норма*»).

Завершающие событие: заключение сохранено (статус «*Предварительный*» или «*Утверждено*»).

Пользователи: врач-рентгенолог.

Процесс:

1. Выбрать исследование из списка с высоким приоритетом.
2. Открыть исследование в режиме работы с протоколами исследования.
3. Ознакомиться с содержимым DICOM SR: краткое руководство пользователя, заключение, детализация находок.
4. Перейти в режим работы с изображениями.
5. Изучить исходную серию и серию с результатами работы ПО на основе ТИИ. Последовательность действий, а также режим отображения врач

выбирает по своему усмотрению. Рекомендованный режим отображения – 2 изображения: из исходной серии и серии от ПО на основе ТИИ.

6. Перейти в режим работы с протоколами. Если патологических изменений не выявлено и врач с этим полностью согласен, то оставить предзаполненный протокол для приоритета «Норма». В противном случае – заполнить поля «Описание» и «Заключение» согласно обнаруженной патологии.

7. Оценить работу ПО на основе ТИИ, оставив обратную связь.

8. Сохранить и завизировать протокол.

При наличии результатов работы более чем одного ПО на основе ТИИ врач имеет право ознакомиться с результатами как одного, так и нескольких ПО на основе ТИИ по своему усмотрению. Обратную связь врач должен дать по тому ПО на основе ТИИ, по которому он проинструктирован давать оценку.

Таким образом, в реальной клинической практике ПО на основе ТИИ должно решать три задачи: приоритизация исследования в рабочем списке врача-рентгенолога, подготовка текстового заключения и маркировка на исследовании патологических областей. Краткая информация о предоставляемых ПО на основе ТИИ данных размещена в таблице 11.

Таблица 11 – Описание и назначение элементов интерфейса диагностического рабочего стола

Элемент	Описание	Назначение
Сортировка	Исследование с патологией маркируется красным цветом, без патологии – бесцветным.	Оптимизация рабочего процесса врача-рентгенолога: в первую очередь описываются исследования с подозрением на патологию
Графическое обозначение	Наличие отметки об обработке исследования ИИ. В поле Modality в общем рабочем списке уведомление через определенное значение тэга (например, Modality = ASMT).	Информирование врача-рентгенолога о завершенном анализе и, как следствие, возможности оценки результатов ИИ-сервиса

Продолжение таблицы 11

Элемент	Описание	Назначение
Серия ИИ	Дополнительная серия, сформированная ИИ и содержащая графическую маркировку патологических областей.	Отображение результатов работы ИИ-сервиса в привычном врачу-рентгенологу формате
Название доп. серии*	Название дополнительной серии должно соответствовать формату «ИИ. <i>Наименование сервиса</i> »	Однозначная идентификация серий, сгенерированных ИИ-сервисом
Синхронизация серий	Возможность параллельного синхронного просмотра нативной серии и серий, сгенерированных ПО на основе ТИИ	Возможность сравнения врачом-рентгенологом результатов ИИ-сервиса с исходными данными
Информация о сервисе	Наличие на изображениях дополнительных серий служебной информации о сервисе: название сервиса, версия ПО, дата анализа, ...	Подтверждение принадлежности результатов конкретному ИИ-сервису
Вероятность(-и) находки(-ок)	Наличие на изображениях дополнительных серий информации о значении вероятности патологических изменений в целом для всего исследования и/или для каждой находки (в%)	Данная информация необходима врачу-рентгенологу для правильной интерпретации значения вероятности находки с учетом обозначенного порога срабатывания для конкретного ИИ-сервиса
Категории находок	В случае, если ИИ-сервис определяет несколько категорий находок, каждая из них должна быть однозначно идентифицируема в дополнительной серии	Данная информация необходима врачу-рентгенологу для правильной интерпретации результатов анализа ИИ-сервиса

Продолжение таблицы 11

Элемент	Описание	Назначение
DICOM SR	Файл DICOM SR должен содержать следующую информацию: Оптимальная для отображения структура DICOM SR: название ПО на основе ТИИ, заключение, детализация находок, краткое руководство пользователя. Данная структура оптимальна, но может быть дополнена иной информацией	Информирование врача-рентгенолога о: 1) принадлежности результатов конкретному ИИ-сервису; 2) общем выводе ИИ-сервиса о всем исследовании в целом, который врач-рентгенолог может использовать при составлении заключения; 3) деталях результатов анализа, которые врач-рентгенолог может использовать при составлении описания; 4) показаниях и ограничениях использования ИИ-сервиса (в том числе при проверке на технологические дефекты); 5) корректном использовании результатов работы ИИ-сервиса**
Шаблон протокола	Наличие предзаполненного заключения (наличие предзаполненного описания – опционально), сформированного ИИ.	Стандартизация составления протоколов описаний врачами-рентгенологами. Автоматизация сбора результатов описания врачами-рентгенологами исследований, проанализированных ИИ-сервисами

Продолжение таблицы 11

Элемент	Описание	Назначение
Обратная связь	Наличие окна обратной связи по каждому ИИ-сервису: 1) выбор ответа: • «Согласен», • «Не согласен» 2) в случае выбора ответа «Не согласен», заполняется текстовое поле «Комментарий» с выпадающим списком: 01. Дефект (технологический). 02. Грубое расхождение в работе сервиса. 03. Неверный диагноз. 04. Неверная локализация внутри целевого органа. 05. Неверная классификация находки. 06. Прочее. Важно! Выбрать можно один или несколько пунктов ошибок или неточностей, допущенных ИИ-сервисом	Получение информации о работе ИИ-сервисов от пользователей. 01. Дефект (технологический). Наличие технологического дефекта, не позволяющего использовать результаты ИИ-сервиса при описании 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 об исследовании). 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 (например, очаговое образование расценено как фиброз). 06. Прочее (указать)

***Ограничения при внедрении программного обеспечения
на основе технологий искусственного интеллекта***

В ППАК допускается только то ПО на основе ТИИ, которое реализовало бесшовную интеграцию с ЕРИС согласно целевой модели, успешно прошедшее тестирование в тестовом контуре (ТПАК).

ПО на основе ТИИ, подключенные к ЕРИС, должны работать только с назначенными им потоками исследований из МО/ДУ. Исследования, направляемые на ПО на основе ТИИ, должны быть обезличены; исследования, содержащие вшитые в изображение персональные данные, маршрутизироваться на ПО на основе ТИИ не могут. Передача исследований за пределы границы Российской Федерации запрещена. Хранение и использование исследований на стороне разработчика ПО на основе ТИИ также запрещены. Допустима только следующая схема работы с исследованием: исследование поступает на вход ПО, осуществляется контроль соответствия входных данных возможностям ПО на основе ТИИ; если исследование не может быть обработано ПО на основе ТИИ – возвращается сообщение об ошибке; если исследование может быть обработано – оно обрабатывается, и в ответ направляются результаты работы ПО на основе ТИИ. После этого исследования безвозвратно удаляются на стороне развертывания ПО на основе ТИИ. Контроль результатов работы врача-рентгенолога и ПО на основе ТИИ осуществляется при проведении мониторингов. Более детально требования к результатам работы ПО на основе ТИИ будут разобраны в последующих главах.

Возможности для постобработки и модификации результатов ПО на основе ТИИ отсутствуют, за исключением процессов деобезличивания.

***Требования к отчетности о работе программного обеспечения
на основе технологий искусственного интеллекта***

При внедрении ПО на основе ТИИ необходимо ведение отчетной документации. Ниже предлагается перечень основных полей, содержащихся в отчетной документации.

1. Наименование ИИ – уникальное название ПО на основе ТИИ.
2. Уникальный идентификатор исследования – уникальный номер, по которому можно идентифицировать обработанное исследование, в системе ЕРИС – UID.
3. Наименование процедуры – код медицинской услуги.
4. МО – наименование медицинской организации.
5. Ae-title – уникальный идентификатор диагностического устройства.
6. Вид исследования: РГ, ФЛГ, ММГ.
7. Дата отправки исследования в PACS – дата и время отправки исследования в PACS, соответствует его фактическому выполнению на диагностическом устройстве.
8. Дата создания заключения – дата и время визирования протокола исследования врачом-рентгенологом.
9. Возраст пациента – количество полных лет пациенту на момент проведения исследования.
10. Время начала обработки моделью ИИ – дата и время начала обработки исследования ПО на основе ТИИ.
11. Время окончания обработки моделью ИИ – дата и время окончания обработки исследования ПО на основе ТИИ.
12. Автор заключения – Ф.И.О. врача-рентгенолога, завизировавшего протокол исследования.
13. Описание исследования – текст из раздела протокола «описание».

14. Заключение по исследованию – текст из раздела протокола «заключение».

15. Автор обратной связи – Ф.И.О. врача-рентгенолога, оставившего обратную связь по данному исследованию;

16. Вероятность патологии – вероятность наличия патологии, которую выставляет ПО на основе ТИИ.

17. Порог срабатывания (англ. «cut-off»), % – уникальное для каждого ПО на основе ТИИ значение вероятности, соответствующей порогу, выше нее исследование классифицируется как патологическое, ниже – как нормальное.

18. Наличие патологии – бинарная оценка исследования в соответствии с пороговым значением из п. 17.

19. Дата и время отправки в сервис-ПУМ – дата и время, когда ЕРИС направил исследование ПО на основе ТИИ.

20. Дата и время получения результата от сервиса-ПУМ, дата и время, когда ПО на основе ТИИ направило результаты обработки исследования в ЕРИС.

Это список основных показателей, которые будут использоваться при оценке ПО на основе ТИИ в рамках настоящей работы.

Оценка результативности предложенной методики внедрения программного обеспечения на основе технологий искусственного интеллекта

Внедрение ПО на основе ТИИ приводит к изменению бизнес-процесса работы ОЛД, однако бизнес-процесс работы рентгенолаборанта остается неизменным.

Пациент приходит в отделение лучевой диагностики по направлению на исследование, которое выдал врач-клиницист. Рентгенолаборант в РИС осуществляет поиск пациента и заполняет необходимые сведения о предстоящем исследовании в случае, если такой пациент найден в базе. После заполнения всех необходимых для исследования данных (выбор вида исследования, оборудования,

на котором оно будет осуществляться) задача на проведение исследования автоматически передается на консоль диагностического устройства (так называемый *worklist*). После выбора на консоли ДУ протокола исследования рентгенолаборант проводит непосредственно исследование. Если связи с РИС нет и рентгенолаборант не может сформировать направление, он должен создать задание на исследование на консоли ДУ. В результате проведения исследования формируется DICOM-файл с изображениями и отправляется в РИС при наличии связи. Если связи с ЕРИС нет, DICOM-файл сохраняется на аппарате и передается в РИС после восстановления связи.

Потенциально при внедрении ПО на основе ТИИ, способного осуществлять контроль качества выполнения рентгеновского исследования, в частности, корректности укладки, данное ПО должно будет подключиться до этапа направления DICOM исследования в РИС и, в случае несоответствия качества проведенного исследования принятым нормам и стандартам, выдать на консоли ДУ информацию о техническом дефекте при выполнении исследования. Также в условиях небольших МО, имеющих локальную PACS-систему, возможно подключение ПО на основе ТИИ напрямую к PACS для контроля качества.

Работа рентгенолаборанта завершается, и на следующем этапе DICOM с исследованием оказывается в РИС (далее по тексту в случае небольшой МО будем подразумевать под РИС PACS-систему с просмотрщиком, при этом бизнес-процессы будут идентичны). В настоящей модели исключаем тот случай, при котором в МО имеется всего одно АРМ для врача-рентгенолога, находящееся на ДУ. Как только исследование оказалось в РИС, оно становится доступно для отправки на ПО на основе ТИИ. Отображенные в результатах показатели времени обработки рентгенографического исследования для рассматриваемых видов исследований не превышают 3–5 минут (с учетом времени на передачу данных, см. таблицу 12), поэтому не оказывают в данном процессе существенного влияния на работу врача-рентгенолога.

Таблица 12 – Время подготовки протокола врачом-рентгенологом и ПО на основе ТИИ в зависимости от вида исследования и года

Вид исследования	Год	Длительность в формате чч:мм:сс		
		Описание врача	Предоставление результатов ИИ	Обработка исследований ПО на основе ТИИ
РГ/ФЛГ	2020	00:26:00	00:02:25	00:00:46
	2021	00:23:00	00:02:29	00:00:14
	2022	00:13:00	00:02:51	00:00:19
ММГ	2020	01:38:00	00:06:20	00:02:43
	2021	04:23:00	00:05:07	00:01:17
	2022	00:28:00	00:04:39	00:00:52

Как видно из приведенных в таблице 12 данных, медианное время подготовки протокола исследования врачом-рентгенологом за последние три года сокращается, однако существенно превышает время готовности результатов обработки ПО на основе ТИИ. При этом большая часть времени уходит на пересылку данных, нежели на обработку их моделью ИИ.

После завершения обработки исследования ПО на основе ТИИ направляет в РИС дополнительную серию (при наличии) и описание исследования. РИС, получив данную информацию, приоритизирует исследование в рабочем списке врача-рентгенолога, предоставляя доступ врачу-рентгенологу к текстовому протоколу. Врач-рентгенолог выбирает исследование с максимальным приоритетом и в области работы с изображениями в РИС просматривает и анализирует изображения исследования, формируя представление о заключении. После этого врачу рентгенологу рекомендовано ознакомиться с результатами работы ПО на основе ТИИ.

В случае согласия с результатами работы врач-рентгенолог может воспользоваться текстовым описанием, подготовленным ПО на основе ТИИ, скопировав его в соответствующие поля протокола исследования. При необходимости он вносит изменения и дополнения в этот протокол. В случае несогласия врач самостоятельно заполняет разделы протокола (рекомендуется

оставить обратную связь по данному исследованию предусмотренным в данной РИС способом). Эта информация на периодической основе может направляться разработчикам ПО на основе ТИИ для повышения его качества. Далее протокол исследования визируется и направляется в МИС.

В настоящей работе была предложена интеграция ПО на основе ТИИ в ЕРИС ЕМИАС. Особенностью данной системы является большой поток исследований. Как показали результаты проведенного анализа, общий поток профилактических исследований составил 4,3 млн за три года, средний поток – 807 тысяч флюорограмм, 422 тысячи рентгенограмм органов грудной клетки и 225 тысяч маммографических исследований. Необходимо отметить, что реальный поток профилактических исследований в ЕРИС значительно выше, однако в рамках настоящей диссертации не было цели охватить его в 100% объеме. Несмотря на такое большое количество исследований, ПО на основе ТИИ удовлетворительно справлялись, обеспечивая непрерывное снижение удельного веса технологических дефектов, связанных с превышением времени обработки исследования. За три года изучения данного процесса наблюдалось ускорение обработки профилактических исследований, причем большая часть времени затрачивалась на пересылку данных, а не на их фактическую обработку ПО на основе ТИИ.

ЕРИС является самой масштабной РИС в России и входит в тройку самых масштабных РИС в мире. По этой причине масштабирование данной схемы интеграции на меньшие РИС не вызовут технических проблем.

Оптимизированная для профилактических исследований методика внедрения ПО на основе ТИИ применима для любых других направлений исследований. За время исследования согласно аналогичной методике были успешно внедрены ПО на основе ТИИ по другим направлениям и видам исследований.

Предложенную методику возможно легко масштабировать на любые другие РИС/PACS-системы. В частности, разработанная методика легла в основу ИИ-

консультанта врача, развернутого на платформе <https://hub.tele-med.ai>. Посредством сервиса автоматического анализа рентгенологических исследований для врачей осуществляется обработка как отдельных исследований, так и всего потока профилактических исследований из медицинских организаций Ямало-Ненецкого автономного округа. Принцип работы hub.tele-med.ai заключается в направлении исследований на специально созданную платформу, которая, по сути, моделирует работу PACS-системы. На стороне hub.tele-med.ai осуществляется анонимизация исследований и направление подлежащих обработке серий на ПО на основе ТИИ, расположенные в облачных ресурсах. ПО на основе ТИИ получает исследование в формате DICOM, обрабатывает его и направляет два файла: DICOM SC с графической разметкой исследований и DICOM SR с текстовым заключением об исследовании. Эти файлы на стороне hub.tele-med.ai добавляются в оригинальное исследование и маршрутизируются обратно в РИС МО.

Заключение по главе 3

В настоящей главе были решены две важные задачи: 1) произведен анализ организации и проведения массовых профилактических исследований, показано в теории, что применение ТИИ для анализа лучевых исследований в теории является действенным способом снижения рисков и оптимизации процессов управления массовыми профилактическими исследованиями; 2) на основании результатов двух масштабных исследований по оценке диагностической точности врачей-рентгенологов при описании профилактических исследований были определены критерии для оценки целесообразности внедрения разработок на основе ИИ в клиническую практику. Для этого на этих же наборах данных протестированы ПО на основе ТИИ, а их результаты сопоставлены между собой. В итоге были получены минимальные значения показателей диагностической точности для допуска ПО на основе ТИИ

к практической эксплуатации, а также были продемонстрированы сопоставимые результаты по отдельным показателям между врачом и ПО на основе ТИИ. ПО на основе ТИИ отличается от врача тем, что может быть доработано с точки зрения тонкой настройки на целевую чувствительность или специфичность. Именно на возможности тонкой настройки базируется модель применения ПО на основе ТИИ для профилактических исследований, которая будет рассмотрена в главе 5.

В настоящем разделе также была описана схема подключения ПО на основе ТИИ к PACS-системе, позволяющая предоставлять результаты врачу-рентгенологу в той системе, с которой он привык работать. Также были детально охарактеризованы все бизнес-процессы с ограничениями и требованиями и представлены собственные результаты, демонстрирующие высокую скорость работы предлагаемой системы внедрения ТИИ. Масштабирование методики внедрения ПО на основе ТИИ в практику рентгеновских отделений, разработанной для городской системы здравоохранения, на федеральный уровень продемонстрировало ее результативность. Однако реализация всей описанной выше методики невозможна без методов и инструментов контроля качества внедрения и работы ПО на основе ТИИ, которые описаны в следующей главе.

Глава 4. СОЗДАНИЕ И ОЦЕНКА РЕЗУЛЬТАТИВНОСТИ СИСТЕМЫ КОНТРОЛЯ КАЧЕСТВА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОФИЛАКТИЧЕСКИХ ИССЛЕДОВАНИЙ

4.1. Методика многоступенчатой системы контроля качества программного обеспечения на основе технологий искусственного интеллекта

Большое количество рассматриваемых в настоящей работе программных решений на основе ТИИ обуславливает необходимость проведения объективного сравнения данных решений между собой. С этой целью осуществлены разработка, внедрение и апробация методики тестирования ПО на основе ТИИ для предварительного отбора и периодического контроля качества работы ИИ.

Разработанная методика представлена на рисунке 18 с указанием характеристики и цели каждого из этапов, а также ответственных участников.

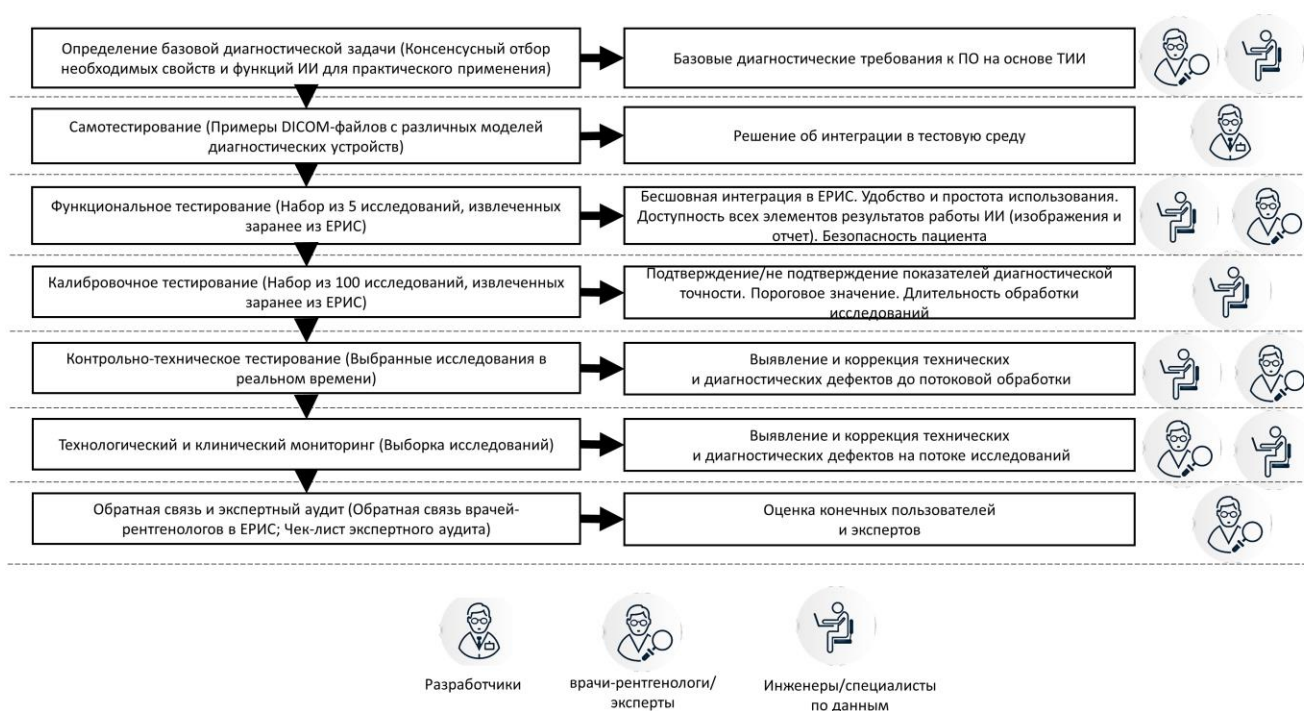


Рисунок 18 – Предлагаемые этапы методики отбора и оценки ИИ-решений для лучевой диагностики при их внедрении в рутинную практику. Слева представлен перечень этапов, справа результат каждого из этапа, пиктограммой показаны участники каждого из этапов

«Этапы функционального и калибровочного тестирований ИИ проводились в тестовой среде (ТПАК), имитирующей работу ЕРИС. Контрольно-техническое тестирование, мониторинг, сбор обратной связи и клинический аудит проводились в ходе валидации ИИ на реальных проспективных данных в ЕРИС» [261, стр. 169–170].

Подготовительный этап для отбора и последующего внедрения искусственного интеллекта

На этом этапе основной фокус был направлен на формализацию ключевой диагностической задачи для ИИ, опираясь на анализ научных публикаций. Затем сформированный консилиум из врачей – экспертов в предметной области обсуждал клинические задачи, предполагаемые к решению с помощью ИИ. Клинические задачи, которые предполагалось решать ПО на основе ТИИ, фиксировались при формировании базовых диагностических требований (БДТ). Таким образом предполагалось сокращение временных затрат при описании исследований, а также повышение точности выявления патологических изменений для массовых диагностических и профилактических исследований.

Базовые диагностические требования для флюорографических и рентгенологических исследований органов грудной клетки. Базовые диагностические требования для флюорографических и рентгенологических исследований органов грудной клетки не отличаются, т.к. в перечень ДУ, подключенных к ПО на основе ТИИ по видам исследования (ФЛГ и РГ ОГК), ограничен цифровыми диагностическими устройствами. При разработке настоящей версии БДТ использовались такие литературные источники, как «Лучевая диагностика органов грудной клетки: национальное руководство» и глоссарий Флейшнеровского сообщества. Первая версия БДТ, разработанная в 2020 году, представлена в Приложении А на рисунке 1.

БДТ по РГ и ФЛГ основывались на экспертном мнении врачей-рентгенологов и содержали в качестве клинической задачи определение одной из семи нозологий, имеющих клиническое значение. Позднее, после разработки инструмента анализа текстовых протоколов (см. подпараграф 4.2), была проведена аналитика частоты встречаемости рентгенологических признаков патологии в заключениях целевой аудитории (в структуре ДЗМ) врачей-рентгенологов, БДТ были скорректированы в части расширения перечня патологических признаков, обязательных для обнаружения ПО на основе ТИИ. Актуальный перечень представлен в Приложении А в таблице А1.

Базовые диагностические требования для маммографии. При подготовке базовых требования для маммографии были использованы методические рекомендации по организации программы популяционного скрининга злокачественных новообразований молочной железы среди женского населения и клинические рекомендации Министерства здравоохранения Российской Федерации «Доброкачественная дисплазия молочной железы» от 2020 года. Представлены требования для двух поднаправлений маммографии: скрининговая и диагностическая, которые отличаются только шкалой. Однако в рамках настоящей работы анализировались бинарные ответы по ММГ. В процессе работы по этому направлению, базовые диагностические требования не изменялись и остаются актуальными с 2020 года. БДТ представлены в Приложении А в таблице А2.

При поддержке инженеров были сформулированы базовые функциональные требования (БФТ) к ИИ. Базовые функциональные требования содержат техническую информацию, позволяющую ПО на основе ТИИ успешно интегрироваться в ЕРИС (рекомендации к информации, передаваемой в Kafka Message), а также предоставлять результат в программной среде, которую использует врач-рентгенолог в повседневной работе в том виде, в котором

ожидает увидеть врач-рентгенолог согласно БДТ. БФТ содержат рекомендации к текстовому описанию (DICOM SR) и рекомендации к дополнительной серии изображений (DICOM SC), все эти требования были разобраны в параграфе 4.1 («методика внедрения»), поэтому здесь не будут дублироваться.

«Привлечение конечного пользователя – врача-рентгенолога – к разработке базовых требований еще на подготовительном этапе позволило отсеивать ИИ-сервисы, которые заведомо не были бы востребованными и удобными в практике» [261, стр. 171].

Этап самотестирования

Разработчикам ПО на основе ТИИ рекомендовалось самостоятельно оценить корректность работы своих разработок. Для этого были подготовлены наборы данных в формате DICOM, полученных с диагностических устройств, подключенных к ЕРИС. Разработчикам предлагалось оценить возможность работы ПО с диагностическими устройствами, соответствующими целевому парку оборудования. При этом процесс самотестирования происходил без какого-либо контроля со стороны заказчика, полностью возлагая ответственность за проверку технической совместимости на разработчиков. Возможность совместимости ПО на основе ТИИ и имеющихся данных позволяла приступить к дальнейшим этапам оценки ПО на основе ТИИ и его интеграции в среду для тестирования, копирующего функционал и интерфейс ЕРИС.

Для проведения самотестирования был подготовлен ряд наборов данных, которые размещены в свободном доступе на сайте mosmed.ai в разделе проведения селф-теста технического «MosMedData PG с признаками патологий ОГК, тип II».

Результаты данного этапа находились исключительно на стороне разработчика ПО на основе ТИИ и освещаться в настоящей работе не будут.

Этап функционального тестирования

После интеграции ИИ в тестовую среду проводилось функциональное тестирование как с технической, так и с клинической точек зрения. Техническая часть отвечает за корректность отображения всех требуемых полей в АРМ врача-рентгенолога, включая точность результатов работы ПО на основе ТИИ. Клиническая часть проводится врачами-экспертами для оценки грубых нарушений в работе ПО на основе ТИИ (например, пропуск явных патологических находок).

Данный вид тестирования направлен на оценку удобства, доступности и полноты функционала ПО на основе ТИИ с пользовательской точки зрения и корректность его интеграции. Функционал ПО анализировался по трем аспектам: триаж и маркировка исследований в рабочем списке, выделение патологических находок и автоматическое заполнение структурированных протоколов DICOM SR. Тестирование включало проверку наличия корректной маркировки явных патологических изменений. Критически важно было обеспечение корректной визуализации изображений, наличия предупреждений о возможных искажениях, а также сохранение оригинальных серий.

Для проведения функционального тестирования был подготовлен ряд наборов данных, информация о которых размещена в разделе «Наборы данных для проведения функционального тестирования» на сайте mosmed.ai.

По мере готовности ПО на основе ТИИ к проведению функционального тестирования на сервис маршрутизируются пять исследований из набора данных для функционального тестирования, содержащих нормальные и патологические случаи. После обработки ПО данного набора специалистами функционального заказчика осуществляется оценка функционала ПО и информирование компании-поставщика ПО на основе ТИИ о результатах. В случае отсутствия значимых замечаний к работе ПО последний может быть допущен к проведению калибровочного тестирования.

По результатам проверки и оценки формируются выводы и заключение.

Результатом функционального тестирования являлся протокол с оценкой полноты реализованного функционала ПО на основе ТИИ для его использования врачами-рентгенологами в ЕРИС ЕМИАС. Протокол может содержать критические и некритические несоответствия, а также незначительные замечания. Примеры замечаний по результатам функционального тестирования ПО на основе ТИИ изложены в таблице 13.

Таблица 13 – Виды замечаний при проведении функционального тестирования ПО на основе ТИИ

Критические несоответствия	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 ПО на основе ТИИ от базовых функциональных требований
Заключения по одному и тому же исследованию противоположные по смыслу	
Отсутствует разметка от ПО на основе ТИИ при указании патологии в заключении. Присутствуют находки вне целевого органа	
Некритические несоответствия	Не влияют на безопасность пациента, но влияют на работу врача. Недоработки функционала, которые отклоняются от общепринятых норм в деятельности врача-рентгенолога
Некорректное отображение подписей в дополнительной серии	
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
Неточные формулировки в DICOM SR	

При проведении функционального тестирования всех ПО на основе ТИИ осуществлялась фиксация несоответствий заявленного функционала, которые классифицировались на значимые и незначимые. Значимые замечания были обязательны для устранения, т.к. не позволяли получить корректный результат работы ПО на основе ТИИ для врача-рентгенолога или затрудняли его работу. Напротив, незначимые не были обязательными к устранению, а носили лишь

рекомендательный характер, они позволяли повысить восприятие результата работы ПО на основе ТИИ. Максимально потребовалось 3 итерации по доработке ПО для устранения всех выявленных критических замечаний. Таким образом, было получено подтверждение факта устранения всех выявленных значимых замечаний при проведении повторного функционального тестирования или на следующем этапе – при проведении выборочного контроля качества исследований из набора данных для калибровочного тестирования.

Если находки расположены вне исследуемой анатомической области (рисунок 19), то они могут ввести пользователя в заблуждение относительно наличия целевых находок в исследовании. Такие дефекты не всегда могут быть выявлены немедицинским персоналом, поэтому требуется привлечение врачей-рентгенологов. В представленном примере при обработке рентгенографии грудной клетки ПО на основе ТИИ произвело анализ за пределами грудной полости и осуществило некорректную маркировку содержимого петель кишки и суммации контуров плечевой кости и клювовидного отростка лопатки. Указанные изменения не входят в перечень целевой патологии, определяемой этим ПО на основе ТИИ. Отдельные грубые нарушения в работе ПО могут быть зафиксированы инженерным персоналом, проводящем оценку функционала ПО на основе ТИИ и привлечение клинических экспертов не требуется (рисунок 20). В данном исследовании присутствуют дефекты, обусловленные выраженным изменением яркости и контрастности изображения, некорректной маркировкой находок и наложением контуров маркировки.

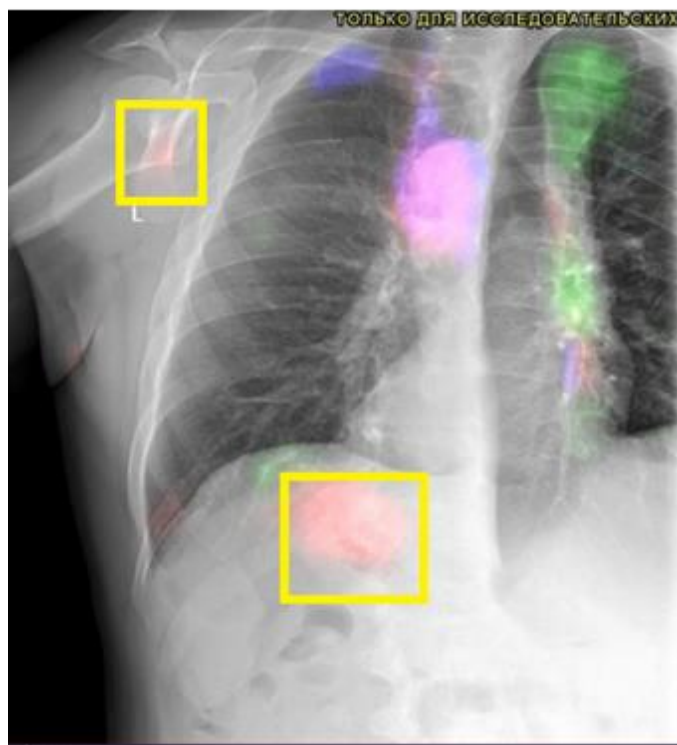
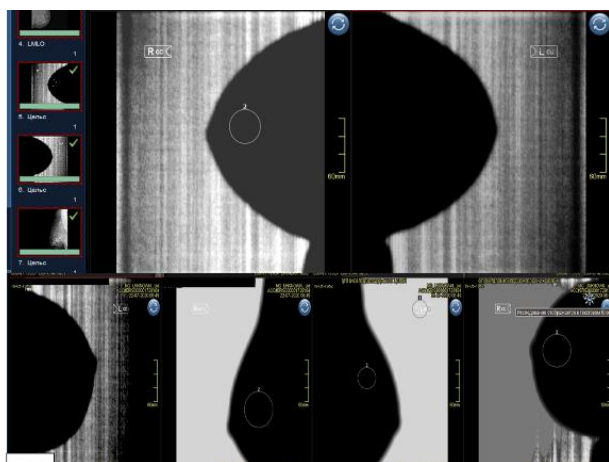
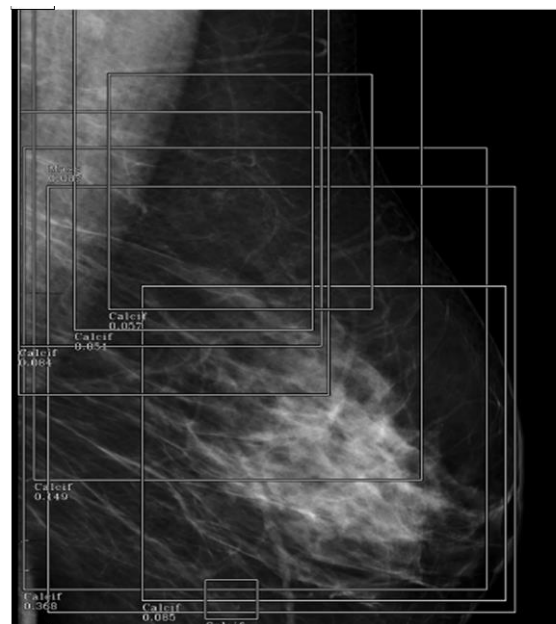


Рисунок 19 – Грубые дефекты в работе ИИ-решений – находки за пределами целевого органа



а



б

Рисунок 20 – Примеры технологических дефектов при обработке маммографического исследования: а – изменение яркости и контрастности изображения; б – чрезмерное количество элементов графической разметки

В ходе функционального тестирования проводится оценка структуры текстового отчета (DICOM SR) для ПО на основе ТИИ на соответствие полноте и порядку, предусмотренным в БФТ; отчет обязательно содержит: сводную информацию о ПО на основе ТИИ; цель или клиническую задачу, решаемую данным ПО; краткое руководство пользователя; описание и заключение, текст которых может быть заимствован врачом для собственного протокола.

Если реализованные функциональные элементы не соответствуют БФТ, то рекомендуется доработать ПО на основе ТИИ. В ходе проведения функционального тестирования нам удалось выявить существенные функциональные ограничения для тестируемого ПО, предоставить обратную связь и детализированные протоколы для разработчиков этого ПО, благодаря чему удалось существенно повысить функциональное качество ПО на основе ТИИ. В случае если реализованные функциональные элементы полностью соответствуют БФТ, то ПО на основе ТИИ допускается к проведению калибровочного тестирования.

Этап калибровочного тестирования

Этап калибровочного тестирования предназначен для оценки прогнозируемой диагностической точности ПО на основе ТИИ на ретроспективном сбалансированном наборе данных из 100 исследований с равным соотношением случаев с патологией и без. Подробное обоснование выбранного объема выборки изложено в параграфе 2.6 «Обоснование объема выборки». Для каждой диагностической задачи были подготовлены специфические наборы данных, что обеспечивало возможность предсказания точности ПО на основе ТИИ для общей популяции в пределах доверительных интервалов. Данные для каждого набора извлекались из ЕРИС, а разметка патологических находок выполнялась врачами-рентгенологами экспертного уровня на основе консенсуса. При этом учитывались данные из электронной

медицинской карты, включая подтверждение морфологическими и другими более точными диагностическими методами или результатами последующих исследований.

Набор данных, размещенный в отдельном контуре для проведения тестирований, направляется ПО на основе ТИИ. Отправка каждого исследования осуществляется последовательно без повторений до тех пор, пока не будут переданы все исследования. ПО на основе ТИИ, приняв исследование, осуществляет его обработку, и возвращает результаты работы в виде вероятности патологии, а также дополнительных серий, содержащих текстовую и графическую информацию. Для расчета показателей диагностической точности используется вероятность наличия патологии, а также эталонное значение разметки (GT). При этом ПО на основе ТИИ получает только отдельные исследования, но не файл с разметкой.

Результатом калибровочного тестирования является характеристическая кривая (ROC-кривая). При анализе ROC-кривой необходимо провести определение оптимальных значений порога активации (так называемого cut-off). В качестве эталонных значений используются показатели, заявленные разработчиком ПО на основе ТИИ и задокументированными в предварительных клинико-технических испытаниях. Для расчета оптимального значения порогового значения используется индекс Юдена (Youden index). Для работы ПО на основе ТИИ на потоке исследований могут применяться пороговые значения, рассчитанные с целью максимизации NPV или PPV. Приоритет той или иной методике определения cut-off должен отдаваться в зависимости от клинической задачи: повышение чувствительности, специфичности или достижение их оптимального соотношения.

При помощи разработанного и описанного выше инструмента выполнения калибровочного тестирования осуществляются автоматический расчет показателей качества ПО на основе ТИИ и формирование протокола калибровочного тестирования. Более подробно работа с платформой описана

в инструкции пользователя по эксплуатации веб-инструмента калибровки ПО на основе ТИИ для работы в ЕРИС ЕМИАС. Пример протокола, формируемого данной платформой, показан на рисунке 21.

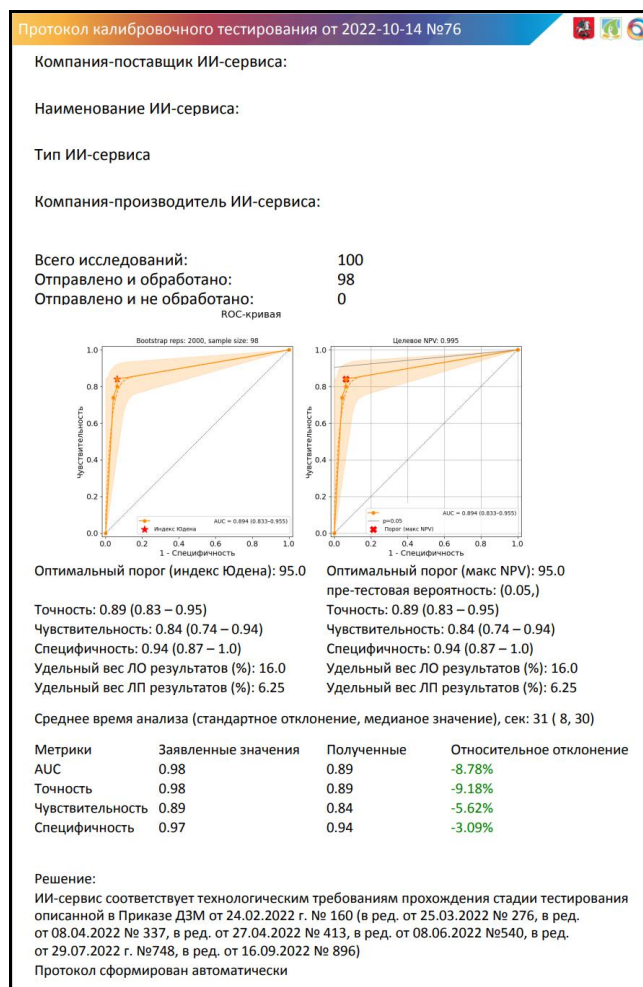


Рисунок 21 – Пример автоматизированного отчета, формируемого по результатам калибровочного тестирования

В указанном документе отражается информация о ПО на основе ТИИ и тестировании, полученных в ходе калибровочного тестирования значениях ключевых диагностических показателей, а также сопоставление с заявленными разработчиком. По результатам проверки и оценки формируются выводы и рекомендации. Также данные об оптимальном значении порога срабатывания cut-off ПО на основе ТИИ передаются в техподдержку ЕРИС, который вносит это в соответствующий справочник.

В ходе калибровочного тестирования также проводится дополнительная ручная проверка функционала ПО на основе ТИИ: оценка ложноотрицательных и ложноположительных срабатываний ПО на основе ТИИ, а также оценка корректности локализации патологических изменений в исследованиях с ИП срабатыванием.

На рисунках 22–25 приведены примеры ЛО и ЛП-срабатываний ПО на основе ТИИ по каждому из рассматриваемых видов исследования.

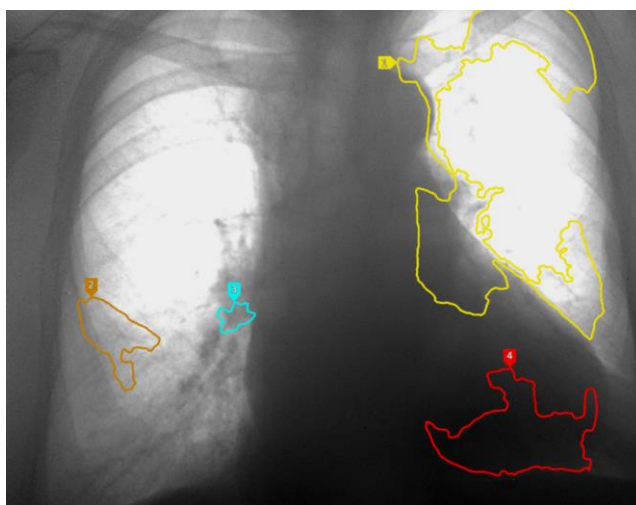


Рисунок 22 – Пример ЛП-срабатывания. Паталогические изменения в легких отсутствуют

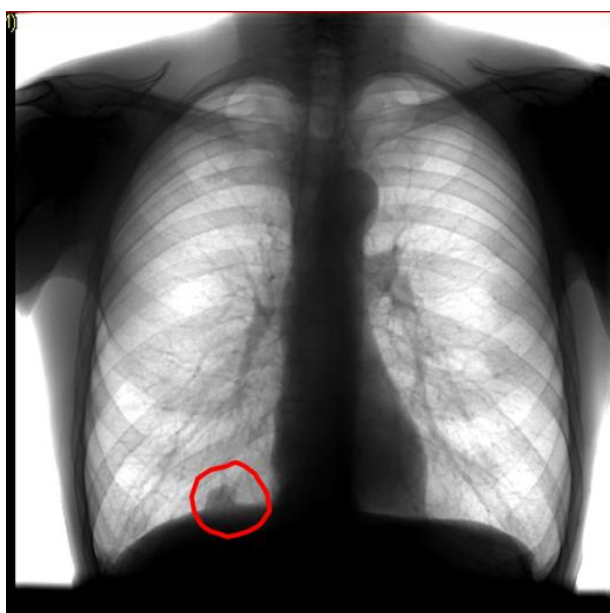


Рисунок 23 – Пример ЛО-срабатывания. ПО на основе ТИИ не локализовало образование нижней доли правого легкого

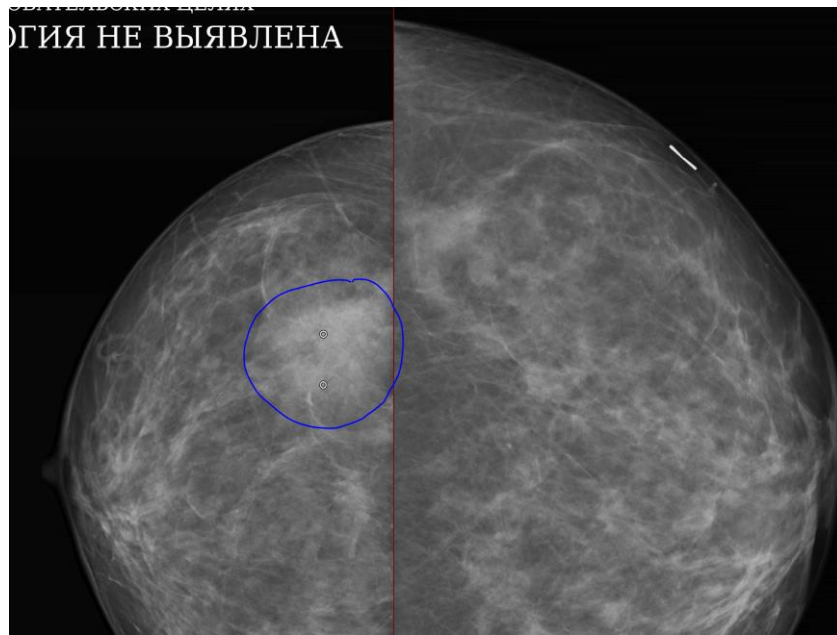


Рисунок 24 – Пример ЛО-срабатывания. ПО на основе ТИИ не локализовало подозрительный участок с целевой патологией (синий контур)



Рисунок 25 – Пример ЛП-срабатывания. ПО на основе ТИИ локализовало сосок в качестве целевой патологии

Анализ выполняется на основании отчета в формате csv(xls), где для каждого исследования указаны следующие данные: уникальный идентификатор

исследования, временные отметки начала и окончания загрузки и обработки исследования, вероятность наличия патологии в исследовании.

Тестирование осуществляли технические специалисты и клинические эксперты. Использовался ROC-анализ для оценки диагностической точности и настройки порогового значения по максимуму индекса Юдена. Также измерялась скорость обработки данных в условиях, близких к реальным. Диагностическая точность (AUROC) выше 0,81 свидетельствовала о возможности дальнейшего внедрения и апробации ПО, ПО со значениями AUC ниже 0,6 считалось непригодным, а результаты AUC в диапазоне 0,61–0,8 указывали на необходимость доработки ПО и повторного тестирования.

Для проведения калибровочных тестирований был подготовлен ряд наборов данных, информация о которых размещена в разделе с калибровочными тестированиями на сайте mosmed.ai

Протокол калибровочного тестирования может содержать:

- критические несоответствия (существенные отклонения от методических рекомендаций и заявленных параметров компанией-поставщиком);
- несоответствия рекомендательного характера (пограничные значения времени, затрачиваемого на принятие исследования, его обработку и передачу результатов анализа по данным ЕРИС).

Результаты проведения калибровочных тестирований представлены в таблице 14.

Таблица 14 – Результаты калибровочных тестирований ПО на основе ТИИ

Вид исследования	ПО на основе ТИИ	AUC	Чувствительность (Юден)	Специфичность (Юден)	Чувствительность (NPV)	Специфичность (NPV)	Дата	AUC (заявлено)	Чувствительность (заявлено)	Специфичность (заявлено)
РГ	ИИ-07	0,91	0,81	0,92	0,93	0,58	08.05.2020	0,93	0,86	0,92
РГ/ФЛГ	ИИ-07	0,95	0,88	0,93	0,92	0,83	26.07.2022			
ФЛГ	ИИ-07	0,81	0,61	0,86	1,00	0,06	21.12.2022			
РГ/ФЛГ	ИИ-07	0,83	0,78	0,78	0,96	0,43	17.01.2022			
ФЛГ	ИИ-02	0,90	0,80	0,92	0,94	0,66	12.11.2020	0,97	0,92	0,94
ФЛГ	ИИ-02	0,95	0,92	0,92	0,92	0,92	23.02.2021	0,90	0,92	0,63
ФЛГ	ИИ-02	0,77	0,74	0,76	0,91	0,44	02.06.2020			
РГ/ФЛГ	ИИ-02	0,89	0,88	0,87	0,88	0,87	16.06.2021			
РГ/ФЛГ	ИИ-02	0,97	0,92	0,94	0,92	0,94	03.02.2022			
РГ	ИИ-02	0,95	0,92	0,92	0,92	0,92	25.10.2021			
РГ	ИИ-02	0,93	0,92	0,86	0,94	0,78	12.10.2021			
РГ	ИИ-02	0,80	0,90	0,69	0,94	0,38	02.06.2020			
РГ	ИИ-02	0,83	0,87	0,78	1,00	0,12	04.04.2022			
РГ	ИИ-06	0,74	0,40	0,96	0,98	0,29	19.10.2020	0,90	0,92	0,63
РГ	ИИ-06	0,86	0,83	0,88	0,96	0,35	17.11.2020			
РГ	ИИ-06	0,71	0,56	0,82	1,00	0,06	08.09.2021			
РГ	ИИ-06	0,72	0,54	0,88	1,00	0,06	08.09.2021			
РГ/ФЛГ	ИИ-06	0,69	0,42	0,90	1,00	0,05	06.09.2021			
ФЛГ	ИИ-06	0,78	0,82	0,67	1,00	0,02	23.10.2020			
ФЛГ	ИИ-06	0,76	0,70	0,76	0,98	0,02	11.11.2020	0,98	0,96	0,93
РГ	ИИ-10	0,67	0,93	0,33	0,98	0,17	17.01.2021			
ММГ	ИИ-10	0,82	0,56	0,98	0,84	0,62	25.02.2021	0,93	0,88	0,89
ММГ	ИИ-04	0,96	0,84	0,94	0,92	0,82	22.10.2020	0,90	0,83	0,95
РГ	ИИ-04	0,91	0,86	0,85	1,00	0,21	23.10.2020	0,92	0,79	0,95
ММГ	ИИ-09	0,75	0,76	0,73	0,88	0,24	30.12.2021	0,88	0,94	0,72
РГ	ИИ-05	0,94	0,91	0,98	0,91	0,98	18.11.2021	0,92	0,90	0,82
ММГ	ИИ-03	0,89	0,74	0,94	0,98	0,28	14.07.2021	0,88	0,83	0,83
ММГ	ИИ-03	0,85	0,88	0,76	1,00	0,04	06.06.2022			
ММГ	ИИ-03	0,86	0,88	0,78	1,00	0,04	09.07.2022			
ММГ	ИИ-03	0,86	0,76	0,84	0,76	0,84	09.07.2022			
ММГ	ИИ-03	0,89	0,74	0,94	1,00	0,04	29.09.2021			
ФЛГ	ИИ-08	0,59	0,73	0,44	0,73	0,44	27.05.2020	0,92	0,90	0,81

Продолжение таблицы 14

Вид исследования	ПО на основе ТИИ	AUC	Чувствительность (Юден)	Специфичность (Юден)	Чувствительность (NPV)	Специфичность (NPV)	Дата	AUC (заявлено)	Чувствительность (заявлено)	Специфичность (заявлено)
ФЛГ	ИИ-08	0,85	0,68	0,94	1,00	0,02	30.12.2020			
РГ/ФЛГ	ИИ-08	0,91	0,90	0,81	0,93	0,71	09.06.2021	0,97	0,87	0,96
РГ	ИИ-08	0,60	0,98	0,23	0,98	0,23	27.05.2020			
РГ	ИИ-08	0,92	0,86	0,86	0,94	0,62	04.08.2020			
ММГ	ИИ-01	0,94	0,92	0,78	0,94	0,72	27.04.2022	0,89	0,82	0,82
ММГ	ИИ-01	0,81	0,68	0,84	0,87	0,45	26.06.2020			
ММГ	ИИ-01	0,73	0,62	0,84	0,96	0,03	26.05.2020			
ММГ	ИИ-01	0,91	0,84	0,86	0,98	0,94	06.09.2021			
РГ/ФЛГ	ИИ-01	0,95	0,88	0,93	0,92	0,83	26.07.2022	0,91	0,80	0,92
РГ/ФЛГ	ИИ-01	0,93	0,85	0,89	0,97	0,38	01.03.2022			
РГ/ФЛГ	ИИ-01	0,91	0,80	0,92	0,97	0,93	06.09.2021			
РГ/ФЛГ	ИИ-01	0,91	0,79	0,92	0,97	0,93	06.09.2021			
ФЛГ	ИИ-01	0,92	0,72	0,90	0,91	0,30	17.06.2020			

Контрольно-техническое тестирование (КТТ) и мониторинг

КТТ и мониторинг технологических дефектов гарантировали стабильность и качество работы ПО на основе ТИИ с разными диагностическими устройствами. КТТ проводилось при первом подключении ИИ к диагностическим устройствам и при обновлениях ПО с целью оценки стабильности работы. Для этих целей формировалась ограниченная выборка исследований – обычно два исследования с каждого из диагностических устройств. Дефекты в работе ПО фиксировались в реальном времени для успешной интеграции и корректного функционирования в ЕРИС. Аналитика дефектов осуществлялась периодически (еженедельно или ежемесячно в зависимости от качества работы ПО), результаты анализа заносятся в протокол, который направлялся разработчикам.

Один из частых дефектов в работе ПО на основе ТИИ был связан с неправильным заполнением тегов DICOM. Эти ошибки никак не влияют на

работу врача-рентгенолога, но оказались критическими для работы ПО. Так, например, при обработке РГ многие ПО на основе ТИИ определяют проекцию снимка исключительно по DICOM-тегам. С учетом регулярного обнаружения некорректно заполненных тегов проекции в метаданных рентгенологического исследования, были составлены алгоритмы определения проекции грудной клетки:

1. Смотрим DICOM-тег (0018, 1400) Acquisition Device Processing Descr:

a) проекция прямая, если значение равно значению из списка:

- ГРУДНАЯ КЛЕТКА, Пр стоя
- ГРУДНАЯ КЛЕТКА, ПЕРЕДНЯЯ Р->А
- postero-anterior^Factory1

b) проекция боковая, если значение равно значению из списка:

- ГРУДНАЯ КЛЕТКА, БОК стоя
- ГРУДНАЯ КЛЕТКА, БОКОВАЯ
- right-lateral^Factory1

2. Если нет DICOM-тега из пункта 1 или нет значения из списка, то смотрим DICOM-тег (0018, 5101) View Position.

c) проекция прямая, если значение

- AP

3. Если нет DICOM-тега из пунктов 1 и 2 или нет значения из списков, то смотрим

DICOM-тег (0054, 0220) View Code Sequence вкладку (0008, 0104) Code Meaning

d) проекция прямая, если значение

- postero-anterior

e) проекция боковая, если значение

- lateral

4. Если нет этих тегов, то проекцию определить невозможно.

Исследования с отдельных диагностических устройств не обрабатывались. В результате проведенной аналитики было выявлено, что все эти диагностические устройства относятся к аналоговой рентгеновской аппаратуре, которая оказалась технически несовместимой с данным ПО на основе ТИИ. Таким образом, замена устаревшего аналогового оборудования на современные цифровые рентгеновские системы не только улучшит качество диагностических изображений, но и облегчит интеграцию и стабильную работу ПО на основе ТИИ.

Во время интеграции ПО на основе ТИИ в ЕРИС в режиме реального времени выявлялись технические дефекты, связанные с задержками обработки исследований и техническими ошибками в полученных от ПО результатах. Эта информация оперативно передавалась разработчикам ПО для повышения качества его работы. После подготовки отчета о КТТ запускается маршрутизация исследований, согласно запланированному потоку данных с ДУ, из группы без технологических дефектов и частично содержащих технологические дефекты.

На этом заканчивается этап тестирования и начинается этап работы ПО на основе ТИИ на потоке данных. В ходе апробации данной методики на практике автор диссертации убедился в том, что от качества прохождения данных этапов (от полноты устранения некритических замечаний и выполнения рекомендаций) зависит эффективность работы ПО на основе ТИИ на потоке. Также для прохождения описанных выше этапов ПО на основе ТИИ могут направляться в случае осуществления различных доработок.

Тестирование при смене версионности

В результате работы ПО на основе ТИИ на потоке могут быть выявлены определенные несоответствия в клинической или технической части. Разработчик ПО устраняет несоответствия путем внесения изменений в работу или самого ядра алгоритма или в обеспечивающие программные модули. С целью обеспечения контроля качества нами была разработана методика оценки

изменений, реализуемых разработчиком на основании заявки о смене версионности. В результате анализа заявки на смену версионности возможны два действия – согласование осуществления смены версионности (в случае изменений, не затрагивающих само ядро ПО на основе ТИИ) или смена версионности через обязательное калибровочное тестирование. Данное тестирование необходимо для обеспечения уровня диагностической точности ПО на основе ТИИ на уровне не хуже, чем в предыдущей версии. Для этого проводится тестирование на том же наборе данных, что и предыдущее калибровочное тестирование. С целью оптимизации проведения тестирования нами была разработана платформа тестирования при смене версионности.

Апробация платформы тестирования при смене версионности сервиса искусственного интеллекта (далее – ПТСВ) проводилась по видам исследований:

- рентгенография,
- флюорография,
- маммография,
- в соответствии с методическими рекомендациями [235].

Для тестирования использовались эталонные наборы данных объемом 100 и 200 исследований в формате DICOM, аналогичные данным для тестирования в тестовом программно-аппаратном комплексе (ТПАК):

- набор данных комбинированный «Рентгенография/флюорография» (РГ/ФЛГ) – 200 исследований (соотношение классов «с целевой патологией» и «без целевой патологии» – 100/100);
- набор данных «Маммография» (ММГ) – 100 исследований (соотношение классов «норма»/«патология» – 50/50).

Объектами тестирования выступало ПО на основе ТИИ, функционирующее в продуктивном программно-аппаратном комплексе Единого радиологического информационного сервиса Единой медицинской информационно-аналитической системы (ППАК ЕРИС ЕМИАС) согласно следующим видам исследований:

- маммография с целью диагностики рака молочной железы. Тестируемые сервисы: ИИ-01 ММГ, ИИ-03ДМ;
- рентгенография и/или флюорография легких с целью определения различных их патологий. Тестируемое ПО на основе ТИИ: ИИ-06, ИИ-01, ИИ-08.

Метод верификации полученных параметров: ROC-анализ на инструменте, используемом для проведения ROC-анализа при тестировании ПО на основе ТИИ в ТПАК с применением классического (эмпирического непараметрического) анализа, по которому происходит расчет AUC, и при контрольном исследовании, полностью совпадает с результатами из программного обеспечения «Stata» (дата контрольного исследования: 23.06.2021 г.).

Контролируемые параметры и характеристики:

- значение площади под ROC-кривой (AUC) с 95% доверительным интервалом;
- пороговые значения;
- точность с 95% доверительным интервалом;
- чувствительность с 95% доверительным интервалом;
- специфичность с 95% доверительным интервалом;
- количество успешно обработанных исследований;
- среднее время анализа одного исследования.

В ходе апробации оценивались качества, позволяющие осуществить полное, автоматизированное проведение калибровочного тестирования с различными сервисами по различным направлениям без необходимости использования ТПАК.

В рамках апробации ПТСВ осуществили 5 калибровочных тестирований, в результате чего было обработано 800 исследований.

В ходе апробации значения основных показателей диагностической точности, полученные на ПТСВ, полностью совпадают со значениями верифицированного инструмента для ROC-анализа диагностических тестов, при использовании которого осуществлялось формирование протокола калибровочного тестирования в ТПАК.

По результатам апробации можно сделать вывод, что ПТСВ:

- доступна и проста в использовании (все оцениваемое ПО смогли пройти тестирование на ПТСВ);
- позволяет минимизировать работу оператора/инженера и общую длительность тестирования путем формирования автоматического отчета и его немедленного направления ПО на основе ТИИ (протокол готовится в автоматическом режиме в формате *.pdf);
- обладает достаточным уровнем безопасности информации благодаря двойной анонимизации (двойная анонимизация при повторном калибровочном тестировании не позволяет ПО на основе ТИИ осуществить подборку значений вероятности патологии).

Также в ходе апробации были выявлены возможности, которые следует реализовать в последующих версиях ПТСВ:

- прохождение тестирований несколькими сервисами одновременно;
- разработка интерфейса для передачи дополнительных данных (например, маски патологических изменений) или предупреждений для пользователей ПТСВ.

В целом апробация ПТСВ показала ее стабильную работу, возможность и эффективность практического применения по всем профилактическим направлениям исследований и достоверность в оценке показателей диагностической точности, эквивалентную приводящемуся в ТПАК калибровочному тестированию, что позволяет рекомендовать ее для дальнейшего применения в целях оценки показателей диагностической точности ПО на основе ТИИ на эталонных наборах данных для лучевых методов исследования.

В этом параграфе была проанализирована методика многоступенчатого тестирования и оценки качества ПО на основе ТИИ, состоящая из ряда этапов: самотестирования; функционального, калибровочного и контрольно-технического тестирований; технологического мониторинга. Приведены типы возникающих дефектов при тестировании и работе на потоке ПО на основе ТИИ.

Представленная система контроля качества, включающая все вышеописанные этапы, легла в основу методики сравнения решений между собой, а также стала основой жизненного цикла ПО на основе ТИИ, который будет детально разобран в параграфе 4.4.

4.2. Разработка инструментов для осуществления контроля качества работы программного обеспечения на основе технологий искусственного интеллекта

Для осуществления настоящего исследования потребовались разработка, апробация и внедрение ряда инструментов контроля качества работы ПО на основе ТИИ: инструмент анализа текстовых протоколов, инструмент ROC-анализа, инструмент проведения калибровочных тестирований и платформа проведения мониторинга работы ПО на основе ТИИ, а также разработка наборов данных для проведения тестирования ПО на основе ТИИ.

Инструмент анализа рентгенологических текстовых протоколов

Настоящий инструмент необходим для автоматизированного анализа текстовых протоколов с целью классификации на исследования «с патологией» и «без патологии». Текстовые протоколы включали описания ММГ, ФЛГ и РГ ОГК из медицинских организаций, подведомственных Департаменту здравоохранения г. Москвы. Все протоколы датировались 2020 годом. Текстовые документы были деперсонализированы до их использования в исследовании. Основная задача заключалась в разработке алгоритма для автоматической обработки и анализа текстовых данных на предмет выявления описаний характерных рентгенологических признаков. Основой для разработки словаря рентгенологических паттернов были базовые диагностические требования, представленные в параграфе 4.1. «Для рентгенографии и флюорографии ОГК целевые патологические признаки включали: плевральный выпот, пневмоторакс,

ателектаз, очаг, инфильтрацию/консолидацию, диссеминацию, полость с распадом или уровнем жидкости, кальцинат и нарушение целостности коркового слоя (перелом). Для маммографии использовалась система анализа и протоколирования результатов лучевых исследований молочной железы BI-RADS 3–6 (Breast Imaging Reporting and Data System) [262]» [263, стр. 377].

Программные и технические решения были разработаны для предварительной разметки текстовых протоколов каждого вида исследования, используя методы обработки естественного языка в сочетании с экспертными оценками врачей-рентгенологов. Разработка алгоритмов предварительной разметки текстовых протоколов проходила в несколько этапов, начиная с определения ключевых слов для выявления определенных патологий и стоп-фраз, перевода этих признаков на машинный язык с использованием Python и соответствующих библиотек (pandas=1.1.3, numpy=1.19.2, re=2.2.1, nltk=3.5.). Этот подход требовал особого внимания и гибкости алгоритма из-за вариативности записи результатов, что подчеркивает важность NLP для извлечения информации из текстовых данных.

Отдельно необходимо выделить задачу классификации протоколов ММГ. По сравнению с направлением РГ/ФЛГ, это была относительно простая задача, которая сводилась к «парсингу» текстовых протоколов на предмет числового значения BI-RADS. Однако при первом использовании этого подхода были выявлены ложные срабатывания из-за различного написания категории BI-RADS (регистр, римские и арабские цифры, ошибки в написании и вариативность пунктуации). По этой причине алгоритм, используемый в работе, был доработан для учета различных вариантов написания и обработки ошибок. Результаты анализа протоколов маммографии приведены в таблице 15.

Таблица 15 – Частота встречаемости классов ММГ по данным анализа протоколов, обработанных ПО на основе ТИИ в 2020–2022 гг.

Класс BI-Rads	Абсолютное количество протоколов	Удельный вес
0	2591	0,33%
1	164763	21,08%
2	538276	68,87%
3	52871	6,76%
4	18988	2,43%
5	2591	0,33%
6	1541	0,20%

Разработанный инструмент позволял дифференцировать класс BI-RADS не только по всему исследованию, но также по молочным железам в отдельности. Также в данном модуле анализа текстовых протоколов был заложен перевод в бинарную классификацию «норма-патология», в том числе с учетом скрининговой шкалы.

Модуль поиска в тексте описаний патологических рентген-признаков в протоколах ФЛГ и РГ ОГК использует словарь ключевых слов для поиска. Процесс разработки модуля включает этапы от предварительной разметки протоколов до определения частоты встречаемости отдельных патологических признаков. Обучение NLP на стандартных формулировках является важным этапом в разработке; это позволяет решить задачу идентификации «нормы», однако без уникальных формулировок невозможно корректно выявлять патологические признаки. По этой причине на следующем этапе осуществляется ручная проверка точности разметки и корректировка алгоритма. Завершающий этап – добавление верифицированных (врачом) данных в базу и итеративное повторение процесса для достижения 98% точности предразметчика, в соответствии с литературными данными [264]. Этот показатель соответствует максимальным по литературным данным достижениям точности NLP при анализе текстов.

«В общей сложности количество протоколов, проанализированных врачом-рентгенологом при разработке алгоритмов-предразметчиков, составляет 977 для маммографии, 3196 – для рентгенографии, 1608 – для флюорографии. Исключению подвергались протоколы, не имевшие заполненных разделов «описание» и «заключение»» [[263], стр. 379]

Перечень РГ-признаков, которые способен выявлять алгоритм:

- Выпот – в протоколе встречаются ключи 'выпот', 'гидроторакс', 'жидкость', 'плеврит', 'эмпиема', 'гидропневмоторакс', 'пневмогидроторакс'.
- Инфильтрация – в протоколе встречаются ключи 'инфильтрац', 'инфильтрат', 'консолидац', 'фокус', 'тень'.
- Диссеминация – в протоколе встречаются ключи 'диссем', 'диссим'.
- Полость – в протоколе встречаются ключи 'полост', 'распад', 'абсцесс'.
- Кальцинат – в протоколе встречаются ключи 'кальцинат', 'оссификат', 'петрификат'.
- Эмфизема – в протоколе встречаются ключ 'эмфизем'.
- Объемные_образования – в протоколе встречаются ключ 'онколог', 'нео', 'с-г', 'сг', 'susp', 't-r', 'зно', 'mts', 'мтс', 'метастаз', 'вторичн', 'образован', 'очаг', 'тень', 'фокус'.
- Очаг – в протоколе встречаются ключи 'очаг', 'фокус', 'тень', 'затемнени'.
- Ателектаз – в протоколе встречаются ключи 'ателектаз', 'ателектат', 'коллапс', 'коллабирован'.
- Пневмоторакс – в протоколе встречаются ключи 'пневмоторакс', 'гидропневмоторакс', 'пневмогидроторакс', 'газ'.
- Консолидированный перелом – в протоколе встречаются ключи 'старый перелом', 'старые перелом', 'консолидирован', 'консолидированные перелом', 'консолидированный перелом', 'консрлидированный перелом', 'сросшиеся перелом', 'сросшийся перелом', 'застарелый перелом', 'застарелые

перелом', 'старый прелом', 'старые прелом', 'консолидированные прелом', 'консолидированный прелом', 'консолидированные перелом', 'консолидированный перелом', 'посттравматическая деформация'.

- Перелом – в протоколе встречаются ключи 'свежий перелом', 'перелом', 'прелом', 'консолидирующийся перелом', 'консолидирующий', 'неконсолидирован'.

Дополнительно был создан перечень «стоп»-слов и «стоп»-фраз: ['без клинически значимых очаговых', 'без свежих очаг', 'без видимых очаг', 'не визуализируется', 'не визуализируются' и другие].

Результатом работы инструмента анализа текстовых протоколов является заполненная таблица, состоящая из столбцов с названием рентгенологических признаков. В случае обнаружения конкретного признака в описании или в заключении в ячейке данного столбца выставляется «1», при отсутствии – ячейка не заполняется. Также в таблицу добавляется дополнительный столбец «истинное значение» (GT), который заполняется «0» в случае отсутствия признаков патологии, «1» – в случае обнаружения хотя бы одного рентгенологического признака. По этому столбцу осуществляется расчет всех показателей диагностической точности.

С использованием данного инструмента были проанализированы описания 461 407 РГ ОГК и ФЛГ, выполненных как в амбулаторном кластере, так и в стационарах во временном интервале 16.06.2020–25.12.2020. Результаты представлены на рисунке 26.

Как видно из представленных на рисунке 26 данных, доля протоколов, содержащих патологические изменения в потоке стационарных пациентов, незначительно превышает долю в амбулаторном потоке. Доминирующим признаком патологии в двух кластерах является кардиомегалия, которая встречается чаще в амбулаторном потоке. На втором месте признаки очаговых образований, которые чаще наблюдаются в стационарном кластере. В тройке лидеров по частоте встречаемости отмечается и рентгенологический признак инфильтрации, также преобладающий в стационарном кластере.

Кроме того, в результате анализа был получен список рентгенологических признаков, встречающихся в виде комбинации в протоколе. Был проанализирован общий поток исследований (результаты представлены в Приложении Б). Количество РГ/ФЛГ-исследований «без патологии»: 356 454, с патологическими признаками: 104 953 (29,4%). Удельный вес считался относительно исследований с патологическими признаками. ПРОПИСНЫМИ БУКВАМИ обозначены клинически значимые патологические признаки. Красным – РГ-признаки вне актуальных на 25.05.2021 базовых диагностических требований. Возрастные изменения не учитываются как патологический признак, если являются единственным признаком. Также отдельно амбулаторный поток и стационарный поток пациентов (Приложение Б). Для амбулаторного потока количество РГ/ФЛГ-исследований «без патологии»: 278 778, с патологическими

признаками: 80 821 (29%). Для стационарного потока количество РГ/ФЛГ-исследований «без патологии»: 77 676, с патологическими признаками: 24 132 (31%).

Инструменты проведения расчета показателей диагностической точности

В рамках настоящей работы необходимо было реализовать несколько инструментов для выполнения расчетов показателей диагностической точности. В их основе лежит ROC-анализ, для удобства пользования все эти инструменты имеют веб-интерфейс.

Инструмент интерактивного ROC-анализа. Разработан удобный и интерактивный инструмент для ROC-анализа, который имеет графический интерфейс и доступен по ссылке: <https://roc-analysis.mosmed.ai>. Данный инструмент разработан для проведения ROC-анализа результатов работы ПО на основе ТИИ, но также может быть использован для любых диагностических тестов. Его функциональные возможности включают: загрузку в веб-инструмент файла в табличном формате, в котором содержатся ответы анализируемого диагностического теста (ПО на основе ТИИ) и корректные ответы (референс-тест) по списку исследований; построение характеристической кривой; определение оптимального порога срабатывания стандартными методами (максимальный индекс Юдена и минимальное расстояние); присвоение типа ответа анализируемого диагностического теста для каждого исследования; расчет набора показателей точности для каждого порогового значения и доверительные интервалы показателей; расчет площади под характеристической кривой с доверительным интервалом. Графический интерфейс инструмента представлен на рисунке 27.

Разработанный веб-инструмент позволяет оценить подробные показатели для каждого порогового значения при наведении курсора на эту точку. Подробная информация включает: отсечку; матрицу ошибок; точность, специфичность и чувствительность с доверительным интервалом; индекс Юдена; значения PPV и NPV. Более специализированные показатели, такие как F1 и MCC для

несбалансированных наборов данных, можно рассчитать вручную из матрицы ошибок [265]. Значения NPV и PPV могут быть полезны для определения оптимальной пороговой точки для целевого или максимального PPV или NPV для задач с известным целевым превалансом, которые также могут использоваться в медицинской диагностике с учетом различных клинических задач [266].

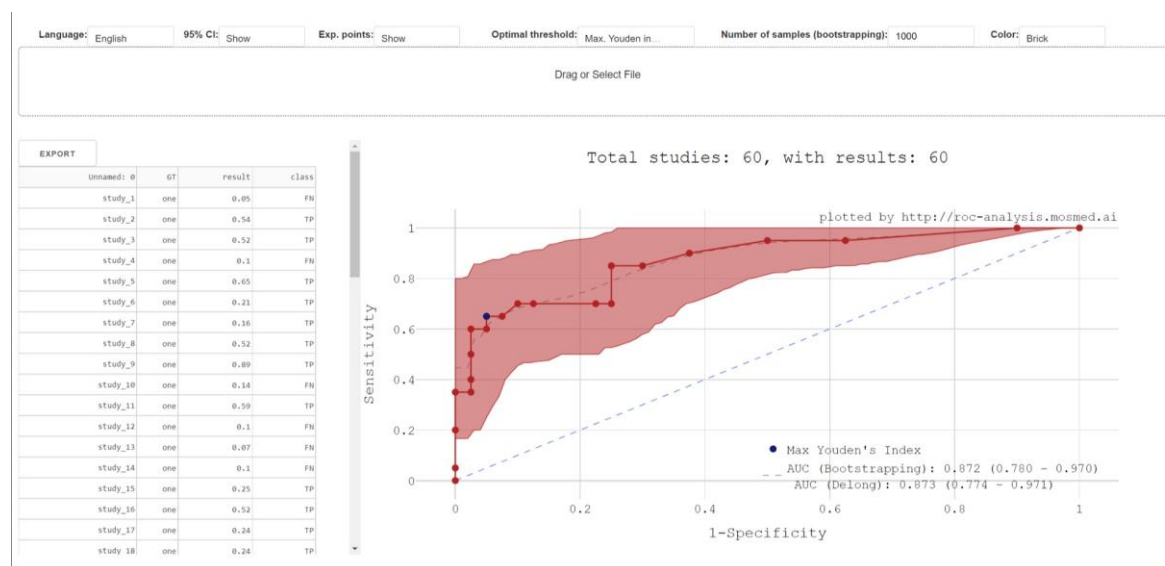


Рисунок 27 – Пример ROC-кривой, построенной с помощью разработанного веб-инструмента. Слева – таблица с определенной категорией для каждого исследования. Синей точкой отмечено оптимальное пороговое значение

Инструмент выполнения калибровочных тестирований. С целью автоматизации проведения калибровочных тестирований (в том числе при смене версии) был разработан дополнительный веб-модуль выполнения калибровочного тестирования. Инструмент «Калибровка ИИ-сервисов» (рисунок 28) позволяет на основе полученных вероятностей патологии на эталонном НД на этапе калибровки ПО на основе ТИИ выполнять расчет параметров диагностической точности, пороговых значений и времени обработки исследований.

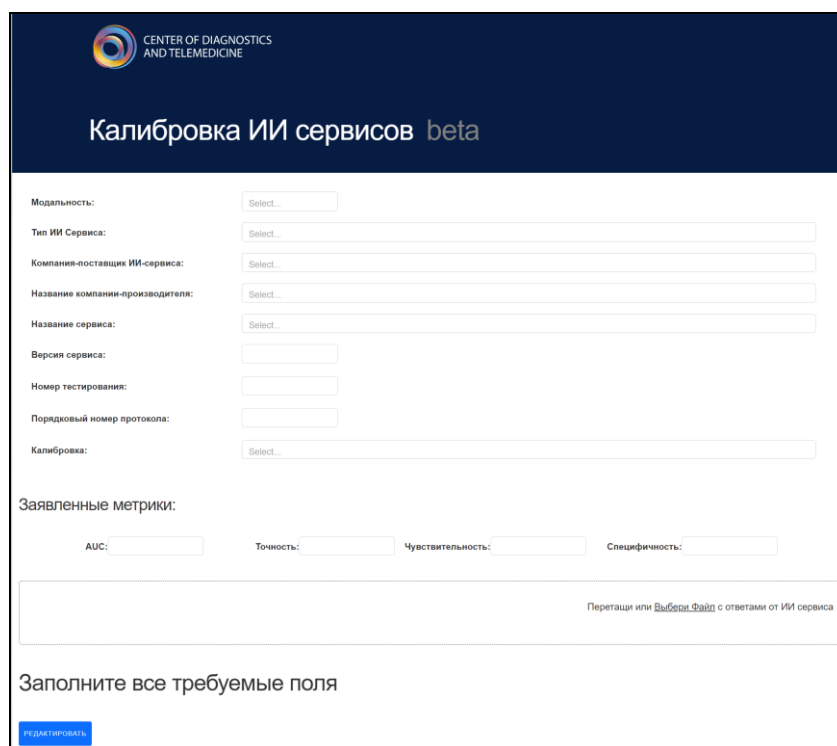


Рисунок 28 – Общий вид платформы «Калибровка ИИ-сервисов»

Инструмент сравнения характеристических кривых. С целью объективного сравнения ПО на основе ТИИ между собой была разработана модификация представленного выше ROC-анализа. Веб-инструмент для попарного сравнения результатов работы ПО на основе ТИИ является специализированным онлайн-сервисом. Для использования сервиса необходимо загрузить файл, содержащий уникальные идентификаторы исследований, соответствующие им бинарные значения наличия или отсутствия патологии, а также результаты, полученные от ИИ-алгоритма. Внешний вид инструмента изображен на рисунке 29.

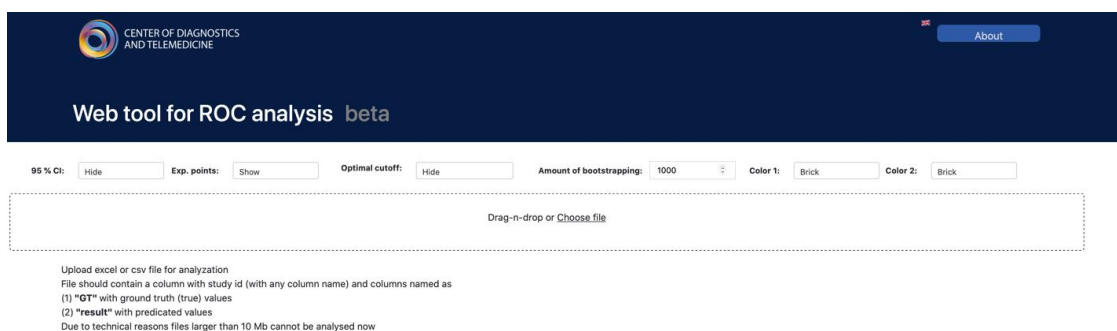


Рисунок 29 – Внешний вид инструмента для сравнения ROC-кривых

После загрузки файлов на интерфейсе отображаются их имена вместе с результатами сравнения по методу ROC AUC через перестановочный тест, включая различия в значениях ROC AUC, p-значение теста, а также графическое представление ROC-кривых, как показано на рисунке 30. Дополнительно инструмент предлагает возможности для настройки визуализации графика, включая цвет, установку оптимального порогового значения, показ экспериментальных точек и доверительных интервалов. Также доступен расчет и отображение площади под ROC-кривой с различными методами вычисления доверительных интервалов.

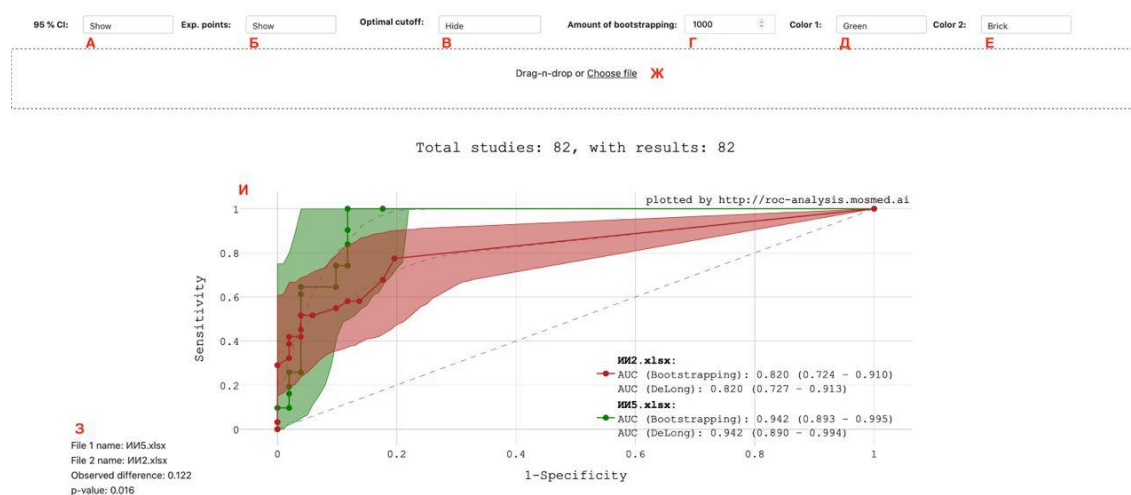


Рисунок 30 – Пример построения ROC-кривых и их сравнения с помощью перестановочного теста. А – настройка отображения доверительного интервала,

Б – отображение экспериментальных точек, В – настройка оптимального порогового значения, Г – количество повторений для бутстрэппинга, Д и Е – цвета первой и второй характеристических кривых соответственно, Ж – поле для загрузки файлов, З – названия загруженных файлов и результат перестановочного теста, И – непосредственно сами графики характеристических кривых

Платформа оценки диагностической точности врача-рентгенолога

С целью оценки показателей диагностической точности «среднего» врача-рентгенолога была разработана специализированная платформа «экспертного пересмотра» с веб-интерфейсом. Платформа основана на Osimis Viewer и адаптирована для оценки диагностической точности при описании профилактических исследований. Работа с платформой состоит из нескольких этапов:

1. Врач-рентгенолог регистрируется на платформе через форму регистрации (все данные записываются в базу данных (БД) “users”) – рисунок 31.

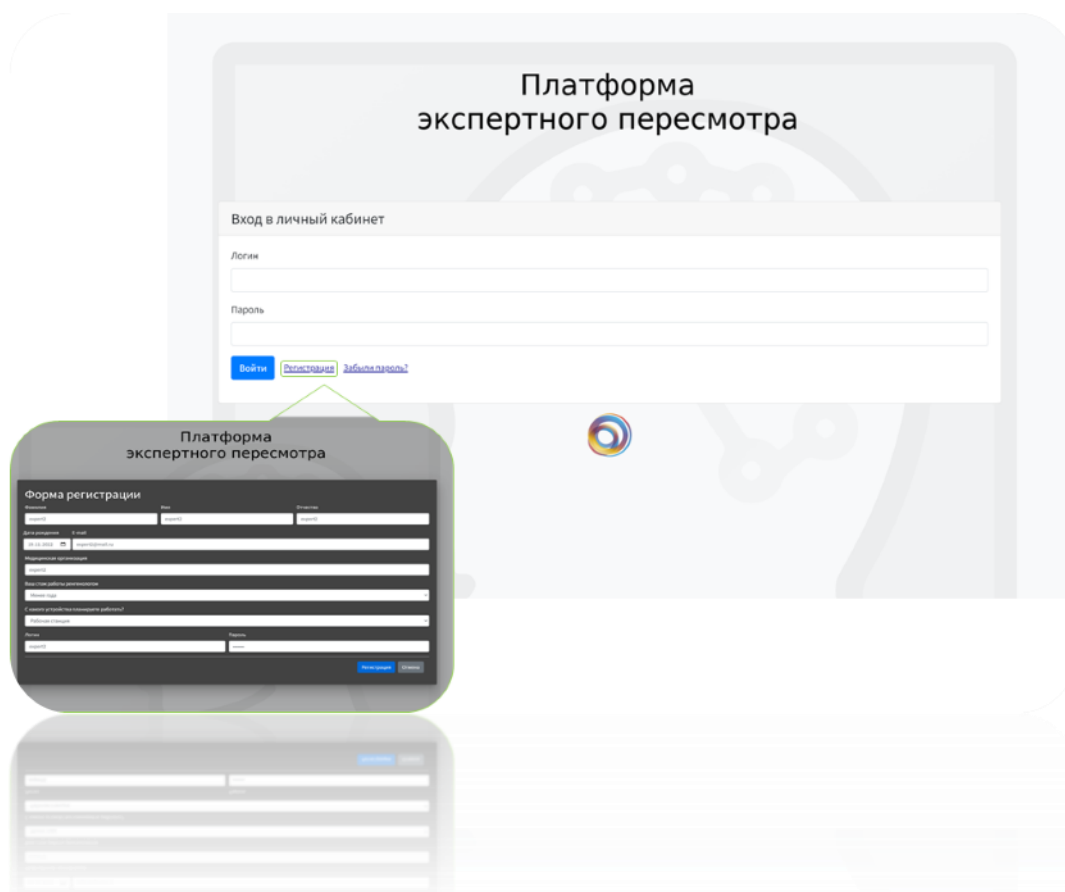


Рисунок 31 – Окно регистрации и входа на платформу оценки диагностической точности (экспертного пересмотра) врача-рентгенолога

2. После регистрации эксперт авторизуется на платформе, знакомится с инструкцией, поставленной диагностической задачей и открывает

просмотровщик с заранее подгруженными исследованиями (рисунок 32), врачу предлагается определить наличие патологии на рентгенограмме пациента из предложенного перечня вариантов ответа. (Данные об исследовании записываются в БД “usersets”.)



Рисунок 32 – Рабочее окно платформы оценки диагностической точности врача-рентгенолога: слева отображаются все доступные для анализа изображения в настоящем исследовании; в центре – просмотровщик изображения с панелью стандартных инструментов, справа – варианты ответа для врача-рентгенолога

3. После оценки экспертом всех исследований в базе данных происходит автоматический расчет параметров точности и специфичности для пользователя, а также потраченного на оценку времени. По завершении разработчик скачивает сохраненные данные и делает рассылку о результатах участникам.

Для отображения исследований в формате DICOM был использован свободно распространяемый код системы передачи и архивации DICOM-изображений Orthanc [130] и совместимый с ним веб-модуль просмотра изображений [130]. Для выгрузки данных о пользователях и разметки набора данных необходимо перейти по закрытому адресу в базу данных MONGO Express. В вкладке reader_study находится 4 базы данных (рисунок 33).

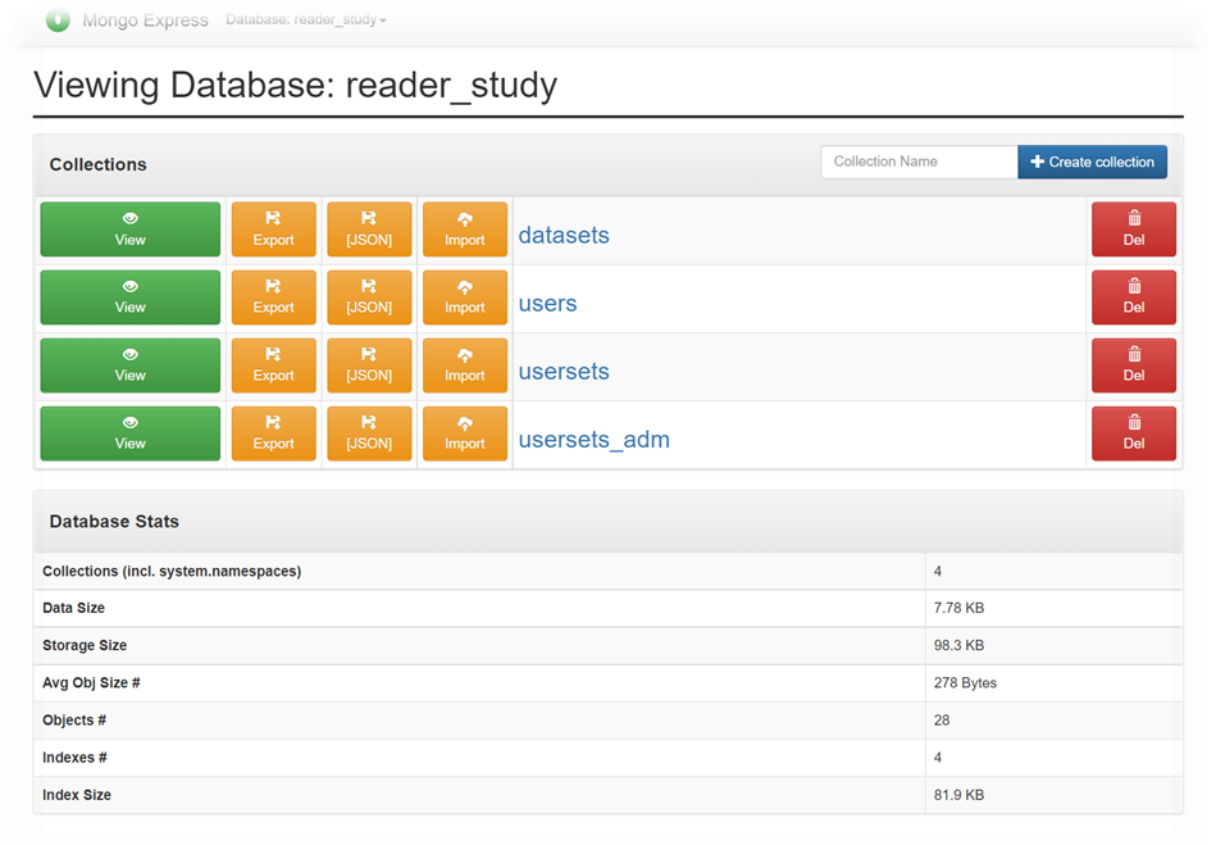


Рисунок 33 – Структура сформированных баз данных

Разработанная платформа применялась на этапах оценки диагностической точности «среднего» врача-рентгенолога, а также на этапе экспертной разметки маммографических исследований.

Инструмент проведения мониторинга работы программного обеспечения на основе технологий искусственного интеллекта

Целью данной платформы является автоматизированный контроль технического качества работы ПО на основе ТИИ. Настоящее решение создано на языке Python и представляет собой платформу с веб-интерфейсом (рисунок 34).

Платформа Мониторинга
Инструмент для отбора исследований и формирования отчета в рамках периодического мониторинга ИИ-сервисов

Выгрузка Лаваль
Выберите выгрузку

ИИ-сервис
Выберите сервис

Отчетный период
Период от → Период до

Модальность
Выберите модальность

Выборка Отчет «а-г» Отчет «а-б» **КТТ**

Выберите дату формирования
19.10.2022

Нажмите для скачивания отчета
СКАЧАТЬ PDF

► Выгрузка результатов из базы данных
► Информация по текущей задаче (отключено)

Рисунок 34 – Общий вид платформы «Платформа Мониторинга»

При помощи указанной платформы осуществляется автоматический расчет технологических параметров ПО на основе ТИИ по всем направлениям эксперимента и формирование отчета КТТ. В указанном документе отражается информация о ПО на основе ТИИ и тестировании, происходит проверка параметров, по результатам которой формируются решения, где предоставлены следующие списки ДУ:

1. Исследования не имеют технологических дефектов, данные ДУ готовы к работе с ИИ.
2. В некоторых исследованиях с ДУ встречаются технологические дефекты, требуется дополнительный анализ работы ПО на основе ТИИ с данной группой ДУ.
3. Исследования с ДУ, содержащие технологические недостатки, не подлежат отправке на ИИ до выявления и устранения проблем. Данные ДУ следует исключить из маршрутизации.

Более подробно работа с платформой описана в инструкции пользователя «Мониторинг».

Позднее данный инструмент был доработан с целью обеспечения многопользовательского режима работы, а также поддержки анализа клинических дефектов. В ходе проспективной работы возникла потребность в анализе удельного веса не только технологических, но и клинических дефектов. Для автоматизации этого процесса была разработана веб-платформа технологического и клинического мониторинга результатов работы алгоритмов анализа цифровых медицинских изображений. Программа предназначена для автоматизации проведения мониторинга технологических и клинических параметров работы ПО на основе ТИИ в лучевой диагностике с целью осуществления периодического контроля качества их работы.

ПО на основе ТИИ, анализируя медицинские изображения в формате DICOM, возвращают ответ в виде описания исследования, вероятности наличия патологии, а также локализации патологических находок. Разработанная веб-платформа принимает на вход: данные в виде таблицы, содержащей технические параметры обработки исследований алгоритмом (идентификатор исследований, временные отметки работы алгоритма, результат работы, протоколы врачебных описаний и т.д.); файл в виде таблицы с результатами ручного анализа врачами ограниченной выборки обработанных исследований (технологические и клинические параметры). Результатом работы программы является: шаблон для ручного анализа врачами с перечнем исследований для проверки технологических и клинических параметров работы алгоритмов; расчет удельного веса исследований, обработанных алгоритмом, с технологическими несоответствиями; расчет удельного веса исследований, некорректно оцененных алгоритмом с клинической точки зрения; формирование отчета по мониторингу в формате PDF и Excel. Внешний вид разработанной платформы представлен на рисунке 35, пример автоматизированного отчета – на рисунке 36.

Информационная система мониторинга						
Реестр Создать новый мониторинг Направление эксперимента ИИ сервисы						
ID	Дата формирования	Направление эксперимента	Источник	Название ИИ сервиса	Период мониторинга	Статус
108	08.02.2023 11:15	определение патологий ОГК	hub	Спейс-BA	01.01.2023 - 01.02.2023	В работе
105	06.02.2023 15:53	рак молочной железы	hub	ТисСМ (МТЛ)	01.01.2023 - 31.01.2023	В работе
104	06.02.2023 15:52	определение патологий ОГК	hub	Третье Мнение ФЛГ	01.01.2023 - 31.01.2023	Отчет сформирован
101	06.02.2023 15:52	определение патологий ОГК	hub	Третье мнение РТ	01.01.2023 - 31.01.2023	Отчет сформирован
102	06.02.2023 15:51	рак молочной железы	hub	Цельс ММЛГ	01.01.2023 - 31.01.2023	Отчет сформирован
101	06.02.2023 15:51	определение патологий ОГК	hub	Цельс ФЛГ	01.01.2023 - 31.01.2023	Отчет сформирован
100	06.02.2023 15:50	определение патологий ОГК	hub	Цельс РТ	01.01.2023 - 31.01.2023	В работе
93	23.01.2023 10:52	определение патологий ОГК	hub	Спейс-BA	24.11.2022 - 31.12.2022	Отчет сформирован
92	23.01.2023 10:47	рак молочной железы	hub	ТисСМ (МТЛ)	24.11.2022 - 31.12.2022	Отчет сформирован
91	23.01.2023 10:46	определение патологий ОГК	hub	Третье Мнение ФЛГ	24.11.2022 - 31.12.2022	Завершен

Рисунок 35 – Веб-интерфейс инструмента проведения мониторинга работы ПО на основе ТИИ

ОТЧЕТ ПО МОНИТОРИНГУ ТЕХНОЛОГИЧЕСКИХ ПАРАМЕТРОВ РАБОТЫ СЕРВИСА		
1. Поставщик ИИ сервиса		
Наименование ИИ сервиса		
Компания производитель ИИ сервиса		
Идентификатор ИИ сервиса		
2. Отчетный период	24.11.2022 - 31.12.2022	
3. Вид исследований	ФЛГ	
4. Клиническая цель	определение патологий ОГК	
5. Общее количество исследований:		
5.1 Направленных на анализ сервису за отчетный период согласно выгрузке из HUB		519
5.2 Из них уникальных согласно выгрузке		519
6. Количество исследований прошедших контроль:		512
6.1 Прошедших выборочный контроль		20
7. Количество исследований:		
7.1 Содержащих технологический дефект "а" шт приложение 1		0
7.2 Содержащих технологический дефект "б" шт приложение 2		32
7.3 Содержащих технологический дефект "в" "г" "д" шт приложение 3		0
7.4 Содержащих технологический дефект "е" шт приложение 4		15
8. Удельный вес исследований:		
8.1 Содержащих технологический дефект "а" относительно 480 исследований %		0 %
8.2 Содержащих технологический дефект "б" относительно 512 исследований %		6 %
8.3 Содержащих технологический дефект "в" "г" "д" относительно 20 исследований %		0 %
9. Количество исследований без дефектов, шт:		480
10. Решение:	участнику Эксперимента необходимо внести изменения в работу Сервиса	
11. Примечания:		
Количество исследований с сообщением об ошибке, полученных от ИИ-сервиса, шт., приложение 5:		7
Дата оформления отчета	23.01.2023 10:46	
Отчет сформирован	автоматически	

Рисунок 36 – Пример отчета, автоматически сформированного инструментом мониторинга

Инструмент тестирования при смене версионности программного обеспечения на основе технологий искусственного интеллекта

Назначение платформы тестирования при смене версионности ПО на основе ТИИ (ПТСВ) искусственного интеллекта: автоматизация проведения калибровочного тестирования сервиса при смене версионности. В основе разработки лежит инструмент проведения калибровочного тестирования, описанного в параграфе 4.1. Отличительной особенностью является наличие модуля прямого взаимодействия с ПО на основе ТИИ.

Общая схема взаимодействия платформы и пользователя приведена на рисунке 37.

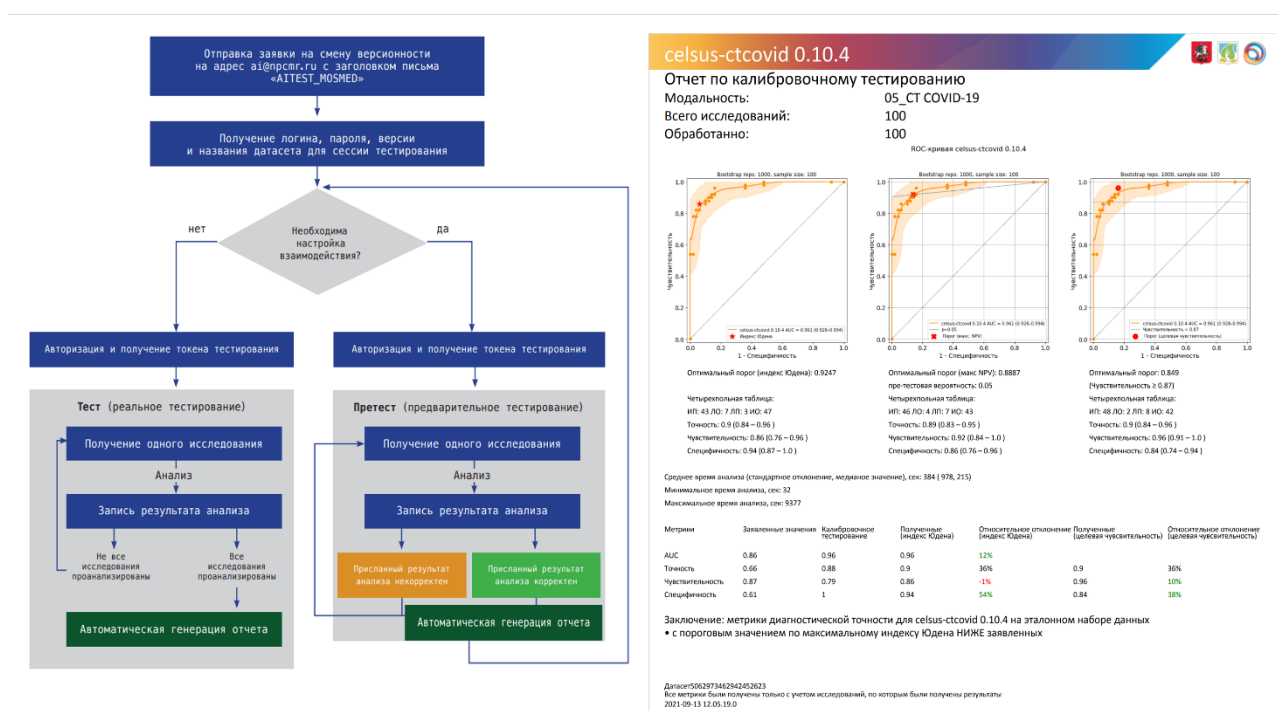


Рисунок 37 – Общая схема взаимодействия ПТСВ и ПО на основе ТИИ и пример подготовленного заключения

Функционал ПТСВ:

— перекодирование данных с подменой идентификационных номеров исследований;

- направление исследований сервису;
- получение ответа от сервиса;
- расчет показателей диагностической точности;
- подготовка протокола заключения.

Работа ПО на основе ТИИ с ПТСВ осуществляется в соответствии с инструкцией. Локализация платформы: внутренний сервер ГБУЗ «НПКЦ ДиТ ДЗМ» (внешний адрес: <https://ai-testing.tele-med.ai/info>).

Отведенное время для проведения тестирования ПО на основе ТИИ: 48 часов с момента выдачи учетных данных.

Платформа тестирования при смене версионности сервиса искусственного интеллекта предназначена для ускорения процесса смены версионности сервисов при необходимости оценки текущих показателей диагностической точности путем проведения калибровочного тестирования.

Оценка применимости разработанного инструментария

Инструмент анализа текстовых протоколов позволил решить несколько практических задач в рамках настоящей работы. С использованием ретроспективного анализа исследований нам удалось рассчитать частоту встречаемости отдельных рентгенологических признаков в потоке исследований. Взяв за основу наиболее часто встречающиеся признаки, а также проведя их клиническую оценку, мы сформировали диагностические требования к ПО на основе ТИИ для анализа РГ ОГК и ФЛГ. Клиническая оценка отдельных РГ-признаков, а также их сочетаний имела большую практическую ценность.

У отдельных рентгенологических признаков при их обнаружении может быть выраженное клиническое значение: «инфильтрация», «объемные образования», «очаг» и т.д. По этой причине обнаружение хотя бы одного из них важно. Напротив, такие рентгенологические признаки, как «фиброз»

и «кальцинат» по отдельности не имеют клинического значения, однако в сочетании с другими рентгенологическими признаками могут обладать выраженной клинической ценностью.

На основании частоты встречаемости рентгенологических признаков были скорректированы базовые диагностические требования к ПО на основе ТИИ для анализа цифровых РГ ОГК и ФЛГ. Действующая версия БДТ содержит следующие признаки патологии: плевральный выпот, пневмоторакс, очаг затемнения, инфильтрация/консолидация, диссеминация, полость, ателектаз, кальцинат/кальцинированная тень, расширение средостения, кардиомегалия, нарушение целостности кортикального слоя и консолидированный перелом. Необходимо отметить, что большая часть из этих признаков выраженно проявляется клинически и не выступает целевыми для профилактических исследований (например, плевральный выпот и пневмоторакс). Вместе с тем нецелесообразно разделять рентгенологические признаки по потокам исследований, т.е. ПО на основе ТИИ, которое может установить очаг затемнения, является целевым для массовых профилактических заболеваний, однако если при этом будет пропущен пневмоторакс, то может иметь неблагоприятные последствия, т.к. в этом случае быстрое принятие врачебного решения имеет высокое значение.

Инструмент анализа текстовых протоколов лег в основу системы оценки диагностической точности ПО на основе ТИИ при работе на потоке данных, позволяя сравнивать результаты работы ПО и врача-рентгенолога на большой выборке исследований. До момента получения приведенных ниже результатов данный программный модуль прошел серию экспертных валидаций на ограниченной выборке исследований – 1000 уникальных протоколов заключений. В ходе этих итераций были выявлены и устранены дефекты в работе модуля, связанные с особенностью подготовки протоколов рентгенологических заключений. Однако возможны единичные случаи некорректного срабатывания

модуля 1–2 на 1000, обусловленные логическими и пунктуационными ошибками врача-рентгенолога при формировании заключения. Например, в заключении могли присутствовать взаимоисключающие формулировки или дифференциальный ряд. В этом случае исследование маркируется как «неопределенное», но в переводе в бинарную оценку этот признак может быть классифицирован как признак патологии. Еще одна важная особенность – пунктуационные ошибки, которые в настоящей версии не могут быть обработаны, однако в последующих версиях была разработана функция (python – класс), позволяющая частично избегать их. Пример значимой ошибки в формировании заключения: *«Очаговые и инфильтративные изменения. Не выявлены Признаки кардиомегалии»*. В этом заключении точка воспринимается алгоритмом как конец смыслового фрагмента, следовательно, анализируется первое предложение, содержащее ключевые слова, указывающие на наличие патологии, на основании которых выставляется лейбл наличия очаговых и инфильтративных изменений. Далее, во втором предложении, присутствует стоп-фраза «не выявлено», которая останавливает работу алгоритма, не позволяя корректно поставить лейбл наличия кардиомегалии. В итоге исследование классифицируется бинарно как патологическое, что является корректным, т.к. есть признаки кардиомегалии, однако лейблы рентгенологических признаков проставлены некорректно.

Таким образом, бинарная оценка протоколов корректна, однако в единичных заключениях наличие рентгенологических признаков может быть некорректно проставлено алгоритмом. Пути уменьшения числа возникновения таких ошибок заключаются в стандартизации формулировок, предзаполненных шаблонах, формализации отдельных признаков рентгенологических паттернов. Это позволит повысить точность сопоставления выявленных находок врачом-рентгенологом и ПО на основе ТИИ.

Вследствие обозначенных выше ограничений инструмента при его применении использовали бинарную оценку исследования. Это полностью соответствовало целевому применению ПО на основе ТИИ для профилактических исследований и предполагало решение бинарной задачи – «норма» или «не норма». Несмотря на это, наличие мультиклассовой классификации протокола по типам выявленных рентгенологических признаков позволяет проводить анализ чувствительности ПО на основе ТИИ к отдельным типам патологии. Это необходимо для предоставления обратной связи разработчикам с целью улучшения ПО на основе ТИИ.

Модуль инструмента анализа текстовых протоколов для маммографических исследований позволяет проводить оценку как по бинарной шкале (есть или нет патологии), так и детально анализировать классы BI-RADS как по исследованию в целом, так и по каждой молочной железе. Кроме этого, модуль разрешает дифференцировать оценки по скрининговой шкале и по диагностической шкале, переводя каждую из них в бинарную. Несмотря на широкие функциональные возможности модуля по анализу текстовых протоколов ММГ-исследований, в настоящей работе была использована лишь бинарная классификация, результаты которой применялись для последующего расчета показателей диагностической точности. В отличие от инструмента анализа протоколов рентгенографии органов грудной клетки точность работы инструмента по анализу маммографических исследований составляла 100%, а те единичные случаи, которые не были обработаны им, не содержали классов патологии по шкале BI-RADS.

Разработанный в рамках настоящей работы инструмент расчета показателей диагностической точности позволил получить показатели диагностической точности, рассматриваемые в работе ПО на основе ТИИ, а также их 95% доверительный интервал. При определении показателей диагностической точности осуществлялся ROC-анализ на основе данных бинарной оценки

исследования, при этом в качестве истинного значения использовалось значение, полученное при подготовке набора данных или на основании заключения врача-рентгенолога. Возможность оценки динамики показателей диагностической точности в зависимости от установленного порогового значения позволили определить оптимальные пороговые значения для каждого из исследуемых ПО как на этапе тестирования, так и на этапе проспективной работы.

Основной задачей при создании инструмента для сравнения точности диагностических алгоритмов, работающих на принципах искусственного интеллекта, был подбор метода для статистического анализа области под ROC-кривой. Существует множество методик для анализа диагностической точности, а также разнообразные программные решения, от программных кодов до готовых инструментов, которые не требуют знания специфических программных языков. Тем не менее доступ ко многим из этих инструментов ограничен, их использование может быть неинтуитивным и сложным. Создавая собственное решение, мы стремились к тому, чтобы оно было простым, доступным и легким в использовании. Ранее мы разработали инструмент для построения ROC-кривых с возможностью расчета доверительных интервалов для площади под кривой с использованием метода Делонга и бутстрэппинга, что позволило успешно проводить калибровку алгоритмов. В контексте сравнения алгоритмов на основе ИИ возникла потребность в инструменте, способном не только выявлять статистически значимые различия, но и вычислять p -значения для сравниваемых данных. Из-за этого было решено использовать перестановочный тест, который легко реализуется на Python – языке, на котором изначально разрабатывался наш инструмент. Перестановочный тест также решает проблему множественного тестирования, что важно при сравнении большого количества моделей.

Анализируя результаты, полученные с помощью перестановочного теста, было замечено, что p -значения иногда были выше, чем при использовании `roc.test`, хотя в некоторых случаях они совпадали или были ниже. Некоторые

исследователи считают метод Делонга, применяемый в `roc.test`, слишком консервативным, что часто приводит к отклонению нулевой гипотезы об отсутствии различий. Эффективность метода Делонга также может зависеть от размера и состава данных, что может объяснять различное поведение p -значений при сравнении методов.

Важно подчеркнуть, что перестановочный тест, как и метод бутстрэппинга, не гарантирует полную воспроизводимость результатов из-за использования псевдослучайных выборок, что может привести к изменению p -значений при повторных тестах.

В результате был разработан модуль для анализа характеристических кривых, который позволяет сравнивать площади под кривыми, построенными на основе одинакового количества данных. Это послужило основой для создания инструментов автоматизации калибровочных тестирований, а также тестирований при смене версионности ПО на основе ТИИ.

Для оценки диагностической точности врача-рентгенолога была разработана платформа «экспертного пересмотра», позволившая разметить РГ ОГК и маммографические исследования. При этом использовалась вероятностная шкала (степень уверенности врача-рентгенолога в том, что исследование содержит признаки целевой патологии). Сопоставив полученные вероятностные оценки с истинными значениями, мы получили показатели диагностической точности врача-рентгенолога.

Практический опыт создания наборов данных в рамках настоящей работы лег в основу многоэтапной методики подготовки наборов данных. Подготовка набора данных начинается с этапа «инициации», в рамках которого фиксируется тип набора данных, его назначение, цель создания набора данных. Далее, на этапе планирования, фиксируется срок проведения работ по созданию набора данных и ответственные участники. На следующем «основном» этапе происходят работы по отбору исследований, их анализу и формированию итогового набора данных.

Затем следует этап регистрации готового набора данных, на котором осуществляется: подготовка документации, в т.ч. структурированного описания, readme-файла и карточки набора данных.

В настоящем параграфе был представлен перечень инструментов: анализ текстовых протоколов; комплекс инструментов для расчета показателей диагностической точности; инструмент для проведения мониторинга, контроля качества при смене версионности; инструмент оценки диагностической точности врача-рентгенолога. Таким образом, в рамках настоящей работы был разработан целый набор инструментов, охватывающий самый широкий спектр задач, связанных с внедрением технологий на основе искусственного интеллекта в медицинскую диагностику.

Большое практическое значение имеет разработанный инструмент текстового анализа врачебных протоколов, позволяющий отслеживать расхождение в трактовке исследования врачом-рентгенологом и ПО на основе ТИИ. Это, как показала проведенная работа, может быть использовано для предоставления обратной связи разработчикам ПО и позволяет выявлять и устранять дефекты в работе ПО на основе ТИИ.

Большую практическую ценность представляет уникальный инструмент оценки диагностической точности врача-рентгенолога, позволяющий определить порог для внедрения ПО на основе ТИИ в клиническую практику, что было подробно изложено ранее в этом параграфе. Инструменты контроля качества имеют важное значение при реализации комплексной системы контроля качества ПО на основе ТИИ, которой будет посвящен следующий параграф.

4.3. Апробация системы контроля качества при проведении профилактических исследований с применением технологий искусственного интеллекта

Методика технологического и клинического мониторингов

Промышленная эксплуатация ПО на основе ТИИ происходила в ППАК, где с целью мониторинга работы ПО в ППАК проводился периодический мониторинг технологических дефектов. Периодический мониторинг осуществлялся для всех исследований, проанализированных ПО на основе ТИИ за отчетный период по дефектам (рисунки 38–41):

- а) время, затрачиваемое на обработку одного исследования, превышает 6,5 минут;
- б) отсутствие результатов проанализированных исследований в ППАК ЕРИС ЕМИАС по данным ЕРИС ЕМИАС;
- в) некорректная работа функционала ПО на основе ТИИ, затрудняющая работу врача-рентгенолога или делающая ее выполнение невозможным с надлежащим качеством;
- г) дефекты, связанные с отображением области изображений;
- д) иные нарушения целостности и содержимого файлов с результатами исследований, обуславливающие ограничение их диагностической интерпретации;
- е) дефекты, связанные с клинической оценкой работы ПО на основе ТИИ.

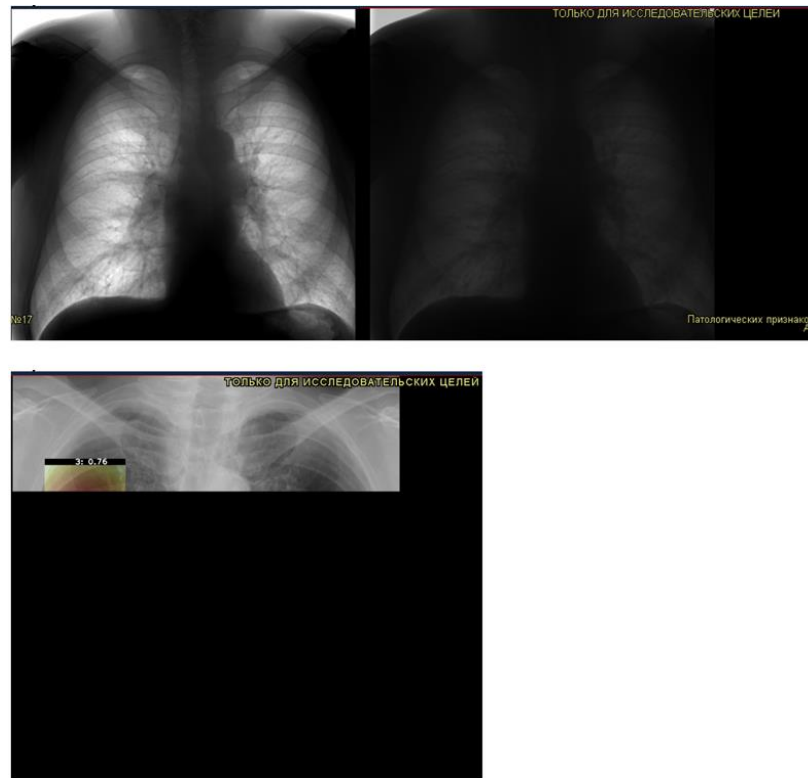


Рисунок 38 – Технологический дефект – на дополнительной серии от ИИ искажение яркости/контрастности (вверху) и обрезка исследования (внизу)

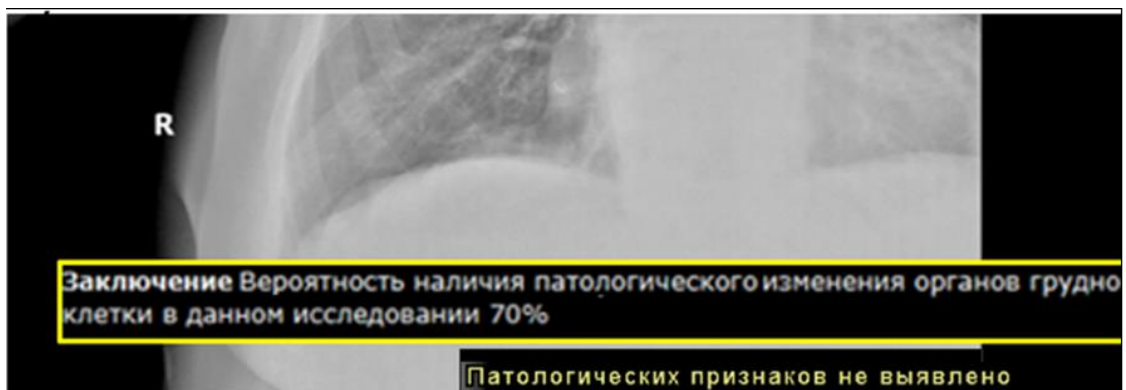


Рисунок 39 – Технологический дефект – на дополнительной серии от ИИ противоречивые формулировки в заключении и на изображении

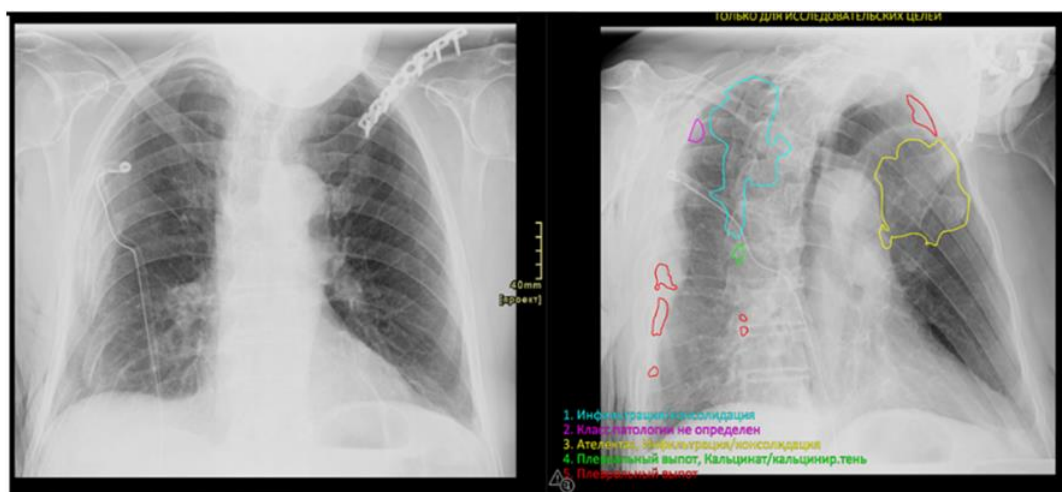


Рисунок 40 – Технологический дефект – ИИ выбрал для обработки неправильную проекцию (слева – подходящая проекция, справа – обработанная ИИ проекция)

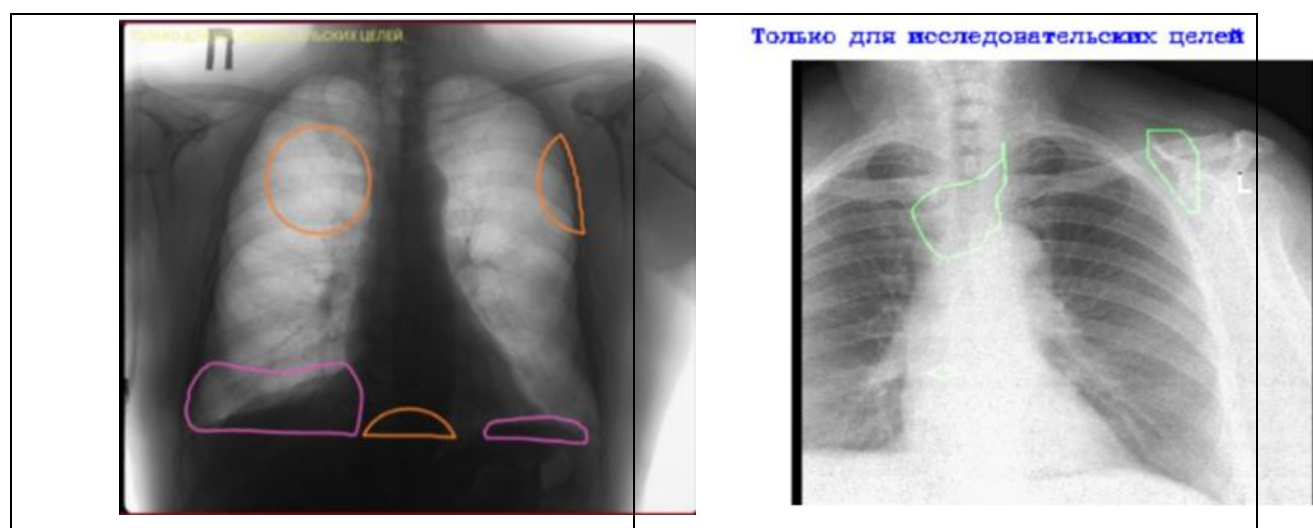


Рисунок 41 – Технологический дефект – ИИ выделил находки за пределами области интереса

Мониторинг необходим для оперативного контроля качества работы ПО на основе ТИИ и недопущения некорректно функционирующих ПО на основе ТИИ врачам-рентгенологам. Определено пороговое значение по удельному весу дефектов: 10%. При превышении указанного значения маршрутизация исследований на данное ПО на основе ТИИ останавливается до момента устранения причин дефектов.

Проведение мониторинга работы ПО на основе ТИИ осуществляется с использованием вышеописанной специализированной платформы «Платформа Мониторинга». Последовательность действий по работе с платформой приведена

на рисунке 42. На первом этапе предлагается выбрать отчетный период, наименование ПО на основе ТИИ и вид исследования. Также необходимо выбрать актуальную выгрузку, которая будет использоваться при формировании отчета. Платформа позволяет выбрать необходимый объем выборки для экспертного анализа врачами-рентгенологами. В случае необходимости может быть подготовлен как сокращенный отчет из дефектов «а» и «б» или полный, в т.ч. с клинической оценкой. При помощи указанной платформы осуществляется автоматический расчет технологических параметров ПО на основе ТИИ и формирование отчета по мониторингу технологических параметров работы ПО.

Рисунок 42 – Общий вид платформы «Платформа Мониторинга»

Также в ходе мониторинга осуществляется клиническая оценка работы ПО на основе ТИИ (дефект группы «е»), которая заключается в сопоставлении заключений по данному исследованию от врача и от ПО на основе ТИИ, а в случае наличия патологических изменений – в оценке совпадения по локализации. Выявленные технологические дефекты и расхождения в клинической оценке исследований заносятся в специально разработанную машиночитаемую форму (рисунок 43).

№	uid	ДУ (кластер)	v1	v2	v3	v4	v5	г1	г2	г3	г4	г5	д1	д2	Трактовка: Полное соответствие	Трактовка: Некорректная оценка	Трактовка: Ложноположительный	Трактовка: Ложноотрицательный	Локализация: Полное соответствие	Локализация: Некорректная оценка	Локализация: Ложноположительный	Локализация: Ложноотрицательный	Патология
1	1.2.840.113619.2.428.3.1930041499.412.1662715257.425	Стационарное																					
2	1.2.40.0.13.1.140836700935241097336075927127531413793	Стационарное																					
3	1.2.40.0.13.1.17690213845763538516817435189540827775	Стационарное																					
4	1.2.40.0.13.1.187195811162793297618658046991843626134	Стационарное																					
5	1.2.40.0.13.1.134633740647811648714437997180089921757	Стационарное																					
6	1.2.40.0.13.1.249372427959682938467233487026509718133	Стационарное																					
7	1.2.40.0.13.1.244965995365728463088675711325886994	Стационарное																					
8	1.2.40.0.13.1.136295058877467157604795303941611554740	Стационарное																					
9	1.2.40.0.13.1.162150460993826595141020274462018443208	Стационарное																					
10	1.2.40.0.13.1.217281132325742678611694139296898024882	Стационарное																					
11	1.2.40.0.13.1.11674490493170685402385276301415408185	Стационарное																					
12	1.2.40.0.13.1.32326971936266249703406989502756309878	Стационарное																					
13	1.2.40.0.13.1.277501542103145000114801236044314629790	Стационарное																					
14	1.2.40.0.13.1.125676883496832328965034265985058312	Стационарное																					
15	1.2.40.0.13.1.208726264519416535716500129047241088550	Стационарное																					
16	1.2.40.0.13.1.218575378872197351037157732601586187151	Стационарное																					
17	1.2.40.0.13.1.271540240066158659854855190131604824919	Стационарное																					
18	1.2.40.0.13.1.27099845695673357195985999596828102777	Стационарное																					
19	1.2.840.113619.2.428.3.1930040987.825.1663555602.971	Стационарное																					
20	1.2.40.0.13.1.25360855439357269547246172434281935522	Стационарное																					
21	1.2.40.0.13.1.19066921615745712190229044285058423021	Стационарное																					
22	1.2.40.0.13.1.95170285926374564805273021052498467957	Стационарное																					
23	1.2.40.0.13.1.23491284414967526404456648834863928902	Стационарное																					
24	1.2.40.0.13.1.116660743291105616775436972916174360362	Стационарное																					
25	1.2.40.0.13.1.112648747101889769936025021477156262320	Стационарное																					
26	1.2.40.0.13.1.126189121374710402904687698421361508390	Стационарное																					
27	1.2.40.0.13.1.98339558098990065458331538973218006	Стационарное																					
28	1.2.840.113619.2.428.3.1930040987.825.1663555596.350	Стационарное																					
29	1.2.40.0.13.1.272416791686260204164206672131298071915	Стационарное																					
30	1.2.40.0.13.1.23874796638923644379091013931919829382	Стационарное																					
31	1.2.40.0.13.1.332737875185432914342291251218761873959	Стационарное																					
32	1.2.40.0.13.1.9441431046699790133485563840297998603	Стационарное																					
33	1.2.40.0.13.1.523042159093936748016795232229791712	Стационарное																					
34	1.2.40.0.13.1.325826350647659177512612350073465679605	Стационарное																					
35	1.2.40.0.13.1.16604526220577046537014077848623260314	Стационарное																					
36	1.2.40.0.13.1.222317285659913022446994220579543283161	Стационарное																					
37	1.2.40.0.13.1.234634172424589036581484291859481818961	Стационарное																					
38	1.2.40.0.13.1.69469044522213667313694107982800504967	Стационарное																					

Рисунок 43 – Форма внутреннего отчета проведения мониторинга работы ПО на основе ТИИ

Для корректной оценки дефектов группы «е» было разработано руководство «Клиническая оценка результатов работы сервиса искусственного интеллекта при проведении мониторинга».

В процессе клинической оценки работы сервиса врач оставляет свое мнение о согласии/несогласии с трактовкой и локализацией, предоставленными в результате работы сервиса.

В форме внутреннего отчета (рисунок 43) возможны следующие варианты:

- Трактовка: Полное соответствие.
- Трактовка: Некорректная оценка.
- Трактовка: Ложноположительный.
- Трактовка: Ложноотрицательный.
- Локализация: Полное соответствие.
- Локализация: Некорректная оценка.
- Локализация: Ложноположительный.
- Локализация: Ложноотрицательный.

Формулировка «Трактовка: Полное соответствие» выбирается в случае полного согласия врача с заключением в DICOM SR от ПО на основе ТИИ.

Формулировка «Трактовка: Некорректная оценка» выбирается в случае частичного согласия врача с заключением в DICOM SR от ПО на основе ТИИ (например, согласие с общим указанием наличия патологических находок, но несогласие с детализацией или согласие с детализацией, но несогласие с общим выводом о вероятности или степени тяжести патологических находок).

Формулировка «Трактовка: Ложноположительный» и «Трактовка: Ложноотрицательный» выбирается в случае полного несогласия врача с заключением в DICOM SR от ПО на основе ТИИ.

Формулировка «Локализация: Полное соответствие» выбирается в случае полного согласия врача с локализацией на дополнительной серии от ПО на основе ТИИ (рисунок 44).

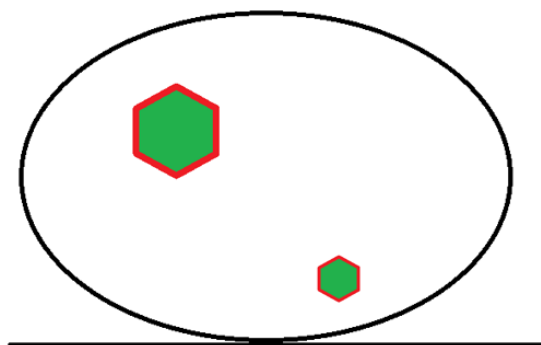


Рисунок 44 – Локализация: Полное соответствие: область целевого органа – черный цвет, истинная патологическая находка – красный контур, локализация патологической находки ПО на основе ТИИ – зеленый цвет

Формулировка «Локализация: Некорректная оценка» выбирается в случае пропуска ПО на основе ТИИ части целевых находок, при условии, что часть была локализована верно (рисунок 45) или некорректной локализации (рисунок 46).

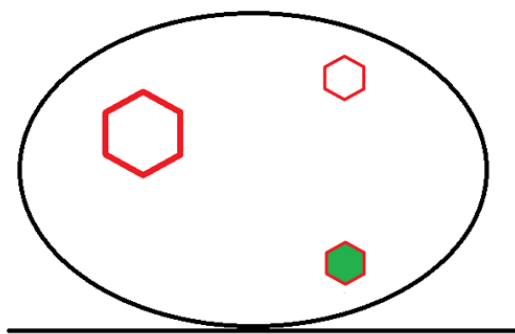


Рисунок 45 – Локализация: Некорректная оценка: область целевого органа – черный цвет, истинная патологическая находка – красный контур, локализация патологической находки ПО на основе ТИИ – зеленый цвет

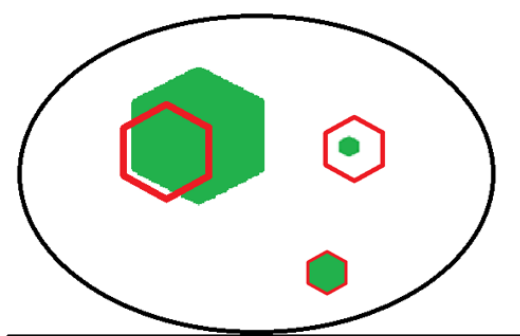


Рисунок 46 – Локализация: Некорректная оценка: область целевого органа – черный цвет, истинная патологическая находка – красный контур, локализация патологической находки ПО на основе ТИИ – зеленый цвет

Формулировка «Локализация: Ложноположительный» выбирается в случае локализации ПО на основе ТИИ патологических находок, которые ими не являются (рисунок 47).

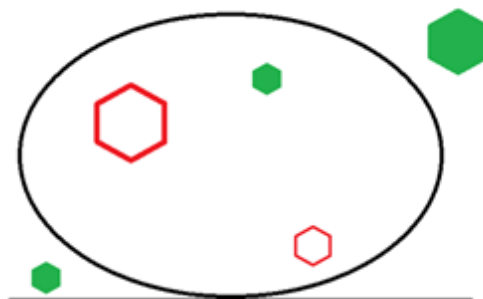


Рисунок 47 – Локализация: Ложноположительный: область целевого органа – черный цвет, истинная патологическая находка – красный контур, ПО на основе ТИИ выявило патологические находки за пределами целевого органа – зеленый цвет

Формулировка «Локализация: Ложноотрицательный» выбирается в случае полного несогласия врача с локализацией на дополнительной серии от ПО на основе ТИИ (рисунок 48).



Рисунок 48 – Локализация: Ложноотрицательный: область целевого органа – черный цвет, истинная патологическая находка – красный контур, ПО на основе ТИИ не выявило патологических находок – зеленый цвет

Заполненная форма, а также список исследований, подлежащих мониторингу, загружаются на платформу, формируя итоговый отчет по мониторингу.

Клиническая оценка формируется путем оценки локализации находок (маркировки) и формирования описания (заключения) для каждого исследования из выборки. Клиническая оценка работы ПО на основе ТИИ является средним арифметическим оценки маркировки и заключения. Оценка маркировки и заключения является средним арифметическим всех оценок по выборке исследований, выраженным в процентах, по критериям, указанным в таблице 16.

Таблица 16 – Критерии и показатели клинической оценки

Критерии оценки	Описание	Баллы	
		маркировка	заключение
Полное соответствие	Отмечены все целевые находки или «норма»	1	1
Некорректная оценка (частичный / избыточный)	Отмечено не менее 1 находки по целевой патологии (или не на всех снимках/проекциях). Неточное оконтуривание находок. Некорректная оценка объема/	0,5	0,5

	количества находок		

Продолжение таблицы 16

Критерии оценки	Описание	Баллы	
		маркировка	заключение
Ложноположительный	Ложная/лишняя находка. Отмечена находка при фактическом полном отсутствии целевой патологии	0,25	0,25
Ложноотрицательный	Пропуск находки. Не отмечено ни одной находки по целевой патологии при фактическом их наличии	0	0

Для каждого направления профилактических исследований рассчитаны пороговые значения показателей клинической оценки работы сервиса.

Допустимые показатели клинической оценки работы ПО на основе ТИИ:

- для маммографии с целью диагностики рака молочной железы – не менее 70%;
- для рентгенографии органов грудной клетки легких с целью диагностики различных патологий – не менее 50%;
- для флюорографии легких с целью с целью диагностики различных патологий – не менее 50%.

Результаты работы программного обеспечения на основе технологий искусственного интеллекта на потоке (технологические дефекты)

Режим апробации и опытной эксплуатации ПО на основе ТИИ осуществлялся в ППАК, где каждый сервис, по мере готовности, настраивался на маршрутизацию определенных видов исследования с марта 2021 года. Получена сводная информация о динамике технологических дефектов по направлениям и сервисам с указанием медианного, среднего, максимального и минимального значений за весь период работы. Указанная графическая информация о среднем

числе технологических дефектов для каждого направления (рисунки 49–52) позволяет проследить тенденцию к снижению числа технологических дефектов, что напрямую свидетельствует о доработке ПО на основе ТИИ и повышении их качества.

С целью количественной оценки данной тенденции были добавлены соответствующие линии тренда (линейные функции вида $k * x + b$). Зная коэффициент k , можно прогнозировать временной промежуток ликвидации технологических дефектов для каждого ПО на основе ТИИ в отдельности или для целого направления. Лидерами по снижению количества технологических дефектов является маммография – ликвидация таких дефектов уже ожидалась в первых месяцах работы в 2022 году. Напротив, ПО на основе ТИИ вида исследования «РГ/ФЛГ» имеют более ровную тенденцию к снижению числа технологических дефектов (рисунок 50).

Приведенные данные свидетельствуют о большом разрыве для РГ ОГК (до 14%), для двух ММГ решений разрыв составил всего 1,6%.

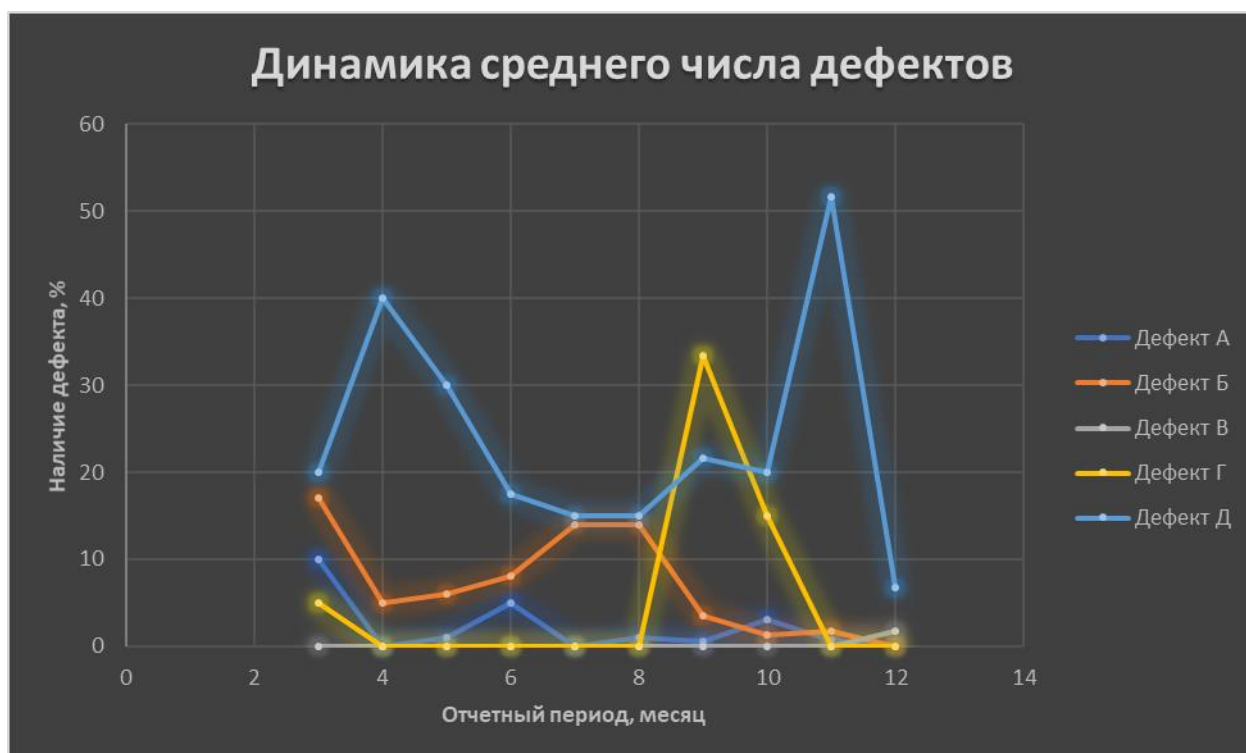


Рисунок 49 – Динамика среднего числа каждого технологического дефекта для сервисов «РГ ОГК»

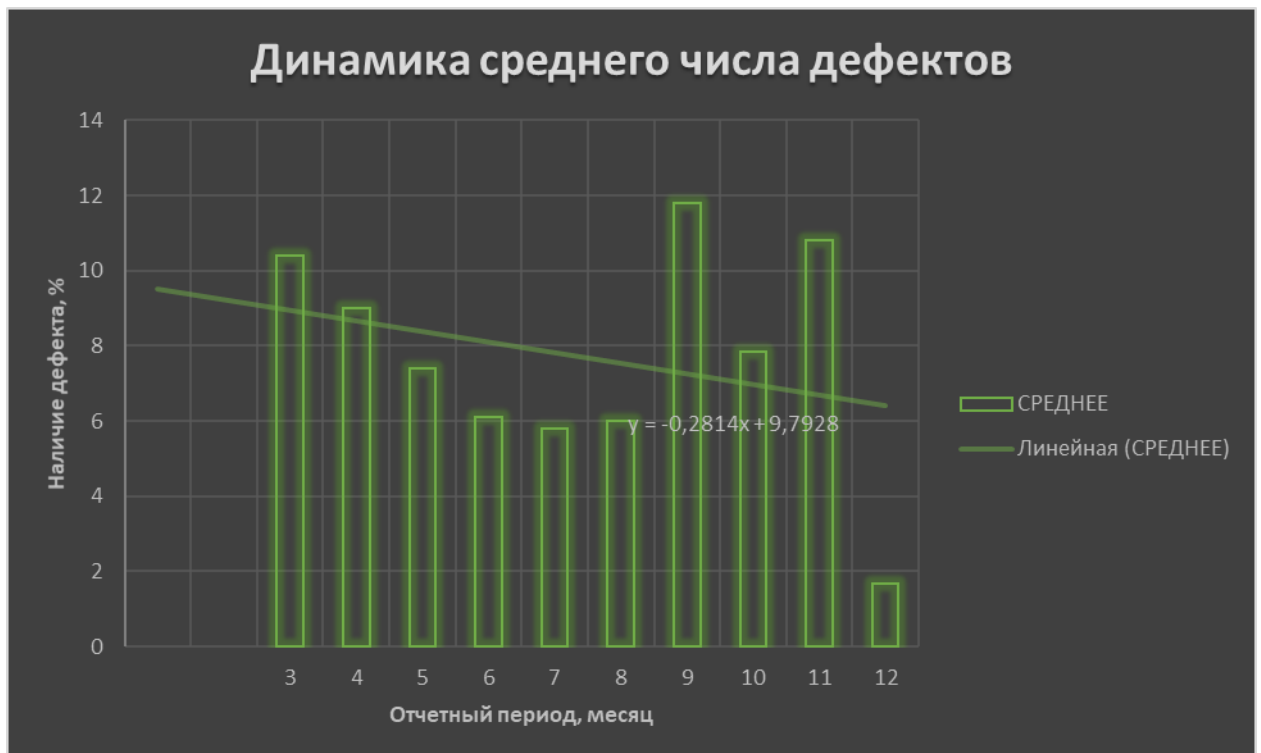


Рисунок 50 – Динамика технологических дефектов для сервисов «РГ ОГК»

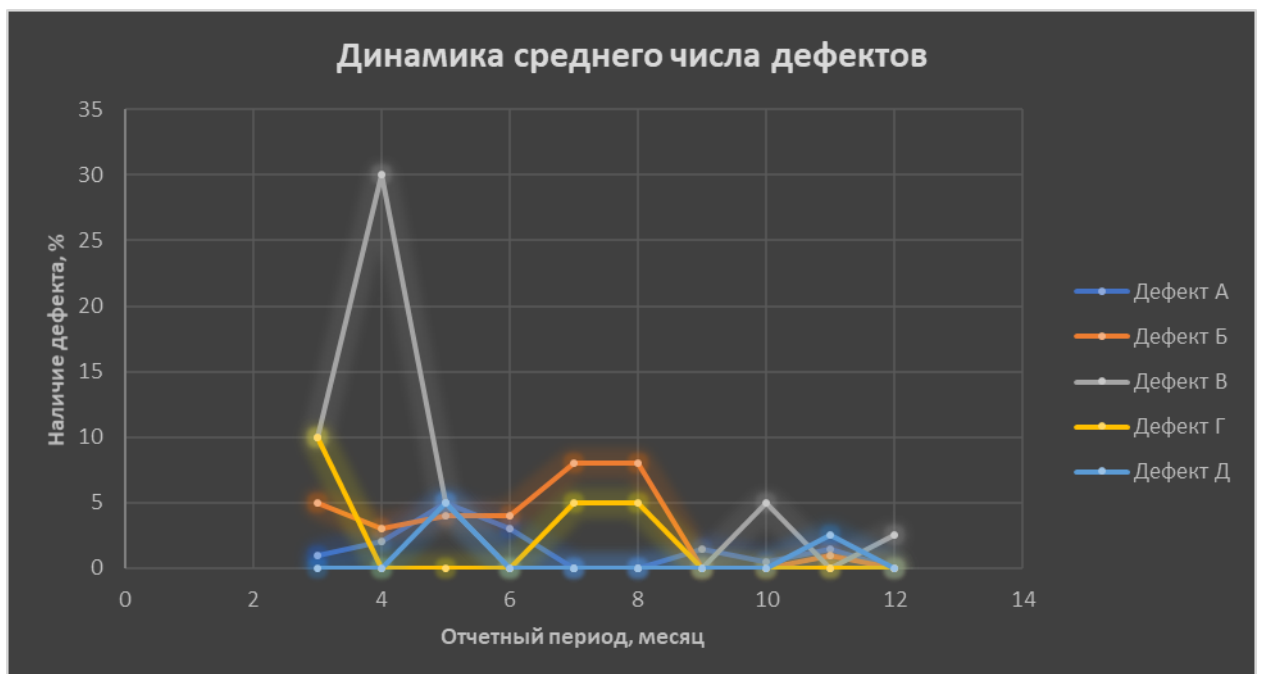


Рисунок 51 – Динамика среднего числа каждого технологического дефекта для сервисов «ММГ РМЖ»



Рисунок 52 – Динамика технологических дефектов для сервисов «ММГ РМЖ»

Как видно из вышеприведенной информации, общая тенденция наличия технологических дефектов имеет убывающую линию тренда, что в первом приближении может говорить об уменьшении количества дефектов и повышении качества результатов работы сервисов. Несмотря на это, была предпринята попытка прогнозирования возникновения технологических дефектов в процессе работы ПО на основе ТИИ. Это актуально для новых разработок, которые еще не применялись на потоке данных.

Методика прогнозирования возникновения дефектов в работе программного обеспечения на основе технологий искусственного интеллекта

На примере двух типов дефектов рассмотрим методику прогнозирования возникновения дефектов в работе ПО на основе ТИИ. Первый тип обозначен как дефекты типа «а» и фиксируется, если время, затрачиваемое ПО на основе ТИИ на анализ одного исследования, превышает 6,5 минут. Второй тип дефектов обозначен как дефекты типа «б» и фиксируется в случае отсутствия результатов анализа исследования в ППАК ЕРИС ЕМИАС. Основным критерий успешного

прохождения тестирования сервисом искусственного интеллекта – суммарное количество дефектов типа «а» и «б», не превышающее 10% от общего числа проанализированных ПО на основе ТИИ данных (на один экземпляр программного обеспечения на основе технологии искусственного интеллекта).

Тестирование каждого экземпляра ПО на основе ТИИ проводилось в период с марта по декабрь 2021 года, данный период времени является достаточно продолжительным и экономически затратным, как для разработчиков медицинского программного обеспечения на основе технологии искусственного интеллекта, так и для лиц, задействованных в тестировании сервисов. В настоящее время существует несколько подходов, позволяющих снизить затраты при эксплуатации различных аппаратных и программно-аппаратных комплексов, основу которых составляет анализ условий эксплуатации за определенный период времени (в частности, зависимость нагрузки на программно-аппаратный комплекс от времени работы комплекса).

В качестве исходных данных для построения прогностической модели использовались результаты работы 24 вариантов ПО на основе ТИИ (были включены ПО, обрабатывающие также КТ ОГК, для повышения качества разрабатываемой прогностической модели) в период с марта по декабрь 2021 года.

Анализ распределения вероятностей возникновения дефектов по критериям Акаике и Байеса показал, что для дефектов типа «А» и «Б» наиболее близким теоретическим распределением является распределение Вейбулла.

На рисунке 53 представлены результаты моделирования вероятности возникновения дефектов и фактическое количество дефектов типа «а» и «б», возникших на сервисах в 2022 году.

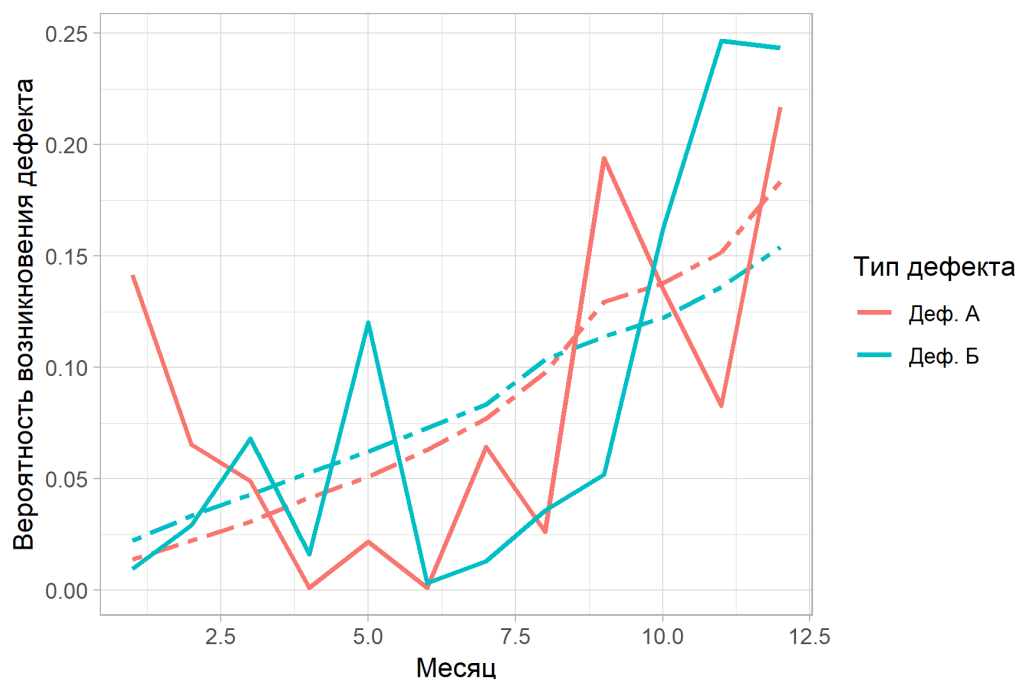


Рисунок 53 – Сравнение результатов прогноза (пунктирная линия) вероятности возникновения технологических дефектов, построенной на основании разработанной модели и реальных данных (непрерывная линия)

Обобщение результатов показывает, что прогнозирование вероятности образования дефектов на основании ограниченного количества неоднородных данных хорошо повторяет основной тренд дефектности, возникающий в 2022 году.

4.4. Инструмент поддержки принятия управленческих решений о выборе программного обеспечения на основе технологий искусственного интеллекта

В условиях цифровой трансформации, при автоматизации производственных процессов организатор здравоохранения сталкивается с проблемой объективного, обоснованного выбора конкретного ПО на основе ТИИ. Принятие такого управленческого решения сопряжено со значительными сложностями: необходимость адекватной формулировки клинико-организационной задачи для автоматизации, учет индивидуальных особенностей и потребностей (медицинской организации, региональной системы

здравоохранения и т.д.), обилие данных – иногда противоречивых – об эффективности тех или иных программных решений.

Созданная нами система контроля безопасности и качества позволяет оценить ПО на основе ТИИ с различных позиций – технических, функциональных, клинико-диагностических. Однако очевидно, что система достаточно сложна, её применение и интерпретация соответствующих результатов требуют временных и прочих затрат. Необходимо создание дополнительного инструмента, который бы позволил интегрировать результаты многосторонней оценки ТИИ, осуществлять бенчмаркинг, объективно и наглядно предоставлять сводные данные организатору здравоохранения для принятия управленческих решений.

Создание такого инструмента осуществлено нами в два этапа. На первом разработана методика рейтингования, на втором – методика усовершенствована, осуществлено её принципиальное развитие, создана матрица зрелости ИИ-сервисов.

Первый этап – методика рейтингования ПО на основе ТИИ. Цель методики – комплексная и периодическая оценка применимости ПО на основе ТИИ в медицинской диагностике, составление рейтинга по основным качественным показателям для оценки возможности дальнейшего использования в системе здравоохранения. Для достижения поставленной цели выполнялась задача аналитики использования ПО на основе ТИИ в системе здравоохранения, которая включала в себя технические (качество и стабильность обмена данными, функциональных возможностей) и клинические (аудит результатов работы, соответствие БДТ, оценка диагностической точности и обратной связи от врачей) аспекты.

В результате практического применения ПО на основе ТИИ был разработан комплексный подход к их оценке, который включает сравнение ПО на основе итогового балла за определенный период. Этот подход основывается на формуле,

включающей четыре ключевых параметра и соответствующие весовые коэффициенты, определенные с использованием методики ранжирования и парного сравнения экспертных оценок [267]. При выборе весов учитывалась статистическая значимость параметров, определенных на ограниченной выборке данных. Были определены две пары параметров: параметры В и С4 (отражающие техническую стабильность и качество работы ПО на основе ТИИ) обеспечивающие 20% оценки итогового балла и пара С2 и С3 (отражают диагностическую точность) и составляющие оставшиеся 80% итогового балла.

Детальная информация о каждом параметре приведена в таблице 17. Формула для расчета итогового балла для ПО на основе ТИИ за отчетный период имеет следующий вид (10):

$$D = B*0,10 + C4*0,10 + C2*0,60 + C3*0,20 \quad (10)$$

Таблица 17 – Детальная информация о расчетах параметров для балльной оценки ПО на основе ТИИ

Параметр	Описание	Формула расчета	Объем выборки	Диапазон значений
В	Выполнение SLA (порог 6,5 минут)	$(6,5-t)*15,39$ 0 при $t \geq 6.5$ минут 100 при $t \rightarrow 0$	Вся за отчетный период	[0, 100]
С2	Минимальная диагностическая точность (AUROC) по результатам проспективного сравнения с текстами протоколов согласно базовым диагностическим требованиям	$(AUROC - ДИ\_AUROC/2)$		(50, 100]
С3	Удельный вес согласия врачей с локализацией находок	С3	Выборка из 20 исследований за отчетный период	[0, 100]
С4	Удельный вес исследований без дефектов «в»–«е»	100 – уд. вес дефектов		[0, 100]

Для включения ПО на основе ТИИ в рейтинг установлены пороговые значения итогового балла, отражающие минимально достаточный уровень качества для конечного пользователя, с установленными минимальными значениями для каждого параметра, включая длительность обработки, диагностическую точность, согласованность решений и долю исследований без дефектов. Пороговые значения для каждого из признаков основаны на методических и клинических рекомендациях [235], а также нормативно-правовых актах [268]. Пороговое значение итогового балла для лидеров по направлениям составляет 80.

Для вида исследования РГ/ФЛГ ОГК итоговый балл находился в диапазоне 60–84. Лидирующие позиции в разные временные интервалы занимали разные ПО на основе ТИИ: ИИ-02, ИИ-01ФЛГ, ИИ-08РГ (рисунок 54). Усредненная оценка всех ПО на основе ТИИ по РГ/ФЛГ ОГК (рисунок 55) также с положительной динамикой. Усредненный итоговый показатель на конец 2021 года составил 77 баллов. Уравнение линии тренда ($0,6452 \cdot x + 70,233$) позволяет предположить достижение среднего уровня по всем ПО на основе ТИИ в 86 баллов к концу 2022 года, что и было подтверждено для одного из ПО: по данным за ноябрь 2022 года ПО на основе ТИИ для анализа РГ ОГК «ИИ-01» достигло значения 87 баллов, что указывает на применимость данной методики и возможность прогнозирования развития решений на основе ТИИ.

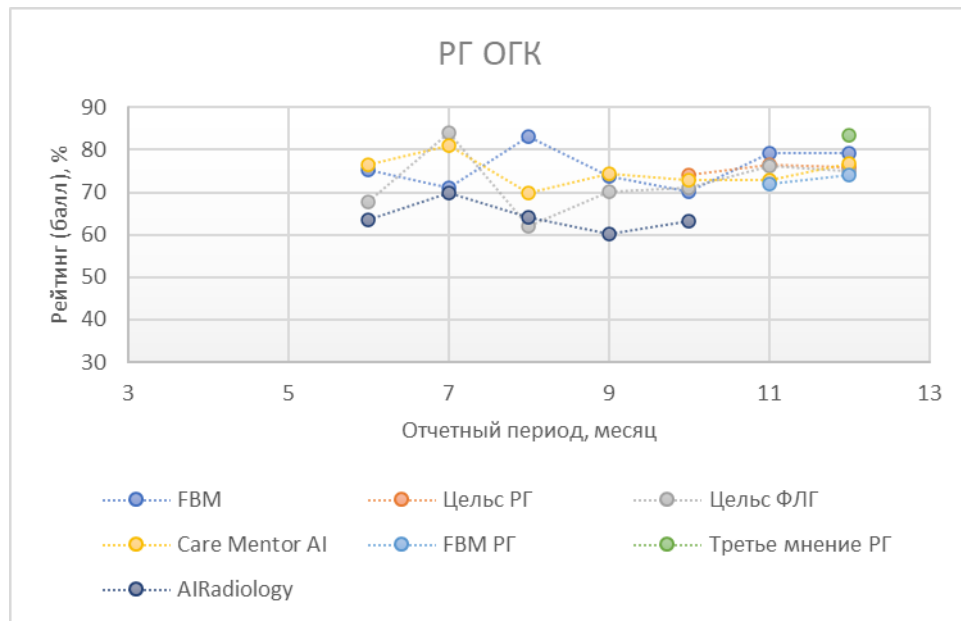


Рисунок 54 – Динамика результатов работы сервисов РГ ОГК



Рисунок 55 – Среднее значение для сервисов РГ ОГК

Вид исследования «ММГ» в 2021 году был представлен двумя ПО на основе ТИИ: ИИ-01 ММГ и ИИ-03ДМ. Итоговый балл находился в диапазоне 59–82. Лидирующие позиции в разные временные интервалы занимали разные ПО на основе ТИИ. Эффективная доработка ПО на основе ТИИ компании «ИИ-01» в конце августа позволила повысить показатели диагностической точности, дальнейшая доработка технической стабильности работы данного ПО на основе

ТИИ дала непрерывный рост итогового балла. Усредненная оценка всех ПО на основе ТИИ по ММГ (рисунки 56, 57) – с положительной динамикой. Усредненный показатель итогового балла на конец 2021 года составил 75 баллов. Уравнение линии тренда ($2,0949 \cdot x + 58,444$) позволяет предположить достижение итогового балла одним из двух ПО на основе ТИИ уровня, близкого к расчетному максимуму в 2022 году. По состоянию на конец 2022 года значимых изменений зарегистрировано не было: ПО на основе ТИИ «ИИ-01» зафиксировалось на отметке в 79 баллов, а ИИ-03 достиг значения в 73 балла.

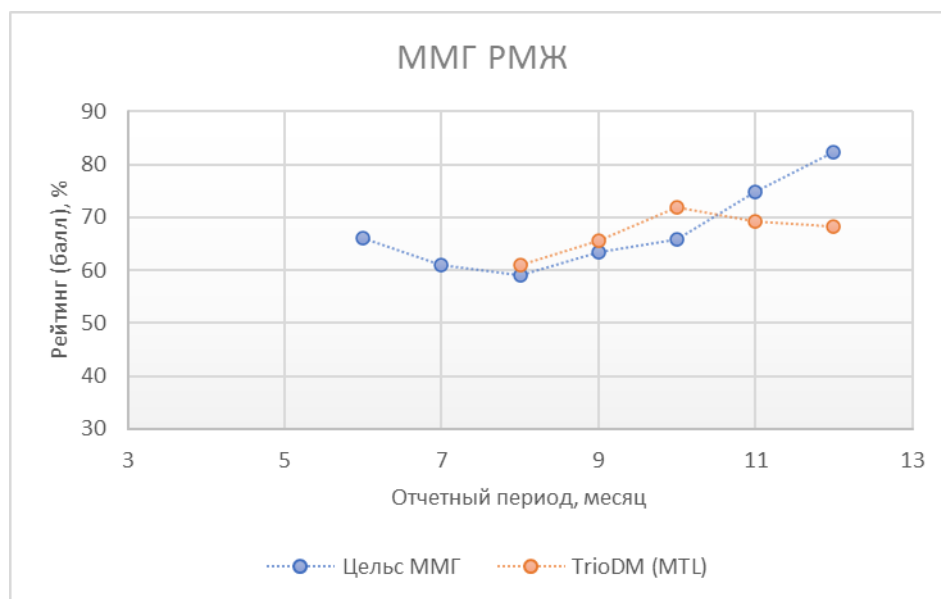


Рисунок 56 – Динамика результатов работы сервисов ММГ РМЖ



Рисунок 57 – Среднее значение для сервисов ММГ РМЖ

Предложенная методика позволила объективно сравнивать между собой решения на основе ТИИ при их эксплуатации на потоке, а открытое сравнение такого ПО – улучшить показатели по отдельным продуктам. Однако, несмотря на то что распределение ПО на основе ТИИ в лидерборде по итоговому баллу коррелирует с выбором врачей-рентгенологов, данная методика учитывает именно интегральную оценку, что не всегда хорошо. Например, увеличение вычислительных мощностей и оптимизация работы ПО на основе ТИИ приводит к повышению технической стабильности и снижает время на обработку исследований, при этом диагностическая точность работы такого ПО может быть невысокой. Напротив, другой разработчик ПО осуществляет дообучение модели в ее основе, показывая существенно большую диагностическую точность, но из-за технических проблем (например, с каналами связи) данное решение уступает предыдущему. При этом для конечного пользователя разница в секундах при обработке исследования может быть незначительна по сравнению с разницей в диагностической точности. ПО этой причине было принято решение о необходимости разработки матрицы зрелости ПО на основе ТИИ, позволяющей визуально оценить все составляющие комплексной оценки.

Второй этап – методика составления матрицы зрелости ИИ-сервисов.

«Качественный этап исследования включал разработку матрицы зрелости ПО на основе ТИИ для сферы здравоохранения на основе литературных данных (теоретический компонент) и собственного практического опыта (эмпирический компонент). Количественный этап исследования включал определение уровня зрелости ПО на основе ТИИ» [269, стр. 80, 81].

Для оценки применимости ПО, основанного на ТИИ, было выделено три основных критерия: процент технологических дефектов (несоответствие БДТ и БФТ), проспективный ROC AUC (диагностическая точность, полученная путем сравнения выводов ПО с заключениями врачей-рентгенологов), а также экспертная клиническая оценка (детальный анализ ограниченного набора исследований для оценки корректности текстовых протоколов, сформированных

ПО, и оценка точности локализации обнаруженных патологий). Эти измеримые и объективные показатели были применены нами для разработки матрицы зрелости – инструмента для оценки уровня развития ПО на основе ТИИ.

Основой для матрицы зрелости служат два параметра, отражающие ценность применения ПО на основе ТИИ: качество (совокупность характеристик, важных для его применения, например, безопасность) и эффективность (мера достижения целей использования ПО). Для визуализации качественной оценки используется четырехсекторная диаграмма (рисунок 58) с координатными осями, где ось X представляет процент технологических ошибок, а ось Y – обратную величину проспективного ROC AUC. Установленные границы «0,81» и «90%» для параметров «ROC AUC» и технологических ошибок соответственно основываются на методических рекомендациях [235] и нормативных актах [268].

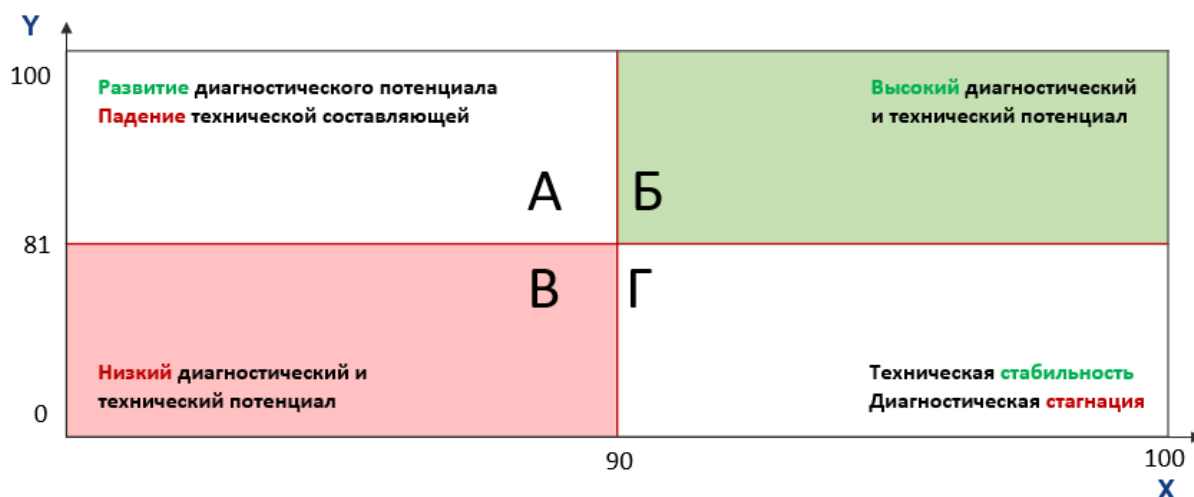


Рисунок 58 – Качественная составляющая матрицы. По горизонтальной оси – процент технологических дефектов; красная черта соответствует пороговому значению в 90%. По вертикальной оси – величина, обратная значению проспективного AUC; красная черта соответствует пороговому значению в 0,81. А, Б, В, Г – наименования зон развития ПО на основе ТИИ

Матрица зрелости ПО на основе ТИИ разделяет программные продукты на четыре категории по уровню их технической стабильности и диагностической эффективности. В первой категории (зона «А») находится ПО, которое еще не обладает технической стабильностью, однако демонстрирует высокие показатели диагностической точности. Вторая категория (зона «В») включает продукты,

которые не достигли желаемых технических и диагностических характеристик. Третья категория (зона «Б») охватывает программы с высоким уровнем технических и диагностических свойств. В четвертой категории (зона «Г») находится ПО, которое имеет невысокую диагностическую эффективность, но технически стабильно.

Качественный аспект матрицы зрелости ПО на основе ТИИ включает в себя оценку стабильности и долговечности. Стабильность – свойство ПО сохранять свое качество при заданных воздействиях на протяжении времени. Такое ПО характеризуется высоким техническим потенциалом (зоны Б, Г);

Долговечность – свойство ПО на основе ТИИ сохранять свое качество при определенных воздействиях и условиях. Такое ПО обладает высоким диагностическим потенциалом (зоны А, В).

Для визуализации эффективности используется пузырьковая диаграмма, где размер пузырька соответствует клинической оценке, а его орбита – разбросу данных. Эффективная составляющая матрицы включает в себя два критерия:

- пригодность – соответствие ПО на основе ТИИ определенным клинико-диагностическим требованиям (диаметр);
- оптимальность – сбалансированная характеристика клинико-диагностического параметра (орбита).

«Матрица зрелости показывает клиническое качество результатов работы ПО на основе ТИИ в зависимости от клинической значимости и технологических дефектов (рисунки 59–61)» [269, стр. 82, 83].

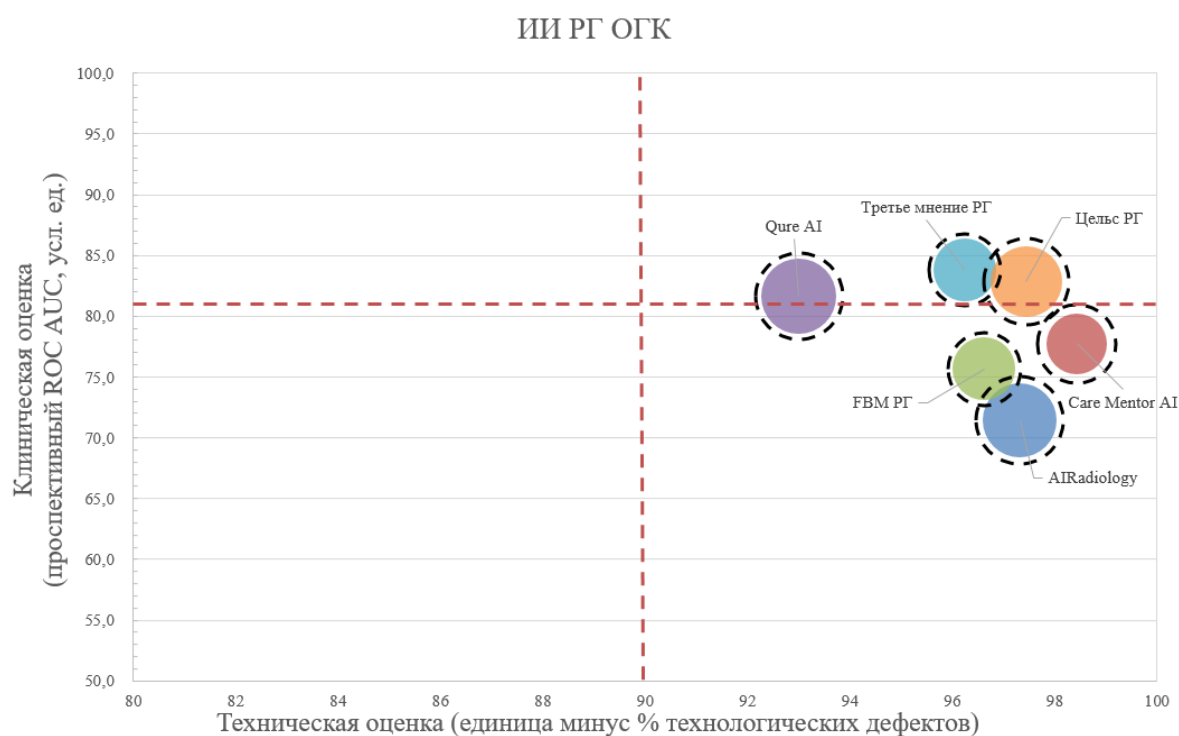


Рисунок 59 – Матрица по направлению «рентгенография (РГ) органов грудной клетки». По горизонтальной оси – процент технологических дефектов (1-тех. деф.,%); по вертикальной оси – величина «ROC AUC»

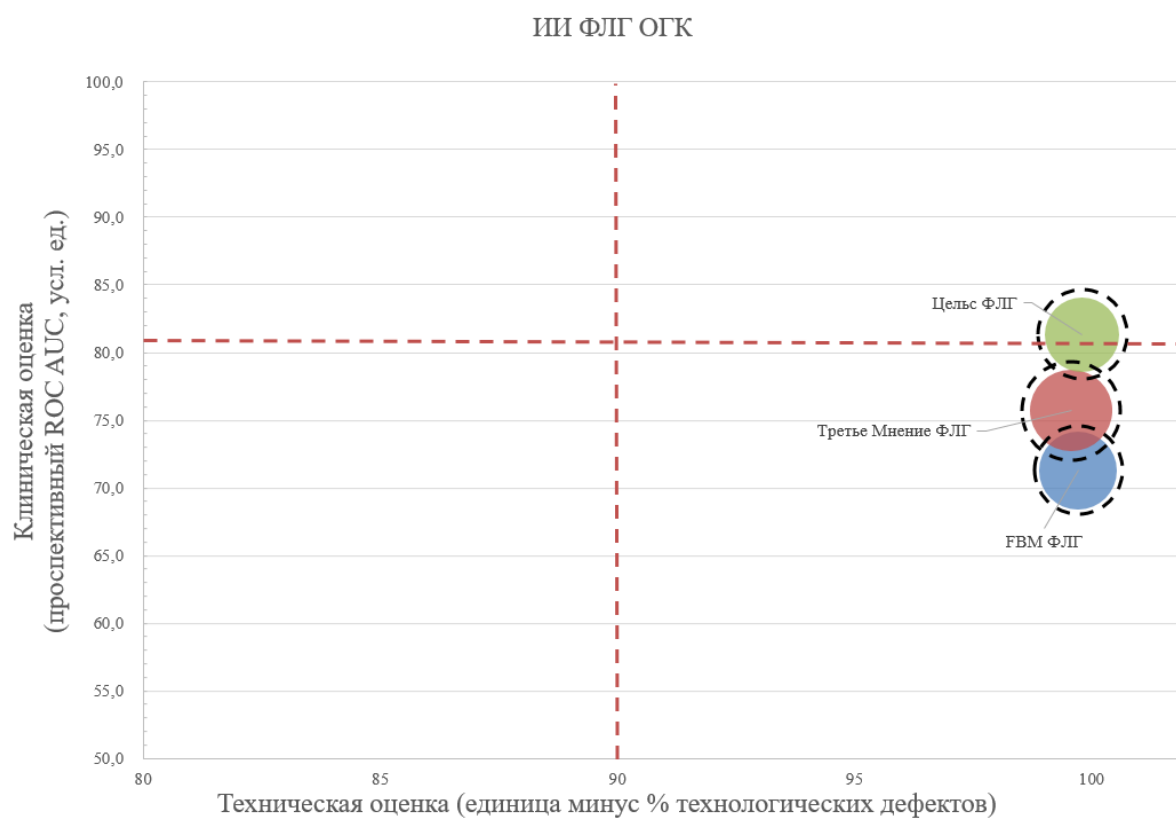


Рисунок 60 – Матрица по направлению «флюорография (ФЛГ)». По горизонтальной оси – процент технологических дефектов (1-тех. деф.,%); по вертикальной оси – величина «ROC AUC»

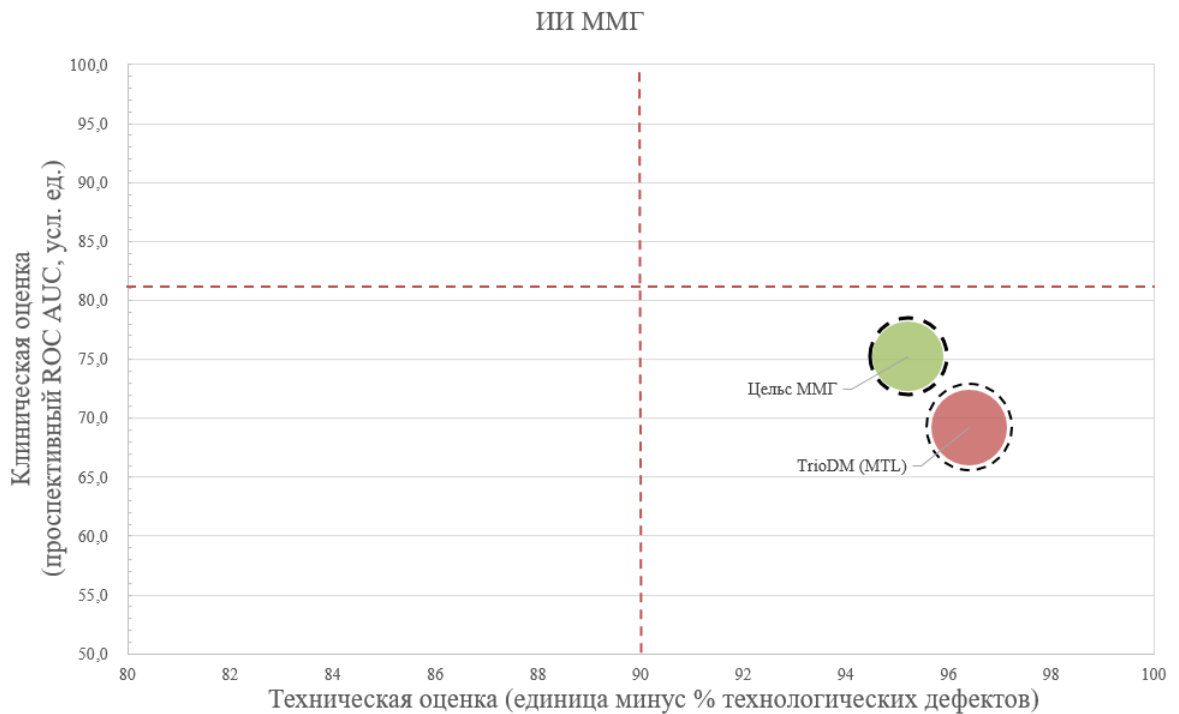


Рисунок 61 – Матрица по направлению «маммография», целевая патология – ЗНО молочной железы. По горизонтальной оси – процент технологических дефектов

Комплексный мониторинг работы программного обеспечения на основе технологий искусственного интеллекта

Для улучшения и контроля качества работы ПО на основе ТИИ в ППАК необходимо анализировать большое количество параметров, такие как время обработки одного исследования, количество проанализированных исследований, аналитика обратной связи в совокупности и по каждому ПО на основе ТИИ. Однако вручную готовить данный анализ слишком трудозатратно и практически не представляется возможным.

Для удобного и своевременного доступа к требуемой оперативной информации был разработан интерактивный интуитивный инструмент «Дашборд», представляющий собой веб-портал по адресу www.dash.mos-med.ai (в настоящее время неактивен, заменен более удобным средством – Yandex Lens) со всей необходимой аналитической информацией. Основными средствами разработки являлся язык программирования Python и открытая библиотека Plotly.

Разумеется, доступ к данной информации ограничен и предоставляется по логину и паролю.

В качестве входных данных используются выгрузка из ЕРИС ЕМИАС с информацией 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, ПО на основе ТИИ и их видам исследований.

Данная разработка полезна как для операционной деятельности, так и для научного анализа результатов работы ПО на основе ТИИ. Примером применения дашборда для операционной деятельности является аналитика медианного времени ожидания и/или обработки одного исследования, это с одной стороны позволяет сравнить ПО на основе ТИИ друг с другом, а с другой, – своевременно выявить технические проблемы, а значит, принять своевременные меры для их устранения.

Кроме этого, инструмент автоматического анализа текстов заключения позволяет реализовать автоматическое сравнение результата работы ПО на основе ТИИ и заключения врача-рентгенолога, в данном случае за истину принимается лейбл, автоматически полученный из заключения врача-радиолога. Пример проспективного сравнения заключений представлен на рисунке 62.

Перспективное сравнение заключения врача-рентгенолога и результаты работы ИИ-сервиса

ИИ-сервисы:

. Количество исследований: 1168153

по порогу maxNPV калибровочного тестирования

по порогу индекса Юдена калибровочного тестирования



Рисунок 62 – Перспективное сравнение заключения врача-рентгенолога и результатов работы ПО на основе ТИИ по различным порогам калибровочного тестирования

Особый интерес представляют собой сравнение калибровочного тестирования и работы ПО на основе ТИИ в перспективном режиме, а также поиск зависимости в точности работы ПО на основе ТИИ от стажа врача или технических параметров оборудования. Последнее поможет определить, требуется ли калибровка ПО на основе ТИИ для каждой модели оборудования или это не приведет к улучшению показателей его работы.

Кроме этой базовой статистики, для пользователя также доступна информация о:

- количестве подключенных к каждому ПО на основе ТИИ медицинских организаций и диагностических устройств;
- количестве проанализированных исследований в день ПО на основе ТИИ и по видам исследований ПО на основе ТИИ;
- количестве результатов, отправленном на анализ и полученном от ПО на основе ТИИ, каждым ПО на основе ТИИ в неделю;
- различных временных показателях;
- показателях диагностической точности;
- данные по обратной связи;

— других количественных показателях, необходимых для контроля качества и принятия управленческих решений.

В настоящее время для расширения возможностей аналитики результатов работы ПО на основе ТИИ «Дашборд» был заменен на инструмент Yandex Datalens, позволяющий более гибко работать с данными. В основе обновленного инструмента заложена логика работы ранее созданного дашборда. Пример отчета представлен на рисунке 63.

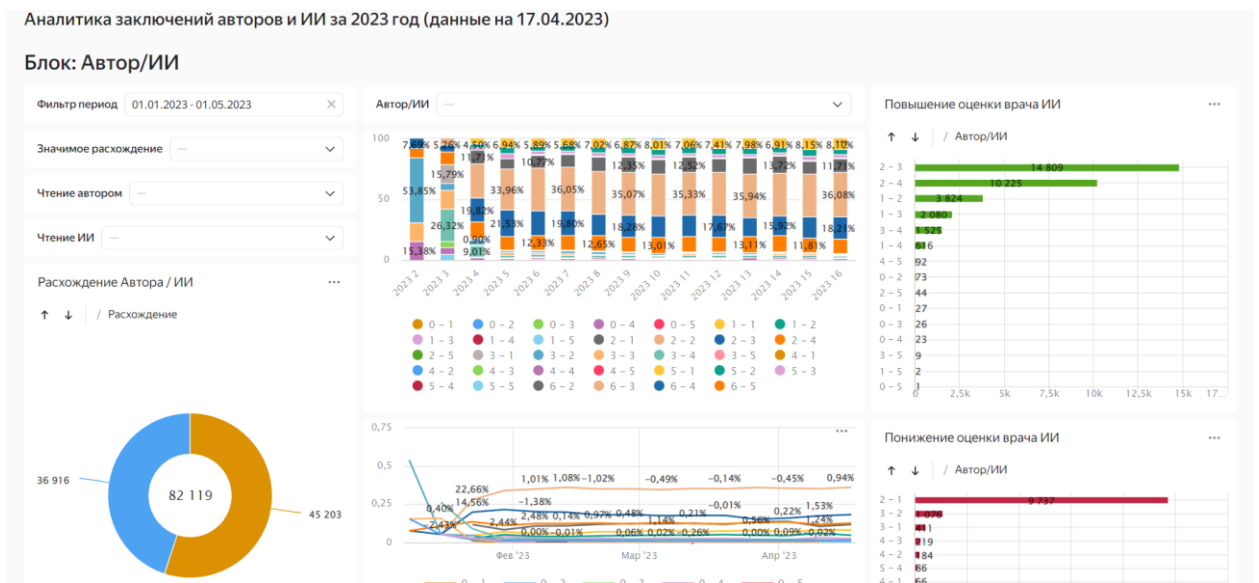


Рисунок 63 – Пример визуализации данных по обработке исследований ПО на основе ТИИ в инструменте Yandex Datalens

Жизненный цикл программного обеспечения на основе технологий искусственного интеллекта как элемент методики комплексной системы оценки

Разработка и внедрение ПО на основе ТИИ для лучевой диагностики требует создания методики отбора такого ПО и контроля его качества при практическом применении. Это обуславливает необходимость создания комплексной методики, которая позволяет оценить ПО в нескольких направлениях, одновременно обеспечивая возможность для разработчиков

полностью раскрывать возможности своих разработок, а пользователям делать осознанный выбор. Первостепенными при внедрении должны быть: научная и клиническая обоснованность, актуальность, удобство использования, снижение трудозатрат рентгенологов и повышение результативности интерпретации исследований, что в конечном итоге способствует достижению цели автоматизации диагностических процессов через применение ТИИ.

В целях эффективного и качественного использования инновационных технологий в области компьютерного зрения для анализа медицинских изображений в масштабах мегаполиса требуется решение вопроса об интеграции ПО на основе ТИИ в рабочий процесс врача-рентгенолога на всех этапах жизненного цикла. До момента начала использования ТИИ врачом-рентгенологом необходимо осуществить контроль корректности работы ПО на основе ТИИ, убедиться в отсутствии технологических и клинических дефектов, способных вызвать существенное затруднение в работе врача-рентгенолога или повлиять на точность диагностики. Однако выполнение контролирующих мероприятий должно быть систематическим, по заранее согласованной схеме. Это является гарантом успешного и продуктивного взаимодействия с разработчиками.

В рамках настоящей работы была разработана и апробирована на практике многоэтапная методика отбора и внедрения ТИИ в практику отделений лучевой диагностики. Данная методика легла в основу жизненного цикла ПО на основе ТИИ (рисунок 64).



Рисунок 64 – Жизненный цикл, инструменты обеспечения контроля качества ПО на основе ТИИ

Оценить эффективность применения жизненного цикла можно по динамике показателей диагностической точности решений на основе ТИИ. Для этого можно воспользоваться тестированием на одном и том же наборе данных, и эти результаты были приведены в таблице 14 для калибровочных тестирований, которые проводились в том числе при смене версионности. При работе на потоке можно отметить влияние на качество работы и на диагностическую точность ряда факторов, поэтому для оценки влияния внедрения многоступенчатой системы контроля качества воспользуемся сопоставлением показателей диагностической точности ПО в ходе проспективной работы. Так, значения AUC при проведении проспективной оценки в 2020 году для ММГ 0,68 (0,65-0,71), для ФЛГ 0,68 (0,662-0,70), для РГ 0,75 (0,74-0,75); в 2023 достигнуты (см. главу 5) следующие показатели: для ММГ по аналогичной методике 0,853 (0,847-0,858), для ФЛГ 0,951 (0,923-0,978), для РГ 0,89 (0,880-0,900). Таким образом, отмечается статистически значимое ($p < 0,05$) увеличение AUC для маммографии на 25,0%, флюорографии – на 41,0%, рентгенографии – на 19,0%. «Предложенная и апробированная в условиях практической работы методика способствует отбору наиболее актуальных и качественных ИИ-решений в лучевой диагностике, что при дальнейшем ее развитии и совершенствовании поможет смягчить недовольство пользователей от несоответствия функционала ИИ ожиданиям,

позволит безболезненно преодолеть точку спада интереса к ИИ и приведет к скорейшему выведению ИИ-решений на безопасный и стабильный уровень производительности в практической деятельности» [261, стр. 170].

Заключение по главе 4

В настоящей главе была решена одна из самых комплексных задач настоящего исследования: разработана система контроля качества ТИИ. Была создана и апробирована методика многоступенчатого тестирования и оценки качества ПО на основе ТИИ, состоящая из ряда этапов: самотестирования; функционального, калибровочного и контрольно-технического тестирований; технологического мониторинга. Внедрение поэтапной методики позволило увеличить диагностическую точность решений на основе искусственного интеллекта для маммографии – на 25,0%, флюорографии – на 41,0%, рентгенографии – на 19,0%. Таким образом, методика продемонстрировала свою эффективность и масштабируемость и легла в основу методики сравнения решений между собой, а также стала основой жизненного цикла ПО на основе ТИИ.

Для реализации методики потребовалась реализация комплекса инструментов: анализ текстовых протоколов; комплекс инструментов для расчета показателей диагностической точности; инструмент для проведения мониторинга, контроля качества при смене версионности; инструмент оценки диагностической точности врача-рентгенолога. Таким образом, в рамках настоящей работы был разработан целый набор инструментов, охватывающий самый широкий спектр задач, связанных с внедрением технологий на основе искусственного интеллекта в медицинскую диагностику.

Завершает текущую главу инструмент поддержки управленческих решений о выборе программных средств на основе ТИИ. Предложено два подхода, базирующихся на ранее разработанных инструментах и методике жизненного

цикла: итоговый балл как интегральный способ оценки и матрица зрелости как инструмент оценки цифровой зрелости. На основе предложенной системы контроля качества можно принимать решение о целесообразности дальнейшего использования ПО на основе ТИИ.

В третьей и четвертой главе представлены основные организационно-технические моменты, позволяющие провести отбор программных решений на основе ТИИ, внедрить в рабочий процесс врача-рентгенолога и обеспечить контроль безопасности и качества их работы, однако до сих пор не был освещен вопрос организации и проведения массовых профилактических и диагностических исследований с применением ТИИ. Эта модель изложена в следующей главе.

Глава 5. МОДЕЛЬ ОРГАНИЗАЦИИ И ПРОВЕДЕНИЯ МАССОВЫХ ПРОФИЛАКТИЧЕСКИХ ИССЛЕДОВАНИЙ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Организация и проведение массовых профилактических исследований с применением ТИИ базируется на применении ИИ-алгоритмов в качестве первого чтения лучевых исследований. Это подразумевает безусловное и тотальное направление результатов профилактических исследований на ИИ-сервисы. В результате формируется заключение и дополнительная графическая серия с результатами работы ИИ, которая становится доступна врачу-рентгенологу. За счет быстрой обработки исследований ИИ-сервисами, результаты фиксируются в радиологической системе до открытия исследования врачом-рентгенологом. Таким образом, врач-рентгенолог может видеть в исследовании как нативное изображение, так и результаты работы ИИ. Этот подход является универсальным, применимым как при описании ММГ, так и РГ/ФЛГ-исследований. Единственная разница будет в особенностях оптимальной настройки алгоритмов ИИ, поэтому ниже рассмотрим применение программных средств на основе ТИИ в отдельности.

5.1. Особенности применения программного обеспечения на основе технологий искусственного интеллекта для маммографии

Нормативно-правовые акты предусматривают обязательный двойной анализ результатов профилактической маммографии, при котором каждое исследование должно быть оценено двумя врачами-рентгенологами независимо друг от друга. Эффективность такой меры подтверждена многочисленными исследованиями, демонстрирующими повышенную частоту выявления ЗНО при использовании двойного просмотра. В предлагаемой модели процесс начинается с назначения профилактической маммографии лечащим врачом, за которым следует

выполнение исследования рентгенолаборантом и передача результатов в цифровом виде в архив медицинских изображений. Далее, первичный анализ результатов проводится с использованием специализированного ПО на базе ТИИ, зарегистрированном в качестве медицинского изделия. В результате работы ПО в информационной системе формируется электронная медицинская запись⁴, которая включает: а) серию изображений с выделением подозрительных участков при помощи графических обозначений или тепловых карт; б) структурированный протокол и руководство для пользователя. Дальнейший просмотр осуществляет врач-рентгенолог, который на основе полученной от ПО информации формирует электронный медицинский документ – протокол описания и заключения.

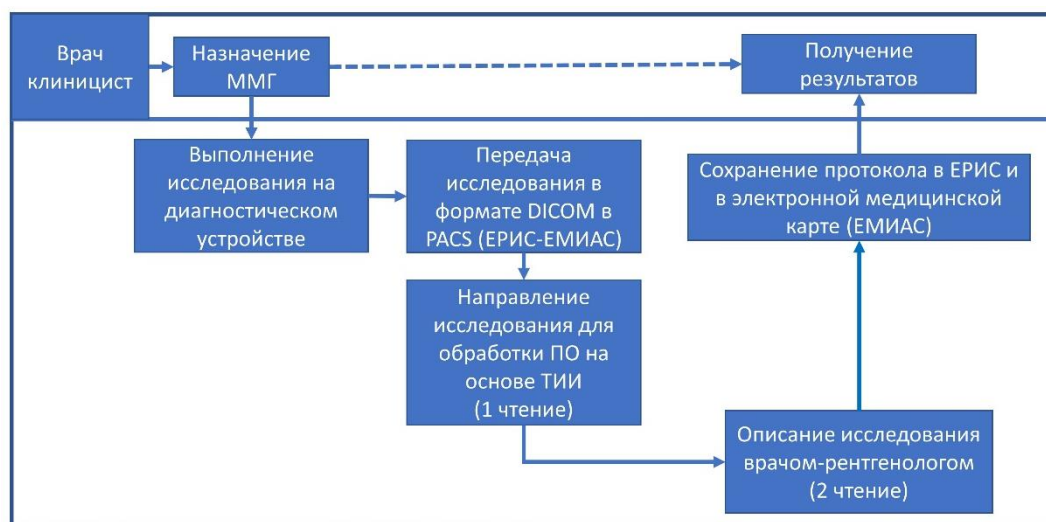


Рисунок 65 – Модель применения программного обеспечения на основе технологии искусственного интеллекта для маммографии

В начальной фазе исследования был выполнен тест на ограниченной выборке из 100 цифровых маммограмм, среди которых 50 содержали патологии. Классификация маммограмм в категории «норма» и «патология» осуществлялась

⁴ Медицинские записи, полученные в автоматическом режиме в результате применения медицинских изделий, допущенных к обращению, не требуют подписания электронной подписью медицинского работника (согласно приказу Министерства здравоохранения Российской Федерации от 07.09.2020 № 947н «Об утверждении Порядка организации системы документооборота в сфере охраны здоровья в части ведения медицинской документации в форме электронных документов»).

на основе согласия между двумя квалифицированными врачами-рентгенологами с опытом работы более пяти лет и подтверждения наличия патологии через трепан-биопсию. Производительность анализируемого ПО показана на рисунке 66, где для разных пороговых значений представлена пара показателей диагностической точности (чувствительность и специфичность).

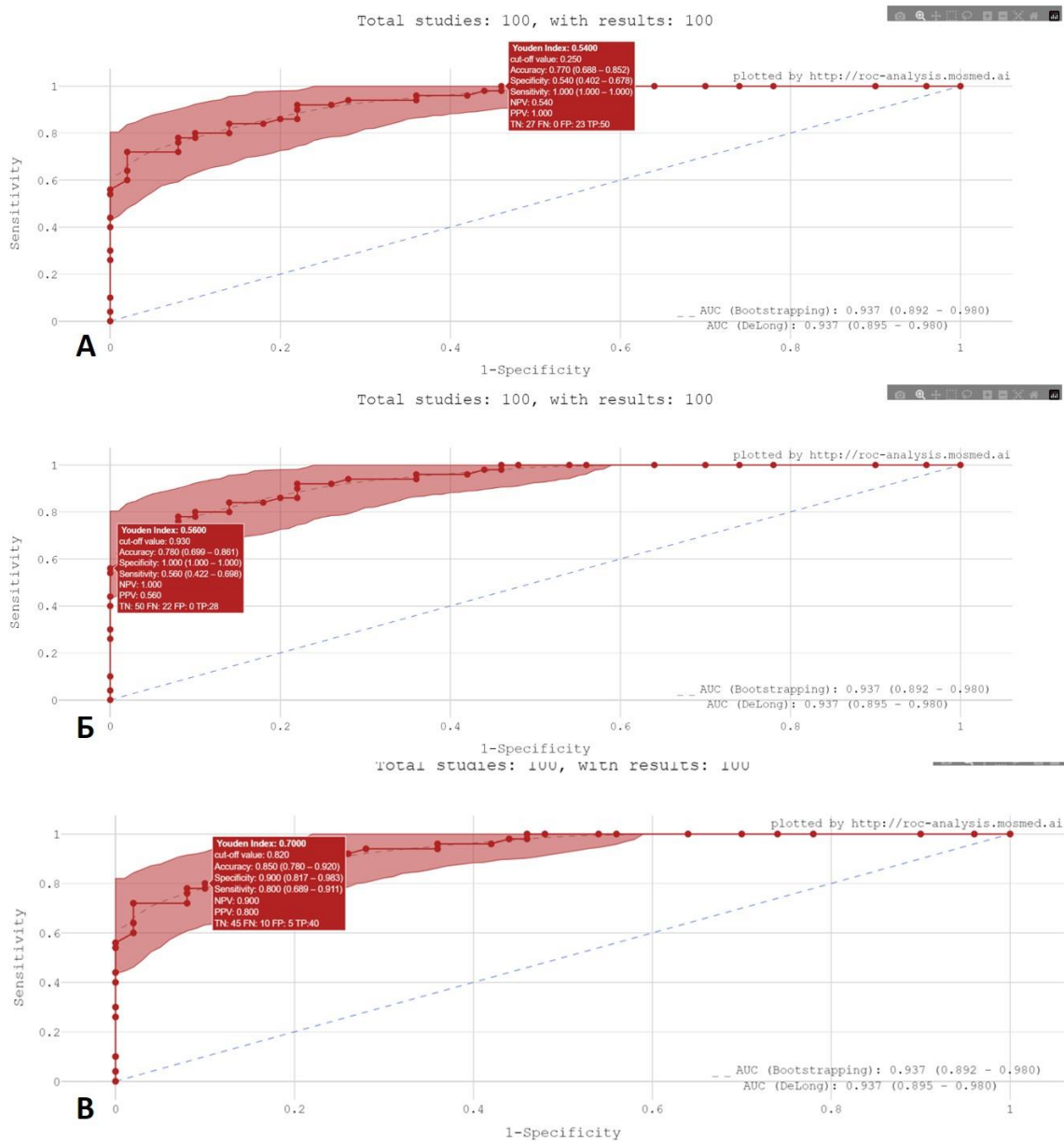


Рисунок 66 – Характеристическая кривая ПО на основе ТИИ. Заливкой показан 95% доверительный интервал. Отдельно выделены экспериментальные значения, соответствующие 100,0% чувствительности (А), 100,0% специфичности (Б) и 80,0% чувствительности (В)

Распределение ММГ по группам (ИП, ИО, ЛО, ЛП) в зависимости от выбранного порогового значения представлено в таблице 18 [270, стр. 99].

Таблица 18 – Четырехпольная таблица для разных пороговых значений

Группа	«Не норма»		«Норма»	
Количество исследований	50		50	
Результат	Истинно положительный	Ложноположительный	Истинно отрицательный	Ложноотрицательный
100% чувствительность	50	23	27 (54,0%*)	0
100% специфичность	28 (56,0%**)	0	50	22
Сбалансированная чувствительность и специфичность	40 (80,0%**)	5	45 (90,0%*)	10
* Процент истинно отрицательных результатов рассчитан относительно объема группы «норма». ** Процент истинно положительных результатов рассчитан относительно объема группы «не норма»				

Установка порога в 0,93 обеспечила 100% специфичность, исключив ложноположительные срабатывания, и позволила ИИ точно идентифицировать 56% патологий из группы «патология». При пороговом значении 0,25, обеспечивающем 100% чувствительность, было достигнуто отсутствие ложноотрицательных результатов и верно определено 54% случаев из группы «норма», что демонстрирует возможности применения такого ПО на основе ТИИ в диагностике в качестве инструмента автономной сортировки исследований.

Установив порог срабатывания в соответствии с методикой оптимизации индекса Юдена (порог = 0,82), были получены следующие результаты: чувствительность достигла 80%, при этом 45 исследований из категории «норма» (90% от общего числа класса) были верно классифицированы как отрицательные,

а 40 исследований из категории «не норма» (80% от общего числа класса) – как положительные. С такими параметрами настройки ПО на основе ТИИ показало чувствительность на уровне 80% (95% ДИ: 68,9–91,1%), специфичность – 90% (95% ДИ: 81,7–98,3%), а общая точность составила 85% (95% ДИ 78,0–92,0%).

Программное обеспечение на основе ТИИ, применяемое для описания и анализа данных маммографических исследований (в рамках услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта»), позволяет при различных настройках пороговых значений достигать или превышать показатели чувствительности, специфичности и точности, наблюдаемые при двойном чтении маммограмм врачами-рентгенологами.

В Российской Федерации, где ежегодно проводится около 8,2 миллиона профилактических маммографических исследований [271], двойное чтение требует значительных временных, кадровых и финансовых ресурсов. Внедрение ТИИ для первичного анализа маммограмм может сократить эти затраты, сохраняя или даже улучшая качество диагностических процедур. Предложены два подхода к настройке ТИИ для оптимизации процесса первичного анализа маммограмм, каждый из которых предусматривает свой метод определения категории изображений и уровня «уверенности» алгоритма в своих выводах. Таким образом, результатом работы ПО на основе ТИИ может быть электронная медицинская запись, содержащая заключение о категории изображения («норма» или «не норма») или категории с указанием степени своей «уверенности» в полученном результате. Общая модель представлена на рисунке 67.

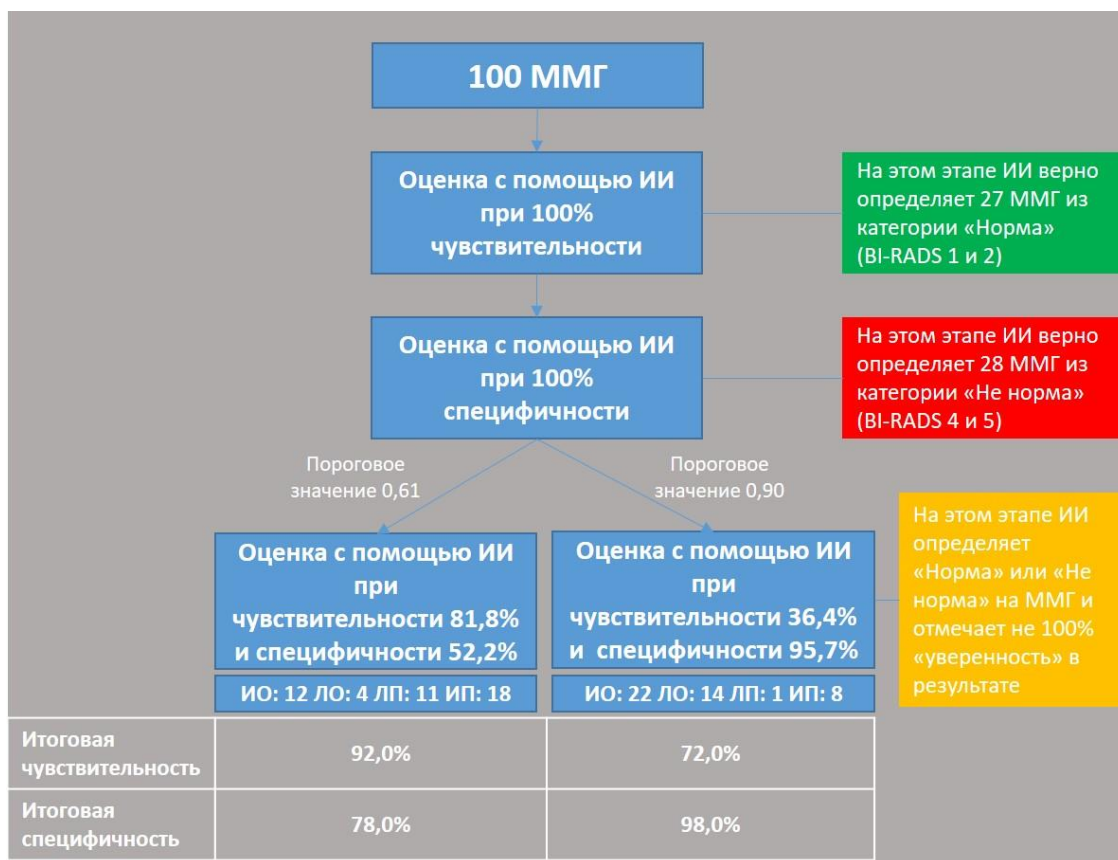


Рисунок 67 – Модель подхода к первому описанию ММГ с помощью ИИ

Выше были рассмотрены показатели диагностической точности ПО на основе ТИИ в зависимости от пороговых значений. Были установлены пороговые значения для достижения максимальной чувствительности и специфичности (0,25 и 0,93 соответственно). Исходя из этих данных, предложено классифицировать исследования в три категории, отражающие различные уровни доверия ТИИ к своим выводам: 1) «зеленый коридор» для значений от 0 до 0,25, указывающих на «норму» с полной уверенностью; 2) «красный коридор» для значений от 0,93 до 1, означающих «не норму» с абсолютной уверенностью; и 3) «желтый коридор» для значений между 0,25 и 0,93, где классификация менее уверена.

Можно рекомендовать включать предсказанное значение и соответствующий цветовой коридор в протокол описания ММГ, подготовленный ПО на основе ТИИ. Это даст врачам-рентгенологам информацию о надежности результатов, повышая их бдительность при анализе случаев из «желтого коридора» и способствуя росту доверия к ТИИ за счет ее возможности точно

классифицировать изображения. Второй подход подразумевает возможность корректировки порогового значения для достижения определенного баланса чувствительности и специфичности в зависимости от клинической задачи, минимизируя число ЛП или ЛО результатов. В зависимости от порогового значения, пара чувствительность-специфичность может быть от 92,0%-78,0% (пороговое значение 0,61) до 72,0%-98,0% (пороговое значение 0,90) [270].

Результаты, полученные на этом этапе, указывают на потенциал автономного применения ПО на основе ТИИ для первого прочтения ММГ. Однако для более эффективного классифицирования исследований на «норму» и «не норму» требуется донастройка ПО.

5.2. Оценка эффективности автоматизации процессов организации и проведения маммографии

В этом параграфе приведены данные за период с 22.07.2020 по 29.12.2022, в рамках которого было оценено 675 200 маммографических исследований. Были исключены отдельные исследования, в т.ч. те, которые не имели заключения. Возраст пациентов составил от 19 до 155 лет ($58,70 \pm 11,39$ лет). Распределение по полу – 480 (0,07%) мужчин, 674 664 женщин (99,92%). В исследовании участвовали 123 медицинские учреждения и 534 врача-рентгенолога. При расчете показателей диагностической точности осуществлялась бинарная классификация: BI-RADS по диагностической шкале – 3, 4, 5, 6 относились к патологии, к норме относились классы 1 и 2. Значения AUC, рассчитанные для всего объема проспективных исследований: ИИ-01 – 0,72, ИИ-03 – 0,69, ИИ-04 – 0,63. Полученные по общему потоку исследований значения оказались довольно низкими. Однако в рамках исследования, как могли видеть в предыдущих главах, ПО на основе ТИИ дорабатывалось и совершенствовалось, повышая в том числе показатели диагностической точности. Из всех рассмотренных решений доработку в модель вносили разработчики ИИ-01, которым удалось повысить

AUC с 0,70 до 0,74. Дальнейшая доработка ПО на основе ТИИ была осуществлена разработчиками уже в 2023 году, на рисунке 68 представлена характеристическая кривая для обновленной версии ИИ-01.

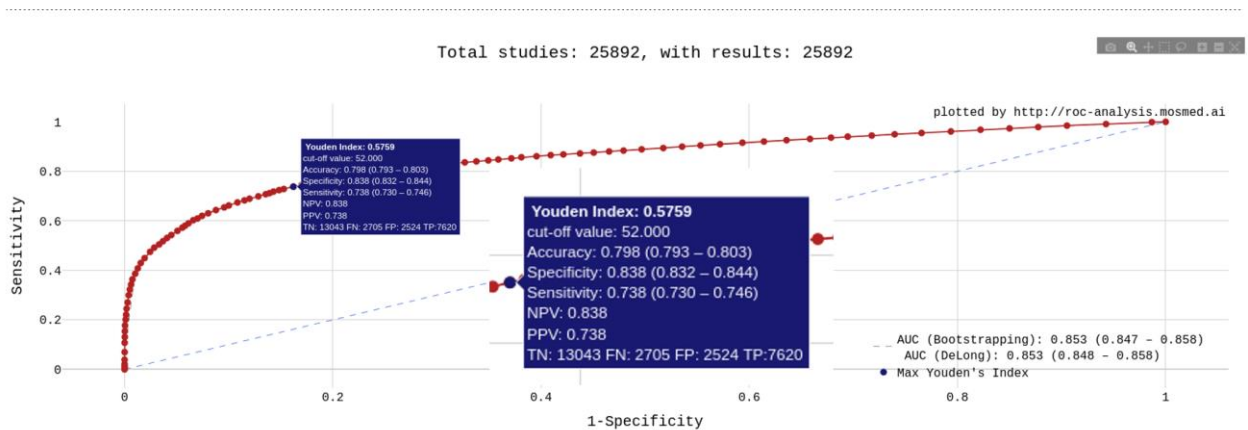


Рисунок 68 – Характеристическая кривая для ПО на основе ТИИ ИИ-01 ММГ, версия, актуальная на 2023 год. К патологии отнесены категории BI-RADS 3–6

Как видно из представленных выше данных, диагностическая точность (AUC) ПО на основе ТИИ в ходе проспективного этапа достигла максимально 0,85 и показало низкие результаты, не соответствующие критериям прохождения в 0,88. Далее были проанализированы причины расхождения показателей и выявлены следующие причины: в ходе проспективного исследования использовались в качестве истинных значений категории BI-RADS, извлеченные из заключения врача-рентгенолога. Первая проблема связана с тем, что врачи, как было показано выше, тоже ошибаются; вторая проблема – с тем, что было принято решение относить категорию BI-RADS 3 к патологии. С учетом того, что процент исследований с патологией (BI-RADS 4–6) составляет менее 5%, объемная категория BI-RADS 3 (6,76% от общего потока исследований) вносит существенный вклад в смещение оценки. Для проверки предложенных гипотез было проведено исследование, в рамках которого осуществлена переклассификация: BI-RADS 1–3 рассматривалась как норма, а BI-RADS 4–6 как патология (характеристическая кривая представлена на рисунке 69).

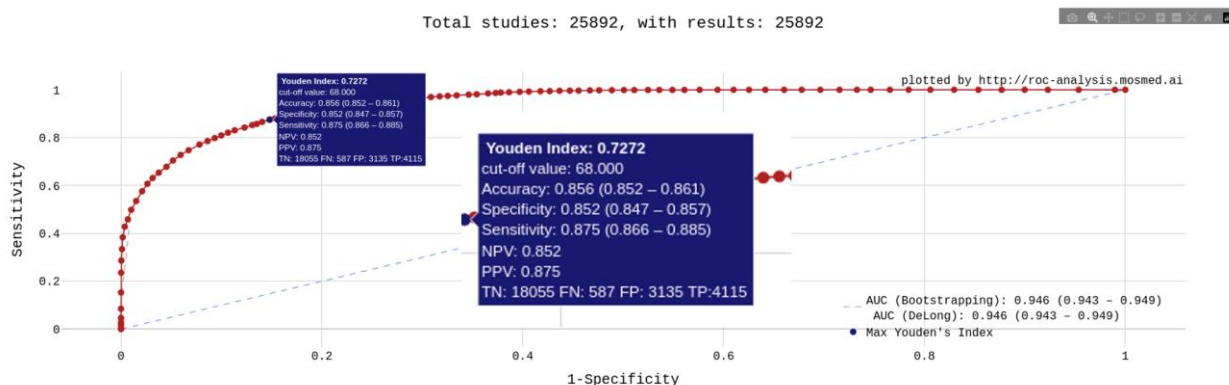


Рисунок 69 – Характеристическая кривая для ПО на основе ТИИ ИИ-01, версия, актуальная на 2023 год. К патологии отнесены категории BI-RADS 4–6

Обновленные данные демонстрируют высокие значения показателей диагностической точности: AUC – 0,946 (95% ДИ 0,943–0,949), чувствительность – 0,875 (95% ДИ 0,866–0,885), специфичность 0,852 (95% ДИ 0,847–0,857) и точность 0,856 (95% ДИ 0,852–0,861). Показатели AUC и чувствительность превысили таковые для «среднего» врача, что позволяет реализовать модель применения ПО на основе ТИИ в качестве первого чтения маммографического исследования.

Результаты работы программного обеспечения на основе технологий искусственного интеллекта в составе медицинской услуги

1. Сравнение оценок маммографии, выставленных ПО на основе ТИИ, врачом-рентгенологом и экспертом, по шкале с категориями 1–5 BI-RADS и бинарным шкалам

В таблице В1 (Приложение В) представлены количественные данные по категоризации BI-RADS, выставленной как врачом, так и ПО на основе ТИИ, по всем 9684 исследованиям. Также представлены результаты врачебной экспертизы выборки из 45 исследований. Изучение средневзвешенных значений и стандартных отклонений выявило отсутствие статистически значимых различий между оценками врача и ИИ по всей выборке. Однако ПО на основе ТИИ чаще

обозначало наличие патологии в исследованиях, чем врач, что демонстрируется в таблице В2 (Приложение В). В то же время в выборке из 45 исследований ситуация обратная из-за специфики формирования выборки. Анализ согласованности между оценками врача и результатами работы ПО на основе ТИИ по 9684 маммографиям выявил следующее: сходство в оценках по шкале BI-RADS составило 43,89%, а по двум бинарным шкалам – 84,10% и 80,69% соответственно (таблица В3, Приложение В). Было выявлено преобладание ситуаций, когда ПО при анализе ММГ повышало категорий BI-RADS по сравнению с оценками врачей. Для отдельных категорий это различие достигло 76,44%, для первой бинарной шкалы – 95,97%, а для второй – 97,34%. Сравнивая результаты ИИ с экспертными оценками, совпадения наблюдались лишь в 33% и 25% случаев соответственно. Тем не менее согласованность между врачом и экспертом была выше для обеих бинарных шкал. Коэффициент Каппа Коэна, используемый для оценки согласованности, показал низкое совпадение между всеми оценками, считаясь «незначительным согласием».

2. Характеристики точности ПО на основе ТИИ относительно врача-рентгенолога и эксперта

Производительность ПО на основе ТИИ была изучена с использованием метрик AUC ROC, чувствительности, специфичности и общей точности в сравнении с оценками врача и эксперта (таблица В4, Приложение В). AUC ROC, рассчитанный для ПО, при сопоставлении с врачебными оценками на всем объеме исследований (9684) показал небольшие различия для бинарных шкал (0,792 и 0,82) и уменьшение до 0,65 для полной шкалы BI-RADS. При сравнении с экспертными оценками значения AUC ROC для ПО были значительно ниже (рисунок 70). Точность врача, измеренная по отношению к эксперту, демонстрирует высокие показатели AUC ROC для бинарных шкал и более низкие для шкалы BI-RADS. Однако расчеты для некоторых категорий проводились на ограниченной выборке, что может влиять на общую достоверность результатов.

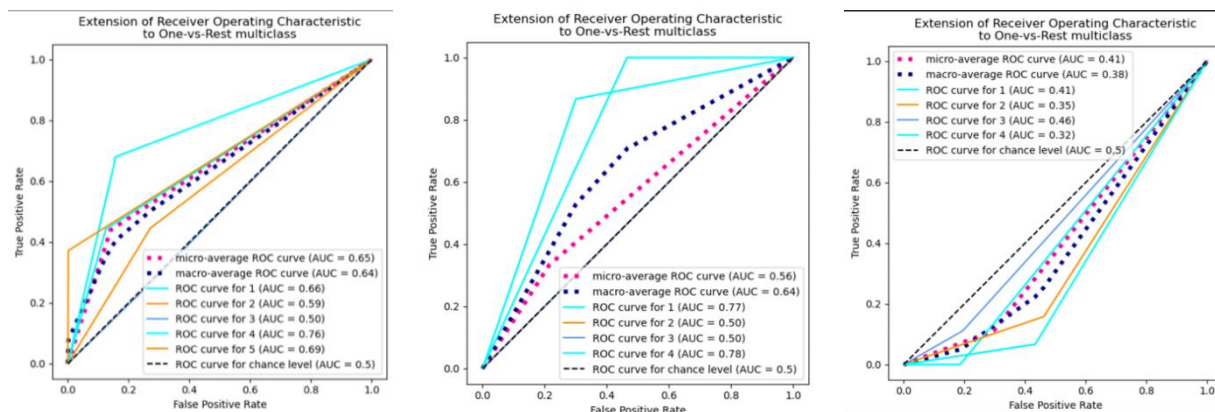


Рисунок 70 – Графики AUC ROC для оценки точности ПО на основе ТИИ относительно врача (1), врача относительно эксперта (2) и ПО на основе ТИИ относительно эксперта (3)

В таблице И5 (Приложение В) представлено количество ложноположительных, ложноотрицательных, истинно положительных и истинно отрицательных оценок ММГ, ПО относительно врача-рентгенолога и эксперта.

Программное обеспечение на основе ТИИ демонстрировало значительно более высокий процент ложноположительных оценок (15,26% и 18,8% по бинарным шкалам) при сравнении с врачебными оценками, в то время как ложноотрицательные результаты были ниже (0,64% и 0,51%). В сравнении с экспертными оценками, ИИ показывал больше ложноотрицательных результатов (31,11% и 48,28%), нежели ложноположительных (28,89% и 34,48%). Врачебные оценки в сравнении с экспертными показали больше ложноположительных, чем ложноотрицательных результатов [259]. Совпадения между оценками ИИ, врача и эксперта в исследованиях при пересмотре представлены на рисунке 71.

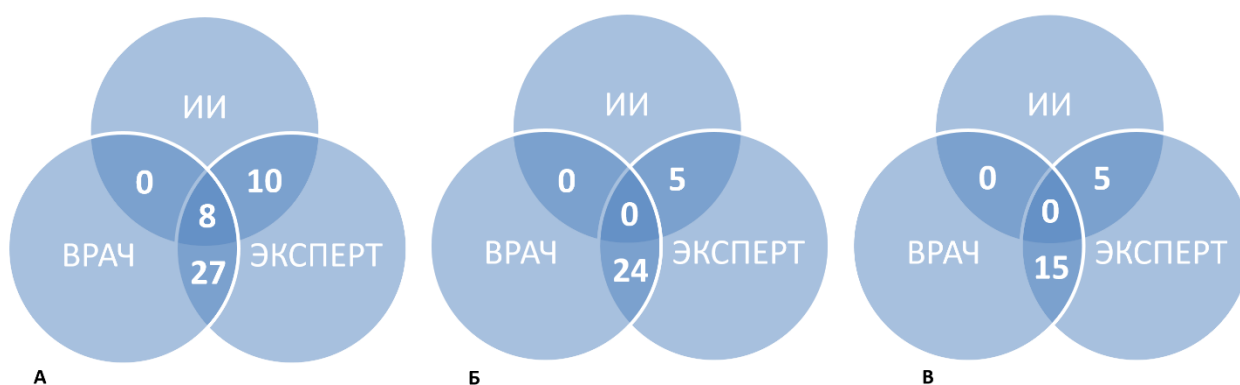


Рисунок 71 – Совпадения оценок ПО на основе ТИИ, врача-рентгенолога и эксперта: в 45 исследованиях, оцененных относительно бинарной шкалы 1 (А); в 29 исследованиях, оцененных относительно бинарной шкалы 2 (Б); в 45 исследованиях, оцененных относительно полной шкалы BI-RADS (В)

Когда была проведена классификация BI-RADS категорий 1–3 как «норма» по бинарной шкале 1, наблюдалось увеличение случаев согласия между врачом, ИИ и экспертом. По этой шкале эксперт согласился с оценкой ИИ в 22,2% случаев, а с врачом – в 60% случаев из 45. В случае с бинарной шкалой 2 эксперт согласился с ИИ в 17,2%, а с врачом – в 82,8% случаев из 29. По шкале BI-RADS согласие с ИИ было в 11,1%, с врачом – в 33,3%, и в 55,6% случаев эксперт не согласился ни с одной из сторон.

Полученные результаты демонстрируют склонность ПО на основе ТИИ к ложным положительным срабатываниям, однако выявленный на столь малой выборке случай корректного срабатывания ПО на основе ТИИ и ложноотрицательное заключение врача-рентгенолога демонстрируют потенциал к снижению случаев пропуска патологии врачом и, следовательно, к повышению качества диагностики. Этот факт подтверждается ранее изложенным (глава 3) сравнением показателей диагностической точности врача и ПО, в рамках которого одно из ПО превосходило «среднего врача» на 3% по точности. Таким образом, очевидна перспективность дальнейшего применения ТИИ для цифровой маммографии в качестве первого чтения.

**Модель применения программного обеспечения на основе технологий
искусственного интеллекта в качестве медицинской услуги
с тонкой настройкой**

После получения объема исследований в размере 25 тысяч единиц была построена характеристическая кривая с площадью под кривой 0,95. Ниже приведена характеристическая кривая (рисунок 72), построенная для ПО на основе ТИИ, работающего в рамках данной услуги.

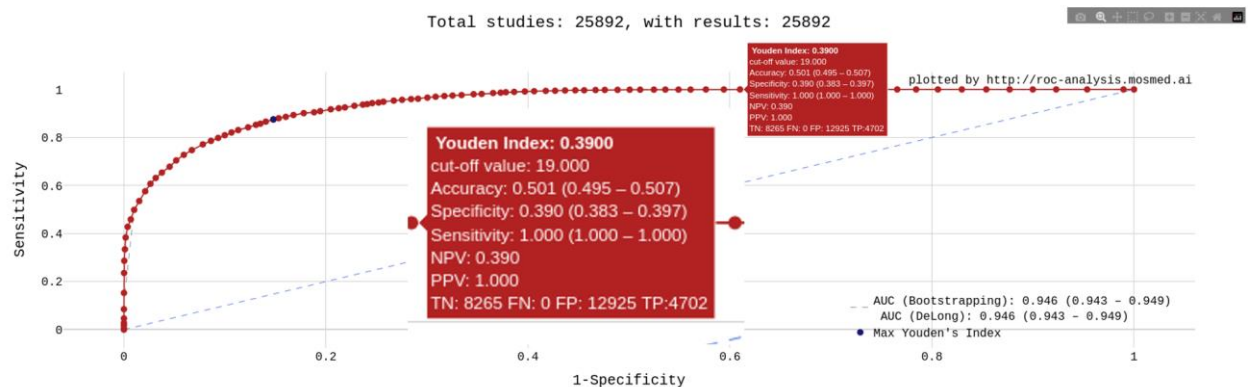


Рисунок 72 – Характеристическая кривая для ПО на основе ТИИ. Синим отмечена точка с оптимальной сбалансированной настройкой. Красным отмечена точка и вынесены показатели диагностической точности, соответствующие 100% чувствительности

Как следует из приведенной характеристической кривой, возможна настройка ПО на 100% чувствительность, при этом корректно классифицировано будет 8265 исследований, что составляет 32% от всего потока. Эта модель является не совсем целевой для маммографии, но может иметь большое практическое значение для других профилактических исследований.

5.3. Особенности применения программного обеспечения на основе технологий искусственного интеллекта для рентгенографии/флюорографии органов грудной клетки

Для повышения эффективности организации здравоохранения при проведении массовых профилактических исследований предлагается добавить две услуги: 1) цифровая флюорография легких, включающая автоматизированную обработку и анализ результатов с помощью ПО на основе ТИИ, сертифицированного как медицинское изделие; 2) анализ и трактовку данных цифровой флюорографии специалистом-рентгенологом. Процесс предусматривает следующие этапы: назначение и запись на исследование, его проведение, а затем автоматизированный анализ с использованием ТИИ, результаты которого в случае отсутствия патологии (т.е. «норма») сохраняются в медицинской информационной системе медицинской организации как электронная медицинская запись (согласно приказу Минздрава России от 07.09.2020 № 947н «Об утверждении Порядка организации системы документооборота в сфере охраны здоровья в части ведения медицинской документации в форме электронных документов») – это и будет результат оказания услуги № 1.

В случае наличия любых отклонений от «нормы» задействуется услуга № 2 – результаты флюорографии направляются на интерпретацию и описание врачом-рентгенологом. Ее результатом является протокол в соответствии с приказом Минздрава России от 09.06.2020 № 560н «Об утверждении Правил проведения рентгенологических исследований» – это результат оказания услуги № 2.

Ответственность

Проблема ответственности при внедрении технологий искусственного интеллекта в медицину вызывает активные обсуждения. Опираясь на опыт проведения Московского эксперимента, предложено, что ответственность за диагностическую точность программного обеспечения на базе ИИ несут как его

разработчики, так и организации, осуществлявшие технические и клинические испытания для получения лицензии на ПО в качестве медицинского изделия.

На данном этапе невозможно достигнуть 100% точности работы ПО на основе ТИИ, по этой причине необходимо разделение ответственности с организацией, проводящей технические и клинические испытания ПО. Такой подход направлен на максимальное соблюдение прав и интересов пациентов (прежде всего в аспекте безопасности медицинской помощи) в отличие от европейской сертификации ПО на основе ТИИ в качестве медицинского изделия, при которой достаточно провести аудит системы обеспечения качества продукции на стороне разработчика, а также представить результаты клинического исследования, выполненного самим же разработчиком. В отечественной системе регистрации медицинского изделия предполагается обязательное проведение клинических испытаний сторонней организацией с целью подтверждения заявленных показателей диагностической точности. Таким образом, в России ПО на основе ТИИ с завышенными показателями не может быть сертифицировано как медицинское изделие. Более жесткие требования компенсируются тем, что показатели диагностической точности и другие технические параметры будут сопоставляться с пороговыми значениями, заданными самим разработчиком в приложенной технической документации. На данный момент существует только один минимальный порог для показателей диагностической точности – 0,81. При переходе полностью на автономный искусственный интеллект в отдельных медицинских услугах необходимо задать минимальный порог показателей диагностической точности для допуска данного ПО к клиническому применению. В рамках настоящей работы предлагаем установление 100,0% чувствительности для ПО на основе ТИИ для автономной сортировки результатов профилактической флюорографии на категории «норма» и «не норма». Такой подход позволит исключить пропуск патологии на потоке данных.

При внедрении в клиническую практику автономного ПО на основе ТИИ необходимо обеспечить входной контроль качества. Для этого предполагается

направить на анализ выборку исследований, репрезентативную целевому потоку исследований. Это позволит убедиться в технической возможности обработки исследований, а также осуществить контроль показателей диагностической точности. Следующим, критически важным шагом является внедрение обязательных проверочных процедур, таких как аудиты, для подтверждения эффективности ПО на основе ТИИ. Это позволит предоставлять разработчикам своевременную обратную связь для дальнейшего совершенствования их продуктов. Очевидно, что реализация таких инициатив потребует изменений в регуляторных нормах. Автоматизация сортировки результатов профилактических рентгенологических исследований позволит сохранить и улучшить массовые скрининги заболеваний, расширить возможности отечественного медицинского оборудования за счет внедрения ТИИ, а также переориентировать значительные финансовые и кадровые ресурсы на решение более сложных диагностических задач.

Проведенное исследование подтверждает, что использование ТИИ для автоматического анализа результатов цифровой флюорографии является перспективным. Предпочтительным вариантом является применение ПО на основе ТИИ для выявления случаев, классифицируемых как «норма», где дополнительный анализ врачом-рентгенологом при высокой чувствительности алгоритма не требуется. Однако результаты, обозначенные как «не норма», должны направляться на детальное описание врачом-рентгенологом.

Предлагается модель оказания медицинской помощи, включающая автоматическую классификацию результатов профилактической флюорографии, обновление медицинских услуг, методы гарантии безопасности и качества медицинской помощи с применением ТИИ, а также решение вопросов ответственности. Эта система позволит увеличить охват населения профилактическими исследованиями, улучшить доступность и качество диагностики. В основу разработки модели легли два ПО на основе ТИИ для

анализа цифровых флюорографий, обозначенных как «ИИ-01» и «ИИ-02» (торговые наименования также анонимизированы).

Модель применения (рисунок 73) подразумевала обработку РГ/ФЛГ-исследований ПО на основе ТИИ с настроенным алгоритмом на 100% чувствительность и направление на анализ врачу-рентгенологу только тех исследований, для которых была выявлена патология.



Рисунок 73 – Модель применения программного обеспечения на основе технологии искусственного интеллекта для рентгенографии/флюорографии органов грудной клетки

Эта модель была проверена на перечне исследований, состоящем из 97 цифровых флюорографий, среди которых 47 – с патологией. Классификация на категории «норма» и «не норма» базировалась на едином мнении двух опытных врачей-рентгенологов (стаж работы более 5 лет). Анализировались такие патологии, как очаг затемнения, плевральный выпот, пневмоторакс, ателектаз и другие. При этом соотношение типов патологий соответствовало их распространенности в популяции, данные о которой получены из медицинских учреждений Москвы. Диагностическая точность ПО на основе ТИИ представлена посредством характеристических кривых (рисунок 74) с выделением значений 100% специфичности (рисунок 75 А, В) и чувствительности (рисунок 75 Б, Г).

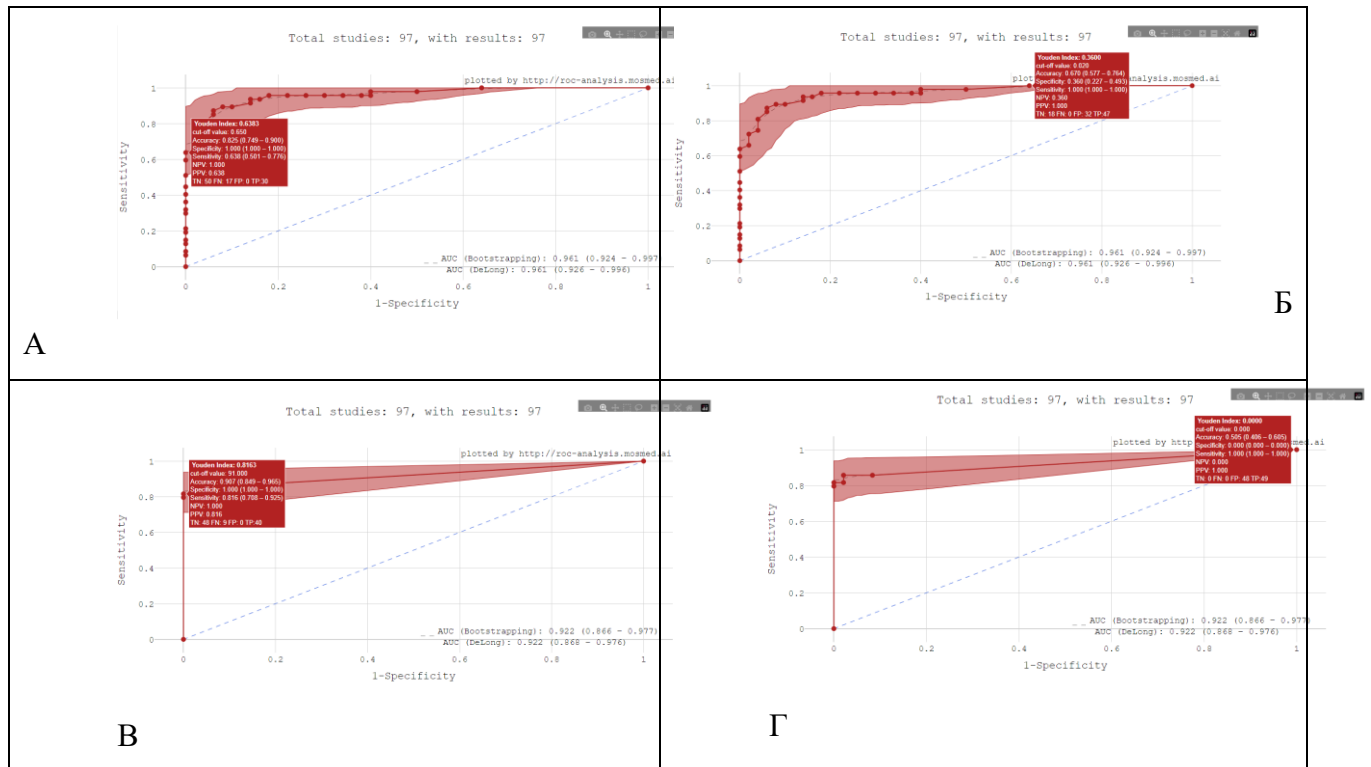


Рисунок 74 – Характеристические кривые двух ПО на основе ТИИ:
А, Б – ИИ-1, В, Г – ИИ-2. Заливкой показан 95% доверительный интервал

Соотношение исследований по группам (ИП, ИО, ЛО, ЛП) в зависимости от порогового значения для каждого из ИИ представлено в таблице 19.

При выставлении порогового значения, соответствующего 100,0% специфичности (пороговые значения в%: 91,0 – для ИИ-2 и 65,0 – для ИИ-1), получаем корректное определение группы исследований «не норма», т.е. нет ЛП-срабатываний. В наборе данных, используемом для тестирования, было 47 исследований группы «не норма». Из этого количества, при предложенной настройке порогов, ИИ-1 корректно определил 30 исследований (63,8% от числа исследований с «не нормой»), а ПО на основе ТИИ ИИ-2 – 40 (85,1% от того же числа). С такими настройками две системы на основе ТИИ показали следующие результаты: ИИ-1 достиг чувствительности на уровне 63,8% (95% ДИ: 50,1–77,6), специфичности – 100,0% и точности – 82,5% (95% ДИ: 74,9–90,0). Вторая система ИИ обеспечила чувствительность на уровне 81,6% (95% ДИ: 70,8–92,5), специфичность – 100% и точность – 90,7% (95% ДИ: 84,9–96,5).

При настройке ПО на основе ТИИ на 100% чувствительность, в рамках настоящего исследования не удалось получить корректного значения для ИИ-2, т.к. полученная вариативность значений процентов патологий на исследуемом наборе данных (количество точек на характеристической кривой – см. рисунок 74) оказался слишком узким. Вследствие этого, единственное значение порога срабатывания, для которого была получена 100,0% чувствительность составило 0, т.е. все исследования отмечались как «не норма». Напротив, для ПО на основе ТИИ ИИ-1 было получено несколько пороговых значений со 100,0% чувствительностью, из которого было определено самое левое значение (рисунок 74 Б), пороговое значение равно 2%. При этом пороговом значении количество ИО срабатываний составило 18 (36,0% от всех исследований группы «норма»). При такой настройке ПО на основе ТИИ получаем чувствительность – 100,0%, специфичность – 36,0% (95%ДИ: 22,7–49,3), точность – 67,0% (95% ДИ: 57,7–76,4) [245].

Таблица 19 – Зависимость количества ложных срабатываний от способа определения порогового значения

Группа		«Не норма»		«Норма»	
Количество исследований		47		50	
ПО на основе ТИИ	Результат в терминах четырехпольной таблицы	Истинно положительный	Ложноположительный	Истинно отрицательный	Ложноотрицательный
ИИ-1	100% чувствительность	47	32	18 (36,0%*)	0
	100% специфичность	30 (63,8%**)	0	50	17
ИИ-2	100% чувствительность	49	48	0	0
	100% специфичность	40 (85,1%**)	0	48	9
* Процент истинно отрицательных результатов рассчитан относительно объема группы «норма». ** Процент истинно положительных результатов рассчитан относительно объема группы «не норма»					

В рамках предложенной модели для обработки данных профилактической флюорографии используется ИИ для первичного анализа изображений. Результаты, соответствующие норме, автоматически регистрируются в медицинских записях без необходимости дополнительного описания врачом-рентгенологом. Те изображения, которые ИИ классифицирует как «не норма», направляются для детального анализа и описания врачом-рентгенологом. При дальнейшем развитии и улучшении алгоритмов ИИ предполагается, что такие случаи будут напрямую передаваться специалистам клинического профиля.

Проведенное исследование на примере двух ПО на основе ТИИ, зарегистрированных в РФ как медицинское изделие, показало возможность настройки каждого из исследуемых решений на 100,0% специфичность. При таком уровне точности, исследования с патологией («не норма») не нуждаются в пересмотре врачом-рентгенологом и подтверждении, могут сразу направляться врачу клинической специальности.

5.4. Оценка эффективности автоматизации процессов организации и проведения флюорографии

В данном разделе будут приведены результаты диагностической точности, основанные на проспективной оценке, используя инструмент текстового анализа протоколов и инструмент ROC-анализа.

Оценка диагностической точности проводилась путем сопоставления результатов работы ПО на основе ТИИ и протокола исследования, подготовленного врачом-рентгенологом. В качестве результата работы ПО на основе ТИИ использовали вероятность наличия патологии, возвращаемую ПО на основе ТИИ по каждому исследованию, с применением разработанного инструмента, описанного в параграфе 4.2.

С целью оценки показателей диагностической точности ПО на основе ТИИ использовали метод ROC-анализа, оценки чувствительности и специфичности –

повторное определение порога срабатывания на основании полученных ROC-кривых.

Оптимальные пороговые значения для расчета чувствительности и специфичности оценивались методом максимизации индекса Юдена. Для анализа флюорографических исследований применялось 4 ПО на основе ТИИ: ИИ-02, ИИ-08 ФЛГ, ИИ-01 ФЛГ. Среднее время работы алгоритма составило $7,37 \pm 25,31$ секунды.

ПО на основе ТИИ при анализе данных выдавали вероятность наличия патологии в диапазоне от 0 до 100%, определение наличия патологии основывалось на пороге срабатывания, установленного для алгоритма. Если вероятность патологии была выше уровня установленного порога срабатывания, то на анализируемых данных имелось патологическое изменение. Пороги срабатывания в процессе доработки ПО изменялись.

Значения представлены в таблице 20.

Таблица 20 – Уровни порога срабатывания, установленные для ПО на основе ТИИ

Рентгенография		Флюорография	
Название ИИ	Порог срабатывания	Название ИИ	Порог срабатывания
ИИ-06	12,0	ИИ-02	20,0
ИИ-07	29,0	ИИ-02 ФЛГ	65,0
ИИ-02	12,0	ИИ-08 ФЛГ	26,0
	30,0 (с 11.08.2021)	ИИ-01 ФЛГ	92,0
ИИ-02 РГ	65,0		48,0 (с 27.07.22)
ИИ-04 DX	21,0	ИИ-01 РГ	92,0
ИИ-05	60,0		
ИИ-08 РГ	26,0		
ИИ-01 РГ	92,0		
	48,0 (с 27.07.22)		
ИИ-01 ФЛГ	92,0		
	48,0 (с 27.07.22)		

Среднее время обработки рентгенографии было выше, чем флюорографии ($55,24 \pm 1721,56$ против $7,37 \pm 25,31$, $p < 0,001$), что связано с настройкой алгоритмов на начальных этапах исследования (флюорография была подключена на поток значительно позже рентгенографии ОГК).

Показатели диагностической точности были рассчитаны для всего потока рентгенографических исследований (таблица 21) и флюорографических исследований (таблица 22), сравнение показателей диагностической точности при обработке алгоритмами рентгенографии и флюорографии приведены в таблице 23.

Таблица 21 – Диагностические показатели для ПО на основе ТИИ, полученные на рентгенографических исследованиях

Название алгоритма	Кол-во исследований (шт.)	Кол-во исследований с патологий (шт.)	AUC	F1	Точность	Чувствительность	Специфичность	Коэф. Каппа-Коэна	Коэф. Метьюса
ИИ-08 РГ	179377	19 274 (10,74%)	0,82	0,35	0,65	0,86	0,63	0,21	0,31
ИИ-06	327214	33372 (10,20%)	0,75	0,26	0,52	0,83	0,49	0,11	0,19
ИИ-04 DX	6390	406 (6,35%)	0,82	0,27	0,72	0,82	0,72	0,18	0,28
ИИ-07	366312	39613 (10,81%)	0,78	0,29	0,53	0,89	0,48	0,13	0,23
ИИ-02	26892	2925 (10,88%)	0,77	0,32	0,62	0,84	0,59	0,18	0,27
ИИ-01 ФЛГ	26059	3046 (11,69%)	0,85	0,51	0,88	0,54	0,93	0,45	0,45
ИИ-01 РГ	231745	16 990 (7,33%)	0,82	0,30	0,86	0,42	0,89	0,23	0,24
ИИ-02 РГ	38726	5397 (13,94%)	0,76	0,45	0,76	0,71	0,76	0,32	0,36
ИИ-05	45061	4743 (10,53%)	0,81	0,37	0,71	0,81	0,69	0,24	0,32

Таблица 22 – Диагностические показатели для ПО на основе ТИИ, полученные на флюорографических исследованиях

Название алгоритма	Кол-во исследований (шт.)	Кол-во исследований с патологией (шт.)	AUC	F1	Точность	Чувствительность	Специфичность	Коэф. Каппа-Коэна	Коэф. Метьюса
ИИ-01 ФЛГ	1473135	34 093 (2,31%)	0,73	0,16	0,93	0,28	0,95	0,13	0,14
ИИ-02	286030	6940 (2,43%)	0,67	0,09	0,74	0,56	0,74	0,05	0,11
ИИ-01 РГ	833	38 (4,56%)	0,75	0,33	0,95	0,29	0,98	0,30	0,30
ИИ-02 ФЛГ	402522	9164 (2,28%)	0,74	0,23	0,92	0,52	0,93	0,20	0,24
ИИ-08 ФЛГ	242810	5322 (2,19%)	0,73	0,10	0,79	0,54	0,79	0,06	0,12

Таблица 23 – Диагностические показатели для ПО на основе ТИИ, полученные при работе ПО на основе ТИИ на рентгенологических и флюорографических исследованиях

Показатели	Рентгенографические исследования	Флюорографические исследования	р
AUC	0,798±0,032	0,724±0,027	р=0,002
F1	0,347±0,080	0,181±0,088	р=0,006
Точность	0,694±0,121	0,864±0,085	р=0,024
Чувствительность	0,745±0,152	0,438±0,127	р=0,004
Специфичность	0,687±0,150	0,877±0,093	р=0,034
Коэф. Каппа Коэна	0,228±0,097	0,148±0,092	р=0,187
Коэф. Метьюса	0,294±0,072	0,183±0,078	р=0,029
Время работы	55.24 ± 1721,56	7.37 ± 25,31	р<0,001

Можно отметить высокие значения показателей AUC, точность, специфичность алгоритмов «ИИ-08 РГ», «ИИ-01 РГ», «ИИ-01 ФЛГ», «ИИ-04», «ИИ-05» и их различие в зависимости от типа диагностики. Так, AUC достоверно выше ($p=0.002$) у алгоритмов, работающих с рентгенологическими

исследованиями, что предположительно обусловлено более широкой доступностью наборов данных для обучения алгоритмов, а также уровнем технического исполнения. При этом точность данных алгоритмов и специфичность достоверно ниже, чем при флюорографических исследования (p=0,024 и 0,034 соответственно). В первую очередь это обусловлено несоответствием количества исследований без патологии и количества исследований с патологическими изменениями при данных методах визуализации, и тем, что данные показатели весьма чувствительны к столь сильному дисбалансу классов.

У показателей, которые не чувствительны к дисбалансу, таких как F1-мера, показатель Каппа-Коэна, коэффициент Метьюса, отмечаются низкие значения, превышающие и приближающиеся к 0,5 у алгоритма «ИИ-01 ФЛГ».

Поэтому далее были проанализированы значения основного показателя – диагностической точности для разных версий (результаты представлены в таблице 24). Необходимо отметить, что большинство актуальных версий АО на основе ТИИ поддерживает работу не только с РГ ОГК, но и с флюорографией.

Таблица 24 – Значения AUC для проспективных исследований

ПО на основе ТИИ_версия	Вид исследования	AUC	Доверительный интервал
ИИ-07_v1	РГ/Флюорограф	0,80	(0,79–0,80)
ИИ-01 ФЛГ_v1	Флюорограф/РГ	0,72	(0,72–0,73)
ИИ-06_v1	РГ	0,74	(0,74–0,74)
ИИ-02_v1	Флюорограф/РГ	0,71	(0,70–0,72)
ИИ-02_v2	Флюорограф/РГ	0,75	(0,75–0,76)
ИИ-04_v1	РГ	0,80	(0,75–0,85)
ИИ-01 ФЛГ_v2	РГ/Флюорограф	0,83	(0,83–0,84)
ИИ-08 РГ_v1	РГ/Флюорограф	0,84	(0,83–0,85)
ИИ-01 РГ_v1	Флюорограф/РГ	0,83	(0,80–0,87)
ИИ-01 РГ_v2 (2023)	Флюорограф/РГ	0,89	(0,88–0,90)

Как следует из таблицы 24, значения показателя диагностической точности (AUC), максимально достигнутые на конец 2022 года, составляли для рентгенографии/флюорографии 0,84, что существенно ниже ранее полученных, указанных в предыдущем параграфе, в том числе для врача-рентгенолога, тем не менее уже в начале 2023 года ИИ-01 РГ_v2 достиг диагностической точности в 0,89. После проведенного анализа расхождения показателей были выявлены следующие причины: в ходе проспективного исследования использовались в качестве истинных протоколы описаний врача-рентгенолога, которые не являются идеальными: многие патологические паттерны могут не указываться в заключении и в описании. Также в рамках выборочного контроля качества были выявлены пропуски патологических признаков. Кроме этого, как было показано в предыдущем параграфе, врачи тоже ошибаются. Поэтому мы провели отдельное углубленное исследование, в рамках которого использовалась актуальная по состоянию на 1 квартал 2023 года версия ПО на основе ТИИ от компании ИИ-01.

Из общего числа 88 048 проанализированных ФЛГ с использованием ПО на основе ТИИ, 87,378 (99,2%) были описаны врачами как «норма», а 670 (0,8%) – как «не норма». В то же время ПО определило 66 043 (75%) как «норма» и 22 005 (25%) – как «не норма». Среди ФЛГ, отнесенных ИИ к категории «норма», 65 979 (99,9%) были также классифицированы врачами как «норма». Были построены характеристические кривые для анализа работы ПО на всей выборке и на тысяче флюорограмм, описанных врачами «высокого доверия»⁵, с акцентом на максимальную чувствительность (рисунок 75), включая экспериментальные значения для оптимальных показателей (по максимальному индексу Юдена) – рисунок 75А и на максимальной чувствительности (рисунок 75Б).

⁵ Врачи «высокого доверия» – специалисты, которые прошли аудит контроля качества и имеют оценку качества работы >95%.

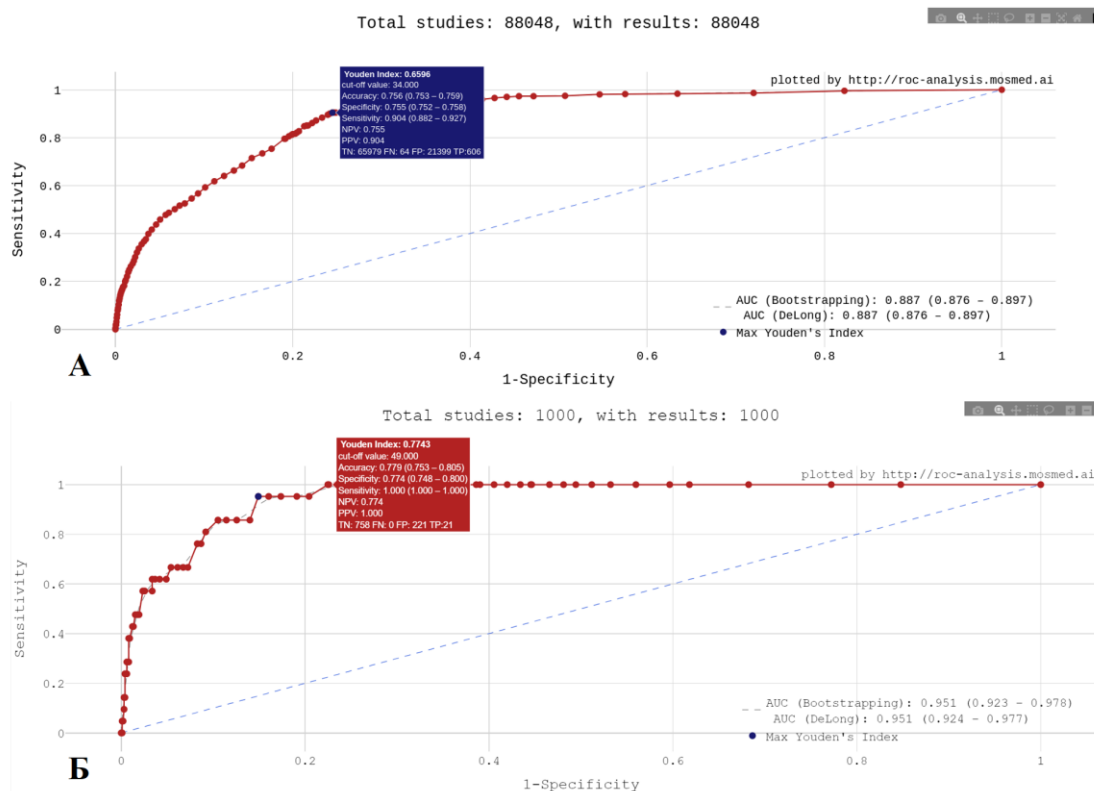


Рисунок 75 – Характеристические кривые ПО на основе ТИИ: А – результат работы на всей выборке при текущих настройках; Б – результат работы на одной тысяче исследований, которые были описаны врачами «высокого доверия», при настройке на максимальную чувствительность

Установка порога в 34,0, соответствующего максимуму индекса Юдена в текущих настройках ПО, позволяет точно классифицировать 65 979 случаев (75%) как «норма», исключая необходимость их дополнительного анализа врачами. Но при этом 64 случая (9,5%) из категории «не норма» будут неправильно идентифицированы как «норма». С таким пороговым значением ИИ показывает диагностическую точность: чувствительность – 90,4% (95% ДИ: 88,2–92,7), специфичность – 75,5% (95% ДИ: 75,2–75,8), точность – 75,6% (95% ДИ: 75,3–75,9).

Оценка работы ИИ на выборке из тысячи ФЛГ, описанных наиболее опытными врачами-рентгенологами, проведена на основе данных первого квартала 2023 года. Установив порог в 49,0 для достижения максимальной чувствительности, ИИ корректно идентифицировал 758 случаев как «норма», что позволяет избежать дальнейшего анализа врачом. При этом ошибочная

классификация исследований из категории «не норма» как «норма» не зафиксирована. Данные настройки обеспечивают чувствительность на уровне 100%, специфичность – 77,4% (95% ДИ: 74,8–80,0), точность – 77,9% (95% ДИ: 75,3–80,5). Значения для четырехпольной таблицы (ИП, ИО, ЛО, ЛП) в зависимости от установленного порогового значения для ПО на основе ТИИ представлены в таблице 25.

Таблица 25 – Четырехпольная таблица результатов работы ПО на основе ТИИ в зависимости от его настроек и характера выборки ФЛГ

Группа	Результат Result	Результаты работы на всей выборке при текущих настройках	Результаты работы на 1000 исследований* при настройке на максимальную чувствительность
«Не норма»	Истинно положительный	606	21
	Ложно- отрицательный	64 9,5%**	0 0%**
«Норма»	Истинно отрицательный	65979	758 77,4%***
	Ложноположительный	21 399	221
* Выборка исследований, описанных врачами «высокого доверия». ** Процент ложноотрицательных результатов рассчитан относительно объема группы «не норма». *** Процент истинно отрицательных результатов рассчитан относительно объема группы «норма».			

Как следует из приведенных данных, ПО на основе ТИИ достигло AUC 0,95, чувствительности – 0,95 (0,861–1,0), специфичности – 0,851 (0,829–0,873), точности – 0,853 (0,831–0,875), что по значениям AUC приблизилось к таковым для врача-рентгенолога, а по показателям чувствительности ПО на основе ТИИ еще пока отстает от таковых для врача-рентгенолога. Однако имеющиеся возможности тонкой настройки позволяют оптимизировать эти показатели под решаемую задачу.

С целью апробации описанной модели было взято ПО на основе ТИИ – лидер по показателям диагностической точности с версией, актуальной на 1 квартал 2023 года, результаты его работы были проанализированы

в сопоставлении с заключениями врачей высокого уровня доверия. Характеристическая кривая с точкой с оптимальной настройкой представлена на рисунке 76.

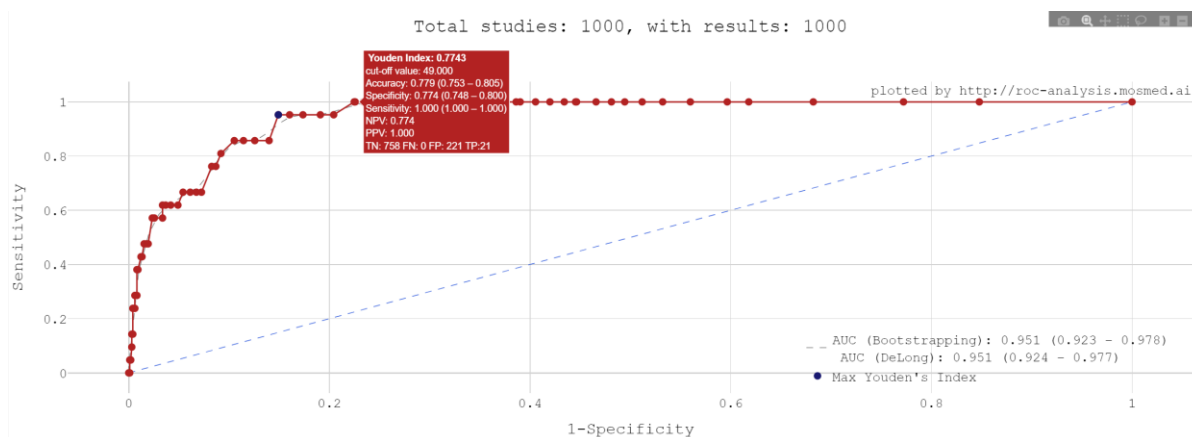


Рисунок 76 – Характеристическая кривая для ПО на основе ТИИ. Синим отмечена точка с оптимальной сбалансированной настройкой. Отмечена точка и вынесены показатели диагностической точности, соответствующие 100% чувствительности

Как следует из приведенной характеристической кривой, возможна настройка ПО на 100% чувствительность, при этом корректно классифицировано будет 758 исследований, что составляет 77,4% от всего потока. Эта модель может являться целевой для флюорографических и рентгенологических исследований, при которых исследования без патологии при условии корректной настройки ПО на основе ТИИ не будут описываться врачом-рентгенологом.

5.5. Оценка влияния применения программного обеспечения на основе технологий искусственного интеллекта на длительность описания результатов профилактических исследований

С целью объективной оценки влияния применения ПО на основе ТИИ на длительность описания профилактических исследований был взят самый первый период внедрения ТИИ в 2020 году. Был выполнен статистический анализ для оценки влияния ПО на основе ТИИ на время, необходимое для подготовки

описаний лучевых исследований до и после его внедрения. Исследование рассматривало три сценария: сокращение, увеличение и отсутствие изменений в продолжительности описаний при использовании ИИ (таблицы 26–28).

Таблица 26 – Длительность описаний результатов ММГ для выявления ЗНО молочной железы

Группа, ИИ- сервис / Параметр	Группа 1	Группа 2	ИИ-1 (ММГ)	ИИ-2 (ММГ)
Количество исследований, абс.	6851	11194	1113	10081
Средняя длительность описаний, мин.	6,45	3,95	4,56	3,88
Медиана	4,0	2,0	3,0	2,0
25-перцентиль	1,0	1,0	1,0	1,0
75-перцентиль	7,0	5,0	5,0	5,0
Средняя разница, мин.	–	-0,66	-0,86	-1,08
Средняя разница, %	–	-4,0	+2,0	-15,0
Значимость увеличения длительности (p)	–	0,89	0,83	0,97
Значимость отсутствия изменения длительности (p)	–	0,22	0,35	0,07
Значимость уменьшения длительности (p)	–	0,11	0,17	0,03

Таблица 27 – Длительность описаний результатов РГ для выявления перечня заболеваний

Группа, ИИ- сервис / Параметр	Группа 1	Группа 2	ИИ-1 (РГ)	ИИ-2 (РГ)	ИИ-2 (РГ)	ИИ-4 (РГ)
Количество исследований, абс.	28034	16288	1156	921	10881	3330
Средняя длительность описаний, мин.	3,96	3,23	4,78	4,46	2,66	4,18
Медиана	2,0	1,0	3,0	3,0	1,0	2,0
25-перцентиль	1,0	1,0	1,0	1,0	0,0	1,0
75-перцентиль	5,0	4,0	6,0	5,0	3,0	5,0
Средняя разница, мин.	–	+0,31	+0,08	+0,46	+0,19	+0,39
Средняя разница, %	–	+25,0	+28,0	+38,0	+17,0	+27,0
Значимость увеличения длительности (p)	–	0,14	0,17	0,06	0,19	0,39
Значимость отсутствия изменения длительности (p)	–	0,28	0,35	0,12	0,39	0,77
Значимость уменьшения длительности (p)	–	0,86	0,83	0,94	0,81	0,61

Таблица 28 – Длительность описаний результатов ФЛГ для выявления перечня заболеваний

Группа, ИИ- сервис / Параметр	Группа 1	Группа 2	ИИ-1 (ФЛГ)	ИИ-2 (ФЛГ)
Количество исследований, абс.	70452	46597	44063	2534
Средняя длительность описаний, мин.	1,52	1,59	1,53	2,6
Медиана	0,0	0,0	0,0	1,0
25-перцентиль	0,0	0,0	0,0	1,0
75-перцентиль	1,0	1,0	1,0	3,0
Средняя разница, мин.	–	+0,17	-8,0	+0,48
Средняя разница, %	–	+1,0	-8,0	-1,0
Значимость увеличения длительности (p)	–	0,0	0,0	0,0
Значимость отсутствия изменения длительности (p)	–	0,0	0,01	0,0
Значимость уменьшения длительности (p)	–	1,0	1,0	1,0

Как следует из приведенных в таблицах 26–28 результатов, применение ПО на основе ТИИ не оказывало статистически значимого влияния на врачей-рентгенологов при описании РГ/ФЛГ исследований ОГК. Один из программных продуктов на основе ТИИ значительно сократил время описания результатов амбулаторной маммографии – на 15%, или 1,08 минуты (критерий Уилкоксона = меньше (“less”), $p = 0,03$), а также уменьшил время подготовки описаний в стационаре на 2 минуты, или 50% (критерий Уилкоксона = меньше (“less”), $p = 0,05$).

Таким образом, применение ПО на основе ТИИ не увеличивает время работы с профилактическими исследованиями (РГ/ФЛГ ОГК) и ускоряет подготовку протокола при описании ММГ на 15,0%.

Заключение по главе 5

В завершающей главе была описана модель организации массовых профилактических исследований с применением ТИИ. Концепция базируется на применении ПО на основе ТИИ в качестве первого чтения лучевого исследования.

В зависимости от вида исследования предполагаются разные настройки ИИ-алгоритмов. Так, для маммографии важно достигнуть баланса чувствительности и специфичности, в этом случае можно будет полностью заменить врача на этапе первого прочтения исследования. Эффективность представленной модели для маммографии была продемонстрирована на ретроспективном этапе, а также в ходе пилотного проспективного исследования. Ключевым показателем эффективности в этом случае являлось достижение целевых показателей диагностической точности, характерных для врача-рентгенолога.

Применительно к флюорографии предлагалась несколько иная настройка программных решений на основе ТИИ, при которой чувствительность ИИ-алгоритма будет стремиться к 100%. В ходе ретроспективного анализа результатов работы ИИ-сервисов по РГ/ФЛГ были показаны значения точности, отличные от таковых для врачей рентгенологов, однако при снижении порога срабатывания ИИ-алгоритма удалось достигнуть чувствительности, близкой к 100%, что указывает на возможность применения машинного обучения для автоматизации описания РГ/ФЛГ исследований с нормой.

ЗАКЛЮЧЕНИЕ

Настоящее научное исследование посвящено решению основных проблем, связанных с внедрением технологий на основе искусственного интеллекта для профилактических лучевых исследований. Переход лучевой диагностики на цифровые технологии позволил существенно расширить инструментарий врача-рентгенолога [132–134]. Начали появляться системы хранения и архивации медицинских изображений (PACS-системы), в которые изображения направлялись напрямую с диагностического устройства, обеспечивая параллельную работу нескольких врачей-рентгенологов со всеми подключенными диагностическими устройствами. Это позволило увеличить скорость работы отделений лучевой диагностики по сравнению с аналоговыми системами. Кроме того, стало возможным сравнение цифровых снимков для оценки динамики патологических процессов наряду с существенным сокращением временных затрат и повышением объективности оценки. Все это положительно сказалось на повышении качества лучевой диагностики.

Следующие важные достижения связаны с возникновением систем обработки цифровых медицинских изображений. Помимо стандартных инструментов настройки яркостно-контрастных показателей изображения, стали появляться системы, позволяющие выделять на рентгенографическом изображении потенциально патологические области, отличающиеся по показателям яркости пикселей от окружающих. Это были первые CAD-системы, по своей сути являющиеся системами поддержки принятия врачебных решений, которые находили применение для выявления признаков туберкулеза на рентгеновских изображениях органов грудной клетки [150,153,154] и признаков, подозрительных на злокачественные образования на рентгенограммах органов грудной клетки [157,158] и маммографических изображениях [161–163]. Однако указанные системы имели ряд ограничений, связанных с незначительным

количеством выявляемых патологических изменений и их диагностической точностью, уступающей точности врача-рентгенолога [149,152].

Методологические основы, лежащие в основе CAD-систем, позволяли с высокой чувствительностью находить целевые патологические паттерны, однако специфичность оставалась крайне низкой [155]. Все CAD-системы имели существенные ограничения на патологические признаки, отличающиеся от целевых, таким образом, разрабатывались отдельные CAD-системы для поиска признаков туберкулеза [153,154], рака легкого [157,158] и другие [159]. В результате исследование могло быть ложно трактовано как «норма». Кроме этого, данные системы обладали невысокой обобщающей способностью, ограничивающей их применение на отличном по качеству и другим техническим параметрам диагностическом оборудовании. Однако после 2010 года появились некоторые исследования, ставящие под сомнение целесообразность применения CAD-систем для профилактических исследований [167–170].

При проведении массовых профилактических исследований основная задача – обнаружить целевую патологию, однако, гораздо большую практическую ценность представляет обнаружение и других патологий. В этом ключе СППВР должна быть направлена не столько на выявление патологии, сколько на выявление нормы. При проведении массовых профилактических исследований большая доля пациентов относится к группе «норма» [67–71]. По данным, полученным в рамках настоящей работы, частота встречаемости патологических признаков в потоке РГ/ФЛГ составляла 29% за счет доминирования кардиомегалии или 18% без учета кардиомегалии. Для маммографии собственные данные указывают на частоту встречаемости BI-RADS 4, 5, 6 не более чем в 3% случаев. При такой низкой частоте встречаемости патологии возрастает вероятность ее пропуска врачом-рентгенологом [65], особенно в условиях его повышенной нагрузки.

После 2010 года, с ростом вычислительных мощностей и их доступности, получили развитие алгоритмы на основе искусственного интеллекта,

анализирующие изображения, в том числе медицинские. В медицинской диагностике технологии на основе искусственного интеллекта стали развиваться относительно недавно, с 2018 года, в связи с повышением доступности наборов медицинских данных [172].

Литературные данные указывают на большие перспективы применения ПО на основе ТИИ для решения задач анализа медицинских изображений по сравнению с CAD-системами. Однако массового применения данных технологий до 2020 года не наблюдалось [177,272]. Большинство исследователей ограничивалось тестированием своих разработок на ограниченных наборах данных, а как максимум – апробацией в рамках одной медицинской организации. Пожалуй, одним из исключений являлась диагностика изменений, характерных для COVID-19 с использованием ПО на основе ТИИ [193, 194].

В систематическом обзоре [205] из 2248 проанализированных работ только одна была посвящена ретроспективному анализу результатов работы, и лишь только одна работа описывала проспективный анализ. Также необходимо подчеркнуть, что в мировой практике не было зафиксировано случаев внедрения ПО на основе ТИИ для профилактических исследований в государственные программы медицинского страхования.

Обобщив вышеизложенное, можно утверждать, что применение ПО на основе ТИИ для профилактических исследований открывает большие возможности для системы здравоохранения и способствует повышению качества диагностики. Однако необходима разработка научно обоснованной системы применения и контроля качества ПО на основе ТИИ.

Проведенное диссертационное исследование стало обобщающим трудом диссертанта и посвящено анализу результатов многолетнего труда, заключающегося в изучении, разработке и усовершенствовании методики внедрения ПО на основе ТИИ в отделения лучевой диагностики медицинских организаций города Москвы. Из шести поставленных задач диссертационного исследования одной из главной являлась разработка многоступенчатой системы

контроля качества ПО на основе ТИИ в масштабах системы здравоохранения мегаполиса. Эта задача была уникальна и по своим масштабам не имела аналогов в мировой практике.

Диссертационная работа была начата в 2020 году и проводилась в рамках НИР «Научное обоснование методологии применения и способов оценки качества интеллектуальных технологий („искусственного интеллекта”) в диагностике», а площадкой для практической апробации стал «Эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы». Достоверность полученных в рамках исследования результатов обеспечивает масштабная выборка: более 150 медицинских организаций, 856 уникальных диагностических устройств, более 6 млн исследований, а также 1410 вовлеченных врачей-рентгенологов.

Методами исследования являлись: цифровая флюорография и рентгенография органов грудной клетки, а также цифровая маммография. Объектом исследования были 12 ПО на основе ТИИ для анализа РГ/ФЛГ- и ММГ-исследований. В качестве параметров оценки использовались характеристики диагностической точности. Подходы к анализу данных заключались в оценке диагностической точности врача-рентгенолога по каждому из рассматриваемых направлений на эталонных наборах данных, сопоставление полученных значений с результатами тестирования ПО на основе ТИИ на этих же наборах данных и проспективная оценка результатов работы ПО на основе ТИИ.

Результаты исследования и полученные заключения соответствовали решению поставленных в диссертации задач, которые коротко отражены ниже.

1. Изучить проблемы управления профилактическими и диагностическими исследованиями, осуществляемыми с применением методов лучевой диагностики. Полученные собственные данные, а также результаты литературного обзора указывают на то, что применение ПО на основе ТИИ для профилактических исследований открывает большие возможности для

системы здравоохранения и способствует повышению качества диагностики за счет снижения нагрузки врачей-рентгенологов и повышения диагностической точности профилактических методов. Вопрос повышения качества описания профилактических исследований является чрезвычайно актуальным: анализ, проведенный по методологии ретроспективных пересмотров [273], показал, что полное соответствие достигается только в 61,0% исследований, но при этом наблюдается прогрессивное снижение числа клинически значимых расхождений до 0,8%. Таким образом, видится целесообразным внедрение ТИИ для снижения количества расхождений за счет снижения нагрузки с врача-рентгенолога.

ТИИ позволяют решить многие проблемы управления массовыми профилактическими исследованиями, при этом не потребуются существенное изменение инфраструктуры. ЕРИС ЕМИАС обладает всеми необходимыми техническими параметрами для подключения решений на основе ТИИ, следовательно, не потребуются существенного изменения инфраструктуры. Результаты анализа потока профилактических исследований свидетельствуют о крайне низкой частоте встречаемости патологических находок, и в этом случае ТИИ могут выступать в роли первичной сортировки исследований, выполняя функцию триажа и предоставляя врачу-рентгенологу предварительный протокол исследования. Это позволяет решить проблему недофинансирования двойного прочтения маммографических исследований и оптимизировать нагрузку на врача-рентгенолога.

2. Оценить применимость технологий искусственного интеллекта в лучевой диагностике путем сравнительного изучения точности диагностических решений, принимаемых врачом самостоятельно и формируемых автоматически. Для решения данной задачи потребовалось провести исследование, целью которого являлась оценка диагностической точности врача-рентгенолога, с использованием специально разработанной для этих целей платформы. Исследование охватило большую когорту врачей с разным стажем работы.

Для направления «флюорография» средняя чувствительность врача-рентгенолога составила 1,0 (1,0–1,0), специфичность – 0,91 (0,83–0,99), точность – 0,96 (0,92–1,0), AUC – 0,97 (0,94–1,0). На аналогичном наборе данных были рассчитаны показатели диагностической точности ПО на основе ТИИ и получены следующие значения показателей: чувствительность – 0,89 (0,80–0,98), специфичность – 0,96 (ДИ: 0,90–1,00), точность – 0,92 (0,87–0,98), AUC – 0,95 (0,90–0,99).

Для направления «рентгенография» средняя чувствительность врача-рентгенолога составила 0,90 (0,82–0,99), специфичность – 0,96 (0,90–1,0), точность – 0,93 (0,88–0,98), AUC – 0,96 (0,93–1,0). Для одного из лучших ПО на основе ТИИ значения показателей диагностической точности составили: чувствительность – 0,86 (0,75–0,96), специфичность – 0,89 (0,80–0,98), точность – 0,88 (0,81–0,94), AUC – 0,91 (0,85–0,97)

Выбор программного обеспечения на базе ИИ для клинического использования требует тщательного анализа поставленной задачи. В контексте скрининговых программ, крайне важно стремиться к достижению максимальной чувствительности, чтобы минимизировать вероятность пропуска патологии [274]. Для функции поддержки принятия решений рентгенологом предпочтительно использовать программы с низким уровнем ложных срабатываний, учитывая, что значения специфичности для врача-рентгенолога традиционно выше его чувствительности [275, 276]. В настоящем исследовании не делалось различий между врачами по стажу, так как предполагается универсальность применения ИИ без учёта опыта специалиста. Необходимо также отметить, что в этом исследовании не оценивалось качество визуализации графической разметки патологических областей и корректность формирования текстового описания исследования, что важно при внедрении технологий ИИ в медицинскую практику [227, 277].

Для направления «маммография» средняя чувствительность врача-рентгенолога составила 0,79 (0,68–0,91), специфичность – 0,94 (0,87–1,0),

точность – 0,87 (0,80–0,93), AUC 0,93 – (0,88–0,98). Для лучшего ПО на основе ТИИ значения показателей диагностической точности на аналогичном наборе данных составили: чувствительность – 0,83 (0,73–0,94), специфичность – 0,96 (0,91–1,0), точность – 0,90 (0,84–0,96), AUC – 0,96 (0,92–0,99).

Выбор ПО на основе ТИИ для клинического применения зависит от его показателей диагностической точности, для оценки которой применяется не только AUC, но также учитывается пара показателей: чувствительность и специфичность. Установленный минимальный порог AUC в 0,81 [235] является условием для применения такого ПО в клинической практике. Таким образом, четыре из пяти ПО на основе ТИИ, рассмотренных в данном исследовании, соответствуют этому критерию и готовы к практическому применению. Вариативность значений AUC в научной литературе колеблется от 0,706 [278] до 0,956 [279]. Однако сравнение с двойным чтением маммограмм выявило, что ИИ-системы в целом уступают врачам-рентгенологам, как, например, показано в обзоре К. Фримена и соавторов [280]. Ни одно из 36 анализируемых решений на основе ИИ не превзошло врача-рентгенолога по AUC. Было показано, что два решения на основе ТИИ продемонстрировали значения AUC, статистически значимо не отличающиеся от «среднего» врача-рентгенолога.

Один из важных выводов данного исследования – диагностическая точность «среднего» врача-рентгенолога при описании ММГ основанная на оценке результатов работы 113 врачей. Зарубежная литература предлагает разнообразные значения AUC для врачей-рентгенологов: 0,778 [281], 0,816 [278], 0,87 [192]. В России этот аспект освещен лишь в одной научной публикации П. И. Павловича и соавторов [250], указывающей на пропуск одного патологического случая из 49 без проведения ROC-анализа. Самое объемное исследование с участием 237 врачей показало чувствительность 90,0% и специфичность 76,0%, AUC 0,88 [282]. Настоящее исследование демонстрирует более высокую точность отечественных врачей по сравнению с многими зарубежными работами. Утверждается, что при интеграции ИИ в анализ маммограмм, необходимо устанавливать строгие

критерии диагностической точности, где алгоритмы с AUC ниже 0,883 (нижняя граница доверительного интервала для «среднего» врача-рентгенолога) могут оказаться неэффективными из-за высокого риска ложных срабатываний, приводя к снижению результативности работы пары «врач + ИИ».

Указанные параметры диагностической точности врача-рентгенолога можно взять за основу для определения минимального порога с целью вхождения ТИИ в массовые профилактические исследования. Однако на практике было установлено, что большое значение имеет не столько сбалансированные значения показателей диагностической точности, сколько возможность настройки ПО на основе ТИИ на 100% чувствительность, что и было экспериментально показано. Практическое применение данной модели изложено в последующих задачах.

Важной составляющей данной задачи является оценка диагностической точности врача-рентгенолога, уникальность которой заключается в большом объеме выборки. Проведение такого исследования стало возможным благодаря созданию специализированной платформы «экспертного пересмотра» [283], позволяющей в удобном интерфейсе со стандартным набором инструментов просматривать цифровое медицинское изображение. Эта платформа позволила не только реализовать независимую экспертную разметку исследования, но и проводить тестирование врачей-рентгенологов для оценки их диагностической точности.

Было реализовано сравнение диагностической точности врача и ПО на основе ТИИ на одних и тех же данных. Благодаря разработанной методики подготовки наборов данных удалось провести объективное сравнение. Так, все рентгенографические и флюорографические исследования, помимо экспертного пересмотра, были подтверждены более точным методом – компьютерной томографией, а маммографические исследования – трепан-биопсией. Полученные в ходе исследования показатели диагностической точности «среднего» врача-рентгенолога рекомендованы для практического применения в качестве пороговых значений при решении вопроса о пригодности ПО на основе ТИИ.

Таким образом, на поставленный в задаче вопрос о применимости ТИИ для массовых профилактических исследований можно дать положительный ответ. ПО на основе ТИИ возможно настроить на целевые значения чувствительности или специфичности, обеспечив тем самым сопоставимые или превосходящие «среднего» врача-рентгенолога показатели диагностической точности, что и будет подтверждено на практике при решении последующих задач.

3. Разработать методику внедрения систем поддержки принятия врачебных решений для автоматизации рабочего процесса врача-рентгенолога при проведении профилактических и диагностических исследований. По данным научной литературы [205], большинство авторов проводили свои исследования на наборах данных, лишь единичные исследователи указывали на возможность практического применения разработанных алгоритмов путем интеграции в бизнес-процесс врача рентгенолога [203]. Стоит отметить, что большинство решений предлагается врачу использовать посредством веб-интерфейса отдельного сервиса или отдельного приложения, развернутого локально на АРМ врача. На практике это означает необходимость врачу-рентгенологу работать одновременно минимум в трех программных продуктах (просмотрщике медицинских изображений, медицинской информационной системе и в интерфейсе программы обработки исследований с ИИ). Каждое из этих решений имеет свои недостатки [215].

Локально развернутое ПО на основе ТИИ требовательно к ресурсам АРМ и при недостаточной вычислительной мощности приведет к увеличению времени обработки исследования, что в свою очередь увеличит время описания исследования врачом-рентгенологом. Напротив, облачное решение лишено этих недостатков, однако оно требует дополнительной безопасности при работе – анонимизации исследований, а также высокой пропускной способности каналов связи. Однако, независимо от выбранного способа реализации, врачу необходимы дополнительные действия по отправке исследований и выгрузки результатов

обработки, причем это все осуществляется в стороннем, непривычном программном обеспечении.

Альтернативный подход заключается в объединении PACS-систем с ПО на основе ТИИ, но на момент начала работы над этим исследованием, в 2020 году, такой подход, по мнению многих зарубежных исследователей, казался далекой перспективой [213, 215]. Теме не менее этот подход позволяет пользователю выбрать наиболее подходящее решение для анализа данного исследования [285], при условии технической поддержки.

В настоящей работе был предложен способ интеграции ПО на основе ТИИ с PACS-системой мегаполиса (ЕРИС). Это является уникальным решением как с технологический, так и с методологической точек зрения. Автоматическая маршрутизация исследований из ЕРИС в ПО на основе ТИИ происходит автономно, по заранее разработанному плану маршрутизации. Это позволяет минимизировать время между поступлением исследования в ЕРИС и появлением результатов его обработки. Главное преимущество: результаты работы ПО на основе ТИИ доступны врачу-рентгенологу в его обычной программной среде – в просмотрщике с любого рабочего места, в том числе из веб-версии просмотрщика, которой пользуются врачи-клиницисты.

В рамках настоящей задачи также был осуществлен выбор оптимального способа представления результата анализа исследования врачу-рентгенологу, что является важной проблемой при внедрении ТИИ в лучевую диагностику [233]. Выбранный способ представления заключался в визуализации патологических областей, выделенных ПО на основе ТИИ на дополнительной графической серии, а также текстового описания на дополнительной вкладке в разделе подготовки протокола описания исследования. Оптимальное наполнение текстовых полей, а также способа визуализации патологических находок было сформировано по результатам экспертного анализа и обратной связи от врачей-рентгенологов. Основные требования к интеграции ПО на основе ТИИ изложены в методических рекомендациях [286].

Отличительной особенностью разработанной методики внедрения ПО на основе ТИИ является возможность предоставления результатов работы от нескольких ПО. Этот функционал позволяет комбинировать различные ПО на основе ТИИ, в зависимости от задачи. Также представленная методика позволила реализовать возможность обработки исследования по запросу, когда врач хочет получить второе мнение, причем весь этот функционал был реализован в штатном ПО. По умолчанию при работе с ПО на основе ТИИ по предложенной схеме врач получает результат анализа ПО автоматически, т.е. при открытии исследования, и ему не нужно совершать дополнительные действия по его выгрузке, анонимизации и направлению в сторону ПО, что увеличивает доступность ИИ-технологий для врача-рентгенолога, оставляя возможность выбора дополнительного инструмента в зависимости от текущих задач.

В предложенной схеме ПО на основе ТИИ размещается в облачном сервисе, что с учетом высокоскоростных линий связи позволило минимизировать время ожидания результатов работы ПО на основе ТИИ. По состоянию на конец 2022 года, медианное время доступности результатов анализа РГ/ФЛГ составило 2 минуты 39 секунд с момента выполнения исследования на диагностическом устройстве, ММГ – 2 минуты. При этом время обработки исследования на стороне ПО на основе ТИИ в большинстве случаев равнялось 7 секундам. За аналогичный отчетный период медианное время (от завершения исследования до визирования протокола врачом-рентгенологом) составило 16 минут. На практике это означает что врач, открывая исследование, практически всегда видит результат ИИ.

Разработанная методика интеграции ПО на основе ТИИ позволила успешно внедрить в клиническую практику решения для обработки профилактических исследований по основным видам исследований: рентгенография органов грудной клетки/флюорография и маммография, а также обеспечить их непрерывное совершенствование. Как показали результаты проведенной работы, общий поток профилактических исследований, обработанных ПО на основе ТИИ, составил 4,3

млн за три года, средний поток исследований в год – 807 тысяч флюорограмм, 422 тысячи рентгенограмм органов грудной клетки и 225 тысяч маммографических исследований. Несмотря на такой большой объем исследований, ПО на основе ТИИ успешно с ним справлялись, обеспечивая непрерывное снижение удельного веса технологических дефектов, связанных с превышением времени обработки исследования, что при таких масштабах является рекордно высоким по сравнению с зарубежными исследованиями.

Предложенная методика может быть легко масштабирована на любые другие РИС/PACS-системы. В частности, она легла в основу ИИ-консультанта врача, развернутого на платформе: <https://hub.tele-med.ai>. Посредством сервиса автоматического анализа рентгенологических исследований для врачей осуществляется обработка как отдельных исследований, так и всего потока профилактических исследований из медицинских организаций Ямало-Ненецкого автономного округа.

Предложенная в рамках настоящего исследования методика позволила также эффективно подключить к ЕРИС ПО на основе ТИИ по другим направлениям и видам исследований. Но самое главное – она продемонстрировала свою тиражируемость за пределы городской системы здравоохранения на федеральный уровень, что доказало ее эффективность и масштабируемость.

4. Обосновать организационно-методические основы контроля безопасности и качества программного обеспечения на основе технологий искусственного интеллекта для профилактических и диагностических лучевых исследований. Как отмечалось выше, большинство исследователей осуществляют проверку качества работы ПО на основе ТИИ на ограниченном наборе данных, а показатели ограничиваются исключительно показателями диагностической точности. Согласно проведенному литературному обзору, в основном применяются показатели AUC, точность, чувствительность и специфичность. При этом ИИ-модели в отличие от CAD-систем используют

вероятностную шкалу оценки. Таким образом, потенциальный пользователь данной модели получает вероятность наличия патологии или патологического признака. Несмотря на то, что ответ в виде вероятности неочевиден для интерпретации, он открывает большие возможности для тонкой настройки ПО на основе ТИИ [214]. Тонкая настройка осуществляется методом ROC-анализа, позволяя задавать оптимальное значение для чувствительности и/или специфичности и получать конкретное пороговое значение для вероятности патологии. Именно с данной целью в рамках настоящей работы был разработан комплекс инструментов, в основе которых лежит ROC-анализ. Это уникальная разработка, т.к. в открытом доступе нет решений с аналогичными возможностями и функционалом.

Разработанный веб-инструмент для проведения ROC-анализа [287] разрешает в интерактивном режиме подбирать оптимальный порог вероятности, отслеживая изменения чувствительности, специфичности и точности в зависимости от выбранного порога. Кроме этого, веб-инструмент позволяет осуществлять автоподбор порогового значения методом максимизации индекса Юдена или методом минимизации расстояния до точки со 100% чувствительностью и специфичностью. Преимуществами этого инструмента являются: доступность, веб-интерфейс, расчет всех показателей диагностической точности с 95% доверительным интервалом и возможность экспорта результата. Интернет-версия инструмента имеет ограничение по количеству строк для анализа в количестве 1000, а локальная версия не имеет ограничений и использовалась для ROC-анализа таблиц с количеством строк более 1 млн.

На основе вышеописанного инструмента были реализованы две платформы: платформа проведения калибровочных тестирований и платформа тестирования при смене версионности. Первая платформа позволяет автономно проводить ROC-анализ данных, полученных на калибровочном наборе данных, подготавливать протокол с заключением о соответствии или не соответствии заявленных значений показателей диагностической точности. Вторая –

представляет собой автономную систему, обеспечивающую прямое подключение к себе ПО на основе ТИИ, осуществляющее обмен диагностическими исследованиями с фиксацией результата работы ПО на основе ТИИ. На основе этих данных осуществляются ROC-анализ, расчет показателей диагностической точности и автоматизированная подготовка протокола с заключением. Указанная платформа использовалась в настоящей работе для контроля за показателями диагностической точности ПО на основе ТИИ при проведении доработок на этапе работы на потоке данных.

Веб-инструмент ROC-анализа был также дополнен инструментом сравнения ROC-кривых, позволяющим достоверно попарно сопоставлять результаты работы ПО на основе ТИИ между ними или с врачами-рентгенологами. Предложенная модификация инструмента ROC-анализа нашла широкое применение в ходе проведения сравнительной оценки качества работы ПО на основе ТИИ.

Разработанные на основе ROC-анализа инструменты позволили стандартизировать и оптимизировать работы по оценке диагностической точности ПО на основе ТИИ, а предоставление инструментов в открытом доступе для научного сообщества позволило ускорить анализ результатов тестов диагностической точности.

Большое практическое значение имеет расчет показателей диагностической точности для врача, это позволяет определить целевые значения для отбора ПО на основе ТИИ. С целью обеспечения объективной оценки врача-рентгенолога, была разработана отдельная платформа [283] с графическим интерфейсом, позволяющая предоставлять врачу-рентгенологу исследования в формате DICOM со всем необходимым набором стандартных инструментов. Платформа поддерживает загрузку наборов данных неограниченного размера, подключая врача-рентгенолога по электронной ссылке посредством любого веб-браузера, а также автоматически осуществляет расчет показателей диагностической точности каждого из врачей. Разработанная платформа применялась в настоящей работе для оценки диагностической точности врачей-рентгенологов

по направлениям РГ/ФЛГ органов грудной клетки, ММГ, а также использовалась для разметки маммографических исследований экспертами.

Важную практическую ценность при внедрении ПО на основе ТИИ для описания профилактических исследований в лучевую диагностику имеет возможность сопоставления результатов работы ПО с результатами работы врача-рентгенолога. Это сопоставление позволяет оценить качество работы ПО на основе ТИИ на большом объеме исследований, а также сделать вывод о целесообразности его внедрения в масштабную клиническую практику, определяет точки дальнейшего развития ПО. Однако для проведения данного сравнения необходим инструмент, позволяющий осуществлять анализ текстовых протоколов врача-рентгенолога. Для этих целей были разработаны два NLP-модуля: для анализа текстовых протоколов РГ/ФЛГ органов грудной клетки и для анализа ММГ-протоколов. Разработанный инструментальный позволил не только провести расчет показателей диагностической точности ПО для проспективной оценки, но и получить точные данные о частоте выявляемости патологических рентген-признаков при проведении профилактических исследований. Данные анализа частоты встречаемости патологических признаков на РГ/ОГК органов грудной клетки позволили сформулировать целевые паттерны для выявления ПО на основе ТИИ. Например, результаты анализа продемонстрировали большую частоту встречаемости «возрастных изменений» – 45% протоколов содержали такую формулировку. При этом клинической ценности «возрастные изменения» не представляют, и срабатывания ПО на основе ТИИ на такие находки не имеют практической ценности тоже, что мы могли наблюдать в начале исследования на потоке данных. Примечательно, что в настоящее время несмотря на то, что данные о возрасте пациентов не передаются для анализа ПО, ложные срабатывания на такие находки сократились. Этому, возможно, способствовало появление алгоритмов на основе компьютерного зрения, позволяющих определять возраст пациента

по РГ ОГК [288–290]. Следовательно, «возрастные изменения» не были включены в перечень патологических признаков.

На этапе работы ПО на основе ТИИ на потоке важно было обеспечить мониторинг качества, для этих целей была разработана автоматизированная система, включающая платформу мониторинга. Данная платформа обеспечивала отбор исследований для мониторинга экспертами, а также автоматический анализ данных для подготовки протокола о мониторинге с формированием заключения.

Таким образом, в рамках настоящей работы были разработаны организационно-методические основы системы контроля безопасности и качества, состоящие из методики и обеспечивающего ее набора инструментов, охватывающие самый широкий спектр задач, связанных с внедрением технологий на основе искусственного интеллекта в медицинскую диагностику.

Несмотря на наличие в литературе отдельных исследований по внедрению ПО на основе ТИИ в клиническую практику рентгеновских отделений, ни в одном из ее источников не приводились данные о разработке комплексной методики контроля качества. В рамках настоящей работы была предложена многоступенчатая система контроля качества, разделенная на пять основных этапов:

1. Подготовительный этап, характеризующийся формированием основных требований к ПО на основе ТИИ. Базовые функциональные требования регламентируют формат и способ предоставления результата конечному пользователю. Базовые диагностические требования описывают клинические задачи, на решение которых должно быть направлено данное ПО на основе ТИИ. В рамках настоящей работы были сформулированы требования к ПО на основе ТИИ для анализа РГ/ФЛГ органов грудной клетки и ММГ-исследований. Кроме требований, на подготовительном этапе осуществляется подготовка наборов данных по названным направлениям для ознакомления разработчиков с особенностью исследований по этим направлениям и проведения последующих тестирований.

2. Этап самотестирования – разработчикам предлагается самостоятельно протестировать свое ПО на основе ТИИ на наборе данных, находящемся в открытом доступе, для оценки готовности решения к тестированиям и последующей работе на потоке.

3. Этап функционального тестирования – направлен на проведение оценки функционала готового ПО на основе ТИИ, а также его соответствия ранее сформированным требованиям.

4. Этап калибровочного тестирования – основной задачей является проверка соответствия параметров диагностической точности заявленным, а также позволяющим осуществлять тонкую настройку ПО на основе ТИИ под перспективные задачи. Тонкая настройка, как показало данное исследование, является важной задачей при внедрении ПО на основе ТИИ для профилактических исследований.

5. При успешном прохождении ПО на основе ТИИ этапов тестирования оно допускается к контрольно-техническому испытанию, в рамках которого осуществляется оценка возможности корректной обработки исследований со всех целевых диагностических устройств. После прохождения этого тестирования поток с успешно пройденных диагностических устройств направляется для обработки на данное ПО. После прохождения данного этапа ПО на основе ТИИ переводится на этап работы на потоке.

Успешное прохождение всех вышеописанных этапов, как показало настоящее исследование, не гарантирует, что ПО на основе ТИИ будет работать на потоке без дефектов, но означает необходимость применения системы контроля качества и мониторинга. В ходе исследования были установлены два основных типа дефектов, возникающих при работе ПО на основе ТИИ на потоке: технологические и клинические. Для практического применения ПО на основе ТИИ большое значение имеет его диагностическая точность, что требует минимизации ошибок, обусловленных в первую очередь пропуском патологии, а также минимизацией числа ложных срабатываний. Клинический мониторинг

осуществляется специально подготовленными врачами-рентгенологами и заключается в оценке по четырех-градационной шкале описания исследований и графической маркировки патологических областей на изображении. Для этих целей формируется псевдослучайная выборка из 80 исследований с обогащением в сторону исследований с патологией. Объем выборки был статистически обоснован и учитывает баланс рисков поставщика и потребителя ТИИ. Несмотря на то, что технологические дефекты не связаны с клинической оценкой изменений, они также влияют на получение результата врачом-рентгенологом. Как было показано в настоящей работе, устранение технологических дефектов не является сложной задачей, а представленная система оценки динамики и прогнозирования технологических дефектов позволяет указывать на них разработчику, а ему оперативно устранять их.

Разработанная методика многоступенчатой системы контроля качества работы ПО на основе ТИИ и его жизненный цикл обеспечили развитие всех решений, участвующих в этом исследовании. Реализация представленного жизненного цикла стала возможной только благодаря разработанному набору инструментов, описанному ранее в предыдущей главе. «Ценной особенностью разработанной методики является обязательное проведение тестирований на локальных данных и последующая валидация на потоке реальных данных в процессе проспективной работы с последующим экспертным аудитом» [261, стр. 170].

Зарубежные метаанализы указывают на ничтожно малое количество публикаций по внешней валидации ИИ [191], критикуя их качество и достоверность [291]. Разочарование от несоответствия ПО на основе ТИИ ожиданиям врачей замедляет прогресс в их внедрении в медицине [292]. Эффективное применение ТИИ требует от врачей-рентгенологов понимания специфики инструментов, что возможно через дополнительное обучение и обновление программ вузовского и послевузовского образования в области ТИИ и работы с данными, что и было реализовано в сотрудничестве с РТУ МИРЭА.

В результате была подготовлена и реализована совместная дополнительная профессиональная программа повышения квалификации «Аспекты разработки и тестирования технологий искусственного интеллекта в здравоохранении».

Методика оценки качества ПО на основе ТИИ, описанная в данной работе для лучевой диагностики, может быть адаптирована для различных сфер медицинской диагностики. Многоступенчатая методика позволяет поэтапно исключать ПО, не отвечающее диагностическим и функциональным требованиям, способствует эффективному использованию ресурсов медицинских учреждений. Современное развитие медицинских ИИ требует новых методов оценки их результативности и качества [293]. Однако лишь некоторые инструменты оценки цифровой зрелости могут быть полностью автоматическими, что усложняет прозрачность управленческих решений из-за отсутствия четких критериев для сравнения такого ПО.

В ходе трехлетнего эксперимента с участием 11 ПО на основе ТИИ была апробирована методика мониторинга и ранжирования, в результате которой были разработаны две системы сравнительной оценки качества работы ПО на основе ТИИ на потоке. Первая из них позволяла получить одно значение итогового балла в диапазоне от 0 до 100 за отчетный период, где чем больше значение, тем лучше оцениваемое решение. Итоговый балл учитывал скорость обработки исследования, удельный вес технологических дефектов, диагностическую точность ПО на основе ТИИ по результатам проспективной работы, а также согласие экспертов, проводящих мониторинг с результатами работы ПО. Эта методика позволяла сравнивать ПО на основе ТИИ между собой, а также формировать лидерборд, являвшийся инструментом открытого сравнения различного ПО на основе ТИИ и способствующий совершенствованию и развитию ПО для медицинской диагностики. Однако эта методика опиралась на один итоговый показатель, что приводило к тому, что в лидерах мог оказаться технически стабильный продукт, но при этом его клиническая ценность уступала конкурентам. Для повышения объективности сравнения и выбора ПО на основе

ТИИ для лучевой диагностики была предложена и апробирована на практике модель матрицы зрелости ПО, которая представляет собой двухмерную систему координат, учитывающую все три показателя: техническую стабильность, диагностическую ценность и клиническую стабильность решения. То решение, которое оказывалось ближе к верхнему правому краю, являлось приемлемым для внедрения в практическое здравоохранение.

По результатам трехлетнего исследования фиксируется положительная динамика показателей количества технологических дефектов, показателей диагностической точности и клинической оценки. Это стало возможным благодаря открытой публикации лидерборда, а разработанная матрица зрелости позволила объективно сравнивать и наглядно представлять текущее состояние ПО на основе ТИИ. Предложенная методика оценки и сравнения ПО не имеет аналогов, была успешно апробирована и внедрена в практику.

5. Обосновать и внедрить модель организации и проведения профилактических и диагностических исследований с применением технологий искусственного интеллекта. В рамках настоящей работы были апробированы два сценария применения ПО на основе ТИИ для массовых профилактических исследований: в качестве инструмента триажа исследований и в качестве первого мнения. Для РГ/ФЛГ-исследований был продемонстрирован очень низкий процент исследований с патологией (менее 20%). В результате врач большую часть времени занят описанием исследований с нормой, а риск пропуска патологии увеличивается [65]. Было предложено использование ПО на основе ТИИ в качестве инструмента фильтрации потока.

Как было уже выше отмечено, алгоритмы, лежащие в основе ПО, представляют результат в виде вероятности наличия патологии, что позволяет провести их тонкую настройку. При настройке ПО на 100% чувствительность возможно отсеять определенный процент исследований с нормой, которые в реальной модели могут уже не направляться врачу-рентгенологу для описания, а в виде электронной медицинской записи фиксироваться в амбулаторной

медицинской карте. На практике реализация такого процесса требует изменения законодательства или установления экспериментально-правового режима.

Полученные в рамках проспективного исследования показатели диагностической точности нельзя назвать высокими для рентгенографии/флюорографии органов грудной клетки, максимальный AUC составил 0,85. Причины низких значений диагностической точности могут быть из-за вероятности ошибок врачей. Кроме того, это совсем не означает низкую клиническую ценность ПО на основе ТИИ. Посредством ROC-анализа, сначала на ограниченном объеме выборки – эталонном наборе данных, а затем в ходе проспективного сравнения, была показана возможность тонкой настройки ПО на основе ТИИ. При тонкой настройке возможна настройка на любой параметр: 100% чувствительность или 100% специфичность. Однако для массовых профилактических исследований клиническую ценность имеет именно 100% чувствительность.

Тонкая настройка ПО на основе ТИИ для анализа флюорографических исследований позволила достигнуть 100% чувствительности, при которой специфичность достигает 77%, что означает: до 77% потока может не пересматриваться врачом-рентгенологом, следовательно, нагрузка на врача-рентгенолога снижается на 77%. В свою очередь, предлагаемая в рамках настоящей работы модель применения ПО на основе ТИИ в качестве маршрутизатора исследований и автономного описания позволит сэкономить колоссальные ресурсы при проведении профилактических исследований без потери качества оказания медицинских услуг. Однако практическая реализация этой модели так и не была осуществлена в силу ограничений со стороны нормативно-правовой базы, не предусматривающей возможность оказания автономной медицинской услуги.

В рамках настоящей работы было реализовано обоснование применимости такой автономной услуги. Схожие результаты могут быть достигнуты при обработке рентгенографии органов грудной клетки, однако на данный момент

обработка исследований на потоке осуществляется в рамках эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы.

Тонкая настройка ПО на основе ТИИ для анализа маммографических исследований позволила достигнуть 100% чувствительности, при которой специфичность составила всего 54%, однако на практике этот показатель будет означать, что большой объем исследований корректно определяется из-за того, что доля исследований без патологии, как было показано в настоящей работе, достигает 90%. Однократный просмотр снижает чувствительность по сравнению с двойным просмотром для всех категорий BI-RADS. Более того, отказ от двойного просмотра несет разнообразные негативные последствия для обследуемых женщин [31, 32], поэтому в настоящей модели предполагается сохранить двойное чтение ММГ-исследований, но при этом заменить врача, выполняющего первое чтение, на ПО на основе ТИИ. Суть предложенной модели заключается в следующем: ПО на основе ТИИ является заменой первого чтения, предоставляя врачу-рентгенологу описание и маркировку патологических областей на маммограмме. Для этого была использована сбалансированная настройка чувствительности и специфичности, позволившая достигнуть показателей, сопоставимых по точности со врачом-рентгенологом и превосходящих его по чувствительности. Применение ТИИ в медицинской диагностике показало, что специализированное ПО, обладающее оптимальным балансом между чувствительностью и специфичностью, достигло уровня чувствительности в 80,0%, что выше, чем при двойном чтении маммографий врачами-рентгенологами, оцененном в 72,0-73,0% [201, 294]. Специфичность ИИ составила 90,0%, сравнимая с показателями при чтении двумя специалистами (88,0-98,0%) [201, 294]. Исследования также указывают на повышение общей точности интерпретации маммографий при использовании ИИ в дополнение к одному врачу-рентгенологу [190,217,295].

Предлагаемая модель применения ПО на основе ТИИ основывается на его использовании в качестве первого чтения исследования. Этот сценарий на практике был реализован для маммографии, т.к. программа скрининга рака молочной железы предполагает двойное описание маммографических исследований.

6. Оценить результативность применения систем поддержки принятия врачебных решений для автоматизации процессов организации и проведения профилактических и диагностических исследований. В среднем в РФ ежегодно выполняется 74 млн профилактических флюорографий [271]. Из этого числа патологические находки содержат в среднем 10,0% исследований (около 7,4 млн в абсолютных числах) [67–71]. Внедрение в клиническую практику предложенной методики настройки ПО на основе ТИИ позволит не пересматривать от 63,8% исследований группы «не норма» (4,7 млн исследований в масштабах страны) до 85,1% этой группы (6,3 млн исследований).

Однако наибольший практический интерес представляет выявление на ФЛГ нормы (то есть сортировка результатов исследований на категории «норма» и «не норма»). В нашем исследовании удалось корректно определить порог для обеспечения 100,0% чувствительность для одного из двух рассмотренных ПО на основе ТИИ. Так, настройка ПО на основе ТИИ ИИ-1 на 100,0% чувствительность позволила корректно выявить 36,0% от всех исследований группы «норма». Пересчитав на объем всех исследований без патологии (66,6 млн – 90,0% от общего числа ФЛГ исследований), получим на общегосударственном уровне ежегодно 24 млн исследований категории «норма», не подлежащих пересмотру. Это в пять раз выше, по сравнению с объемом исследований с патологией, которые не подлежали бы пересмотру.

В настоящей работе не изучались другие показатели точности, связанные с классификацией патологии или с ее локализацией. Однако это и не является

критическим, если конечная цель ПО на основе ТИИ заключается в автономном поиске и описании исследований с «нормой».

В контексте автоматизации медицинских диагностических процессов уже реализовано автономное создание заключений по рентгенографии ОГК, когда изображения без патологий обрабатываются и описываются без участия врача, например, с помощью системы «ChestLink®». Эта система автоматически генерирует отчеты для случаев без патологии («норма») и перенаправляет снимки с признаками патологии врачу-рентгенологу для дальнейшего описания [204]. Результаты указывают на потенциал для широкого применения ТИИ в медицинской диагностике, подчеркивая необходимость доработки ПО для улучшения выявления «нормы» и донастройки порогов срабатывания алгоритмов под специфические клинические задачи. Полученные результаты указывают на возможность такой оптимизации и демонстрируют высокие показатели целесообразности применения ТИИ в качестве автономной медицинской услуги с высокими показателями диагностической эффективности.

Как подтвердили результаты проведенного исследования, ПО на основе ТИИ осуществляет подготовку описания ММГ-исследований быстрее, чем это делает врач-рентгенолог. Соответственно, врач, открывая исследование, уже видит результаты работы ПО на основе ТИИ – это и предполагается в качестве первого чтения. Проведенное исследование демонстрирует высокие показатели диагностической точности ПО на основе ТИИ, превосходящие отмеченные в литературе для врача-рентгенолога и сравнивающиеся с полученными в рамках вышеописанного исследования по оценке диагностической точности. На основании этого можно утверждать, что ТИИ позволят повысить выявляемость злокачественных новообразований молочной железы.

Так, в ходе ретроспективного анализа, лучший из алгоритмов смог достигнуть AUC, чувствительности и специфичности, соответствующим таковым для врача-рентгенолога. В результате была подготовлена технологическая карта новой медицинской услуги «Описание и интерпретация данных

маммографического исследования с использованием искусственного интеллекта» с кодом услуги 1601 (согласно Тарифному соглашению на 2023 год от 30 декабря 2022 года), которая и была внедрена в систему обязательного медицинского страхования.

Полученные предварительные данные продемонстрировали высокие показатели диагностической точности (проспективный AUC = 0,85), однако недостаточные для убедительного доказательства эффективности применения ТИИ. При этом, как следует из литературных данных, BI-RADS 3 характеризуется низкой (2%) вероятностью патологии [296]. Нельзя исключить, что ПО на основе ТИИ корректно классифицирует исследования с BI-RADS 3 на потенциально злокачественные и доброкачественные, как это уже было отмечено отдельными исследователями [297, 298]. В результате, показатели диагностической точности существенно улучшились, достигли значений в 0,95 и превышают таковые для «среднего» врача-рентгенолога.

В рамках исследования по оценке длительности подготовки описания профилактического исследования оценивался один из первых периодов внедрения ПО на основе ТИИ. Это позволило выделить и объективно сравнить группы врачей, которые не использовали ТИИ и которые использовали такие решения. В результате было показано, что для таких направлений как РГ/ФЛГ ОГК статистически значимых изменений в длительности подготовки протокола выявлено не было. Потенциально, в целевой модели применения ПО на основе ТИИ на потоке исследований это приведет к сокращению времени на подготовку описания, т.к. большая часть исследований с нормой будет отфильтровываться ПО. Для маммографии, в которой предполагается использование ПО на основе ТИИ в качестве первого чтения скорость подготовки описания возрастает, как было показано на 15%.

В заключение хотелось бы отметить, что настоящее исследование является прорывным, в том числе на международном уровне: в то время, как зарубежные исследователи только задаются вопросами о потенциальных перспективах и

целесообразности применения ТИИ для описания профилактических исследований [198], в рамках данной работы эти технологии уже были внедрены, и показана их эффективность. Таким образом, в настоящей работе не только были разработаны теоретические положения о применении ТИИ в качестве системы поддержки принятия врачебных решений и возможности автономной сортировки исследований, но и приведено решение научной проблемы в области применения технологий искусственного интеллекта в медицинской диагностике, позволяющее повысить качество профилактических и диагностических лучевых исследований. Анализируя перспективы дальнейшей разработки темы и учитывая тренды в научной литературе [299, 300], необходимо выделить два основных направления: расширение видов исследований, результаты которых будут анализироваться с применением ПО на основе ТИИ и реализация мероприятий по повышению диагностического качества работы ПО. Что касается других видов исследований, то это уже было показано в ряде смежных работ, а также в рамках Эксперимента. Так, предложенная методика многоступенчатого контроля качества продемонстрировала свою применимость не только для других проекционных видов исследований, но также с успехом применяется для исследований КТ и МРТ. При этом остаются открытыми вопросы моделей применения ПО на основе ТИИ для КТ органов грудной клетки, брюшной полости и т.д. Также остаются открытыми вопросы дальнейшего повышения диагностической точности ПО на основе ТИИ и реализация так называемого автономного ИИ.

ВЫВОДЫ

1. К основным проблемам управления массовыми профилактическими лучевыми исследованиями относятся: высокая нагрузка и недостаточный уровень качества работы врача-рентгенолога; риск ограничения доступности медицинской помощи в силу кадрового дефицита и несовершенства системы финансирования; необходимость обеспечения бесперебойности и надежности функционирования профилактического направления в здравоохранении. С 2020 по 2022 год в Москве отмечалось увеличение количества профилактических исследований в среднем на 85 %. Для совершенствования организации и управления целесообразна автоматизация отдельных процессов массовых профилактических исследований.

2. Технологии искусственного интеллекта применимы для автоматизации массовых профилактических и диагностических лучевых исследований, так как показатели диагностической точности врача и решения на основе искусственного интеллекта достоверно не различимы (AUC программного обеспечения на основе технологии искусственного интеллекта для маммографии – 0,96, флюорографии – 0,95, рентгенографии – 0,91; точность врача для маммографии – 0,93 ($p = 0,22$), флюорографии – 0,97 ($p = 0,12$) рентгенографии – 0,96 ($p = 0,46$)).

3. Разработана организационная технология автоматизации рабочего процесса врача-рентгенолога при проведении профилактических и диагностических исследований, основанная на бесшовной интеграции технологий искусственного интеллекта в информационную систему в сфере здравоохранения, унификации технических и диагностических требований к результатам работы программного обеспечения, совокупности методических мероприятий, которая позволяет врачу-рентгенологу пользоваться инструментами автоматизации рабочего процесса в привычной программной среде. Результативность применения выражается в сокращении времени подготовки протокола исследования на 15 % и достигается за счет автоматизации подготовки текстового описания исследования и предоставления дополнительного изображения с

выделенными патологическими областями. Данная технология была апробирована на 11 решениях на основе технологии искусственного интеллекта для рентгенографии и флюорографии легких, на 5 решениях для маммографии, а также показала применимость для других видов исследований.

4. Разработана методика и программные средства контроля качества программного обеспечения на основе технологий искусственного интеллекта для профилактических исследований, включающая следующие этапы: самотестирование, функциональное, калибровочное, контрольно-технологическое тестирование, мониторинг технологических и клинических показателей. Мероприятия по непрерывному повышению качества обеспечили рост диагностической точности программного обеспечения для анализа профилактических и диагностических лучевых исследований (для маммографии – на 25,0 %, флюорографии – на 41,0 %, рентгенографии – на 19,0 %).

5. Модель организации и проведения массовых профилактических и диагностических исследований с применением технологий искусственного интеллекта включает специальные сценарии автоматизации описания результатов профилактических лучевых исследований молочной железы и органов грудной клетки, совокупность мероприятий по их реализации, контролю качества. Модель реализована в виде медицинской услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта», обеспечивающей сокращение трудозатрат на описание маммографического исследования в 2 раза при сохранении диагностической точности двойного описания (чувствительность – 0,875, специфичность – 0,852).

6. Применение технологий искусственного интеллекта в качестве систем поддержки принятия врачебных решений для профилактической маммографии позволило сократить число клинически значимых пропусков патологии на 3,0 %.

7. Автоматизация описания результатов флюорографии/рентгенографии органов грудной клетки способствует повышению производительности врача-рентгенолога на 77,0 %.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. Органам исполнительной власти субъектов Российской Федерации в сфере охраны здоровья:

1.1 Технологии на основе искусственного интеллекта рекомендуется применять в качестве инструмента автоматизации при проведении массовых профилактических и диагностических лучевых исследований.

1.2 Программное обеспечение на основе технологий искусственного интеллекта для лучевой диагностики должно быть интегрировано в информационную систему в сфере здравоохранения субъекта РФ, а результаты его работы рекомендовано предоставлять в виде двух дополнительных серий: текстовой (DICOM SR) с описанием и заключением, графической (DICOM SC) с визуальным отображением выявленных патологических изменений.

2. Руководителям медицинских организаций:

2.1 Целесообразно организовать контроль качества работы программного обеспечения на основе технологий искусственного интеллекта на всем протяжении его жизненного цикла и обеспечить на этапе внедрения функциональное, калибровочное, контрольно-технологическое тестирование; на этапе эксплуатации – мониторинг технологических и клинических показателей.

2.2 В качестве минимальных показателей диагностической точности программного обеспечения на основе технологий искусственного интеллекта для интерпретации результатов массовых профилактических лучевых исследований рекомендовано использовать следующие показатели диагностической точности (AUC): для маммографии – 0,88, рентгенографии – 0,93, флюорографии – 0,94.

2.3 Рекомендовано использование программного обеспечения на основе технологий искусственного интеллекта в качестве первого чтения при анализе маммографии и в качестве инструмента фильтрации исследований без патологии при рентгенографии/флюорографии органов грудной клетки.

СПИСОК СОКРАЩЕНИЙ, АББРЕВИАТУР И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

АРМ	– автоматизированное рабочее место
БДТ	– базовые диагностические требования
БФТ	– базовые функциональные требования
ГБУЗ	– государственное бюджетное учреждение здравоохранения
«НПКЦ ДиТ ДЗМ»	города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы»
ДЗМ	– Департамент здравоохранения города Москвы
ДИ	– доверительный интервал
ДУ	– диагностическое(ие) устройство(а)
ЕГИСЗ	– Единая государственная информационная система в сфере здравоохранения
ЕМИАС	– Единая медицинская информационно-аналитическая система города Москвы
ЕРИС	– Единый радиологический информационный сервис
ЗНО	– злокачественные новообразования
ИИ	– искусственный интеллект
ИП	– истинно положительное
ИТ	– информационные технологии
ИО	– истинно отрицательное
КМОП	– комплементарная структура металл-оксид-полупроводник
ККТ	– контрольно-техническое тестирование
КТ	– компьютерная томография
ЛО	– ложноотрицательное
ЛП	– ложноположительное
МГМУ	– Московский государственный медицинский университет

МЖ	– молочная железа
МИС	– медицинская информационная система
МКБ-10	– Международная классификация болезней 10-го пересмотра
МО	– медицинская организация
ММГ	– маммография
МРТ	– магнитно-резонансная томография
МРЦЛД	– Московский референс-центр лучевой диагностики
НД	– набор данных
ОБП	– органы брюшной полости
ОГК	– органы грудной клетки
ОЛД	– отделение лучевой диагностики
ОМС	– обязательное медицинское страхование
ОНМК	– острое нарушение мозгового кровообращения
ПЗС	– приборы с зарядовой связью
ПО	– программное обеспечение
ППАК	– продуктивный программно-аппаратный комплекс
ПТСВ	– платформа тестирования при смене версионности
ПУМ	– продукт управления моделями
РГ	– рентгенография
РИС	– радиологическая информационная система
РМЖ	– ЗНО молочной железы
РФ	– Российская Федерация
СППВР	– система поддержки принятия врачебных решений
ТИИ	– технологии искусственного интеллекта
ТПАК	– тестовый программно-аппаратный контур
ФГБОУ ВО «МИРЭА	– федеральное государственное бюджетное образовательное учреждение высшего образования «Российский технологический университет»
ФЛГ	– флюорография

ФОТ	– фонд оплаты труда
ХНИЗ	– хронические неинфекционные заболевания
ЦАМИ	– Центральный архив медицинских изображений
ЭМК	– электронная медицинская карта
ЯНАО	– Ямало-Ненецкий автономный округ
API	– Application Programming Interface, описание способов взаимодействия одной компьютерной программы с другими
AUC	– Area Under Curve, площадь под кривой
BI-RADS	– Breast Imaging Reporting and Data System, система отчетности и данных по визуализации молочной железы
CAD	– computer aided diagnostic, компьютерное обнаружение, также называемое компьютерной диагностикой
CCD	– Charge-Coupled Device, прибор с зарядовой связью
CR	– Computed Radiography, компьютерная радиография
DICOM	– Digital Imaging and Communications in Medicine, стандарт создания, хранения, передачи и визуализации медицинских изображений и документов
DICOM SC	– secondary capture, дополнительное изображение/серия изображений в DICOM
DR Digital Radiography	– цифровая радиография
FDA	– Food and Drug Administration, управление по контролю качества пищевых продуктов и лекарственных средств
GT	– Graund Truth, эталонное значение разметки
HIS	– Hospital Information Systems, медицинская информационная система
HL7	– Health Level 7, седьмой уровень для здравоохранения
IHE	– Integration the Healthcare Enterprise, интеграция предприятий в здравоохранении

JSON	– JavaScript Object Notation, текстовый формат обмена данными, основанный на JavaScript
Kafka	– распределенный программный брокер сообщений с открытым исходным кодом
NLP	– Natural Language Processing, компьютерные технологии обработки естественного языка
OS	– operating system, операционная система (Windows, Linux, X)
PACS	– Picture Archiving and Communication System, система архивирования и передачи изображений
PNG	– Portable Network Graphics, растровый формат хранения графической информации
PSA	– Photo-Diode Sensors Array/ Multi-CCD Detector Method
PubMed	– биомедицинская онлайн-библиотека
REST	– Representational State Transfer, передача репрезентативного состояния
RIS	– Radiology Information System, Радиологическая информационная система
TFT	– Thin Film Transistor
XML	– eXtensible Markup Language, расширяемый язык разметки

СПИСОК ЛИТЕРАТУРЫ

1. Еганян Р. А., Калинина А. М., Измайлова О. В. [и др.]. Первичная и повторная диспансеризация определенных групп взрослого населения: динамика алиментарно-зависимых факторов риска хронических неинфекционных заболеваний // Профилактическая медицина. – 2019. – Т. 22, № 4. – С. 14–21.
2. Концевая А. В., Сапунова И. Д., Панов С. А. [и др.]. Использование ценовой политики в первичной профилактике хронических неинфекционных заболеваний: налоги на сахаросодержащие напитки // Профилактическая медицина. – 2019. – Т. 22, № 5. – С. 16–22.
3. Красникова Н. В., Шеметова Г. Н., Казбан О. Г. [и др.]. Медико-социальная эффективность работы центров здоровья по выявлению факторов риска основных хронических неинфекционных заболеваний // Саратовский научно-медицинский журнал. – 2019. – Т. – 15, № 1. – С. 77–82.
4. Лобеев А. В. Новые подходы в профилактике хронических неинфекционных заболеваний у лиц трудоспособного возраста // Медицина труда и промышленная экология. – 2020. – Т. 60, № 11. – С. 821–823.
5. Максикова Т. М., Калягин А. Н. Профилактика хронических неинфекционных заболеваний в России: возможности оптимизации // Менеджер здравоохранения. – 2019. – № 4. – С. 34–40.
6. Aceves B., Ingram M., Nieto C., et al. Non-communicable disease prevention in Mexico: policies, programs and regulations // Health Promot. Int. – 2020. – Vol. 35, № 2. – P. 409–421.
7. Lane M. M., Davis J. A., Beattie S., et al. Ultraprocessed food and chronic noncommunicable diseases: A systematic review and meta-analysis of 43 observational studies // Obes. Rev. – 2021. – Vol. 22, № 3.
8. Tapia-Conyer R., Saucedo-Martinez R., Mujica-Rosales R., et al. A Policy Analysis on the Proactive Prevention of Chronic Disease: Learnings from the

Initial Implementation of Integrated Measurement for Early Detection (MIDO) // *Int. J. Heal. Policy Manag.* – 2017. – Vol. 6, № 6. – P. 339–344.

9. Состояние онкологической помощи населению России в 2021 году / под ред. А. Д. Каприна, В. В. Старинского, А. О. Шахзадовой. – М., 2022. – 239 с.

10. Драпкина О. М., Концевая А. В., Калинина А. М. [и др.]. Профилактика хронических неинфекционных заболеваний в Российской Федерации. Национальное руководство 2022. Диспансеризации населения // *Кардиоваскулярная терапия и профилактика.* – 2022. – Т. 21. – № 4. – С. 5–232.

11. Кобякова О. С., Деев И. А., Куликов Е. С. [и др.]. Хронические неинфекционные заболевания: эффекты сочетанного влияния факторов риска // *Профилактическая медицина.* – 2019. – Т. 22. – № 2. – С. 45–50.

12. Buse K., Tanaka S., Hawkes S. Healthy people and healthy profits? Elaborating a conceptual framework for governing the commercial determinants of non-communicable diseases and identifying options for reducing risk exposure // *Global. Health.* – 2017. – Vol. 13, № 1. – P. 34.

13. Du M., Zhu L., Liu M., et al. Mutual Associations of Healthy Behaviours and Socioeconomic Status with Respiratory Diseases Mortality: A Large Prospective Cohort Study // *Nutrients.* – 2023. – Vol. 15, № 8. – P. 1872.

14. Williams J., Allen L., Wickramasinghe K., et al. A systematic review of associations between non-communicable diseases and socioeconomic status within low- and lower-middle-income countries // *J. Glob. Health.* – 2018. – Vol. 8, № 2.

15. Дроздова Л. Ю. Лищенко О. В., Раковская Ю. С. [и др.]. Деятельность центров общественного здоровья и медицинской профилактики при эпидемии COVID-19 // *Профилактическая медицина.* – 2020. – Т. 23, № 7. – С. 33–36.

16. Марцевич С. Ю., Кутишенко Н. П., Лукина Ю. В. [и др.]. Самоконтроль и лечение хронических неинфекционных заболеваний в условиях пандемии COVID-19. Консенсус экспертов Национального общества доказательной фармакотерапии и Российского общества профилактики

неинфекционных заболеваний // Кардиоваскулярная терапия и профилактика. – 2020. – Т. 19, № 3. – С. 295–301.

17. Фролова И. А., Абаева О. П., Романов С. В. [и др.]. Изменение доступности медицинской помощи пациентам с хроническими неинфекционными заболеваниями в период пандемии COVID-19: обзор литературы // Социология медицины. – 2022. – Т. 21, № 2. – С. 211–222.

18. Azarpazhooh M. R., Morovatdar N., Avan A., et al. COVID-19 Pandemic and Burden of Non-Communicable Diseases: An Ecological Study on Data of 185 Countries // J. Stroke Cerebrovasc. Dis. – 2020. – Vol. 29, № 9. – P. 105089.

19. Basu S. Non-communicable disease management in vulnerable patients during Covid-19 // Indian J. Med. Ethics. – 2020. – Vol. 5, № 2. – P. 103–105.

20. Palmer K., Monaco A., Kivipelto M., et al. The potential long-term impact of the COVID-19 outbreak on patients with non-communicable diseases in Europe: consequences for healthy ageing // Aging Clin. Exp. Res. 2020. – Vol. 32, № 7. – P. 1189–1194.

21. Голубев Н. А., Огрызко Е. В., Тюрина Е. М. [и др.]. Особенности развития службы лучевой диагностики в Российской Федерации за 2014–2019 года // Современные проблемы здравоохранения и медицинской статистики. – 2021. – № 2. – С. 356–376.

22. Тюрин И. Е. Лучевая диагностика в Российской Федерации // Онкологический журнал: лучевая диагностика, лучевая терапия. – 2018. – Т. 1, № 4. – С. 43–51.

23. Шелехов П. В. Кадровая ситуация в лучевой диагностике // Современные проблемы здравоохранения и медицинской статистики. – 2019. – № 1. – С. 265–275.

24. Шелехов П. В. Эффективность использования оборудования лучевой диагностики в субъектах Российской Федерации // Менеджер здравоохранения. – 2017. – № 5. – С. 33–41.

25. Рожкова Н. И., Ройтберг П. Г., Варфоломеева А. А. [и др.].

Сегментационная модель скрининга рака молочной железы на основе нейросетевого анализа рентгеновских изображений // Сеченовский вестник. – 2020. – Т. 11, № 3. – Р. 4–14.

26. Chen Y., James J. J., Michalopoulou E., et al. Performance of Radiologists and Radiographers in Double Reading Mammograms: The UK National Health Service Breast Screening Program // Radiology. – 2023. – Vol. 306, № 1. – P. 102–109.

27. Von Euler-Chelpin M., Lillholm M., Napolitano G., et al. Screening mammography: benefit of double reading by breast density // Breast Cancer Res. Treat. – 2018. – Vol. 171, № 3. – P. 767–776.

28. Posso M., Puig T., Carles M., et al. Effectiveness and cost-effectiveness of double reading in digital mammography screening: A systematic review and meta-analysis // Eur. J. Radiol. – 2017. – Vol. 96. – P. 40–49.

29. Schünemann H. J., Lerda D., Quinn C., et al. Breast Cancer Screening and Diagnosis: A Synopsis of the European Breast Guidelines // Ann. Intern. Med. – 2020. – Vol. 172, № 1. – P. 46.

30. Глушкова И. В., Кошкарров А. А., Мурашко Р. А. [и др.]. Региональная радиологическая информационная система Краснодарского края: организация работы референсного клинико-диагностического центра по патологии молочной железы // Врач и информационные технологии. – 2018. – № S1. – С. 18–27.

31. Кошкарров А. А., Мурашко Р. А., Елишев В. Г. [и др.]. Особенности распределенного хранения медицинских изображений в онкологической службе в рамках создания единого цифрового контура // Врач и информационные технологии. – 2020. – № S1. – С. 15–27.

32. Кошкарров А. А., Пеннер Д. В., Рубцова И. Т. Региональная радиологическая информационная система: новые возможности распределенного хранения медицинских изображений в Краснодарском крае // Тезисы докладов международного конгресса «Информационные технологии медицине». – 2017.

33. Пшекарский Д. С., Ющук В. Н. Опыт организации КТ-исследований в условиях эпидемии COVID-19 // Лучевая диагностика и терапия. – 2020. – № 2.

– С. 116–118.

34. Bashshur R. L., Krupinski E. A., Thrall J. H., et al. The Empirical Foundations of Teleradiology and Related Applications: A Review of the Evidence // *Telemed. e-Health*. – 2016. – Vol. 22, № 11. – P. 868–898.

35. Seo H. S., Park J. S., Oh Y.-Wh. et al. Peer Review of Teleradiology at a Teleradiology Clinic: Comparison of Unacceptable Diagnosis and Clinically Significant Discrepancy between Radiology Sections and Imaging Modalities // *J. Korean Soc. Radiol.* – 2021. – Vol. 82, № 6. – P. 1545.

36. Weikert T., Sommer G., Tamm M., et al. Centralized expert HRCT Reading in suspected idiopathic pulmonary fibrosis: Experience from an Eurasian teleradiology program // *Eur. J. Radiol.* – 2019. – Vol. 121. – P. 108719.

37. Гусев А. В., Астапенко Е. М., Иванов И. В. [и др.]. Принципы формирования доверия к системам искусственного интеллекта для сферы здравоохранения // *Вестник Росздравнадзора*. – 2022. – №2. – С. 25–33.

38. Collins G. S., Dhiman P., Navarro C. L. A., et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence // *BMJ Open*. – 2021. – Vol. 11, № 7. – P. e048008.

39. Hardy M., Harvey H. Artificial intelligence in diagnostic imaging: impact on the radiography profession // *Br. J. Radiol.* – 2020. – Vol. 93, № 1108. – P. 20190840.

40. Milne-Ives M., de Cock C., Lim E., et al. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review // *J. Med. Internet Res.* – 2020. – Vol. 22, № 10. – P. e20346.

41. Shahzad R., Ayub B., Siddiqui M. A. R. Quality of reporting of randomised controlled trials of artificial intelligence in healthcare: a systematic review // *BMJ Open*. – 2022. – Vol. 12, № 9. – P. e061519.

42. Stanfill M. H., Marc D. T. Health Information Management: Implications of Artificial Intelligence on Healthcare Data and Information Management // *Yearb.*

Med. Inform. – 2019. – Vol. 28, № 01. – P. 056–064.

43. Da Costa Vieira R. A., Biller G., Uemura G., et al. Breast cancer screening in developing countries // Clinics. – 2017. – Vol. 72, № 4. – P. 244–253.

44. Basu P., Ponti A., Anttila A., et al. Status of implementation and organization of cancer screening in The European Union Member States – Summary results from the second European screening report // Int. J. Cancer. – 2018. – Vol. 142, № 1. – P. 44–56.

45. Васильева И. А., Белиловский Е. М., Борисов С. Е. [и др.]. Заболеваемость, смертность и распространенность как показатели бремени туберкулеза в регионах ВОЗ, странах мира и в Российской Федерации. Ч. 1. Заболеваемость и распространенность туберкулеза // Туберкулез и болезни легких. – 2017. – Т. 95, № 6. – С. 9–21.

46. Пядушкина Е. А., Авксентьева М. В., Омеляновский В. В. Зарубежные рекомендации по раннему выявлению хронических неинфекционных заболеваний: опыт Великобритании и США // Медицинские технологии. Оценка и выбор. Фонд развития социальной политики и здравоохранения. – 2016. – № 3 (25). – С. 20–32.

47. Таранцова А. В. Пути оптимизации донозологического скрининга хронических неинфекционных заболеваний (к 10-летнему юбилею центров здоровья в России) // Евразийский кардиологический журнал. – 2019. – № S1. – С. 69–70.

48. Приказ Министерства здравоохранения Российской Федерации от 27.04.2021 № 404н «Об утверждении Порядка проведения профилактического медицинского осмотра и диспансеризации определенных групп взрослого населения» // Электронный фонд правовых и нормативно-технических документов [Электронный ресурс]. – URL: <https://docs.cntd.ru/document/607124051> (дата обращения: 20.06.2022).

49. Sun L, Legood R, Sadique Z, Dos-Santos-Silva I, Yang L. Cost-effectiveness of risk-based breast cancer screening programme, China. Bull World

Health Organ. 2018 Aug 1;96(8):568-577. doi: 10.2471/BLT.18.207944. Epub 2018 Jun 28. PMID: 30104797; PMCID: PMC6083393. <https://doi.org/10.2471%2FBLT.18.207944>.

50. National cancer control programmes: policies and managerial guidelines // World Health Organization [Электронный ресурс]. – 2002. – URL: <https://apps.who.int/iris/handle/10665/42494> (дата обращения: 19.06.2022).

51. Датий А. В., Фесюн О. А. Актуальность классической рентгенологии в клинической практике // Актуальные вопросы медицинской практики: сборник научных трудов. – М.: ООО «Медицина», 2018. – С. 28–32.

52. Барковский А. Н., Ахматдинов Р. Р. [и др.]. Дозы облучения населения Российской Федерации в 2019 году: информационный сборник. – СПб.: НИИРГ им. проф. Рамзаева, 2020. – 70 с.

53. Black W. C., Gareen I. F., Soneji S. S., et al. Cost-Effectiveness of CT Screening in the National Lung Screening Trial // N. Engl. J. Med. – 2014. – Vol. 371, № 19. – P. 1793–1802.

54. Manser R., Irving L. B., Stone C., et al. Screening for lung cancer // Cochrane database Syst. Rev. – 2013. – Vol. 2013, № 6.

55. Aberle D. R., Adams A. M., Berg Ch. D., et al. National Lung Screening Trial Research T // Reduc. lung-cancer Mortal. with low-dose Comput. Tomogr. Screening // N Engl J Med. – 2011. – Vol. 365, № 5. – P. 395–409.

56. Ямщиков О. Н., Дробышева Н. И., Емельянова Н. В. [и др.]. Ранняя диагностика рака легких. Литературный обзор // Онкологический журнал: лучевая диагностика, лучевая терапия. – 2022. – Т. 5, № 1. – С. 74–82.

57. Струнина А. А., Соколов Д. В., Соколов В. В. [и др.]. Современные подходы к диагностике предопухолевой патологии и раннего рака легкого // Хирургия. Журнал им. НИ Пирогова. – 2018. – № 3–2. – С. 27–32.

58. Bajre M. K., Pennington M., Woznitza N. et al. Expanding the role of radiographers in reporting suspected lung cancer: a cost-effectiveness analysis using a decision tree model // Radiography. – 2017. – Vol. 23, № 4. – P. 273–278.

59. De Groot P. M., Brett W., Carter et al. Pitfalls in Chest Radiographic Interpretation: Blind Spots // *Semin. Roentgenol.* – 2015. – Vol. 50, № 3. – P. 197–209.
60. Waite S., Scott J., Gale B. et al. Interpretive error in radiology // *Am. J. Roentgenol.* – 2017. – Vol. 208, № 4. – P. 739–749.
61. Fardanesh M., White C. Missed Lung Cancer on Chest Radiography and Computed Tomography // *Semin. Ultrasound, CT MRI.* – 2012. – Vol. 33, № 4. – P. 280–287.
62. Raoof S., Feigin D., Sung A., et al. Interpretation of Plain Chest Roentgenogram // *Chest.* – 2012. – Vol. 141, № 2. – P. 545–558.
63. Ciatto S., Ambrogetti D., Bonardi R., et al. Second reading of screening mammograms increases cancer detection and recall rates. Results in the Florence screening programme // *J. Med. Screen.* – 2005. – Vol. 12, № 2. – P. 103–106.
64. Морозов С. П., Ветшева Н. Н., Диденко В. В. Организация программы популяционного скрининга злокачественных новообразований молочной железы среди женского населения: методические рекомендации / Серия «Лучшие практики лучевой диагностики». – М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2020. – 44 с.
65. Evans K. K., Birdwell R. L., Wolfe J. M. If You Don't Find It Often, You Often Don't Find It: Why Some Cancers Are Missed in Breast Cancer Screening // *PLoS One* / ed. M. J. Proulx – 2013. – Vol. 8, № 5. – P. e64366.
66. Морозов С. П., Владзимирский А. В., Ледихова Н. В. Телерадиология в Российской Федерации: достигнутый уровень // *Врач и информационные технологии.* – 2019. – № 2. – С. 67–73.
67. Карпов А. В., Болотских В. В., Карпов Д. С. Значение цифрового флюорографического обследования в ранней диагностике легочных форм туберкулеза в современных условиях // *Туберкулез и социально значимые заболевания.* – 2019. – Т. 4. – С. 50–51.
68. Ломаков С. Ю., Строгонов Е. А. Современные аспекты организации профилактических лучевых исследований методом флюорографии // *Международный академический вестник.* – 2019. – Т. 11(43). – С. 4–6.

69. Сайфулин М.Х., Аверенкова Н. С., Тарасова Л. Г. Эффективность массовой флюорографии как метода раннего выявления туберкулеза легких // Медицинский Альянс. – 2015. – Т. 1. – С. 50–51.
70. Александрович С. С., Геннадьевич Р. С., Валерьевна О. О. Медико-экономическая эффективность выявления больных туберкулезом методом флюорографии с использованием двойного независимого чтения // Социальные аспекты здоровья населения. – 2013. – Т. 34, № 6. – С. 1–10.
71. Цыбикова Э. Б., Зубова Н. А. Оценка эффективности массовых периодических осмотров, направленных на выявление туберкулеза // Туберкулез и болезни легких. – 2016. – Т. 94, № 4. – С. 13–19.
72. Yi T., Pan I., Collins S., et al. DICOM Image ANalysis and Archive (DIANA): an Open-Source System for Clinical AI Applications // J. Digit. Imaging. – 2021. – Vol. 34, № 6. – P. 1405–1413.
73. Fazal M. I., Patel E. M., Tye J., et al. The past, present and future role of artificial intelligence in imaging // Eur. J. Radiol. – 2018. – Vol. 105. – P. 246–250.
74. Pierce J. D., Rosipko B., Youngblood L., et al. Seamless Integration of Artificial Intelligence Into the Clinical Environment: Our Experience With a Novel Pneumothorax Detection Artificial Intelligence Algorithm // J. Am. Coll. Radiol. – 2021. – Vol. 18, № 11. – P. 1497–1505.
75. Mun S. K., Wong K. H., Lo Sh.-Ch. B., et al. Artificial Intelligence for the Future Radiology Diagnostic Service // Front. Mol. Biosci. – 2020. – Vol. 7. – P. 614258.
76. Haubold J. Artificial Intelligence in radiology: What can be expected in the next few years? // Radiologe. – 2020. – Vol. 60, № 1. – P. 64–69.
77. Nguyen N. H., Nguyen H. Q., Nguyen N. T., et al. Deployment and validation of an AI system for detecting abnormal chest radiographs in clinical settings // Front. Digit. Heal. – 2022. – Vol. 4.
78. Шурыгин В. П., Дергилев А. П., Зеленский М. И. Цифровая рентгенография в клинической практике // Вестник радиологии и рентгенологии.

– 1992. – № 1. – С. 27.

79. Линденбратен Л. Д. Очерки истории российской рентгенологии: славное прошлое, неожиданное настоящее, предвидимое будущее. – М.: ООО «Видар», 1995. – 283 с.

80. Портной Л. М. Сегодняшнее место цифровой рентгенологии в диагностике легочной патологии // Медицинская газета. – 1999. – № 58. – С. 10–11.

81. Блинов Н. Н., Мазуров А. И. О национальном проекте «Здоровье» в свете технического переоснащения рентгенологической службы России // Радиология-практика. – 2007. – № 2. – С. 53–56.

82. Портной Л. М., Вяткина Е. И., Стащук Г. А. К вопросу организации и внедрения в практическое здравоохранение России цифровой рентгенографии легких // Вестник радиологии и рентгенологии. – 2000. – № 5. – С. 10–19.

83. Sonoda M., Takano M., Miyahara J., et al. Computed radiography utilizing scanning laser stimulated luminescence // Radiology. – 1983. – Vol. 148, № 3. – P. 833–838.

84. Street R. A., Nelson S., L. Antonuk, et al. Amorphous silicon sensor arrays for radiation imaging // MRS Online Proc. – 1990. – Vol. 192. – P. 441–452.

85. Floyd Jr C. E., Warp R. J., Dobbins 3rd J. T., et al. Imaging characteristics of an amorphous silicon flat-panel detector for digital chest radiography // Radiology. – 2001. – Vol. 218, № 3. – P. 683–688.

86. Kotter E., Langer M. Digital radiography with large-area flat-panel detectors // Eur. Radiol. – 2002. – Vol. 12, № 10. – P. 2562–2570.

87. Мазуров А. И. Семинар «Технические средства рентгенологии» // Медицинская техника. – 1999. – № 2. – Р. 48.

88. Блинов Н. Н., Мазуров А. И. Что впереди? // Медицинская техника. – 2006. – № 5. – С. 3–6.

89. Белова И. Б., Китаев В. М. Сравнительная оценка флюорографии, пленочной рентгенографии и малодозовой цифровой рентгенографии грудной

полости // Проблемы туберкулеза. – 2000. – № 6. – С. 23–27.

90. Бердяков Г. И., Ртищева Г. М., Кокуев А. Н. Особенности конструкции и использования цифровых рентгеновских аппаратов для исследования легких // Медицинская техника. – 1998. – № 5. – С. 35–40.

91. Бучнев А. А., Шурыгин В. П. Цифровая рентгенография как новый метод диагностической визуализации // Труды 7-й международной конференции по компьютерной графике и визуализации. – 1997. – С. 177–180.

92. Чикирдин Э. Г. Развитие цифровой техники для рентгенодиагностики // Медицинская техника. – 1998. – № 3. – С. 36–39.

93. Блинов Н. Н. Проблемы переоснащения службы лучевой диагностики после выполнения Федеральной программы «Здоровье» // Радиология-практика. – 2008. – № 5. – С. 57–62.

94. Мазуров А. И., Пахарьков Г. Н. Оптимизация технических средств рентгенологической службы лечебно-профилактических учреждений. – СПб.: «Книжный Дом», 2008. – 352 с.

95. Рожкова Н. И., Кочетова Г. П. Динамика развития диагностической службы Российской Федерации за 2002–2010 гг. // Медицинская визуализация. – 2012. – № 4. – С. 11–16.

96. Зеликман М. И. Цифровые системы в медицинской рентгенодиагностике. – Медицина, 2007. – 207 с.

97. Matsuura N., Zhao W., Huang Z. et al. Digital radiology using active matrix readout: Amplified pixel detector array for fluoroscopy // Med. Phys. – 1999. – Vol. 26, № 5. – P. 672–681.

98. Lu J. P., van Schuylenbergh K., Ho J., et al. Flat panel imagers with pixel level amplifiers based on polycrystalline silicon thin-film transistor technology // Appl. Phys. Lett. – 2002. – Vol. 80, № 24. – P. 4656–4658.

99. Элинсон М. Б. Анализ преимуществ цифровых рентгеновских аппаратов перед пленочными // Медицинская техника. – 2005. – № 5. – С. 37–39.

100. Kim H. K., Cunningham I., Yin Zh., et al. On the development of digital

radiography detectors: A review // Int. J. Precis. – 2008. – Vol. 9. – С. 86–100.

101. Чернецов В. Б. Обзор рынка рентгеновского оборудования или как выбрать рентгеновский аппарат // Поликлиника. – 2018. – № 5–1. – С. 18–22.

102. Iared W., Shigueoka D. C., Torloni M. R., et al. Avaliação comparativa entre a mamografia digital e mamografia em filme: revisão sistemática e metanálise // Sao Paulo Med. J. – 2011. – Vol. 129, №4.

103. Држевецкая К. С. Обзор подходов к массовому скринингу рака молочной железы в России и мире // Российский электронный журнал лучевой диагностики. – 2020. – Т. 10, № 4. – С. 225–236.

104. Корженкова Г. П., Долгушин Б. И. Опыт использования цифровой маммографии // Опухоли женской репродуктивной системы. – 2011. – № 1. – С. 37–41.

105. Мануйлова О. О. Сравнительный анализ диагностических возможностей CR-системы и переносного плоскопанельного цифрового детектора для маммографии // Радиология-практика. – 2021. – № 3. – С. 24–30.

106. Jayadevan R., Armada M. J., Shaheen R. et al. Optimizing Digital Mammographic Image Quality for Full-Field Digital Detectors: Artifacts Encountered during the QC Process // RadioGraphics. – Vol. 35, № 7. P. 2080–2089.

107. Акиев Р. М., Труфанов Г. Е., Багненко С. С. Лучевая диагностика. – М.: ГЭОТАР-Медиа, 2007. – 416 с.

108. Баженова Ю. В. Современные аспекты деятельности службы лучевой диагностики в Российской Федерации // Сибирский медицинский журнал. – 2015. – Т. 134, № 3. – С. 78–81.

109. Федеральный закон «Об основах охраны здоровья граждан в Российской Федерации» от 21.11.2011 № 323-ФЗ // Электронный фонд правовых и нормативно-технических документов [Электронный ресурс]. – URL: <https://docs.cntd.ru/document/902312609> (дата обращения: 15.06.2022).

110. Шахабов И. В., Мельников Ю. Ю., Смышляев А. В. Трендовый анализ деятельности подразделений лучевой диагностики в медицинских организациях,

оказывающих помощь в амбулаторных условиях // Вестник Ивановской медицинской академии. – 2020. – Т. 25, № 1. – С. 17–19.

111. Программа модернизации первичного звена здравоохранения Московской области на 2021–2025 годы [Электронный ресурс]. URL: <https://mz.mosreg.ru/download/document/9127123> (дата обращения: 21.07.2022).

112. Мазуров А. И. Борьба с рассеянным излучением в цифровых рентгеновских аппаратах // Увидеть невидимое: сб. науч. трудов. – 2012. – Вып. 2. – С. 85–97.

113. Кривошеев И. В. Гатиятуллин А. М., Халиков Э. Ф. [и др.]. Рациональное использование рентгенологического оборудования // ХСІХ Международные научные чтения (Памяти И. И. Сикорского). – М., 2021. – С. 64–69.

114. Блинов Н. Н., Мазуров А. И. Новые реальности в современной рентгенотехнике // Медицинская техника. – 2003. – № 5. – С. 3–6.

115. Elze T., Taylor C., Bex P. J. An evaluation of organic light emitting diode monitors for medical applications: Great timing, but luminance artifacts // Med. Phys. – 2013. – Vol. 40, № 9. – P. 92701.

116. Marchessoux C., de Paepe L., Vanovermeire O., et al. Clinical evaluation of a medical high dynamic range display // Med. Phys. – 2016. – Vol. 43. – P. 4023–4031.

117. Камышанская И. Г., Черемисин В. М. Результаты клинического использования метода цветового контрастирования цифровых рентгенограмм // Лучевая диагностика и терапия. – 2022. – Т. 12, № 4. – С. 83–98.

118. Mustra M., Delac K., Grgic M. Overview of the DICOM standard // 2008 50th International Symposium ELMAR. – 2008. – Vol. 1. – P. 39–44.

119. The DICOM Standard: Current Edition [Электронный ресурс]. URL: <https://www.dicomstandard.org/current> (дата обращения: 08.01.2023).

120. Каперусов С. Ю. PACS-система архивирования и передачи изображений в лучевой диагностике // Радиология-практика. – 2007. – № 3.

– С. 69–73.

121. Bauman R. A., Gell G., Dwyer S. J. Large picture archiving and communication systems of the world. Part 1 // J. Digit. Imaging. – 1996. – Vol. 9, № 3. – P. 99–103.

122. Морозов С. П., Переверзев М. О. Обзор текущего состояния и основных требований к PACS-системам // Врач и информационные технологии. – 2013. – № 3. – С. 17–29.

123. Камышанская И. Г., Емельянов О. В., Черемисин В. М. [и др.]. Настоящее и будущее радиологической информационной системы Мариинской больницы // Сб. материалов конференции. – 2012.

124. Черемисин В. М., Камышанская И. Г., Элинсон А. С. [и др.]. PACS компании «Электрон» // Здоровье и болезнь. Спецвыпуск 4. Алматинский радиологический форум «Инновационные диагностические технологии в медицине» (Алматы, 14 июня 2012). – 2012. – № 3 (105). – С. 35–45.

125. Радиологическая информационная система «АПК АрхиМед» [Электронный ресурс]. – URL: <http://med-ray.ru> (дата обращения: 04.01.2023).

126. Луч-С // Tiara Medical: официальный сайт [Электронный ресурс]. – URL: <https://www.tiaramed.ru/catalog/sistemy-obrabotki-izobrazheniy/luch-s> (дата обращения: 04.01.2023).

127. PACS Система // Компания «Комета»: официальный сайт [Электронный ресурс]. – URL: <http://www.kometa-med.ru/company> (дата обращения: 04.01.2023).

128. Комплекс программ Sciberia // Sciberia: официальный сайт [Электронный ресурс]. – URL: <https://sciberia.io/ru> (дата обращения: 04.01.2023).

129. Впервые в России разработан PACS на российском программном и аппаратном обеспечении // АРМИТ: Ассоциация развития медицинских информационных технологий: официальный сайт [Электронный ресурс]. – URL: <https://www.armit.ru/news/9165-vpervye-v-rossii-razrabotan-pacs-na-rossiyskom-programmnom-i-apparatnom-obespechenii> (дата обращения: 04.01.2023).

130. Orthanc – DICOM Server // Orthanc: официальный сайт [Электронный ресурс]. – URL: <https://www.orthanc-server.com> (дата обращения: 01.12.2020).
131. Jodogne S. The Orthanc Ecosystem for Medical Imaging // J. Digit. Imaging. – 2018. – Vol. 31, № 3. – P. 341–352.
132. Koenker R. Improved conspicuity of key X-ray findings using advanced post-processing techniques: clinical examples // Kontraste (Hamburg). – 2005. – Vol. 49, – № 3. – P. 4–11.
133. Sheng B.-yong, Mao L., Zhou Sh., et al. Optimization of digital chest radiography image post-processing in diagnosis of pneumoconiosis // Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi. Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi. – 2013. – Vol. 31, № 11. – P. 825–829.
134. Kristensen S. V., Outzen C., Grau L. M., et al. Can advanced edge enhancement software improve image quality to visualise tubes, catheters and wires in digital chest radiographs? // Radiography. – 2022. – Vol. 29, № 1. – P. 165–170.
135. Камышанская И. Г. Роль инновационных цифровых технологий в оптимизации лучевой диагностики стационара (опыт внедрения ПАКС) // Российский электронный журнал лучевой диагностики. – 2016. – Т. 6, № 3. – С. 88–105.
136. Морозов С. П., Переверзев М. О. Лучевая диагностика – авангард информатизации здравоохранения // Российский электронный журнал лучевой диагностики. – 2013. – Т. 3, № 3. – С. 41–50.
137. Дубровин А. В., Кошкарров А. А. От PACS к телерадиологии // Врачи и информационные технологии. – 2017. – № 3. – С. 106–111.
138. Imhof H. et al. Change in process management by implementing RIS, PACS and flat-panel detectors // Radiologe. – 2002. – Vol. 42, № 5. – P. 344–350.
139. Bakker A.R. HIS, RIS, and PACS // Comput. Med. Imaging Graph. – 1991. Vol. 15, № 3. – P. 157–160.
140. Морозов С. П., Владзимирский А. В., Ледихова Н. В. [и др.]. Телемедицинские технологии (телерадиология) в службе лучевой диагностики /

Серия «Лучшие практики лучевой и инструментальной диагностики». – М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2018. – 58 с.

141. Постановление Правительства Российской Федерации от 05.05.2018 №555 «О единой государственной информационной системе в сфере здравоохранения» // Электронный фонд правовых и нормативно-технических документов [Электронный ресурс]. – URL: <https://docs.cntd.ru/document/557308809> (дата обращения: 15.02.2022).

142. Постановление Правительства Российской Федерации от 12.04.2018 № 447 «Об утверждении Правил взаимодействия иных информационных систем, предназначенных для сбора, хранения, обработки и предоставления информации, касающейся деятельности медицинских организаций и предоставляемых ими услуг, с информационными системами в сфере здравоохранения и медицинскими организациями» // Электронный фонд правовых и нормативно-технических документов [Электронный ресурс]. – URL: <https://docs.cntd.ru/document/557172596> (дата обращения: 15.01.2022).

143. Приказ Министерства здравоохранения Российской Федерации от 07.09.2020 №947н «Об утверждении Порядка организации системы документооборота в сфере охраны здоровья в части ведения медицинской документации в форме электронных документов» // Электронный фонд правовых и нормативно-технических документов [Электронный ресурс]. – URL: <https://docs.cntd.ru/document/565911145> (дата обращения: 15.01.2022).

144. Приказ Министерства здравоохранения Российской Федерации от 09.06.2020 №560н «Об утверждении Правил проведения рентгеновских исследований» // Электронный фонд правовых и нормативно-технических документов [Электронный ресурс]. – URL: <https://docs.cntd.ru/document/565342962> (дата обращения: 15.01.2022).

145. Галаянц С. Врач может описывать исследование там, где ему удобно // Журнал Vademecum. – 2020. – 28 декабря [Электронный ресурс]. – URL: https://vademec.ru/article/sergey_morozov_-_vrach_mozhet_opisyvat_issledovanie_tam-

_gde_ему_удобно (дата обращения: 08.01.2023).

146. Deborah R. B. Buyers move cautiously into digital radiography // *Diagn. Imaging*. – 2001. – P. 51–54.
147. Lodwick G. S., Keats T. E., Dorst J. P. The coding of roentgen images for computer analysis as applied to lung cancer // *Radiology*. – 1963. – Vol. 81, № 2. – P. 185–200.
148. Qin C., Yao D., Shi Y., et al. Computer-aided detection in chest radiography based on artificial intelligence: a survey // *Biomed. Eng. Online*. – 2018. – Vol. 17, № 1. – P. 1–23.
149. Jaeger S., Karargyris A., Antani S., et al. Detecting tuberculosis in radiographs using combined lung masks // *Annu Int Conf IEEE Eng Med Biol Soc*. – 2012. – P. 4978–4981.
150. Kulkarni S., Jha S. Artificial intelligence, radiology, and tuberculosis: a review // *Acad. Radiol*. – 2020. – Vol. 27, № 1. – P. 71–75.
151. Xu T. et al. Novel coarse-to-fine dual scale technique for tuberculosis cavity detection in chest radiographs // *EURASIP J. Image Video Process*. – 2013. – Vol. 2013. – № 1. – P. 1–18.
152. Jaeger S., Karargyris A., Candemir S., et al. Automatic tuberculosis screening using chest radiographs // *IEEE Trans. Med. Imaging*. – 2013. – Vol. 33, № 2. – P. 233–245.
153. Hogeweg L., Mol Ch., de Jong P. A., et al. Fusion of local and global detection systems to detect tuberculosis in chest radiographs // *International conference on medical image computing and computer-assisted intervention*. – 2010. – Vol. 13, № 3. – P. 650–657.
154. Rahman M. T., Codlin A. J., Rahman Md M., et al. An evaluation of automated chest radiography reading software for tuberculosis screening among public- and private-sector patients // *Eur. Respir*. – 2017. – Vol. 49, № 5. – P. 1602159.
155. Melendez J., Sánchez C. I., Philipsen R. H. H., et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and

clinical information // Sci. Rep. – 2016. – Vol. 6. – P. 25265.

156. McAdams H. P., Samei Eh., Dobbins J. 3rd, et al. Recent advances in chest radiography // Radiology. – 2006. – Vol. 241, № 3. – P. 663–683.

157. Chen J. J., White C. S. Use of CAD to Evaluate Lung Cancer on Chest Radiography // J. Thorac. Imaging. – 2008. – Vol. 23, № 2. – P. 93–96.

158. White C. S., Flukinger Th., Jeudy J., et al. Use of a computer-aided detection system to detect missed lung cancer at chest radiography // Radiology. – 2009. – Vol. 252, № 1. – P. 273–281.

159. Van Ginneken B., Hogeweg L., Prokop M. Computer-aided diagnosis in chest radiography: Beyond nodules // Eur. J. Radiol. – 2009. – Vol. 72, № 2. – P. 226–230.

160. Van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning // Radiol. Phys. – 2017. – Vol. 10, № 1. – P. 23–32.

161. Jiang Y., Nishikawa R. M., Wolverton D. E., et al. Malignant and benign clustered microcalcifications: automated feature analysis and classification // Radiology. – 1996. – Vol. 198, № 3. – P. 671–678.

162. Medical Devices; List of Premarket Approval Actions // Federal Register [Электронный ресурс]. – 1998. – December 30. – URL: <https://www.federalregister.gov/documents/1998/12/30/98-34347/medical-devices-list-of-premarket-approval-actions> (accessed: 08.01.2023).

163. Rao V. M., Levin D. C., Parker L., et al. How widely is computer-aided detection used in screening and diagnostic mammography? // J. Am. Coll. Radiol. – 2010. – Vol. 7, № 10. – P. 802–805.

164. Ko J. M., Nicholas M. J., Mendel J. B., et al. Prospective assessment of computer-aided detection in interpretation of screening mammography // Am. J. Roentgenol. – 2006. – Vol. 187, № 6. – P. 1483–1491.

165. Taplin S. H., Rutter C. M., Lehman C. D. Testing the effect of computer-assisted detection on interpretive performance in screening mammography // Am. J.

Roentgenol. – 2006. – Vol. 187, № 6. – P. 1475–1482.

166. Gur D., Sumkin J. H. CAD in screening mammography // Am. J. Roentgenol. – 2006. – Vol. 187, № 6. – P. 1474.

167. Nishikawa R. M., Schmidt R. A., Linver M. N., et al. Clinically missed cancer: how effectively can radiologists use computer-aided detection? // Am. J. Roentgenol. – 2012. – Vol. 198, № 3. – P. 708–716.

168. Fenton J.J. et al. Effectiveness of computer-aided detection in community mammography practice // J. Natl. Cancer Inst. – 2011. – Vol. 103, № 15. – P. 1152–1161.

169. Cole E.B. et al. Impact of computer-aided detection systems on radiologist accuracy with digital mammography // AJR. Am. J. Roentgenol. – 2014. – Vol. 203, № 4. – P. 909.

170. Kohli A., Jha S. Why CAD Failed in Mammography // J. Am. Coll. Radiol. – 2018. – Vol. 15, № 3. – P. 535–537.

171. Chan H.-P., Samala R. K., Hadjiiski L. M. CAD and AI for breast cancer-recent development and challenges // Br. J. Radiol. – 2020. – Vol. 93, № 1108. – P. 20190580.

172. Oakden-Rayner L. Exploring Large-scale Public Medical Image Datasets // Acad. Radiol. – 2020. – Vol. 27, № 1. – P. 106–112.

173. Wang X., Peng Y., Lu L. et al. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – May. – P. 3462–3471.

174. Irvin J., Rajpurkar P., Ko M., et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison // Proc. AAAI Conf. Artif. Intell. – 2019. – Vol. 33, № 01. – P. 590–597.

175. Johnson A.E.W.W., Pollard T. J., Berkowitz S. J., et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports // Sci. Data. – 2019. – Vol. 6, № 1. – P. 317.

176. Bustos A., Pertusa A., Salinas J.-M., et al. PadChest: A large chest x-ray image dataset with multi-label annotated reports // *Med Image Anal.* – 2020. – Vol. 66. – P. 101797.
177. Kelly B. S., Judge C., Bollard S. M., et al. Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE) // *Eur. Radiol.* – 2022. – Vol. 32, № 11. – P. 7998–8007.
178. Основы менеджмента медицинской визуализации / под ред. проф. С. П. Морозова. – М.: ГЭОТАР-Медиа, 2020. – 432 с.
179. Ranschaert E. R., Morozov S., Algra P. R. Artificial intelligence in medical imaging: opportunities, applications and risks. – Springer, 2019. – 373 p.
180. Beam A. L., Kohane I. S. Big Data and Machine Learning in Health Care // *JAMA.* – 2018. – Vol. 319, № 13. – P. 1317.
181. Gianfrancesco M. A., Tamang S., Yazdany J., et al. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data // *JAMA Intern. Med.* – 2018. – Vol. 178, № 11. – P. 1544.
182. Pande T., Cohen C., Pai M., et al. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review // *Int. J. Tuberc. Lung Dis.* – 2016. – Vol. 20, № 9. – P. 1226–1230.
183. Harris M., Qi A., Jeagal L., et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis // *PLoS One.* – 2019. – Vol. 14, № 9. – P. e0221339.
184. Santosh K., Allu S., Rajaraman, S. et al. Advances in Deep Learning for Tuberculosis Screening using Chest X-rays: The Last 5 Years Review // *J. Med. Syst.* – 2022. – Vol. 46, № 11. – P. 82.
185. Aggarwal R., Sounderajah V., Martin Gю, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis // *NPJ Digit Med.* – 2021. – Vol. 4, № 1. – P. 65.
186. Siddiqui S., Arifeen M., Hopgood A., et al. Deep Learning Models for the Diagnosis and Screening of COVID-19: A Systematic Review // *SN Comput. Sci.* –

2022. – Vol. 3, № 5. – P. 397.

187. Meedeniya D., Kumarasinghe H., Kolonne Sh., et al. Chest X-ray analysis empowered with deep learning: A systematic review // *Appl. Soft Comput.* – 2022. – Vol. 126. – P. 109319.

188. Tzeng I.-S., Hsieh P.-Ch., Su W.-L., et al. Artificial Intelligence-Assisted Chest X-ray for the Diagnosis of COVID-19: A Systematic Review and Meta-Analysis // *Diagnostics.* – 2023. – Vol. 13, № 4. – P. 584.

189. Jia L.-L., Zhao J.-X., Pan N. N., et al. Artificial intelligence model on chest imaging to diagnose COVID-19 and other pneumonias: A systematic review and meta-analysis // *Eur. J. Radiol. Open.* – 2022. – Vol. 9. – P. 100438.

190. Schaffter T., Buist D. S. M., Lee Ch. I., et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms // *JAMA.* – 2020. – Vol. 3, № 3. – P. e200265.

191. Liu X., Faes L., Kale A. U., et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis // *Lancet Digit. Heal.* – 2019. – Vol. 1, № 6. – P. e271–e297.

192. Rodriguez-Ruiz A., Lang K., Gubern-Merida A., et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists // *JNCI J. Natl. Cancer Inst.* – 2019. – Vol. 111, № 9. – P. 916–922.

193. Chen Y., Stavropoulou Ch., Narasinkan R., et al. Professionals' responses to the introduction of AI innovations in radiology and their implications for future adoption: a qualitative study // *BMC Health Serv. Res.* – 2021. – Vol. 21, № 1. – P. 813.

194. Блинов Д.С., Лобищева А.Е., Варфоломеева А.А. [и др.]. Нейросетевая интерпретация рентгенологического изображения грудной клетки: современные возможности и источники ошибок // *Проблемы стандартизации в здравоохранении.* – 2019 – № 9–10. С. 4–9.

195. Жуков Е. А., Блинов Д. С., Леонтьев В. С. [и др.]. Система компьютерного зрения для анализа обзорных рентгенограмм органов грудной

клетки: возможности искусственного интеллекта в обнаружении патологических изменений и инородных тел // Врач. – 2020. – Т. 31, № 5. – С. 34–41.

196. Гаврилов П.В., Смольникова У.А. Оценка диагностической точности системы автоматического анализа цифровых рентгенограмм легких при выявлении округлых образований // Альманах клинической медицины. - 2021. - Т. 49. - №6. - С. 359-364.

197. Смольникова У.А. Сравнительный анализ выявления округлых образований легких на рентгенограмме врачом и программой автоматического анализа изображений, “Вестник Центрального научно-исследовательского института туберкулеза” // Вестник ЦНИИТ. – 2021 – № 1(2021). С. 81–82.

198. Ram S., Bodduluri S. Implementation of Artificial Intelligence – Assisted Chest X-ray Interpretation: It Is About Time // Ann. Am. Thorac. Soc. – 2023. – Vol. 20, № 5. – P. 641–642.

199. Hickman S. E., Woitek R., Le E. P. V., et al. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis // Radiology. – 2022. – Vol. 302, № 1. – P. 88–104.

200. Liu J., Lei J., Ou Y., et al. Mammography diagnosis of breast cancer screening through machine learning: a systematic review and meta-analysis // Clin. Exp. Med. – 2022. – October 15.

201. Rozhkova N. I., Roitberg P. G., Varfolomeeva A. A., et al. Neural network-based segmentation model for breast cancer X-ray screening // Sechenov Med. J. – 2021. – Vol. 11, № 3. – P. 4–14.

202. AI for radiology [Электронный ресурс]. – URL: https://grand-challenge.org/aiforradiology/?subspeciality=Chest&modality=Xray&ce_under=All&ce_class=All&fda_class=All&sort_by=lastmodified&search= (дата обращения: 24.12.2022).

203. CE Mark for First Autonomous AI Medical Imaging Application [Электронный ресурс]. – URL: <https://oxipit.ai/news/first-autonomous-ai-medical-imaging-application> (дата обращения: 24.12.2022).

204. Miró Catalina Q., Fuster-Casanovas A., Solé-Casals J., et al. Developing an Artificial Intelligence Model for Reading Chest X-rays: Protocol for a Prospective Validation Study // JMIR Res. Protoc. – 2022. – Vol. 11, № 11. – P. e39536.

205. Ahmad H. K., Milne M. R., Buchlak, Q. D. et al. Machine Learning Augmented Interpretation of Chest X-rays: A Systematic Review // Diagnostics. – 2023. – Vol. 13, № 4. – P. 743.

206. Кушнирчук И. И., Коваленко Е. И., Болгарев Д. В. [и др.]. Возможности искусственного интеллекта в мониторинге физического здоровья // Известия Российской Военно-медицинской академии. – 2019. – Т. 38, № S3. – С. 110–113.

207. Бородулина Е. А. Искусственный интеллект в выявлении туберкулеза: возможности и перспективы // Vrach (Doctor). – 2020. – Т. 31, № 5. С. 30–33.

208. Харлов В. В. Применение нейронных сетей для распознавания туберкулеза по флюорографии / Фундаментальные и прикладные научные исследования: актуальные вопросы, достижения и инновации: сб. ст. XXIX Международной научно-практической конференции. В 2 ч. Ч. 1. – Пенза: МЦНС «Наука и Просвещение», 2019. – С. 13–19.

209. Кривихин Д. С., Федорова А. Е., Котилевец И. Д. Разработка системы диагностики заболеваний, основанные на базах рентгенограмм / Хроники цифровых трансформаций: материалы 2-х межкафедральных круглых столов и межкафедрального научного семинара. – Волгоград: Изд. дом «Сириус», 2022. – С. 38–42.

210. Владзимирский А. В., Кудрявцев Н. Д., Кожихина Д. Д. [и др.]. Эффективность применения технологий искусственного интеллекта для двойных описаний результатов профилактических исследований легких // Профилактическая медицина. – 2022. – Т. 25, № 7. – С. 7.

211. Wu J. T., Wong K. L., Gur Y., et al. Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents // JAMA. – 2020. – Vol. 3, № 10. – P. e2022779.

212. Alexander A., Jiang A., Ferreira C., et al. An Intelligent Future for Medical Imaging: A Market Outlook on Artificial Intelligence for Medical Imaging // J. Am. Coll. Radiol. – 2020. – Vol. 17, № 1. – P. 165–170.
213. Majkowska A., Mittal S., Steiner D. F., et al. Chest Radiograph Interpretation with Deep Learning Models: Assessment with Radiologist-adjudicated Reference Standards and Population-adjusted Evaluation // Radiology. – 2020. – Vol. 294, № 2. – P. 421–431.
214. Adams S. J., Henderson R. D. E., Yi X., et al. Artificial Intelligence Solutions for Analysis of X-ray Images // Can Assoc Radiol J. – 2021. – Vol. 72, № 1. – P. 60–72.
215. McKinney S. M., Sieniek M., Godbole V., et al. International evaluation of an AI system for breast cancer screening // Nature. Nature Research. – 2020. – Vol. 577, № 7788. – P. 89–94.
216. Kim H.-E., Kim H. H., Han B.-K., et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study // Lancet Digit. Heal. – 2020. – Vol. 2, № 3. – P. e138–e148.
217. Leibig C., Brehmer M., Bunk S., et al. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis // Lancet Digit. Heal. – 2022. – Vol. 4, № 7. – P. e507–e519.
218. Elmore J. G., Jackson S. L., Abraham L., et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy // Radiology. – 2009. – Vol. 253, № 3. – P. 641–651.
219. Lehman C.D., Wellman R. D., Buist D. S. M., et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection // JAMA. – 2015. – Vol. 175, № 11. – P. 1828–1837.
220. Raya-Povedano J. L., Romero-Martín S., Elías-Cabot E., et al. AI-based Strategies to Reduce Workload in Breast Cancer Screening with Mammography and Tomosynthesis: A Retrospective Evaluation // Radiology. – 2021. – Vol. 300, № 1. – P. 57–65.

221. Khunte M. et al. Trends in clinical validation and usage of US Food and Drug Administration-cleared artificial intelligence algorithms for medical imaging // *Clinical Radiology*. – 2023. – Т. 78. – №. 2. – С. 123-129. <https://doi.org/10.1016/j.crad.2022.09.122AI>.

222. Seibert J. A. Projection X-ray Imaging: Radiography, Mammography, Fluoroscopy // *Health Phys.* – 2019. – Vol. 116, № 2. – P. 148–156.

223. 224. 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.

224. Romero-Martín S., Elías-Cabot E., Raya-Povedano J. L., et al. Stand-Alone Use of Artificial Intelligence for Digital Mammography and Digital Breast Tomosynthesis Screening: A Retrospective Evaluation // *Radiology*. – 2022. – Vol. 302, № 3. – P. 535–542.

225. Henriksen E. L., Carlsen J. F., Vejborg I. M., et al. The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review // *Acta radiol.* – 2019. – Vol. 60, № 1. – P. 13–18.

226. Морозов С. П., Владзимирский А. В., Ледихова Н. В. [и др.]. Использование сервисов на основе технологии искусственного интеллекта при проведении описаний рентгенологических снимков / Серия «Лучшие практики лучевой диагностики». – М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2020. – 48 с.

227. Thammarach P., Khaengthanyakan S., Vongsurakrai S., et al. AI Chest 4 All // 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020. – Vol. 2020. – P. 1229–1233.

228. Singh R., Kalra M. K., Nitiwarangkul Ch., et al. Deep learning in chest radiography: Detection of findings and presence of change // *Public Library of Science*. – 2018. – Vol. 13, № 10. – P. e0204155.

229. WHO issues first global report on Artificial Intelligence (AI) in health and six guiding principles for its design and use // WHO: official website. – 2021. – June 28

[Электронный ресурс]. – URL: <https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use> (дата обращения: 04.01.2023).

230. Van Leeuwen K. G., Matthieu S. S., Rutten J. C. M., et al. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence // *Eur. Radiol.* – 2021. – Vol. 31, № 6. – P. 3797–3804.

231. Recht M. P., Dewey M., Dreyer K., et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations // *Eur. Radiol.* – 2020. – Vol. 30, № 6. – P. 3576–3584.

232. Логунова Т. А., Арзамасов К. М., Кирпичев Ю. С. [и др.]. Методика опроса пациентов и врачей при внедрении интеллектуальных технологий в практическое здравоохранение // Центр диагностики и телемедицины: официальный сайт. – 2021. – 22 июня [Электронный ресурс]. – URL: <https://telemedai.ru/biblioteka-dokumentov/metodika-oprosa-pacientov-i-vrachej-pri-vnedrenii-intellektualnyh-tehnologij-v-prakticheskoe-zdravooohranenie> (дата обращения: 04.01.2023).

233. Browning T., O'Neill Th., Ng Y., et al. Special Considerations for Integrating Artificial Intelligence Solutions in Urban Safety-Net Hospitals // *J. Am. Coll. Radiol.* – 2020. – Vol. 17, № 1. – P. 171–174.

234. DICOM 3.0 Conformance Statement for DICOM. Anonymization Orchestrating Medical Imaging Workflow. – 2019. [Электронный ресурс]. – URL: <https://laurelbridge.com/pdf/Dicom-Anonymization-Conformance-Statement.pdf> (дата обращения: 23.03.2023).

235. Morozov S. P., Vladzimirsky A.V., Klyashtorny V. G., et al. Clinical acceptance of software based on artificial intelligence technologies (radiology) / Series “Best practices in medical imaging”. – 2019. – Vol. 57, № 495.

236. Youden W. J. Index for rating diagnostic tests // *Cancer.* – 1950. – Vol. 3, №1. – P. 32–35.

237. Zou K. H., Tempany C. M., Fielding J. R., et al. Original smooth receiver operating characteristic curve estimation from continuous data: Statistical methods for

analyzing the predictive value of spiral CT of ureteral stones // Acad. Radiol. – 1998. – Vol. 5, № 10. – P. 680–687.

238. Ruopp M. D., Perkins N. J., Whitcomb B. W., et al. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection // Biometrical J. – 2008. – Vol. 50, № 3. – P. 419–430.

239. El Naqa I., Boone J. M., Benedict S. H., et al. AI in medical physics: guidelines for publication // Med. Phys. – 2021. – Vol. 48, № 9. – P. 4711–4714.

240. SPSS Software // IBM [Электронный ресурс]. – URL: <https://www.ibm.com/spss> (дата обращения: 15.04.2023).

241. Stata 14 // Stata [Электронный ресурс]. – URL: <https://www.stata.com/stata14/> (дата обращения: 15.04.2023).

242. MedCalc software description [Электронный ресурс]. URL: <https://www.medcalc.org/features> (дата обращения: 15.04.2023).

243. Goksuluk D., Korkmaz S., Zararsiz G., et al. easyROC: An Interactive Web-tool for ROC Curve Analysis Using R Language Environment // R J. – 2016. – Vol. 8, № 2. – P. 213–230.

244. Pauly M., Asendorf T., Konietzschke F. Permutation-based inference for the AUC: A unified approach for continuous and discontinuous data // Biom. J. – 2016. – Vol. 58, № 6. – P. 1319–1337.

245. К. М. Арзамасов, С. С. Семенов, Д. Ю. Кокина [и др.]/ Критерии применимости компьютерного зрения для профилактических исследований на примере рентгенографии и флюорографии органов грудной клетки // Медицинская физика. – 2022. – № 4(96). – С. 56-63. – DOI 10.52775/1810-200X-2022-96-4-56-63.

246. Патент № 2697733 С1 Российская Федерация, МПК G06Т 1/40, G06Т 7/11, G06К 9/50. Система обработки рентгенографических изображений и вывода результата пользователю: № 2019118035: заявл. 10.06.2019: опубл. 19.08.2019 / И. И. Чернин, В. И. Чернин, С. И. Чернин, Николаев Н. А., Никитин Е. Д., Капнинский А. А.; заявитель ООО «Медицинские Скрининг Системы».

– 15 с.: ил.

247. Карпов О. Э., Бронов О. Ю., Капнинский А. А. [и др.]. Компаративное исследование результатов анализа данных цифровой маммографии системы на основе искусственного интеллекта «Цельс» и врачей-рентгенологов // Вестник Национального медико-хирургического Центра им. Н. И. Пирогова. – 2021. – Т. 16, № 2. – С. 86–92.

248. Павлович П. И., Бронов О. Ю., Капнинский А. А. [и др.]. Сравнительное исследование результатов анализа данных цифровой маммографии системы на основе искусственного интеллекта «Цельс» и врачей-рентгенологов // Digital Diagnostics. – 2021. – Т. 2, № 2S. – С. 22–23.

249. Андрианов Н. Г. [и др.]. Информационно-измерительная система автоматизированного выявления заболеваний органов грудной клетки на базе мобильного флюорографического комплекса // III-я Международная конференция «Инжиниринг & Телекоммуникации–2016». – Долгопрудный: МФТИ, 2016. – Сер. 30. – С. 185–191.

250. Морозов С. П., Владзимирский А. В., Ледихова Н. В. [и др.]. Оценка диагностической точности системы скрининга туберкулеза легких на основе искусственного интеллекта // Туберкулез и болезни легких. – 2018. – Т. 96, № 8. – С. 42–49.

251. Klassen V. I., Safin A. A., Maltsev A. V., et al. AI-based screening of pulmonary tuberculosis: diagnostic accuracy // J. eHealth Technol. Appl. – 2018. – Vol. 16, № 1. – P. 28–32.

252. Sirazitdinov I., Kholiavchenko M., Mustafaev T., et al. Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database // Comput. Electr. Eng. – 2019. – Vol. 78. – P. 388–399.

253. Ebrahimian S., Homayounieh F., Rockenbach M. A. B. C., et al. Artificial intelligence matches subjective severity assessment of pneumonia for prediction of patient outcome and need for mechanical ventilation: a cohort study // Sci. Rep. Nature Research. – 2021. – Vol. 11, № 1. – P. 858.

254. Mushtaq J., Pennella R., Lavalley S., et al. Initial chest radiographs and artificial intelligence (AI) predict clinical outcomes in COVID-19 patients: analysis of 697 Italian patients // *Eur. Radiol.* – 2021. – Vol. 31, № 3. – P. 1770–1779.

255. Nash M., Kadavigere R., Andrade J., et al. Deep learning, computer-aided radiography reading for tuberculosis: a diagnostic accuracy study from a tertiary hospital in India // *Sci. Rep.* – 2020. – Vol. 10, № 1. – P. 210.

256. Putha P., Tadepalli M., Reddy B., et al. Can Artificial Intelligence Reliably Report Chest X-Rays? Radiologist Validation of an Algorithm trained on 2.3 Million X-Rays: Preprint. – 2018. 19 July [Электронный ресурс]. – URL: https://www.researchgate.net/publication/326505882_Can_Artificial_Intelligence_Reliably_Report_Chest_X-Rays_Radiologist_Validation_of_an_Algorithm_trained_on_12_Million_X-Rays (дата обращения: 05.01.2023).

257. Бородачев С. М. Статистические методы в управлении качеством. – Екатеринбург: Изд-во Уральского университета, 2016. – 87 с.

258. С. Ф. Четвериков, К. М. Арзамасов, А. Е. Андрейченко [и др.]/ Подходы к формированию выборки для контроля качества работы систем искусственного интеллекта в медико-биологических исследованиях // *Современные технологии в медицине.* – 2023. – Т. 15, № 2. – С. 19-27. – DOI 10.17691/stm2023.15.2.02.

259. Ю. А. Васильев, А. В. Владзимирский, К. М. Арзамасов [и др.] / Первые 10000 маммографических исследований, выполненных в рамках услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта» // *Менеджер здравоохранения.* – 2023. – № 8. – С. 54-67. – DOI 10.21045/1811-0185-2023-8-54-67.

260. К. М. Арзамасов, Ю. А. Васильев, А. В. Владзимирский [и др.] / Применение компьютерного зрения для профилактических исследований на примере маммографии // *Профилактическая медицина.* – 2023. – Т. 26, № 6. – С. 117–123. – DOI 10.17116/profmed202326061117.

261. Андрейченко А.Е., Арзамасов К.М., Кирпичев Ю.С., Четвериков С.Ф.

Комплексная система оценки качества решений на основе ИИ для медицинской диагностики // Альманах Искусственный Интеллект. 2022. № 11. С. 168–173.

262. D’Orsi C. J. ACR BI-RADS atlas: breast imaging reporting and data system. – American College of Radiology, 2013.

263. Д. Ю. Кокина, В. А. Гомболевский, К. М. Арзамасов [и др.] / Возможности и ограничения использования инструментов машинной обработки текстов в лучевой диагностике // Digital Diagnostics. – 2022. – Т. 3, № 4. – С. 374–383. – DOI 10.17816/DD101099.

264. Caliskan D., Zierk J., Kraska D., et al. First Steps to Evaluate an NLP Tool’s Medication Extraction Accuracy from Discharge Letters // Stud Health Technol Inform. – 2021. – May 24. – Vol. 278. – P. 224–230.

265. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation // BMC Genomics. – 2020. – Vol. 21, № 1. – P. 6.

266. Thomas G., Kenny L.C., Baker Ph. N., et al. A novel method for interrogating receiver operating characteristic curves for assessing prognostic tests // Diagnostic Progn. Res. – 2017. – Vol. 1, № 1. – P. 17.

267. Коробов В. Б. Сравнительный анализ методов определения весовых коэффициентов «влияющих факторов» // Социология: методология, методы, математические модели. – 2005. № 20. – С. 54–73.

268. Документы об эксперименте по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы // Центр диагностики и телемедицины: официальный сайт [Электронный ресурс]. – URL: <https://mosmed.ai/ai/docs> (дата обращения: 15.04.2023).

269. И. А. Тыров, Ю. А. Васильев, К. М. Арзамасов [и др.] / Оценка зрелости технологий искусственного интеллекта для здравоохранения: методология и ее применение на материалах Московского Эксперимента по компьютерному зрению в лучевой диагностике // Врач и информационные

технологии. – 2022. – № 4. – С. 76-92. – DOI 10.25881/18110193_2022_4_76.

270. Ю. А. Васильев, И. А. Тыров, А. В. Владзимирский [и др.] / Двойной просмотр результатов маммографии с применением технологий искусственного интеллекта: новая модель организации массовых профилактических исследований // Digital Diagnostics. – 2023. – Т. 4, № 2. – С. 93-104. – DOI 10.17816/DD321423.

271. Отчет за 2020 г. главного внештатного специалиста Минздрава России по лучевой и инструментальной диагностике Тюрина И. Е. 2020. [Электронный ресурс]. – URL: https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/056/620/original/%D0%9E%D1%82%D1%87%D0%B5%D1%82_%D0%B7%D0%B0_%2020_%D0%B3%D0%BE%D0%B4_%D0%A2%D1%8E%D1%80%D0%B8%D0%BD.pdf?1624967722 (дата обращения: 04.05.2023).

272. Jin D., Parmar Ch., Quackenbush J., et al. Artificial intelligence in radiology // Artificial Intelligence in Medicine. – 2021. – P. 265–289.

273. Морозов С. П., Ветшева Н. Н., Ледихова Н. В. Оценка качества рентгенорадиологических исследований / Серия «Лучшие практики лучевой диагностики». – Вып. 48. – М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2019. – 47 с.

274. Thrall J.H. et al. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success // J. Am. Coll. Radiol. – 2018. – Vol. 15, № 3. – P. 504–508.

275. Yoo H., Lee S. H., Arru Ch. D., et al. AI-based improvement in lung cancer detection on chest radiographs: results of a multi-reader study in NLST dataset // Eur. Radiol. – 2021. – Vol. 31, № 12. – P. 9664–9674.

276. Hwang E. J., Nam J. G., L. W. Hyeon et al. Deep Learning for Chest Radiograph Diagnosis in the Emergency Department // Radiology. – 2019. – Vol. 293, № 3. – P. 573–580.

277. Морозов С. П., Владзимирский А. В., Шулькин И. М. [и др.]. Целесообразность применения технологий искусственного интеллекта в лучевой диагностике // Врач и информационные технологии. – 2022. – Vol. 1. – С. 12–29.

278. Sasaki M., Tozaki M., Rodríguez-Ruiz A., et al. Artificial intelligence for

breast cancer detection in mammography: experience of use of the ScreenPoint Medical Transpara system in 310 Japanese women // *Breast Cancer*. – 2020. – Vol. 27, № 4. – P. 642–651.

279. Salim M. et al. External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms // *JAMA*. – 2020. – Vol. 6, № 10. – P. 1581.

280. Freeman K., Geppert J., Stinton Ch., et al. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy // *BMJ*. – 2021. – Vol. 374. – P. 1872.

281. Wu N., Phang J., Park J., et al. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening // *IEEE Trans. Med. Imaging*. Institute of Electrical and Electronics Engineers Inc., 2020. Vol. 39, № 4. P. 1184–1194.

282. Chen Y. et al. Performance of a Breast Cancer Detection AI Algorithm Using the Personal Performance in Mammographic Screening Scheme // *Radiology*. 2023. Vol. 308, № 3.

283. Евтеева К. Б., Павлов Н. А., Анрейченко А. Е. [и др.]. Свидетельство о государственной регистрации программы для ЭВМ № 2022617275 Российская Федерация. Платформа экспертного пересмотра: веб-инструмент разметки проекционных методов лучевой диагностики на уровне исследования : № 2022615948 : заявл. 05.04.2022 : 2022.

284. Chang P. J. Moving Artificial Intelligence from Feasible to Real: Time to Drill for Gas and Build Roads // *Radiology*. – 2020. – Vol. 294, № 2. – P. 432–433.

285. Parekh S. Selecting an AI marketplace for radiology: key considerations for healthcare providers // *Signify Research*. – 2019. – October 17 [Электронный ресурс]. – URL: <https://www.signifyresearch.net/medical-imaging/selecting-ai-marketplace-radiology-key-considerations-healthcare-providers> (дата обращения: 05.01.2023).

286. Морозов С. П., Абуладзе Л. Р., Андрейченко А. Е. [и др.]. Базовые рекомендации к работе сервисов искусственного интеллекта для лучевой диагностики: методические рекомендации / Серия «Лучшие практики лучевой

и инструментальной диагностики». – Вып. 119. – М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2022. – 68 с.

287. Свидетельство о государственной регистрации программы для ЭВМ : RU 2022617324. Веб-инструмент для выполнения ROC-анализа результатов диагностических тестов : № 2022616046 : заявл. 05.04.2022 : опубл. 19.04.2022 / Морозов С. П., Андрейченко А. В., Четвериков С. Ф., Арзамасов К. М., Кирпичев Ю. С., Семенов С. С., Курицын С. О., Логунова Т. А.

288. Ieki H., Ito K., Saji M. 3, et al. Deep learning-based age estimation from chest X-rays indicates cardiovascular prognosis // Commun. Med. – 2022. – Vol. 2, № 1. – P. 1–12.

289. Yi P. H., Wei J., Kim T. K., et al. Radiology “forensics”: determination of age and sex from chest radiographs using deep learning // Emerg. Radiol. – 2021. – Vol. 28, № 5. – P. 949–954.

290. Chibane S., Mahar S., Das B., et al. Accuracy of Patient Age Estimation from Frontal Chest Radiographs of Adult Patients // Open J. Radiol. – 2022. – Vol. 12, № 02. – P. 37–48.

291. Kim D. W., Jang H. Y., Kim K. W., et al. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers // Korean J. Radiol. – 2019. – Vol. 20, № 3. – P. 405.

292. Digital Health Hype Circle 2021 // Thought Leadership for HealthTech. – 2021. – March 14 [Электронный ресурс]. – URL: <https://www.healthcare.digital/single-post/digital-health-hype-cycle-2021> (дата обращения: 15.04.2023).

293. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) // Department of Health and Human Services (United States) [Электронный ресурс]. – 2019. – URL: <https://cohenhealthcarelaw.com/2020/06/proposed-regulatory-framework-for-modifications-to-artificial-intelligence-machine-learning-ai-ml-based->

software-as-a-medical-device-samd (дата обращения: 15.04.2023).

294. Rubin M. N., Fusaro M. V., Scurlock C., et al. Telemedicine in the ICU: clinical outcomes, economic aspects, and trainee education // J. Telemed. Telecare. – 2016. – Vol. 19, № 2. – P. 530–534.

295. Wan Y., Tong Y., Liu Y., et al. Evaluation of the Combination of Artificial Intelligence and Radiologist Assessments to Interpret Malignant Architectural Distortion on Mammography // Front. Oncol. – 2022. – Vol. 12.

296. Lee K. A., Talati N., Oudsema, R. et al. BI-RADS 3: Current and Future Use of Probably Benign // Curr. Radiol. Rep. – 2018. – Vol. 6, № 2. – P. 5.

297. Browne J. L., Pascual M. Á., Perez J., et al. AI: Can It Make a Difference to the Predictive Value of Ultrasound Breast Biopsy? // Diagnostics. – 2023. – Vol. 13, № 4. – P. 811.

298. Dratsch T., Chen X., Mehrizi M. R., et al. Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance // Radiology. – 2023. – Vol. 307, №4. – P. e222176.

299. 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.

300. Najjar, R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. Diagnostics 2023, 13, 2760.

ПРИЛОЖЕНИЯ

Приложение А

№	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
1	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 Б. Положительные результаты верификации хотя одной из приоритетных нозологий		Обязательно – вероятность наличия во всем исследовании хотя бы одного рентгенологического признака из перечня А. Обязательно – локализация найденных признаков (если применимо)	Дробное или целое число Тепловая карта/Контур/Маска/др.	Apache Kafka Message + DICOM DICOM
			Нет признаков патологии: ни одного рентгенологического признака из перечня А.		Опционально – наименование одного или нескольких рентгенологических признаков	Текст	DICOM

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
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

Рисунок 1 – Базовые диагностические требования к результатам ПО на основе ТИИ по флюорографии и рентгенографии органов грудной клетки, актуальные на 2020 год

Таблица А1 – Базовые диагностические требования к результатам работы ПО на основе ТИИ по направлению «флюорография» и «рентгенография» органов грудной клетки

Вид исследования - рентгенография (флюорография)		
Наименование	РГ ОГК (ФЛГ) - выявление спектра патологических изменений	
Клиническая задача, решаемая ПО на основе ТИИ	Определение наличия и локализации рентгенографических признаков (минимум 7), коррелирующих с приоритетной нозологией (из перечня ниже):	
	1. Туберкулез (МКБ-10: A15–A16, A19).	
	2. Пневмония, гнойные и некротические состояния (МКБ-10: J10–J18, J80–J86)	
	3. Гидроторакс (МКБ-10: J94, R09.1).	
	4. Пневмоторакс (МКБ-10: S27.0).	
	5. Ателектаз (МКБ-10: J98.1).	
	6. Объемные образования (МКБ-10: D38.1– D38.4, C34–C39).	
	7. Перелом ребра (ребер), грудины и грудного отдела позвоночника (МКБ-10: S22)	
	8. Кардиомегалия (МКБ-10: I51.7).	
	9. Патология средостения (МКБ-10: D15.2, D38.3, I71)	
Диагностические требования к выявлению патологии	Наличие хотя бы одного рентгенологического признака из перечня	Характеристики признаков
	Плевральный выпот	Однородное затемнение в нижних отделах легких с практически горизонтальным уровнем, не определяется анатомическая структура наружно-нижних отделов легких – реберно-диафрагмальный синус

Продолжение таблицы А1

	Пневмоторакс	Однородное просветление в периферических отделах верхних долей легких, соответствующее распределению воздуха в замкнутом пространстве, в зоне интереса не определяется (смещен) легочный рисунок
	Очаг затемнения	Локальная зона затемнения легочной ткани с нарушением дифференцировки легочного рисунка, не соответствующая анатомическому перибронхиальному распределению, возможен значительный разброс локализации/размеров/контуров/ формы
	Инфильтрация/консолидация	Локальная зона затемнения легочной ткани с полным/неполным нарушением дифференцировки легочного рисунка, соответствующая анатомическому перибронхиальному/сегментарному/ долевого распределению
	Диссеминация	Множественные однотипные субсантиметровые локальные зоны затемнения легочной ткани, соответствующие анатомическому перибронхиальному распределению
	Полость	Локальная зона затемнения легочной ткани с нарушением дифференцировки легочного рисунка, не соответствующая анатомическому перибронхиальному распределению, возможен разброс локализации/размеров/контуров/формы, возможно наличие центрального просветления, горизонтального уровня раздела содержимого «газ–жидкость»

Продолжение таблицы А1

Вид исследования – рентгенография (флюорография)		
	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
	Кальцинат/кальцинированная тень в легких	Локальное однородное высокоинтенсивное (высокоплотное) затемнение с четкими контурами
	Расширение средостения	Расширение тени средостения на всем протяжении в обе стороны, локальное расширение тени средостения с одной стороны, повышение прозрачности тени средостения, затемнение на фоне тени средостения, смещение тени средостения
	Кардиомегалия	Увеличение показателя кардиоторакального индекса (отношение поперечного размера сердца, за исключением жировой подушки верхушки сердца, к внутреннему размеру грудной клетки) более 0,5
	Нарушение целостности кортикального слоя	Локальное прерывание линии наружного контура кости, возможно с наличием смещения/расхождения костных отломков
	Консолидированный перелом	Деформация контуров костных структур, костная мозоль
	<i>Примечание</i>	Признаки, наличие которых требует быстрого принятия врачебного решения

Продолжение таблицы А1

Вид исследования - рентгенография (флюорография)		
Требования к отображению находок на дополнительной серии изображений	Способ маркировки отдельных находок – оконтуривание или оконтуривание с заливкой	
	Способ маркировки областей/зон – цветовая карта или оконтуривание	
	Числовая индикация находок	
Содержание ответа ПО на основе ТИИ		Формат ответа ПО на основе ТИИ
Раздел «Описание»	Вероятность наличия патологии в исследовании в целом	Число
	Вероятность наличия каждого рентгенологического признака	Число
	Локализация и однозначная цифровая идентификация найденных признаков	Тепловая карта/ контур/ маска/ др.
Раздел «Заключение»	Вероятность наличия патологии в исследовании в целом	Число
	Вероятность наличия каждого рентгенологического признака	Число
	Локализация и однозначная цифровая идентификация найденных признаков	Тепловая карта/ контур/ маска/ др.

Таблица А2 – Базовые диагностические требования к результатам работы ПО на основе ТИИ по направлению «маммография»

Вид исследования – маммография (ММГ)		
Наименование	Скрининговая ММГ – выявление признаков рака молочной железы	
Клиническая задача, решаемая ПО на основе ТИИ	Определение признаков наличия и локализации находок, характерных для рака молочной железы	
Диагностические требования к выявлению патологии	BI-RADS 0 – для скрининга РМЖ (соответствует BI-RADS 3–5 ACR 2013 г.)	
Требования к отображению находок на дополнительной серии изображений	Способ маркировки отдельных находок – оконтуривание или оконтуривание с заливкой	
	Способ маркировки областей/зон – цветовая карта или оконтуривание	
	Числовая индикация находок	
Содержание ответа ПО на основе ТИИ		Формат ответа ПО на основе ТИИ
Раздел «Описание»	Определение и локализация находок, свидетельствующих о наличии во всем исследовании признаков рака молочной железы	Тепловая карта/ контур/ маска/ др.
	Оценка по шкале BI-RADS 0–2, (BI-RADS, 0 – находки, характерные для BI-RADS 3–5 (по классификации ACR 2013 г.)	Текст

Продолжение таблицы А2

Вид исследования - маммография (ММГ)		
	Указание вероятности злокачественности находок («Вероятность злокачественности __»)	Число
Раздел «Заключение»	Определение и локализация находок, свидетельствующих о наличии во всем исследовании признаков рака молочной железы	Тепловая карта/ контур/ маска/ др.
	Оценка по шкале BI-RADS 0–2, (BI-RADS, 0 – находки, характерные для BI-RADS 3–5 (по классификации ACR 2013 г.)	Текст
	Указание вероятности злокачественности находок («Вероятность злокачественности __»)	Число
Наименование	Диагностическая ММГ - выявление признаков рака молочной железы	
Клиническая задача, решаемая на основе ТИИ	Определение признаков наличия и локализации находок, характерных для рака молочной железы	
Диагностические требования к выявлению патологии	BI-RADS 3–5 ACR 2013	
Требования к отображению находок на дополнительной серии изображений	Способ маркировки отдельных находок – оконтуривание или оконтуривание с заливкой	
	Способ маркировки областей/зон – цветовая карта или оконтуривание	
	Числовая индикация находок	

Продолжение таблицы А2

Вид исследования - маммография (ММГ)		
	Изображения в серии, содержащие патологические находки, должны быть обозначены на ScrollBar (отдельная полоса визуализации уровня просмотра серии в дополнительной серии)	
Содержание ответа ПО на основе ТИИ		Формат ответа ПО на основе ТИИ
Раздел «Описание»	Определение и локализация находок, свидетельствующих о наличии во всем исследовании признаков рака молочной железы	Тепловая карта/ контур/маска/др.
	Оценка по шкале BI-RADS 0–2, (BI-RADS, 0 – находки, характерные для BI-RADS 3–5 (по классификации ACR 2013 г.)	Текст
	Указание вероятности злокачественности находок («Вероятность злокачественности __»)	Число
Раздел «Заключение»	Определение и локализация находок, свидетельствующих о наличии во всем исследовании признаков рака молочной железы	Тепловая карта/ контур/маска/др.
	Оценка по шкале BI-RADS 0–2, (BI-RADS, 0 – находки, характерные для BI-RADS 3–5 (по классификации ACR 2013 г.)	Текст
	Указание вероятности злокачественности находок («Вероятность злокачественности __»)	Число

Приложение Б

	Возрастные изменения	КАРДИОМЕГАЛИЯ	ОЧАГ	ИНФИЛЬТРАЦИЯ	ОБЪЕМНЫЕ ОБРАЗОВАНИЯ	ФИБРОЗ	ПНЕВМОНИЯ	КАЛЬЦИНАТ	СМОЛКА	ВЫЛОТ	ТУБЕРКУЛЕЗ	КОМБИНИРОВАННЫЙ ПЕРИОМ	ПЕРИОМ	РАСПИРЕНИЕ КОРНЕЙ	РАСПИРЕНИЕ СРЕДОСТЕНИЯ	ПНЕВМОТОРАКС	АТЕЛЕКТАЗ	ПРЯЖА	РАССЕИВАНИЕ	ПОМОСТЬ	ПНЕВМОПЕРИТЕЗИМ
Количество протоколов всего, шт	75653	50103	31753	28572	20292	15314	8325	6866	4453	3748	2649	2334	1872	1369	1008	911	753	555	168	136	24
Доля протоколов всего, %	72.1	47.7	30.3	27.2	19.3	14.6	7.9	6.5	4.2	3.6	2.5	2.2	1.8	1.3	1	0.9	0.7	0.5	0.2	0.1	0
Количество протоколов с единственными признаком, шт	34322*	12431	4928	2924	295	4050	1475	1846	2679	576	148	1081	1055	331	18	206	78	126	1	6	7
Доля протоколов с единственными признаком, %	32.7	11.8	4.7	2.8	0.3	3.9	1.4	1.8	2.6	0.5	0.1	1	1	0.3	0	0.2	0.1	0.1	0	0	0
Количество протоколов с комбинацией признаков, шт	41331	37672	26825	25548	19997	11264	6850	5020	1774	3172	2501	1253	817	1038	990	705	675	429	167	130	17
Доля протоколов с комбинацией признаков, %	39.4	35.9	25.6	23.3	16.5	12	6.1	5.1	3.2	2.6	2	1.9	1.5	1.1	0.9	0.8	0.6	0.5	0.2	0.1	0
1. Возрастные изменения, КАРДИОМЕГАЛИЯ, %	23.6	23.6	10.6	10.6	10.6	2.3															0
2. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %																					
3. Возрастные изменения, ФИБРОЗ, %	2.3	1.6				2.3															
4. Возрастные изменения, КАРДИОМЕГАЛИЯ, ФИБРОЗ, %	1.6	1.6				1.6															
5. КАРДИОМЕГАЛИЯ, ФИБРОЗ, %	1.4	1.4				1.4															
6. Инфильтрация, ПНЕВМОНИЯ, %	1.2			1.4			1.4														
7. Возрастные изменения, кальцинат, %								1.2													
8. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, ПНЕВМОНИЯ, %			1.2	1.2	1.2		1.2														
9. Инфильтрация, ОЧАГ, %			1	1.1	1																
10. КАРДИОМЕГАЛИЯ, ОЧАГ, %			1	1	1																
11. Возрастные изменения, ИНФИЛЬТРАЦИЯ, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %	1	1	1	1	1																
12. Возрастные изменения, КАРДИОМЕГАЛИЯ, ОЧАГ, %	0.8	0.8	0.8			0.8															
13. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, ФИБРОЗ, %	0.5	0.5		0.8	0.8	0.8															
14. Возрастные изменения, кальцинат, КАРДИОМЕГАЛИЯ, %						0.5		0.5													
15. Инфильтрация, кальцинат, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %			0.5	0.5	0.5			0.5													
Прочие комбинации (удел < 0.5), %	8.4	7	8.6	6.7	2.4	5.9	3.5	2.9	3.2	2.6	2	1.9	1.5	1.1	0.9	0.8	0.6	0.5	0.2	0.1	0

Рисунок 1 – Распределение РГ-признаков в общем потоке исследований

ВОЗРАСТНЫЕ ИЗМЕНЕНИЯ	КАРДИОМЕТАБОЛИЗМ	ОЧАГ	ИНФИЛЬТРАЦИЯ	ОБЪЕМНЫЕ ОБРАЗОВАНИЯ	ФИБРОЗ	КАЛЬЦИНАТ	ПНЕВМОНИЯ	СКОМОЛ	КОМБИНИРОВАННЫЙ ПЕРИОД	ПЕРИОД	ТВЕРЖУЩЕ	ВЫНОТ	РАСШИРЕНИЕ КОРНЕЙ	РАСШИРЕНИЕ СРЕДОСТЕНИЯ	ПРИЖА	АТЕСКЛЕЗ	ПНВМОТОРИК	ДИССЕМИНАЦИЯ	ПОПОСЬ	ПНВМОТОПОСЬ
Количество протоколов всего, шт	39743	24003	21432	15302	10506	5394	5004	3532	1665	1393	1375	1283	1137	808	420	391	196	54	37	0
Доля протоколов всего, %	79,1	49,2	29,7	26,5	18,9	6,7	6,2	4,4	2,1	1,7	1,7	1,6	1,4	1	0,5	0,5	0,2	0,1	0	0
Количество протоколов с единственным признаком, шт	30199*	9632	3934	2784	241	2667	1109	2116	785	940	81	277	303	12	93	63	85	0	1	0
Доля протоколов с единственным признаком, %	37,4	11,9	4,9	3,4	0,3	3,3	1,4	2,6	1	1,2	0,1	0,3	0,4	0	0,1	0,1	0,1	0	0	0
Количество протоколов с комбинацией признаков, шт	33747	30111	20069	18648	15061	7839	3895	1416	880	453	1294	1056	834	796	327	328	111	54	36	0
Доля протоколов с комбинацией признаков, %	41,7	37,3	24,8	23,1	15,6	10,4	4,8	3,2	1,1	1,3	1,4	1,3	1,1	0,9	0,4	0,4	0,2	0,1	0	0
Из них:																				
1. Возрастные изменения, КАРДИОМЕТАБОЛИЗМ, %	25,8	25,8	11	11	11															
2. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %	2,2		11		2,2															
3. Возрастные изменения, ФИБРОЗ, %	1,8	1,8			1,8															
4. Возрастные изменения, КАРДИОМЕТАБОЛИЗМ, ФИБРОЗ, %	1,4				1,4															
5. КАРДИОМЕТАБОЛИЗМ, ФИБРОЗ, %							1,1													
6. Инфильтрация, ПНЕВМОНИЯ, %	1,1						1,1													
7. Возрастные изменения, ИНФИЛЬТРАЦИЯ, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %			1,1	1,1	1,1															
8. Инфильтрация, ОЧАГ, %			1	1	1															
9. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, ПНЕВМОНИЯ, %			1	1	1		1													
10. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, ФИБРОЗ, %			0,7	0,7	0,7					0,7										
11. Возрастные изменения, кальцинат, %	1,5					1,5														
12. КАРДИОМЕТАБОЛИЗМ, ОЧАГ, %		1,2	1,2																	
13. Возрастные изменения, КАРДИОМЕТАБОЛИЗМ, ОЧАГ, %	0,9	0,9	0,9																	
14. Возрастные изменения, кальцинат, КАРДИОМЕТАБОЛИЗМ, %	0,6	0,6				0,6														
15. Инфильтрация, ОЧАГ, ПНЕВМОНИЯ, %			0,5	0,5			0,5													
16. Возрастные изменения, СКОМОЛ, %	0,5							0,5												
17. Инфильтрация, кальцинат, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %			0,5	0,5	0,5	0,5														
Прочие комбинации (уд.вес < 0,5), %	7,3	5,6	6,9	6,2	1,3	4,3	2,2	2,2	2,7	1,1	1,3	1,4	1,3	0,9	0,4	0,4	0,2	0,1	0	0

Рисунок 2 – Распределение РГ-признаков в амбулаторном потоке исследований

	ВОЗРАСТНЫЕ ИЗМЕНЕНИЯ	КАРДИОМЕГАЛИЯ	ОЧАГ	ИНФИЛЬТРАЦИЯ	ОБЪЕМНЫЕ ОБРАЗОВАНИЯ	ФИБРОЗ	ПНЕВМОНИЯ	ВЫПОТ	КАЛЬЦИНАТ	ТВЕРДУШЕЗ	СКОПОС	ПНЕВМОТОРАК	КОМПЛЕКТОВАННЫЙ ПЕРИОМ	ПЕРИОМ	АТЕЛЕКТА	РАСШИРЕНИЕ КОРНЕЙ	РАСШИРЕНИЕ СРЕДОСТИЯ	ПРЯМАЯ	ДИССЕМИНАЦИЯ	ПОЛОСТЬ	ПНЕВМОТЕРАПЕВИЗМ
Количество протоколов всего, шт	11707	10360	7750	7140	4990	4808	3221	2665	1472	1274	921	715	669	479	362	232	200	135	114	99	24
Доля протоколов всего, %	48,5	42,9	32,1	29,6	20,7	19,9	13,8	10,2	6,1	5,3	3,8	3	2,8	2	1,5	1	0,8	0,6	0,5	0,4	0,1
Количество протоколов с единственным признаком, шт	4123*	2799	994	140	54	1383	366	349	342	67	563	121	296	115	15	28	6	33	1	5	7
Доля протоколов с единственным признаком, %	17,1	11,6	4,1	0,6	0,2	5,7	1,5	1,4	1,4	0,3	2,3	0,5	1,2	0,5	0,1	0,1	0	0,1	0	0	0
Количество протоколов с комбинацией признаков, шт	7584	7561	6756	7000	4936	3425	2955	2116	1130	1207	358	594	373	364	347	204	194	102	113	94	17
Доля протоколов с комбинацией признаков, %	31,4	31,3	26,4	25,5	18,4	18,4	12,4	8,8	4,9	4,7	3,3	2,5	2,5	1,8	1,4	0,9	0,7	0,6	0,5	0,4	0,1
1. Возрастные изменения, КАРДИОМЕГАЛИЯ, %	16,1	16,1	8,9	8,9	8,9																
2. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %						2,5															
3. Возрастные изменения, фиброз, %	2,5	1				1															
4. Возрастные изменения, КАРДИОМЕГАЛИЯ, фиброз, %	1	1,5				1,5	2,2														
5. КАРДИОМЕГАЛИЯ, фиброз, %																					
6. Инфильтрация, ПНЕВМОНИЯ, %				2,2																	
7. Возрастные изменения, инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %	0,6		0,6	0,6	0,6																
8. Инфильтрация, ОЧАГ, %			1,4	1,4																	
9. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, ПНЕВМОНИЯ, %			1,8	1,8	1,8		1,8														
10. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, фиброз, %			1	1	1	1															
11. ОЧАГ, фиброз, %		0,9	1,1	0,9		1,1	0,9														
12. Инфильтрация, КАРДИОМЕГАЛИЯ, ПНЕВМОНИЯ, %		0,7				0,7															
13. КАРДИОМЕГАЛИЯ, ПНЕВМОНИЯ, %																					
14. ОЧАГ, ТВЕРДУШЕЗ, фиброз, %		0,7				0,7				0,7											
15. Возрастные изменения, инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %	0,6	0,6						0,6													
16. Возрастные изменения, выпот, %	0,5							0,5													
17. Выпот, КАРДИОМЕГАЛИЯ, %	0,5	0,5						0,5													
18. Возрастные изменения, КАРДИОМЕГАЛИЯ, РАСШИРЕНИЕ СРЕДОСТИЯ, %	0,5							0,5													
19. Выпот, инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, ПНЕВМОНИЯ, %		0,5	0,5	0,5	0,5		0,5	0,5	0,5												
20. Кальцинат, КАРДИОМЕГАЛИЯ, %									0,5												
21. Выпот, инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, %			0,5	0,5	0,5			0,5													
22. Инфильтрация, ОБЪЕМНЫЕ ОБРАЗОВАНИЯ, ОЧАГ, ТВЕРДУШЕЗ, %			0,5	0,5	0,5																
Прочие комбинации (Узвес < 0,5), %	9,6	9	9,4	7,2	4,6	10,6	6,3	6,2	4,4	3,5	3,3	2,5	2,5	1,8	1,4	0,9	0,2	0,6	0,5	0,4	0,1

Рисунок 3 – Распределение РГ-признаков в амбулаторном потоке исследований

[illegible]

Таблица В2 – Распределение по бинарным шкалам категорий BI-RADS, выставленных ПО на основе ТИИ, врачом и экспертом в выборках из 10 417 и 60 исследований

Шкала	Оценщик результата	Количество исследований в группе 0 (% от общего количества)	Количество исследований в группе 1 (% от общего количества)	Всего исследований
Бинарная шкала 1	Врач	9446 (97,54)	238 (2,46)	9684
	ПО на основе ТИИ	8030 (82,92)	1654 (17,08)	
Бинарная шкала 2	Врач	6604 (96,90)	211 (3,10)	6815
	ПО на основе ТИИ	5358 (78,62)	1457 (21,38)	
Бинарная шкала 1	Эксперт	30 (66,67)	15 (33,33)	45
	Врач	22 (48,89)	23 (51,11)	
	ПО на основе ТИИ	31 (68,89)	14 (31,11)	
Бинарная шкала 2	Эксперт	14 (48,28)	15 (51,72)	29
	Врач	11 (37,93)	18 (62,07)	
	ПО на основе ТИИ	18 (62,07)	11 (37,93)	

Таблица ВЗ – Оценка согласованности (совпадений и расхождений, а также показателей завышения, занижения и коэффициента Каппа Коэна) между врачом и ПО на основе ТИИ, ПО на основе ТИИ и экспертом, врачом и экспертом относительно шкалы BI-RADS и бинарных шкал

Шкала	Группы сравнения	Количество совпадений (% от всех ММГ)	Расхождения			Коэффи- циент Каппа Коэна	Общее количество ММГ
			Количество случаев (% от всех ММГ)	из них			
				Количество завышений* (% от всех расхождений)	Количество занижений* (% от всех расхождений)		
Бинарная шкала 1	ПО и врач	8144 (84,10)	1540 (15,9)	1478 (95,97)	62 (4,03)	0,15	9684
	ПО и эксперт	18 (40)	27 (60)	13 (48,15)	14 (51,85)	-0,372	45
	Врач и эксперт	35 (77,78)	10 (22,22)	9 (90)	1 (10)	0,558	
Бинарная шкала 2	ПО и врач	5499 (80,69)	1316 (19,31)	1281 (97,34)	35 (2,66)	0,166	6815
	ПО и эксперт	5 (17,24)	24 (82,76)	10 (41,67)	14 (58,33)	-0,641	29
	Врач и эксперт	24 (82,76)	5 (17,24)	4 (80)	1 (20)	0,652	
Шкала BI- RADS	ПО и врач	4250 (43,89)	5434 (56,11)	4154 (76,44)	1280 (23,56)	0,12	9684
	ПО и эксперт	5 (11,11)	40 (88,89)	20 (50)	20 (50)	-0,248	45
	Врач и эксперт	15 (33,33)	30 (66,67)	10 (33,33)	20 (66,67)	0,142	

* Повышение/понижение категории, выставленной врачом, ПО на основе ТИИ; повышение/понижение категории, выставленной экспертом, врачом или ПО на основе ТИИ

Таблица В4 – Характеристики точности ПО на основе ТИИ относительно врача-рентгенолога и эксперта; врача-рентгенолога относительно эксперта

Шкала	Объект оценки относительно истинного значения (врача или эксперта)	AUC (micro)	Чувствительность	Специфичность	Точность	Всего исследований
Бинарная шкала 1	ПО относительно врача	0,792	0,74	0,844	0,84	9684
	ПО относительно эксперта	0,316	0,067	0,567	0,4	45
	Врач относительно эксперта	0,817	0,933	0,7	0,777	
Бинарная шкала 2	ПО относительно врача	0,82	0,834	0,806	0,807	6815
	ПО относительно эксперта	0,176	0,067	0,286	0,172	29
	Врач относительно эксперта	0,824	0,933	0,714	0,828	
Шкала BI-RADS	ПО относительно врача	0,65	0,85	0,846	0,83	9684
	ПО относительно эксперта	0,41	0,4	0,802	0,702	45
	Врач относительно эксперта	0,54	0,927	0,694	0,878	

Таблица В5 – Четырехпольная таблица результатов оценки маммографии (ММГ)

Шкала	Группы сравнения	Количество ЛП оценок (% от всех ММГ)	Количество ЛО оценок (% от всех ММГ)	Количество ИП оценок (% от всех ММГ)	Количество ИО оценок (% от всех ММГ)	Всего ММГ
Бинарная шкала 1	ПО относительно врача	1478 (15,26)	62 (0,64)	176 (1,82)	7968 (82,28)	9684
	ПО относительно эксперта	13 (28,89)	14 (31,11)	1 (2,22)	17 (37,78)	45
	Врач относительно эксперта	9 (20)	1 (2,22)	14 (31,11)	21 (46,67)	45
Бинарная шкала 2	ПО относительно врача	1281 (18,8)	35 (0,51)	176 (2,58)	5323 (78,11)	6815
	ПО относительно эксперта	10 (34,48)	14 (48,28)	1 (3,45)	4 (13,79)	29
	Врач относительно эксперта	4 (13,79)	1 (3,45)	14 (48,28)	10 (34,48)	29