

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

На правах рукописи

ВАСИЛЬЕВ ЮРИЙ АЛЕКСАНДРОВИЧ

**ИННОВАЦИОННЫЕ ДИАГНОСТИЧЕСКИЕ И ОРГАНИЗАЦИОННЫЕ
ТЕХНОЛОГИИ В РЕНТГЕНОЛОГИИ**

3.1.25. Лучевая диагностика

3.3.9. Медицинская информатика

Диссертация на соискание ученой степени
доктора медицинских наук

Научные консультанты:
доктор медицинских наук,
профессор Сеницын В.Е.
доктор медицинских наук,
Владимирский А.В.

Москва – 2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
Глава 1. СИСТЕМАТИЗАЦИЯ РЕЗУЛЬТАТОВ НАУЧНЫХ ИССЛЕДОВАНИЙ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВОЙ ДИАГНОСТИКЕ	23
1.1. Актуальные проблемы лучевой диагностики как предпосылки для автоматизации.....	23
1.2. Зарубежные исследования применимости и качества искусственного интеллекта в лучевой диагностике: систематический обзор метаанализов.....	30
1.3. Развитие технологий искусственного интеллекта в лучевой диагностике в Российской Федерации	40
Глава 2. МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ	55
Глава 3. ОБОСНОВАНИЕ ВОЗМОЖНОСТИ И СПОСОБОВ АВТОМАТИЗИРОВАННОГО АНАЛИЗА РЕЗУЛЬТАТОВ ЛУЧЕВЫХ ИССЛЕДОВАНИЙ.....	70
3.1. Анализ диагностической точности технологий искусственного интеллекта в лучевой диагностике	70
3.2. Обоснование возможности автоматизированного анализа результатов рентгенографии	81
3.3. Обоснование возможности автоматизированного анализа результатов маммографии	112
3.4. Обоснование возможности автоматизированного анализа результатов компьютерной томографии	122
3.5. Обоснование возможности автоматизированного анализа результатов магнитно-резонансной томографии	168
3.6. Обобщение способов автоматизированного анализа результатов лучевых исследований	178
Глава 4. КЛИНИЧЕСКАЯ ЦЕЛЕСООБРАЗНОСТЬ, БЕЗОПАСНОСТЬ И	

КАЧЕСТВО АВТОМАТИЗИРОВАННОГО АНАЛИЗА РЕЗУЛЬТАТОВ ЛУЧЕВЫХ ИССЛЕДОВАНИЙ	184
4.1. Оценка клинической значимости поддержки принятия врачебных решений с применением искусственного интеллекта	184
4.2. Оценка безопасности и качества автономного применения искусственного интеллекта в профилактических лучевых исследованиях	191
4.2.1. Обоснование автономной морфометрии результатов рентгенографии стопы	192
4.2.2. Проспективное клиническое исследование автономной сортировки результатов профилактических лучевых исследований	195
Глава 5. ОРГАНИЗАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОБЕСПЕЧЕНИЯ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РЕНТГЕНОЛОГИИ .	209
ЗАКЛЮЧЕНИЕ	225
ВЫВОДЫ	247
ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ.....	250
СПИСОК СОКРАЩЕНИЙ.....	252
СПИСОК ЛИТЕРАТУРЫ.....	253
ПРИЛОЖЕНИЯ.....	284

ВВЕДЕНИЕ

Актуальность темы исследования

Лучевая диагностика – ключевой и неотъемлемый компонент современного здравоохранения, обеспечивающий колоссальный вклад в скрининг, диагностику, стадирование, контроль динамики и результативности лечения подавляющего числа нозологий. Потребность в лучевых исследованиях возрастает постоянно. В период 2014–2019 гг. в России отмечался ежегодный прирост абсолютного количества таких исследований на 2,5–3,0 %, а в 2019–2020 гг. и в период пандемии COVID-19 – на 4,6–8,2 %. Меняется структура исследований: растет количество компьютерных и магнитно-резонансных томографий (КТ, МРТ), в том числе выполняемых в амбулаторно-поликлинических медицинских организациях. Снижается интенсивность роста числа рентгенологических исследований (Голубев с соавт., 2021, Берген с соавт., 2022, Завадовский с соавт., 2022, Митькова с соавт., 2022, Попова с соавт., 2022, Серова с соавт., 2020, Терновой с соавт., 2020, Тюрин с соавт., 2018, Шахабов с соавт., 2020).

На востребованность и количество лучевых исследований значимо влияет профилактическое направление медицины. Согласно действующим нормативно-правовым актам, именно лучевые методы (рентгенография, флюорография, маммография) служат основными способами скрининга онкологических заболеваний, туберкулеза и иных социально значимых заболеваний. С одной стороны, в интересах обеспечения общественного здоровья требуется наращивание объемов соответствующих профилактических исследований. С другой – такое наращивание однозначно связано с колоссальной нагрузкой на службы и подразделения лучевой диагностики. При этом врачи-рентгенологи сталкиваются с однотипными описаниями преимущественно нормальных результатов исследований. Это масштабная, стереотипная работа плохо сказывается как на отдельных профессиональных траекториях (так называемое «выгорание»), так и на экономике здравоохранения в целом (постоянно растущие

затраты наиболее ценного ресурса – высококвалифицированных специалистов) (Богородская с соавт., 2023, Газиева с соавт., 2022, Држевецкая с соавт., 2022, Ливзан с соавт., 2023, Цыбикова с соавт., 2023, Belue et al., 2022, Cao et al., 2023, Chamberlin et al., 2021, Chi et al., 2021, Hoffman et al., 2020, Li et al., 2022, Michalopoulou et al., 2023, Mohan et al., 2020, Vang et al., 2018).

Рост количества исследований в том числе стал возможен благодаря развитию парка оборудования, что отмечается несколькими исследователями параллельно. На этом фоне все более значительным становится вопрос эффективности использования такого парка. Выявлена проблематика недостаточных объемов применения современных методик (прежде всего исследований с контрастным усилением), отсутствия зависимости между оснащенностью и загрузкой. Последнее чаще всего связывают с низкой результативностью организационных процессов, отсутствием единого подхода к управлению, однако, здесь критичное влияние оказывает кадровый дефицит (Голубев с соавт., 2021, Лантух с соавт., 2023, Шелехов с соавт., 2017, 2019, 2021, 2023).

Авторы, изучающие состояние лучевой диагностики в России, отмечают сохраняющуюся проблему кадрового дефицита. В целом темпы роста потребности в лучевых исследованиях и парка оборудования принципиальным образом опережают рост количества и тем более компетенций врачей-рентгенологов (Ветшева с соавт., 2017, Голубев с соавт., 2021, Сура с соавт., 2022, Шелехов с соавт., 2019, 2021). Значимой проблемой становится необходимость повышения охвата населения профилактическими лучевыми исследованиями, а также обеспечение должного уровня их качества и доступности. Здесь вновь сдерживающим барьером оказывается кадровый дефицит.

Примечательно, что ключевые тренды – рост востребованности, количества исследований и парка оборудования на фоне неустраняемого кадрового дефицита – полностью характерны для всех стран с развитой экономикой (Cho et al., 2020, da Silva et al., 2021, Li et al., 2022, Schlemmer et al., 2018).

Можно утверждать, что сложилась ситуация, когда переход к новым

формам организации работы подразделений и служб лучевой диагностики становится обязательным. Без такого перехода дальнейшее развитие предметной области невозможно.

Принципиальным отличием современной лучевой диагностики является ее широчайшая цифровизация. Интенсивный технический прогресс обеспечил практически полный переход к использованию цифровой диагностической аппаратуры. Полностью очевидным дальнейшим шагом стало создание автоматизированных рабочих мест, объединение их от уровня цифровых архивов отдельных медицинских организаций до централизованных архивов медицинских изображений уровня крупных административно-территориальных образований (в Российской Федерации – субъектов) (Андреев с соавт., 2021, Глушкова с соавт., 2018, Гусев с соавт., 2017, 2022, Кармазановский с соавт., 2022, Кошкаров с соавт., 2020, Филь с соавт., 2022). Высокий уровень цифровизации создает основу для внедрения принципиально новых способов и методик как организации, так и оказания медицинской помощи.

В контексте изложенных выше проблем следующий шаг может и должен быть сделан за счет внедрения актуального, нового поколения технологий автоматизации – так называемых «технологий искусственного интеллекта» (ТИИ). Проблематика искусственного интеллекта (ИИ) в здравоохранении и, в частности, в лучевой диагностике является одной из центральных в современной научной литературе. Однако, несмотря на значительное количество исследований, целый ряд проблемных вопросов остается нерешенным.

Степень разработанности темы исследования

По проблеме применения ИИ в лучевой диагностике публикуются разнообразные стратегические прогнозы, обзоры, исследования по этике, роли и месту ИИ в процессах работы врачей-рентгенологов, проблематике формирования доверия к технологиям ИИ. Интересны попытки юридического анализа проблематики искусственного интеллекта в контексте лучевой диагностики в

России (Казакова с соавт., 2023). Однако конкретные предложения не формируются. Полагаем, что причиной отсутствия конкретики служит недостаточный объем научных знаний о реальных возможностях и ограничениях ИИ в лучевой диагностике.

В мире проводятся различного масштаба опросы врачей-рентгенологов о перспективах, текущем статусе применения ИИ в их работе, а также о предпочтительных клинических задачах, направлениях диагностики для автоматизации (Agrawal et al., 2023, Alsharif et al., 2022, Coppola et al., 2021). Приведенная статистика интересна; вместе с тем она может служить лишь некоторым обозначением направления для дальнейших исследований. В целом такие работы не раскрывают ключевые вопросы качества, надежности и значимости технологий искусственного интеллекта в лучевой диагностике.

В глобальной перспективе результаты исследований применимости и качества ИИ в лучевой диагностике показывают достаточно высокую диагностическую точность технологий искусственного интеллекта (Abesi et al., 2023, Adamou et al., 2022, Bedrikovetski et al., 2021, Campello et al., 2023, Din et al., 2023, Dumitrescu et al., 2022, Hickman et al., 2023, Islam et al., 2022, Komolafe et al., 2023, Kim et al., 2021, Kuo et al., 2022, Lex et al., 2023, Liu et al., 2023, Menon et al., 2023, Nabizadeh et al., 2023, Poly et al., 2021, Potipimpanon et al., 2022, Rescinito et al., 2023, Sadr et al., 2023, Soffer et al., 2023, Sugibayashi et al., 2023, Taib et al., 2023, Thong et al., 2023, Zhang et al., 2023, Zheng et al., 2022, Yoon et al., 2023). Однако эти результаты в подавляющем большинстве случаев получены в исследованиях с некорректным дизайном, способом их проведения и отчетностью, что фактически гарантированно приводит к систематическим ошибкам и переоценке результативности алгоритмов. Подчеркнем преобладание ретроспективных исследований на дискретных наборах данных, а также отсутствие полноценных исследований в реальных клинических условиях.

В России отмечается научно-исследовательская активность в области развития технологий искусственного интеллекта в лучевой диагностике (Алиев с соавт. 2022, Аль-Хайдари с соавт. 2022, Баранов с соавт. 2020, Бастанов с соавт.

2020, Блинов с соавт. 2019, Борзов с соавт. 2022, Глазнев с соавт. 2017, Данилов с соавт. 2022, Елигулашвили с соавт. 2022, Зуков с соавт. 2023, Костин с соавт. 2019, Кошелев с соавт. 2023, Лисиенкова с соавт. 2022, Мазо с соавт. 2021, Макиев с соавт. 2020, Мелдо с соавт. 2018-2020, Моисеенко с соавт. 2018, Падалко с соавт. 2019, Рожкова с соавт. 2020, Ройтберг с соавт. 2020, Скоробогач с соавт. 2023, Смольникова с соавт. 2022, Солодкий с соавт. 2023, Филиппова с соавт. 2023, Шарифуллин с соавт. 2020). Отмечается подавляющее количество научных публикаций технического и математического характера, имеющих ограниченное значение для медицинской науки и практики. Во многих работах допускаются типовые методологические ошибки: необоснованные объемы выборок, отсутствие внешней валидации, несоответствие принятым стандартам для выполнения и описания диагностических исследований. Исследования об ИИ в лучевой диагностике клинической направленности отличаются сугубо ретроспективным дизайном, что не позволяет получить достоверные данные о точности и надежности решений на основе искусственного интеллекта в реальных производственных условиях.

Цель исследования: повышение качества исследований и результативности производственных процессов в рентгенологии на основе применения технологий искусственного интеллекта.

Задачи исследования:

1. Изучить диагностическую точность технологий искусственного интеллекта при работе в условиях реальной клинической практики.
2. Обосновать способы автоматизированного анализа результатов лучевых исследований в контексте решения конкретных клинических задач.
3. Оценить клиническую значимость поддержки принятия решений врачом-рентгенологом с применением искусственного интеллекта; разработать организационную технологию для профилактики дефектов в работе врача-рентгенолога.
4. Обосновать возможность автономного применения технологий

искусственного интеллекта в лучевой диагностике; изучить безопасность и качество такого применения в условиях реальной клинической практики.

5. Разработать общий подход и организационные технологии для обеспечения безопасной и качественной автоматизации на основе технологий искусственного интеллекта производственных процессов в рентгенологии.

6. Оценить медицинскую, социальную и экономическую эффективность организационных технологий искусственного интеллекта, применяемых в рентгенологии.

Научная новизна

Систематизированы результаты метаанализов ($n = 38$) в предметной области, позволившие установить средние уровни точности ИИ в лучевой диагностике по литературным данным: медиана чувствительности – 88,7 %, специфичности – 89,5 %, площади под характеристической кривой (AUROC от англ. Area Under Receiver Operating Characteristic) – 93,8 %.

Впервые в условиях реальной клинической практики определены средние значения точности технологий ИИ при анализе результатов лучевых исследований (медиана AUROC – 0,855, точности – 0,840, чувствительности – 0,926, специфичности – 0,775). Установлена динамическая тенденция постепенного повышения чувствительности ИИ-сервисов на фоне некоторого снижения их специфичности. Существует значительная вариативность диапазонов и динамики показателей точности для каждой из модальностей и для отдельных ИИ-сервисов.

Установлено, что лишь 6,0 % ИИ-сервисов сохраняют стабильность показателей диагностической точности при переходе к работе с реальными клиническими данными; в 84,0 % случаев выявлено снижение точности. Тем самым показана необходимость обязательного проспективного клинического исследования технологий ИИ как этапа разработки, предшествующего клиническим испытаниям в целях регистрации медицинского изделия.

Впервые в проспективном дизайне проанализирована работа 32 ИИ-сервисов в реальных клинических условиях, позволившая обосновать основные способы автоматизированного анализа результатов лучевых исследований, к которым относятся: поддержка принятия врачебных решений, морфометрия (в том числе автономная), автономная сортировка.

Доказаны возможность и целесообразность использования систем поддержки принятия врачебных решений (СППВР) на основе ИИ при описании результатов компьютерной томографии, рентгенографии и магнитно-резонансной томографии с целью диагностики онкологических, демиелинизирующих, дегенеративно-дистрофических заболеваний, а также ряда острых состояний.

Доказаны возможность и целесообразность применения автоматизированной морфометрии при описании результатов компьютерной томографии и рентгенографии для выполнения измерений диаметров магистральных сосудов, объемов кровоизлияния или жидкости в полости, а также для выявления нарушений со стороны опорно-двигательной системы.

Доказаны возможность и целесообразность применения автономной сортировки для результатов профилактической рентгенографии (флюорографии) органов грудной клетки и профилактической маммографии.

Доказана применимость систем поддержки принятия решений на основе искусственного интеллекта для профилактики гиподиагностики в работе врача-рентгенолога.

Впервые установлена структура потенциально предотвратимых за счет применения технологий искусственного интеллекта дефектов, возникающих при интерпретации и описании результатов лучевых исследований врачом-рентгенологом.

Показана возможность автономной морфометрии результатов рентгенографии стопы с целью выявления продольного плоскостопия.

Впервые в реальных клинических условиях доказана возможность автономной сортировки результатов массовых профилактических исследований (маммографии, рентгенографии/флюорографии органов грудной клетки).

Согласованность решений медицинских изделий на основе ИИ и врачей-рентгенологов (проводящих первичное описание, экспертов) при сортировке высока – каппа Коэна $> 0,99$.

Доказано, что медицинские изделия на основе технологий ИИ превосходят врача-рентгенолога при сортировке результатов массовых профилактических исследований: удельный вес ложных отнесений результатов исследований к категории «норма» со стороны ИИ составляет 0,08 % как для маммографии, так и для рентгенографии/флюорографии; уровень клинически значимых ложных отнесений составляет 0,02 %.

Научно обоснован многоуровневый подход к организации внедрения и применения технологий искусственного интеллекта с целью автоматизации и повышения результативности производственных процессов в рентгенологии, для его реализации разработаны организационные технологии создания и внедрения технологий искусственного интеллекта в рентгенологии, реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта.

Проведена оценка медицинской, социальной и экономической эффективности организационных технологий искусственного интеллекта, применяемых в рентгенологии.

Теоретическая и практическая значимость работы

Разработана научная концепция организации внедрения и применения искусственного интеллекта в рентгенологии, обеспечивающая повышение качества исследований и результативности производственных процессов, а также позволяющая получить медицинский, социальный и экономический эффект.

Обоснованы конкретные способы применения технологий искусственного интеллекта при сортировке, интерпретации и анализе результатов лучевых исследований в разрезе модальностей и конкретных клинических задач.

Обосновано применение СППВР на основе ИИ для описания результатов компьютерной томографии, рентгенографии и магнитно-резонансной томографии

с целью диагностики онкологических, демиелинизирующих, дегенеративно-дистрофических заболеваний (деформирующий остеоартроз и т. д.), некоторых острых состояний.

Обосновано применение инструментов морфометрии на основе ИИ при описании результатов компьютерной томографии и рентгенографии для выполнения измерений диаметров магистральных сосудов, объемов кровоизлияния или жидкости в полости, для выявления нарушений со стороны опорно-двигательной системы.

Обоснованы безопасность и качество автономной сортировки результатов профилактических лучевых исследований (уровень гиподиагностики 0,08 %, клинически значимых расхождений – 0,02 %, согласованность диагностических решений $> 0,99$).

Выявлено отсутствие достоверных различий при измерениях угла свода стопы на результатах рентгенографии, выполненных ИИ-сервисом и врачами-рентгенологами; тем самым доказана возможность автономной морфометрии с целью выявления продольного плоскостопия.

Установлены средние значения и динамика диагностической точности ИИ в условиях применения в реальной клинической практике, которые в том числе могут применяться при нормировании требований к медицинским изделиям на основе искусственного интеллекта.

Разработан подход к профилактике гиподиагностики в работе врача-рентгенолога, заключающийся в применении систем поддержки принятия врачебных решений на основе ИИ и опирающийся в том числе на структуру потенциально предотвратимых дефектов. Разработана соответствующая организационная технология обеспечения безопасности лучевых исследований на основе обязательного предварительного машинного анализа их результатов.

Выявлен феномен снижения точности ТИИ при начале работы в условиях реальной клинической практики, который не должен расцениваться как неблагоприятное событие при использовании медицинского изделия, а должен послужить источником информации для тонкой настройки конкретного решения

на основе ИИ.

Разработаны организационные технологии создания и внедрения технологий искусственного интеллекта в рентгенологии, реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта которые могут применяться в деятельности учреждений здравоохранения, разработчиков ИИ, а также – при формировании государственной политики в области цифрового здравоохранения.

Реализация организационных технологий позволила получить медицинский эффект, состоящий в увеличении на 28,7 % случаев выполнения лучевых исследований, при проведении которых использованы технологии искусственного интеллекта, а также в снижении удельного веса выявляемых при внутреннем контроле качества медицинской помощи случаев клинически значимых расхождений с 6,0 % до 1,6–1,7 %. Социальная эффективность состоит в статистически значимом ($p = 0,0024$, $p = 0,0131$) росте вовлеченности врачей-рентгенологов в 5,5–6,6 раз в 2022 и 2023 гг. по сравнению с 2021 г.

Научный подход и разработанные организационные технологии позволили получить медико-социальный эффект без увеличения финансовых затрат, а также экономический эффект, заключающийся в создании возможностей и условий для допуска к использованию медицинских изделий на основе технологий искусственного интеллекта.

Методология и методы исследования

Организация, структура, методика и дизайн диссертационной работы определялись целью исследования и заключалась в решении актуальной научной проблемы принципиального повышения производительности и качества лучевой диагностики путем научного обоснования подхода к организации внедрения и применения технологий искусственного интеллекта с целью автоматизации и повышения результативности производственных процессов в рентгенологии.

Научное исследование проведено в 2022–2023 гг. в медицинских организациях Департамента здравоохранения города Москвы (МО ДЗМ), оказывающих первичную медико-санитарную и специализированную помощь взрослому населению в амбулаторных или стационарных условиях. Информационно-аналитическое изучение достигнутого уровня научных исследований в предметной области выполнено в дизайне выборочного обзора литературы для источников на русском языке и систематического обзора метаанализов для источников на английском языке. Исследование диагностической точности технологий ИИ выполнено в два этапа: в дизайне ретроспективного диагностического исследования на эталонных наборах данных и в дизайне проспективного многоцентрового клинического исследования. Оценка клинической целесообразности систем поддержки принятия врачебных решений выполнена в дизайне обсервационного исследования; обоснование безопасности и качества автономной сортировки результатов профилактических лучевых исследований – в дизайне проспективного одноцентрового клинического исследования.

Диссертационное исследование выполнено на основе системного подхода с применением методов научного познания (аналитических, диагностических, математико-статистических).

Основные положения, выносимые на защиту

1. В условиях реальной клинической практики средние значения точности технологий ИИ при анализе результатов лучевых исследований составляют: AUROC – $0,846 \pm 0,099$ (медиана – 0,855), точность – $0,81 \pm 0,122$ (медиана – 0,840), чувствительность – $0,907 \pm 0,080$, (медиана – 0,926), специфичность – $0,746 \pm 0,177$ (медиана – 0,775).

2. Апробация в формате проспективного клинического исследования является обязательным этапом разработки ИИ-сервиса; она направлена на получение объективной информации для доработки и усовершенствования

сервиса и предшествует клиническим испытаниям в целях регистрации медицинского изделия.

3. Основными способами автоматизированного анализа результатов лучевых исследований являются поддержка принятия врачебных решений, морфометрия (в том числе автономная), автономная сортировка.

4. СППВР на основе ИИ целесообразно применять при описании результатов компьютерной томографии, рентгенографии и магнитно-резонансной томографии с целью диагностики онкологических, демиелинизирующих, дегенеративно-дистрофических заболеваний, а также ряда острых состояний; инструменты автоматизированной морфометрии – при описании результатов компьютерной томографии и рентгенографии для выполнения измерений диаметров магистральных сосудов, объемов кровоизлияния или жидкости в полости, а также для выявления нарушений со стороны опорно-двигательной системы; автономная сортировка должна применяться при интерпретации результатов профилактической рентгенографии (флюорографии) органов грудной клетки и профилактической маммографии.

5. Клиническая значимость поддержки принятия решений врачом-рентгенологом с применением искусственного интеллекта состоит в предотвращении клинически значимых дефектов (прежде всего пропуска признаков новообразований (49,2 %) и неспецифических воспалительных изменений (24,6 %)) при интерпретации и описании результатов лучевых исследований.

6. В аспекте мероприятий по контролю безопасности и качества медицинской помощи факт применения технологий искусственного интеллекта врачом-рентгенологом целесообразно рассматривать как критерий качества оказания медицинской помощи.

7. Автономная сортировка может применяться для категорирования результатов профилактической маммографии и профилактической рентгенографии/флюорографии органов грудной клетки на «норму» и «не норму». В реальных клинических условиях согласованность решений

медицинских изделий на основе ИИ и врачей-рентгенологов при сортировке результатов профилактических исследований высока (каппа Коэна > 0,99).

8. Удельный вес дефектов в работе медицинских изделий на основе технологий ИИ, состоящих в ложном отнесении результатов профилактической маммографии и профилактической рентгенографии/флюорографии органов грудной клетки к категории «норма», составляет 0,08 %, а клинически значимых ложных отнесений – 0,02 %; при сортировке результатов массовых профилактических исследований ИИ по точности превосходит врача-рентгенолога.

9. Эффективная автоматизация производственных процессов медицинских организаций, связанных с выполнением, интерпретацией и описанием результатов рентгенологических исследований, возможна на основе научно обоснованного подхода, который реализуется посредством организационных технологий (создания и внедрения технологий искусственного интеллекта в рентгенологии; реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта) и позволяет существенно улучшить качество исследований.

Внедрение результатов исследования

Результаты диссертационного исследования:

1) Внедрены в деятельность медицинских организаций государственной системы здравоохранения г. Москвы ГБУЗ г. Москвы «Городская поликлиника №195 ДЗМ» (акт внедрения б/н от 25 сентября 2023 г.), ГБУЗ г. Москвы «Городская поликлиника №220 ДЗМ» (акт внедрения б/н от 01 октября 2023 г.). Соответствующие акты внедрения представлены в приложении.

2) Стали основой для нормативно-правовых актов – приказа органа исполнительной власти в сфере здравоохранения субъекта Российской Федерации (Департамента здравоохранения города Москвы) от 16.02.2023 № 134 «Об утверждении Порядка и условий проведения эксперимента по использованию

инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы».

3) Внедрены в Территориальную программу государственных гарантий бесплатного оказания гражданам медицинской помощи в городе Москве (см. приложение № 6 к Тарифному соглашению на 2023 год от 30.12.2022).

4) Внедрены в практическое здравоохранение в виде информационно-технологического взаимодействия Департамента здравоохранения города Москвы и департамента здравоохранения Ямало-Ненецкого автономного округа. В рамках взаимодействия медицинским организациям округа предоставлены доступ к результатам анализа медицинских изображений с использованием сервисов компьютерного зрения (соглашение №10-04-279/22 от 21.12.2022).

5) Внедрены в практическое здравоохранение в виде методических рекомендаций, утвержденных органом исполнительной власти в сфере здравоохранения субъекта РФ:

- методические рекомендации № 20 «Клинические испытания систем искусственного интеллекта (лучевая диагностика)», утверждены Экспертным советом по науке Департамента здравоохранения г. Москвы 10 апреля 2023 г., протокол №4;
- методические рекомендации № 32 «Оценка зрелости технологий искусственного интеллекта для здравоохранения», утверждены Экспертным советом по науке Департамента здравоохранения г. Москвы 04 августа 2023 г., протокол №9.

6) Внедрены в педагогический процесс ГБУЗ «Научно-практический клинический центр диагностики и телемедицины Департамента здравоохранения города Москвы» (акт внедрения б/н от 15 сентября 2023 г.) (см. приложение).

- 7) Послужили основой национальных стандартов Российской Федерации:
- ГОСТ Р 59921.0-2022. Системы искусственного интеллекта в клинической медицине. Основные положения.

- ГОСТ Р 59921.1-2022. Системы искусственного интеллекта в клинической медицине. Часть 1. Клиническая оценка.
- ГОСТ Р 59921.4-2021. Системы искусственного интеллекта в клинической медицине. Часть 4. Оценка и контроль эксплуатационных параметров.
- ГОСТ Р 59921.6-2021. Системы искусственного интеллекта в клинической медицине. Часть 6. Общие требования к эксплуатации.
- ГОСТ Р 59921.7-2022. Системы искусственного интеллекта в клинической медицине. Алгоритмы анализа медицинских изображений. Методы испытаний. Общие требования.
- ПНСТ 777-2022. Системы искусственного интеллекта в клинической медицине. Часть 10. Процессы жизненного цикла.

Степень достоверности и апробация результатов

Проведение научного исследования в рамках диссертационной работы было одобрено на заседании Независимого этического комитета Московского регионального отделения Российского общества рентгенологов и радиологов от 23 июня 2022г. №6/2022.

Достоверность выполненного исследования определяется достаточным объемом первичных данных (ИИ-сервисов – $n = 32$, результатов лучевых исследований в проспективном многоцентровом исследовании – $n = 6\ 563\ 753$, результатов профилактических лучевых исследований в проспективном одноцентровом исследовании – $n = 258\ 718$, клинических случаев в экспериментальном – $n = 100$ и наблюдательном – $n = 60$), применением адекватных методов исследования и статистической обработки полученных результатов.

Материалы диссертации доложены и обсуждены на российских и зарубежных научных мероприятиях: Московская межрегиональная конференция с международным участием «Проблемы ранней диагностики, лечения,

маршрутизации и реабилитации больных со ЗНО органов головы и шеи» (г. Москва, 03.02.2023), Всероссийская научно-практическая конференция по искусственному интеллекту в здравоохранении и системам поддержки принятия врачебных решений (ITM.AI) (г. Москва, 09.02.2023), VI Всероссийский научно-образовательный конгресс с международным участием «Онкорadiология, лучевая диагностика и терапия» (г. Москва, 20.02.2022, 11-12.02.2023), Всероссийская научно-практическая конференция «Современные аспекты развития здравоохранения 2023-2025» (г. Москва, 24.03.2023), Многопрофильная конференция по развитию сотрудничества между Москвой и Бангкоком (Королевство Таиланд, г. Бангкок, 03.04.2023), Форум по управлению интернетом (RIGF 2023) (г. Москва, 06.04.2023), Научно-практическая конференция Диагностика и лечение опухолей торакоабдоминальной локализации (г. Москва, 14.04.2023), XI Петербургский международный юридический форум (г. Санкт-Петербург, 11.05.2023), II Евразийский экономический форум с участием представителей стран-участников ЕАЭС (Киргизия, г. Бишкек, 16.05.2023; Казахстан, г. Астана 17.05.2023; Беларусь, г. Минск, 23.05.2023), IV Всероссийская научно-практическая конференция «Инновационные технологии в комплексной диагностике и лечении рака молочной железы» (г. Москва, 08.06.2023), XI международный конгресс «Оргздрав-2023» (г. Москва, 08.06.2023), IV Конференция онкологов Московской области (г. Москва, 09.06.2023), Международная Конференция «Smart-клиника: путь к цифровой трансформации» (г. Москва, 19.06.2023), «Нейрофорум-2023» (г. Москва, 23.06.2023), 10-й Евразийский радиологический Форум (Казахстан, г. Астана, 29.06.2023), Заседание Комитета по работе с некоммерческими профессиональными объединениями и разработке клинических рекомендаций АОР (г. Москва, 17.07.2023), Московский Урбанистический форум (г. Москва, 04.08.2023; 29.08.2023; 31.08.2023), VI международный форум онкологии и радиотерапии «Ради жизни — FOR LIFE» (г. Москва, 20-22.09.2022; 11-13.09.2023), III Международная научно-практическая конференция «3D-технологии в клинической анатомии и медицине» (г. Москва, 22.09.2023), XXXI

Ежегодная конференция ДиаМА с международным участием «Диагностические центры: от профилактики до высоких технологий» (г. Ставрополь, 29.09.2023), Российский диагностический саммит (г. Москва, 06.09.2022; 04.10.2023), Форум «Биотехмед» (г. Москва, 10.10.2023), XXIII и XXIV Конгресс «Информационные технологии в медицине» ИТМ 2023 (г. Москва, 13.10.2022; 12.10.2023), IV Национальный междисциплинарный конгресс «Времена года. Женское здоровье – от юного до серебряного и золотого возраста» (г. Москва, 19.10.2023), Научно-практическая конференция «Инструментальная диагностика на стыке науки и практики» (г. Москва, 20.10.2023), XVI Всероссийский национальный конгресс лучевых диагностов и терапевтов «Радиология – 2022» (г. Москва, 25.05.2022), Всероссийская конференция «Обеспечение потребностей национального здравоохранения. Лето-2022» (г. Москва, 09.06.2022), Летняя академия онкологов России (г. Обнинск, 17.06.2022), II ежегодная конференция «Medical Business 2022: бизнес для человека» (г. Москва, 22.06.2022), «Медицинский форум – Калининград – 2022» (г. Калининград, 26.08.2022), Всероссийская конференция «ИТМ Восток 2022» (г. Якутск, 02.09.2022), Международная конференция «Здравоохранение, образование и спорт в эру Возрождения новой эпохи Могущественного государства» (Туркменистан, г. Ашхабад, 09.10.2022), 2-й саммит по цифровой медицине и информационным технологиям в здравоохранении «ЦИФРОАЙТИМЕД» (г. Москва, 12.10.2022), VII Всероссийский Конгресс онкопациентов (г. Москва, 19.10.2022), Осенняя академия онкологов России (г. Сочи, 28.10.2022), Конгресс Российского общества рентгенологов и радиологов (г. Москва, 08.11.2022 и 09.11.2022), IV Международный муниципальный форум стран БРИКС+ (г. Санкт-Петербург, 24.11.2022), Конференция AI Journey 2022 (г. Москва, 24.11.2022), Конференция «Аддитивные технологии для медицины: мировой и российский опыт» Международного научно-практического форума «Российская неделя здравоохранения – 2022» (г. Москва, 05.12.2022), Конгресс Ассоциации организаторов здравоохранения в онкологии (г. Москва, 09.12.2022), Конференция Russian Internet Week (г. Москва, 12.12.2022), Конференция

«Радиология в Сочельник» (г. Москва, 23.12.2022).

Заслушивание основных результатов диссертационной работы состоялось на Расширенном заседании Ученого совета (совместно с заседанием Кафедры лучевой диагностики с курсом клинической радиологии ФГБУ "Национальный медико-хирургический центр им. Н.И. Пирогова" Минздрава России), протокол от 29 ноября 2023 г. №10/2023.

Личное участие автора в проведении исследования

Автор лично сформулировал ключевую научную проблему, имеющую высокую актуальность, выполнил планирование исследования, определил цель и задачи, выбрал дизайн и подобрал методы исследования. Непосредственно осуществлял сбор первичных данных, также лично принимал участие в экспертных мероприятиях (пересмотрах, технологическом и клиническом мониторинге), далее провел систематизацию и анализ данных с формулировкой итоговых и заключительных положений, выводов и рекомендаций по итогам научно-исследовательской работы, а также – с обобщением подхода и организационных технологий, что явилось решением ключевой научной проблемы.

Публикации

По материалам диссертационного исследования опубликовано 28 печатных работ в отечественных и зарубежных изданиях, из них 23 – в изданиях, рекомендованных ВАК при Минобрнауки России, 14 – в изданиях, входящих в международные базы данных Web of Science и Scopus. Получено 9 авторских свидетельств на базы данных. Опубликовано 1 монография, 2 методические рекомендации.

Объем и структура диссертации

Диссертация изложена на 288 страницах, состоит из вступления, аналитического обзора литературы (в том числе в дизайне метаанализа), главы о методологии, материалах и методах исследования, 5 глав собственных исследований, заключения, выводов, практических рекомендаций, списка сокращений и списка цитируемой литературы (254 источника, в том числе 139 отечественных и 115 зарубежных). Работа иллюстрирована 40 таблицами, 44 рисунками и дополнена приложением.

Глава 1. СИСТЕМАТИЗАЦИЯ РЕЗУЛЬТАТОВ НАУЧНЫХ ИССЛЕДОВАНИЙ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЛУЧЕВОЙ ДИАГНОСТИКЕ

1.1. Актуальные проблемы лучевой диагностики как предпосылки для автоматизации

Лучевая диагностика – ключевой и неотъемлемый компонент современного здравоохранения. Вклад рентгенологических и радиологических методов исследований в скрининг, диагностику, стадирование, контроль динамики и эффективности лечения подавляющего числа нозологий трудно переоценить [11, 50, 126]. Интенсивный технический прогресс обеспечил практически полный переход к цифровой диагностике, то есть к масштабному использованию цифровой диагностической аппаратуры. Полностью очевидным дальнейшим шагом стало создание автоматизированных рабочих мест, объединение их от уровня цифровых архивов отдельных медицинских организаций до централизованных архивов медицинских изображений уровня крупных административно-территориальных образований (в Российской Федерации – субъектов) [36, 128].

Потребность в лучевых исследованиях возрастает постоянно. В период 2014–2019 гг. в России отмечался ежегодный прирост абсолютного количества соответствующих исследований на 2,5–3,0 %, в 2019–2020 гг. и в период пандемии новой коронавирусной инфекции прирост составил от 4,6 до 8,2 %. При этом довольно динамично меняется их структура: рентгенологические и ультразвуковые исследования по-прежнему составляют основной массив, однако каждый год их абсолютное число сокращается в пользу более сложных и клинически значимых – компьютерной и магнитно-резонансной томографии [30, 98, 126].

Примечательно, что рост количества исследований во многом происходит за счет медицинских услуг, оказываемых населению в амбулаторных условиях.

В течение ряда лет отмечается рост числа компьютерных и магнитно-резонансных томографий (КТ, МРТ), выполняемых в амбулаторно-поликлинических медицинских организациях, в 4,3 и 2 раза соответственно; в 4 раза возросло количество МРТ, проводимых в условиях дневного стационара. На этом фоне количество амбулаторных рентгенологических исследований возросло не в несколько раз, а лишь примерно на 10,0 % [135]. Определенный негативный вклад в указанный рост вносят необоснованные направления на исследования. По нашим собственным данным, удельный вес необоснованных направлений в амбулаторных условиях может достигать 10,7 % для рентгенографии, 17,5 % для маммографии, 37,8 % для КТ. В подавляющем большинстве случаев такая ситуация обусловлена несоблюдением клинических рекомендаций. Вместе с тем нет доказательств положительного или отрицательного влияния выполненных по необоснованным направлениям исследований на состояние здоровья пациентов [70]. Не умаляя важности системной и постоянной организационно-методической работы с врачами клинических специальностей, утверждаем, что рост потребности в лучевых исследованиях и увеличение их числа является необратимым явлением. Более того, по мере развития гибридных методик и новых способов обработки данных (например, радиомики) указанные тенденции могут только усилиться в ближайшие годы.

Еще одним фактором, влияющим на количество проводимых лучевых исследований, является профилактическое направление медицины. Согласно действующим нормативно-правовым актам, именно лучевые методы (рентгенография, флюорография, маммография) служат основными способами скрининга онкологических заболеваний, туберкулеза и иных социально значимых заболеваний. С одной стороны, в интересах обеспечения общественного здоровья требуется наращивание объемов соответствующих профилактических исследований [13, 24, 45, 73, 131, 209, 211, 240]. С другой – такое наращивание однозначно связано с колоссальной нагрузкой на службы и подразделения лучевой диагностики. При этом врачи-рентгенологи сталкиваются с однотипными

описаниями преимущественно нормальных результатов исследований. Эта масштабная, стереотипная работа плохо сказывается как на отдельных профессиональных траекториях (так называемое «выгорание»), так и на экономике здравоохранения в целом (постоянно растущие затраты наиболее ценного ресурса – высококвалифицированных специалистов) [156].

Еще более усугубляет ситуацию следующий момент. Развитие методов КТ и МРТ позволяет осуществлять скрининг многих социально значимых заболеваний, а также выявлять рентгенологически определяемые факторы риска. Примерами могут служить низкодозовая КТ и бипараметрическая МРТ для скрининга злокачественных новообразований легкого и предстательной железы соответственно. На результатах стандартной КТ, в рамках концепции оппортунистического скрининга, можно выявлять факторы риска болезней системы кровообращения (например, коронарного кальциноза), признаки остеопороза, иных хронических неинфекционных заболеваний [32, 33, 96, 152, 157, 162, 181, 201, 233]. Однако масштабное внедрение таких новых профилактических подходов фактически не происходит, прежде всего в силу экономических и ресурсных ограничений. Если дополнительно загрузить существующий парк оборудования, как правило, возможно, то обеспечить требуемое количество врачей-рентгенологов, обладающих нужными компетенциями, – нет. Складывается ситуация, когда текущие ресурсные возможности лучевой диагностики создают критичные ограничения для дальнейшего развития медицинской профилактики, а значит – для достижения целей в области снижения смертности от онкологических и кардиологических заболеваний.

Рост количества исследований в том числе стал возможен благодаря развитию парка оборудования, что отмечается несколькими исследователями параллельно. На этом фоне все более значительным становится вопрос эффективности использования такого парка. Выявлена проблематика недостаточных объемов применения современных методик (прежде всего исследований с контрастным усилением), отсутствие зависимости между

оснащенностью и загрузкой. Последнее чаще всего связывают с низкой эффективностью организационных процессов, отсутствием единого подхода к управлению, однако здесь критичное влияние оказывает кадровый дефицит [30, 136, 138]. Авторы, изучающие состояние лучевой диагностики в России, отмечают сохраняющуюся проблему кадрового дефицита: снижение укомплектованности занятыми должностями, высокий коэффициент совместительства, несоответствие темпов роста кадрового обеспечения росту парка оборудования и потребности в исследованиях. Изучение состояния кадровой службы лучевой диагностики в России показало, что на фоне роста кадров на 19 % имеет место 12-процентный дефицит занимаемых должностей к штатному числу врачей-рентгенологов. Значительны неравномерность обеспеченности врачами-рентгенологами, а также интенсивность деятельности таких специалистов по субъектам Российской Федерации [136].

Темпы роста потребности в лучевых исследованиях и парка оборудования принципиальным образом опережают рост количества врачей-рентгенологов. Можно утверждать, что сложилась ситуация, когда переход к новым формам организации работы подразделений и служб лучевой диагностики становится обязательным. Без такого перехода дальнейшее развитие предметной области становится невозможным.

Но не только количество специалистов лучевой диагностики является проблемой. В литературе все большее внимание обращают на качество медицинской помощи – то есть на уровень компетенций медицинского персонала [119]. Выявлена тенденция снижения доли врачей-рентгенологов, радиотерапевтов, радиологов и специалистов ультразвуковой диагностики, имеющих квалификационную категорию [30]. Ранее обсуждались вопросы значимого влияния начального профиля подготовки на качество работы врачей-рентгенологов (качество работы специалистов, прошедших профессиональную переподготовку, было относительно ниже) [21]. Примечательно, что, по данным систематического обзора 2022 года, основными способами повышения качества в лучевой диагностике в мире являются информатизация, обучение, меры по

оптимизации протоколирования, развитие программ безопасности. Очевидно, что малое внимание уделяется персонализированной поддержке врача-рентгенолога, хотя приоритизация внедрения информационных технологий создает техническую основу для таковой (например, за счет очевидного широкого внедрения систем поддержки принятия решений, построенных на современных математических и технических подходах) [186].

Примечательно, что ключевые тренды – рост востребованности, количества исследований и парка оборудования на фоне неустраняемого кадрового дефицита – полностью характерны для всех стран с развитой экономикой [163, 166, 200, 226].

Решить проблематику количественного и качественного дефицита в лучевой диагностике может автоматизация определенных звеньев производственных процессов, отдельных процедур. Таковыми потенциально могут стать автоматические измерения, поддержка в принятии решений (в том числе как своеобразная «профилактика» неточностей и дефектов в работе врача-рентгенолога), автоматизированная сортировка и классификация пациентов (например, по стандартизированным шкалам оценки). Данные гипотетические утверждения, безусловно, нуждаются в детальном изучении.

На примерах смежных сфер, таких как лабораторная диагностика, известно, что автоматизация положительно сказывается на производительности и качестве труда медицинского персонала, повышает доступность медицинской помощи, открывает новые возможности для пациентов [31, 40, 130, 149, 215].

Текущий уровень цифровизации лучевой диагностики (наличие соответствующего парка, автоматизированных рабочих мест, сетей передачи данных, радиологических информационных систем, интегрированных в информационные системы в сфере здравоохранения различных уровней, наконец, централизованных архивов медицинских изображений) служит мощным фундаментом для внедрения актуального поколения технологий автоматизации – так называемых «технологий искусственного интеллекта».

В соответствии с Указом Президента Российской Федерации от 10.10.2019

№ 490 (<http://kremlin.ru/acts/bank/44731>) искусственный интеллект (ИИ) – это комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений.

Технологии искусственного интеллекта (ТИИ) – это технологии, основанные на использовании искусственного интеллекта, включая компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта. Сразу отметим, что для сферы лучевой диагностики наибольшее значение имеет компьютерное зрение для распознавания и интерпретации диагностических изображений. На втором месте – обработка естественного языка и распознавание речи для различных форм работы с медицинской документацией.

Проблематика искусственного интеллекта в здравоохранении, без сомнений, является одной из центральных в современной научной литературе.

Публикуются разнообразные стратегические прогнозы, обзоры, исследования по этике, роли и месту ИИ в процессах работы врачей-рентгенологов, проблематике формирования доверия к технологиям ИИ [34, 39, 176, 189, 202, 207, 210, 251]. Интересны попытки юридического анализа проблематики искусственного интеллекта в контексте лучевой диагностики в России. Авторы таких работ полагают основными ограничениями непрозрачность процесса принятия решений ТИИ (так называемый «эффект черного ящика») и высокую частоту ошибок со стороны соответствующих алгоритмов. По результатам социологических исследований демонстрируется недостаточный уровень доверия к ИИ как со стороны врачей, так и со стороны пациентов.

Декларируется «объективная необходимость создания эффективных правовых механизмов, предусматривающих меры ответственности за ошибочные решения ИИ, защищающие права врачей и пациентов при работе данных программам» [54, 55]. Однако конкретные предложения не формируются. Полагаем, что причиной отсутствия конкретики служит недостаточный объем научных знаний о реальных возможностях и ограничениях ИИ в лучевой диагностике.

По данным Европейского сообщества радиологов, по состоянию на 2019 год до 20,0 % соответствующих специалистов уже применяли в своей работе ТИИ, а 30,0 % планировали сделать это в ближайшее время. При этом потенциально наиболее подверженными внедрению ИИ модальностями были определены маммография, КТ и МРТ, а ключевыми клиническими задачами – скрининг, стадирование (в том числе повторное) в онкологии, количественные измерения биомаркеров [171]. По данным на 2022 год, ситуация несколько изменилась: реальный удельный вес врачей-рентгенологов – активных пользователей ИИ составил 40,0 %. Больше половины врачей использовали искусственный интеллект при интерпретации результатов лучевых исследований (детекция и маркировка специфических патологических проявлений). Примечательно, что почти в 40,0 % случаев такого использования целью была приоритизация описаний исследований для повышения доступности медицинской помощи в особых случаях. Постпроцессинг, в том числе реконструкция изображений и измерения, составил 28,6 % случаев применения ИИ [170]. Приведенная статистика интересна; вместе с тем она может служить лишь некоторым обозначением направления для дальнейших исследований, так как строится на добровольном социологическом опросе (как и результаты иных подобных исследований распространенности и характера применения ИИ в лучевой диагностике в Америке и Азии [144, 145, 165]). В процитированных работах не раскрыты ключевые вопросы качества, надежности и значимости технологий искусственного интеллекта в лучевой диагностике.

Далее нами проведен углубленный анализ текущей ситуации с применением технологий искусственного интеллекта в лучевой диагностике в России и за рубежом.

1.2. Зарубежные исследования применимости и качества искусственного интеллекта в лучевой диагностике: систематический обзор метаанализов

В глобальной перспективе отмечается интенсивная научно-исследовательская деятельность по различным аспектам создания, применения и оценки технологий искусственного интеллекта в рентгенологии и радиологии. С учетом значительного числа научных публикаций в предметной области нами принято решение провести систематизацию и совокупный анализ статей, подготовленных в дизайне метаанализа.

В систематический обзор, методика которого подробно приведена в главе 2, включены 38 метаанализов [140, 141, 142, 147, 150, 151, 155, 167, 168, 172, 177, 180, 185, 188, 193, 196, 198, 199, 203, 208, 213, 219, 220, 223, 225, 229, 232, 234, 237, 243, 246, 247, 248, 249, 250, 252, 253, 254].

Из общего числа 47,0 % (38) работ посвящены диагностике онкологических заболеваний, в том числе злокачественных новообразований (ЗНО) пищеварительной системы, – 5 (28,0 %), заболеваний дыхательной системы – 4 (22,0 %), репродуктивной системы – 4 (22,0 %), нервной системы – 2 (11,0 %), эндокринной системы – 2 (11,0 %), злокачественных опухолей брюшинного пространства – 1 (6,0 %). Диагностике поражений мозга (рассеянного склероза, аневризм, инсульта, внутричерепного кровоизлияния) были посвящены 11,0 % (4) метаанализов. Такое же количество работ отмечается в области стоматологии, из них три посвящены обнаружению и сегментации структур и одна – принятию решений об удалении зубов. Всего три метаанализа (8,0 %) посвящены диагностике поражений легких при новой коронавирусной инфекции. Такое же количество публикаций представлено по теме лучевой диагностики переломов, преимущественно – длинных костей конечностей. По одной работе (т. е. по 3,0 %) пришлось на темы лучевой диагностики легочной эмболии, пневмоторакса, остеопороза, колоректальных полипов, ожогов; также один метаанализ рассматривал проблематику оценки риска задержки внутриутробного развития.

Таким образом, максимальные усилия ученых в мире сосредоточены на создании технологий ИИ для выявления по результатам лучевых исследований онкологических заболеваний и болезней центральной нервной системы.

В большинство метаанализов – 20 из 38 (53,0 %) – включены исследования нескольких различных модальностей, в остальных 43,0 % – лучевые исследования только одной. Компьютерной томографии (КТ) посвящено 45,0 % (18) метаанализов; причем как единственному методу – только три. ИИ для магнитно-резонансной томографии (МРТ) изучался в 29,0 % (11) метаанализов, причем как единственный метод – в трех. Автоматизация анализа рентгенографии исследовалась в 16,0 % (6) метаанализов, при этом только в одном из них – как единственный метод. Отдельно рентгенография органов грудной клетки исследована в 11,0 % (4) метаанализов, при этом в двух из них – как единственный метод. ИИ в контексте проведения позитронно-эмиссионной томографии, совмещенной с КТ (ПЭТ КТ), исследован в 5,0 % (2). Автоматизированному анализу цифровой маммографии было посвящено также 5,0 % (2) исследований (причем в одно из них включили изучение томосинтеза). Применимость технологий ИИ в ультразвуковой диагностике (УЗД) стала объектом исследования в 16,0 % (6) метаанализов, как единственный метод – в четырех из них. Конусно-лучевая компьютерная томография (КЛКТ) изучена в 8,0 % (3) стоматологических метаанализов. В одной (3,0 %) работе об обнаружении легочной эмболии изучали исключительно КТ-ангиографию. Обращает на себя внимание тот факт, что в пяти обзорах не было корректных данных о модальностях.

Таким образом, большинство исследователей пытаются систематизировать данные о результативности ИИ для совокупности модальностей (чаще всего – КТ, рентгенографии, ультразвуковой диагностики), преимущественно в контексте решения конкретной клинической задачи. Обращает на себя внимание малое количество работ, систематизирующих вопросы автоматизации маммографии, что явно не соответствует клинической, демографической и социально-экономической значимости данной разновидности лучевых исследований.

Абсолютное большинство исследователей изучают диагностическую точность технологий ИИ – это первая по распространенности научная задача. В 92,0 % (35) метаанализов в качестве основных диагностических параметров точности использованы стандартные, общепринятые показатели – чувствительность, специфичность и площадь под ROC-кривой (AUROC). Соответствующие данные нами извлечены и систематизированы (таблица 1.1); далее рассчитали интервал обобщенных значений каждого показателя между метаанализами. Использовались при этом только данные, где были в наличии средние значения. Если показатели приводились по подгруппам, то брали среднее значение. Таким образом, при обобщении данных 38 метаанализов, комплексно отображающих ситуацию с изучением искусственного интеллекта в лучевой диагностике, значение чувствительности ИИ колебалось в диапазоне 75,4–97,9 %, в среднем оно составило 89,0 % (медиана – 88,7 %). Значение специфичности изменялось от 74,0 до 97,6 %, составив в среднем 88,8 % (медиана – 89,5 %). Диапазон значений площади под характеристической кривой был 75,0–98,0 %, среднее значение – 92,0 % (медиана – 93,8 %).

Таблица 1.1 – Сводные данные о показателях диагностической точности технологий искусственного интеллекта для лучевой диагностики по данным систематических обзоров и метаанализов [140, 141, 142, 147, 150, 151, 155, 167, 168, 172, 177, 180, 185, 188, 193, 196, 198, 199, 203, 208, 213, 219, 220, 223, 225, 229, 232, 234, 237, 243, 246, 247, 248, 249, 250, 252, 253, 254]

Первый автор, год	Чувствительность ИИ, %	Специфичность ИИ, %	AUROC ИИ, %
Yoon et al., 2023	80,6 (95%-й ДИ 74,3–85,7)	85,7 (95%-й ДИ 74,1–92,6)	87,0–90,0 (min–max)
Hickman et al., 2023	75,4 (95%-й ДИ 65,6–83,2)	90,6 (95%-й ДИ 82,9–95,0)	89,0 (95%-й ДИ 84,0–98,0)

Продолжение таблицы 1.1

Первый автор, год	Чувствительность ИИ, %	Специфичность ИИ, %	AUROC ИИ, %
Thong et al., 2023	94,6 (95%-й ДИ 91,4–96,7)	93,6 (95%-й ДИ 88,5–96,6)	–
Zheng et al., 2022	66,1–83,8 (min–max)	59,8–77,4 (min–max)	79,0 (95%-й ДИ 77,0–82,0) – 83,0 (95%-й ДИ 78,0–88,0)
Liu et al., 2023	87,0 (95%-й ДИ 82,0–90,0)	87,0 (95%-й ДИ 82,0–91,0)	93,0 (95%-й ДИ 91,0–95,0)
Sugibayashi et al., 2023	84,0 (95%-й ДИ 79,0–89,0)	96,0 (95%-й ДИ 94,0–98,0)	97,0 (95%-й ДИ 96,0–98,0)
Kuo et al., 2022	91,0 (95%-й ДИ: 84,0–95,0)	91,0 (95%-й ДИ 81,0–95,0)	96,0 (95%-й ДИ 94,0–98,0)
Zhang et al., 2023	90,0 (95%-й ДИ 87,0–92,0)	92,0 (95%-й ДИ 90,0–94,0)	97,0 (95%-й ДИ 95,0–98,0)
Lex et al., 2023	89,3±8,5	87,5±9,9	–
Potipimpanon et al., 2022	86,0 (95%-й ДИ 81,0–91,0)	78,0 (95%-й ДИ 73,0–83,0)	89,0 (95%-й ДИ 86,0–92,0)
Campello et al., 2023	81,7 (95%-й ДИ 77,2–85,4)	84,8 (95%-й ДИ 76,0–90,8)	–
Nabizadeh et al., 2023	92,0 (95%-й ДИ 90,0–95,0)	93,0 (95%-й ДИ 90,0–96,0)	93,0 (95%-й ДИ 89,0–96,0)
Din et al., 2023	91,2 (95%-й ДИ 82,2–95,8)	–	93,6
Jørgensen et al., 2023	96,0 (95%-й ДИ 93,0–97,0)	97,0 (95%-й ДИ 90,0–99,0)	98,0 (95%-й ДИ 97,0–99,0)
Kim et al., 2021	77,0 (95%-й ДИ 70,0–83,0)	74,0 (95%-й ДИ 64,0–82,0)	–
Adamou et al., 2022	–	–	–

Продолжение таблицы 1.1

Первый автор, год	Чувствительность ИИ, %	Специфичность ИИ, %	AUROC ИИ, %
Agarwal et al., 2023	90,0 (95%-й ДИ 85,0–94,0)	90,0 (95%-й ДИ 83,0–95,0)	80,0–99,1
Bedrikovetski et al., 2021a	–	–	80,8 (95%-й ДИ 73,9–87,6) – 91,7 (95%-й ДИ 88,2–95,2)
Soffer et al., 2023	88,0 (95%-й ДИ 80,3–92,7)	86,0 (95%-й ДИ 75,6–92,4)	85,0–95,0
Yang et al., 2022	–	–	–
Gao et al., 2021	96,0 (95%-й ДИ 93,0–100,0)	95,0 (95%-й ДИ 91,0–99,0)	86,0–100,0
Rescinito et al., 2023	84,0 (95%-й ДИ 80,0–88,00)	87,0 (95%-й ДИ 83,0–90,0)	–
Wang et al., 2022	87,0 (95%-й ДИ 85,0–89,0)	85,0 (95%-й ДИ 82,0–87,0)	94,0 (95%-й ДИ 91,0–96,0)
Poly et al., 2021	95,0 (95%-й ДИ 94,0–95,0)	96,0 (95%-й ДИ 96,0–97,0)	98,0
Komolafe et al., 2023	91,0 (95%-й ДИ 88,0–93,0)	92,0 (95%-й ДИ 88,0–94,0)	95,0 (95%-й ДИ 88,0–92,0)
Abesi et al., 2023	–	–	–
Badr et al., 2022	93,0 (95%-й ДИ 84,0–97,0)	–	–
Xing et al., 2023	–	–	75,0 (95%-й ДИ 67,0–84,0)
Menon et al., 2023	86,5 (95%-й ДИ 81,1–90,6) – 86,7 (95%-й ДИ 81,4–90,7)	76,1 (95%-й ДИ 69,9–81,4) – 87,1 (95%-й ДИ 78,0–92,8)	–
Taib et al., 2023	90,8–97,9	84,4–97,6	–

Продолжение таблицы 1.1

Первый автор, год	Чувствительность ИИ, %	Специфичность ИИ, %	AUROC ИИ, %
Sadr et al., 2023	92,5 (95%-й ДИ 86,2–96,0)	85,2 (95%-й ДИ 81,0–88,5)	–
Islam et al., 2022	93,8	91,7	96,0
Evangelista et al., 2022	84,0 (95%-й ДИ 58,0–100,0)	89,0 (95%-й ДИ 74,0–98,0)	92,0 (95%-й ДИ 72,0–100,0)
Xue et al., 2022	88,0 (95%-й ДИ 85,0–90,0)	81,0 (95%-й ДИ 74,0–86,0)	92,0 (95%-й ДИ 89,0–94,0)
Xu et al., 2022	88,0 (95%-й ДИ 85,0–90,0)	85,0 (95%-й ДИ 82,0–88,0)	93,0 (95%-й ДИ 91,0–95,0)
Dumitrescu et al., 2022	92,0 (95%-й ДИ 89,0–95,0)	90,0 (95%-й ДИ 83,0–94,0)	95,0
Bedrikovetski et al., 2021b	72,0–100,0	56,6–76,2	79,8–89,5
Xu et al., 2021	84,8 (95%-й ДИ 69,2–93,2) – 94,3 (95%-й ДИ 92,7–95,5)	89,4 (95%-й ДИ 63,1–97,7) – 96,5 (95%-й ДИ 94,6–97,7)	95,0 (95%-й ДИ 93,0–97,0) – 98,0 (95%-й ДИ 96,0–99,0)

В подавляющем большинстве (36 из 38) метаанализов отмечается высокая диагностическая точность ИИ – как чувствительности (доли истинно положительных результатов), так и специфичности (доли истинно отрицательных результатов), а также AUROC, отражающей соотношение между истинно положительными и ложноположительными результатами. Высокие значения показателей точности ИИ выявлены при диагностике ЗНО, остеопороза, переломов, COVID-19, а также в области нейровизуализации. Вместе с тем фактически все приводимые результаты получены в ретроспективном дизайне, на дискретных наборах данных; выводы многочисленных авторов совершенно не подкреплены информацией о работе технологий ИИ в реальных клинических условиях отделений и центров лучевой диагностики.

Второй по распространенности исследовательской задачей является сравнение диагностической точности врачей-рентгенологов и технологий

искусственного интеллекта. Эта задача рассматривается в 26,0 % (10) метаанализов, пять из которых имеют онкологическую направленность. Надо подчеркнуть, что в 50,0 % этих метаанализов отсутствовали данные о квалификации, опыте врачей-рентгенологов, что, несомненно, снижает объективность результатов. В одной работе (Campello et al.) диагностическую точность ИИ оценивали по нативным изображениям, а врачей – по изображениям с контрастным усилением [155]; данную работу рассмотрели отдельно.

Опираясь на данные из семи указанных выше работ (где соответствующая информация была представлена в полном объеме), определили интервалы показателей диагностической точности для ИИ и для врачей [252, 180, 232, 198, 253, 220, 248].

Значение чувствительности ИИ колебалось в диапазоне 75,4–91,0 %, в среднем составляя 85,2 % (медиана – 86,0 %); для врачей-рентгенологов диапазон составил 73,0–94,0 %, среднее – 84,4 % (медиана – 85,0 %). Специфичность ИИ колебалась от 78,0 до 96,0 %, в среднем достигая 89,5 % (медиана – 91,0 %); для врачей диапазон составил 82,0–98,0 %, среднее – 90,0 % (медиана – 94,0 %).

Диапазон значений площади под характеристической кривой для ИИ был 89,0–97,0 %, среднее составило 93,5 % (медиана – 92,5 %); соответственно для врачей – 85,0–98,0 %, среднее – 92,8 % (медиана – 94,0 %).

Лишь в двух метаанализах приведены уровни статистической значимости при сравнении диагностических параметров ИИ и врачей. В обзоре Yoon et al., 2023, рассматривающем диагностику ЗНО посредством цифровой маммографии и томосинтеза, показано, что чувствительность ИИ (80,6 %) достоверно выше ($p = 0,031$), чем 143 суммарно включенных врачей (73,6 %); показатели специфичности ИИ (85,7 %) и врачей (89,6 %) не различаются достоверно ($p = 0,221$); значения AUROC не различаются достоверно ($p = 0,152$) между ИИ (87–90 %) и врачами (81–96 %) [252].

В обзоре Nickman et al., 2023, по аналогичной тематике, но без томосинтеза, показано, что чувствительность ИИ (75,4 %) и 211 суммарно включенных врачей (73 %) не различается достоверно ($p = 0,7$); специфичность ИИ (90,6 %) и врачей

(88,6 %) не различается достоверно ($p = 0,73$) [180].

Еще в четырех метаанализах, рассматривающих диагностику пневмоторакса с использованием рентгенографии грудной клетки и КТ [232], переломов костей конечностей с использованием рентгенографии и КТ [198, 253], узлов щитовидной железы с использованием ультразвукового исследования [220], указано отсутствие разницы в значениях диагностических показателей между ИИ и врачами; при этом приведены сами показатели, но отсутствуют значения p -value.

В целом отмечается крайне высокая гетерогенность количества врачей, данных об их стаже, квалификации, компетенциях. Отсутствуют принятые подходы к описанию параметров отбора врачей-рентгенологов для участия в исследованиях. Складывается четкое убеждение, что единственным критерием отбора является личная заинтересованность и энтузиазм конкретного врача.

Метаанализ Campello et al., 2023, рассматривающий диагностику злокачественных новообразований печени с использованием ультразвукового исследования, отличается тем, что врачи анализировали изображения, полученные с помощью контрастного усиления, а ИИ – нативные изображения. Авторы этой работы отмечают высокие диагностические показатели ИИ, но уровни значимости вновь отсутствуют [155].

Таким образом, по данным зарубежных авторов, диагностическая точность ИИ не уступает диагностической точности врачей, а в некоторых случаях может превосходить ее. В частности, показано, что диагностическая точность ИИ сравнима с показателями врачей-рентгенологов экспертного уровня, превосходя показатели «обычных» врачей [248]. Тем не менее число работ, проводящих прямые количественные сравнения между ИИ и врачами, даже в глобальной перспективе по-прежнему мало. Существенно снижает качество работ и отсутствие корректных характеристик компетенций включаемых в исследования врачей.

Так же, как и в отечественной научной литературе, зарубежные авторы крайне поверхностно изучили результаты внедрения ИИ в работу отделений

лучевой диагностики. Доказанные результаты внедрения ИИ приведены только в одном (3 %) метаанализе [142]. Его авторы выявили три оригинальные статьи с доказанными результатами внедрения ИИ в клиническую практику. Все указанные работы относятся к сфере нейровизуализации с использованием КТ. Два научных коллектива поместили модель ИИ в начало клинического пути – она осуществляет предварительную сортировку перед интерпретацией результатов исследования врачом-рентгенологом. В одной из указанных работ благодаря ИИ сокращено с 512 до 19 минут среднее время описания «несрочных» исследований (в эту категорию ИИ помещал результаты, подлежащие пересмотру). Во второй работе отмечено достоверное снижение за счет автоматизации среднего времени описаний для амбулаторных (с 674 до 70 минут, $p < 0,001$) и стационарных больных (с 390 до 352 минут, $p = 0,002$). Вместе с тем для экстренных случаев сокращение длительности подготовки протокола не выявлено. Необходимо отметить, что обе работы проигнорировали анализ потенциального вреда и отсрочки оказания медицинской помощи из-за ложноотрицательных результатов. Между тем удельный вес таких результатов составил 7,5 и 11,6 % соответственно [142].

В двух работах, включенных в метаанализ Agarwal et al., 2023, программное обеспечение на основе технологий ИИ использовано для второго просмотра после интерпретации результатов лучевого исследования врачом-рентгенологом. Посредством автоматизированного анализа выявлено до 1,2 % ложноотрицательных результатов врачей-рентгенологов, что позволило провести мероприятия по повышению качества работы отделений лучевой диагностики [142].

Во всех обобщающих публикациях отмечается высокий потенциал для внедрения ИИ в практику, вместе с тем это утверждение обычно является декларативным. Реальные результаты приводятся только в одном метаанализе; причем соответствующее обобщение строится лишь на трех оригинальных статьях. Результаты использования ИИ в клинической практике неоднозначны. С одной стороны, автоматизация дала экономию времени при «несрочных»

исследованиях, но такого эффекта нет в неотложных ситуациях. В качестве инструмента верификации качества работы врачей-рентгенологов ИИ эффективно выявлял ложноотрицательные результаты, но одновременно точность выявления ложноположительных решений врача оказалась неудовлетворительной.

В процессе изучения 38 метаанализов применения ИИ в лучевой диагностике нами выявлены серьезные методологические дефекты многих научных работ, существенно влияющие на результаты и выводы.

Прежде всего, отмечается критично недостаточное число проспективных исследований (преобладает дизайн «случай – контроль») и работ с внешней валидацией данных (что, впрочем, отмечается и авторами самих метаанализов). Далее, фиксируется значительный разброс в размерах выборок пациентов между исходными публикациями, достигающий иногда 3–4 порядков. В некоторых статьях выборка экстремально мала и может не превышать двух десятков пациентов. Отмечается необоснованность критериев исключения, что также служит важным фактором снижения качества исследований. Типовой ошибкой является использование перекрывающихся наборов данных (выборок) в качестве обучающих, тестовых и валидирующих.

Довольно часто недостаточно детальное описание методики исследования приводит к снижению ценности его результатов. В ряде исследований неясен характер и способ заслепления, количество и компетенции врачей, время между индексным и референсным тестами. В целом фиксируется низкий уровень стандартизации дизайна и методов исследований, при том что неоднородность подходов и методов приводит к высокой неоднородности диагностических оценок.

Надо отметить, что в более чем 50,0 % метаанализов включены лучевые исследования разных модальностей, что чрезвычайно повышает неоднородность данных и затрудняет их обобщение.

В большинстве метаанализов риск предвзятости публикаций оценен как низкий. Однако отсутствие публикаций с отрицательными результатами дает основание пересмотреть градацию данного риска в пользу более значимой его

выраженности.

Таким образом, в глобальной перспективе результаты исследований применимости и качества ИИ в лучевой диагностике показывают достаточно высокую диагностическую точность технологий искусственного интеллекта, однако эти результаты в подавляющем большинстве случаев получены в исследованиях с некорректным дизайном, способом проведения и отчетностью, что фактически гарантированно приводит к систематическим ошибкам и переоценке эффективности алгоритмов. Еще раз подчеркнем преобладание ретроспективных исследований на дискретных наборах данных, а также отсутствие полноценных исследований в реальных клинических условиях.

1.3. Развитие технологий искусственного интеллекта в лучевой диагностике в Российской Федерации

Вопросы применения технологий искусственного интеллекта в различных клинических направлениях активно изучаются российскими учеными. Ведутся исследования применимости ИИ в гематологии [87], урологии и онкоурологии [99], кардиологии и функциональной диагностике [105, 111], дерматологии (преимущественно для скрининга злокачественных новообразований кожных покровов) [110], гастроэнтерологии [102], неонатологии [129], патоморфологии [8, 25], лабораторной диагностике [37, 40]; также разрабатываются прогностические системы поддержки принятия врачебных решений [38]. Отдельным перспективным направлением можно считать применение ИИ при проведении эндоскопических исследований желудочно-кишечного тракта и мочевыводящих путей [41, 65, 109]. Особенность здесь состоит в необходимости автоматизированного анализа динамичных видеоизображений в режиме реального времени с немедленным отображением результатов. Схожая проблематика существует и для ультразвуковой диагностики [71, 124]. Достаточно успешны разработки на основе ИИ для ретроспективного анализа электронных медицинских карт. Соответствующие решения в разных форматах

внедрялись в субъектах РФ для выполнения задач профилактической медицины [26, 68]. Особенно значителен прогресс в офтальмологии, где показана принципиальная достижимость для ИИ точности врача-офтальмолога при интерпретации изображений глазного дна в контексте скрининга диабетической ретинопатии, глаукомы и некоторых иных наиболее распространенных заболеваний [44, 58, 66, 86].

В сфере лучевой диагностики также отмечается научная и публикационная активность.

Разработана и протестирована модель глубокого обучения (подход 3D-классификации с помощью модели DenseNet) для типирования глиом на результатах МРТ головного мозга. Авторы использовали один набор данных ($n = 707$), разделив его на две части: 80,0 % – для обучения, 20,0 % – для тестирования. В таких условиях достигнута точность 83,0 %, площадь под характеристической кривой составила 0,95. Авторы отметили достижение цели – принципиальная возможность использования ИИ для конкретной клинической задачи в области лучевой диагностики достигнута [42].

В 2020 году опубликованы результаты ретроспективной оценки точности программного обеспечения на основе ИИ, имеющего статус медицинского изделия (RU.96876180.62.01.29-01). Исследование проведено на результатах 75 флюорографий. Авторы отметили 100,0 % чувствительность этой разработки, снабдили свою статью эмоциональными высказываниями («ИИ прекрасно распознавал патологии органов грудной клетки») и рекомендовали продукт к дальнейшей клинической валидации. Несколько странно звучит данная рекомендация в отношении уже зарегистрированного медицинского изделия. Явными ограничениями исследования являются: ретроспективный характер, малый объем выборки (ничем не обоснованный предварительно), а также довольно произвольное использование общепринятых показателей диагностической точности [7].

В 2022 году опубликованы результаты совместной разработки ФГБУ «НМИЦ колопроктологии им А.Н. Рыжих» Минздрава России и одной из

коммерческих компаний. С использованием набора данных из 900 результатов магнитно-резонансной томографии прямой кишки разработана «базовая модель искусственного интеллекта» на основе нейросетей SegResNet, TransUnet, 3D Unet. На исходном наборе данных (то есть без внешней валидации или проверки на новых данных) получена точность 77,0 %, чувствительность – 98,1 %, специфичность – 45,1 %, положительная прогностическая ценность – 72,9 %, отрицательная прогностическая ценность – 94,1 %. Низкую специфичность авторы объяснили высоким удельным весом «ложноположительных результатов у здоровых пациентов» (то есть при анализе изображений без признаков онкологической патологии). Авторы наметили пути дальнейшего развития своей разработки (улучшение специфичности, расширение анализируемых параметров, экспериментирование с параметрами обучения, увеличение набора данных) [48]. Однако они обошли вниманием необходимость внешней валидации; также проигнорирован аспект оценки точности и надежности разработки в проспективном режиме, в реальных клинических условиях.

Достаточно объемным исследованием последних лет стала серия научных работ А.А. Мелдо с соавт. Обосновав актуальность применения технологий ИИ для выявления признаков злокачественных новообразований на результатах КТ ОГК, авторы подробно описали этапы разработки технического решения на основе сиамских нейронных сетей, включая признаки формы, внутреннюю структуру и архитектуру дифференциальной диагностики. Также были предложены оригинальные математические методики для классификации анализируемых объектов на изображении, способ понижения размерности данных для повышения эффективности и скорости обучения моделей. Приводятся показатели точности авторской системы: чувствительность – 75,0–93,0 %, специфичность – 84,0–97,0 %, точность – 81,0–95,0 % (колебания обусловлены конкретной архитектурой нейросети) [78, 79, 80, 81, 85]. Вместе с тем принятая информация о процессе оценки точности, объемах и характеристиках использованных данных, внешней независимой проверке, наконец, клиническом применении системы не представлена. Авторы фокусируются на технологических

аспектах, алгоритмизации процессов дифференциальной диагностики, но вопросы воспроизводимости результатов, работы на реальных клинических данных полностью упущены.

Опубликовано краткое сообщение (тезисы) о применении в Красноярском крае программного обеспечения на основе ИИ для ретроспективного пересмотра результатов КТ ОГК, выполненных в период пандемии COVID-19. Из 10 500 исследований 484 (4,6 %) случая автоматически классифицированы как «подозрительные». После проверки врачами-рентгенологами 45 КТ-снимков отнесены к «группе пациентов с рентгенологическими признаками или подозрением на рак легкого». Полученные результаты авторы использовали как обоснование «эффективности гибридного подхода» к анализу результатов КТ ОГК [52].

Особое внимание российские исследователи уделяют проблематике применения технологий ИИ для анализа результатов профилактической маммографии.

Рожкова Н. И. с соавт., 2020, опубликовали результаты разработки и изучения диагностической значимости сегментационной нейросетевой модели детекции патологических изменений молочных желез. С применением набора данных из размеченных результатов 15 486 цифровых маммографий реализована модель на нейронной архитектуре ResNet50 с выводом результата при помощи градиентного бустинга. Точность сегментации определяли по коэффициенту сходства, вероятность злокачественности находок вычисляли с помощью показателей бинарной классификации. Результаты: разработана система, представленная сегментационной моделью на основе нейросетевой архитектуры. Модель выявляла критичные рентгенологические признаки (новообразования, кальцинаты, локальную перестройку структуры или асимметрию) с высокой точностью: коэффициент сходства – 0,8176 и выше при пороговых значениях на выходных нейронах сети 0,1 и 0,15. Авторы заявили, что результаты машинной сегментации и разметки изображений врачом-рентгенологом эквивалентны по своей точности как минимум в отношении выявления новообразований,

внеочаговых кальцинатов и внутрижелезистых лимфатических узлов. Полученные результаты послужили обоснованием возможности применения разработанного решения как системы поддержки принятия врачебных решений при анализе и интерпретации профилактических маммографических исследований [106]. Это чрезвычайно важное исследование, логичным продолжением которого должно послужить исследование точности и надежности авторской разработки в реальных клинических условиях.

Выполнены обзор и сравнение отдельных шести отечественных и зарубежных решений на основе ИИ для автоматизированного анализа результатов маммографии и улучшения выявляемости злокачественных новообразований молочной железы. Приведены данные о функциональных возможностях, а также о диагностической точности. Вместе с тем сопоставимость показателей вызывает определенные сомнения, так как они получены в разных условиях: как на собственных наборах данных разработчиков, так и в ходе независимой валидации. В целом работа подтверждает сложившееся мнение о значительном потенциале ИИ как системы поддержки принятия врачебных решений, направленной на снижение рисков пропуска значимой патологии и повышение производительности труда врачей-рентгенологов [116].

Представляет интерес публикация, появившаяся в процессе развития одного из отечественных программных продуктов на основе ИИ, впоследствии получившего статус медицинского изделия. Для создания системы анализа результатов рентгенографии органов грудной клетки использовали находящийся в свободном доступе набор данных «ChestX-ray8» и три разные архитектуры (нейросеть с обучением методом обратного распространения ошибки, нейронную сеть, работающую по конкурентному принципу, глубокую сверточную нейросеть). Для первого варианта архитектуры достигнута точность в 81,03 % при минимальном среднеквадратическом отклонении. Второй вариант достигал большее среднее значение точности (90,12 %), но величина погрешности была более высокой. Третий вариант показал максимальную точность и минимальную величину погрешности, но был сопряжен со значительными ресурсными

затратами на обучение [12]. Особую ценность работе придают сделанные авторами выделение, структурирование и анализ причин ошибок нейронных сетей. Это важный методический момент, часто упускаемый из виду в научных публикациях. Тщательный разбор ошибок позволяет пошагово и системно повышать точность работы ИИ в лучевой диагностике.

Встречаются отдельные работы, представляющие собой смешение обзора и тестирования некоего конкретного программного продукта (например, для автоматизированного анализа результатов конусно-лучевой компьютерной томографии в стоматологической практике) [14, 63]. Однако результаты апробации излагаются без достаточного объема статистического материала.

На этом фоне положительно выделяется исследование, опубликованное в 2022 году и посвященное сравнительному изучению различных систем ИИ для выявления очагов и округлых образований в легких. Авторы выбрали четыре программных продукта и анонимизировали их. Для тестирования использованы два оригинальных зарегистрированных набора данных из 150 цифровых рентгенограмм легких с различными верифицированными патологиями и из 5000 цифровых рентгенограмм без патологических признаков. На основе наборов данных сформированы три выборки с распространенностью патологических признаков 3,0, 6,0 и 50,0 %. Тем самым авторы смоделировали ситуации рентгенологического скрининга (первые две выборки) и диагностического пульмонологического центра (третья выборка). Включенные программные продукты последовательно протестированы. Для выборки 1 общая точность колебалась в диапазоне 90,0–94,0 %, площадь под характеристической кривой – 0,723–0,825; для выборки 2 соответствующие диапазоны составили 89,0–97,0 % и 0,750–0,911; для выборки 3 – 77,0–89,0 % и 0,770–0,890. Выявлены существенные различия в диагностической точности при работе ИИ на разных выборках. В разных ситуациях некоторые системы не могли преодолеть минимальный пороговый уровень показателя «площадь под характеристической кривой», либо показатели точности достаточно серьезно варьировались. Доказана необходимость отбора программного продукта четко под задачи и специфику

конкретной медицинской организации. Авторы сообщили, что показатели точности всех программных продуктов сопоставимы с результатами тестирования врачей-рентгенологов. Такое тестирование было проведено и опубликовано тем же авторским коллективом ранее (показатели чувствительности и специфичности врачей были 76,0 и 72,3 % соответственно) [115]. В процитированной публикации проведена независимая валидация четырех программных решений на основе технологий ИИ для лучевой диагностики. К сожалению, подобных работ исчезающе мало. Позиция авторов состояла в объективном установлении точности ИИ и в нежелании сравнивать конкретные продукты. Тем не менее отсутствие информации о протестированных разработках несколько снижает ценность публикации. Минимальной альтернативой могло стать хотя бы перечисление протестированных продуктов с последующей анонимизацией результатов тестирования. Подчеркнем особо, что выявленные авторами зависимости точности от характера выборки (фактически претестовой вероятности патологии) служат серьезным обоснованием необходимости проведения исследования технологий ИИ в реальных клинических условиях.

Остаются малоизученными аспекты влияния автоматизации на основе ИИ на производительность, качество работы врачей-рентгенологов, организацию и результативность лучевой диагностики. На экстремально малой выборке ($n = 5$) показано, что автоматизация рабочего процесса врача-рентгенолога за счет внедрения ИИ для анализа результатов рентгенографии ОГК оптимизирует трудозатраты, сокращает время на проведение рутинных процедур примерно на 30,0 %, повышает производительность труда (то есть увеличивает количество описываемых исследований) [107].

Пандемия новой коронавирусной инфекции создала колоссальные вызовы для систем здравоохранения всех стран мира; особую роль в борьбе с этим заболеванием сыграла лучевая диагностика. Глобально к выявлению, дифференциальной диагностике и контролю динамики течения COVID-19 применялись – с точки зрения рентгенологии – различные подходы. Часть стран сфокусировалась на применении компьютерной томографии, другая часть отдала

предпочтение рентгенографии. В России, в силу наличия особенностей субъектов, применялись обе модальности, также велись исследования в аспекте использования магнитно-резонансной томографии [242]. В любом случае ситуация привлекла значительное внимание разработчиков технологий ИИ. В мире был зафиксирован скачок числа публикаций о разработках и применении искусственного интеллекта для диагностики и оценки тяжести COVID-19 по результатам лучевых исследований [74, 77].

Необходимо подчеркнуть негативный факт. В период пандемии в России появился целый ряд оригинальных разработок – программных продуктов на основе ИИ для автоматизированного выявления признаков COVID-19 и оценки тяжести по результатам компьютерной томографии (прогресс в этой области наглядно демонстрирует сетевой каталог ИИ-сервисов – https://mosmed.ai/service_catalog/); часть таких разработок даже получили статус медицинского изделия. Вместе с тем научных публикаций о соответствующих решениях, их испытаниях и применении фактически не появилось. Отечественные авторы ограничивались литературными обзорами [74, 77, 127]. Единичные сообщения о применении технологий ИИ носили сугубо ретроспективный характер и основывались на малых объемах наблюдений (от 16 до 187 пациентов) [75, 118].

Отдельно стоит отметить работу Скоробогач с соавт., 2023, в которой технологии машинного обучения применены как метод исследования – с их помощью авторы выявили варианты течения вирусной пневмонии COVID-19, проведя автоматизированный количественный анализ результатов КТ ОГК. В качестве инструмента использовалось свободно доступное программное приложение с открытым кодом «3D Slicer» [113]. Единственным недостатком работы можно считать отсутствие информации о точности использованного программного решения в отношении определения объема поражения легочной ткани. Впрочем, это небольшое методическое упущение вовсе не снижает общую положительную оценку прогрессивного подхода, примененного авторами.

Использование ИИ, машинного обучения как метода обусловило развитие

радиомики – отдельного направления в современной лучевой диагностике, основанного на текстурном анализе изображений. Соответствующим проблемам посвящено значительное количество научных публикаций как в России, так и в мире в целом [2, 15, 16, 57, 59, 92, 93, 117, 120]. Однако радиомика, будучи отдельным и самостоятельным направлением в диагностике, находится за рамками нашего исследования.

Фактически единичная публикация посвящена вопросу создания наборов данных для обучения ИИ в лучевой диагностике острых нарушений мозгового кровообращения (ОНМК). Важно отметить, что набор ($n = 220$) включает не только результаты лучевых исследований (нативная КТ, КТ-ангиография), но и клинические данные – тип ОНМК, наличие сопутствующих заболеваний и осложнений, тактика лечения, длительность госпитализации, исход. Относительно разметки: на каждой серии изображений были оконтурены и протегированы области, соответствующие прямым и косвенным признакам ОНМК. Авторы полагают использовать сформированный набор для создания технологий ИИ для решения задач диагностики типа ОНМК, оценки объема поражения, а также – для прогноза степени неврологического дефицита [134].

Представлен опыт создания набора данных цифровых рентгенограмм позвоночного столба. Разметка данных выполнена всего лишь одним врачом-рентгенологом. Посредством набора разработан алгоритм сегментации костных структур и построения касательных к межпозвонковым дискам для последующего автоматизированного анализа и оценки степени тяжести идиопатического сколиоза. Достигнуто корректное распознавание угла Кобба более чем на 85,0 % рентгенограмм. Положительной оценки заслуживает взвешенная позиция авторов, утверждающих, что данное исследование представляет собой лишь первый шаг к созданию полноценного программного продукта для морфометрии [72].

Только в единичных работах рассматривается возможность обогащения визуализации клиническими данными. Например, изучена возможность совместного анализа результатов КТ легких и «текстовых радиологических данных» с целью повышения точности автоматизированного выявления

эмфиземы. Авторы использовали набор данных из 111 случаев с визуализацией и текстовыми данными и 108 случаев только с результатами КТ. Использование информации о локализации патологического проявления из текстовых данных позволило авторам повысить точность своей модели ИИ на 11 % (с 0,66 до 0,77), также получено повышение F-меры обнаружения эмфиземы с 0,55 до 0,73 [114]. Авторы подчеркивают экспериментальный характер своей работы, соответственно, проспективного применения разработки в реальных клинических условиях нет.

Обзорно рассмотрена проблематика создания наборов данных КТ-исследований легких. Опубликованы результаты разработки пакета программных средств, состоящего из приложений и скриптов конвертации для автоматизации процесса разметки медицинских изображений и подготовки обучающих наборов данных. Вместе с тем сведения о практическом использовании столь актуальной разработки отсутствуют [49].

Несмотря на представленные результаты, количество российских публикаций о применении и результативности ИИ в лучевой диагностике крайне мало. Подавляющее большинство статей в предметной области представляют собой обзоры [22, 43, 51, 80, 108, 123] или материалы технического характера.

Действительно, значительная группа научных статей и коротких сообщений посвящена описанию опыта разработки технологий ИИ для лучевой диагностики, основных параметров соответствующих математических моделей. Такие публикации создают лишь некую предварительную основу для изучения проблематики ИИ в лучевой диагностике [1, 3, 9, 28, 62, 76, 91]. Большинство из них носят сугубо «математический» характер, не соответствуют принятым в доказательной медицине дизайнам и совершенно не отвечают запросам медицинской науки и практики.

В частности, разработан алгоритм на основе сверточной нейросети U-Net для диагностики идиопатической нормотензивной гидроцефалии по четырем признакам, выявляемым на результатах магнитно-резонансной томографии. Основная клиническая цель определена как поддержка принятия решения о

необходимости хирургического лечения. Для разработки модели ИИ использован набор данных из 1800 МРТ около 600 пациентов. Тестирование точности проведено на группе из 90 пациентов с диагностированной целевой патологией. Авторы не приводят точных формулировок, но можно предположить, что решения алгоритма по 90 случаям были верифицированы тремя «врачами лучевой диагностики». В качестве показателя использовался коэффициент Соренсена-Дайса. По утверждению исследователей, точность алгоритма составила 90,0 %. Авторы указали, что алгоритм был интегрирован в чат-бот в одном из популярных интернет-мессенджеров. В чат-бот необходимо загрузить архив с DICOM-файлами, после чего происходит анализ и формирование рекомендаций [1]. Реализован весьма спорный подход, так как и методически, и технически, и юридически более корректным представляется размещение подобных решений путем интеграции с медицинскими информационными системами и/или государственными информационными системами в сфере здравоохранения субъектов РФ. Из текста публикации невозможно установить, относились ли 90 случаев, использованных для тестирования, к выборке для обучения нейросети. Также остается непонятным механизм поддержки принятия клинического решения нейрохирургом при использовании данной разработки. Собственно, валидация алгоритма самими авторами не включала изучение влияния разработки на решения врачей (то есть согласованность решений не проанализирована).

Некоторые авторы пытаются предлагать некие математические подходы, представляющие собой, скорее, решающие правила, но именуемые «искусственным интеллектом». Либо, декларируя разработку ИИ, они фокусируются на деталях аспектах клинических алгоритмов дифференциальной диагностики [3].

Предложен программный комплекс на основе «глубоких сверточных нейронных сетей» для анализа результатов рентгенографии с целью выявления и диагностики ревматоидного артрита. Авторы представили этапы работы их системы: нормализация рентгенограмм и подготовка к сегментации, собственно сегментация изображений и детекция признаков участков, оценка стадии

заболевания по одной из принятых шкал. Информация о наборе данных не приводится вовсе, лишь констатируется факт нехватки данных. Компенсация дефицита достигнута путем «аугментации данных» (нормализация, вращение, изменение резкости, добавление гауссового шума). Заявлены «устойчивая детекция» искомым признаковых участков и общая точность на уровне 71,0 % (на фоне полного отсутствия информации о наборе данных) [9].

С использованием публично доступного набора данных «BraTS» проведено сравнение методов классификации МРТ головного мозга: по базовым примитивам контура, на основе сверточной нейронной сети, с бинарным классификатором, на основе сверточной предобученной нейронной сети Xception. Проведя вычислительный эксперимент, авторы показали, что именно последний метод обладает наивысшей точностью [84].

Опубликовано краткое сообщение о разработке сегментационной нейросетевой модели для анализа результатов маммографии. Указаны некоторые технические аспекты, однако заявленной в цели исследования «оценки клинической эффективности» не проведено [76].

Выполнен сравнительный анализ методов классификации легочных узлов по результатам КТ органов грудной клетки и информативности соответствующих признаков (формы и текстуры). Авторы использовали набор данных «LIDS», находящийся в открытом доступе, из которого выбрали 321 изображение. Для классификации, также в сравнении, изучены различные алгоритмы машинного обучения: искусственная нейронная сеть, наивный байесовский классификатор, случайный лес, методы опорных векторов и k-ближайших соседей. Для каждого подхода определены показатели точности. Авторы полагают, что добились точности машинной классификации, «на 10 % превосходящей результаты опытных врачей-радиологов» [28]; вместе с тем данные о включенных врачах, уровни значимости выявленных различий они не приводят.

Предложена разработка на основе глубинного машинного обучения для сегментации структур легких на КТ, предназначенная для выявления признаков диссеминированной формы туберкулеза. Автор использовал набор данных из 11

результатов КТ с предварительно сегментированными легкими. Обучение собственной нейросети проведено только на одном кейсе, тестирование – на оставшихся 10. В дополнение к столь поверхностному подходу автор использовал нехарактерный для биомедицинской статистики показатель F1. Чувствительность и специфичность разработки составили 0,61 и 0,5–0,78 соответственно [62].

С применением набора данных из 41 исследования мультиспиральной компьютерной томографии с контрастированием разработана нейросеть для 3D-моделирования образований почек и прилежащих структур. При этом для обучения использовано 32 исследования, для валидации – 9. Для оценки точности использован коэффициент Соренсена-Дайса с моделями ANNet, DynUNet и с тремя вариантами модели nnU-Net. Максимально достигнутые показатели составили для нормальной паренхимы почки – 0,89, новообразований почки – 0,58, артерий – 0,86, вен – 0,80, мочеточников – 0,80. Отмечены (вполне ожидаемые) трудности обнаружения новообразований по сравнению с определением паренхимы и сосудов. Авторы запланировали увеличить объем набора данных до 300 исследований и применить операции постобработки для улучшения модели [132].

В целом публикации математического характера довольно типичны: как по подходам и структуре, так и по своим методическим дефектам [10, 23, 27, 47, 56, 82, 101, 103, 121, 125]. За основу берется актуальная клиническая задача (внутричерепные кровоизлияния, новообразования в легких, туберкулез и т. д.) и публично доступный набор данных (КТ, МРТ, рентгенография). Впрочем, иногда в подобных работах и изначально выбранная задача носит довольно частный характер [53]. Далее приводится детальное описание процессов математического моделирования, построения одной или нескольких нейросетей с разными архитектурами, тестирования. Часто авторы фокусируются на проблематике предварительной подготовки данных, а в условиях их дефицита – на «аугментации» и довольно дискуссионных для сферы здравоохранения синтетических наборах [133]. Результаты валидации, как правило, представляются с использованием показателей, непринятых в области

биомедицинской статистики. Из числа стандартных показателей довольно часто выбирается один; чаще всего – чувствительность, что легко объясняется манипуляцией с результатами – именно этот показатель обычно наиболее высок. Общий дизайн таких публикаций не соответствует требованиям к диагностическим исследованиям. Клиническая значимость соответствующих разработок является чрезвычайно сомнительной. Еще одной проблемой сугубо математических работ является частое отсутствие в составе авторов представителей медицинских специальностей. Отсюда вытекают некорректные подходы к дизайну исследования и интерпретации результатов, необоснованные задачи и откровенные ошибки (например, выражение «легочное помутнение»).

Разработана модель ИИ для автоматизированного анализа результатов рентгенографии органов грудной клетки. Для обучения использованы наборы данных из Японии и США, находящиеся в открытом доступе (общее количество использованных изображений составило 112 120). Ядро разработки представляет собой ансамбль из 10 нейросетей на основе архитектуры DenseNet-121. Авторы не приводят ни одного показателя диагностической точности, лишь утверждают, что модель может обработать 32 снимка в секунду и демонстрирует точность, сопоставимую с точностью врача, а по некоторым патологическим признакам и превышающую ее. В публикации сочетаются безапелляционность и бездоказательность, довольно типичные для многих математических публикаций об ИИ в лучевой диагностике [91].

Изредка подобные технические публикации дополняются результатами пилотных проектов по апробации разработок [46].

В частности, изучена возможность применения ИИ в качестве системы поддержки принятия решения врачей-рентгенологов для выявления узлов размером более 6 мм в легких при КТ органов грудной клетки на примере одной из коммерчески доступных российских разработок. При этом проводилось сравнение работы врачей в «обычное время» и в условиях пандемии новой коронавирусной инфекции. На выборке из 150 пациентов получены значения точности ИИ: чувствительность – 1,0, специфичность – 0,88 и AUC – 0,94. В

группе также из 150 пациентов, но с признаками COVID-19 (КТ-0 – КТ-1) показатели составили 0,93, 0,81 и 0,86 соответственно. Снижение специфичности авторы объяснили увеличением ложноположительных результатов за счет множественных уплотнений легкого вследствие коронавирусного поражения, которые и были приняты алгоритмом за легочные узлы [97]. Явными недостатками указанных исследований являются: ретроспективный дизайн, ограниченная выборка (максимум около 150–200 пациентов) и конфликт интересов, так как в авторские коллективы входят представители компаний – разработчиков решений на основе ИИ.

Таким образом, в России отмечается научно-исследовательская активность в области развития технологий искусственного интеллекта в лучевой диагностике. Подавляющее количество научных публикаций технического и математического характера имеют ограниченное значение для медицинской науки и практики. Во многих работах допускаются типовые методологические ошибки: необоснованные объемы выборок, отсутствие внешней валидации, несоответствие принятым стандартам для выполнения и описания диагностических исследований. Исследования об ИИ в лучевой диагностике клинической направленности отличаются сугубо ретроспективным дизайном, что не позволяет получить достоверные данные о точности и надежности решений на основе искусственного интеллекта в реальных производственных условиях.

Глава 2. МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ

В основу диссертационного исследования положен системный подход, подразумевающий рассмотрение проблемы качественного развития лучевой диагностики за счет внедрения технологий искусственного интеллекта как целостного комплекса (системы) взаимосвязанных и взаимодействующих элементов (клинических, методологических, организационных, технологических и прочих). В частности, нами применены следующие принципы системного подхода:

1. Системность. Каждый указанный элемент имеет такой ключевой признак всей системы, как единство главной цели: качественное развитие лучевой диагностики для улучшения состояния здоровья населения.

2. Иерархичность. Все элементы системы находятся во взаимодействии, при этом они имеют разные виды взаимного подчинения и связей; например, методологические элементы находятся выше организационных и технологических.

3. Целостность. Указанная выше проблема рассмотрена как единая система, но параллельно и как подсистема для вышестоящего уровня (лучевая диагностика как часть и компонент всей системы здравоохранения).

4. Структуризация. Диссертационное исследование подразумевает научный анализ указанных элементов, выявление их связей, взаимного влияния и значения.

5. Множественность. Диссертационное исследование включает различные методы научного познания, применяемые в отношении объекта и предмета исследования и реализуемые в дизайне исследования со смешанными методами.

Диссертационное исследование выполнено в рамках научного Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы, проведение которого было одобрено Независимым этическим комитетом МРО РОПР (протокол

№ 2/2020 от 20.02.2020). Для включения в исследование пациенты подписывали специальную форму информированного добровольного согласия, также использовалась специальная информационная брошюра (соответствующие материалы представлены публично в публикациях ранее [61]).

При систематизации и описании результатов диссертационного исследования руководствовались рекомендованными международным сообществом «The EQUATOR (Enhancing the QUality and Transparency Of health Research) Network» (<https://www.equator-network.org>) методологиями PRISMA-ScR [238], CONSORT-AI [204], StaRI [218], STARD 2015 [153], а также рекомендациями по научному описанию создания и валидации модели машинного обучения [35].

Научное исследование проведено в период с 01.01.2022 по 30.09.2023 в медицинских организациях Департамента здравоохранения города Москвы (МО ДЗМ), оказывающих первичную медико-санитарную и специализированную помощь взрослому населению в амбулаторных или стационарных условиях.

Дизайн: исследование со смешанными методами (параллельное применение количественных и качественных подходов для всестороннего изучения проблемы; в рамках диссертации этапы количественных и качественных исследований имеют равный статус, а их результаты используются для обобщений [227]).

Источники первичных данных:

1. Информационная система в сфере здравоохранения субъекта Российской Федерации «Единый радиологический информационный сервис Единой медицинской информационно-аналитической системы города Москвы (ЕРИС ЕМИАС)».

2. Форма ФСН № 30 «Сведения о медицинской организации (годовая)» (приказ Росстата от 14.01.2013 № 13).

3. Нормативно-правовые акты в сфере здравоохранения.

4. Научные публикации в рецензируемых журналах.

В исследование включены следующие программные продукты на основе ТИИ (n = 33):

1. ООО «КэреМенторЭйАй», CareMentor AI.
2. ООО «ФтизисБиоМед», FBM.
3. ООО «Медицинские скрининг системы», Цельс.
4. ООО «Платформа Третье Мнение», Третье мнение.
5. ООО «Платформа Третье Мнение», Программный модуль для анализа исследований компьютерной томографии органов грудной клетки (аорта).
6. ООО «Платформа Третье Мнение», Программный модуль для анализа исследований компьютерной томографии грудной клетки человека (гидроторакс).
7. ООО «Платформа Третье Мнение», Программный модуль для анализа исследований компьютерной томографии органов грудной клетки (легочный ствол).
8. АНО ВО «Университет Иннополис», AI-Radiology.
9. ООО «Честная медицина», qXR (qure.ai).
10. ООО «КэреМенторЭйАй», Система нейросетевая CareMentor AI для определения продольного плоскостопия по данным боковой рентгенография стопы под нагрузкой.
11. ООО «Оксиджен Технолоджиес Рус», Oxytech Spine XR Scoliosis.
12. ООО «Синапс Тех», Синапс Нейро.
13. ООО «СиВижинЛаб», CVisionRad – Knee Arthrosis.
14. «АО Медицинские технологии Лтд», Трио ДМ.
15. ООО «Медицинские скрининг системы», Цельс ММГ.
16. ООО «Платформа Третье Мнение», Третье мнение ММГ.
17. ООО «Интеллоджик», NTechMed СТ Brain.
18. ООО «Медицинские скрининг системы». Цельс.
19. ООО «Гаммамед-Софт». Multivox ASPECTS.
20. ООО «Интел Диагностик», Система поддержки принятия диагностических решений «АИ Диагностик» (гидроторакс).
21. ООО «Интел Диагностик», Система поддержки принятия диагностических решений «АИ Диагностик» (остеопороз).
22. ООО «Интел Диагностик», Система поддержки принятия

диагностических решений «АИ Диагностик» (позвоночник).

23. ООО «СиВижинЛаб», CVL – Chest CT Pleural Effusion.

24. ООО «АЙРА Лабс», Pulm Trunk-IRA.

25. ООО «Интел Диагностик», Система поддержки принятия диагностических решений «АИ Диагностик» (легочной ствол).

26. ООО «Интел Диагностик», Система поддержки принятия диагностических решений «АИ Диагностик» (аорта).

27. ООО «СиВижинЛаб», CVL – Chest CT Pulmonary trunk.

28. ООО «СиВижинЛаб», CVL – Abdomen CT Aorta, CVL – Chest CT Aorta.

29. ООО «АЙРА Лабс», Aorta-IRA.

30. ООО «АЙРА Лабс», Adrenal-IRA.

31. ООО «МИП «Биномикс-Рэй», HealthOST, HealthVCF.

32. ООО «АЙРА Лабс», Genant-IRA.

33. ООО «ИМВИЖН», IMV MS.

В тексте диссертации перечисленные ИИ-сервисы анонимизированы в целях устранения предубеждений и более объективного представления результатов.

Критерии включения:

1. Возраст старше 18 лет.
2. Пол мужской или женский.
3. Наличие в карте пациента направления на проведение и результатов проведения одного из видов лучевых исследований.

4. Результаты в стандарте DICOM следующих видов лучевых исследований: рентгенографии, флюорографии, маммографии профилактической, компьютерной томографии, магнитно-резонансной томографии.

Критерии невключения:

1. Лучевые исследования иной модальности.
2. Отсутствие подписанного информированного согласия на участие в исследовании.

Описание вмешательства. При интерпретации и описании результатов лучевых исследований применялось программное обеспечение на основе технологий искусственного интеллекта (далее – ИИ-сервисы), интегрированное в государственную информационную систему в сфере здравоохранения субъекта РФ – Единый радиологический информационный сервис автоматизированной информационной системы города Москвы «Единая медицинская информационно-аналитическая система города Москвы» (далее – ЕРИС ЕМИАС).

Были установлены следующие функциональные задачи ИИ-сервисов:

- локализовать патологические находки на изображениях в виде маркировок на дополнительной серии изображений, эквивалентных оригинальным изображениям исследования;

- классифицировать выявленные находки с указанием вероятности патологии и/или выполнить измерения (морфометрия);

- автоматически сформировать отчет о результате работы ИИ-сервиса с предлагаемым шаблоном описания лучевого исследования.

В соответствии с условиями научного исследования ИИ-сервисы могли иметь государственную регистрацию в качестве медицинского изделия либо представлять собой экспериментальную разработку. Соответствующие ограничения были указаны в добровольном информированном согласии пациентов на участие в исследовании, в соответствии с рекомендациями Независимого комитета по биоэтике (см. далее).

ИИ-сервисы применялись врачами-рентгенологами при описании результатов лучевых исследований в ЕРИС ЕМИАС. Перед внедрением каждого нового сервиса проводилось обучение пользователей.

При обращении пациента в МО ДЗМ осуществлялось стандартное обследование, в процессе которого лечащий врач устанавливал наличие показаний для лучевого исследования и формировал направление. Назначенное лучевое исследование выполнялось в установленном порядке, его результаты в формате DICOM автоматически сохранялись в ЕРИС ЕМИАС. Далее эти результаты маршрутизировались на определенный ИИ-сервис (в соответствии с

модальностью и видом лучевого исследования), который выполнял автоматический их анализ. Результаты такого анализа возвращались в ЕРИС ЕМИАС и появлялись на автоматизированном рабочем месте врача-рентгенолога параллельно с исходными данными лучевого исследования. Результаты работы ИИ-сервиса представляли собой: дополнительную серию изображений с маркировкой патологических находок, дополнительными данными (например, результатами измерений) и обязательным информационным сообщением «Только для исследовательских целей» (в соответствии с рекомендацией Независимого комитета по биоэтике); проект текстового описания результатов интерпретации данного лучевого исследования (технически – в виде DICOM SR).

Врач-рентгенолог, проводящий интерпретацию и описание исследования в ЕРИС ЕМИАС, мог открыть дополнительную серию; изучить результаты работы программного обеспечения на основе технологий искусственного интеллекта; по своему усмотрению использовать, не использовать или частично использовать эти результаты при формировании протокола исследования. Исходно обращение к результатам работы ИИ-сервиса было добровольным, зависело от индивидуального профессионального решения врача.

Диссертационное исследование состоит из четырех этапов. Общий ход исследования представлен на диаграмме (рисунок 2.1).



Рисунок 2.1 – Диаграмма хода исследования

Первый этап включал информационно-аналитическое изучение достигнутого уровня научных исследований в области применения искусственного интеллекта в лучевой диагностике. В силу высокой гетерогенности данных для источников на русском языке проведен выборочный обзор источников. Ввиду значительного количества публикаций в предметной области на английском языке (свыше 15 700 статей в PubMed только за период 2020-2023 гг.) принято решение провести систематический обзор метаанализов, в которых уже обобщены основные достижения и проблемы по отдельным аспектам предметной области.

Выборочный обзор литературных источников проведен по следующим критериям включения:

- научная статья (оригинальное исследование или обзор), опубликованная в рецензируемом журнале или сборнике трудов конференции;
- тематическая монография;
- соответствие тематике исследования.

Поиск источников проведен в библиографической базе Российского индекса научного цитирования «eLibrary». Глубина поиска составила 15 лет. Отобранные по критериям включения статьи проанализированы в полнотекстовом виде, выполнено реферирование, извлеченные данные проанализированы, упорядочены и систематизированы.

Систематический обзор метаанализов проведен в дизайне зонтичного систематического обзора литературы (в соответствии с методологией PRISMA-ScR).

Стратегия литературного поиска. Поиск осуществлялся по англоязычным статьям в библиографической базе Национальной медицинской библиотеки США «PubMed» и был ограничен временным интервалом 2021–2023 гг. Поисковый запрос выглядел следующим образом: “systematic review * OR meta-analysis * AND Neoplasm* OR Neoplasm, Lung* OR Radiotherapy * OR Cancer* OR Cancer, Lung* OR Pneumothorax* OR Fracture Detection* OR Fractures* OR bone* OR liver lesion* OR Thyroid nodule* OR Brain* OR Aneurysm* OR Lymph node* OR

Metastasis OR Osteoporosis OR Cervical Cancer OR dentistry” AND “Intelligence, Artificial OR Computational Intelligence OR Intelligence, Computational OR Machine Intelligence OR Intelligence, Machine OR Computer Reasoning OR Reasoning, Computer OR AI (Artificial Intelligence) OR Computer Vision Systems OR Computer Vision System OR System, Computer Vision OR Systems, Computer Vision OR Vision System, Computer OR Vision Systems, Computer OR Knowledge Acquisition (Computer) OR Neural network”.

Критерии включения: метаанализы диагностической точности ИИ в лучевой диагностике. Названия и аннотации найденных статей были проанализированы, по результатам была сформирована выборка для полнотекстового анализа.

Извлечение информации и оценка методологического качества исследований. Из полных текстов отобранных статей была извлечена следующая информация:

- 1) библиометрические данные (имя первого автора, название статьи, год выхода, DOI, название журнала, импакт-фактор журнала, страна проведения исследования);
- 2) параметры включенных исследований (модальность, объем выборки, доля проспективных исследований, наличие тестирования на внешних данных);
- 3) показатели качества включенных исследований (риск систематической ошибки, неоднородность данных, объективность критериев включения статей в обзор – «publication bias»);
- 4) значения диагностической точности ИИ; результаты сравнения диагностической точности ИИ с диагностической точностью врачей; информация о числе врачей и уровне их квалификации;
- 5) результаты внедрения.

Для оценки качества включенных систематических обзоров был использован инструмент AMSTAR-2 [228].

Поиск литературы и отбор работ. Результаты поискового запроса включали 2855 источников, для которых был проведен первичный анализ названий и аннотаций. В результате применения критериев включения было

исключено 2817 работ, которые не являлись обзорами либо принадлежали к иным предметным областям: медицина животных и гуманитарные науки. В итоговый анализ включено 38 метаанализов.

Результаты выборочного и систематического обзоров проанализированы совместно; это позволило в том числе уточнить задачи и детализировать ход исследования, а в дальнейшем сопоставить собственные результаты с литературными.

Второй этап представляет собой двухэтапное диагностическое исследование (в соответствии с методологией STARD 2015 [153]).

Первый подэтап диагностического исследования представляет собой ретроспективное диагностическое исследование программного обеспечения на основе технологий искусственного интеллекта (ИИ-сервисов) для анализа и интерпретации результатов лучевых исследований.

Индекс-тест: ИИ-сервис, интегрированный в тестовый контур ЕРИС ЕМИАС.

Референс-тест: эталонные наборы данных с разметкой по оригинальной методологии [104, 90, 67].

Для каждой клинической задачи использовался отдельный набор данных из 100 результатов исследований соответствующей модальности и вида. Подготовка и разметка эталонных наборов данных выполнена по научно обоснованной методологии [104, 90, 67]. Результаты разметки были недоступны для разработчиков ИИ-сервисов.

Базовые пороговые значения показателей диагностической точности взяты из утвержденных методических рекомендаций [60], они составили: чувствительность – 0,81; специфичность – 0,81; площадь под характеристической кривой – 0,81.

Второй подэтап диагностического исследования представляет собой проспективное, многоцентровое диагностическое исследование ИИ-сервисов для анализа и интерпретации результатов лучевых исследований.

В работу включены результаты 6 563 753 лучевых исследований по модальностям: рентгенография, флюорография, маммография, компьютерная томография, магнитно-резонансная томография.

Индекс-тест: ИИ-сервис, интегрированный в промышленный контур ЕРИС ЕМИАС для работы с реальными клиническими данными и прошедший необходимые технологические тестирования.

Референс-тест: заключение врача-рентгенолога, непосредственно проводившего интерпретацию и описание исследования в МО ДЗМ.

За весь период работы каждого ИИ-сервиса ежемесячно проводился технологический и клинический мониторинг по оригинальной методологии [61]. Соответствующие выборки формировались последовательно, проспективно, с включением исследований каждого вида, проанализированных ИИ-сервисами в период исследования. Результаты мониторингов за каждый квартал были проанализированы совместно, сформированы средние значения, которые и были использованы в диссертации.

На каждом подэтапе, по результатам работы каждого включенного в исследование ИИ-сервиса формировалась четырехпольная таблица; осуществлялось построение характеристической кривой; рассчитывались показатели «чувствительность», «специфичность», «площадь под характеристической кривой». При вычислениях устанавливался оптимальный порог активации по индексу Юдена [60].

Проведен анализ изменений диагностической точности ИИ-сервисов в динамике, для чего осуществлено построение модели смешанных эффектов для многократных измерений.

Также использована описательная статистика.

Данные, полученные на каждом из подэтапов, анализировались во взаимной связи.

Третий этап также состоит из двух компонентов, непосредственно не связанных между собой.

Первый компонент: изучение клинической целесообразности систем поддержки принятия врачебных решений (СППВР) выполнено в дизайне observationalного исследования.

Период исследования: 01.01.2022–30.09.2023.

Из ЕРИС ЕМИАС отобраны результаты компьютерной томографии, первично описанные врачами-рентгенологами, проанализированные ТИИ и прошедшие пересмотр врачами-экспертами в рамках мероприятий по контролю качества. Критерии включения:

- пол мужской или женский, возраст старше 18 лет;
- модальность, анатомическая область: компьютерная томография (КТ) органов грудной клетки;
- в ЕРИС ЕМИАС содержатся результаты компьютерной томографии ОГК и их описание;
- результаты компьютерной томографии ОГК обработаны ИИ-сервисами;
- результаты компьютерной томографии ОГК подверглись ретроспективному анализу в рамках мероприятий по контролю качества, при этом выявлены клинически значимые расхождения [89];
- патологические признаки, обусловившие клинически значимые расхождения, предусмотрены требованиями к результатам работы ТИИ (могут быть выявлены при автоматизированном анализе).

Критерии невключения:

- суть клинически значимого расхождения состояла в гипердиагностике или некорректной оценке динамики патологических изменений;
- патология, ставшая причиной клинически значимого расхождения, не предусмотрена требованиями к результатам работы ТИИ.

Включенные исследования проанализированы на основе следующего положения. Патологический признак был выявлен ИИ-сервисом непосредственно после выполнения исследования, а затем – врачом-экспертом в рамках планового контроля качества, но пропущен врачом-рентгенологом, выполнявшим первичное описание. Соответственно, такой признак послужил причиной клинически

значимого расхождения при интерпретации результатов данного лучевого исследования.

Использованы описательные методы статистики.

Второй компонент:

А) Экспериментальное обоснование возможности автономной морфометрии.

Оценка возможности автономной морфометрии выполнена на материале автоматизированного анализа результатов рентгенографии стопы с целью выявления продольного плоскостопия. Суть методики состояла в экспериментальном сравнении результатов автоматизированной морфометрии и врачебной разметки.

Использованы результаты работы ИИ-сервиса и 4 врачей-рентгенологов со стажем работы от 5 до 10 лет. Набор данных включал деперсонализированные результаты 100 рентгенографий, выполненных 100 различным пациентам, случайным образом отобранным в ЕРИС ЕМИАС.

Для снижения субъективного фактора при проведении разметки врачами-рентгенологами каждое уникальное изображение повторялось 4 раза и перемешивалось случайным образом с остальными изображениями, т. е. врачи проводили пересмотр 400 изображений, из которых 100 было уникальными. Те же 100 уникальных изображений были направлены на анализ ИИ-сервису для морфометрии.

Для сравнения результатов применен t-критерий Стьюдента.

Б) Проспективное одноцентровое клиническое исследование безопасности и качества автономной сортировки, осуществляемой ИИ-сервисами для анализа результатов профилактической маммографии, профилактической рентгенографии (флюорографии) органов грудной клетки.

Период исследования: 01.08.2023–31.10.2023.

В работу включены результаты 258 718 профилактических лучевых исследований по модальностям: флюорография, рентгенография, маммография.

Индекс-тест: ИИ-сервис, интегрированный в тестовый контур ЕРИС ЕМИАС.

Референс-тест № 1: протокол, оформленный врачом-рентгенологом.

Референс-тест № 2: экспертный пересмотр квалифицированным врачом-рентгенологом с субспециализацией.

Включенные ИИ-сервисы были настроены на чувствительность 1,0 (95%-й ДИ 1,0–1,0); значение специфичности было несущественно. В период исследования ИИ-сервисы осуществляли сортировку результатов лучевых исследований на две категории: «норма» и «не норма».

С применением инструмента для автоматизированного анализа текста протоколов проводилось сопоставление решений о категорировании ИИ-сервисов и врачей-рентгенологов. Клинические случаи расхождений направлялись на экспертный пересмотр квалифицированным врачом-рентгенологом с субспециализацией. По результатам пересмотра определялась правильность категорирования результатов данного исследования.

Использована описательная статистика и коэффициент межэкспертной согласованности каппа Коэна.

Четвертый этап исследования носил аналитический характер и включал систематизацию всех полученных результатов и формирование концептуальных положений, выводов, рекомендаций. Используются аналитические методы исследования.

Методы исследования

В диссертации использованы следующие научные методы: аналитические (анализ, синтез, индукция, дедукция); анализ и интерпретация результатов рентгенорадиологических исследований; оценка диагностической точности; математические и статистические.

Статистический анализ. В диссертации использованы различные методы статистического и математического анализа. Выбор конкретных методов и критериев осуществлен с учетом руководств по биомедицинской статистике

[20, 94, 100]. В частности, применены:

1. Описательная статистика (число непропущенных значений (N), минимальное значение (Min), максимальное значение (Max), арифметическое среднее (Mean), стандартное отклонение (SD), медиана (Med), первый и третий квартили (Q1 и Q3)).

2. Показатели диагностической точности: площадь под характеристической кривой (AUROC), точность, чувствительность, специфичность [60].

3. Критерии дисперсии Фишера (z-критерий), t-критерий Стюдента.

4. Коэффициент межэкспертной согласованности каппа Коэна.

5. Модель смешанных эффектов для многократных измерений (в качестве зависимой переменной включалось значение соответствующей диагностической характеристики, в качестве факторов – модальность (фиксированный эффект), сервис (случайный эффект), квартал оценки (повторяющийся фактор), а также взаимодействие модальности с кварталом; для каждого фактора дана оценка регрессионного коэффициента на линейной шкале и 95%-й ДИ).

6. Для определения значимости отличий использовали «р» – вероятность случайного характера различий (результаты значимы при $p < 0,05$), 95%-й доверительный интервал (ДИ) (результаты значимы в пределах интервала).

Обработка и анализ данных выполнены в статистической программе «MedCalc 18» (MedCalc Software Ltd, Бельгия).

Глава 3. ОБОСНОВАНИЕ ВОЗМОЖНОСТИ И СПОСОБОВ АВТОМАТИЗИРОВАННОГО АНАЛИЗА РЕЗУЛЬТАТОВ ЛУЧЕВЫХ ИССЛЕДОВАНИЙ

3.1. Анализ диагностической точности технологий искусственного интеллекта в лучевой диагностике

Технологии искусственного интеллекта представляют собой способ автоматизации производственных операций и процессов с целью получения измеримых положительных результатов. Исходя из этого, необходимо обосновать конкретные способы применения таких технологий в отношении типовых клинических и трудовых задач, возникающих перед врачом-рентгенологом. Эмпирически предположили, что основных способов применения технологий ИИ в контексте интерпретации и описаний результатов лучевых исследований может быть три:

1. Сортировка. Задача ТИИ: выявление результатов исследований, содержащих патологические изменения, и их маркировка в рабочем списке врача-рентгенолога в программном обеспечении автоматизированного рабочего места.

2. Система поддержки принятия врачебных решений (СППВР). Задача ТИИ: выявить и маркировать патологию на диагностическом изображении, подготовить проект ее текстового описания. При интерпретации результатов исследования врач-рентгенолог использует указанные результаты для ускорения и обеспечения качества (недопущения дефектов) своей работы.

3. Морфометрия. Задача ТИИ: выполнение принятых измерений анатомических структур и/или патологических проявлений. При интерпретации результатов исследования врач-рентгенолог не выполняет измерения самостоятельно, а руководствуется автоматически сформированными данными.

Обоснование возможности решения соответствующих задач может базироваться на достижениях технологиями ИИ определенного уровня диагностической точности (вновь в аспекте той или иной клинической задачи).

Отметим, что ранее эмпирическим путем был установлен базовый пороговый уровень каждого показателя точности ИИ в 0,81 [60].

В период исследования получены первичные данные о точности технологий искусственного интеллекта при анализе результатов лучевых исследований различных модальностей.

Для формирования общей картины и объективной оценки текущего уровня и тенденций развития точности технологий ИИ в предметной области нами выполнено двухэтапное исследование.

Первый этап: средствами описательной статистики нами обобщены показатели диагностической точности для каждой модальности в отдельности, соответствующие результаты представлены в таблице 3.1.

Таблица 3.1 – Описательная статистика показателей диагностической точности технологий искусственного интеллекта по модальностям лучевой диагностики

Показатель	Параметр	Модальность					
		РГ	ММГ	ФЛГ	КТ	МРТ	Суммарно
AUROC	N	61	19	24	68	7	179
	Mean	0,824	0,826	0,809	0,885	0,854	0,846
	SD	0,109	0,075	0,076	0,089	0,118	0,099
	95%-й ДИ	(0,796; 0,851)	(0,790; 0,862)	(0,777; 0,842)	(0,863; 0,907)	(0,744; 0,963)	(0,832; 0,861)
	Min	0,369	0,637	0,663	0,660	0,639	0,369
	Max	0,977	0,940	0,950	0,995	0,948	0,995
	Med	0,833	0,830	0,798	0,908	0,899	0,855
	Q1	0,789	0,803	0,773	0,823	0,739	0,795
	Q3	0,908	0,872	0,849	0,956	0,948	0,920

Продолжение таблицы 3.1

Показатель	Параметр	Модальность					
		РГ	ММГ	ФЛГ	КТ	МРТ	Суммарно
Точность	N	61	19	24	68	7	179
	Mean	0,801	0,723	0,729	0,865	0,861	0,810
	SD	0,095	0,136	0,115	0,115	0,090	0,122
	95%-й ДИ	(0,776; 0,825)	(0,658; 0,789)	(0,680; 0,778)	(0,837; 0,893)	(0,777; 0,944)	(0,792; 0,828)
	Min	0,567	0,375	0,504	0,469	0,721	0,375
	Max	0,950	0,863	0,912	0,990	0,954	0,990
	Med	0,800	0,762	0,727	0,900	0,890	0,840
	Q1	0,742	0,637	0,655	0,818	0,750	0,742
	Q3	0,880	0,840	0,817	0,948	0,938	0,900
Чувствительность	N	60	19	24	67	7	177
	Mean	0,889	0,870	0,901	0,935	0,912	0,907
	SD	0,074	0,110	0,080	0,064	0,099	0,080
	95%-й ДИ	(0,870; 0,909)	(0,817; 0,923)	(0,867; 0,934)	(0,919; 0,951)	(0,821; 1,004)	(0,895; 0,919)
	Min	0,611	0,636	0,735	0,694	0,739	0,611
	Max	0,989	1,000	1,000	1,000	1,000	1,000
	Med	0,903	0,867	0,906	0,950	0,958	0,926
	Q1	0,847	0,792	0,858	0,902	0,840	0,860
	Q3	0,947	1,000	0,964	0,987	1,000	0,974
Специфичность	N	61	19	24	68	7	179
	Mean	0,716	0,712	0,655	0,810	0,785	0,746
	SD	0,181	0,156	0,152	0,165	0,193	0,177
	95%-й ДИ	(0,669; 0,762)	(0,637; 0,787)	(0,591; 0,720)	(0,770; 0,850)	(0,607; 0,964)	(0,720; 0,772)
	Min	0,083	0,312	0,329	0,387	0,480	0,083
	Max	0,980	0,940	0,930	1,000	0,940	1,000

Продолжение таблицы 3.1

Показатель	Пара- метр	Модальность					
		РГ	ММГ	ФЛГ	КТ	МРТ	Суммарно
	Med	0,724	0,746	0,636	0,852	0,875	0,775
	Q1	0,646	0,627	0,563	0,693	0,545	0,627
	Q3	0,827	0,796	0,787	0,960	0,935	0,884

Примечание: N – число непропущенных значений (количество эпизодов измерения значений показателей), Min – минимальное значение, Max – максимальное значение, Mean – арифметическое среднее, SD – стандартное отклонение, Med – медиана, Q1 и Q3 – первый и третий квартили.

Среднее значение AUROC оказалось наиболее высоким для компьютерной томографии – $0,885 \pm 0,089$, медиана составила 0,908. Наименьшие значения отмечены для флюорографии: среднее – $0,809 \pm 0,076$, медиана – 0,798. Примечательно, что диапазон минимальных и максимальных значений был наиболее выражен для рентгенографии: от 0,369 до 0,977. Вместе с тем в половине случаев значение AUROC этой модальности попадало в диапазон 0,789–0,908, то есть было приемлемо высоким. Для остальных модальностей отмечено наличие минимальных значений AUROC в диапазоне 0,637–0,663. Такой уровень говорит о крайне низкой диагностической ценности, приближающейся к «угадыванию». Вместе с тем максимальные значения AUROC составили 0,940–0,995, что говорит о достижении рядом ИИ-сервисов высокой диагностической ценности, приближающейся или даже идентичной врачу-рентгенологу при решении отдельных клинических задач. Медиана AUROC была наименьшей для флюорографии – 0,798, наибольшей для КТ – 0,908. Отметим, что также для КТ половина всех значений AUROC попадала в диапазон 0,823–0,956; его верхняя и нижняя границы были самыми высокими по сравнению с иными модальностями. Базовый пороговый уровень каждого показателя точности составляет 0,81 [60]. Исходя из этого, выше порогового значения находились 50,0 % значений AUROC только для модальности «компьютерная томография». Надо подчеркнуть, что для трех модальностей – РГ, КТ и МРТ – значение третьего квартиля превышало 0,9.

Для двух иных модальностей оно было ниже.

Среднее значение обобщенного AUROC для всех модальностей составило $0,846 \pm 0,099$, медианы – $0,855$. Указанные значения можно считать одними из объективных параметров уровня развития технологий ИИ для лучевой диагностики на момент выполнения диссертации. Вместе с тем вариативность значений этого показателя точности подчеркивают уровни первого и третьего квартилей – $0,795$ и $0,920$ соответственно. Они ограничивают интервал, в котором находится $50,0\%$ значений обобщенного AUROC.

Среднее значение точности было наибольшим для КТ ($0,865 \pm 0,115$) и МРТ ($0,861 \pm 0,090$), наименьшим для маммографии ($0,723 \pm 0,136$) и флюорографии ($0,729 \pm 0,115$) (в таблице 3.1 приведены 95%-е доверительные интервалы). Диапазоны значений точности для всех модальностей, за исключением МРТ, были довольно обширными. Минимальные значения колебались на уровне $0,375$ – $0,567$, максимальные – на уровне $0,863$ – $0,990$. Это подчеркивало разнообразие уровней качества реализации конкретных прикладных решений – ИИ-сервисов; вместе с тем говорило о в целом технически достигнутом уровне точности, сопоставимом с врачом-рентгенологом или даже превосходящем его при решении отдельных клинических задач. Минимальное значение точности МРТ составило $0,721$ и было принципиально более высоким, чем у других модальностей. Максимальное достигло $0,954$ и также было относительно высоким. Выше базового порогового значения $0,81$ находились $50,0\%$ значений точности только для модальности «компьютерная томография». Примечательно, что лишь для КТ и МРТ значение третьего квартиля превышало $0,9$. Для остальных модальностей оно было ниже. Выше порогового значения $0,81$ была медиана опять для модальностей КТ и МРТ, а также максимально приблизилась к порогу – для РГ.

Среднее значение обобщенного показателя точности для всех модальностей составило $0,81 \pm 0,122$, медианы – $0,840$; их также можно считать объективными параметрами уровня развития технологий ИИ для лучевой диагностики на момент выполнения диссертации. Вместе с тем вариативность значений этого показателя точности подчеркивает тот факт, что $50,0\%$ значений обобщенной точности

находятся в интервале от 0,742 до 0,900.

Наибольшее среднее значение чувствительности вновь получено для модальности компьютерная томография ($0,935 \pm 0,064$), наименьшее – для маммографии ($0,87 \pm 0,11$).

Обращают на себя внимание максимальные значения этого показателя. У всех модальностей, за исключением рентгенографии, они достигали 1,0. Впрочем, и у РГ оно было очень высоким и составляло 0,989. На этом фоне лишь у флюорографии и у МРТ минимальное значение превысило 0,7. У всех иных модальностей оно было ниже (хуже всего у рентгенографии – 0,611). Базовое пороговое значение 0,81 превысили медианы чувствительности всех модальностей. Вместе с тем только у ММГ медиана была ниже 0,9. Напротив, наибольшие значения получены для МРТ – 0,958 и КТ – 0,950.

Также у всех модальностей 50,0 % значений чувствительности находились выше базового порогового уровня, исключение составила только маммография – первый квартиль для этой модальности равнялся 0,792. Вместе с тем только для маммографии и МРТ уровень третьего квартиля достиг 1,0, превысив, хоть и не всегда принципиально, уровни иных модальностей.

Среднее значение обобщенной чувствительности для всех модальностей составило $0,907 \pm 0,080$, медиана – 0,926. Эти значения можно считать одними из объективных параметров уровня развития технологий ИИ для лучевой диагностики на момент выполнения диссертации. Обращает на себя внимание «сжатый» и высокий диапазон, в котором лежит половина значений этого показателя: уровни первого и третьего квартилей составляют 0,86 и 0,974 соответственно.

Среднее значение специфичности оказалось наибольшим для КТ – $0,810 \pm 0,165$, наименьшим для флюорографии – $0,655 \pm 0,152$. Особого внимания заслуживает тот факт, что лишь для КТ и МРТ в пределах 95%-го доверительного интервала был преодолен базовый пороговый уровень. Для остальных модальностей верхняя граница доверительного интервала не превысила 0,78. Вместе с тем максимальные значения специфичности для всех модальностей

были высоки – от 0,93 до 1,0. Но при этом катастрофически низкими были минимальные значения – от 0,083 до 0,480. Также лишь для КТ и МРТ медиана специфичности превысила базовый пороговый уровень в 0,81. Половина значений специфичности не попали в диапазон, включающий или превышающий базовый пороговый уровень, для модальностей маммография и флюорография. Для остальных трех модальностей характерна выраженная разница значений первого и третьего квартилей: от примерно 0,6 до 0,8, от примерно 0,5 до 0,9. Полученные результаты говорят о существенной вариативности специфичности.

Среднее значение обобщенной специфичности для всех модальностей составило $0,746 \pm 0,177$, медианы – 0,775. Таким образом, только для этого показателя не был преодолен (или хотя бы включен в доверительный интервал) базовый пороговый уровень. Указанные значения приходится считать одними из объективных параметров уровня развития технологий ИИ для лучевой диагностики на момент выполнения диссертации. В целом специфичность отличается колоссальной вариативностью для каждой модальности.

Второй этап решения данной задачи состоял в изучении динамики показателей точности, в том числе по модальностям, а также в объективном сравнении точности ИИ для анализа результатов исследований различных модальностей. Для решения такой комплексной задачи использовано построение модели смешанных эффектов для многократных измерений (зависимая переменная – значение соответствующей диагностической характеристики, факторы – модальность (фиксированный эффект), ИИ-сервис (случайный эффект), квартал оценки (повторяющийся фактор), а также взаимодействие модальности с кварталом. Для каждого фактора дана оценка регрессионного коэффициента на линейной шкале и 95%-й ДИ.

Результаты представлены в таблицах 3.2–3.3.

Таблица 3.2 – Результаты анализа данных AUROC и точность с помощью смешанной модели

Фактор	β	t	p	95%-й ДИ для β
AUROC				
Время, на 1 квартал	-0,006	-1,31	0,193	-0,015; 0,003
Модальность (РГ = реф.)				
ММГ	-0,047	-1,13	0,262	-0,130; 0,036
ФЛГ	-0,024	-0,64	0,521	-0,099; 0,050
КТ	0,095	3,12	0,002	0,035; 0,156
МРТ	-0,042	-0,64	0,525	-0,172; 0,088
Взаимодействие				
Модальность*Время				
ММГ	0,013	1,44	0,152	-0,005; 0,031
ФЛГ	-0,000	-0,05	0,957	-0,018; 0,017
КТ	-0,010	-1,51	0,133	-0,023; 0,003
МРТ	0,021	1,45	0,148	-0,008; 0,050
Точность				
Время, на 1 квартал	0,001	0,23	0,815	-0,010; 0,013
Модальность (РГ = реф.)				
ММГ	-0,081	-1,62	0,107	-0,180; 0,018
ФЛГ	-0,079	-1,69	0,093	-0,171; 0,013
КТ	0,125	3,75	<0,001	0,059; 0,191
МРТ	0,004	0,05	0,957	-0,148; 0,156
Взаимодействие				
Модальность*Время				
ММГ	0,001	0,09	0,931	-0,022; 0,024
ФЛГ	0,002	0,19	0,852	-0,020; 0,024
КТ	-0,020	-2,36	0,020	-0,036; -0,003
МРТ	0,015	0,86	0,392	-0,019; 0,049

Таблица 3.3 – Результаты анализа данных чувствительности и специфичности с помощью смешанной модели

Фактор	β	t	p	95%-й ДИ для β
Чувствительность				
Время, на 1 квартал	0,013	3,36	0,001	0,005; 0,021
Модальность (РГ = реф.)				
ММГ	0,006	0,16	0,870	-0,062; 0,073
ФЛГ	0,052	1,69	0,093	-0,009; 0,114
КТ	0,095	3,86	<0,001	0,046; 0,144
МРТ	0,053	1,00	0,317	-0,051; 0,158
Взаимодействие				
Модальность*Время				
ММГ	0,003	0,45	0,656	-0,011; 0,018
ФЛГ	-0,011	-1,48	0,140	-0,025; 0,004
КТ	-0,011	-1,98	0,049	-0,022; -0,000
МРТ	-0,008	-0,71	0,478	-0,031; 0,015
Специфичность				
Время, на 1 квартал	-0,023	-2,80	0,006	-0,040; -0,007
Модальность (РГ = реф.)				
ММГ	-0,124	-1,66	0,098	-0,270; 0,023
ФЛГ	-0,105	-1,56	0,121	-0,238; 0,028
КТ	0,107	2,02	0,045	0,002; 0,211
МРТ	-0,101	-0,87	0,386	-0,329; 0,128
Взаимодействие				
модальность*Время				
ММГ	0,029	1,80	0,074	-0,003; 0,062
ФЛГ	0,013	0,86	0,391	-0,017; 0,044
КТ	-0,009	-0,76	0,447	-0,032; 0,014
МРТ	0,047	1,83	0,069	-0,004; 0,097

Показатель площади под характеристической кривой значимо не изменялся во времени, то есть выраженная динамика для него отсутствовала (p для фактора времени = 0,193). В сравнении с иными модальностями значительно более высокие показатели AUROC зафиксированы для компьютерной томографии

($\beta = 0,095$, 95%-й ДИ 0,035–0,156, $p = 0,002$).

Аналогично показатель точности оказался более высоким для КТ ($\beta = 0,125$, 95%-й ДИ 0,059–0,191, $p < 0,001$). Однако для всех остальных модальностей отсутствовала достоверная динамика во времени, а для КТ выявлено значимое ухудшение показателя точности на 0,021 единицы в квартал ($\beta = -0,020$, 95%-й ДИ от $-0,036$ до $-0,003$, $p = 0,020$). Вместе с тем при суммировании данных для всех модальностей достоверная динамика отсутствовала (p для фактора времени = 0,815); «вклад» динамики КТ не повлиял на общую ситуацию.

В отличие от предыдущих двух показателей, для чувствительности выявлена статистически значимая положительная динамика: зафиксирован прирост на 0,013 единиц в квартал (p для фактора времени = 0,001). При этом основной «вклад» сделан модальностями «рентгенография» и «маммография». В аспекте сравнения лидирующую позицию также занимает КТ ($\beta = 0,095$, 95%-й ДИ 0,046; 0,144, $p < 0,001$).

На фоне положительной динамики чувствительности отмечается отрицательная динамика специфичности: статистически значимое снижение на 0,023 единиц в квартал (p для фактора времени = 0,006). И в этой ситуации также отрицательная динамика в основном обусловлена РГ и КТ. С точки зрения сравнения с иными модальностями КТ сохраняет и здесь лидирующую позицию ($\beta = 0,107$, 95%-й ДИ 0,002–0,211, $p = 0,045$).

При обобщении анализа динамики и уровней показателей точности посредством модели смешанных эффектов для многократных измерений выраженного улучшения показателей диагностической точности ИИ по модальностям не отмечается. Вместе с тем четко проявляется тенденция «перенастройки» систем с целью повышения чувствительности за счет снижения специфичности. Целесообразность такого подхода – во всяком случае применительно к отдельным видам лучевых исследований – будет показана и доказана в последующих разделах диссертации. В целом именно компьютерная томография обладает явно наилучшими диагностическими характеристиками по сравнению с остальными модальностями, однако отмечается постепенное

сокращение этого разрыва.

Таким образом, при совокупном анализе диагностической точности технологий ИИ при анализе результатов лучевых исследований установлены средние значения обобщенных показателей: AUROC для всех модальностей – $0,846 \pm 0,099$, (медиана – 0,855), точность – $0,81 \pm 0,122$ (медиана – 0,840), чувствительность – $0,907 \pm 0,080$, (медиана – 0,926), специфичность – $0,746 \pm 0,177$ (медиана – 0,775). Во временной динамике отмечается постепенное повышение уровня чувствительности на 0,013 единиц в квартал ($p = 0,001$) на фоне также достоверного снижения специфичности на 0,023 единиц в квартал ($p = 0,006$).

В разрезе модальностей наблюдается преобладание точности технологий ИИ для анализа результатов компьютерной томографии. Однако во временной динамике выявлено значимое ухудшение показателя точности на 0,021 единицы в квартал ($p = 0,020$).

В целом полученные обобщенные результаты свидетельствуют о значительной вариативности и внушительных диапазонах колебания показателей точности. Это говорит о необходимости перехода от общего к частному.

Конкретные предложения по способам применения искусственного интеллекта в лучевой диагностике могут быть научно обоснованы только путем детального сравнительного анализа точности и клинической результативности отдельных ИИ-сервисов. Этому вопросу посвящены следующие разделы, в которых – для обоснования возможности применения технологий искусственного интеллекта для анализа и интерпретации результатов лучевых исследований, а также для обоснования конкретных способов такого использования – проведено двухэтапное (ретро- и проспективное) диагностическое исследование (дизайн, ход, материал и методы подробно описаны в главе 2).

3.2. Обоснование возможности автоматизированного анализа результатов рентгенографии

Рентгенография и флюорография органов грудной клетки. В контексте автоматизированного анализа результатов рентгенографии (РГ) или флюорографии (ФЛГ) органов грудной клетки с целью диагностики ряда заболеваний (прежде всего – социально значимых) была сформулирована следующая клиническая задача – определение наличия и локализации рентгенографических признаков, коррелирующих с приоритетной нозологией из перечня: 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); 8) кардиомегалия (I51.7); 9) патология средостения (D15.2, D38.3, I71); а именно:

– определение наличия рентгенологических признаков: плевральный выпот; пневмоторакс; очаг затемнения; инфильтрация/консолидация; диссеминация; полость; ателектаз; кальцинат/кальцинированная тень в легких; расширение средостения; кардиомегалия; нарушение целостности кортикального слоя; консолидированный перелом;

– локализация и идентификация патологических находок.

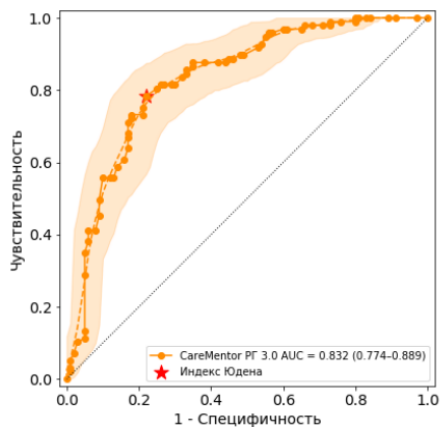
Необходимо отметить особенность ИИ-сервисов данной группы. Первоначально клиническая задача объединяла обе модальности – рентгенографию и флюорографию. В процессе технической реализации моделей ИИ у разработчиков появилось два типа программного обеспечения: сортирующего модальности на входе и анализирующего обе; анализирующего только рентгенографию. Для второго типа обработка флюорографии была невозможна в силу технических причин, изучение которых не является задачей диссертации. В контексте научного обоснования применимости технологий ИИ в лучевой диагностике сочли прагматичным базироваться на клинической задаче и рассматривать обе модальности одновременно.

Ретроспективное исследование диагностической точности шести¹ ИИ-сервисов для автоматизированного анализа результатов РГ/ФЛГ органов грудной клетки выполнено на эталонном наборе данных (n = 200). Построены характеристические кривые, результаты расчета показателей точности сведены в общую таблицу (рисунок 3.1, таблица 3.4).

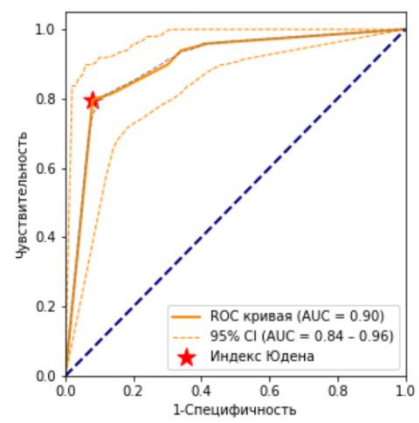
Таблица 3.4 – Показатели диагностической точности ИИ-сервисов для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при ретроспективном тестировании на эталонном наборе данных

Показатели	ИИ-сервис		
	ИИ-1	ИИ-2	ИИ-3
AUROC	0,83	0,9	0,95
95%-й ДИ	0,77–0,89	0,84–0,96	0,921–0,979
Точность	0,78	0,86	0,9
95%-й ДИ	0,7–0,87	0,77–0,92	0,860–0,940
Чувствительность	0,78	0,8	0,88
95%-й ДИ	0,7–0,87	0,66–0,9	0,810–0,940
Специфичность	0,78	0,92	0,93
95%-й ДИ	0,7–0,86	0,81–0,98	0,880–0,980
–	ИИ-4	ИИ-5	ИИ-6
AUROC	0,916	0,86	0,94
95%-й ДИ	0,878–0,954	0,79–0,94	0,885–0,994
Точность	0,85	0,86	0,95
95%-й ДИ	0,81–0,9	0,77–0,92	0,9–0,99
Чувствительность	0,9	0,83	0,91
95%-й ДИ	0,84–0,96	0,70–0,93	0,83–0,99
Специфичность	0,81	0,88	0,98
95%-й ДИ	0,73–0,89	0,75–0,95	0,94–1,0

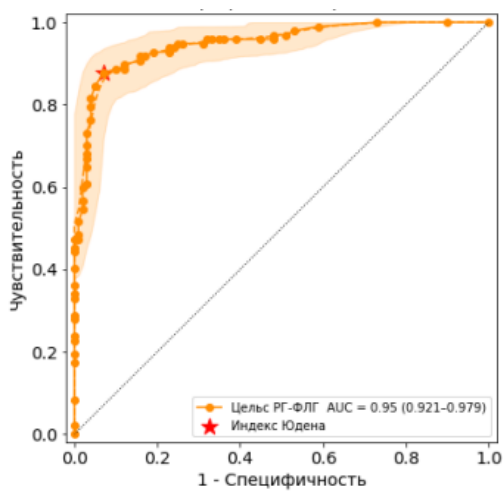
¹ ООО «КэреМенторЭйАй», CareMentor AI. ООО «ФтизисБиоМед», FBM. ООО «Медицинские скрининг системы», Цельс. ООО «Платформа Третье Мнение», Третье мнение. АНО ВО «Университет Иннополис», AI-Radiology. ООО «Честная медицина», qXR (Qure.ai).



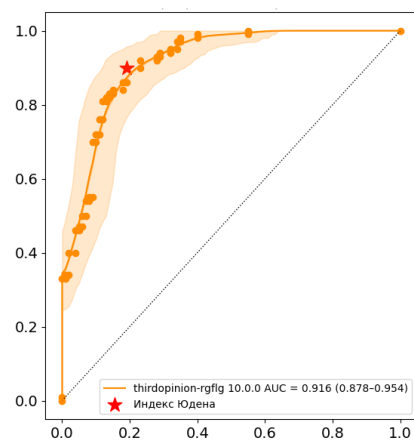
ИИ-1



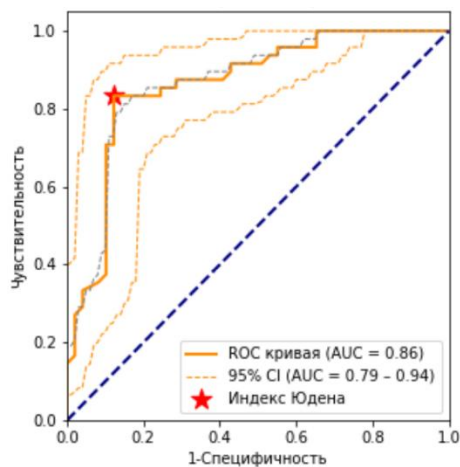
ИИ-2



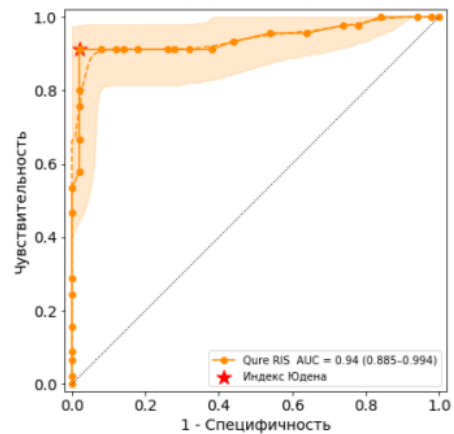
ИИ-3



ИИ-4



ИИ-5



ИИ-6

Рисунок 3.1 – Характеристические кривые для ИИ-сервисов для анализа результатов РГ/ФЛГ органов грудной клетки (оптимальный порог по индексу Юдена)

При тестировании на новых данных наименьшее значение AUROC (0,83 (95%-й ДИ 0,77–0,89)) зафиксировано у ИИ-1, наибольшие – у ИИ-3 и ИИ-6 (0,95 (95%-й ДИ 0,921–0,979) и 0,94 (95%-й ДИ 0,885–0,994) соответственно). Такая же тенденция отмечалась и для прочих показателей диагностической точности. Полученные результаты сопоставлены с заявленными базовыми значениями показателей (рисунок 3.2).

На новых данных ИИ-1 действительно продемонстрировал ухудшение точности: примерно на 7,0–13,0 % снизились все показатели, особенно сильно – чувствительность (на 13,3 %) и общая точность (на 10,9 %). Вместе с тем уровень гипо- и гипердиагностики был чрезвычайно высок – по 22,0 %.

У ИИ-2 и ИИ-3 отмечен рост значений показателей диагностической точности. Исключение у ИИ-2 составила чувствительность: она снизилась на 12,4 %. У этого же сервиса принципиально возросла специфичность – с 0,66 до 0,92, то есть на 39,4 %. Уровень гиподиагностики был также высок (20,0 %), а гипердиагностики – принципиально ниже (8,0 %). У ИИ-3 увеличение значений было относительно небольшим, но уверенным: AUROC возросла на 3,3 %, точность – на 9,8 %, чувствительность – на 7,3 %, специфичность – на 8,1 %. Однако данный сервис отличался низким – относительно ИИ-1 и ИИ-2 – уровнем пропусков патологии (7,0 %) и гипердиагностики (12,0 %).

Также неудовлетворительная ситуация зафиксирована у ИИ-4. Удельный вес ложноотрицательных случаев составил целых 10,0 %, а гипердиагностики – 19,0 %, при том что показатели разнонаправленно колебались в пределах 1,0–2,0 %, за исключением снизившейся на 3,5 % специфичности.

На этом фоне у ИИ-5 и ИИ-6 удельный вес ложноотрицательных результатов также был высок: 17,0 % и 12,0 % соответственно. А уровень ложноположительных результатов несколько меньше – 9,0 % и 2,0 % соответственно. Динамика показателей ИИ-5 была схожа с ИИ-2. Отмечен более значительный рост точности (на 16,7 %) и, особенно, специфичности (на 38,8 %). Вместе с тем чувствительность и AUROC снизились на 10,3 % и 5,2 % соответственно.

У ИИ-6 при тестировании на новых данных зафиксирован стабильный уровень площади под характеристической кривой и уверенный рост прочих показателей. Особенно значительно увеличились точность и специфичность – примерно на 10,0 %.

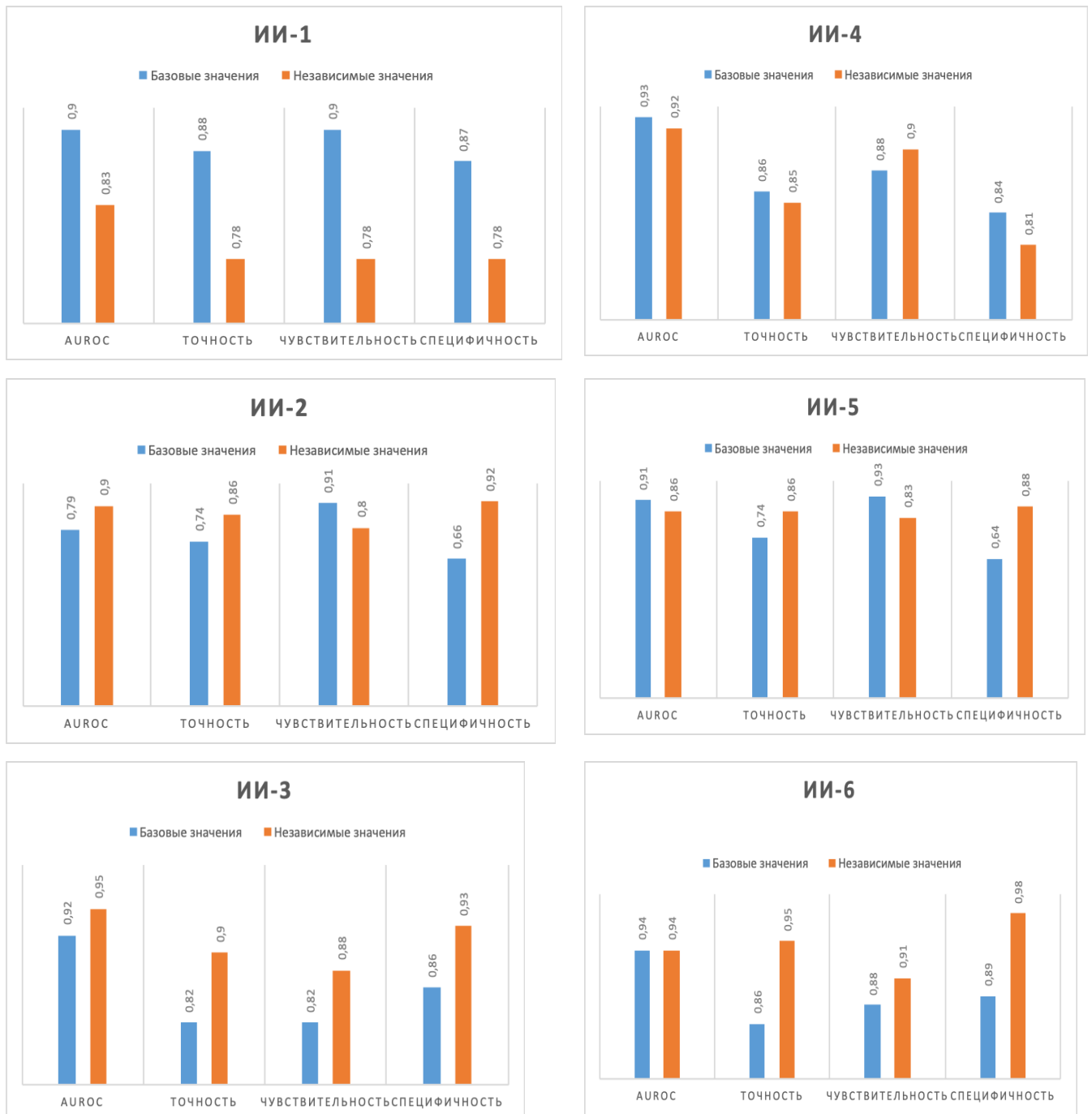


Рисунок 3.2 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервисов для анализа результатов РГ/ФЛГ органов грудной клетки

Перспективное исследование выполнено в условиях работы ИИ-сервисов с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией, проведено от пяти до семи измерений точности, в том числе для каждой модальности отдельно; результаты представлены в таблицах 3.5–3.12.

Таблица 3.5 – Показатели диагностической точности ИИ-1, ИИ-5, ИИ-6 для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при проспективном наблюдении в 2022 году

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.
ИИ-1 (РГ)	AUROC	0,687	0,783	0,793	0,806
	95%-й ДИ	0,613–0,759	0,727–0,839	0,732–0,852	0,741–0,872
	Точность	0,625	0,658	0,679	0,733
	95%-й ДИ	0,564–0,686	0,598–0,718	0,620–0,738	0,677–0,789
	Чувствительность	0,833	0,932	0,929	0,88
	95%-й ДИ	0,734–0,933	0,857–1,000	0,833–1,000	0,790–0,970
	Специфичность	0,565	0,597	0,646	0,695
	95%-й ДИ	0,493–0,636	0,528–0,666	0,582–0,711	0,629–0,760
ИИ-5 (РГ)	AUROC	–	0,795	0,703	0,789
	95%-й ДИ	–	0,728–0,864	0,639–0,767	0,728–0,850
	Точность	–	0,762	0,65	0,762
	95%-й ДИ	–	0,697–0,828	0,590–0,710	0,709–0,816
	Чувствительность	–	0,908	0,9	0,781
	95%-й ДИ	–	0,837–0,978	0,841–0,959	0,710–0,853
	Специфичность	–	0,663	0,471	0,741
	95%-й ДИ	–	0,568–0,758	0,389–0,554	0,660–0,822
ИИ-6 (РГ)	AUROC	–	0,799	0,809	0,788
	95%-й ДИ	–	0,734–0,867	0,758–0,863	0,730–0,847
	Точность	–	0,75	0,746	0,665

Продолжение таблицы 3.5

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.
	95%-й ДИ	–	0,683–0,817	0,691–0,801	0,606–0,725
	Чувствительность	–	0,774	0,889	0,927
	95%-й ДИ	–	0,670–0,878	0,82–0,957	0,859–0,996
	Специфичность	–	0,735	0,673	0,588
	95%-й ДИ	–	0,647–0,822	0,60–0,746	0,518–0,659

Таблица 3.6 – Показатели диагностической точности ИИ-1, ИИ-5, ИИ-6 для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при проспективном наблюдении в 2023 году

ИИ-сервис	Показатели	Период		
		I кв. 2023 г.	II кв. 2023 г.	III кв. 2023 г.
ИИ-1 (РГ)	AUROC	0,924	0,872	0,933
	95%-й ДИ	0,886–0,963	0,825–0,920	0,894–0,973
	Точность	0,876	0,825	0,85
	95%-й ДИ	0,834–0,917	0,777–0,873	0,795–0,905
	Чувствительность	0,975	0,908	0,958
	95%-й ДИ	0,940–1,000	0,843–0,973	0,902–1,000
	Специфичность	0,827	0,787	0,804
	95%-й ДИ	0,769–0,885	0,724–0,849	0,730–0,877
ИИ-5 (РГ)	AUROC	0,566	0,938	0,918
	95%-й ДИ	0,489–0,636	0,899–0,977	0,858–0,980
	Точность	0,567	0,838	0,875
	95%-й ДИ	0,504–0,629	0,78–0,895	0,803–0,947
	Чувствительность	0,911	0,906	0,846
	95%-й ДИ	0,836–0,985	0,805–1,0	0,707–0,985
	Специфичность	0,462	0,82	0,889
	95%-й ДИ	0,390–0,534	0,754–0,887	0,805–0,973
ИИ-6 (РГ)	AUROC	0,836	0,74	–
	95%-й ДИ	0,782–0,889	0,677–0,802	–
	Точность	0,8	0,717	–

Продолжение таблицы 3.6

ИИ-сервис	Показатели	Период		
		I кв. 2023 г.	II кв. 2023 г.	III кв. 2023 г.
	95%-й ДИ	0,749–0,851	0,660–0,774	–
	Чувствительность	0,791	0,978	–
	95%-й ДИ	0,705–0,877	0,949–1,0	–
	Специфичность	0,805	0,551	–
	95%-й ДИ	0,743–0,868	0,471–0,631	–

Таблица 3.7 – Показатели диагностической точности ИИ-2 для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при проспективном наблюдении в 2022 году

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.
ИИ-2 (РГ)	AUROC	0,856	0,688	0,806	0,778
	95%-й ДИ	0,763–0,949	0,615–0,755	0,757–0,854	0,728–0,829
	Точность	0,75	0,571	0,696	0,742
	95%-й ДИ	0,655–0,845	0,508–0,633	0,638–0,754	0,686–0,797
	Чувствительность	0,889	0,87	0,919	0,856
	95%-й ДИ	0,770–1,000	0,781–0,960	0,852–0,987	0,783–0,928
	Специфичность	0,679	0,484	0,618	0,673
	95%-й ДИ	0,554–0,805	0,412–0,556	0,547–0,689	0,598–0,748
ИИ-2 (ФЛГ)	AUROC	0,776	0,663	0,798	0,79
	95%-й ДИ	0,717–0,835	0,591–0,736	0,745–0,851	0,722–0,857
	Точность	0,654	0,504	0,625	0,55
	95%-й ДИ	0,594–0,714	0,441–0,567	0,564–0,686	0,487–0,613
	Чувствительность	0,957	0,955	1	1
	95%-й ДИ	0,900–1,000	0,906–1,000	1,000–1,000	1,000–1,000
	Специфичность	0,58	0,329	0,505	0,502
	95%-й ДИ	0,511–0,650	0,259–0,400	0,433–0,578	0,436–0,569

Таблица 3.8 – Показатели диагностической точности ИИ-2 для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при проспективном наблюдении в 2023 году

ИИ-сервис	Показатели	Период		
		I кв. 2023 г.	II кв. 2023 г.	III кв. 2023 г.
ИИ-2 (РГ)	AUROC	0,844	0,977	0,864
	95%-й ДИ	0,796–0,891	0,960–0,996	0,806–0,921
	Точность	0,767	0,938	0,844
	95%-й ДИ	0,713–0,820	0,900–0,975	0,787–0,900
	Чувствительность	0,842	0,989	0,957
	95%-й ДИ	0,747–0,937	0,968–1,000	0,916–0,998
	Специфичность	0,743	0,866	0,687
	95%-й ДИ	0,680–0,806	0,784–0,947	0,575–0,798
ИИ-2 (ФЛГ)	AUROC	0,843	0,937	0,775
	95%-й ДИ	0,796–0,891	0,904–0,971	0,694–0,856
	Точность	0,704	0,912	0,825
	95%-й ДИ	0,646–0,762	0,877–0,948	0,766–0,884
	Чувствительность	0,971	0,974	0,921
	95%-й ДИ	0,932–1,000	0,949–0,999	0,868–0,973
	Специфичность	0,594	0,805	0,661
	95%-й ДИ	0,520–0,668	0,721–0,888	0,540–0,782

Таблица 3.9 – Показатели диагностической точности ИИ-3 для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при проспективном наблюдении в 2022 году

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.
ИИ-3 (РГ)	AUROC	0,824	0,781	0,772	0,844
	95%-й ДИ	0,755–0,893	0,715–0,850	0,710–0,832	0,784–0,905
	Точность	0,76	0,758	0,725	0,781
	95%-й ДИ	0,701–0,819	0,704–0,812	0,669–0,781	0,717–0,845
	Чувствительность	0,787	0,611	0,767	0,848
	95%-й ДИ	0,670–0,904	0,499–0,724	0,679–0,854	0,762–0,935
	Специфичность	0,752	0,821	0,7	0,734
	95%-й ДИ	0,683–0,820	0,764–0,879	0,627–0,773	0,645–0,823

Продолжение таблицы 3.9

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.
ИИ-3 (ФЛГ)	AUROC	0,798	0,798	0,841	0,806
	95%-й ДИ	0,715–0,882	0,738–0,857	0,785–0,897	0,731–0,883
	Точность	0,683	0,658	0,792	0,762
	95%-й ДИ	0,624–0,742	0,598–0,718	0,740–0,843	0,697–0,828
	Чувствительность	0,857	0,95	0,765	0,735
	95%-й ДИ	0,728–0,987	0,882–1,000	0,648–0,881	0,611–0,858
	Специфичность	0,66	0,6	0,799	0,775
	95%-й ДИ	0,597–0,724	0,532–0,668	0,742–0,856	0,697–0,852

Таблица 3.10 – Показатели диагностической точности ИИ-3 для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при проспективном наблюдении в 2023 году

ИИ-сервис	Показатели	Период		
		I кв. 2023 г.	II кв. 2023 г.	III кв. 2023 г.
ИИ-3 (РГ)	AUROC	0,84	0,807	0,939
	95%-й ДИ	0,791–0,887	0,754–0,862	0,902–0,974
	Точность	0,754	0,738	0,881
	95%-й ДИ	0,700–0,809	0,682–0,793	0,831–0,931
	Чувствительность	0,926	0,798	0,988
	95%-й ДИ	0,877–0,975	0,725–0,872	0,965–1,000
	Специфичность	0,614	0,683	0,76
	95%-й ДИ	0,531–0,697	0,601–0,764	0,663–0,857
ИИ-3 (ФЛГ)	AUROC	0,833	0,843	0,679
	95%-й ДИ	0,776–0,887	0,789–0,896	0,594–0,767
	Точность	0,767	0,808	0,656
	95%-й ДИ	0,713–0,820	0,759–0,858	0,583–0,730
	Чувствительность	0,859	0,794	0,892
	95%-й ДИ	0,785–0,933	0,728–0,861	0,817–0,968
	Специфичность	0,716	0,828	0,495
	95%-й ДИ	0,645–0,787	0,754–0,903	0,394–0,595

Таблица 3.11 – Показатели диагностической точности ИИ-4 для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при проспективном наблюдении в 2022 году

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.
ИИ-4 (РГ)	AUROC	0,807	0,869	0,886	0,839
	95%-й ДИ	0,744–0,870	0,822–0,916	0,842–0,930	0,789–0,887
	Точность	0,742	0,771	0,858	0,787
	95%-й ДИ	0,686–0,797	0,718–0,824	0,814–0,902	0,736–0,839
	Чувствительность	0,821	0,905	0,761	0,892
	95%-й ДИ	0,721–0,922	0,832–0,977	0,659–0,863	0,834–0,950
	Специфичность	0,717	0,723	0,896	0,698
	95%-й ДИ	0,652–0,782	0,657–0,789	0,850–0,941	0,618–0,777
ИИ-4 (ФЛГ)	AUROC	0,77	0,718	0,737	0,855
	95%-й ДИ	0,671–0,873	0,643–0,795	0,672–0,803	0,807–0,901
	Точность	0,675	0,637	0,546	0,796
	95%-й ДИ	0,572–0,778	0,577–0,698	0,483–0,609	0,745–0,847
	Чувствительность	0,905	0,75	1	0,838
	95%-й ДИ	0,779–1,000	0,622–0,878	1,000–1,000	0,768–0,909
	Специфичность	0,593	0,612	0,414	0,763
	95%-й ДИ	0,468–0,719	0,544–0,680	0,343–0,485	0,691–0,835

Таблица 3.12 – Показатели диагностической точности ИИ-4 для анализа результатов РГ/ФЛГ органов грудной клетки, полученные при проспективном наблюдении в 2023 году

ИИ-сервис	Показатели	Период		
		I кв. 2023 г.	II кв. 2023 г.	III кв. 2023 г.
ИИ-4 (РГ)	AUROC	0,856	0,739	0,766
	95%-й ДИ	0,804–0,910	0,677–0,802	0,694–0,835
	Точность	0,863	0,713	0,75
	95%-й ДИ	0,819–0,906	0,655–0,770	0,683–0,817
	Чувствительность	0,974	0,952	0,961
	95%-й ДИ	0,948–0,999	0,915–0,989	0,918–1,000
	Специфичность	0,674	0,452	0,554
	95%-й ДИ	0,577–0,772	0,361–0,543	0,447–0,661

Продолжение таблицы 3.12

ИИ-сервис	Показатели	Период		
		I кв. 2023 г.	II кв. 2023 г.	III кв. 2023 г.
ИИ-4 (ФЛГ)	AUROC	0,901	0,785	0,725
	95%-й ДИ	0,858–0,944	0,730–0,840	0,647–0,804
	Точность	0,871	0,75	0,694
	95%-й ДИ	0,828–0,913	0,695–0,805	0,622–0,765
	Чувствительность	0,983	0,94	0,906
	95%-й ДИ	0,959–1,000	0,896–0,983	0,835–0,978
	Специфичность	0,764	0,573	0,552
	95%-й ДИ	0,689–0,839	0,486–0,660	0,453–0,652

В первом временном периоде, в условиях работы с реальными клиническими данными ИИ-1 для анализа результатов РГ органов грудной клетки продемонстрировал выраженное ухудшение точности: показатель площади под характеристической кривой снизился на 17,2 %, точности – на 19,9 %, специфичности – на целых 27,6 % относительно уровня, полученного при ретроспективной оценке. Лишь для чувствительности отмечен небольшой рост – на 6,8 %. Различия между значениями AUROC при ретроспективном и первом проспективном наблюдениях носили статистически значимый характер (z-критерий Фишера 2,9662, $p = 0,0030$). Точно так же различия достигли статистической значимости для точности (t-критерий $-27,883$, $p < 0,0001$) и специфичности (t-критерий $-37,264$, $p < 0,0001$).

В последующие периоды, очевидно на фоне интенсивной технической доработки, ИИ-1 продемонстрировал медленный, но в целом уверенный рост показателей точности. К последнему периоду наблюдения AUROC достигла 0,933 (95%-й ДИ 0,894–0,973), теперь уже статистически значимо превысив полученный при ретроспективной оценке уровень на 12,4 % (z-критерий Фишера $-2,8104$, $p = 0,0049$).

Аналогичная ситуация отмечена для точности и специфичности: прирост относительно исходного уровня составил 9,0 % (t-критерий 13,081, $p < 0,0001$) и

3,1 % (t-критерий 4,102, $p = 0,0001$). Особенно значителен был рост чувствительности – он достиг 22,8 % (t-критерий 34,511, $p < 0,0001$).

Отличием от ретроспективного этапа стал выраженный рост удельного веса случаев гипердиагностики ИИ-1. Этот показатель был стабильно высоким, колебался в пределах 25,7–72,8 % (в среднем – $46,9 \pm 18,3$ %). Неустойчивой была ситуация с пропуском патологии. В трех периодах наблюдений таких случаев не было вовсе. Однократно удельный вес ложноотрицательных случаев достиг 4,1 %. В прочих периодах он колебался от 1,7 до 2,4 %.

Характерной негативной чертой стал низкий процент согласий врачей-экспертов с решениями ИИ-1 при оценке его клинической результативности. В первые два периода наблюдения полное согласие отмечено всего лишь в 3,0 и 1,2 % случаев соответственно. Затем ситуация несколько улучшилась, однако процент согласий все равно не превысил 40,0. Высок был удельный вес случаев несогласий с локализацией и маркировкой обнаруженных патологических признаков: от 20,1 до 42,9 % (соответствующие расхождения определялись в каждом периоде).

В условиях работы с реальными клиническими данными ИИ-2 продемонстрировал стабильность точности при анализе результатов рентгенографии (отсутствовали достоверные различия значений AUROC, $p = 0,4358$), но ее снижение при анализе результатов флюорографии (AUROC значимо опустилась – на 13,8 %, z-критерий Фишера 2,8882, $p = 0,0039$). К последнему периоду наблюдений AUROC для рентгенографии оставался немного ниже исходного уровня (на 4,0 %), но статистической значимости эти различия по-прежнему не достигали ($p = 0,3958$). Для флюорографии ситуация осталась стабильно негативной. «Разрыв» сохранился на прежнем уровне и с прежним уровнем значимости.

Особый интерес представляет динамика чувствительности и специфичности. Для обеих модальностей при переходе к работе с потоком данных зафиксирован устойчивый рост чувствительности. В частности, в первом периоде наблюдений он составил для рентгенографии 11,1 % (t-критерий 9,872,

$p < 0,0001$), для флюорографии целых 19,6 % (t-критерий 21,483, $p < 0,0001$). В двух периодах зафиксирован уровень чувствительности при анализе результатов ФЛГ в 1,0 (95%-й ДИ 1,0–1,0). К концу наблюдений высокий уровень данного показателя устойчиво наблюдался для обеих модальностей. Различия с исходным значением, определенным при ретроспективном анализе, были достоверны и составляли для РГ плюс 19,6 % (t-критерий 21,925, $p < 0,0001$), для ФЛГ плюс 15,1 % (t-критерий 16,455, $p < 0,0001$). Ситуация со специфичностью была противоположной. Для обеих модальностей во все периоды наблюдения сохранялся значимый, иногда даже критичный «разрыв» с исходным значением. Для РГ в первом периоде он составил 26,2 %, в последнем – 25,3 % (t-критерий – 30,013 и –31,177, $p < 0,0001$). Для ФЛГ – 37,0 % и 28,2 % соответственно (t-критерий –56,595 и –33,001, $p < 0,0001$).

С учетом такого дисбаланса показателей ожидаемо высоким был удельный вес гипердиагностики. С первого по пятый период наблюдений для рентгенографии он составлял от 35,7 до 68,2 %, для флюорографии – от 26,2 до 76,2 %. Вместе с тем ИИ-2 постоянно подвергался технической доработке и совершенствованию моделей ИИ. Благодаря этому в два заключительных периода уровень гипердиагностики резко снизился до 7,2–9,4 % для РГ и 4,3–7,5 % для ФЛГ. Обращает на себя внимание небольшой процент пропусков патологии. В ряде периодов таковых случаев не было совсем. В среднем удельный вес ложноотрицательных результатов колебался в пределах 1,0–2,0 %. Однократно встретились своеобразные «выбросы»: для РГ – до 3,4 % в первом периоде, для ФЛГ – до 3,6 % в последнем. Пропорционально низким был удельный вес случаев полного соответствия (вплоть до 7,0 %). Лишь в двух последних периодах наблюдений он возрос до 82,5–87,1 %.

Очень схожую динамику продемонстрировал ИИ-3. При переходе к работе с потоком реальных клинических данных показатель чувствительности анализа результатов рентгенографии уменьшился на 10,6 % (t-критерий –13,239, $p < 0,0001$). Затем, после некоторой нестабильности, начал расти и в последнем периоде значимо превысил исходный уровень на 12,3 % (t-критерий 28,305,

$p < 0,0001$).

Чувствительность для анализа результатов ФЛГ продемонстрировала стабильность: снижение в первом периоде, хотя и было значимым, но составило лишь 2,6 % (t-критерий $-3,037$, $p = 0,0027$). К последнему периоду оно немного превышало исходный уровень (на 1,4 %, t-критерий $2,245$, $p = 0,0260$).

На этом фоне специфичность ИИ-3 критично упала, как для рентгенографии, так и для флюорографии. В первом периоде снижение составило для РГ 15,6 %, а для ФЛГ – целых 29,0 % (t-критерий $-37,906$ и $-59,874$ соответственно, $p < 0,0001$). Далее этот показатель оставался стабильно низким. К последнему периоду наблюдения для флюорографии он и вовсе достиг критично низкого уровня, опустившись на 46,8 % (t-критерий $-72,391$, $p < 0,0001$). Для рентгенографии «разрыв» был меньше – 18,3 %, но все равно оставался значительным и значимым (t-критерий $-29,022$, $p < 0,0001$).

Отметим, что площадь под характеристической кривой для ИИ-3 при анализе результатов РГ снизилась в первом периоде на 13,3 % (t-критерий $3,2996$, $p = 0,0010$), однако затем постепенно возросла и практически достигла исходного значения (при этом различия в 1,2 % стали незначимыми, $p = 0,6409$). Напротив, для флюорографии снижение AUROC к концу наблюдений составило целых 28,5 % (t-критерий $5,8222$, $p < 0,0001$) от исходного значения.

При ретроспективном анализе ИИ-3 отличался высоким уровнем гипо- и гипердиагностики. При проспективном формате работы отмечена столь же негативная ситуация с гипердиагностикой. Для рентгенографии удельный вес ложноположительных результатов составлял 13,7–67,6 % (в среднем $32,1 \pm 18,7$ %, медиана 24,8 %); для ФЛГ лишь в последнем периоде наблюдения получено минимальное значение в 3,2 %, в остальных – гипердиагностика составляла от 10,2 до 73,0 % (в среднем $39,3 \pm 21,2$ %, медиана 39,0 %). Фактически в каждом периоде отмечались случаи пропуска патологии. Однако, в отличие от ретроспективного тестирования, для рентгенографии их удельный вес был минимален: от 0,6 до 1,3 %. Но для флюорографии проблема ложноотрицательных результатов сохранялась. Их удельный вес варьировался от

0,7 до 13,8 %, в среднем составляя $3,6 \pm 4,8$ % (медиана – 2,3 %).

Аналогичная динамика зафиксирована также для ИИ-4: резкое падение AUROC и прочих показателей в первом периоде, более выраженное для анализа рентгенографии, чем ФЛГ. Для РГ снижение площади под характеристической кривой, точности и чувствительности было примерно одинаковым – около 11,0–12,0 %, а специфичности было меньшим – 8,7 %. Все различия были статистически значимыми ($p < 0,0001$). Для ФЛГ AUROC снизилась на 16,0 %, точность на 20,6 %, а специфичность – на целых 26,8 %. Различия также были статистически значимыми ($p < 0,0001$). Вместе с тем чувствительность практически осталась на прежнем уровне (отмечены колебания около 0,6 %, $p = 0,4492$). Далее ИИ-4 продемонстрировал общее низкое качество, так как к последнему периоду наблюдений «разрыв» значений показателей в проспективном формате работы и исходного уровня, полученного при ретроспективном тестировании, не только сохранился, но и усугубился. Например, для РГ значение специфичности снизилось на 31,6 %, а для ФЛГ – на 31,9 %. На этом фоне лишь показатель чувствительности оказался стабильным. Для флюорографии он в целом удержался на уровне первого периода ($p = 0,2319$), а для рентгенографии даже немного увеличился – на 6,8 % (t-критерий 15,204, $p < 0,0001$). С точки зрения проспективной оценки клинической результативности у ИИ-4 отмечались те же проблемы, что и у ИИ-2 и ИИ-3: постоянный, значительный процент гипердиагностики (для рентгенографии – 16,4–57,7 %, для флюорографии – 14,7–63,5 %). Уровень пропусков патологии был явно ниже при анализе результатов РГ (в двух периодах – по 0,6 %, в одном – 3,3 %, в остальных – ноль), чем результатов ФЛГ (только один период был «нулевым», в одном отмечен критичный «выброс» до 29,5 %, в остальных – удельный вес ложноотрицательных ответов колебался от 0,6 до 2,9 %).

ИИ-5 продемонстрировал некоторое снижение точности в первом периоде наблюдений, которое, впрочем, не достигало статистической значимости (для AUROC $p = 0,2082$). В последующем «разрыв» для всех показателей усугубился. На фоне технологической доработки в двух последних периодах наблюдений

значения показателей практически сравнялись с исходными, установленными при ретроспективном тестировании. Так, точность в первом периоде снизилась на 11,4 % (t-критерий $-18,051$, $p < 0,0001$), во втором – на 24,4 % (t-критерий $-39,914$, $p < 0,0001$), а в последнем – на 1,7 % (t-критерий $2,660$, $p < 0,0085$) превысила исходное значение. Аналогичная «траектория» у специфичности. На этом фоне чувствительность при переходе к проспективному формату работы возросла на 9,4 % (t-критерий $10,423$, $p < 0,0001$). Затем последовали колебания в довольно выраженном диапазоне от 0,781 до 0,911. В последнем периоде зафиксирован уровень 0,846, на 1,9 % превышающий исходный ($p = 0,0994$).

Динамика ИИ-6 в целом совпадала с тенденциями ИИ-3 при анализе результатов ФЛГ. С первого же периода наблюдений началось выраженное снижение AUROC, точности и чувствительности, достигшее к последнему периоду 21,3 %, 24,5 % и 43,8 % соответственно (t-критерий $4,7273$, $-60,076$, $-96,980$, $p < 0,0001$).

Чувствительность, после значимого падения в первом периоде на 15,0 % (t-критерий $-19,436$, $p < 0,0001$), продемонстрировала в последнем периоде наблюдения также статистически достоверное превышение исходного уровня на 7,5 % (t-критерий $14,323$, $p < 0,0001$). Вместе с тем необходимо отметить неустойчивость значения данного показателя, так как в другие периоды наблюдения отмечался его разброс от 0,774 до 0,927.

Оба сервиса – ИИ-5 и ИИ-6 – имели стабильный и значительный уровень случаев гипердиагностики: 20,8–64,0 % у ИИ-5, 25,3–40,4 % у ИИ-6 (без «нулевых» периодов). Встречаемость случаев пропуска патологии у сервисов различалась и довольно сильно варьировалась. У ИИ-5 было два «нулевых» периода, один период с «выбросом» до 11,7 %, в остальных – удельный вес ложноотрицательных случаев составил 3,0–6,3 %. У ИИ-6 «нулевой» период был только один, в остальных – уровень колебался от 0,7 до 5,0 %.

Таким образом, для ИИ-сервисов, осуществляющих анализ результатов РГ и ФЛГ (как отдельно, так и одновременно), при работе с реальным потоком данных отмечается значительное снижение показателей точности, критично

влияющее на клиническую значимость их применения. С другой стороны, на фоне статистически значимого падения специфичности типичным является стабильность или даже значимый рост чувствительности (вплоть до 1,0 95%-й ДИ 1,0–1,0). Это дает основание предложить применять такие ИИ-сервисы для сортировки результатов рентгенологических исследований органов грудной клетки, в том числе в контексте модели «автономного искусственного интеллекта» (данный вопрос подробно будет изучаться далее).

Рентгенография стопы. В контексте автоматизированного анализа результатов рентгенографии стопы с целью диагностики продольного плоскостопия сформулирована следующая клиническая задача – выявление на боковой проекции наличия как минимум одного рентгенологического признака, коррелирующего с наличием продольного плоскостопия:

- определение значения угла свода стопы степени I–III в соответствии с принятой классификацией;
- определение значения высоты свода стопы степени I–III в соответствии с принятой классификацией;
- графическое представление оценки угла и высоты свода стопы;
- наличие информации о количественной оценке угла свода стопы и высоты свода стопы с определением степени изменений.

Ретроспективное исследование диагностической точности одного² ИИ-сервиса для автоматизированного анализа рентгенографии стопы с целью диагностики продольного плоскостопия выполнено на эталонном наборе данных (n = 100).

Построена характеристическая кривая (рисунок 3.3). Значение площади под характеристической кривой составило 0,93 (95%-й ДИ 0,88–0,98). Определены показатели: общая точность – 0,93 (95%-й ДИ 0,88–0,98), чувствительность – 0,98 (95%-й ДИ 0,93–1,0), специфичность – 0,89 (95%-й ДИ 0,79–0,98). Удельный вес

² ООО «КэреМенторЭйАй», Система нейросетевая CareMentor AI для определения продольного плоскостопия по данным боковой рентгенография стопы под нагрузкой.

ложноотрицательных результатов оказался относительно небольшим и составил 2,0 %; уровень ложноположительных был умеренным (11,0 %).

Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.4).

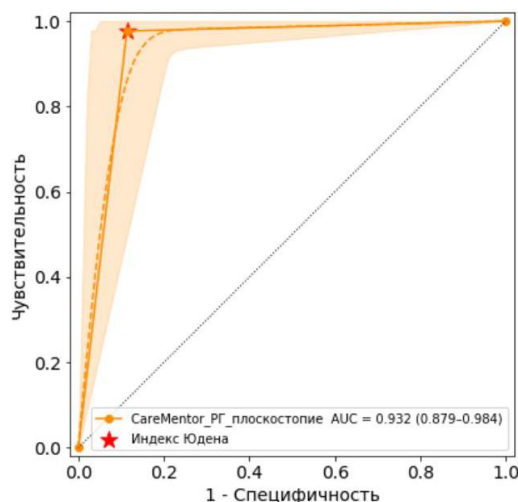


Рисунок 3.3 – Характеристическая кривая для ИИ-сервиса для анализа результатов рентгенографии стопы с целью диагностики продольного плоскостопия (оптимальный порог по индексу Юдена)

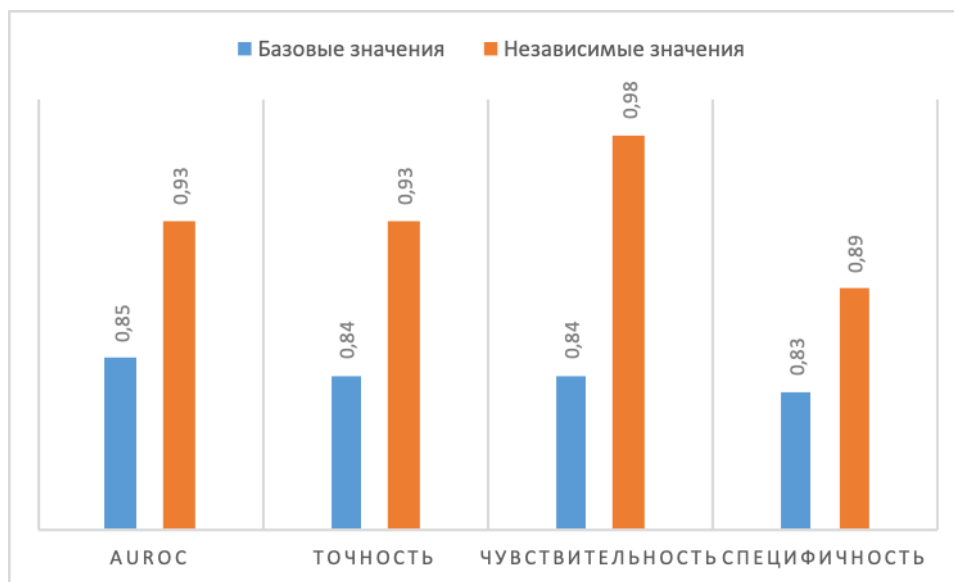


Рисунок 3.4 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов рентгенографии стопы с целью диагностики продольного плоскостопия

Обращает на себя внимание увеличение значений всех показателей

точности, продемонстрированное ИИ-сервисом при тестировании на новых данных (что само по себе является нечастым явлением).

Прирост AUROC составил 9,4 %, точности – 10,7 %. Наиболее значительно увеличилась чувствительность – на 16,7 %, менее всего – специфичность (только на 7,2 %).

Успешное превышение порогового значения AUROC для данного вида исследований обеспечило переход ИИ-сервиса к работе с реальными клиническими данными и, соответственно, ко второму этапу оценки диагностической точности.

Проспективное исследование выполнено в условиях работы ИИ-сервиса с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях. В соответствии с методологией, проведены четыре измерения точности в проспективном формате; результаты представлены в таблице 3.13.

Таблица 3.13 – Показатели диагностической точности ИИ-сервиса для анализа результатов рентгенографии стопы с целью диагностики продольного плоскостопия, полученные при проспективном наблюдении

Показатели	Период			
	III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
AUROC	0,817	0,833	0,657	0,816
95%-й ДИ	0,742–0,891	0,776–0,889	0,611–0,704	0,758–0,872
Точность	0,887	0,892	0,642	0,887
95%-й ДИ	0,839–0,936	0,852–0,931	0,581–0,702	0,848–0,927
Чувствительность	0,974	0,971	0,956	0,988
95%-й ДИ	0,945–1,000	0,946–0,996	0,919–0,994	0,972–1,000
Специфичность	0,659	0,696	0,357	0,643
95%-й ДИ	0,519–0,799	0,587–0,804	0,273–0,441	0,531–0,755

В первом временном периоде, в условиях работы с реальными клиническими данными ИИ-сервис для анализа результатов рентгенографии стопы с целью диагностики продольного плоскостопия продемонстрировал

ухудшение точности: показатель площади под характеристической кривой снизился на 12,2 %, точности – на 4,6 %, специфичности – на целых 26,0 % относительно уровня, полученного при ретроспективной оценке. Минимальным было снижение чувствительности (0,6 %).

Различия между значениями AUROC при ретроспективном и первом проспективном наблюдениях носили статистически значимый характер (z-критерий Фишера 2,4685, $p = 0,0136$). Также достоверным было снижение значений общей точности (t-критерий $-11,388$, $p < 0,0001$), чувствительности (t-критерий $-2,458$, $p = 0,0149$) и специфичности (t-критерий $-25,771$, $p < 0,0001$).

Такая ситуация сохранилась в следующем периоде, а в третьем периоде наблюдения произошло выраженное падение точности, в частности AUROC на 29,4 %. После технической доработки восстановился прежний уровень точности. Вместе с тем он был ниже, чем выявленный при ретроспективном анализе. Площадь под характеристической кривой была меньше исходной на 12,3 % (z-критерий Фишера 2,9469, $p = 0,0032$), специфичность – на целых 27,8 % (t-критерий $-31,367$, $p < 0,0001$); наименьший «разрыв» был у точности – лишь 4,6 % (t-критерий $-12,311$, $p < 0,0001$). На этом фоне динамика чувствительности была положительной, отмечался прирост, хоть и минимальный, до уровня 0,988 (95%-й ДИ 0,972–1,000).

При проспективной оценке клинической результативности установлено, что удельный вес случаев с полным соответствием решений врачей-экспертов и ИИ-сервиса колебался в пределах 57,0–97,0 % (причем в трех из четырех периодов он был выше 90,0 %).

В одном периоде зафиксирован критично высокий уровень случаев с некорректной оценкой локализации – 43,0 %. В остальных случаях он составлял лишь 3,0–6,0 %. Очень важно подчеркнуть, что случаев гипердиагностики и пропусков патологии вовсе не зафиксировано.

Таким образом, при автоматизированном анализе результатов рентгенографии стопы с целью диагностики продольного плоскостопия потенциально достижим уровень точности, близкий к точности врача-

рентгенолога (на потоке данных AUROC может превышать 0,83, чувствительность высока и в пределах доверительного интервала тяготеет к 1,0, случаи гипо- и гипердиагностики отсутствуют).

Полученные данные позволяют говорить о потенциальном результате. Технологии ИИ могут применяться для автоматизированного анализа результатов рентгенографии стопы с целью диагностики продольного плоскостопия. В контексте данной клинической задачи ТИИ могут использоваться для морфометрии (измерения угла и высоты свода стопы).

На этом этапе нами выявлено недостаточно высокое качество конкретного прикладного программного обеспечения на основе технологий ИИ. Вместе с тем сама по себе задача морфометрии стопы с целью диагностики продольного плоскостопия является крайне тривиальной, высокостандартизированной и дискретной. Решение именно таких задач необходимо максимально автоматизировать для высвобождения рабочего времени врача-рентгенолога для решения более когнитивно сложных задач. Поэтому проблема автоматизированного анализа результатов рентгенографии стопы с целью диагностики продольного плоскостопия рассмотрена нами более детально далее, при изучении автономного искусственного интеллекта.

Рентгенография позвоночника. В контексте автоматизированного анализа результатов РГ позвоночника с целью диагностики сколиоза была сформулирована следующая клиническая задача – выявление в прямой проекции наличия рентгенологических признаков, коррелирующих с наличием сколиоза:

- наличие угловой деформации оси позвоночника (I степень – 1–10 градусов, II степень – 11–25 градусов, III степень – 26–50 градусов, IV степень – более 50 градусов);
- локализация находок;
- количественная оценка углов деформации оси позвоночника (абсолютные значения и/или степень сколиоза);

– направление сколиотической дуги (при наличии двух разнонаправленных С-сколиозов – формирование заключения «S-сколиоз» с измерением угловой деформации вправо и влево).

Первым этапом ретроспективное исследование диагностической точности двух³ ИИ-сервисов (далее в этом разделе – ИИ-1 и ИИ-2) для автоматизированного анализа результатов РГ позвоночника с целью выявления сколиоза и морфометрии углов деформации выполнено на эталонном наборе данных (n = 100).

Построены характеристические кривые (рисунок 3.5).

Для ИИ-1 значение AUROC составило 0,819 (95%-й ДИ 0,741–0,897); этот и прочие показатели точности были ниже установленных пороговых значений, что потребовало технологической доработки модели ИИ, положенной в основу изучаемого сервиса. После соответствующего улучшения проведено повторное тестирование. Определены показатели: площадь под характеристической кривой – 0,912 (95%-й ДИ 0,855–0,970), общая точность – 0,900 (95%-й ДИ 0,840–0,960), чувствительность – 0,880 (95%-й ДИ 0,790–0,970), специфичность – 0,910 (95%-й ДИ 0,840–0,990). Удельный вес ложноотрицательных случаев (гиподиагностика) составил 12,0 %, ложноположительных (гипердиагностика) – 8,5 %.

Для ИИ-2 значение площади под характеристической кривой при первом же тестировании составило 0,855 (95%-й ДИ 0,781–0,928). Определены показатели: общая точность – 0,850 (95%-й ДИ 0,78–0,93), чувствительность – 0,932 (95%-й ДИ 0,889–0,989), специфичность – 0,880 (95%-й ДИ 0,79–0,98). Показатели превышали пороговые значения, ИИ-2 был сразу допущен к работе в реальных клинических условиях. Удельный вес случаев гиподиагностики составил 17,4 %, гипердиагностики – 11,6 %.

Для обоих ИИ-сервисов полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.6).

³ 1) ООО «Оксиджен Технолджиес Рус». Oxytech Spine XR Scoliosis. 2) ООО «Синапс Тех». Синапс Нейро.

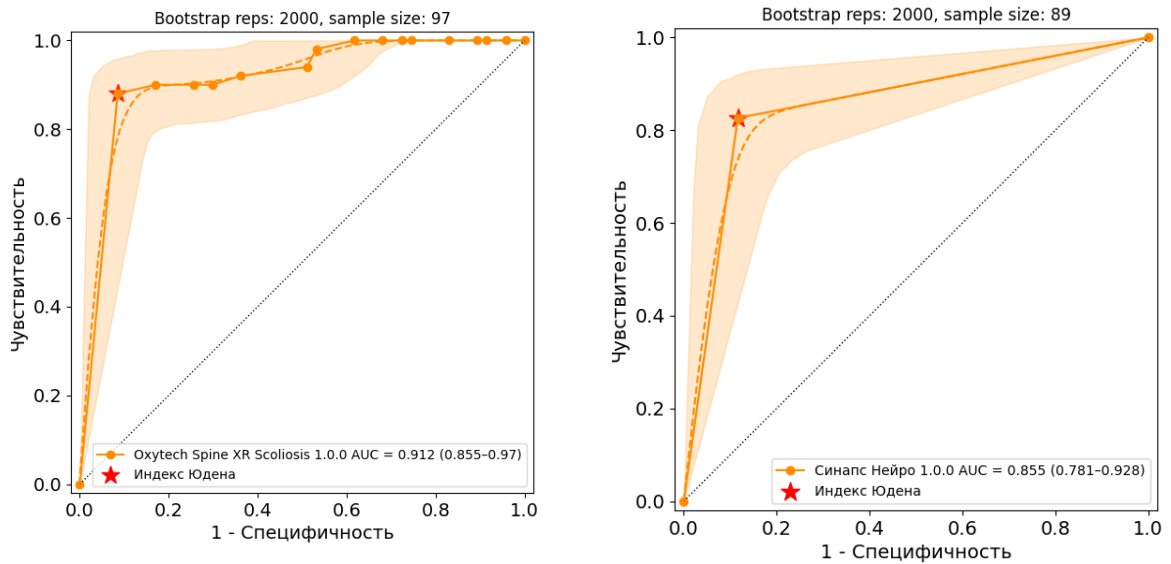


Рисунок 3.5 – Характеристическая кривая для ИИ-сервиса для анализа результатов РГ позвоночника (оптимальный порог по индексу Юдена)

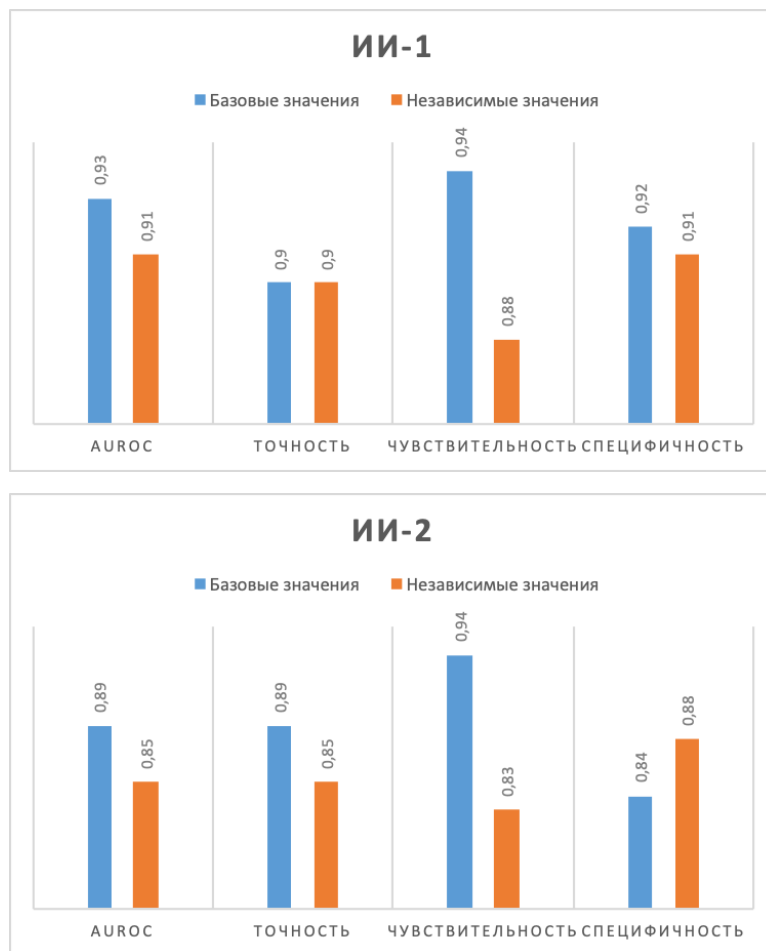


Рисунок 3.6 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов РГ позвоночника

При ретроспективном исследовании ИИ-2 продемонстрировал худшие результаты по гипер- и, особенно, гиподиагностике. Также у него снижение показателей точности относительно исходных было более выраженным: AUROC и точность уменьшились на 4,5 % (у ИИ-1 – на 2,2 %, точность не изменилась); чувствительность – на 11,7 % (у ИИ-1 – на 6,3 %). Единственным положительным отличием можно считать некоторый прирост специфичности – на 4,8 %. Однако полученные значения не пересекли установленный порог, что позволило перейти ко второму этапу.

Вторым этапом выполнено проспективное исследование в условиях работы обоих ИИ-сервисов с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены по два измерения точности в проспективном формате для каждого ИИ-сервиса; результаты представлены в таблице 3.14.

Таблица 3.14 – Показатели диагностической точности ИИ-сервиса для анализа результатов РГ позвоночника, полученные при проспективном наблюдении

ИИ-сервис	Показатели	Период	
		I кв. 2023 г.	II кв. 2023 г.
ИИ-1	AUROC	0,369	0,488
	95%-й ДИ	0,164–0,567	0,353–0,624
	Точность	0,906	0,917
	95%-й ДИ	0,861–0,951	0,882–0,952
	Чувствительность	0,941	0,961
	95%-й ДИ	0,904–0,978	0,935–0,986
	Специфичность	0,143	0,083
	95%-й ДИ	0,000–0,402	0,000–0,240

Продолжение таблицы 3.14

ИИ-сервис	Показатели	Период	
		I кв. 2023 г.	II кв. 2023 г.
ИИ-2	AUROC	0,833	0,827
	95%-й ДИ	0,742–0,922	0,761–0,892
	Точность	0,863	0,842
	95%-й ДИ	0,787–0,938	0,795–0,888
	Чувствительность	0,941	0,849
	95%-й ДИ	0,877–1,000	0,800–0,899
	Специфичность	0,724	0,805
	95%-й ДИ	0,561–0,887	0,684–0,926

В первом временном периоде, в условиях работы с реальными клиническими данными ИИ-1 для анализа результатов РГ позвоночника продемонстрировал критическое, статистически значимое ($p < 0,0001$) ухудшение точности: показатель площади под характеристической кривой снизился на 59,5 %, специфичности – на целых 84,3 % относительно значений, полученных при ретроспективной оценке. Сохранилась высокая чувствительность, но это означало лишь то, что подавляющее число решений ИИ-1 было ложноположительными. ИИ-1 был отправлен на доработку, затем, после предусмотренных процедур, повторно введен в эксплуатацию. Однако результаты принципиально не изменились: AUROC составляла всего 0,488 (95%-й ДИ 0,353–0,624), специфичность приближалась к нулю; ИИ-1 выдавал преимущественно ложноположительные результаты.

Результаты работы ИИ-2 были иными. При переходе к проспективному формату работы AUROC изменился на 2,6 %, что было статистически незначимо ($p = 0,7106$). Аналогичная ситуация отмечена и для общей точности – показатель оставался стабильным.

Однако в первом периоде наблюдения зафиксировано достоверное снижение специфичности на 17,7 % (t-критерий $-14,873$, $p < 0,0001$). Во втором

разрыв с уровнем, полученным при ретроспективной оценке, сократился до 8,5 %, но по-прежнему оставался значимым (t-критерий $-8,891$, $p < 0,0001$).

На этом фоне чувствительность в первом периоде оставалась высокой, но во втором, также статистически значимо, снизилась на 8,9 % (t-критерий $-21,812$, $p < 0,0001$).

При проспективной оценке клинической результативности ИИ-2 установлено, что удельный вес случаев с полным соответствием решений врачей-экспертов и ИИ-сервиса колебался в пределах 82,9–89,5 %; случаев с некорректной оценкой локализации – от 6,6 до 17,8 %. Во втором периоде наблюдения выявлено 3,95 % ложноотрицательных случаев. При клиническом мониторинге случаев гипердиагностики не зафиксировано, однако отрицательная динамика специфичности свидетельствует об их несомненном наличии во всей проспективной выборке.

Таким образом, при автоматизированном анализе РГ позвоночника с целью диагностики сколиоза уровень точности, сопоставимый с врачом-рентгенологом, в настоящее время не достижим. Модели, лежащие в основе ИИ-сервиса, отличаются тенденцией к гипердиагностике (высокому удельному весу ложноположительных результатов), что создает повышенную и неоправданную нагрузку на врача, а также отрицательно сказывается на доверии специалистов к технологиям ИИ.

Рентгенография коленного сустава. В контексте автоматизированного анализа результатов РГ коленного сустава с целью выявления дегенеративно-дистрофических нарушений была сформулирована следующая клиническая задача – определение наличия рентгенологических признаков (как минимум одного), коррелирующих с наличием деформирующего артроза суставов:

- уменьшение высоты рентгеновской суставной щели сустава (I стадия);
- наличие краевых костных разрастаний (остеофитов) 2 мм и более по суставным поверхностям сустава (II стадия);

- выраженное сужение рентгеновской суставной щели более чем в 2 раза по сравнению с нормой в сочетании с субхондральным остеосклерозом (II стадия);
- резкое сужение суставной щели менее 1 мм, вплоть до отсутствия (III стадия);
- локализация найденного признака (как минимум одного) во всем исследовании;
- определение степени выраженности изменений – 1–3 стадии деформирующего остеоартроза.

Первый этап оценки диагностической точности одного⁴ ИИ-сервиса для автоматизированного анализа результатов РГ коленного сустава с целью выявления дегенеративно-дистрофических нарушений (признаков деформирующего остеоартроза) выполнен в формате ретроспективного исследования на эталонном наборе данных ($n = 100$).

Построена характеристическая кривая (рисунок 3.7). Значение площади под характеристической кривой составило 0,780 (95%-й ДИ 0,70–0,87). Данный и иные показатели точности были ниже установленных пороговых значений, что потребовало значительной технологической доработки модели ИИ, положенной в основу изучаемого программного обеспечения. После соответствующего улучшения проведено повторное тестирование. Определены показатели: AUROC возросла до 0,914 (95%-й ДИ 0,857–0,972), общая точность – 0,880 (95%-й ДИ 0,81–0,94), чувствительность – 0,800 (95%-й ДИ 0,68–0,91), специфичность – 0,960 (95%-й ДИ 0,91–1,0). Удельный вес случаев гиподиагностики был довольно высок и составил 20,0 %, гипердиагностики – всего 4,0 %.

Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.8).

⁴ ООО «СиВижинЛаб». CVisionRad - Knee Arthrosis

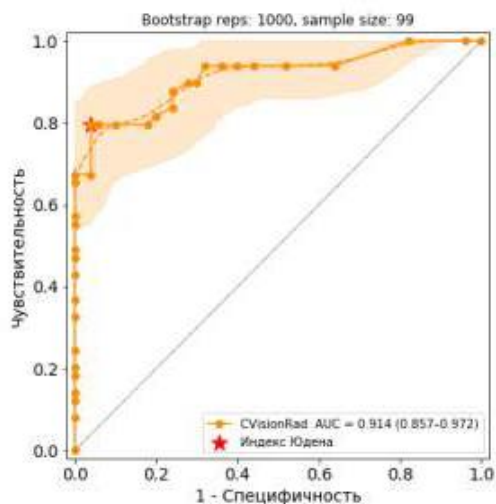


Рисунок 3.7 – Характеристическая кривая для ИИ-сервиса для анализа результатов РГ коленного сустава (оптимальный порог по индексу Юдена)

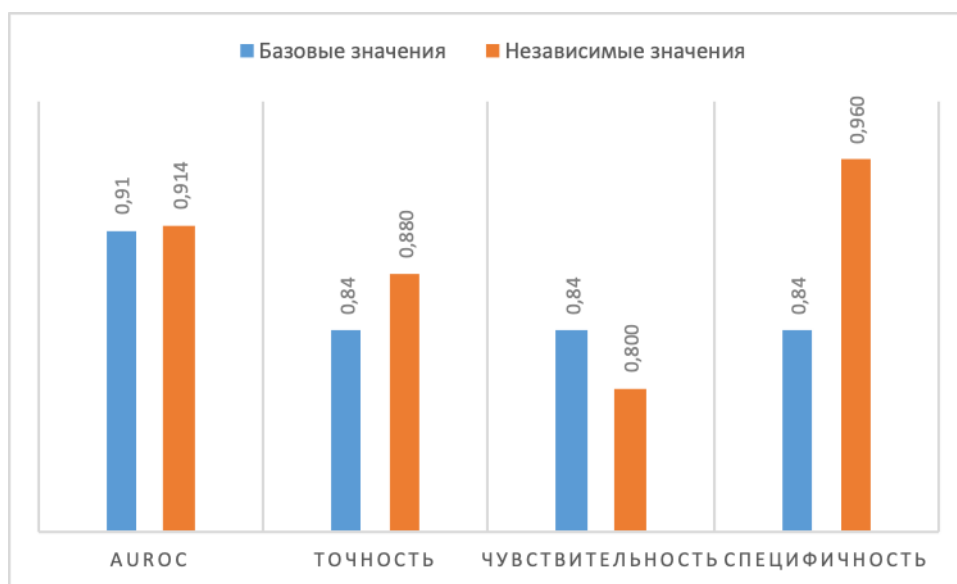


Рисунок 3.8 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов РГ коленного сустава

При ретроспективном тестировании на новых данных ИИ-сервис для анализа результатов РГ коленного сустава продемонстрировал отличную стабильность и воспроизводимость результатов своей работы. Показатели AUROC и точность возросли (0,4 % и 4,8 % соответственно), отмечается некоторое снижение чувствительности (на 4,8 %). Однако на этом фоне значительно возросла специфичность – на 14,3 %. Это проявилось минимальным удельным весом случаев гипердиагностики, указанных выше.

Второй этап оценки диагностической точности выполнен как проспективное

исследование в условиях работы ИИ-сервиса с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены четыре измерения точности в проспективном формате; результаты представлены в таблице 3.15.

Таблица 3.15 – Показатели диагностической точности ИИ-сервиса для анализа результатов РГ коленного сустава, полученные при проспективном наблюдении

Показатели	Период			
	III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
AUROC	0,934	0,908	0,966	0,928
95%-й ДИ	0,895–0,975	0,815–1,004	0,933–0,998	0,894–0,963
Точность	0,912	0,887	0,946	0,912
95%-й ДИ	0,877–0,948	0,818–0,957	0,917–0,974	0,87–0,948
Чувствительность	0,91	0,864	0,94	0,895
95%-й ДИ	0,867–0,952	0,777–0,952	0,905–0,976	0,852–0,939
Специфичность	0,921	0,952	0,958	0,98
95%-й ДИ	0,854–0,987	0,861–1,000	0,912–1,000	0,940–1,000

В первом временном периоде, в условиях работы с реальными клиническими данными ИИ-сервис для анализа результатов РГ коленного сустава продемонстрировал устойчивость и воспроизводимость результатов своей работы. При сравнении AUROC статистически значимые различия отсутствовали ($p = 0,9140$). Такая ситуация сохранялась во все периоды наблюдения. Показатели точности и специфичности также продемонстрировали стабильность при начале работы ИИ-сервиса в реальных клинических условиях, а затем, к четвертому периоду наблюдения – по сравнению с первым – даже фиксируется статистически значимый рост и точности – на 3,6 % (t-критерий $-6,148$, $p < 0,0001$), и специфичности – на 2,1 % (t-критерий $7,250$, $p < 0,0001$).

Показатель чувствительности при переходе к работе на потоке данных

статистически значимо возрос на 8,0 % (t-критерий $-8,064$, $p < 0,0001$). В последующие периоды также отмечался его рост на 11,9–17,5 % от уровня, выявленного при ретроспективной оценке. Все соответствующие значения статистически значимо различались с исходным ($p < 0,0001$).

При проспективной оценке клинической результативности установлено, что удельный вес случаев с полным соответствием решений врачей-экспертов и ИИ-сервиса колебался в пределах 15,4–94,2 %; случаев с некорректной оценкой локализации – от 1,7 до 63,1 %. Гипердиагностика отсутствовала. Однако удельный вес ложноотрицательных случаев был значителен: от 11,4 до 21,6 %, лишь в третьем периоде наблюдения он был нулевым. Наиболее качественно ИИ-сервис работал в первом и третьем периодах наблюдения: полное соответствие – в 85,0 и 94,2 % случаев, гиподиагностика – в 13,3 и 0 %; наименее качественно – во втором (минимальный уровень полного соответствия и максимальный – гиподиагностики). С точки зрения клинической результативности ИИ-сервис для анализа результатов РГ коленного сустава функционировал нестабильно.

Таким образом, при автоматизированном анализе результатов РГ коленного сустава с целью выявления дегенеративно-дистрофических нарушений (признаков деформирующего остеоартроза) потенциально достижим уровень точности, сопоставимый с врачом-рентгенологом (на потоке данных AUROC составляет 0,908–0,966, наименьшее значение случаев гиподиагностики равно нулю, случаев гипердиагностики не было). При работе в реальных клинических условиях отмечается стабильность значений AUROC и статистически значимый ($p < 0,0001$) рост значений иных показателей диагностической точности.

Положительной характеристики заслуживает стабильность ИИ-сервиса при переходе от ретроспективного к проспективному формату работы. Вместе с тем в большинстве случаев отмечается наличие недопустимо высокого уровня ложноотрицательных случаев (от 11,4 до 21,6 %, за исключением одного периода наблюдения).

Полученные данные позволяют говорить лишь о потенциальном результате. Технологии ИИ могут применяться для автоматизированного анализа результатов

РГ коленного сустава с целью выявления признаков деформирующего остеоартроза и классификации степени тяжести. В контексте данной клинической задачи ТИИ могут применяться в качестве системы поддержки принятия врачебных решений.

3.3. Обоснование возможности автоматизированного анализа результатов маммографии

В контексте автоматизированного анализа результатов профилактической маммографии (ММГ) с целью скрининга была сформулирована следующая клиническая задача – определение признаков наличия и локализации находок, характерных для злокачественных новообразований (ЗНО) молочной железы в соответствии с категориями BIRADS 3-5:

- определение рентгенологической плотности структуры по ACR (каждой молочной железы);
- определение наличия, типа (доброкачественные/подозрительные), распределения и локализации кальцинатов;
- определение наличия и локализации (квадрант) образований;
- определение нарушения и локализации (квадрант) архитектоники ткани;
- определение измененных аксиллярных лимфатических узлов;
- заключение по классификации BI-RADS по каждой молочной железе.

Ретроспективное исследование диагностической точности трех⁵ ИИ-сервисов для автоматизированного анализа результатов ММГ с целью скрининга ЗНО выполнено на эталонном наборе данных (n = 100).

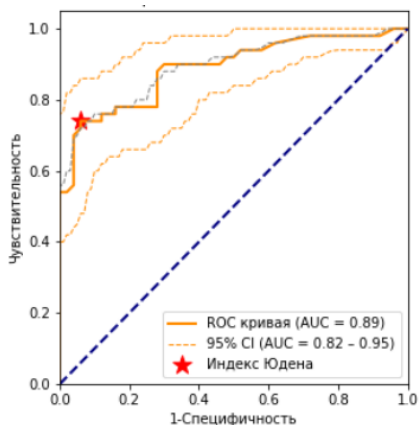
Построены характеристические кривые, значения показателей сведены в таблицу (рисунок 3.9, таблица 3.16). Наиболее высокое значение AUROC зафиксировано у ИИ-2; статистически значимо оно отличалось от такового у ИИ-3 (z-критерий Фишера 2,6746, p = 0,0075), в остальных случаях достоверных

⁵ «АО Медицинские технологии Лтд», Трио ДМ. ООО «Медицинские скрининг системы», Цельс ММГ. ООО «Платформа Третье Мнение», Третье мнение ММГ.

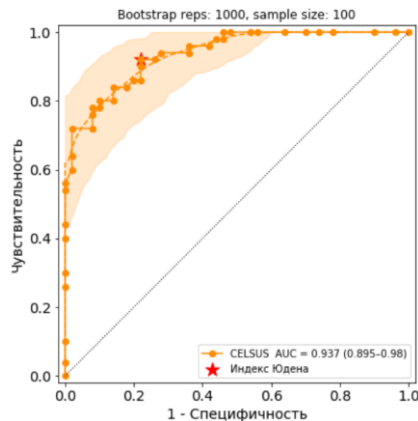
различий не было. Максимальная специфичность отмечена у ИИ-1, а точность и чувствительность – у ИИ-2. Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.10).

Таблица 3.16 – Показатели диагностической точности ИИ-сервисов для анализа результатов профилактической ММГ с целью скрининга ЗНО молочной железы, полученные при ретроспективном тестировании на эталонном наборе данных

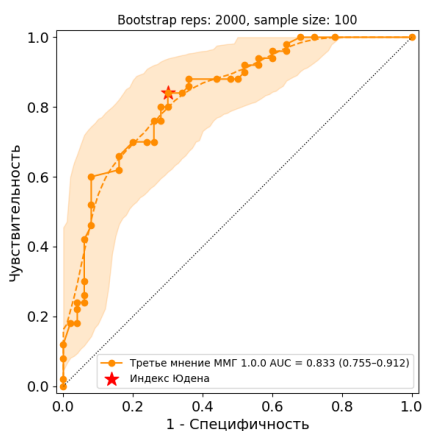
Показатели	ИИ-сервис		
	ИИ-1	ИИ-2	ИИ-3
AUROC	0,89	0,94	0,83
95%-й ДИ	0,82–0,95	0,895–0,98	0,775–0,912
Точность	0,84	0,85	0,77
95%-й ДИ	0,75–0,91	0,78–0,92	0,69–0,85
Чувствительность	0,74	0,85	0,84
95%-й ДИ	0,6–0,85	0,78–0,92	0,74–0,94
Специфичность	0,94	0,78	0,7
95%-й ДИ	0,83–1,0	0,67–0,89	0,57–0,83



ИИ-1



ИИ-2



ИИ-3

Рисунок 3.9 – Характеристические кривые для ИИ-сервисов анализа результатов профилактической ММГ с целью скрининга ЗНО молочной железы (оптимальный порог по индексу Юдена)

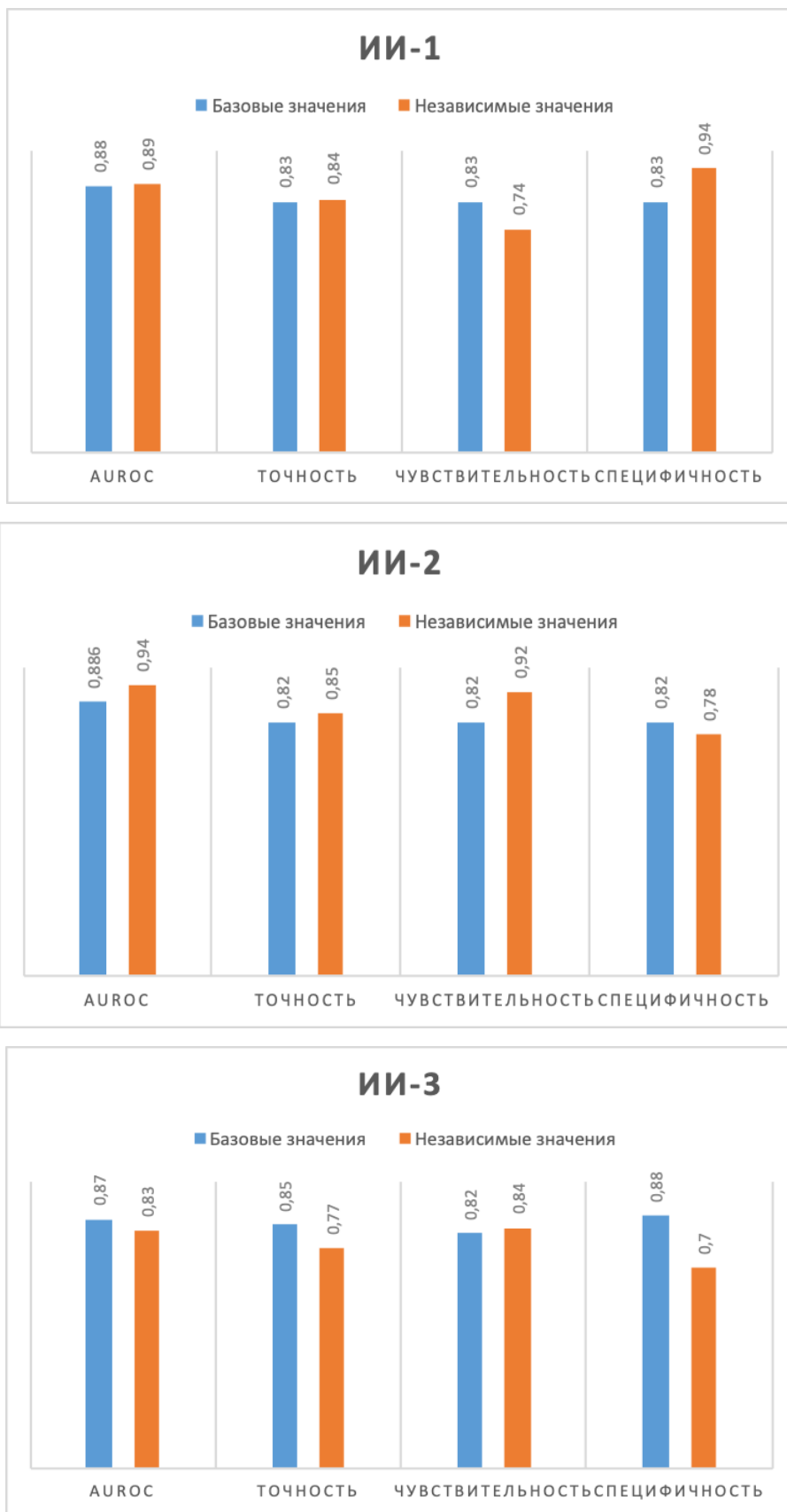


Рисунок 3.10 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервисов для анализа результатов профилактической ММГ с целью скрининга ЗНО молочной железы

При тестировании на новых данных наихудшие результаты продемонстрировал ИИ-3: все показатели точности, за исключением чувствительности, снизились; наиболее значительным было снижение специфичности – на 20,5 %. Вполне объяснимо, что такое падение точности самым негативным образом сказалось на удельном весе как ложноположительных, так и ложноотрицательных результатов. В частности, пропуски патологии зафиксированы в 16,0 % случаев, а гипердиагностика – в 30,0 %.

На этом фоне ИИ-1 и ИИ-2 продемонстрировали большую стабильность, а частично и даже некоторое улучшение. AUROC и общая точность ИИ-1 фактически остались на исходном уровне (повышение едва превышало 1,0 %), значительно возросла специфичность – на 13,3 %. Негативным моментом стало снижение чувствительности, причем на целых 10,9 %. У ИИ-2 зафиксировано снижение специфичности на 4,9 %, а значения остальных показателей увеличились, причем особенно возросла чувствительность – на 12,2 %. Стабильность и некоторая положительная динамика точности сопровождалась негативной ситуацией с ложноположительными и ложноотрицательными результатами. В частности, у ИИ-1 удельный вес ложноотрицательных результатов составил 26,0 %, гипердиагностика зафиксирована в 6,0 % случаев. ИИ-2 отличался принципиально меньшей гипо- (8,0 %), но более высокой гипердиагностикой (22,0 %).

Таким образом, два из трех изученных прикладных решений на основе ИИ продемонстрировали стабильность и хорошую воспроизводимость результатов работы при тестировании на новых данных.

Второй этап оценки точности проведен в проспективном формате, в условиях работы ИИ-сервисов с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях.

В соответствии с методологией проведены от двух до семи измерений точности в проспективном формате; результаты представлены в таблицах 3.17–3.18.

Таблица 3.17 – Показатели диагностической точности ИИ-сервисов для анализа результатов профилактической маммографии, полученные при проспективном наблюдении в 2022 году

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.
ИИ-1	AUROC	0,771	0,661	0,819	0,795
	95%-й ДИ	0,668–0,871	0,499–0,828	0,710–0,934	0,617–0,964
	Точность	0,733	0,446	0,792	0,85
	95%-й ДИ	0,677–0,789	0,383–0,509	0,740–0,843	0,772–0,928
	Чувствительность	0,75	0,909	0,733	0,636
	95%-й ДИ	0,590–0,910	0,739–1,000	0,510–0,957	0,352–0,921
	Специфичность	0,731	0,424	0,796	0,884
	95%-й ДИ	0,671–0,791	0,360–0,488	0,743–0,848	0,809–0,960
ИИ-2	AUROC	0,637	0,846	0,83	0,813
	95%-й ДИ	0,497–0,778	0,788–0,903	0,734–0,928	0,742–0,886
	Точность	0,375	0,758	0,767	0,608
	95%-й ДИ	0,300–0,450	0,704–0,812	0,713–0,820	0,547–0,670
	Чувствительность	0,938	1	0,792	0,943
	95%-й ДИ	0,819–1,000	1,000–1,000	0,629–0,954	0,866–1,000
	Специфичность	0,312	0,746	0,764	0,551
	95%-й ДИ	0,237–0,388	0,689–0,802	0,707–0,821	0,483–0,619

Таблица 3.18 – Показатели диагностической точности ИИ-сервисов для анализа результатов профилактической маммографии, полученные при проспективном наблюдении в 2023 году

ИИ-сервис	Показатели	Период		
		I кв. 2023 г.	II кв. 2023 г.	III кв. 2023 г.
ИИ-1	AUROC	0,872	0,869	0,803
	95%-й ДИ	0,805–0,942	0,767–0,976	0,688–0,915
	Точность	0,787	0,637	0,744
	95%-й ДИ	0,736–0,839	0,577–0,698	0,676–0,811
	Чувствительность	0,867	1	0,8

Продолжение таблицы 3.18

ИИ-сервис	Показатели	Период		
		I кв. 2023 г.	II кв. 2023 г.	III кв. 2023 г.
	95%-й ДИ	0,695–1,000	1,000–1,000	0,598–1,000
	Специфичность	0,782	0,627	0,738
	95%-й ДИ	0,728–0,836	0,565–0,689	0,666–0,810
ИИ-2	AUROC	0,889	0,902	0,839
	95%-й ДИ	0,826–0,952	0,792–1,012	0,688–0,995
	Точность	0,842	0,863	0,594
	95%-й ДИ	0,795–0,888	0,819–0,906	0,518–0,670
	Чувствительность	0,818	0,909	1
	95%-й ДИ	0,687–0,950	0,739–1,000	1,000–1,000
	Специфичность	0,845	0,86	0,581
	95%-й ДИ	0,796–0,895	0,815–0,905	0,503–0,658
ИИ-3	AUROC	–	0,86	0,829
	95%-й ДИ	–	0,775–0,950	0,728–0,932
	Точность	–	0,762	0,725
	95%-й ДИ	–	0,697–0,828	0,656–0,794
	Чувствительность	–	1	1
	95%-й ДИ	–	1,000–1,000	1,000–1,000
	Специфичность	–	0,752	0,712
	95%-й ДИ	–	0,683–0,820	0,641–0,784

При переходе к работе в реальных клинических условиях ИИ-1 продемонстрировал устойчивое снижение показателей. В первом периоде наблюдения AUROC достоверно снизилась на 13,4 % (z-критерий Фишера 1,9352, $p = 0,043$). Затем последовали постоянные колебания этого показателя, в том числе, периодически, с падением ниже порогового уровня. Лишь к пятому периоду наблюдения AUROC стабилизировалась, хотя и оставалась на 9,8 % меньше значения, полученного при ретроспективном тестировании (различия при этом утратили значимость $p = 0,1923$).

Аналогичной была динамика общей точности. Исходное падение на 12,7 % (t-критерий $-20,649$, $p < 0,0001$) фактически сохранялось во все периоды наблюдения, периодически ухудшаясь до совершенно критичного уровня; минимальное значение в 0,446 (95%-й ДИ 0,383–0,509) зафиксировано во втором периоде наблюдения. Полностью идентичной была динамика специфичности, отличие состояло лишь в проценте падения – в первом периоде значение этого показателя снизилось на 22,2 % (t-критерий $-37,839$, $p < 0,0001$) и оставалось фактически таким же до завершения исследования. На этом фоне лишь чувствительность продемонстрировала устойчивость и рост. В первом периоде она незначимо возросла – на 1,4 % ($p = 0,3704$). Далее колебалась, вплоть до критичного падения до уровня 0,636 (95%-й ДИ 0,352–0,921). Однако в три заключительных периода наблюдения стала возрастать, стабилизовавшись в пределах 0,8–1,0 и значимо отличаясь от исходного уровня (t-критерий $4,572$, $p < 0,0001$).

При проспективной оценке клинической результативности установлено, что удельный вес случаев с полным соответствием решений врачей-экспертов и ИИ-1 был довольно низким и колебался в пределах 35,8–74,7 %. За исключением только одного периода фиксировались пропуски патологии: минимальные значения составили 0,5 и 0,9 %, максимальное – 5,0 %. Сервис отличался высоким уровнем гипердиагностики (что вполне объяснимо дисбалансом чувствительности и специфичности); соответствующее минимальное значение составило 6,3 %, максимальное – 36,0 %. Стабильно высоким был процент некорректной локализации и маркировки признаков патологических нарушений – в среднем 24,1 %.

Очень схожей была динамика показателей работы ИИ-2. Сразу же после перехода к работе в реальных клинических условиях произошло критичное падение AUROC на 32,2 % (z-критерий Фишера $4,0459$, $p = 0,0001$), точности – на 55,9 % (t-критерий $-85,229$, $p < 0,0001$), специфичности – на 60,0 % (t-критерий $-66,171$, $p < 0,0001$). В дальнейшем ситуация улучшилась. Площадь под характеристической кривой превысила пороговое значение, далее колебалась и

установилась к концу наблюдений на уровне 0,839 (95%-й ДИ 0,688–0,995), который был ниже исходного на 10,8 % ($p = 0,2139$). Аналогичной была динамика общей точности и специфичности, но с еще более выраженными падениями критичного характера. К заключительному периоду исследования «разрыв» точности с исходным уровнем сократился, но лишь до 30,1 % (t-критерий $-45,575$, $p < 0,0001$), а специфичности – до 25,5 % (t-критерий $-27,861$, $p < 0,0001$). На этом фоне чувствительность ИИ-2 с первого же периода начала расти. Лишь с однократным «провалом» на 6,8 % в одном периоде. К концу наблюдения прирост составил 17,7 % (t-критерий $42,030$, $p < 0,0001$); значение этого показателя достигло 1,0 (95%-й ДИ 1,000–1,000).

При проспективной оценке клинической результативности обращает на себя внимание общая положительная динамика постепенного уменьшения удельного веса как гипо-, так и гипердиагностики. Для наглядности результаты клинических мониторингов ИИ-1 и ИИ-2 представлены на графиках (рисунок 3.11).

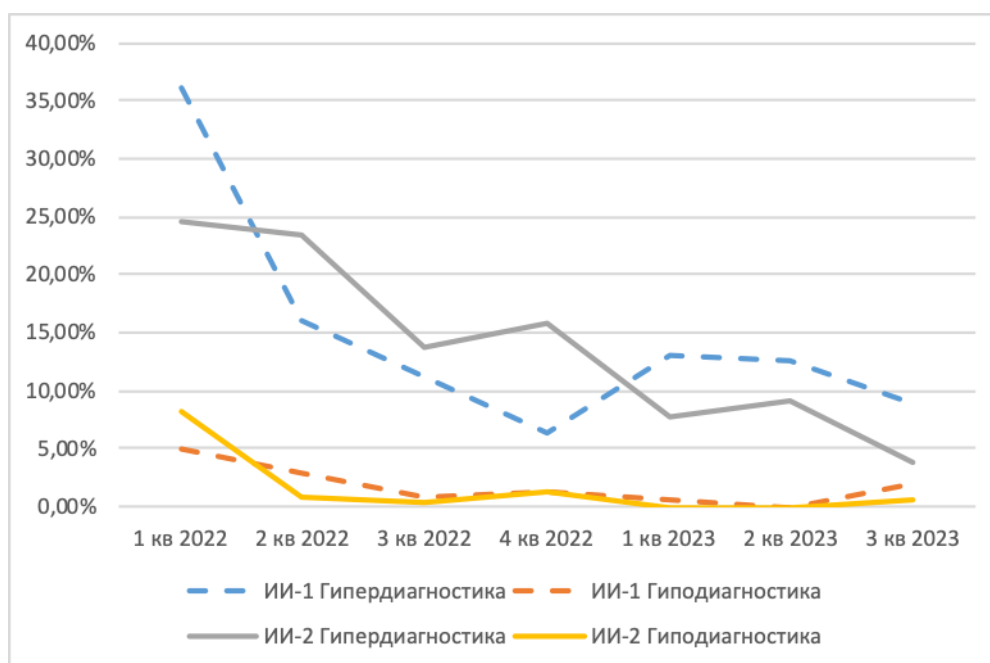


Рисунок 3.11 – Динамика удельного веса ложноотрицательных и ложноположительных решений ИИ-1 и ИИ-2 для анализа результатов профилактической ММГ с целью скрининга ЗНО молочной железы

В первом периоде наблюдения удельный вес случаев гипердиагностики ИИ-2 составлял 24,6 %. Затем он постоянно снижался, достигнув минимального уровня в 3,8 % в последнем периоде наблюдения. Уровень пропусков патологии в первом периоде был 8,2 %, однако в дальнейшем этот показатель снизился до минимума, составляя от 0,6 до 1,3 %, а в двух периодах – 0. На этом фоне ИИ-1 выглядит более нестабильным с точки зрения клинической результативности.

При работе в реальных клинических условиях ИИ-3 отличался стабильностью всех показателей. Не имели значимости различия AUROC ($p = 0,5967, 0,9873$). Только во втором периоде значимо – на 5,8 % – уменьшилась точность (t-критерий 7,937, $p < 0,0001$). Специфичность продемонстрировала рост на 7,4 % (t-критерий 6,758, $p < 0,0001$), затем вернулась к исходному уровню ($p = 0,1249$). Чувствительность ИИ-3 в проспективном формате работы уверенно составляла 1,0 (95%-й ДИ 1,000–1,000). Данный сервис отличал нулевой уровень пропусков патологии. Тем не менее гипердиагностика составляла от 16,3 до 23,8 %; высок был процент некорректной локализации патологических находок.

Таким образом, при автоматизированном анализе результатов профилактической ММГ с целью скрининга ЗНО молочной железы отмечается достижение наивысшего уровня чувствительности (1,0). Дисбаланс с низкой специфичностью обуславливает почти постоянно имеющийся процент пропуска патологии – в среднем $1,5 \pm 2,2$ % (медиана – 0,7 %). Вместе с тем, по литературным данным, удельный вес клинически значимых дефектов в работе врача-рентгенолога при интерпретации результатов профилактической ММГ составляет от 3,5 до 30,0 % [148, 183, 169, 173, 206, 217].

Негативным аспектом является стабильно высокий процент гипердиагностики (в среднем $15,2 \pm 8,3$ %, медиана – 13,4 %), что делает применение таких ИИ-сервисов малопривлекательным из-за создания повышенной и неоправданной нагрузки на врача.

Вместе с тем полученные данные позволяют говорить не только о потенциальном, но и прикладном результате. Технологии ИИ могут применяться для автоматизированного анализа результатов профилактической ММГ с целью

скрининга ЗНО молочной железы в качестве системы сортировки, в том числе в контексте так называемого «автономного искусственного интеллекта» (этот вопрос подробно будет изучен далее).

3.4. Обоснование возможности автоматизированного анализа результатов компьютерной томографии

Компьютерная томография головного мозга. В контексте автоматизированного анализа результатов КТ головного мозга с целью диагностики внутричерепного кровоизлияния была сформулирована следующая клиническая задача – определение наличия кровоизлияния и автоматический подсчет его объема в мл или см кубических:

- наличие на нативных изображениях рентгенологических признаков, соответствующих кровоизлияниям по типам: эпидуральных, субдуральных, субарахноидальных или внутримозговых;
- определение локализации кровоизлияния;
- определение типа кровоизлияния;
- подсчет объема в мл или см кубических.

Ретроспективное исследование диагностической точности двух⁶ ИИ-сервисов для автоматизированного анализа результатов КТ головного мозга (далее – ИИ-1 и ИИ-2) с целью диагностики внутричерепного кровоизлияния выполнено на эталонном наборе данных ($n = 100$).

Построена характеристическая кривая (рисунок 3.12). Для ИИ-1 значение площади под характеристической кривой составило 0,904 (95%-й ДИ 0,834–0,974). Определены показатели: общая точность – 0,890 (95%-й ДИ 0,83–0,95), чувствительность – 0,910 (95%-й ДИ 0,82–0,99), специфичность – 0,880 (95%-й ДИ 0,78–0,97). Удельный вес ложноотрицательных результатов составил 9,1 %, а ложноположительных – 12,5 %.

⁶ 1) ООО «Интеллоджик», NTechMed CT Brain. 2) ООО «Медицинские скрининг системы», Цельс.

Для ИИ-2: площадь под характеристической кривой – 0,960 (95%-й ДИ 0,92–1,0), точность – 0,920 (95%-й ДИ 0,87–0,98), чувствительность – 0,890 (95%-й ДИ 0,8–0,98), специфичность – 0,960 (95%-й ДИ 0,9–1,0). Удельный вес ложноотрицательных результатов был чрезвычайно мал и составил лишь 0,11 %. Удельный вес случаев гипердиагностики был также минимален – 0,04 %. Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.13).

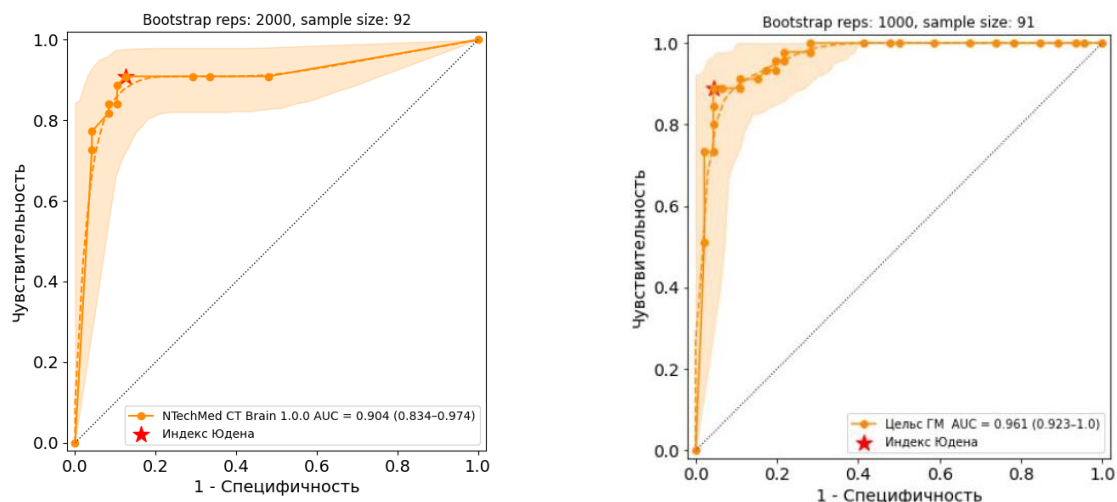


Рисунок 3.12 – Характеристические кривые для ИИ-сервисов для анализа результатов КТ головного мозга с целью диагностики внутричерепного кровоизлияния (оптимальный порог по индексу Юдена):
слева – ИИ-1, справа – ИИ-2

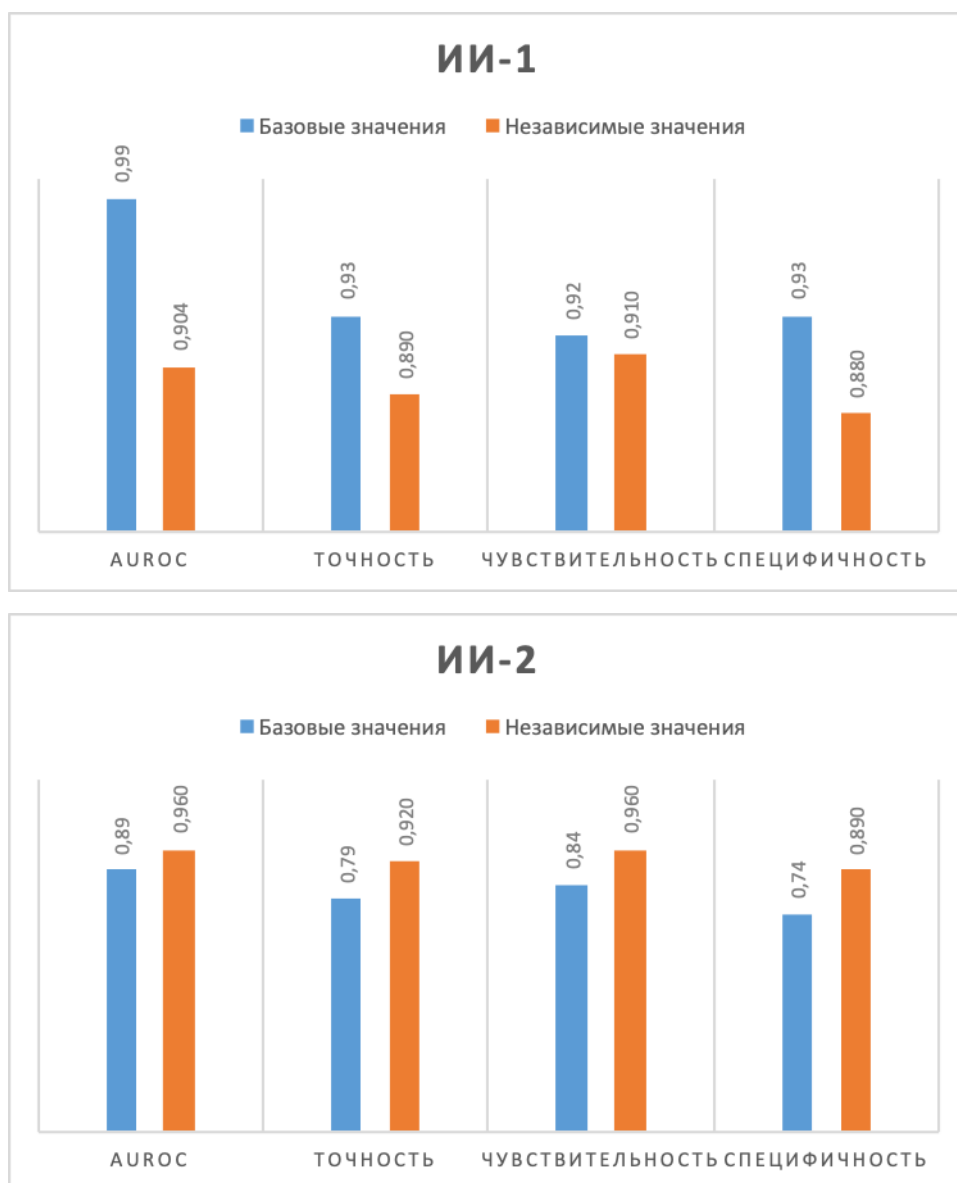


Рисунок 3.13 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервисов для анализа результатов КТ головного мозга с целью диагностики внутримозгового кровоизлияния

Обращает на себя внимание явное снижение значений показателей точности, полученных в ходе независимого ретроспективного тестирования ИИ-1. В частности, относительного базового значение площади под характеристической кривой снизилось на 8,7 %, точности – на 4,3 %, специфичности – на 5,4 %; лишь снижение чувствительности были минимальным (1,1 %). На этом фоне ИИ-2 продемонстрировал принципиально лучшее качество и воспроизводимость результатов своей работы на новых данных. Показатели не просто остались

стабильными, но значительно возросли: AUROC на 7,9 %, точность на 16,5 %, чувствительность и специфичность на 14,3 и 20,3 % соответственно.

Полученные результаты позволили перейти ко второму, проспективному этапу.

Проспективное исследование обоих ИИ-сервисов проведено в условиях их работы с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены два измерения точности ИИ-1 и четыре измерения точности ИИ-2 в проспективном формате; результаты представлены в таблице 3.19.

Таблица 3.19 – Показатели диагностической точности ИИ-сервиса для анализа результатов КТ головного мозга с целью диагностики внутричерепного кровоизлияния, полученные при проспективном наблюдении

ИИ-сервис	Показатели	Период			
		III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
ИИ-1	AUROC	-	-	0,957	0,930
	95%-й ДИ	-	-	0,862–1,046	0,895–0,964
	Точность	-	-	0,988	0,825
	95%-й ДИ	-	-	0,963–1,000	0,777–0,873
	Чувствительность	-	-	0,857	0,941
	95%-й ДИ	-	-	0,598–1,000	0,877–1,000
	Специфичность	-	-	1,000	0,794
	95%-й ДИ	-	-	1,000–1,000	0,736–0,851
ИИ-2	AUROC	0,952	0,935	0,919	0,929
	95%-й ДИ	0,926–0,978	0,901–0,971	0,883–0,955	0,896–0,959
	Точность	0,901	0,825	0,863	0,856

Продолжение таблицы 3.19

ИИ-сервис	Показатели	Период			
		III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
	95%-й ДИ	0,864–0,939	0,777–0,873	0,819–0,906	0,812–0,900
	Чувствительность	0,937	0,941	0,861	0,863
	95%-й ДИ	0,892–0,982	0,885–0,997	0,796–0,926	0,801–0,926
	Специфичность	0,871	0,779	0,864	0,849
	95%-й ДИ	0,814–0,928	0,717–0,841	0,805–0,922	0,787–0,912

В условиях работы с реальными клиническими данными у ИИ-1 значение AUROC незначимо ($p = 0,3689$, $p = 0,5138$), колебалось в пределах 2,9–5,9 %.

Показатель точности значимо возрос на 11,0 % (t-критерий 30,283, $p < 0,0001$), однако во втором периоде наблюдения – снизился на 7,3 % по отношению к уровню, полученному при ретроспективной оценке (также значимо – t-критерий -15,829, $p < 0,0001$).

Показатель чувствительности в первом периоде уменьшился на 5,8 %, однако в дальнейшем значимо возрос (t-критерий 5,559, $p < 0,0001$) относительно полученного на первом этапе исследования на 3,4 %.

Схожая тенденция наблюдалась для специфичности: принципиальное улучшение в первом периоде (рост на 13,6 %, t-критерий 24,775, $p < 0,0001$) сменилось падением на целых 9,8 % (t-критерий -14,699, $p < 0,0001$).

Указанные негативные изменения были подтверждены проспективной оценкой клинической результативности. В сравнении между первым и вторым периодами наблюдения удельный вес исследований с полным соответствием снизился с 75,7 до 43,1 %, а исследований с некорректной локализацией – возрос с 5,1 до 10,1 %. Положительной тенденцией стало исчезновение ложноотрицательных случаев (в первом периоде их было 1,3 %). Однако ухудшение специфичности привело к значительному росту гипердиагностики с 18,0 до 46,7 %. В конечном итоге работа ИИ-1 на потоке исследований была прекращена.

В условиях работы с реальными клиническими данными у ИИ-2 значение AUROC так же, как и у ИИ-1, несущественно ($p = 0,1354$, $p = 0,3566$) колебалось в пределах 0,8–4,3 % с тенденцией к снижению.

На этом фоне отмечено значимое ухудшение общей точности: в первом периоде на 6,2 % (t-критерий -14,614, $p < 0,0001$), в четвертом – на 10,3 % (t-критерий -24,233, $p < 0,0001$).

Обращает на себя внимание динамика чувствительности: в первые два периода отмечается достоверное снижение примерно на 3,0 % ($p < 0,0001$); однако, далее показатель уверенно растет – в четвертом периоде на 5,7 % от исходного уровня (t-критерий 9,120, $p < 0,0001$).

Динамика специфичности была исключительно отрицательной: в первый период наблюдения она снизилась на 10,0 %, а к четвертому – на 18,9 % от исходного, установленного в ретроспективном исследовании. Конечно, все указанные изменения были статистически значимы (t-критерий -22,853 и -41,502 соответственно, $p < 0,0001$).

Негативные тенденции основных показателей диагностической точности подтверждаются результатами проспективной оценки клинической результативности.

Удельный вес исследований с полным соответствием изначально был низким – всего 53,9 %; в четвертом периоде наблюдения он упал до 35,0 %. Гипердиагностика, наоборот, возросла с 15,4 % до 46,3 %. Единственной, хотя и безусловно важной, положительной тенденцией стало исчезновение в третьем и четвертом периодах наблюдения ложноотрицательных случаев (в первом и втором их было 2,6 % и 1,3 % соответственно).

ИИ-2 характеризовался устойчивостью при переходе от ретроспективного к проспективному формату работы, однако в дальнейшем его показатели начали существенно ухудшаться, теряя клиническую значимость.

Таким образом, при автоматизированном анализе результатов КТ головного мозга с целью диагностики внутричерепного кровоизлияния потенциально достижим уровень точности, сопоставимый с врачом-рентгенологом (на потоке

данных AUROC составляет 0,92–0,96, возможно полное отсутствие случаев гиподиагностики, максимальный их удельный вес достигает 2,6 %). Вместе с тем обе изученные модели ИИ демонстрируют негативную тенденцию выраженного ухудшения специфичности, что приводит к росту случаев гипердиагностики, превысивших 46,0 % в обоих случаях. Парадоксально, что при переходе в реальные клинические условия оба ИИ-сервиса продемонстрировали высокую устойчивость и воспроизводимость результатов своей работы, однако в дальнейшем работали нестабильно, показатели точности постоянно ухудшались (вплоть до прекращения работы ИИ-1).

Полученные данные позволяют говорить только лишь о потенциальном результате. Технологии ИИ потенциально могут применяться для автоматизированного анализа результатов КТ головного мозга с целью диагностики внутричерепного кровоизлияния в качестве системы поддержки принятия врачебных решений и для морфометрии (подсчета объема кровоизлияния).

В контексте автоматизированного анализа результатов КТ головного мозга с целью диагностики острого нарушения мозгового кровообращения (ОНМК) ишемического типа была сформулирована следующая клиническая задача – определение наличия острого ишемического инсульта и его оценка в баллах по шкале ASPECTS⁷:

- наличие на нативных изображениях рентгенологических признаков острого ишемического инсульта;
- указание бассейна областей острого ишемического инсульта;
- оценка указанных выше областей по ASPECTS (0–10 баллов), если затронут бассейн средней мозговой артерии.

Ретроспективное исследование диагностической точности одного⁸ ИИ-сервиса для автоматизированного анализа результатов КТ головного мозга с целью диагностики и классификации ОНМК ишемического типа выполнено на эталонном наборе данных (n = 100).

⁷ Pexman J. H., Barber P. A., Hill M. D., et al.. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke // AJNR Am J Neuroradiol. 2001. Vol. 22, №8. P. 1534–1542.

⁸ ООО «Гаммамед-Софт». Multivox ASPECTS.

Построена характеристическая кривая (рисунок 3.14). Значение площади под характеристической кривой составило 0,855 (95%-й ДИ 0,779–0,930). Определены показатели: общая точность составила 0,840 (95%-й ДИ 0,77–0,91), чувствительность – 0,860 (95%-й ДИ 0,76–0,96), специфичность – 0,820 (95%-й ДИ 0,71–0,93). Удельный вес ложноотрицательных результатов оказался довольно высоким и составил 14,0 %, в то время как ложноположительных – 18,0 %.

Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.15).

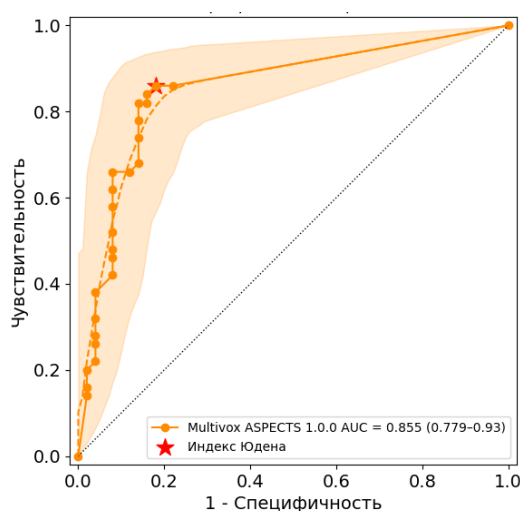


Рисунок 3.14 – Характеристическая кривая для ИИ-сервиса для анализа результатов КТ головного мозга с целью диагностики и классификации ОНМК ишемического типа (оптимальный порог по индексу Юдена)

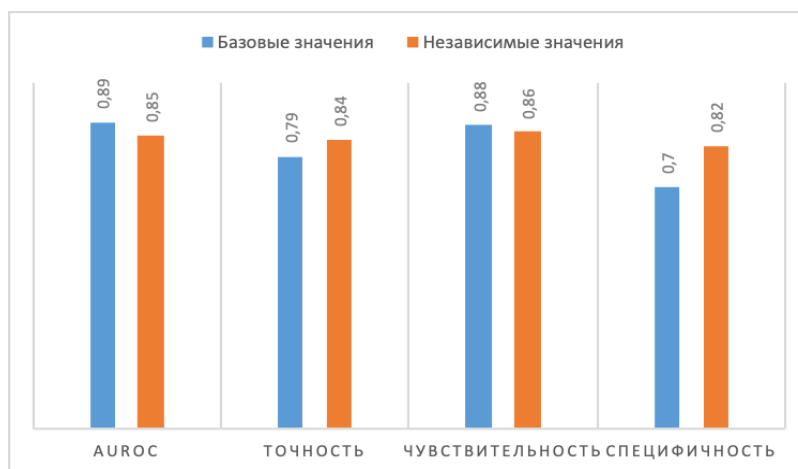


Рисунок 3.15 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов КТ головного мозга с целью диагностики и классификации ОНМК ишемического типа

Отмечаются незначительные колебания AUROC, точности и чувствительности (в пределах 2,0–6,0 %). Обращает на себя внимание рост специфичности на 17,2 %, выявленный при тестировании ИИ-сервиса на новых данных.

На втором этапе ИИ-сервис анализировал поток результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного). Выполнено проспективное диагностическое исследование. Проведены два измерения точности, результаты представлены в таблице 3.20.

Таблица 3.20 – Показатели диагностической точности ИИ-сервиса для анализа результатов КТ головного мозга с целью диагностики и классификации ОНМК ишемического типа, полученные при проспективном наблюдении

Показатели	Период	
	I кв. 2023 г.	II кв. 2023 г.
AUROC	0,660	0,697
95%-й ДИ	0,556–0,762	0,618–0,776
Точность	0,537	0,694
95%-й ДИ	0,474–0,601	0,622–0,765
Чувствительность	0,850	0,960
95%-й ДИ	0,694–1,000	0,906–1,000
Специфичность	0,509	0,573
95%-й ДИ	0,443–0,575	0,480–0,665

В первом временном периоде, в условиях работы с реальными клиническими данными, ИИ-сервис для анализа результатов КТ головного мозга продемонстрировал выраженное ухудшение точности: показатель площади под характеристической кривой снизился на целых 22,8 % (z-критерий Фишера 2,9928, $p = 0,0028$), точности – на 36,1 % (t-критерий -59,570, $p < 0,0001$), специфичности – на 37,9 % (t-критерий -46,034, $p < 0,0001$) относительно уровня, полученного при ретроспективной оценке. Незначительно снизилась лишь чувствительность (на 1,1 %, $p = 0,3240$). Низкая точность обусловила

приостановку работы ИИ-сервиса. После технической доработки он возобновил свою работу, но, как показал второй период наблюдения, точность автоматизированного анализа осталась в целом неудовлетворительной. Площадь под характеристической кривой оставалась значимо (z-критерий Фишера 2,8339, $p = 0,0046$) ниже исходного уровня, а также сильно ниже порогового.

Столь же низки были точность и специфичность (t-критерий -26,931 и -32,068 соответственно, $p < 0,0001$). Единственным положительным моментом стал достоверный рост чувствительности выше исходного на 11,6 % (t-критерий 17,358, $p < 0,0001$).

Однако в сопоставлении с результатами оценки клинической результативности это не изменило общую негативную картину. Повышенная чувствительность на фоне недостаточной специфичности привела к значительному числу случаев гипердиагностики (31,9–45,0 %), также часто отмечалась некорректная локализация патологических находок (11,3–42,5 %). Общий объем исследований с полным соответствием составил лишь 42,5 % и 25,5 % в первом и втором периодах наблюдения соответственно. Случаи ложноотрицательных решений не превысили 1,3 % и зафиксированы только в первом периоде.

Таким образом, при автоматизированном анализе результатов КТ головного мозга с целью диагностики ишемического ОНМК и классификации по шкале ASPECTS уровень точности, сопоставимый с врачом-рентгенологом, в настоящее время не достижим. Модель, лежащая в основе ИИ-сервиса, отличается крайне низкой общей точностью (при работе с реальными клиническими данными AUROC достоверно снизилась ниже 0,7) и некоторой тенденцией к гипердиагностике.

Компьютерная томография органов грудной клетки. В контексте автоматизированного анализа результатов КТ органов грудной клетки с целью выявления гидроторакса была сформулирована следующая клиническая задача – определение наличия свободной жидкости (выпота) в плевральных полостях:

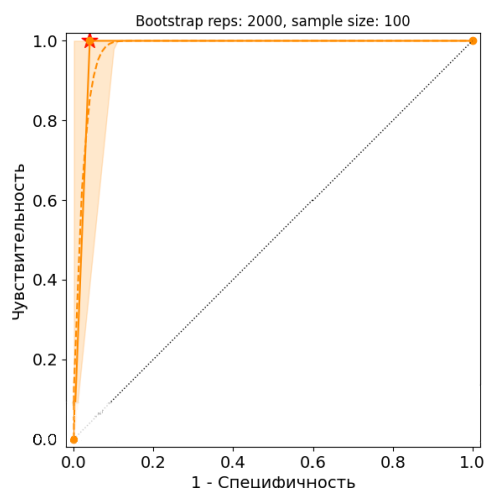
- наличие на нативных изображениях в плевральной полости серповидного скопления жидкостного содержимого (выпота) плотностью 0–30 НУ в гравитационно зависимых отделах грудной клетки;
- локализация найденных патологических находок;
- объем плеврального выпота (мл) для каждого легкого;
- среднее значение плотности выпота (НУ) для каждого легкого.

Ретроспективное исследование диагностической точности трех⁹ ИИ-сервисов для автоматизированного анализа результатов КТ органов грудной клетки с целью выявления гидроторакса выполнено на эталонном наборе данных (n = 100). Построены характеристические кривые, полученные показатели сведены в таблицу (таблица 3.21, рисунок 3.16).

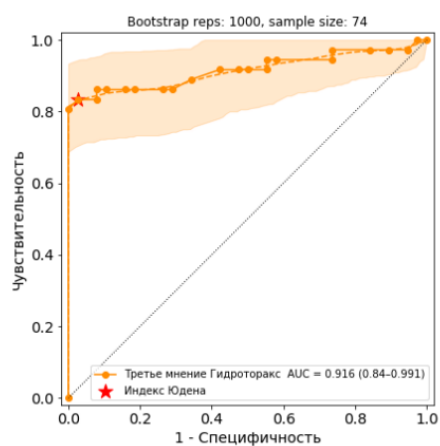
Таблица 3.21 – Показатели диагностической точности ИИ-сервисов для анализа результатов КТ органов грудной клетки с целью выявления гидроторакса, полученные при ретроспективном тестировании на эталонном наборе данных

Показатели	ИИ-сервис		
	ИИ-1	ИИ-2	ИИ-3
AUROC	0,980	0,920	0,890
95%-й ДИ	0,953–1,000	0,840–0,990	0,823–0,948
Точность	0,980	0,910	0,890
95%-й ДИ	0,95–1,0	0,840–0,970	0,82–0,95
Чувствительность	1,000	0,830	0,960
95%-й ДИ	1,000–1,000	0,710–0,960	0,9–1,0
Специфичность	0,960	0,970	0,810
95%-й ДИ	0,91–1,0	0,920–1,000	0,7–0,92

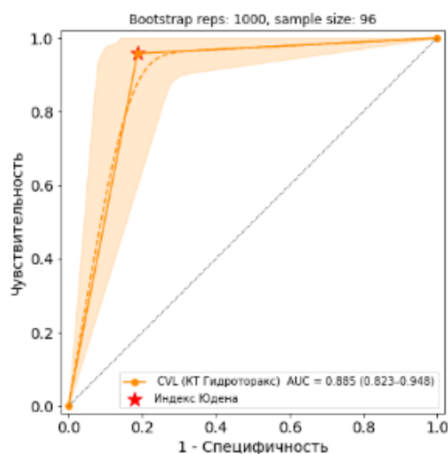
⁹ ООО «Интел Диагностик», система поддержки принятия диагностических решений «АИ Диагностик»; ООО «Платформа Третье Мнение», программный модуль для анализа исследований компьютерной томографии грудной клетки человека; ООО «СиВижинЛаб», CVL – Chest CT Pleural Effusion.



ИИ-1



ИИ-2



ИИ-3

Рисунок 3.16 – Характеристические кривые ИИ-сервисов для анализа результатов КТ органов грудной клетки с целью выявления гидроторакса (оптимальный порог по индексу Юдена)

При тестировании на новых данных наибольшие значения показателей диагностической точности, за исключением специфичности, продемонстрировал ИИ-1; наименьшие, за исключением чувствительности, – ИИ-3.

AUROC ИИ-1 и ИИ-2 не различались между собой значимо ($p = 0,1346$); в то время, как AUROC ИИ-3 была значимо меньше двух иных сервисов (z-критерий Фишера 2,6419 и 0,6023, $p = 0,0082$, 0,5470 соответственно).

Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.17).

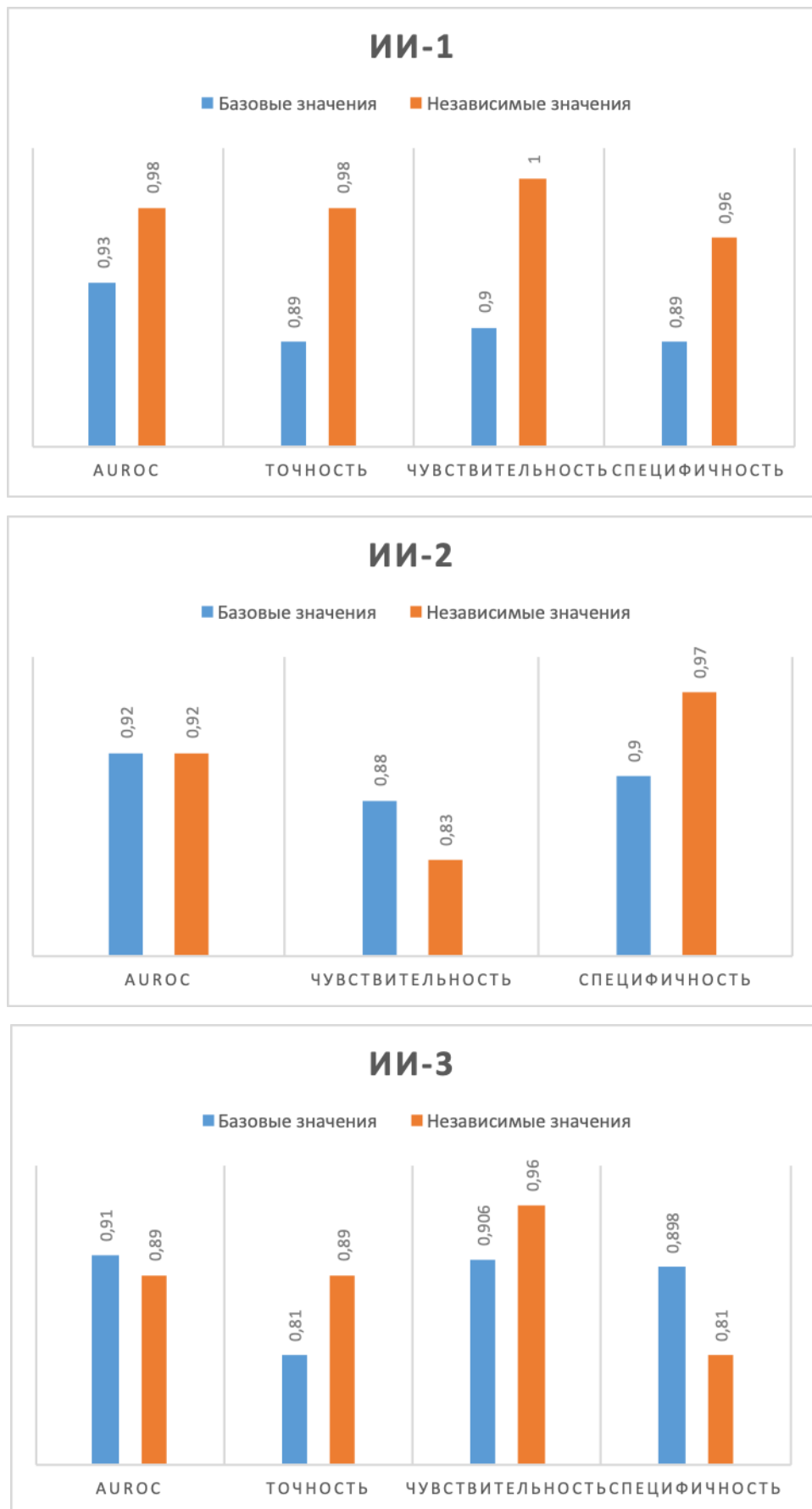


Рисунок 3.17 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов КТ органов грудной клетки с целью выявления гидроторакса

Примечательно, что ИИ-1 продемонстрировал не просто устойчивость, но положительную динамику, то есть увеличение показателей: AUROC возросла на 5,4 %, точность – на 10,1 %, чувствительность – на 11,1 %, специфичность – на 7,9 %. При этом не зафиксировано ни одного ложноотрицательного результата. Гипердиагностика составила 4,0 %.

ИИ-2 продемонстрировал стабильность уровня AUROC и рост на 7,8 % специфичности, однако при этом на 5,7 % снизилась чувствительность. Ложноотрицательные результаты отмечены в 17,0 % случаев, а гипердиагностика – в 3,0 %.

ИИ-3 продемонстрировал нестабильность: часть показателей увеличилась, часть уменьшилась. Особо значительным было снижение специфичности (на 9,8 %), которое обусловило появление 0,04 % ложноотрицательных и 0,19 % ложноположительных результатов. Прирост чувствительности на 5,9 % привел к связанному увеличению общей точности на 9,9 %. Однако AUROC уменьшилась на 2,2 %.

В целом можно отметить удовлетворительную устойчивость и воспроизводимость результатов работы ИИ-сервисов на новых данных в условиях ретроспективного тестирования.

Проспективное исследование выполнено в условиях работы ИИ-сервисов с потоком результатов лучевых исследований, проведенных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией, проведены от двух до трех измерений точности в проспективном формате; результаты представлены в таблице 3.22.

Таблица 3.22 – Показатели диагностической точности ИИ-сервиса для анализа результатов КТ органов грудной клетки с целью выявления гидроторакса, полученные при проспективном наблюдении

ИИ-сервис	Показатели	Период			
		III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
ИИ-1	AUROC	-	0,706	0,952	0,985
	95%-й ДИ	-	0,667–0,746	0,896–1,005	0,965–1,005
	Точность	-	0,614	0,963	0,975
	95%-й ДИ	-	0,553–0,676	0,921–1,000	0,941–1,000
	Чувствительность	-	0,988	1,000	1,000
	95%-й ДИ	-	0,964–1,000	1,000–1,000	1,000–1,000
	Специфичность	-	0,425	0,903	0,970
	95%-й ДИ	-	0,348–0,502	0,799–1,000	0,928–1,000
ИИ-2	AUROC	0,925	0,856	0,955	-
	95%-й ДИ	0,868–0,984	0,783–0,930	0,908–1,000	-
	Точность	0,946	0,863	0,938	-
	95%-й ДИ	0,917–0,974	0,819–0,906	0,884–0,991	-
	Чувствительность	0,897	0,778	0,941	-
	95%-й ДИ	0,818–0,975	0,642–0,914	0,862–1,000	-
	Специфичность	0,962	0,877	0,935	-
	95%-й ДИ	0,934–0,989	0,832–0,922	0,863–1,000	-
ИИ-3	AUROC	0,672	-	0,894	-
	95%-й ДИ	0,614–0,732	-	0,836–0,954	-
	Точность	0,469	-	0,875	-
	95%-й ДИ	0,391–0,546	-	0,803–0,947	-
	Чувствительность	0,957	-	1	-
	95%-й ДИ	0,873–1,000	-	1,000–1,000	-
	Специфичность	0,387	-	0,787	-
	95%-й ДИ	0,305–0,468	-	0,670–0,904	-

При переходе к работе с реальными клиническими данными ИИ-1 – лидер ретроспективного этапа – продемонстрировал выраженное ухудшение показателей точности. Площадь под характеристической кривой значимо

уменьшилась на 28,0 % (z-критерий Фишера 11,6847, $p < 0,0001$), общая точность – на 37,4 % (t-критерий -98,009, $p < 0,0001$). Критично снизилась специфичность – на 55,8 % (t-критерий -107,916, $p < 0,0001$). Незначительно уменьшилась лишь чувствительность. При оценке клинической результативности в 68,0 % зафиксировано несогласие врачей-экспертов с локализацией патологических находок, в 2,0 % – ложноположительные результаты. Столь низкие показатели потребовали технической доработки ИИ-сервиса, которая оказалась довольно результативной. В течение двух периодов наблюдения все показатели точности фактически сравнялась со значениями, полученными при ретроспективном тестировании ($p = 0,7508, 0,0190, 1,000$). Тем не менее сохранилась проблема с точностью маркировки обнаруженной патологии – в 68,7 % случаев были отмечены соответствующие дефекты. Значительно возрос удельный вес случаев гипердиагностики (во втором периоде – 50,0 %, в третьем – 12,5 %). Впрочем, положительным моментом было отсутствие гиподиагностики. Таким образом, ИИ-1 при переходе к проспективному формату работы продемонстрировал критичное ухудшение точности и качества; после доработки значения показателей возвратились к исходным, достаточным для клинического применения данного ИИ-сервиса.

Схожая ситуация наблюдалась у ИИ-3. В первом же периоде AUROC значимо снизилась на 24,5 % (z-критерий Фишера 4,9714, $p < 0,0001$). Однако затем показатель почти вернулся к исходному значению ($p = 0,9273$). Такой же была динамика общей точности: падение на 47,3 % (t-критерий -76,171, $p < 0,0001$) с последующим возвратом к приемлемому уровню (хотя и ниже исходного). Специфичность снизилась критично, на 52,2 % (t-критерий -58,054, $p < 0,0001$), однако, в последующем ее значение не только возобновилось, но и возросло на 2,8 % от исходного (t-критерий -2,637, $p = 0,0091$). Падение показателей в первом периоде сопровождалось закономерным ростом удельного веса гиподиагностики – 16,3 %, что, впрочем, было полностью устранено в следующем периоде. Также характерной чертой ИИ-3, как и ИИ-1, было постоянное наличие ложноположительных решений (8,8–15,0 %). В целом, менее

чем в половине случаев (47,5–48,8 %) врачи-эксперты выражали полное согласие с решениями этого сервиса.

Надо отметить, что первоначальное падение точности было столь значительным, что работа ИИ-3 на потоке данных была приостановлена. Сервис вернулся к проспективному формату лишь после существенной технической доработки.

При переходе к работе с реальными клиническими данными ИИ-2 продемонстрировал хорошую устойчивость. Значение AUROC фактически осталось на исходном уровне ($p = 0,9177$), а к третьему периоду наблюдений, хотя и статистически незначимо, но все же возросло на 3,8 %. Сразу обозначился существенный рост чувствительности: в первом периоде наблюдения – на 8,1 % (t-критерий 8,600, $p < 0,0001$), в заключительном – на 13,4 % (t-критерий 14,816, $p < 0,0001$). На этом фоне несколько снизилась специфичность; хотя максимальное снижение составило лишь 3,7 %, различия имели статистическую значимость (t-критерий -7,937, $p < 0,0001$). На 3,0–3,9 % статистически значимо возрастала общая точность, начиная с первого же периода наблюдения (t-критерий 9,752, $p < 0,0001$). Вместе с тем, как ни парадоксально, анализ клинической результативности выявил негативную картину.

Гиподиагностика все время была характерным признаком ИИ-2. В первом периоде наблюдения удельный вес соответствующих ложноотрицательных результатов был 4,4 %, во втором – 14,3 %, в третьем – 8,1 %. Чрезвычайно высоким было количество исследований с некорректной локализацией и маркировкой патологических находок (от 21,8 до 81,1 % случаев). Факт корректного обнаружения патологии в большинстве ситуаций (значения показателей точности) сочетался с неправильной маркировкой и измерениями, а также пропусками патологии (результаты оценки клинической результативности).

Таким образом, при автоматизированном анализе результатов КТ органов грудной клетки с целью выявления гидроторакса точность ИИ-сервисов достаточно высока (на потоке данных максимальная AUROC достигает 0,985, чувствительность – 1,0, специфичность – 0,970). Вместе с тем серьезным

недостатком соответствующих моделей ИИ является стабильное наличие ложноположительных результатов (в среднем – 13,8 % случаев относятся к гипердиагностике).

Полученные данные позволяют утверждать, что технологии ИИ могут применяться для автоматизированного анализа результатов КТ органов грудной клетки с целью выявления гидроторакса; потенциально – в качестве системы поддержки принятия врачебных решений и для морфометрии (определение объема плеврального выпота). С прикладной точки зрения, модели, лежащие в основе ИИ-сервисов, отличаются нестабильностью, тенденцией к гипердиагностике.

В контексте автоматизированного анализа результатов КТ органов грудной клетки с целью диагностики болезней системы кровообращения была сформулирована следующая клиническая задача – определение расширения и количественное определение диаметра легочного ствола:

- наличие на нативных изображениях расширения диаметра легочного ствола ≥ 29 мм;
- локализация найденных признаков патологии;
- значение наибольшего диаметра легочного ствола (мм).

Ретроспективное исследование диагностической точности четырех¹⁰ ИИ-сервисов для автоматизированного анализа результатов КТ органов грудной клетки с целью выявления расширения легочного ствола выполнено на эталонном наборе данных ($n = 100$).

Построены характеристические кривые, значения показателей сведены в таблицу (таблица 3.23, рисунок 3.18).

¹⁰ ООО «АЙРА Лабс», Pulm Trunk-IRA; ООО «Интел Диагностик», система поддержки принятия диагностических решений «АИ Диагностик»; ООО «Платформа Третье Мнение», программный модуль для анализа исследований компьютерной томографии органов грудной клетки; ООО «СиВижинЛаб», CVL – Chest CT Pulmonary trunk.

Таблица 3.23 – Показатели диагностической точности ИИ-сервисов для анализа результатов КТ органов грудной клетки с целью выявления расширения легочного ствола, полученные при ретроспективном тестировании на эталонном наборе данных

Показатели	ИИ-сервис			
	ИИ-1	ИИ-2	ИИ-3	ИИ-4
AUROC	0,990	0,990	0,986	0,868
95%-й ДИ	0,990–1,000	0,970–1,000	0,971–1,000	0,804–0,931
Точность	0,970	0,990	0,940	0,870
95%-й ДИ	0,930–1,000	0,97–1,0	0,890–0,990	0,8–0,94
Чувствительность	0,940	1,000	0,920	0,760
95%-й ДИ	0,870–1,000	1,0–1,0	0,840–1,000	0,63–0,88
Специфичность	1,000	0,980	0,960	0,980
95%-й ДИ	1,000–1,000	0,94–1,0	0,910–1,000	0,94–1,0

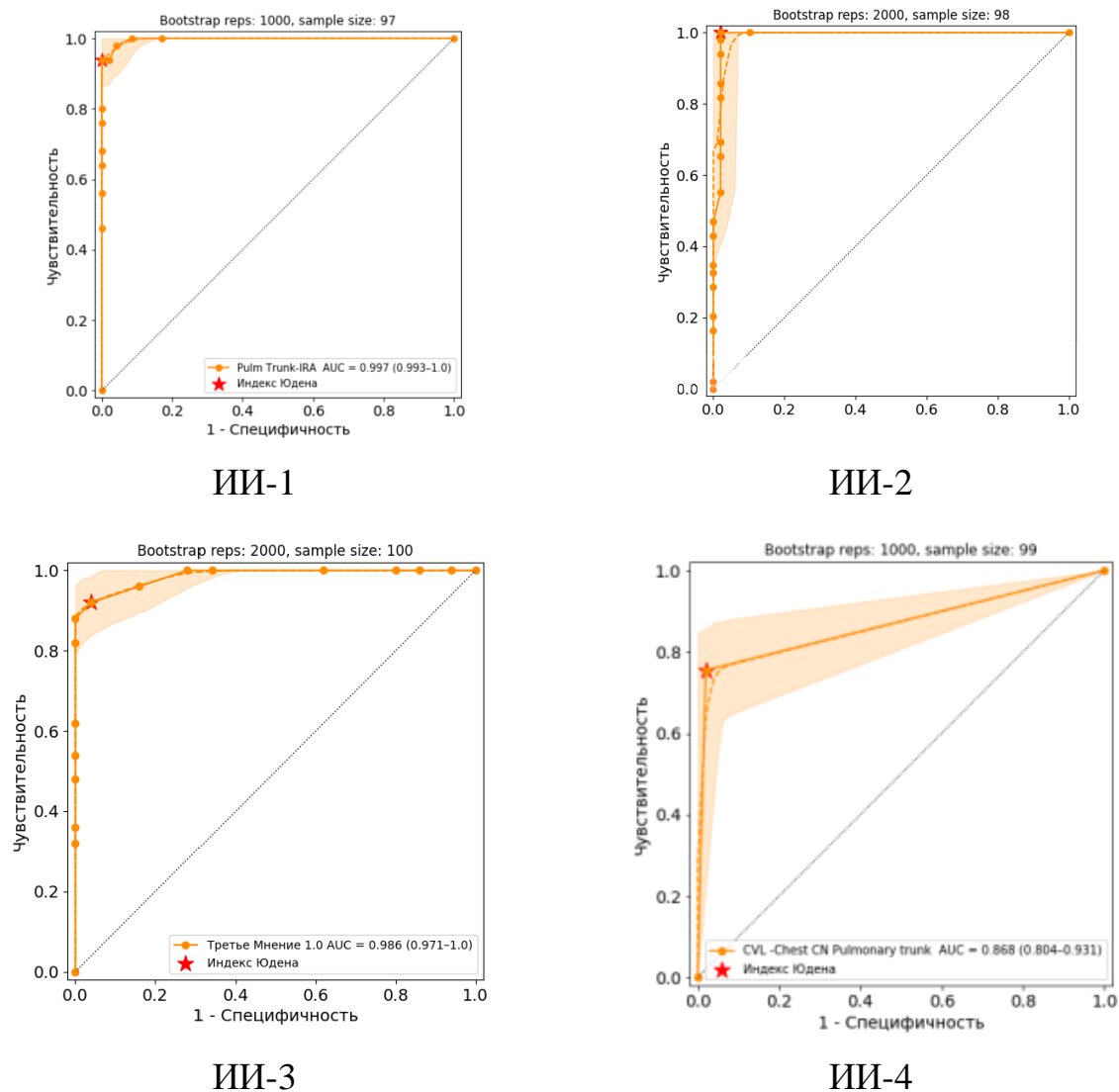


Рисунок 3.18 – Характеристические кривые для ИИ-сервисов для анализа результатов КТ органов грудной клетки с целью выявления расширения легочного ствола (оптимальный порог по индексу Юдена)

При тестировании на новых данных ИИ-сервисы показали достаточно высокие показатели точности. Максимальные значения AUROC (0,99) зафиксированы у ИИ-1 и ИИ-2, минимальное (0,868; $p = 0,0002, 0,0004$) – у ИИ-4. Также у этого сервиса были минимальны значения точности и чувствительности. Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.19).

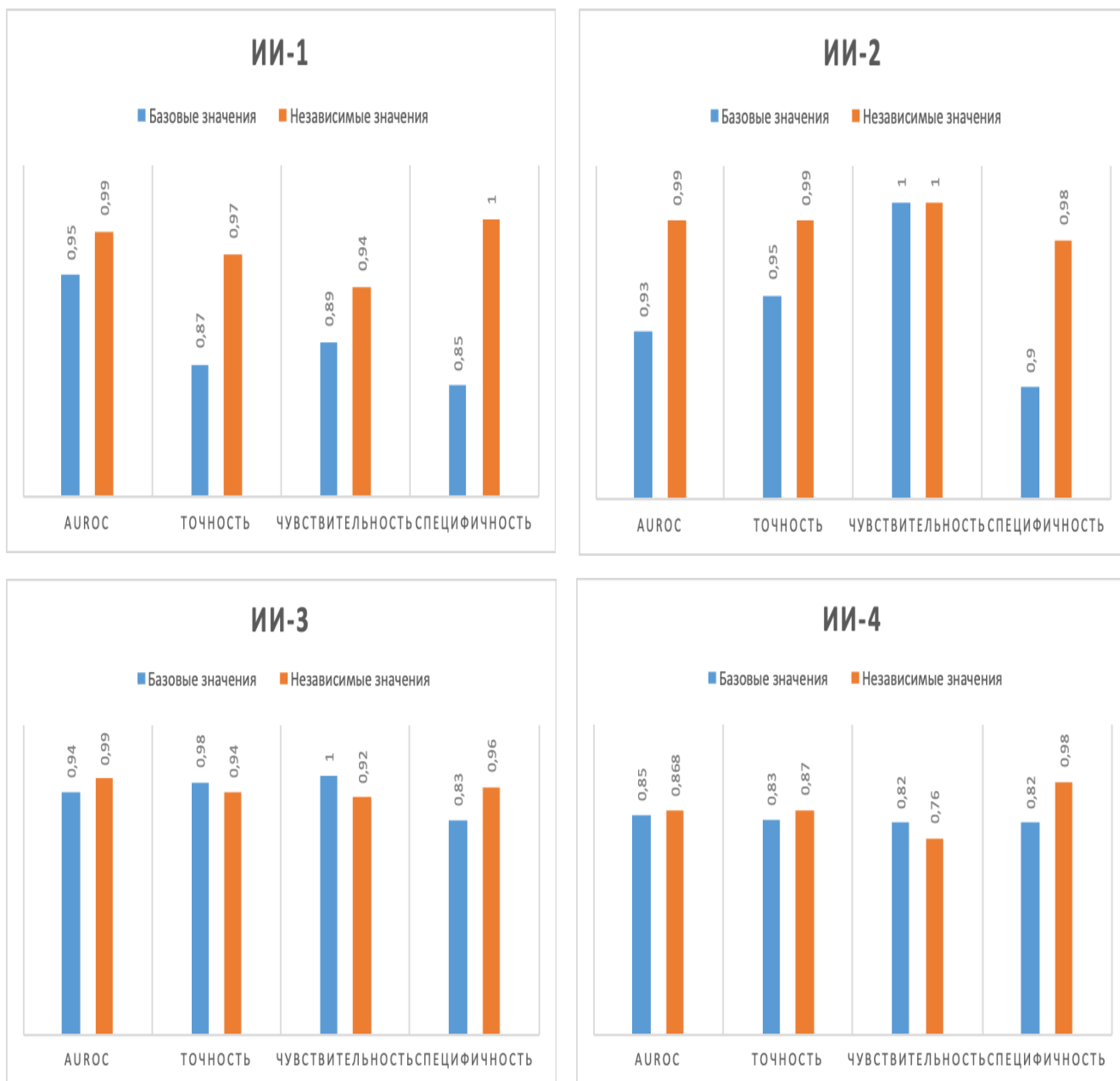


Рисунок 3.19 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов КТ органов грудной клетки с целью выявления расширения легочного ствола

ИИ-1 продемонстрировал рост показателей точности по отношению к заявленным. Наиболее значительно приросли точность (+11,5 %) и специфичность (+17,7 %). Схожая динамика зафиксирована у ИИ-2, здесь максимальный прирост произошел у специфичности (+8,9 %). ИИ-3 и ИИ-4 объединяли колебания трех показателей и существенный прирост именно специфичности – на 15,6 % и 19,5 % соответственно. Но также оба этих сервиса характеризовались очень высоким уровнем пропусков патологии. Удельный вес

ложноотрицательных результатов составил 8,0 % и 24,0 % соответственно. На этом фоне у ИИ-1 отмечено 6,0 % подобных результатов, а у ИИ-2 они вовсе отсутствовали. Ложноположительные результаты зафиксированы у всех сервисов, за исключением ИИ-1. Максимальный уровень был у ИИ-3 (4,0 %).

Все полученные показатели AUROC превышали пороговое значение для данного вида исследований, поэтому все ИИ-сервисы были допущены к работе с реальными клиническими данными и, соответственно, ко второму этапу оценки диагностической точности.

Проспективное исследование выполнено в условиях работы ИИ-сервисов с потоком результатов лучевых исследований, проведенных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены одно или два измерения точности в проспективном формате; результаты представлены в таблице 3.24.

Таблица 3.24 – Показатели диагностической точности ИИ-сервиса для анализа результатов КТ органов грудной клетки с целью выявления расширения легочного ствола, полученные при проспективном наблюдении

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.
ИИ-1	AUROC	0,948	-	-	-
	95%-й ДИ	0,903–0,993	-	-	-
	Точность	0,938	-	-	-
	95%-й ДИ	0,900–0,975	-	-	-
	Чувствительность	0,940	-	-	-
	95%-й ДИ	0,897–0,983	-	-	-
	Специфичность	0,930	-	-	-
	95%-й ДИ	0,854–1,000	-	-	-
ИИ-2	AUROC	-	-	0,878	0,858
	95%-й ДИ	-	-	0,823–0,935	0,726–0,985
	Точность	-	-	0,875	0,877
	95%-й ДИ	-	-	0,833–0,917	0,818–0,957
	Чувствительность	-	-	0,902	0,935
	95%-й ДИ	-	-	0,857–0,946	0,874–0,997
	Специфичность	-	-	0,806	0,722
	95%-й ДИ	-	-	0,711–0,901	0,515–0,929
ИИ-3	AUROC	-	-	0,823	-
	95%-й ДИ	-	-	0,752–0,893	-
	Точность	-	-	0,817	-
	95%-й ДИ	-	-	0,768–0,866	-
	Чувствительность	-	-	0,862	-
	95%-й ДИ	-	-	0,812–0,912	-
	Специфичность	-	-	0,678	-
	95%-й ДИ	-	-	0,559–0,797	-
ИИ-4	AUROC	0,719	0,794	-	-
	95%-й ДИ	0,628–0,814	0,726–0,862	-	-
	Точность	0,762	0,85	-	-
	95%-й ДИ	0,669–0,856	0,795–0,905	-	-

Продолжение таблицы 3.24

ИИ-сервис	Показатели	Период			
		I кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.
	Чувствительность	0,938	0,954	-	-
	95%-й ДИ	0,869–1,000	0,914–0,993	-	-
	Специфичность	0,5	0,635	-	-
	95%-й ДИ	0,327–0,673	0,504–0,765	-	-

В условиях работы с реальными клиническими данными ИИ-1 продемонстрировал устойчивость: AUROC фактически не изменилась ($p = 0,9921$), на прежнем уровне осталась чувствительность. Вместе с тем на 7,0 % снизилась специфичность (t-критерий -16,801, $p < 0,0001$), на 3,3 % – точность (t-критерий -11,484, $p < 0,0001$). На этом фоне, при оценке клинической результативности, не отмечено случаев гипердиагностики, но в 26,3 % локализация патологических находок на изображении была некорректной. Также в 1,3 % случаев выявлены пропуски патологии (ложноотрицательные решения).

В первом временном периоде, в условиях работы с реальными клиническими данными ИИ-2 для анализа результатов КТ органов грудной клетки с целью выявления расширения легочного ствола продемонстрировал ухудшение точности: показатель площади под характеристической кривой снизился на 11,3 % (z-критерий Фишера 3,7866, $p = 0,0002$). Во втором периоде ситуация еще более усугубилась, достоверное снижение достигло 13,3 % (z-критерий Фишера 1,9846, $p = 0,0472$). В обоих периодах также наблюдалось снижение точности примерно на 11,0 %. Статистически значимо снизилась чувствительность на 9,8 % (t-критерий -38,585, $p < 0,0001$). Во втором периоде «разрыв» сократился до 6,5 %, но сохранил статистическую значимость ($p < 0,0001$).

Динамика специфичности имела негативный характер: в обоих периодах происходило постоянное снижение, сначала на 17,8 %, затем на 26,3 %

(t-критерии -30,886 и -21,657 соответственно, $p < 0,0001$). При проспективной оценке клинической результативности установлено, что удельный вес случаев с полным соответствием решений врачей-экспертов и ИИ-сервиса составил 62,9–78,3 %. При этом стабильно фиксировались случаи пропуска патологии – 1,5–1,6%, а также гипердиагностики – 10,2–14,5 %.

ИИ-3 продемонстрировал схожую с ИИ-1 динамику с падением специфичности на 29,4 %, точности – на 13,1%, AUROC – на 16,5 % ($p < 0,0001$). Уровень гипердиагностики достиг 15,8 %, а в 37,2 % случаев зафиксировано несогласие врачей-экспертов с локализацией и маркировкой патологии, выполненной ИИ-3.

При ретроспективном тестировании у ИИ-4 выявлены самые низкие значения показателей точности. В условиях реального потока клинических данных этот ИИ-сервис также продемонстрировал крайне негативную динамику. В первом временном периоде AUROC снизилась на 17,2 % (z-критерий Фишера 2,5934, $p = 0,0095$), точность – на 12,4 % (t-критерий -16,823, $p < 0,0001$). На этом фоне чувствительность продемонстрировала достоверный рост на 23,4–25,5 % (t-критерий 24,091, $p < 0,0001$). Однако падение специфичности носило критичный характер: снижение составило 49,9 % (t-критерий -48,040, $p < 0,0001$) в первом периоде и 35,2 % (t-критерий -45,373, $p < 0,0001$) во втором.

Именно у ИИ-4 при проспективной оценке клинической результативности выявлен наивысший удельный вес пропуска патологии – до 4,9 %, также гипердиагностики – от 9,5 до 24,2 %. Некорректную локализацию патологии врачи-эксперты отметили в 16,1–20,6 % случаев.

Таким образом, при автоматизированном анализе результатов КТ органов грудной клетки с целью выявления расширения легочного ствола уровень точности, сопоставимый с врачом-рентгенологом, в настоящее время не достижим. Модели, лежащие в основе ИИ-сервиса, отличаются низкой точностью, нестабильностью, высоким уровнем гипердиагностики и ошибками в локализации патологических находок.

Компьютерная томография органов грудной клетки и органов брюшной полости. В контексте автоматизированного анализа результатов КТ органов грудной и/или брюшной полости с целью диагностики аневризмы аорты была сформулирована следующая клиническая задача – определение на нативных изображениях расширения восходящего и нисходящего отделов грудной аорты, брюшного отдела аорты:

- выявление патологических признаков: наибольший диаметр восходящей части грудной аорты от 40 до 49 мм включительно в аксиальной плоскости считать дилатацией, ≥ 50 мм считать аневризмой; наибольший диаметр нисходящей части грудной аорты ≥ 40 мм в аксиальной плоскости считать аневризмой; наибольший диаметр брюшного отдела аорты от 25 до 29 мм включительно считать дилатацией, ≥ 30 мм считать аневризмой брюшного отдела аорты;
- локализация найденных признаков патологии;
- измерение диаметров восходящего и нисходящего отделов аорты на каждом срезе, брюшной аорты в аксиальной плоскости на каждом срезе (мм);
- указание наибольших диаметров восходящего и нисходящего отделов аорты, брюшного отдела аорты (мм).

На первом этапе выполнено ретроспективное исследование диагностической точности пяти¹¹ ИИ-сервисов для анализа результатов КТ органов грудной ($n = 3$) и брюшной ($n = 2$) полостей с целью диагностики аневризмы аорты (далее – в подразделе ИИ-1, ИИ-2, ИИ-3, ИИ-4, ИИ-5). Исследование выполнено на эталонном наборе данных ($n = 100$). Построены характеристические кривые, полученные показатели точности сведены в таблицу (таблица 3.25, рисунок 3.20). Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.21).

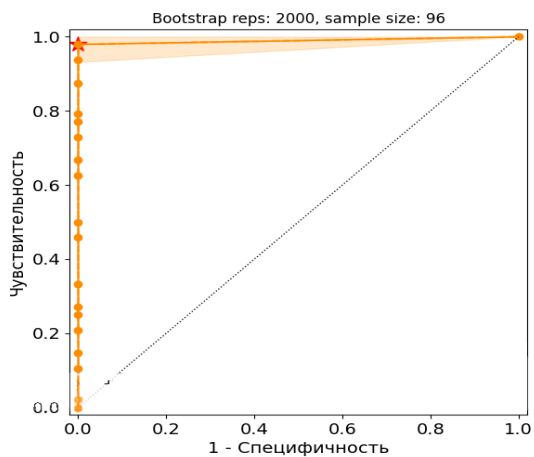
¹¹ ООО «Интел Диагностик», система поддержки принятия диагностических решений «АИ Диагностик»; ООО «Платформа Третье Мнение», программный модуль для анализа исследований компьютерной томографии органов грудной клетки; ООО «СиВижинЛаб», CVL – Abdomen CT Aorta, CVL – Chest CT Aorta; ООО «АЙРА Лабс», Aorta-IRA.

Значения всех показателей всех ИИ-сервисов отличались очень высоким уровнем, так как превышали 0,9. Максимальное значение площади под характеристической кривой отмечено у ИИ-5 – 1,0 (95%-й ДИ 0,999–1,0), минимальное – 0,92 (95%-й ДИ 0,86–0,97) у ИИ-4.

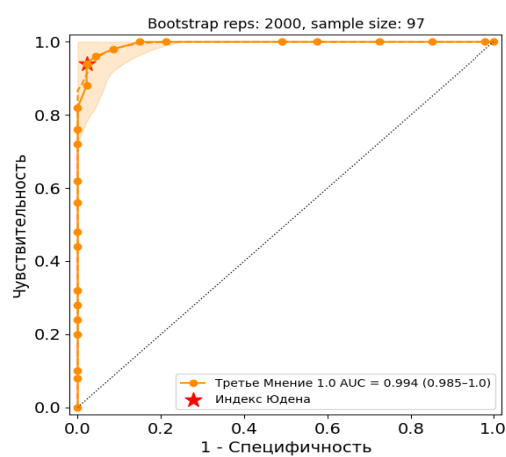
Максимальные точность (0,99, 95%-й ДИ 0,970–1,000), чувствительность (0,98, 95%-й ДИ 0,940–1,000) и специфичность (1, 0,95%-й ДИ 1,0–1,0) зафиксирована у ИИ-1 и ИИ-5. Минимальные значения трех перечисленных показателей зафиксированы у ИИ-4, а чувствительности также у ИИ-2; они составили 0,92 (95%-й ДИ 0,86–0,97), 0,94 (95%-й ДИ 0,870–1,000) и 0,9 (95%-й ДИ 0,81–0,98) соответственно.

Таблица 3.25 – Показатели диагностической точности ИИ-сервисов для анализа результатов КТ органов грудной и/или брюшной полости с целью выявления аневризмы, полученные при ретроспективном тестировании на эталонном наборе данных

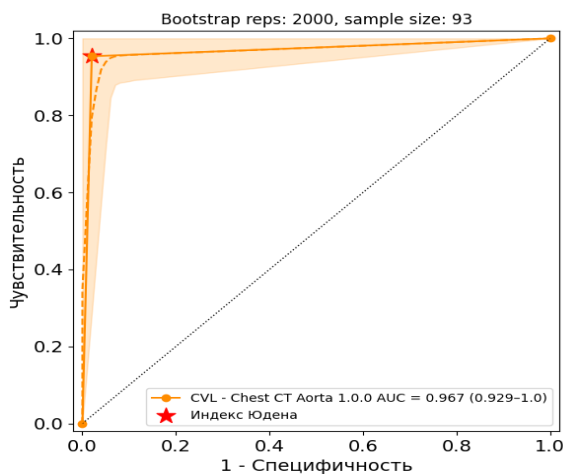
Показатели	ИИ-сервис				
	ИИ-1	ИИ-2	ИИ-3	ИИ-4	ИИ-5
AUROC	0,99	0,994	0,967	0,92	1,000
95%-й ДИ	0,969–1,000	0,985–1,000	0,929–1,000	0,86–0,97	0,999–1,0
Точность	0,99	0,96	0,97	0,92	0,99
95%-й ДИ	0,970–1,000	0,920–1,000	0,93–1,0	0,86–0,97	0,97–1,0
Чувствительность	0,98	0,94	0,95	0,94	0,980
95%-й ДИ	0,940–1,000	0,870–1,000	0,89–1,0	0,87–1,0	0,94–1,0
Специфичность	1,0	0,98	0,98	0,9	1,0
95%-й ДИ	1,0–1,0	0,940–1,000	0,94–1,0	0,81–0,98	1,0–1,0



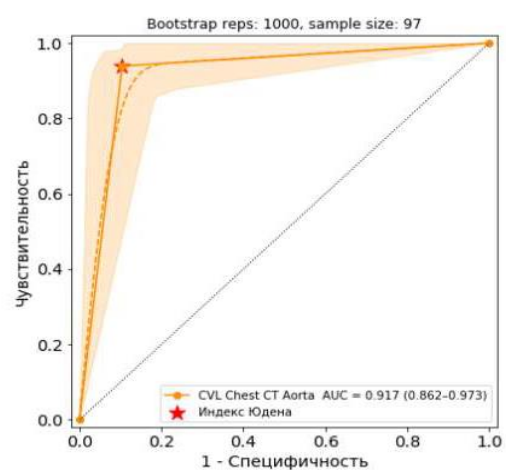
ИИ-1



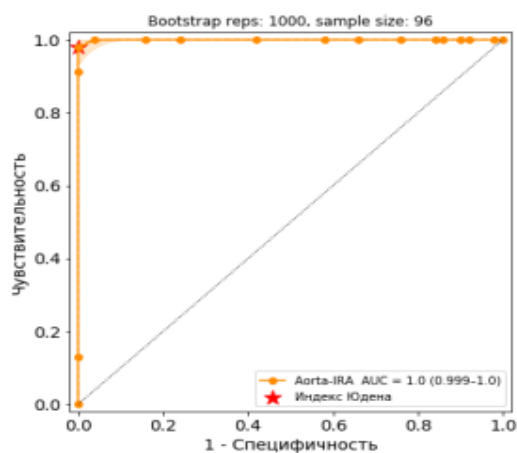
ИИ-2



ИИ-3



ИИ-4



ИИ-5

Рисунок 3.20 – Характеристические кривые ИИ-сервисов для анализа результатов КТ органов грудной и/или брюшной полости с целью выявления аневризмы (оптимальный порог по индексу Юдена)

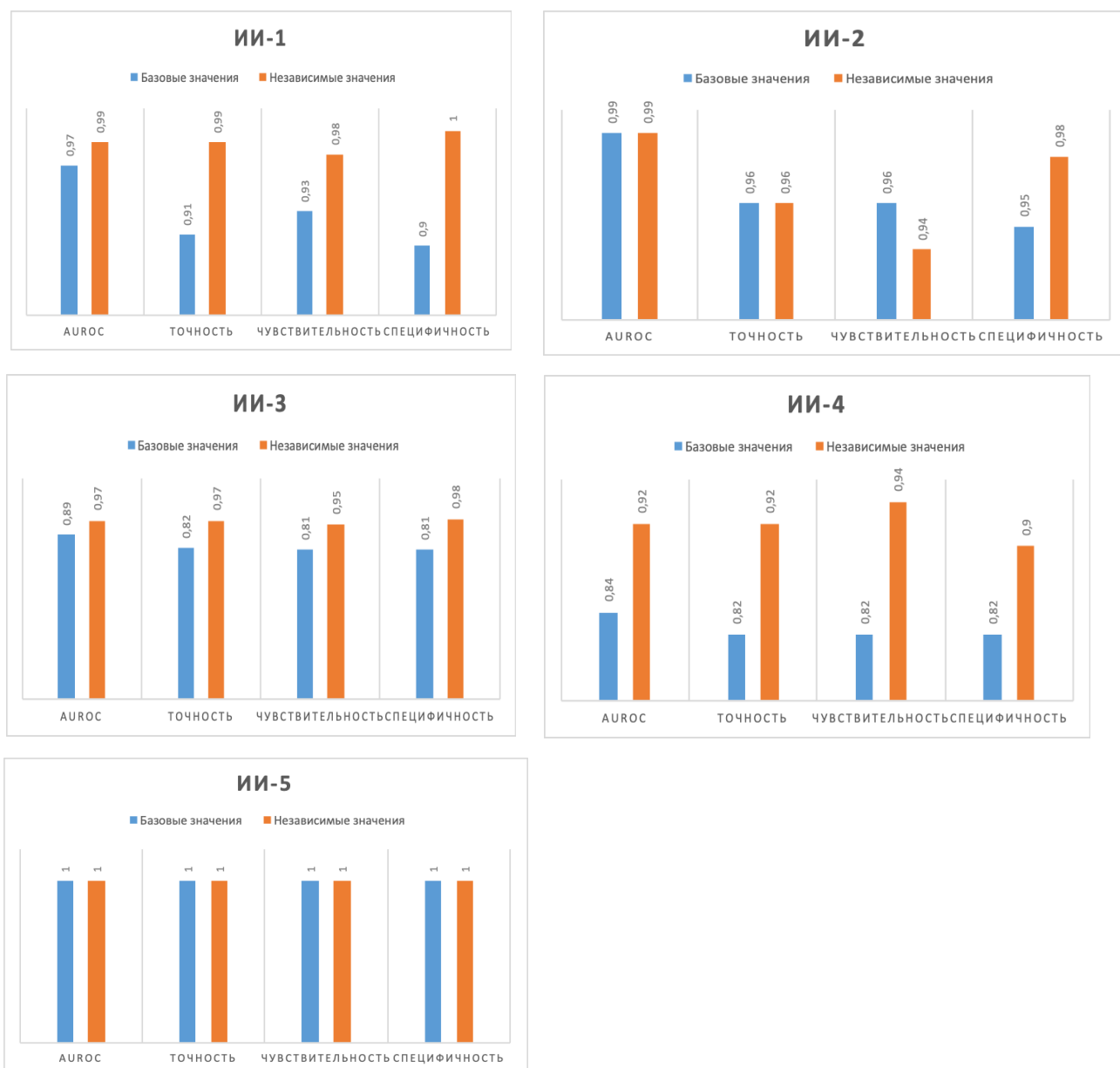


Рисунок 3.21 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов КТ органов грудной и/или брюшной полости с целью выявления аневризмы

При тестировании на эталонном наборе данных фактически все ИИ-сервисы продемонстрировали достаточно хорошую воспроизводимость результатов своей работы. Кроме того, для ИИ-1, ИИ-3 и ИИ-4 были получены даже более высокие показатели точности, чем ранее установленные производителем. Так, значение точности ИИ-1 возросло на 8,8 %, а специфичности – на 10,0 %. Точность ИИ-3 повысилась на 18,3 %, чувствительность – на 17,3 %, специфичность – на целых 21,0 %. Площадь под характеристической кривой ИИ-4 возросла на 9,5 %, почти

на столько же увеличилась специфичность, а точность и чувствительность возросли на 12,2 и 14,6 % соответственно. Вместе с тем ИИ-1 и особенно ИИ-3 отличались уровнем гиподиагностики (то есть ложноотрицательных решений). Соответствующие значения составили 2,1 % и 4,7 % соответственно. На этом фоне удельный вес ложноотрицательных ответов ИИ-4 составил лишь 6,0 %, а ложноположительных – 10,0 %. Стабильным оказался ИИ-2 (колебания чувствительности и специфичности в пределах 2,0–3,0 %). Вместе с тем он оказался «лидером» по уровню ложноотрицательных результатов – 6,0 %. При этом уровень гипердиагностики достиг 2,1 %. Максимально стабильным оказался ИИ-5. При тестировании на новых данных отмечено лишь незначительное снижение точности и чувствительности на 1,0 и 2,0 % (при том что заявленные значения всех показателей равнялись 1,0). Недостатком этого сервиса явились 2 случая ложноотрицательных результатов (на фоне нуля случаев гипердиагностики).

Таким образом, при тестировании на новых данных ИИ-сервисы для анализа результатов КТ органов грудной и/или брюшной полости с целью диагностики аневризмы аорты показали не только стабильность, но и положительную динамику. Уровень гиподиагностики составил порядка 6,0 %.

Вторым этапом изучения диагностической точности ИИ-сервисов для выявления аневризмы аорты на результатах КТ органов грудной и/или брюшной полости стало проспективное исследование в условиях работы ИИ с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены от двух до четырех измерений точности в проспективном формате; результаты представлены в таблице 3.26.

Таблица 3.26 – Показатели диагностической точности ИИ-сервиса для анализа результатов КТ органов грудной и/или брюшной полости с целью выявления аневризмы, полученные при проспективном наблюдении

ИИ-сервис	Показатели	Период			
		III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
ИИ-1	AUROC	-	0,691	0,812	-
	95%-й ДИ	-	0,622–0,758	0,718–0,902	-
	Точность	-	0,547	0,787	-
	95%-й ДИ	-	0,485–0,610	0,698–0,877	-
	Чувствительность	-	0,928	0,694	-
	95%-й ДИ	-	0,866–0,989	0,544–0,845	-
	Специфичность	-	0,397	0,864	-
	95%-й ДИ	-	0,324–0,469	0,762–0,965	-
ИИ-2	AUROC	-	0,81	0,905	-
	95%-й ДИ	-	0,758–0,863	0,818–0,989	-
	Точность	-	0,758	0,9	-
	95%-й ДИ	-	0,704–0,812	0,834–0,966	-
	Чувствительность	-	0,97	0,906	-
	95%-й ДИ	-	0,937–1,000	0,827–0,984	-
	Специфичность	-	0,604	0,889	-
	95%-й ДИ	-	0,523–0,686	0,770–1,000	-
ИИ-3	AUROC	-	-	0,738	-
	95%-й ДИ	-	-	0,688–0,785	-
	Точность	-	-	0,637	-
	95%-й ДИ	-	-	0,585–0,690	-
	Чувствительность	-	-	0,9	-
	95%-й ДИ	-	-	0,824–0,976	-
	Специфичность	-	-	0,577	-

Продолжение таблицы 3.26

ИИ-сервис	Показатели	Период			
		III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
	95%-й ДИ	-	-	0,517–0,637	-
ИИ-4	AUROC	0,924	0,914	0,993	-
	95%-й ДИ	0,887–0,960	0,854–0,971	0,978–1,007	-
	Точность	0,938	0,912	0,988	-
	95%-й ДИ	0,907–0,968	0,851–0,974	0,963–1,000	-
	Чувствительность	0,993	0,854	1	-
	95%-й ДИ	0,979–1,000	0,745–0,962	1,0–1,0	-
	Специфичность	0,854	0,974	0,985	-
	95%-й ДИ	0,784–0,925	0,925–1,000	0,957–1,000	-
ИИ-5	AUROC	0,823	0,854	0,868	0,763
	95%-й ДИ	0,761–0,887	0,777–0,927	0,808–0,926	0,707–0,818
	Точность	0,725	0,909	0,921	0,742
	95%-й ДИ	0,656–0,794	0,864–0,953	0,888–0,955	0,686–0,797
	Чувствительность	0,984	0,979	0,993	0,846
	95%-й ДИ	0,953–1,000	0,950–1,000	0,979–1,000	0,748–0,944
	Специфичность	0,557	0,814	0,818	0,713
	95%-й ДИ	0,458–0,656	0,723–0,905	0,742–0,894	0,648–0,777

При переходе к работе в реальных клинических условиях ИИ-1, ИИ-2 и ИИ-3 продемонстрировали значимое ухудшение показателей точности. AUROC у них снизилась на 30,0 % (z-критерий Фишера 8,4029, $p < 0,0001$), 18,5 % (z-критерий Фишера 6,5884, $p < 0,0001$) и 23,7 % (z-критерий Фишера 7,4679, $p < 0,0001$) соответственно. Столь же значительно упал уровень специфичности: у ИИ-1 – на целых 60,3 % (t-критерий -145,726, $p < 0,0001$), у ИИ-2 – на 38,4 % (t-критерий -76,788, $p < 0,0001$), у ИИ-3 – на 41,1 % (t-критерий -107,451, $p < 0,0001$). Также значительно ухудшился уровень общей точности.

В такой ситуации ИИ-3 вовсе прекратил работу на потоке лучевых исследований. ИИ-1 и ИИ-2 после технических доработок смогли повысить

уровень AUROC. Он все равно оставался достоверно ниже, чем полученный при ретроспективном тестировании (у ИИ-1 – на 18,0 % (z-критерий Фишера 3,7395, $p = 0,0002$), у ИИ-2 – на 9,0 % (z-критерий Фишера 2,0075, $p = 0,0447$). Однако позволял сервисам продолжить работу в проспективном формате. Также достоверно снизился разрыв для значений специфичности (до 13,6 % у ИИ-1 и 9,3 % у ИИ-2). Тем не менее в дальнейшем эти ИИ-сервисы прекратили работу на потоке по техническим причинам, не связанным с проведением нашего научного исследования. Дополнительно укажем, что ИИ-2 отличался высоким уровнем гиподиагностики (11,0 %), а ИИ-1 – гипердиагностики (23,0 %).

ИИ-4 успешно работал на потоке реальных данных, однако значения показателей точности постоянно колебались. При переходе на проспективный формат работы его AUROC осталась фактически на прежнем уровне ($p = 0,9055$), а в дальнейшем даже достоверно увеличилась на 7,9 % (z-критерий Фишера -2,5155, $p = 0,0119$). Аналогичная ситуация наблюдается для общей точности. Чувствительность же ИИ-4 в первом периоде наблюдения вовсе возросла на 5,6 % (t-критерий 15,738 $p < 0,0001$). Затем последовал достоверный спад, сменившийся уверенным подъемом до 1,0 (95%-й ДИ 1,0–1,0); различия с исходным уровнем, определенным ретроспективно, были значимы (t-критерий 18,105, $p < 0,0001$). Лишь показатель специфичности в первом периоде наблюдения, то есть при переходе к работе в реальных клинических условиях, достоверно снизился на 5,1 % (t-критерий -7,779, $p < 0,0001$). Однако в последующих периодах благодаря интенсивной доработке ИИ-сервиса специфичность не только восстановилась, но и превысила исходный уровень на 9,4 % (t-критерий 18,872, $p < 0,0001$).

При проспективной оценке клинической результативности ИИ-4 установлено, что удельный вес случаев с полным соответствием решений врачей-экспертов и ИИ колебался в пределах 74,0–98,0 %. В первом и втором периодах наблюдения фиксировались случаи гипердиагностики (1,0 % и 11,0 % соответственно), а вот в третьем отмечено 7,0 % ложноотрицательных результатов. Такая динамика подчеркивает общую нестабильность работы ИИ-4.

В целом отрицательная динамика отмечена для ИИ-5. При переходе на обработку потока реальных клинических данных показатель AUROC достоверно снизился, причем на 17,7 % (z-критерий Фишера 5,5068, $p < 0,0001$), а к заключительному периоду – на 23,7 % (z-критерий Фишера 8,3698, $p < 0,0001$). Столь негативным было и снижение чувствительности: в первом периоде – на 26,8 % (t-критерий -66,057, $p < 0,0001$), в заключительном – на 25,1 % (t-критерий -76,105, $p < 0,0001$). Критичным было снижение специфичности: на 44,3 % (t-критерий -78,402, $p < 0,0001$).

К заключительному периоду наблюдения «разрыв» сократился до 28,7 %, но все равно оставался критичным и статистически значимым (t-критерий -77,961, $p < 0,0001$). Лишь чувствительность вначале продемонстрировала относительную устойчивость. В первом периоде наблюдения она фактически не изменилась ($p = 0,06$), затем были отмечены колебания и выраженное снижение на 13,7 % от исходного уровня, определенного ретроспективно (t-критерий -23,109, $p < 0,0001$). Вместе с тем для ИИ-5 при проспективной оценке клинической результативности не зафиксировано ни одного случая ложноотрицательного результата. Стабильно отмечена гипердиагностика: от 2,7 до 8,6 % случаев. В целом же, полное соответствие определено для 67,3–93,1 % случаев.

Суммарно решение задачи анализа результатов КТ органов грудной и/или брюшной полости с целью выявления аневризмы оказалось невозможным для 3 ИИ-сервисов из 5. Два оставшихся на потоке исследований продемонстрировали выраженную нестабильность и общую негативную динамику.

Таким образом, при автоматизированном выявлении аневризмы аорты потенциально достижим достаточно высокий уровень точности (AUROC более 0,9), в целом сопоставимый с врачом-рентгенологом. Отмечается нестабильная и неэффективная работа конкретных программных продуктов на основе ИИ.

Полученные данные позволяют утверждать, что технологии ИИ потенциально могут применяться для автоматизированного анализа результатов

КТ органов грудной и/или брюшной полости с целью выявления аневризмы как СППВР и для морфометрии, однако, требуется принципиальное развитие программного обеспечения для устойчивого функционирования в реальных клинических условиях.

В контексте автоматизированного анализа результатов КТ органов грудной или брюшной полости целью диагностики злокачественных новообразований была сформулирована следующая клиническая задача – выявление образований надпочечников:

- измерение толщины тела и ножек надпочечников, мм;
- наличие образования тела или ножек надпочечника с размером по короткой оси ≥ 10 мм на нативных изображениях;
- определение локализации образований надпочечников;
- измерение аксиального размера по короткой оси наибольшего образования надпочечников (при наличии), мм.

Ретроспективное исследование диагностической точности одного¹² ИИ-сервиса для автоматизированного анализа результатов КТ ОГК/ОБП с целью выявления новообразований надпочечников выполнено на эталонном наборе данных ($n = 100$). Построена характеристическая кривая (рисунок 3.22). Значение площади под характеристической кривой составило 0,995 (95%-й ДИ 0,986–1,000). Определены показатели: общая точность составила 0,980 (95%-й ДИ 0,95–1,0), чувствительность – 1,0 (95%-й ДИ 1,0–1,0), специфичность – 0,960 (95%-й ДИ 0,9–1,0). Ложноотрицательных результатов не выявлено. Удельный вес гипердиагностики составил 4,2 %.

Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.23).

¹² ООО «АЙРА Лабс», Adrenal-IRA.

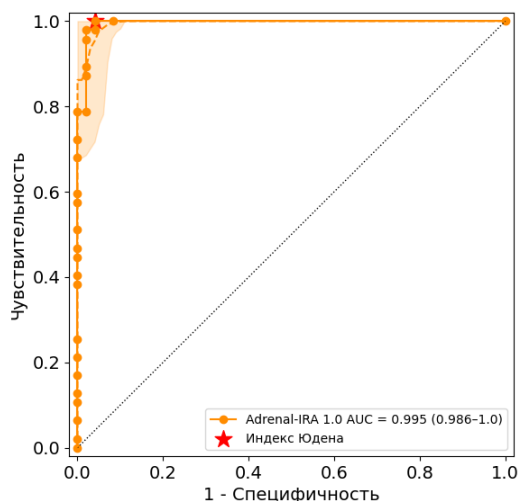


Рисунок 3.22 – Характеристическая кривая для ИИ-сервиса для анализа результатов КТ ОГК/ОБП с целью выявления новообразований надпочечников (оптимальный порог по индексу Юдена)

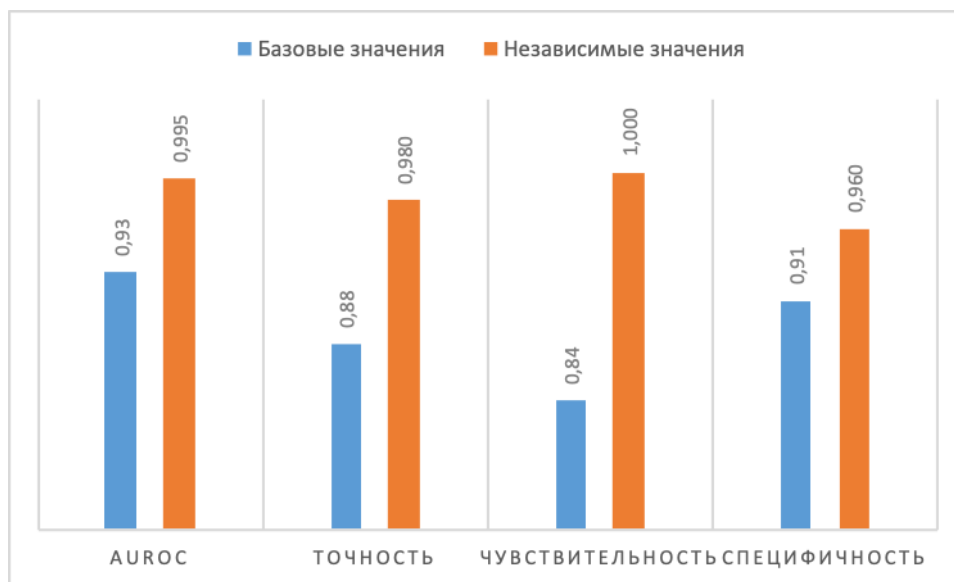


Рисунок 3.23 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов КТ ОГК/ОБП с целью выявления новообразований надпочечников

Обращает на себя внимание выраженный рост значений показателей точности, полученных в ходе независимого ретроспективного тестирования на новых данных. В частности, относительно базового значение площади под характеристической кривой увеличилось на 7,0 %, общей точности – на 11,4 %, специфичности – на 5,5 %; особенно возросла чувствительность – на целых 19,1 %.

На втором этапе проведено проспективное исследование в условиях работы ИИ-сервиса с потоком результатов лучевых исследований, выполненных

в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены три измерения точности в проспективном формате; результаты представлены в таблице 3.27.

Таблица 3.27 – Показатели диагностической точности ИИ-сервиса для анализа результатов КТ ОГК/ОБП с целью выявления новообразований надпочечников, полученные при проспективном наблюдении

Показатели	Период		
	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
AUROC	0,952	0,809	0,893
95%-й ДИ	0,920–0,984	0,752–0,864	0,846–0,940
Точность	0,95	0,779	0,892
95%-й ДИ	0,923–0,978	0,727–0,832	0,853–0,931
Чувствительность	0,98	0,944	0,992
95%-й ДИ	0,957–1,000	0,904–0,984	0,978–1,000
Специфичность	0,904	0,596	0,771
95%-й ДИ	0,845–0,964	0,506–0,687	0,692–0,850

При переходе к проспективному формату работы ИИ-сервис для анализа результатов КТ ОГК/ОБП с целью выявления новообразований надпочечников продемонстрировал хорошую воспроизводимость результатов своей работы на новых данных. Вместе с тем в первом временном периоде отмечено некоторое снижение показателей точности. Различия между значениями AUROC при ретроспективном и первом проспективном наблюдениях (снижение на 4,3 %) носили статистически значимый характер (z-критерий Фишера 2,5730, $p = 0,0101$). Точность снизилась на 3,1 %, чувствительность – на 2,0 %, специфичность – на 5,8 % относительно уровня, полученного при ретроспективной оценке. Все различия носили достоверный характер (t-критерий -14,997, 18,245, -13,446 соответственно, $p < 0,0001$). В последнем периоде наблюдения «разрыв» с исходным уровнем возрос: для AUROC до 10,3 % (z-критерий Фишера 4,2073, $p < 0,0001$), точности – до 9,0% (t-критерий -35,957, $p < 0,0001$). Наибольшим было

уменьшение специфичности – на 19,7 % (t-критерий -14,265, $p < 0,0001$); минимально оно было выражено у чувствительности – 0,8 % (t-критерий -38,288, $p < 0,0001$).

На фоне такой динамики обращает на себя внимание прогресс ИИ-сервиса в повышении клинической результативности. Если в первых двух периодах наблюдения удельный вес пропусков патологии и гипердиагностики колебался в диапазонах 3,9–9,5 % и 1,1–3,9 %, то в последнем периоде ложноотрицательных и ложноположительных результатов не было совсем. Удельный вес случаев с полным соответствием колебался от 70,7 % до 82,9 %.

Таким образом, при автоматизированном анализе результатов КТ ОГК/ОБП с целью выявления новообразований надпочечников потенциально достижим уровень точности, близкий к точности врача-рентгенолога (на потоке данных максимальная AUROC составляет 0,952, уровень гипо- и гипердиагностики может быть нулевым). С прикладной точки зрения, уровень функционирования конкретного программного решения должен быть повышен.

Полученные данные позволяют утверждать, что технологии ИИ могут применяться для автоматизированного анализа результатов КТ ОГК/ОБП с целью выявления новообразований надпочечников. В контексте данной клинической задачи ТИИ могут использоваться в качестве системы поддержки принятия врачебных решений и для морфометрии (измерения толщины тела и ножек надпочечников, размеров новообразований).

В контексте автоматизированного анализа результатов КТ органов грудной и/или брюшной полости с целью выявления признаков остеопороза была сформулирована следующая клиническая задача – определение наличия и локализации компрессионных переломов тел позвонков с деформацией более 25 % (по полуколичественной шкале Genant, 2–3 степень) и снижения минеральной плотности костной ткани в телах позвонков в интервале Th11–L3 (оптимально L1–L2) согласно критериям ACR 2018, позиции ISCD 2019 на нативных изображениях¹³:

¹³ Федеральные клинические рекомендации по диагностике, лечению и профилактике остеопороза / сост. Г. А. Мельниченко, Ж. Е. Белая, Л. Я. Рожинская [и др.] // Проблемы

- выявление и маркировка (нумерация) позвонков, имеющих деформацию $\geq 25\%$;
- графическое отображение высоты тела позвонка в передних, средних и задних отделах (контур) для всех проанализированных позвонков;
- числовое отображение степени компрессии позвонков в процентах (для позвонков с деформацией $\geq 25\%$) и указание степени по классификации Genant;
- опционально – измерение минеральной плотности (или NU) губчатого вещества тел позвонков Th12–L3 с указанием номера позвонка.

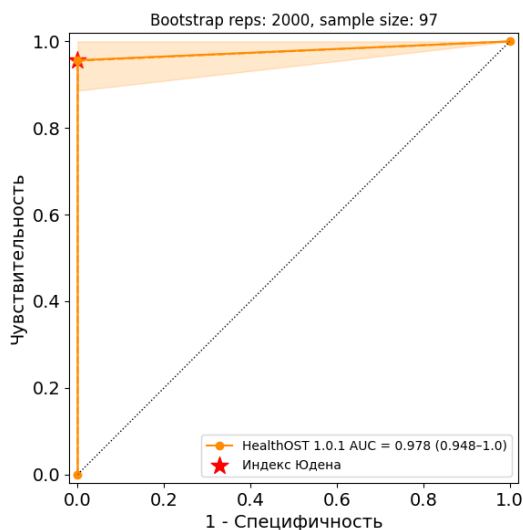
На первом этапе ретроспективное исследование диагностической точности четырех¹⁴ ИИ-сервисов для анализа результатов КТ органов грудной ($n = 1$) и брюшной ($n = 3$) полостей с целью выявления признаков остеопороза выполнено на эталонном наборе данных ($n = 100$). Построены характеристические кривые, показатели точности сведены в таблицу (таблица 3.28, рисунок 3.24).

Таблица 3.28 – Показатели диагностической точности ИИ-сервисов для анализа результатов КТ органов грудной и/или брюшной полости с целью выявления признаков остеопороза, полученные при ретроспективном тестировании на эталонном наборе данных

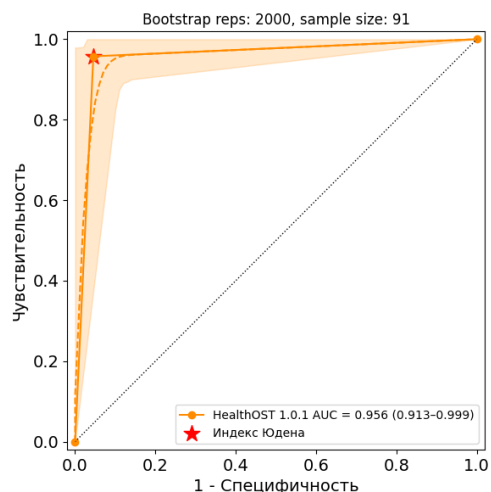
Показатели	ИИ-сервис			
	ИИ-1	ИИ-2	ИИ-3	ИИ-4
AUROC	0,978	0,910	0,982	0,978
95%-й ДИ	0,948–1,000	0,85–0,97	0,958–1,000	0,953–1,0
Точность	0,980	0,910	0,950	0,960
95%-й ДИ	0,95–1,0	0,83–0,96	0,91–0,99	0,92–1,0
Чувствительность	0,960	0,860	0,980	0,910
95%-й ДИ	0,9–1,0	0,72–0,95	0,969–1,000	0,84–0,99
Специфичность	1,000	0,950	0,920	1,000
95% ДИ	1,0–1,0	0,85–0,99	0,85–1,0	1,0–1,0

эндокринологии. – 2017. – Т. 63 – С. 392–426; Клинические рекомендации Министерства здравоохранения Российской Федерации «Патологические переломы, осложняющие остеопороз». – М., 2018. ID614. – Режим доступа: https://cr.minzdrav.gov.ru/schema/614_1.

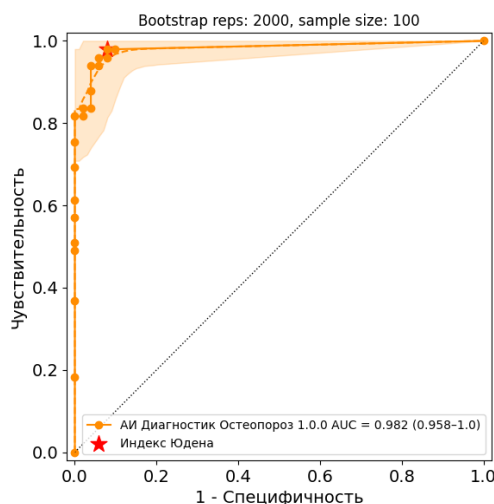
¹⁴ ООО «МИП «Биномикс-Рэй»; HealthOST, HealthVCF; ООО «Интел Диагнос



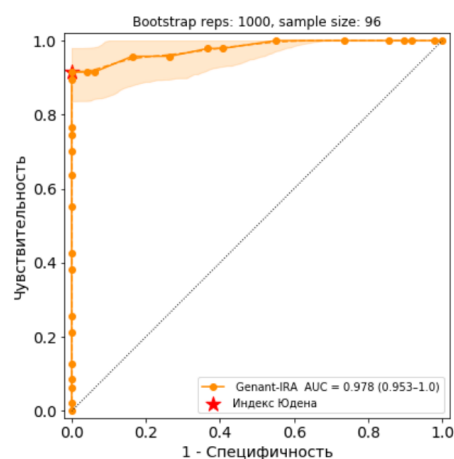
ИИ-1



ИИ-2



ИИ-3



ИИ-4

Рисунок 3.24 – Характеристические кривые ИИ-сервисов для анализа результатов КТ органов грудной и/или брюшной полости с целью выявления признаков остеопороза (оптимальный порог по индексу Юдена)

На данном этапе при попарном сравнении статистически значимо различались между собой значения AUROC ИИ-2 всех остальных сервисов (z-критерий Фишера 2,0382, 2,2200, 2,0684, $p = 0,0264$, 0,0386, 0,0415 соответственно). В остальных случаях значимых различий не было.

У ИИ-3 отмечено наивысшее значение AUROC и чувствительности. При этом удельный вес ложноотрицательных результатов составлял 2,1 %, а ложноположительных – 7,9 % (последнее было обусловлено самым низким значением специфичности в 0,920 (95%-й ДИ 0,85–1,0)). У ИИ-1 и ИИ-4 были

почти одинаковые AUROC, лишь немного «отстающие» от предыдущего сервиса. Вместе с тем по чувствительности и особенно по специфичности они явно превосходили ИИ-3. У ИИ-1 отмечено 4,4 % случаев гиподиагностики, у ИИ-4 – 9,0 %. У обоих сервисов не было случаев гипердиагностики, что логично следует из значений специфичности 1,0 (95%-й ДИ 1,0–1,0). У ИИ-2 отмечено большее количество как гипо- (15,0 %), так и гипердиагностики (5,0 %) соответственно, что можно было объяснить преобладанием специфичности над чувствительностью.

Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.25).

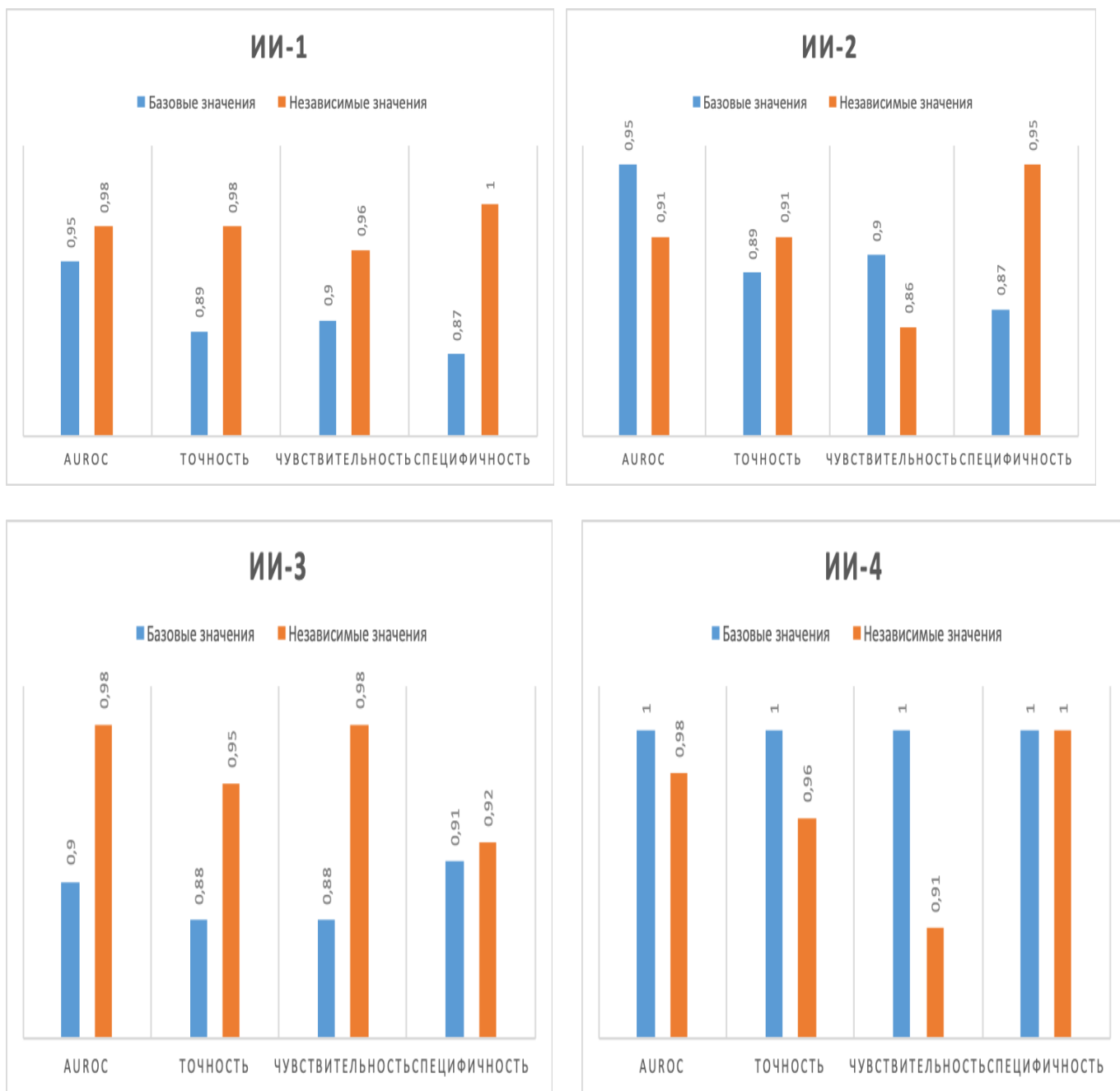


Рисунок 3.25 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервисов для анализа результатов КТ органов грудной и/или брюшной полости с целью выявления признаков остеопороза

Особенно примечательны результаты ИИ-1 и ИИ-3: у обоих сервисов все показатели при ретроспективном тестировании на новых данных продемонстрировали рост. Значения AUROC возросли на 3,2 и 8,9 % соответственно. У ИИ-1 особо значительным было увеличение специфичности (+15,0 %), а у ИИ-3 – чувствительности (+11,4 %).

У ИИ-2 показатели колебались в пределах 2,0–4,0 %, лишь специфичность

продемонстрировала значительный рост на 9,0 %. Напротив, сервис ИИ-4 продемонстрировал снижение всех показателей точности, за исключением специфичности. Более всего уменьшилась чувствительность – на 9,0 %.

В целом ИИ-сервисы продемонстрировали устойчивость и хорошую воспроизводимость результатов собственной работы при анализе новых данных.

На втором проспективном этапе проведено исследование в условиях работы ИИ-сервиса с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены от двух до шести измерений точности в проспективном формате; результаты представлены в таблице 3.29.

Таблица 3.29 – Показатели диагностической точности ИИ-сервисов для анализа результатов КТ органов грудной и/или брюшной полости с целью выявления признаков остеопороза, полученные при проспективном наблюдении

ИИ-сервис	Показатели	Период					
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
ИИ-1	AUROC	-	-	-	0,832	0,786	-
	95%-й ДИ	-	-	-	0,782–0,884	0,723–0,847	-
	Точность	-	-	-	0,819	0,700	-
	95%-й ДИ	-	-	-	0,759–0,878	0,600–0,800	-
	Чувствительность	-	-	-	0,986	1,000	-
	95%-й ДИ	-	-	-	0,960–1,000	1,000–1,000	-
	Специфичность	-	-	-	0,678	0,571	-
	95%-й ДИ	-	-	-	0,580–0,776	0,442–0,701	-
ИИ-2	AUROC	0,811	0,816	-	-	-	-
	95%-й ДИ	0,760–0,861	0,771–0,862	-	-	-	-
	Точность	0,787	0,812	-	-	-	-
	95%-й ДИ	0,724–0,851	0,763–0,862	-	-	-	-
	Чувствительность	1,000	0,957	-	-	-	-
	95%-й ДИ	1,000–1,000	0,921–0,994	-	-	-	-
	Специфичность	0,622	0,675	-	-	-	-
	95%-й ДИ	0,522–0,722	0,592–0,758	-	-	-	-

Продолжение таблицы 3.29

ИИ-сервис	Показатели	Период					
		I кв. 2022 г.	II кв. 2022 г.	III кв. 2022 г.	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
ИИ-3	AUROC	-	-	-	-	0,775	0,848
	95%-й ДИ	-	-	-	-	0,693– 0,855	0,799–0,898
	Точность	-	-	-	-	0,856	0,812
	95%-й ДИ	-	-	-	-	0,802– 0,911	0,763–0,862
	Чувствительность	-	-	-	-	0,988	0,966
	95%-й ДИ	-	-	-	-	0,965– 1,000	0,932–0,999
	Специфичность	-	-	-	-	0,707	0,669
	95%-й ДИ	-	-	-	-	0,604– 0,810	0,587–0,752
ИИ-4	AUROC	0,887	0,92	0,918	0,962	0,967	0,905
	95%-й ДИ	0,817–0,954	0,848–0,990	0,876– 0,960	0,932–0,992	0,936– 0,999	0,860–0,949
	Точность	0,912	0,9	0,917	0,959	0,956	0,925
	95%-й ДИ	0,869–0,956	0,834–0,966	0,882– 0,952	0,933–0,984	0,925– 0,988	0,892–0,958
	Чувствительность	0,989	0,957	0,986	0,987	0,99	0,987
	95%-й ДИ	0,968–1,000	0,898–1,000	0,966– 1,000	0,969–1,000	0,971– 1,000	0,968–1,000
	Специфичность	0,812	0,824	0,822	0,909	0,897	0,822
	95%-й ДИ	0,719–0,904	0,695–0,952	0,747– 0,896	0,849–0,969	0,818– 0,975	0,743–0,901

При работе в реальных клинических условиях сервис ИИ-1 продемонстрировал выраженное ухудшение показателей точности. AUROC достоверно снизилась на 19,6 % (z-критерий Фишера 5,5975, $p = p < 0,0001$) и 14,9 % (z-критерий Фишера 4,9989, $p = p < 0,0001$).

Аналогично произошло значимое снижение точности. Но самое выраженное ухудшение коснулось специфичности: в первом периоде наблюдения снижение составило 42,9 % (t-критерий -58,042, $p < 0,0001$), во втором – 32,2 % (t-критерий -

57,569, $p < 0,0001$). С точки зрения клинической результативности, выявлено значительное количество случаев с некорректным определением локализации патологии (до 41,7 % в первом периоде наблюдения).

Фактически аналогичная ситуация отмечена для сервиса ИИ-2. Зафиксировано некоторое, хотя и значимое ($p = 0,0134$, $p = 0,0144$), снижение AUROC примерно на 10,0 %. Крайне негативно выглядит ситуация со специфичностью – зафиксировано ее падение на 34,5 % (t-критерий -48,727, $p < 0,0001$) и 29,0 % (t-критерий -46,366, $p < 0,0001$) в первом и втором периодах наблюдения соответственно.

У обоих сервисов критичное снижение специфичности на фоне высокой чувствительности вело к подавляющей гипердиагностике, что создавало неадекватную нагрузку на врачей-рентгенологов и резко ухудшало их приверженность к использованию ИИ-1 и ИИ-2. В конечном итоге, оба сервиса прекратили работу в реальных клинических условиях.

Изначальный лидер точности ИИ-3 полностью повторил траекторию двух предыдущих сервисов: достоверное снижение AUROC, критичное снижение специфичности на фоне сохраняющейся высокой чувствительности привели к остановке работы сервиса в проспективном режиме.

Вместе с тем надо отметить, что по результатам оценки клинической результативности ИИ-1, ИИ-2 и ИИ-3 фактически не давали ложноотрицательных результатов, то есть не пропускали действительно имеющуюся патологию. Лишь во втором периоде наблюдения у ИИ-1 и ИИ-2 появилось по 1,5 % и 1,3 % случаев с гиподиагностикой.

Наибольшую стабильность и наилучшее качество продемонстрировал ИИ-4. При переходе на проспективный формат работы все-таки зафиксировано значимое снижение AUROC на 9,3 % (z-критерий Фишера 2,4630, $p = 0,0138$). Однако затем точность была увеличена и стабильно установилась на уровне более 0,9; в частности, в последнем периоде наблюдения площадь под характеристической кривой составила 0,905 (95%-й ДИ 0,860–0,949), а максимальное значение этого параметра было зафиксировано в пятом периоде

наблюдения – 0,967 (95%-й ДИ 0,936–0,999), лишь на 1,1 % уступающее исходному на ретроспективном этапе.

Аналогичная динамика отмечена для общей точности, а вот чувствительность значимо возросла и стабильно оставалась на уровне более 0,98 с третьего периода наблюдения.

С учетом данных первых трех ИИ-сервисов интересным представляется сравнительное изучение специфичности. При переходе к проспективному формату работы специфичность ИИ-4 падает на 18,8 % (t-критерий -35,610, $p < 0,0001$), однако, затем увеличивает и стабильно остается на уровне более или равно 0,82. Лишь в четвертом и пятом периодах наблюдения фиксируется внезапный рост до 0,909 до 0,897 с последующим возвратом к «плато». С одной стороны, отсутствует значимый прогресс, с другой – в отличие от трех иных ИИ-сервисов обеспечен стабильный минимальный уровень специфичности для нормального функционирования ИИ-4 в реальных клинических условиях. Еще раз подчеркнем, что к последнему периоду наблюдения отличия от исходного уровня сохраняют статистическую значимость (t-критерий -39,478, $p < 0,0001$), но остаются в пределах некоего «плато» и обеспечивают нормальную работу сервиса.

При проспективной оценке клинической результативности ИИ-4 установлено, что удельный вес случаев с полным соответствием решений врачей-экспертов и ИИ-сервиса колебался в пределах 89,8–95,0 % (причем ниже 92,0 % он опустился лишь один раз).

Удельный вес случаев с некорректной оценкой локализации составил 1,9–5,9 %. Гипердиагностика встретилась в 1,3–5,0 %, а ложноотрицательные случаи зафиксированы только во втором и пятом периодах наблюдения (1,28 и 0,64 % соответственно).

Длительность работы в реальных клинических условиях (6 периодов наблюдения) позволила корректно определить средние значения характеристик клинической результативности ИИ-4: полное соответствие – $92,7 \pm 1,7$ % (медиана – 93,0 %), некорректная локализация – $3,8 \pm 1,5$ % (медиана – 3,6 %),

гипердиагностика – $3,1 \pm 1,4$ % (медиана – 3,5 %), гиподиагностика – $0,32 \pm 0,54$ % (медиана – 0,0).

Таким образом, при автоматизированном анализе результатов КТ органов брюшной полости с целью выявления признаков остеопороза потенциально достижим уровень точности, сопоставимый с врачом-рентгенологом (на потоке данных AUROC составляет в среднем 0,93, наименьшее значение случаев гиподиагностики равно нулю, а гипердиагностики – 1,2 %). Существует возможность при работе в реальных клинических условиях поддерживать стабильные значения показателей точности и качества, обеспечивающие приемлемую функциональность ИИ-сервиса.

Вместе с тем в прикладном аспекте лишь один из четырех программных продуктов для работы с изображениями органов брюшной полости соответствовал требованиям и работал стабильно. Это свидетельствует о наличии значительных технических трудностей при обучении и развитии моделей ИИ для анализа результатов КТ с целью выявления признаков остеопороза.

Полученные данные позволяют утверждать, что технологии ИИ могут применяться для автоматизированного анализа результатов КТ органов брюшной полости с целью выявления признаков остеопороза (компрессионных переломов позвонков). В контексте данной клинической задачи ТИИ могут использоваться в качестве системы поддержки принятия врачебных решений и для морфометрии (определение степени деформации позвонков).

3.5. Обоснование возможности автоматизированного анализа результатов магнитно-резонансной томографии

Магнитно-резонансная томография позвоночника. В контексте автоматизированного анализа результатов МРТ пояснично-крестцового отдела позвоночника с целью выявления дегенеративно-дистрофических изменений была сформулирована следующая клиническая задача – определение наличия и локализация на T2WI в сагиттальной и аксиальной проекциях МРТ-признаков

(минимум одного), коррелирующих с наличием дегенеративно-дистрофических изменений межпозвонковых дисков пояснично-крестцового отдела позвоночника:

– для нативных исследований – дорзальные выпячивания межпозвонковых дисков за пределы дисковых пространств (края замыкательных пластин смежных позвонков) в зоне исследования с распространением их в просвет позвоночного канала, в соответствии с Lumbar disc nomenclature: version 2.0¹⁵;

– при наличии признака производится автоматическое измерение дурального мешка на уровне всех межпозвонковых дисков: на аксиальных изображениях – передне-заднего и фронтального размеров, площадь просвета дурального мешка на уровне дисков; на сагиттальных изображениях – передне-заднего размера.

Ретроспективное исследование диагностической точности одного¹⁶ ИИ-сервиса для автоматизированного анализа результатов МРТ пояснично-крестцового отдела позвоночника с целью выявления дегенеративно-дистрофических изменений, точнее – протрузий и грыж межпозвоночных дисков, стеноза позвоночного канала выполнено на эталонном наборе данных (n = 100).

Построена характеристическая кривая; оптимальный порог (индекс Юдена) определен как 95,0 (рисунок 3.26). Значение площади под характеристической кривой составило 0,894 (95%-й ДИ 0,833–0,955). Определены показатели: общая точность составила 0,89 (95%-й ДИ 0,83–0,95), чувствительность – 0,84 (95%-й ДИ 0,74–0,94), специфичность – 0,94 (95%-й ДИ 0,87–1,0). Удельный вес ложноотрицательных результатов оказался довольно высоким и составил 16,0 %, в то время как ложноположительных – 6,25 %.

Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.27).

тик», система поддержки принятия диагностических решений «АИ Диагностик»; ООО «АЙРА Лабс», Genant-IRA.

¹⁵ Williams A. L., Murtagh F. R., Rothman S. L., Sze G. K. Lumbar disc nomenclature: version 2.0 // AJNR Am J Neuroradiol. – 2014. – Vol. 35, №11. – P. 20–29. – DOI:10.3174/ajnr.A4108.

¹⁶ ООО «Интел Диагностик», система поддержки принятия диагностических решений «АИ Диагностик».

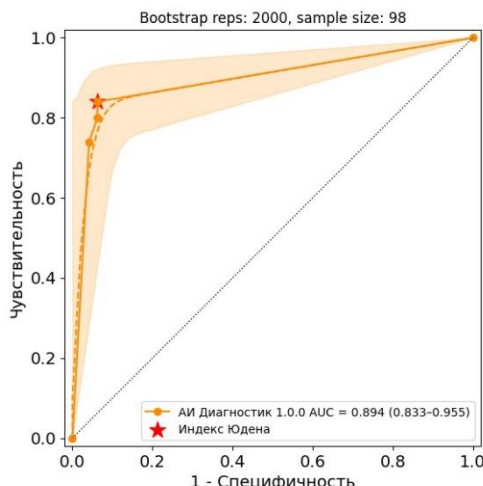


Рисунок 3.26 – Характеристическая кривая для ИИ-сервиса для анализа результатов МРТ пояснично-крестцового отдела позвоночника (оптимальный порог по индексу Юдена)

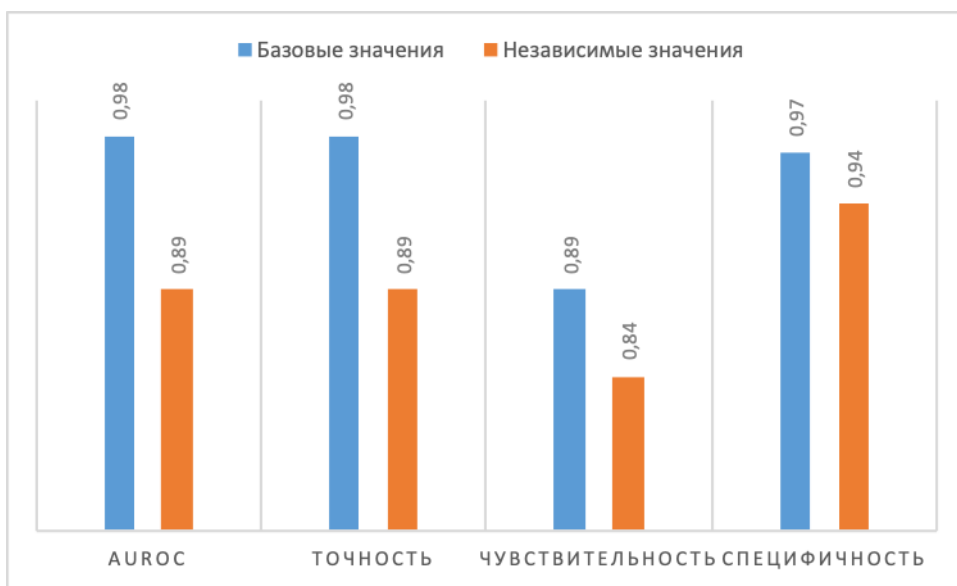


Рисунок 3.27 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов МРТ пояснично-крестцового отдела позвоночника

Обращает на себя внимание явное снижение значений показателей точности, полученных в ходе независимого ретроспективного тестирования. В частности, относительного базового значение площади под характеристической кривой снизилось на 8,78 %, общей точности – на 9,18%, чувствительности на – 5,62 %, а специфичности – на 3,09 %.

Тем не менее показатель AUROC превышал пороговое значение для данного вида исследований, поэтому ИИ-сервис был допущен к работе с реальными клиническими данными и, соответственно, ко второму этапу оценки диагностической точности.

Проспективное исследование осуществлено в условиях работы ИИ-сервиса с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены три измерения точности в проспективном формате; результаты представлены в таблице 3.30.

Таблица 3.30 – Показатели диагностической точности ИИ-сервиса для анализа результатов МРТ пояснично-крестцового отдела позвоночника, полученные при проспективном наблюдении

Показатели	Период		
	IV кв. 2022 г.	I кв. 2023 г.	II кв. 2023 г.
AUROC	0,639	0,909	0,948
95%-й ДИ	0,522–0,761	0,844–0,976	0,893–1,005
Точность	0,721	0,871	0,954
95%-й ДИ	0,664–0,778	0,828–0,913	0,928–0,981
Чувствительность	0,739	0,87	0,958
95%-й ДИ	0,680–0,797	0,826–0,915	0,931–0,985
Специфичность	0,545	0,875	0,92
95%-й ДИ	0,337–0,754	0,743–1,000	0,814–1,000

В первом временном периоде, в условиях работы с реальными клиническими данными ИИ-сервис для анализа результатов МРТ пояснично-крестцового отдела позвоночника продемонстрировал выраженное ухудшение точности: показатель площади под характеристической кривой снизился на 28,5 %, точности – на 19,0 %, чувствительности – на 12,0 %, специфичности – на целых 42,0 % относительно уровня, полученного при ретроспективной оценке.

Различия между значениями AUROC при ретроспективном и первом

проспективном наблюдении носили статистически значимый характер (z-критерий Фишера 3,7252, $p = 0,0002$).

Также достоверным было снижение значений общей точности (t-критерий $-37,628$, $p < 0,0001$), чувствительности (t-критерий $-15,684$, $p < 0,0001$) и специфичности (t-критерий $-35,083$, $p < 0,0001$).

Такие значения показателей были ниже порогового уровня – это привело к остановке работы ИИ-сервиса на потоке реальных клинических данных с целью технической доработки. Соответствующие мероприятия были успешны, при возобновлении работы с потоком результатов исследований ИИ-сервис продемонстрировал значительное улучшение показателей точности. При наблюдениях в два последующих периода AUROC возросла на 1,7–6,1 % относительно уровня, полученного при ретроспективной оценке. В третьем периоде также значительно возросла чувствительность (на 14,1 %), превысив при этом «психологическую» отметку в 95,0 %. Фактически, до исходного уровня возросла и специфичность. Значительно, на 7,2 %, возрос показатель общей точности, достигнув уровня 0,954.

Принципиальный рост диагностической точности анализируемого ИИ-сервиса подтверждается достоверностью различий между значениями AUROC в первом и третьем периодах проспективного наблюдения (z-критерий Фишера 4,5892, $p < 0,0001$). Аналогичная ситуация фиксируется для общей точности (t-критерий $-64,981$, $p < 0,0001$), чувствительности (t-критерий $-59,588$, $p < 0,0001$) и специфичности (t-критерий $-28,796$, $p < 0,0001$).

При проспективной оценке клинической результативности установлено, что удельный вес случаев с полным соответствием решений врачей-экспертов и ИИ-сервиса колебался в пределах 40,5–81,6 %; случаев с некорректной оценкой локализации – от 13,2 до 36,7 %. Гипердиагностика в объеме 3,6 % зафиксирована только в одном временном периоде. Удельный вес ложноотрицательных случаев составил в первых двух периодах 3,9 и 1,3 % соответственно, а в третьем – резко возрос до 22,8 %.

Таким образом, при автоматизированном анализе результатов МРТ

пояснично-крестцового отдела позвоночника с целью выявления протрузий и грыж межпозвоночных дисков потенциально достижим уровень точности, сопоставимый с врачом-рентгенологом (на потоке данных AUROC составляет 0,909–0,948, наименьшее значение случаев гиподиагностики – 1,3 %, гипердиагностики – 0). Вместе с тем работа конкретного продукта – программного обеспечения – отличается нестабильностью. При начале работы на потоке реальных данных отмечается типичное, статистически значимое ухудшение точности и качества ($p = 0,0002$, $p < 0,0001$). В дальнейшем за счет доработки продукта выявлен также существенный ($p < 0,0001$) рост значений всех показателей диагностической точности, в том числе AUROC. Однако этому сопутствует значительный рост числа пропусков патологии (удельный вес гиподиагностики достигает 22,8 %).

Полученные данные позволяют утверждать, что технологии ИИ могут применяться для автоматизированного анализа результатов МРТ пояснично-крестцового отдела позвоночника с целью выявления протрузий и грыж межпозвоночных дисков, стеноза позвоночного канала. В контексте данной клинической задачи ТИИ могут использоваться в качестве системы поддержки принятия врачебных решений и для морфометрии (определения размеров выпячиваний межпозвоночных дисков).

Магнитно-резонансная томография головного мозга. В контексте автоматизированного анализа результатов МРТ головного мозга с целью диагностики рассеянного склероза была сформулирована следующая клиническая задача – определение наличия очагов демиелинизации как основного признака рассеянного склероза и дифференциация от других патологий:

– выявление патологических признаков: для нативных исследований – наличие ≥ 1 патогномоничных гиперинтенсивных очагов демиелинизации не менее 3 мм по длинной оси на T2 и/или FLAIR двух или более локализаций: юкстакортикальные или субкортикальные, перивентрикулярные,

инфратенториальные (критерии McDonald¹⁷); для исследований с контрастным усилением – наличие ≥ 1 патогномоничных очагов, накапливающих контрастное вещество, на постконтрастных T1-изображениях;

– оконтуривание очагов демиелинизации с цветовой дифференциацией по локализациям; обязательное выделение очагов, накапливающих контрастное вещество;

– подсчет общего количества очагов демиелинизации на бесконтрастных сериях и подсчет количества очагов, накапливающих контрастное вещество.

Ретроспективное исследование диагностической точности одного¹⁸ ИИ-сервиса для автоматизированного анализа результатов МРТ головного мозга с целью выявления признаков рассеянного склероза выполнено на эталонном наборе данных (n = 100).

Построена характеристическая кривая; оптимальный порог (индекс Юдена) определен как 0,99 (рисунок 3.28). Значение площади под характеристической кривой составило 0,739 (95%-й ДИ 0,670–0,810). Определены показатели: общая точность составила 0,750 (95%-й ДИ 0,660–0,830), чувствительность – 1,000 (95%-й ДИ 1,000–1,000), специфичность – 0,480 (95%-й ДИ 0,330–0,620). Ложноотрицательных результатов не зафиксировано вовсе, а показатель ложноположительных (гипердиагностика) был критично высок – 53,0 %. Полученные результаты сопоставлены с заявленными базовыми значениями показателей диагностической точности (рисунок 3.29).

¹⁷ Применение критериев диагностики и контроля рассеянного склероза по MAGNIMS / сост. В. А. Гомболевский, А. Ш. Лайпан, А. Н. Шапиев [идр.] // Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып.11. – М., 2018. – 12 с.

¹⁸ ООО «ИМВИЖН», IMV MS.

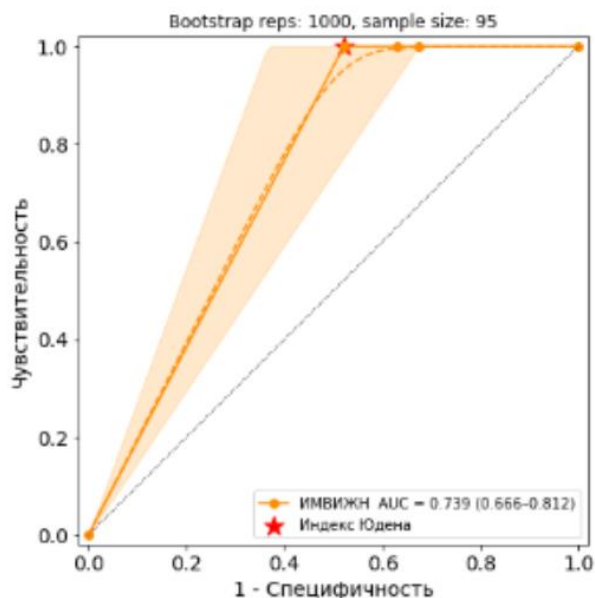


Рисунок 3.28 – Характеристическая кривая для ИИ-сервиса для анализа результатов МРТ головного мозга (оптимальный порог по индексу Юдена)

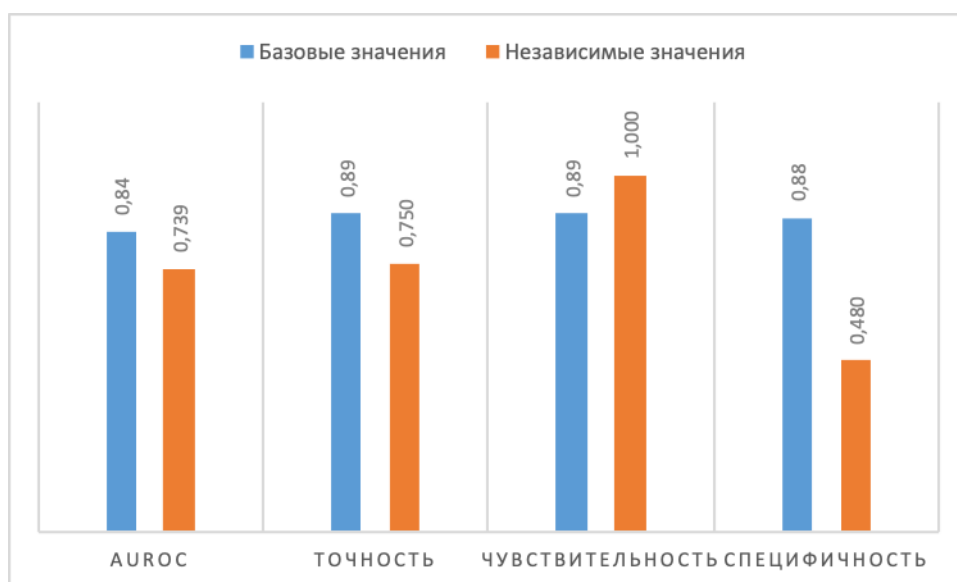


Рисунок 3.29 – Сравнение базовых и независимых показателей диагностической точности ИИ-сервиса для анализа результатов МРТ головного мозга

Обращает на себя внимание явное снижение значений показателей точности, полученных в ходе независимого ретроспективного тестирования. В частности, относительного базового значение площади под характеристической кривой снизилось на 12,0 %, общей точности – на 15,7%, специфичности – на целых 45,5 %. На этом фоне значение чувствительности получено больше заявленного на 12,4 %.

Тем не менее показатель AUROC превышал пороговое значение

для данного вида исследований, поэтому ИИ-сервис был допущен к работе с реальными клиническими данными и, соответственно, ко второму этапу оценки диагностической точности.

Перспективное исследование проведено в условиях работы ИИ-сервиса с потоком результатов лучевых исследований, выполненных в медицинских организациях г. Москвы при оказании помощи в амбулаторных условиях и условиях стационара (в том числе дневного).

В соответствии с методологией проведены два измерения точности в проспективном формате; результаты представлены в таблице 3.31.

Таблица 3.31 – Показатели диагностической точности ИИ-сервиса для анализа результатов МРТ головного мозга, полученные при проспективном наблюдении

Показатели	Период	
	I кв. 2023 г.	II кв. 2023 г.
AUROC	0,948	0,899
95%-й ДИ	0,903–0,995	0,852–0,945
Точность	0,938	0,900
95%-й ДИ	0,884–0,991	0,854–0,946
Чувствительность	1,000	0,978
95%-й ДИ	1,000–1,000	0,947–1,000
Специфичность	0,935	0,880–0,990
95%-й ДИ	0,803	0,710–0,895

В первом временном периоде, в условиях работы с реальными клиническими данными ИИ-сервис для анализа результатов МРТ головного мозга с целью выявления признаков рассеянного склероза продемонстрировал значительное улучшение точности: показатель площади под характеристической кривой возрос на 28,3 %, точности – на 25,1 %, специфичности – на целых 94,8 % относительно уровня, полученного при ретроспективной оценке. Показатель чувствительности остался на прежнем максимальном уровне.

Различия между значениями AUROC при ретроспективном и первом проспективном наблюдениях носили статистически значимый характер

(z-критерий Фишера $-4,8906$, $p < 0,0001$). Также достоверным было увеличение значений общей точности (t-критерий $33,779$, $p < 0,0001$) и специфичности (t-критерий $52,072$, $p < 0,0001$). Чувствительность оставалась максимальной.

Измерения точности в следующий временной период показали некоторое снижение значения AUROC (до $0,899$), однако оно не имело статистической значимости (z-критерий Фишера $1,4683$, $p = 0,1420$). Аналогичной ситуация была для чувствительности и общей точности. На этом фоне значение специфичности существенно понизилось до $0,803$ (95%-й ДИ $0,710-0,895$) (t-критерий $21,503$, $p < 0,0001$).

При проспективной оценке установлена высокая клиническая результативность ИИ-сервиса для анализа результатов МРТ головного мозга с целью выявления рассеянного склероза: не выявлено ни одного ложноотрицательного случая. А в период работы на реальных клинических данных происходило явное техническое улучшение модели ИИ, проявившееся снижением удельного веса случаев гипердиагностики и некорректной локализации патологических проявления с $14,3\%$ и $71,4\%$ соответственно до нуля. Во втором временном периоде полное соответствие решений врачей-экспертов и ИИ-сервиса имело место в $100,0\%$.

Таким образом, при автоматизированном анализе результатов МРТ головного мозга с целью выявления рассеянного склероза потенциально достижим уровень точности и клинической результативности, сопоставимый с врачом-рентгенологом (на потоке данных AUROC составляет $0,899-0,948$, удельный вес случаев гиподиагностики равен нулю, а случаев гипердиагностики уменьшился до нуля). При работе в реальных клинических условиях показатели точности характеризуются стабильностью, а клинической результативности – улучшением, на котором даже статистически значимое снижение специфичности не сказалось негативно.

Полученные данные позволяют говорить не только о потенциальном, но и прикладном результате. Технологии ИИ могут применяться для автоматизированного анализа результатов МРТ головного мозга с целью

выявления признаков рассеянного склероза, а уровень технологической готовности позволяет реализовать эффективно функционирующий программный продукт. В контексте данной клинической задачи ТИИ могут применяться в качестве системы поддержки принятия врачебных решений и для морфометрии (подсчета количества очагов демиелинизации).

3.6. Обобщение способов автоматизированного анализа результатов лучевых исследований

Нами проанализирована диагностическая точность технологий искусственного интеллекта при решении ряда клинических задач, актуальных с точки зрения работы врача-рентгенолога. Исследование включало два этапа. На первом, ретроспективном, этапе ИИ-сервисы тестировались на эталонных наборах данных. Причем такие данные были новыми для сервиса, то есть не использовались при его разработке или тестировании. На втором, проспективном, этапе ИИ-сервисы работали в реальных клинических условиях; по разработанной ранее методологии технологического и клинического мониторинга осуществлялись периодические измерения точности. Полученные результаты позволили нам утверждать следующее.

Технологии ИИ могут и должны применяться для автоматизации решения клинических задач лучевой диагностики, однако уровни прикладной готовности достаточно сильно варьируются.

Полученные результаты позволили нам качественно пересмотреть трактовку «сортировки» как одного из способов применения технологий ИИ при интерпретации и описании результатов лучевых исследований.

Наибольшее количество прикладных решений разработано для автоматизированного анализа результатов рентгенографии (флюорографии) органов грудной клетки и профилактической маммографии. Нами достоверно показана характерная их особенность: достижение стабильно высокого уровня чувствительности на фоне средней специфичности. Такое сочетание

потенциально устраняет возможность ложноотрицательного результата, то есть ТИИ гарантированно не пропустит результаты исследования с патологическими изменениями. Вместе с тем точность дальнейшей дифференциации таких изменений остается средней. На основе этого предлагается переход от прежнего понимания «сортировки» к идее автономного ИИ при проведении профилактических исследований.

Известно, что профилактическое направление здравоохранения является чрезвычайно важным, в том числе с точки зрения снижения смертности от онкологических, кардиологических и иных социально значимых заболеваний. Охват населения профилактическими мероприятиями – важный показатель результативности деятельности органов управления здравоохранения различного уровня. С одной стороны, всегда отмечается необходимость увеличения такого охвата, с другой – это неизбежно приводит к росту нагрузки на систему здравоохранения. В контексте массовых скринингов колоссальные ресурсы тратятся на оказание медицинских услуг преимущественно здоровому населению. Особенно актуально это утверждение для лучевой диагностики. Профилактические рентгенологические исследования органов грудной клетки и молочной железы – основные методы массовых обследований населения. Они отличаются высокой степенью стандартизации; к их интерпретации предъявляются особенные требования (прежде всего – двойные просмотры результатов маммографии); подавляющее большинство результатов таких исследований содержат «норму». Такая ситуация явным образом требует автоматизации для оптимизации ресурсов лучевой диагностики и системы здравоохранения в целом. Важным будет и высвобождение кадровых ресурсов – врачей-рентгенологов – для работы с более когнитивно сложными задачами и современными модальностями.

Ранее предложен следующий подход: технологии ИИ с максимальной чувствительностью (1,0 (95%-й ДИ 1,0–1,0)) должны применяться для автономной сортировки результатов массовых профилактических исследований (РГ/ФЛГ и ММГ) на две категории: «норма» и «не норма». Ко второй категории

относятся все случаи любых отклонений (патологические состояния, последствия перенесенных заболеваний и операций, возрастные и врожденные особенности и т.д.). При этом, с учетом максимальной чувствительности ТИИ, исследования категории «норма» не требуют описания врачом-рентгенологом, а сразу могут быть сохранены в карте пациента для дальнейшего использования врачом клинической специальности. Потенциально такой подход на десятки процентов уменьшит нагрузку на кадровый ресурс лучевой диагностики, позволит оптимизировать финансовую составляющую. Важным преимуществом подхода является возможность нарастить охват населения профилактическими мероприятиями без создания дополнительной нагрузки на ресурсы системы здравоохранения.

Трактовки двух иных способов – СППВР и морфометрии – не изменились.

В таблице 3.32 обобщены способы автоматизированного анализа для разных модальностей и видов исследований.

Таблица 3.32 – Способы автоматизированного анализа результатов лучевых исследований

№	Модальность, вид исследования	СППВР	Морфометрия	Автономная сортировка
1.	РГ (ФЛГ) органов грудной клетки	+	–	++
2.	РГ стопы	–	++	+
3.	РГ позвоночника	*	*	–
4.	РГ коленного сустава	++	*	–
5.	Маммография	+	–	++

Продолжение таблицы 3.32

№	Модальность, вид исследования	СППВР	Морфометрия	Автономная сортировка
6.	КТ головного мозга:			
	– внутричерепное кровоизлияние	*	*	–
	– ОНМК	*	*	–
7.	КТ органов грудной клетки:			
	– гидроторакс	*	*	–
	– легочный ствол	*	*	–
8.	КТ томография органов грудной клетки и брюшной полости:			
	– аорта	+	++	–
	– надпочечники	++	+	–
	– позвоночник	++	+	–
9.	МРТ позвоночника	++	+	–
10.	МРТ головного мозга (рассеянный склероз)	++	+	–
Примечание. «++» – основной способ применения; «+» – дополнительный способ применения; «*» – требуется технологическое совершенствование моделей ИИ; «–» – не применимо				

Достигнутый уровень развития технологий искусственного интеллекта как способа автоматизации задач анализа, интерпретации и описания результатов лучевых исследований позволяет использовать их для повышения производительности и качества работы врача-рентгенолога в виде:

1. Системы поддержки принятия врачебных решений при описании результатов:

– магнитно-резонансной томографии головного мозга с целью диагностики и оценки динамики рассеянного склероза;

- магнитно-резонансной томографии позвоночника с целью выявления дегенеративно-дистрофических изменений;

- компьютерной томографии органов грудной клетки и брюшной полости с целью выявления новообразований надпочечников, компрессионных переломов позвонков и признаков снижения минеральной плотности костной ткани (остеопороза);

- рентгенографии коленного сустава с целью стадирования деформирующего остеоартроза.

2. Инструмента морфометрии при описании результатов:

- компьютерной томографии органов грудной клетки и брюшной полости с целью выявления аневризмы аорты;

- рентгенографии стопы с целью выявления продольного плоскостопия.

3. Инструмента автономной сортировки при интерпретации результатов:

- профилактической рентгенографии (флюорографии) органов грудной клетки;

- профилактической маммографии.

Требуют дальнейшего улучшения математические модели, лежащие в основе прикладных решений для автоматизированного анализа:

- рентгенографии позвоночника с целью диагностики сколиоза;

- компьютерной томографии головного мозга с целью выявления и измерения объема внутричерепного кровоизлияния, диагностики и оценки по принятой шкале ишемического инсульта;

- компьютерной томографии органов грудной клетки с целью выявления и морфометрии гидроторакса, морфометрии легочного ствола.

В рамках нашего исследования переход ИИ-сервисов на работу с реальными клиническими данными состоялся 38¹⁹ раз. При этом в 84,0 % (32) случаев зафиксировано снижение площади под характеристической кривой

¹⁹ Это число отличается от количества включенных ИИ-сервисов, так как факты перехода отдельно для модальностей «рентгенография» и «флюорография» были подсчитаны отдельно.

– интегрального показателя диагностической точности. Лишь в 16,0 % (6) случаев ИИ-сервисы продемонстрировали устойчивость или рост AUROC. Данный факт позволяет утверждать, что при разработке ИИ-сервиса его апробация в формате проспективного клинического исследования является обязательной. При этом такая апробация не может быть оформлена как клинические испытания в целях регистрации медицинского изделия. В результате проспективного клинического исследования должны появиться объективные оценки точности и результативности и предприняты усилия по доработке ИИ-сервиса с повторным, также проспективным, его тестированием. Лишь после этого, уже как отдельное мероприятие, целесообразно проводить клинические испытания в целях регистрации медицинского изделия.

Таким образом, нами была доказана применимость ТИИ для решения отдельных клинических задач; при этом лидирующим по числу видов исследований является способ применения в виде СППВР. Клиническая значимость использования таких СППВР изучена далее, в главе 4.

Выше была представлена идея автономной сортировки результатов профилактических лучевых исследований. Параллельно нами получены результаты, свидетельствующие о высокой точности отдельных ТИИ для морфометрии, прежде всего – для диагностики плоскостопия. Достигнутый уровень качества и скорости измерений, а также минимальная (если не сказать «отсутствующая») когнитивная сложность интерпретации результатов такого исследований, позволяют их рассмотреть и в контексте автономного применения технологий искусственного интеллекта. Обоснование автономной сортировки и автономной морфометрии с применением технологий искусственного интеллекта требует отдельного исследования, которому посвящена глава 4.

Глава 4. КЛИНИЧЕСКАЯ ЦЕЛЕСООБРАЗНОСТЬ, БЕЗОПАСНОСТЬ И КАЧЕСТВО АВТОМАТИЗИРОВАННОГО АНАЛИЗА РЕЗУЛЬТАТОВ ЛУЧЕВЫХ ИССЛЕДОВАНИЙ

4.1. Оценка клинической значимости поддержки принятия врачебных решений с применением искусственного интеллекта

Очевидно, что любая система поддержки принятия решений служит для повышения точности и обоснованности соответствующих действий человека. Показателем результативности такой системы можно полагать объем предотвращенного ущерба, то есть разницу между ущербом от неточных решений, принятых человеком самостоятельно и принятых совместно, на основе данных конкретной системы. Корректно работающая СППВР предотвращает ошибки при принятии решений, тем самым не допуская или снижая ущерб.

В контексте автоматизации лучевой диагностики полагаем, что система поддержки принятия врачебных решений должна потенциально снижать частоту дефектов в работе врача-рентгенолога – прежде всего пропусков, а также гипердиагностики. В этом и заключается основная клиническая значимость таких решений. Для доказательства этого утверждения мы пошли от обратного.

В государственном здравоохранении г. Москвы постоянно и системно проводятся мероприятия по контролю и обеспечению безопасности и качества медицинской помощи. В лучевой диагностике соответствующая деятельность ведется экспертным отделом ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ» методом ретроспективного пересмотра результатов лучевых исследований по утвержденной методологии [89]. Наша гипотеза состояла в следующем:

- в результатах пересмотров содержатся случаи с клинически значимыми расхождениями;
- эти случаи были не только первично описаны врачом-рентгенологом, но и проанализированы ИИ-сервисом;

– врач принял ложноотрицательное решение (которое привело к гиподиагностике и клинически значимому расхождению) из-за игнорирования результатов работы ИИ, свободно доступных на автоматизированном рабочем месте.

В этом контексте мы изучаем клиническую значимость СППВР на основе ИИ с позиции потенциально предотвратимых клинически значимых дефектов в работе врача-рентгенолога.

В обсервационном дизайне выполнено исследование потенциального влияния СППВР на качество работы врача-рентгенолога. Согласно методологии (подробно см. главу 2) включены и проанализированы случаи, когда существенный патологический признак был выявлен ИИ-сервисом сразу после выполнения исследования и врачом-экспертом в рамках планового контроля качества, но пропущен врачом-рентгенологом, выполнявшим первичное описание (также сразу после непосредственного выполнения исследования). Указанный признак стал причиной клинически значимого расхождения при осуществлении пересмотров в рамках контроля качества.

Всего за период проведено 49 994 ретроспективных анализов, в том числе 30 524 результата компьютерной томографии всех анатомических областей. Клинически значимые расхождения выявлены в результатах 587 КТ органов грудной клетки.

В соответствии с критериями включения выявлено 60 исследований с клинически значимыми расхождениями и релевантными результатами работы технологий ИИ (рисунок 4.1), что составляет 10,2 % от числа случаев КТ органов грудной клетки с таким типом дефектов.

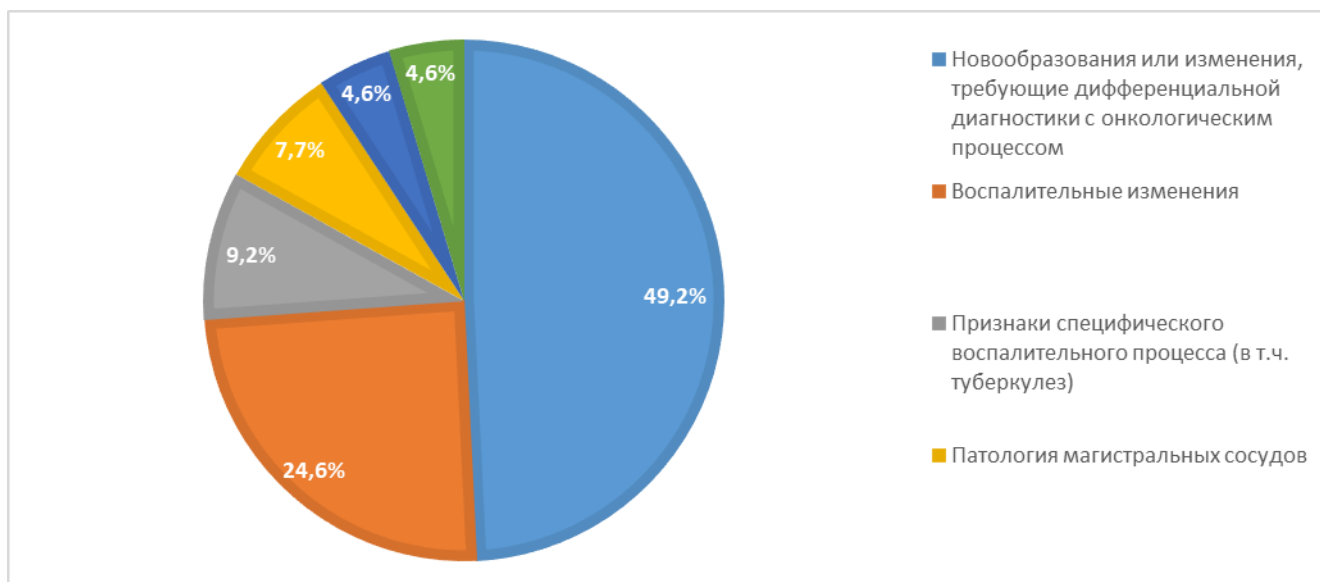


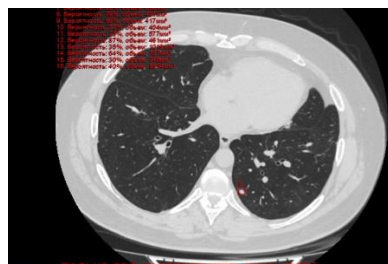
Рисунок 4.1 – Удельный вес причин клинически значимых расхождений – потенциально предотвратимых ложноотрицательных решений врача-рентгенолога

Определена следующая структура потенциально предотвратимых дефектов. В большинстве случаев зафиксирован пропуск признаков новообразований (как типичных, так и требующих дифференциальной диагностики именно онкологического характера) – 49,2 % (32), также значительный удельный вес составила гиподиагностика неспецифических воспалительных изменений в легких – 24,6 % (16). Принципиально ниже было количество пропусков признаков туберкулеза или иного специфического воспалительного процесса – 9,2 % (6) и патологии магистральных сосудов (аневризмы аорты, расширения легочного ствола) – 7,7 % (5). В одинаковом объеме обнаружены пропуски плеврального выпота и иной патологии (значимый коронарный кальциноз, компрессионный перелом позвонка) – по 4,6 % (3) каждая категория.

Некоторые клинические примеры наглядно иллюстрируют проблематику (рисунки 4.2–4.4).



А



Б



В

Рисунок 4.2 – Пациент М., мужчина, 60 лет: А – ложноотрицательное решение врача, пропущен очаг в С10 левого легкого; Б, В – истинно положительное решение ИИ-сервиса, выявлен очаг в С10



А



Б

Рисунок 4.3 – Пациент Х., мужчина, 75 лет: А – ложноотрицательное решение врача, пропущен очаг в С10 правого легкого; Б – истинно положительное решение ИИ-сервиса, выявлен очаг в С10



А



Б

Рисунок 4.4 – Пациент Н., женщина, 33 года: А – ложноотрицательное решение врача, пропущен очаг в С1 правого легкого; Б – истинно положительное решение ИИ-сервиса, выявлен очаг в С1

Все указанные патологические находки были верно обнаружены, идентифицированы и маркированы ИИ-сервисами. Игнорирование врачом-рентгенологом результатов работы искусственного интеллекта привело к формированию ложноотрицательного решения и по прошествии времени – к клинически значимому расхождению. Мы не ставили перед собой задачу проследить судьбу пациентов. Сам факт выявления потенциально предотвратимых дефектов при описании результатов лучевых исследований (особенно связанных с онкологической диагностикой) служит убедительным доказательством необходимости применения врачом-рентгенологом ИИ-сервисов как систем поддержки принятия врачебных решений.

Таким образом, системы поддержки принятия врачебных решений на основе технологий искусственного интеллекта позволяют предотвратить клинически значимые дефекты (пропуск патологии) при интерпретации и описании результатов лучевых исследований. В структуре потенциально предотвратимых дефектов лидируют признаки новообразований (49,2 %) и неспецифических воспалительных изменений (24,6 %). Обращает на себя внимание удельный вес случаев гиподиагностики, связанный с некорректными или не проведенными измерениями диаметров магистральных сосудов (7,7 %).

Продемонстрированные возможности профилактики гиподиагностики следует считать основным доказательством клинической значимости СППВР на основе ИИ для врачей-рентгенологов.

В аспекте совершенствования методологий и нормативно-правового обеспечения в области контроля безопасности и качества медицинской помощи факт применения СППВР на основе ТИИ врачом-рентгенологом может рассматриваться как критерий качества оказания медицинской помощи, во всяком случае применительно к определенным нозологиям или клиническим ситуациям.

С позиции организации и управления медицинской помощью полагаем возможным предложить организационную технологию обеспечения безопасности лучевых исследований на основе обязательного предварительного машинного анализа результатов таковых исследований (рисунок 4.5).



Рисунок 4.5 – Схема организационной технологии обеспечения безопасности лучевых исследований на основе обязательного предварительного машинного анализа их результатов

Организационная технология направлена на профилактику дефектов в работе врача-рентгенолога при интерпретации и описании результатов лучевых исследований. Предварительным условием для ее реализации является наличие интеграции диагностических устройств в информационную систему (оптимально – на уровне субъекта, с созданием централизованного архива медицинских изображений, минимально – на уровне отдельной медицинской организации).

Первый элемент организационной технологии – подбор и интеграция медицинских изделий на основе технологий искусственного интеллекта в

указанную выше систему. Выбор конкретных медицинских изделий, в том числе в контексте решаемых ими клинических задач, должен быть сделан ответственным руководителем здравоохранения. В основу управленческого решения могут быть положены результаты предыдущих мероприятий по контролю качества работы врачей-рентгенологов (фокус на профилактику наиболее частых дефектов), данные по заболеваемости, неблагоприятным исходами и т.д.

Второй элемент – подготовка, согласование и утверждение локальных нормативно-правовых актов. В соответствующих документах должно быть закреплено требование по обязательному ознакомлению врача-рентгенолога с результатами автоматизированного анализа результатов лучевого исследования при его интерпретации и описании. Факт ознакомления может подтверждаться отметкой в информационной системе, разумеется при наличии такой технической возможности. Также, в локальных нормативно-правовых актах должно быть учтено, что внедрение обязательного ознакомления скорее всего приведет к некоторому увеличению длительности подготовки протокола врачом-рентгенологом. Однако, в этом случае значительный рост качества (за счет недопущения пропусков патологии) полностью оправдывает незначимое снижение производительности.

Третий элемент – это обучение врачей при внедрении новой технологии, разработка и обеспечение доступности материалов для методического сопровождения.

Четвертый элемент – мероприятия по контролю качества. Здесь предлагается такая форма, как аудит «врача и ИИ». Стандартный пересмотр уполномоченным врачом-экспертом результатов лучевого исследования и протокола врача (известный и массовый способ контроля качества) дополняется обязательным пересмотром результатов работы медицинского изделия на основе ИИ. В задачи врача-эксперта входит сравнительный анализ решений врача-рентгенолога и медицинского изделия с ИИ, выявление и оценка расхождений.

Результаты аудита становятся основой для совершенствования локальных нормативно-правовых актов, материалов для сопровождения врачей, но главное –

позволяют выстраивать новые мероприятия по непрерывному повышению качества работы врачей-рентгенологов и соответствующих структурных подразделений медицинских организаций.

Изучение эффективности обоснованной организационной технологии будет проведено в следующей главе.

4.2. Оценка безопасности и качества автономного применения искусственного интеллекта в профилактических лучевых исследованиях

Прикладные решения на основе технологий искусственного интеллекта для автоматизированного анализа результатов отдельных видов лучевых исследований (профилактических, а также рентгенографии стопы с целью выявления продольного плоскостопия) отличаются чувствительностью, теоретически сопоставимой со значением данного показателя среднестатистического врача-рентгенолога (0,95–1,0). Это позволило нам предложить идею применения ТИИ при анализе соответствующих исследований в автономном режиме.

Так называемый автономный искусственный интеллект подразумевает наличие технологической и методологической возможности передачи полномочий принятия части решений, обычно принимаемых человеком, математической модели. Относительно указанных выше лучевых исследований такая возможность появилась, как было показано ранее. При анализе результатов профилактических исследований (РГ/ФЛГ органов грудной клетки, маммографии) возможно осуществление автономной сортировки на две категории: «норма» и «не норма». При анализе результатов рентгенографии стопы – автономная морфометрия с последующей сортировкой также на две указанных категории. Исследования категории «норма» не интерпретируются и не описываются врачом-рентгенологом. Соответствующие результаты работы ИИ (очевидно являющегося медицинским изделием, допущенным к обращению в соответствии с действующим законодательством) сохраняются в карте пациента

в виде электронной медицинской записи, полученной в автоматическом режиме и не требующей подписания электронной подписью медицинского работника (в соответствии с приказом Министерства здравоохранения РФ от 07.09.2020 № 947н «Об утверждении Порядка организации системы документооборота в сфере охраны здоровья в части ведения медицинской документации в форме электронных документов»). С учетом средней специфичности данных ИИ-сервисов (0,6–0,8) исследования категории «не норма» направляются на интерпретацию и описание врачом-рентгенологом в стандартном порядке.

Применительно к рентгенографии стопы валидация идеи автономной морфометрии не проводилась, поэтому она и вошла в задачи нашего исследования.

Применительно к профилактическим лучевым исследованиям предварительная валидация идеи в лабораторных условиях на ограниченных наборах данных проведена в рамках иного исследования с ретроспективным дизайном [18, 19] и в диссертации не рассматривается. Своей задачей мы определили валидацию идеи в клинических условиях, то есть в проспективном дизайне, на потоке реальных данных.

4.2.1. Обоснование автономной морфометрии результатов рентгенографии стопы

Для оценки возможности автономной морфометрии выполнено экспериментальное исследование, заключающееся в математическом сравнении результатов автоматизированной морфометрии результатов рентгенографии стопы с целью выявления продольного плоскостопия и соответствующей разметки врачами-рентгенологами.

В таблице 4.1 представлено количество исследований, имеющих статистически значимые различия при сравнении врачей-рентгенологов между собой, сравнении врача и морфометрического ИИ-сервиса.

Таблица 4.1 – Количество статистически значимых различий при попарном сравнении

Сравниваемые пары	Количество различающихся результатов измерений	
	Абс.	Относит.
Врач 1 – Врач 2	21	22,6
Врач 1 – Врач 3	15	16,1
Врач 1 – Врач 4	16	17,2
Врач 2 – Врач 3	21	22,6
Врач 2 – Врач 4	16	17,2
Врач 3 – Врач 4	8	8,6
Среднее между врачами	16	17,4
Врач 1 – ИИ	30	32,2
Врач 2 – ИИ	36	38,7
Врач 3 – ИИ	32	34,4
Врач 4 – ИИ	18	19,4
Среднее между врачами и ИИ	29	31,2

Из результатов, приведенных в таблице 4.1, следует, что согласованность между врачами-рентгенологами наступает в 82,6 % случаев, между врачами и морфометрическим ИИ-сервисом – в 68,8 % случаев. Однако минимальное среднеквадратическое отклонение между мнениями экспертов составило 0,60, максимальное отклонение – 1,76 (на одном исследовании среднеквадратичное отклонение составило 28,26). Минимальное среднеквадратичное отклонение между экспертами и ИИ-сервисом составило 0,44, а максимальное – 1,33. Таким образом, средний уровень расхождения между результатами измерений врачами-рентгенологами и морфометрическим ИИ-сервисом ниже, чем средний уровень расхождения результатов измерений между врачами.

Значение среднеквадратичного отклонения измерений находится в приемлемых для клинических исследований величинах (исключение могут составлять пограничные случаи отнесения к норме или патологии и степеням патологии.) В таблице 4.2 представлены средние значения по всем исследованиям угла свода стопы, оцененные для каждого врача и морфометрического ИИ-сервиса, а также соответствующие среднеквадратичные отклонения.

Таблица 4.2 – Средние значения и среднеквадратичные отклонения измерений угла свода стопы (n = 100)

Эксперт	Средний угол, градусы	Среднеквадратичное отклонение, абс.
Врач 1	136,2876	10,69955
Врач 2	136,2204	10,68868
Врач 3	134,8582	17,18463
Врач 4	134,9332	14,46209
ИИ-сервис	135,5413	11,15616

По результатам проведенной оценки средних значений и среднеквадратических отклонений можно сделать вывод, что результаты измерений угла свода правой стопы, полученные автоматизировано, не отличаются от результатов врачей-рентгенологов.

С учетом приведенных выше результатов проспективного клинического исследования диагностической точности и клинической результативности можно утверждать, что достигнутый уровень развития ТИИ для лучевой диагностики позволяет реализовать автономную морфометрию, по крайней мере для оценки результатов рентгенографии стопы с целью выявления продольного плоскостопия.

Вместе с тем необходимо отметить, что, с прикладной точки зрения, конкретный ИИ-сервис, включенный в исследование, требует некоторого усовершенствования и максимального повышения чувствительности.

Таким образом, потенциально ТИИ способны работать полностью автономно, проводя необходимые измерения (морфометрию) и создавая заключение о наличии и степени плоскостопия. Важным преимуществом использования данного подхода является объективизация измерений, поскольку, как было показано в исследовании, результат повторного измерения врача отличается от первоначального, тогда как результат измерения сервиса всегда одинаков и не зависит от количества повторений.

4.2.2. Проспективное клиническое исследование автономной сортировки результатов профилактических лучевых исследований

В проспективном формате за период исследования ИИ-сервисами проанализированы результаты 49 221 *профилактических маммографий*, выполненных в амбулаторных условиях в медицинских организациях Департамента здравоохранения г. Москвы.

К категории «норма» при автоматизированном анализе было отнесено 7 028 ММГ, что составило 14,3 %. К категории «не норма» – 42 193, или 85,7 % ММГ. В таблице 4.3 представлены данные о количестве и результатах автоматизированного анализа включенными в исследование ИИ-сервисами.

Таблица 4.3 – Итоги работы ИИ-сервисов для автоматизированного анализа результатов профилактической маммографии

ИИ-сервис	Количество проанализированных исследований, абс.	Количество исследований категории «норма»		Количество исследований категории «не норма»	
		Абс.	Относит.	Абс.	Относит.
ИИ-1	10 881	1 045	9,6	9 836	90,4
ИИ-2	2 150	1 109	51,6	1 041	48,4
ИИ-3	36 190	4 874	13,5	31 316	86,5
Итого	49 221	7028	14,3	42193	85,7

Наибольшее количество исследований было проанализировано ИИ-3 (36 190), наименьшее (2 150) – ИИ-2. Различия были обусловлены внешними обстоятельствами: либо технологическими ограничениями конкретного прикладного решения (ИИ-1), либо задействованием основных ресурсов сервиса на оказание медицинской услуги с применением технологий ИИ в рамках системы обязательного медицинского страхования г. Москвы (ИИ-2). Больше количество исследований к категории «норма» отнес ИИ-2 (51,6 %), наименьшее – ИИ-1 (9,6 %). Соответствующий удельный вес результатов профилактической маммографии, отнесенной по итогам автоматизированного анализа к категории «не норма», составил 48,4 % и 90,4 % соответственно.

В контексте этой части диссертационного исследования сопоставление ИИ-сервисов между собой не играет критичной роли. Более важным является сопоставление результатов ИИ с решениями врачей-рентгенологов.

Всего к категории «норма» было отнесено 7 028 (14,3 %) случаев, к «не норма» – 42 193 (85,7 %). В рамках идеи автономной сортировки случаи «не нормы» направляются на интерпретацию и описание врачом-рентгенологом. Поэтому мы подвергли детальному анализу, напротив, случаи, отнесенные к категории «норма». Собственно, очевидно, что ключевой задачей для ИИ в текущем контексте является именно правильное определение результатов маммографии, вовсе не содержащих каких-либо патологических признаков, аномалий и артефактов.

На первом этапе для каждого клинического случая ($n = 7028$) проведено сопоставление решений ИИ и врача-рентгенолога, проводившего самостоятельное первичное описание, об отнесении результатов данного исследования к категории по шкале «BI-RADS».

Полное согласие ИИ и врача-рентгенолога зафиксировано в 99,0 % (6 960) случаев, расхождения выявлены в 1,0 % (69) случаев.

На этом этапе согласованность решений ИИ и врачей-рентгенологов, проводивших первичные описания результатов профилактической ММГ, определена нами уже как высокая – коэффициент каппа Коэна составил 0,994 (95%-й ДИ 0,992–0,996).

На втором этапе проведен пересмотр врачом-экспертом с субспециализацией по маммографии результатов исследований ($n = 69$), в которых выявлены расхождения решения ИИ и врача-рентгенолога, проводившего первичное описание.

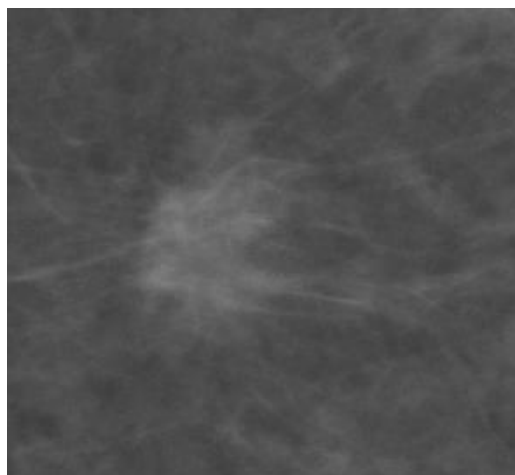
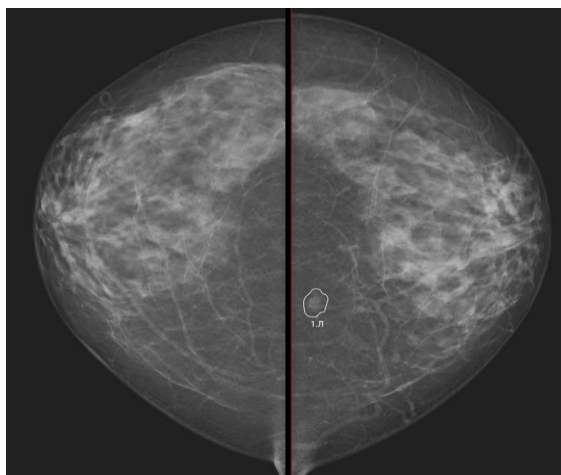
Решение ИИ «норма» было подтверждено в 41,2% (29) случаев, решение врача «не норма» – в 58,8 % (40). Таким образом, ИИ допустил неправильную сортировку в 40 случаях, из которых врачом-экспертом 80,0 % (32) были отнесены к категории BI-RADS-3, а лишь 17,5 % (7) – к категории BI-RADS-4 и 2,5 % (1) – к категории BI-RADS-5.

На этом этапе также согласованность всех решений была очень высока. Для решений врачей-рентгенологов, проводивших первичные описания результатов профилактической ММГ, и врачей-экспертов коэффициент каппа Коэна составил 0,99564 (95%-й ДИ 0,99397–0,99732), а для технологий ИИ и врачей-экспертов – 0,99649 (95%-й ДИ 0,99499–0,99799).

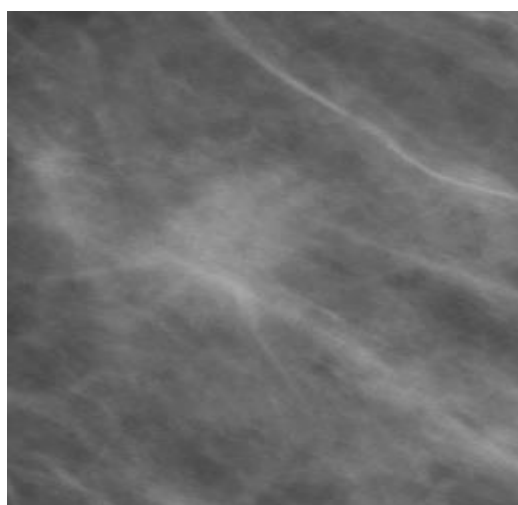
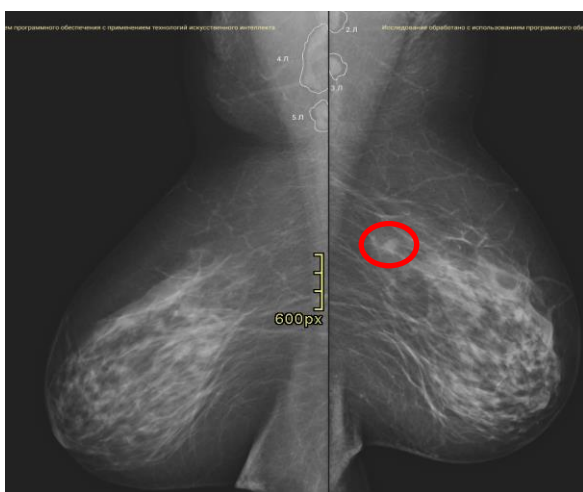
Именно 8 случаев, для которых врач-эксперт установил истинные категории BI-RADS-4 и BI-RADS-5, мы полагаем ситуациями клинически значимых расхождений.

На 49 221 проанализированных результатов профилактической маммографии удельный вес ошибочной автономной сортировки с применением технологий ИИ составил 0,08 % (40); при этом клинически значимое расхождение ИИ-сервисы допустили лишь в 8 случаях, что составило 0,02% от общего количества проанализированных исследований. В 99,92 % (49 181) случаев автономная сортировка была выполнена полностью корректно.

На рисунках 4.6–4.8 представлены клинические примеры.

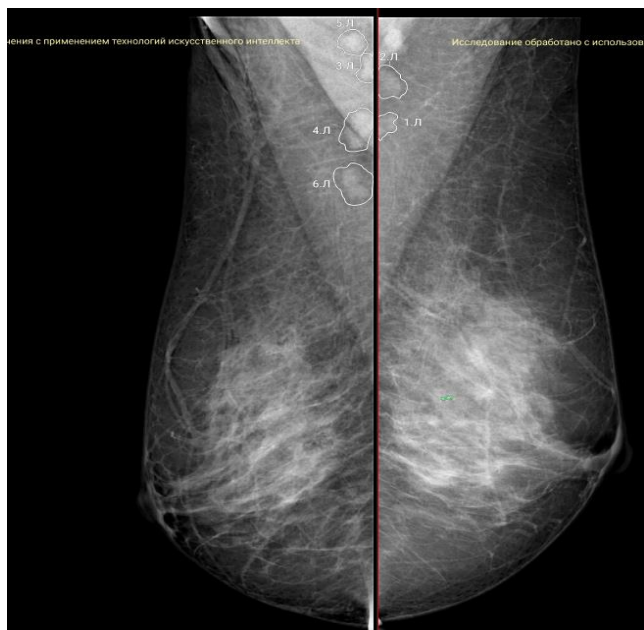


А

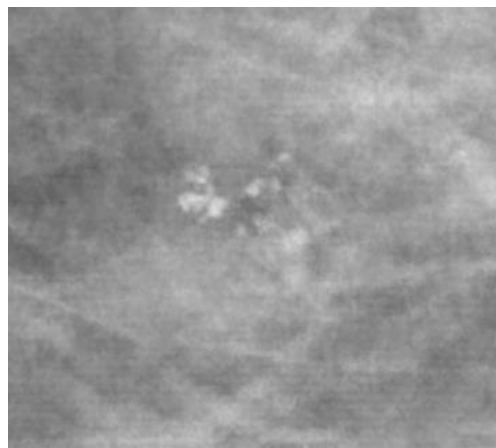
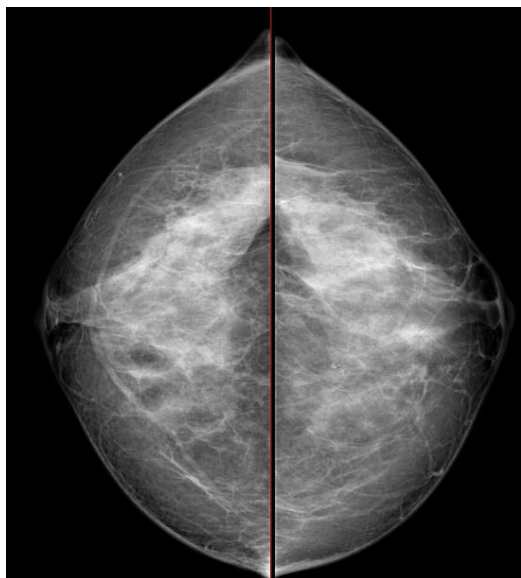


Б

Рисунок 4.6 – Пациент К., женщина, 54 г., профилактическая маммография: А – ложноотрицательное решение ИИ-сервиса, неправильная классификация патологического участка слева BI-RADS-2; Б – врачом-экспертом определена категория BI-RADS-4c

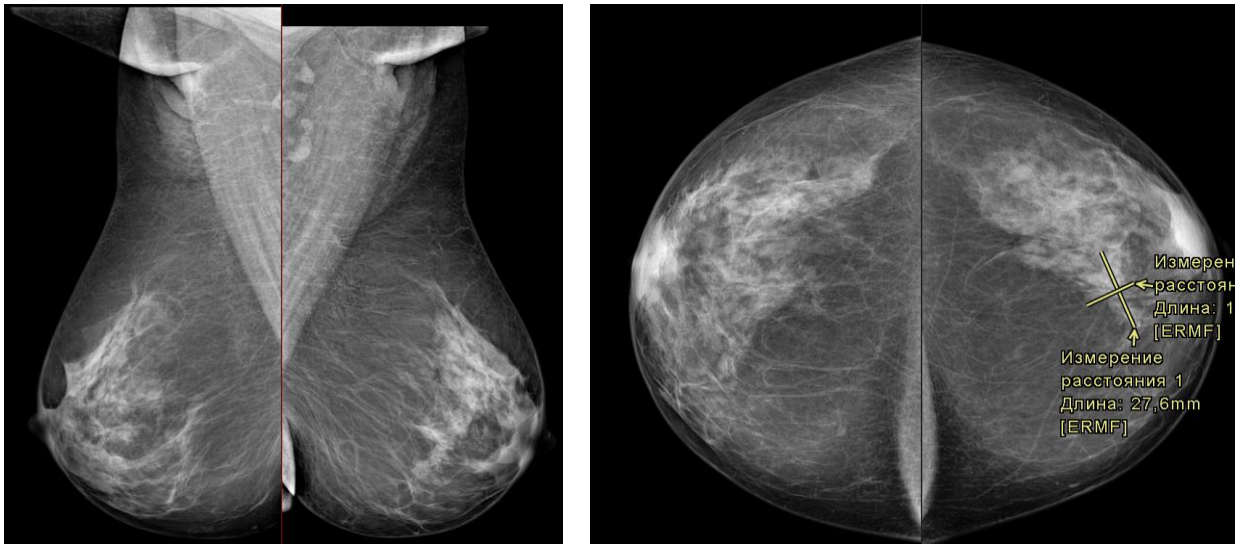


А



Б

Рисунок 4.7 – Пациент К., женщина, 46 л., профилактическая маммография:
 А – ложноотрицательное решение ИИ-сервиса, пропущены патологические
 кальцинаты, неправильная классификация патологического участка
 слева BI-RADS-2; Б – врачом-экспертом определена категория BI-RADS-4b



А



Б

Рисунок 4.8 – Пациент Щ., женщина, 57 л., профилактическая маммография:
 А – ложноположительное решение врача-рентгенолога, некорректное определение патологического участка в прямой проекции (суммация теней), неправильная классификация слева BI-RADS 4; Б – корректная оценка ИИ-сервисом, определена категория BI-RADS-2

Вполне очевиден тезис, что диагностическая точность в любом случае и всегда должна равняться 100,0 %. Объективно такое утверждение может быть принято как экзистенциальная стратегическая цель. Однако в реальности всегда остается некий процент несовершенства и дефектов. В развитии диагностики полагаем нужным стремиться: 1) к сопоставимости новых и старых методов;

2) к превышению точности новых методов над старыми со стратегическим стремлением всех показателей точности к единице. Все это совершается при безусловном обеспечении безопасности пациента, соблюдении норм законодательства, биоэтики и т. д.

Полученный удельный вес дефектуры ИИ необходимо сопоставить с объективными данными о точности врача-рентгенолога при интерпретации результатов профилактической маммографии.

По литературным данным, удельный вес клинически значимых дефектов в работе врача-рентгенолога при интерпретации результатов профилактической ММГ составляет от 3,5 % до 30,0 % [148, 169, 173, 183, 206, 217]. Подобные исследования обычно проводятся на крайне малых группах врачей-рентгенологов (до 10 человек); в целом гетерогенность таких работ очень высока и произвести их систематизацию проблематично.

Ранее в ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ» проведено специальное исследование с целью определить средние показатели точности «среднестатистического» врача-рентгенолога при интерпретации результатов профилактической маммографии. На выборке 113 врачей-рентгенологов установлены искомые средние уровни (таблица 4.4) [5, 146, 241].

Таблица 4.4 – Научно обоснованные средние показатели точности врача-рентгенолога при интерпретации результатов профилактической маммографии [5, 146, 241]

Показатели точности	Значение
AUROC	0,928
95%-й ДИ	0,883–0,976
Чувствительность	0,792
95%-й ДИ	0,677–0,907
Специфичность	0,940
95%-й ДИ	0,874–1,000

Исходя из научно обоснованных средних показателей удельный вес дефектов в работе врача-рентгенолога при интерпретации результатов

профилактической маммографии составляет 3,0–7,2 %. Выявленный нами удельный вес дефектов в работе ИИ-сервисов, имеющих статус медицинского изделия и работающих на потоке данных в реальных клинических условиях, при автономной сортировке составляет 0,08 % (клинически значимых – 0,02 %).

Таким образом, автономная сортировка результатов профилактической маммографии превышает по своей безопасности и точности усредненного врача-рентгенолога. Согласованность решений врача-рентгенолога, выполняющего первичное описание, врача-эксперта и ИИ-сервиса чрезвычайно высока (каппа Коэна > 0,99). Практическая реализация соответствующей медицинской услуги потенциально позволит оптимизировать ресурсы системы здравоохранения, сохранив или даже повысив качество профилактической маммографии.

В проспективном формате за период исследования ИИ-сервисами проанализированы результаты 209 497 **профилактических исследований органов грудной клетки (рентгенографий и флюорографий)**, выполненных в амбулаторных условиях в медицинских организациях Департамента здравоохранения г. Москвы.

К категории «норма» при автоматизированном анализе было отнесено 117 041 РГ/ФЛГ, что составило 55,9 %; к категории «не норма» – 92 456, что составило 44,1 %. В таблице 4.5 представлены данные о количестве и результатах автоматизированного анализа включенными в исследование ИИ-сервисами.

Таблица 4.5 – Итоги работы ИИ-сервисов для автоматизированного анализа результатов профилактической рентгенографии/флюорографии

ИИ-сервис	Количество проанализированных исследований, абс.	Количество исследований категории «норма»		Количество исследований категории «не норма»	
		Абс.	Относит.	Абс.	Относит.
ИИ-1	28 200	16 213	57,5	11 987	42,5
ИИ-2	91 659	44 539	48,6	47 120	51,4
ИИ-3	89 638	56 289	62,8	33 349	37,2
Итого	209 497	117041	55,9	92456	44,1

Наибольшее количество исследований было проанализировано ИИ-2 (91 659), наименьшее – ИИ-1 (28 200). Различия были обусловлены технологическими ограничениями конкретного прикладного решения (ИИ-1).

Большее количество исследований к категории «норма» отнес ИИ-3 (62,8 %), наименьшее – ИИ-2 (48,6 %). Соответствующий удельный вес результатов РГ/ФЛГ, отнесенных по итогам автоматизированного анализа к категории «не норма», составил 37,2 % и 51,4 % соответственно.

Как было сказано выше, детальное сравнение работы ИИ-сервисов в этой задаче не является необходимым. Сопоставлены результаты работы ИИ с решениями врачей-рентгенологов.

Всего к категории «норма» было отнесено 117 041 (55,9 %) случаев, к «не норма» – 92 456 (44,1 %). Детально проанализированы случаи, отнесенные ИИ-сервисами к категории «норма».

На первом этапе для каждого клинического случая категории «норма» ($n = 117\,041$) проведено сопоставление решений ИИ и врача-рентгенолога, осуществлявшего самостоятельное первичное описание. Исследованы решения о наличии патологических признаков, коррелирующих с социально значимыми заболеваниями и рядом иных клинически значимых состояний.

Полное согласие ИИ и врача-рентгенолога зафиксировано в 99,7 % (116 667) случаев, расхождения выявлены в 0,3 % (374) случаев.

На этом этапе согласованность решений ИИ и врачей-рентгенологов, проводивших первичные описания результатов профилактической РГ/ФЛГ, уже была высокой – коэффициент каппа Коэна составил 0,988 (95%-й ДИ 0,987–0,990).

На втором этапе проведен пересмотр врачом-экспертом с субспециализацией по торакальной радиологии результатов исследований ($n = 374$), в которых выявлены расхождения решения ИИ и врача-рентгенолога, проводившего первичное описание.

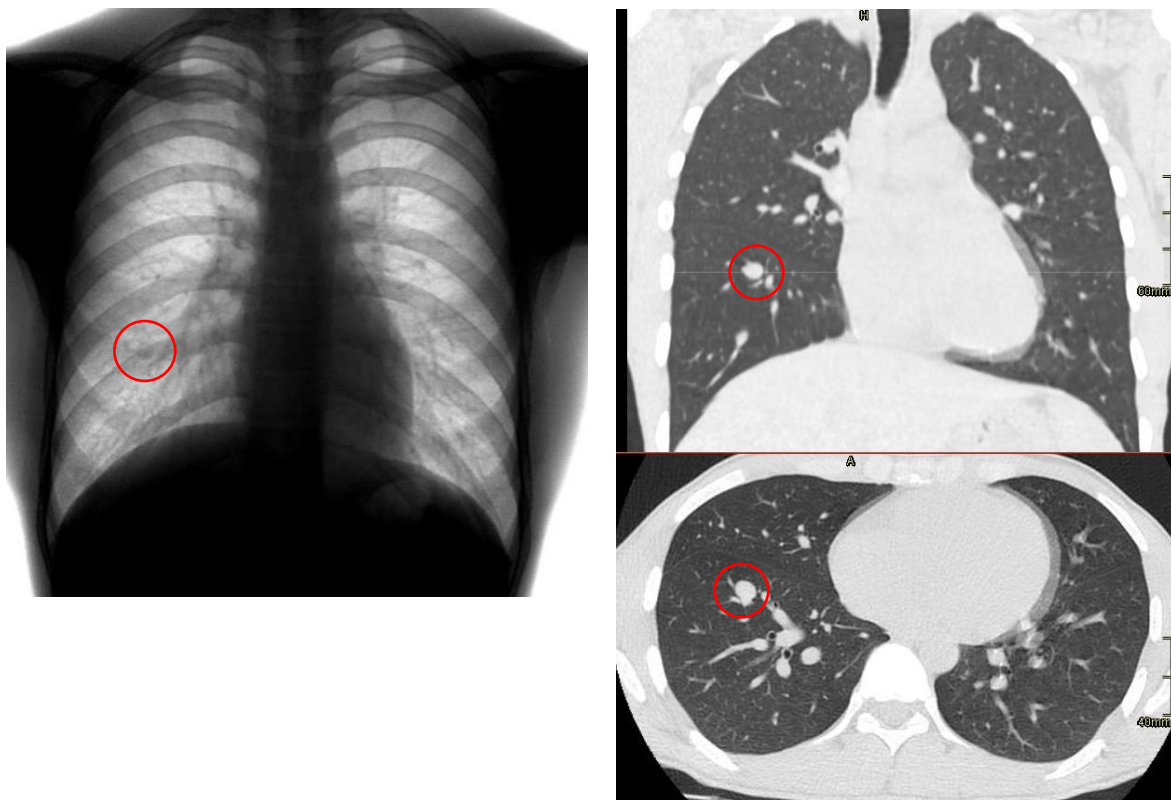
Решение ИИ «норма» было подтверждено в 56,7 % (212) случаях, решение врача «не норма» – в 43,3 % (162). Таким образом, ИИ допустил неправильную

сортировку в 162 случаях, из которых врачом-экспертом в 29,0 % (47) случаев определена клинически значимая патология, а в 71,0 % (115) – клинически незначимые отклонения от нормы. К последним относились кальцинаты, консолидированные переломы и иные состояния, непредусмотренные первоначальной клинической задачей (например, неосложненная диафрагмальная грыжа, фиброз, пневмосклероз, консолидированный перелом ребра и т.д.). В качестве клинически значимых расхождений выявлены: очаги – 76,6% (36), инфильтрация/консолидация – 8,5% (4), образование – 6,4% (3), диссеминация – 4,3% (2), пневмоторакс или расширение средостения – по 2,1% (по 1 случаю).

На этом этапе также согласованность всех решений была очень высока. Для решений врачей-рентгенологов, проводивших первичные описания результатов профилактической РГ/ФЛГ, и врачей-экспертов коэффициент каппа Коэна составил 0,9915 (95%-й ДИ 0,9906–0,9925), а для технологий ИИ и врачей-экспертов – 0,9968 (95%-й ДИ 0,9962–0,9974).

Из 209 497 проанализированных результатов профилактических лучевых исследований органов грудной клетки удельный вес ошибочной автономной сортировки с применением технологий ИИ составил лишь 0,08% (162), при чем клинически значимые расхождения выявлены только в 0,02 % (47) случаев. В 99,92 % (209 335) случаев автономная сортировка была выполнена полностью корректно.

На рисунках 4.9–4.11 представлены клинические примеры.



А

Б

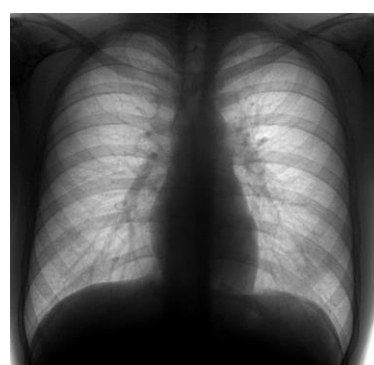
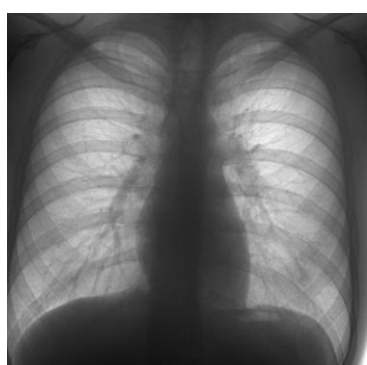
Рисунок 4.9 – Пациент Р., мужчина, 23 г., профилактическое рентгенологическое исследование органов грудной клетки: А – ложноотрицательное решение ИИ-сервиса, пропущена очаговая тень диаметром 9 мм в IV межреберье правого легкого; Б – результаты КТ ОГК, выполненной для верификации находки по назначению врача-рентгенолога



А

Б

Рисунок 4.10 – Пациент А., мужчина, 48 л., профилактическое рентгенологическое исследование органов грудной клетки: А – ложноотрицательное решение ИИ-сервиса, пропущены очаговые уплотнения легочной ткани верхней доли левого легкого; Б – результаты КТ ОГК, выполненной для верификации находки по назначению врача-рентгенолога



А

Б

Рисунок 4.11 – Пациент В., мужчина, 19 л., профилактическое рентгенологическое исследование органов грудной клетки: А – ложноположительное решение врача-рентгенолога, заподозрено образование с левой стороны; Б – проверка врачом-экспертом, на архивной рентгенографии 2022 г. определяется тень соска

Как и в случае с маммографией, удельный вес дефектуры ИИ при автономной сортировке сопоставлен с объективными данными о точности врача-рентгенолога при интерпретации результатов профилактической РГ/ФЛГ.

Ранее в ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ» проведено специальное исследование с целью определить средние показатели точности «среднестатистического» врача-рентгенолога при интерпретации результатов профилактических исследований органов грудной клетки. На выборке 254 врачей-рентгенологов установлены искомые средние уровни (таблица 4.6) [6, 146, 241].

Таблица 4.6 – Научно обоснованные средние показатели точности врача-рентгенолога при интерпретации результатов профилактической рентгенографии (флюорографии) органов грудной клетки [6, 146, 241]

Показатели точности	Модальность	
	Рентгенография	Флюорография
AUROC	0,96	0,97
95%-й ДИ	0,93–1,0	0,94–1,0
Чувствительность	0,9	1,0
95%-й ДИ	0,82–0,99	1,0–1,0
Специфичность	0,96	0,91
95%-й ДИ	0,9–1,0	0,83–0,99

Исходя из научно обоснованных средних показателей удельный вес дефектов в работе врача-рентгенолога при интерпретации результатов профилактической РГ/ФЛГ составляет 3,0–4,0 %. Выявленный нами удельный вес дефектов в работе ИИ-сервисов, имеющих статус медицинского изделия и работающих на потоке данных в реальных клинических условиях, при автономной сортировке составляет 0,08 %.

Можно утверждать, что автономная сортировка результатов профилактической рентгенографии (флюорографии) органов грудной клетки превышает по своей безопасности и точности усредненного врача-рентгенолога. Согласованность решений врача-рентгенолога, выполняющего первичное описание, врача-эксперта и ИИ-сервисов чрезвычайно высока (каппа

Коэна $> 0,99$). Практическая реализация соответствующей медицинской услуги потенциально позволит оптимизировать ресурсы системы здравоохранения, сохранив или даже повысив качество профилактических исследований органов грудной клетки.

Таким образом, в результате проспективного клинического исследования установлено, что согласованность решений ИИ-сервисов (медицинских изделий) и врачей-рентгенологов (как линейных сотрудников, так и экспертов) высока – каппа Коэна $> 0,99$. Удельный вес дефектов, то есть ложных отнесений результатов исследований к категории «норма» составляет 0,08 % для всех проанализированных модальностей; при этом удельный вес ложных решений, имеющих клиническую значимость составляет лишь 0,02 %. Указанный уровень дефектов значительно ниже, чем у врача-рентгенолога, исходя из научно обоснованных усредненных показателей диагностической точности. На реальных клинических данных доказана безопасность и целесообразность автономной сортировки результатов профилактических лучевых исследований с применением технологий искусственного интеллекта.

Глава 5. ОРГАНИЗАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОБЕСПЕЧЕНИЯ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РЕНТГЕНОЛОГИИ

На основании проведенных исследований, отдельных данных литературы и собственного практического опыта организации и применения ИИ в рентгенологии разработан *подход* к организации внедрения и применения технологий искусственного интеллекта с целью автоматизации и повышения результативности производственных процессов в рентгенологии. Его суть состоит в многоуровневном рассмотрении проблематики предметной области.

Первый уровень решения проблемы — это применение обоснованного клинического целеполагания и стандартизированных подходов к разработке ИИ для рентгенологии.

Второй уровень – это обязательная апробация в реальных клинических условиях с накоплением информации для доработки ИИ-сервиса, его максимальной адаптации для решения конкретной задачи в клиническом контексте.

Третий уровень – унификация способов и клинического контекста применения ИИ. Под последним понимается совокупность модальностей лучевых исследований, групп нозологий, анатомические области и структуры. Например, для СППВР доказана возможность и целесообразность применения при интерпретации результатов МРТ, выполненной с целью диагностики демиелинизирующего заболевания – рассеянного склероза; для морфометрии – при интерпретации результатов КТ, для измерения диаметра отделов аорты с целью выявления аневризмы; для автономной сортировки – категорирования результатов профилактической маммографии и т. д.

Четвертый уровень представляет собой «встраивание» унифицированных способов применения ИИ в стандартный процесс выполнения, интерпретации и описания результатов лучевого исследования.

Для практической реализации подхода, на основе результатов исследования

предложены организационные технологии: а) создания и внедрения технологий искусственного интеллекта в рентгенологии; б) реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта.

Совместное использование предложенных организационных технологий позволит обеспечить системную и результативную автоматизацию этапов производственных процессов медицинских организаций, связанных с выполнением, интерпретацией и описанием результатов лучевых исследований.

Организационная технология создания и внедрения технологий искусственного интеллекта в рентгенологии.

Технология обеспечивает реализацию I–III уровней подхода, она основана на унификации и алгоритмизации процессов разработки, обеспечения качества, а также способов применения технологий ИИ (схематически представлена на рисунке 5.1).

Ключевым элементом является постановка клинической задачи. Она выполняется на основе совокупного анализа информации из нескольких источников. Первым служит база данных, позволяющая проанализировать структуру и востребованность лучевых исследований на уровне административно-территориальной единицы (оптимально – субъекта РФ). Чаще всего нужные статистические данные могут быть получены из централизованного архива медицинских изображений, аналитических систем в составе информационной системы в сфере здравоохранения субъекта РФ. Вторым источником служат результаты внутреннего и ведомственного контроля качества. В ходе соответствующих мероприятий в работе врачей-рентгенологов могут выявляться типичные дефекты и проблемы, носящие системный характер. Обобщение таких данных позволяет сформировать предложения по автоматизации определенных действий врача-рентгенолога и поддержке решений.

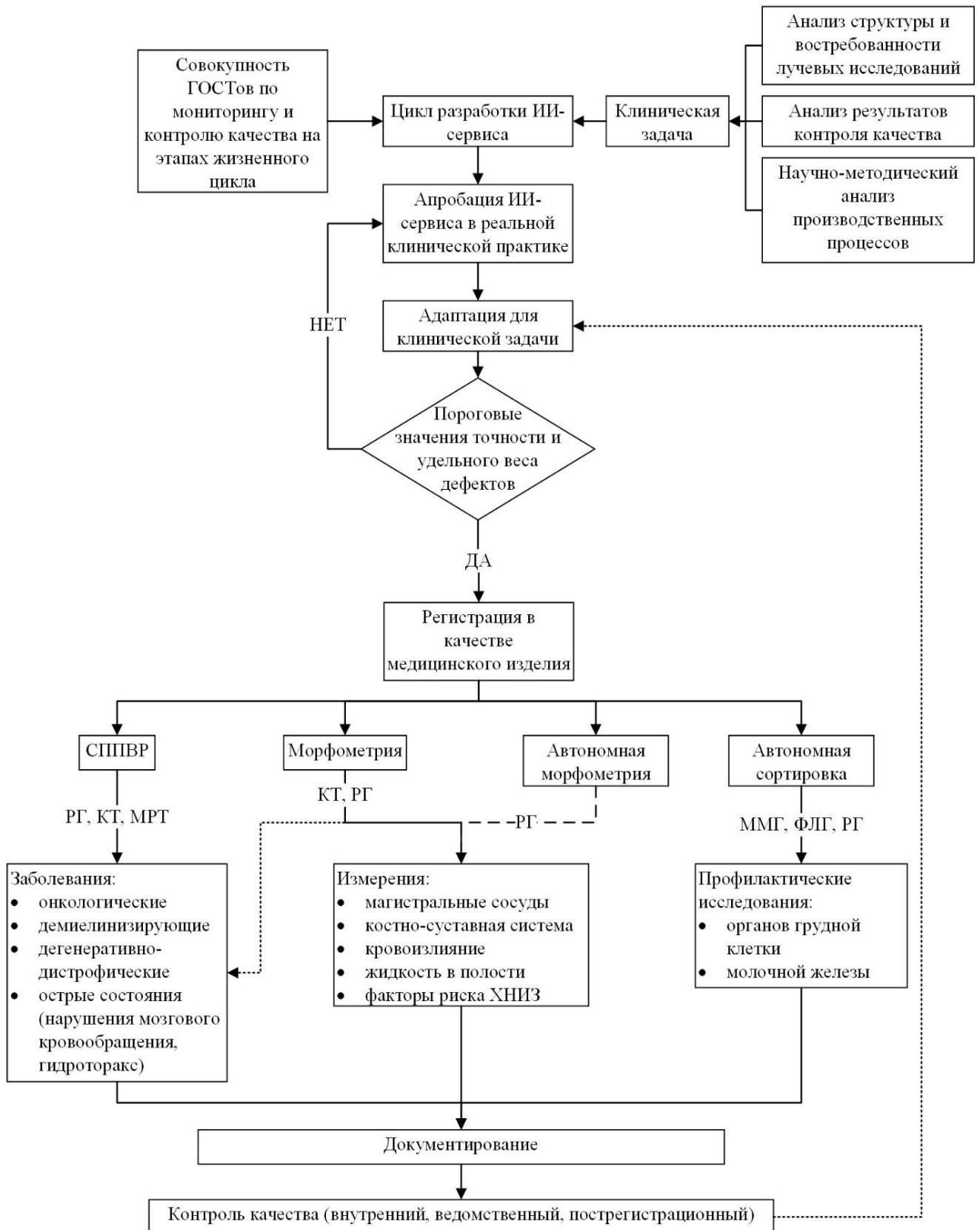


Рисунок 5.1 – Организационная технология создания и внедрения технологий искусственного интеллекта в рентгенологии

Наконец, третьим источником служит оригинальный анализ производственных процессов в лучевой диагностике. Таковым может служить

собственный эмпирический опыт или специальная научно-исследовательская работа.

На основе обоснованной клинической задачи осуществляется разработка конкретного ИИ-сервиса. Для поддержки этого процесса применяется совокупность национальных стандартов, регламентирующих этапы жизненного цикла соответствующих разработок, подготовку нужных наборов данных, этапную оценку качества и т.д.

Следующим обязательным и критично важным элементом полагаем тестирование ИИ-сервиса в реальных клинических условиях, обязательно в формате именно проспективного клинического исследования. В ходе такового можно максимально адаптировать ИИ-сервис для решения целевой клинической задачи и добиться требуемого уровня показателей точности и клинической результативности. По нашему практическому опыту этому очень способствует тесное взаимодействие разработчиков с пользователями – врачами-рентгенологами, причем именно в проспективном варианте.

Как было показано выше, клиническая апробация является обязательным этапом разработки ИИ-сервиса. Она не может подменять собой клинические испытания в целях регистрации медицинского изделия, так как направлена на получение объективной информации о точности и результативности ИИ в реальных клинических условиях. По результатам апробации осуществляется доработка и технологическое усовершенствование.

По достижении нужного качества осуществляется государственная регистрация ИИ-сервиса в качестве медицинского изделия по предусмотренной законодательством процедуре.

Следующий элемент – собственно клиническое использование. С позиции методологии выделены 4 основные способа применения ИИ в лучевой диагностике: поддержка принятия врачебных решений, морфометрия, автономная морфометрия, автономная сортировка.

Проведенное нами проспективное исследование работы ИИ-сервисов в условиях реальной клинической практики позволило нам вывести следующие

положения. Достигнутый уровень развития технологий ИИ позволяет использовать их в виде:

1. Системы поддержки принятия врачебных решений при описании результатов компьютерной томографии, рентгенографии и магнитно-резонансной томографии с целью диагностики онкологических, демиелинизирующих, дегенеративно-дистрофических заболеваний, а также ряда острых состояний, в частности, гидроторакса, острых нарушений мозгового кровообращения по геморрагическому и ишемическому типам.

2. Инструмента морфометрии при описании результатов компьютерной томографии и рентгенографии для выполнения измерений диаметров магистральных сосудов, объемов кровоизлияния или жидкости в полости (например, в плевральной), а также для выявления нарушений со стороны опорно-двигательной системы (например, плоскостопия). Возможной задачей для морфометрии является определение факторов риска хронических неинфекционных заболеваний (ХНИЗ), например, плотности паренхимы печени, кальциевого индекса, объема паракардиального жира и т.д. Однако именно этот вопрос в диссертации не рассматривается. В настоящее время достигнутый технологический уровень позволяет использовать морфометрию только с валидацией измерений врачом-рентгенологом, то есть в совокупности с СППВР (такое взаимодействие на рисунке 5.1 обозначено стрелкой с мелким пунктиром). Вместе с тем наметилась тенденция возможности введения автономной морфометрии с целью выявления нарушений со стороны опорно-двигательной системы (на рисунке 5.2 обозначено стрелкой с крупным пунктиром), соответствующие доказательства приведены в тексте диссертации.

3. Инструмента автономной сортировки при интерпретации результатов профилактической рентгенографии (флюорографии) органов грудной клетки и профилактической маммографии.

Следующими элементами являются документирование (детально этот аспект будет разобран далее) и контроль качества. Мероприятия по мониторингу и обеспечению качества медицинской помощи, оказываемой с применением

медицинских изделий на основе искусственного интеллекта, осуществляются в форме внутреннего контроля в медицинских организациях, ведомственного контроля на уровне органов исполнительной власти в сфере здравоохранения субъектов РФ, наконец – в форме пострегистрационного мониторинга медицинских изделий со стороны уполномоченных органов и учреждений.

Результаты мероприятий по контролю качества доводятся до разработчика и применяются для совершенствования ИИ-сервиса, лучшей адаптации его для решения целевой клинической задачи.

Организационная технология реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта

Технология обеспечивает реализацию IV уровня подхода, она основана на унификации и алгоритмизации процессов проведения, интерпретации и описания результатов лучевых исследований с применением технологий ИИ (схематически представлена на рисунке 5.2).

Назначение лучевого исследования осуществляется лечащим врачом на основе клинических рекомендаций, иных стандартизирующих и методологических документов, а также, возможно, с применением специальных СППВР. Исследование выполняется рентгенолаборантом. При этом целесообразно использовать автоматизированный контроль качества, то есть специальные ИИ-сервисы для быстрой оценки корректности выполнения исследования (укладки пациента, наличия артефактов и проч.). Развитие таких инструментов находится, в целом, на ранних стадиях и не входит в задачи диссертационной работы.

Результаты исследования сохраняются в централизованном или локальном архиве медицинских изображений в соответствующей информационной системе и направляются на анализ релевантным ИИ-сервисом. В зависимости от модальности, вида исследования и соответствующего способа применения ИИ происходит разделение на оказание стандартной медицинской услуги (основанной на существующем подходе к медицинскому вмешательству²⁰) либо

²⁰ Медицинское вмешательство – выполняемые медицинским работником и иным

автономной медицинской услуги (то есть на вмешательстве, выполняемом без участия медицинского или иного уполномоченного работника, и имеющем профилактическую направленность).

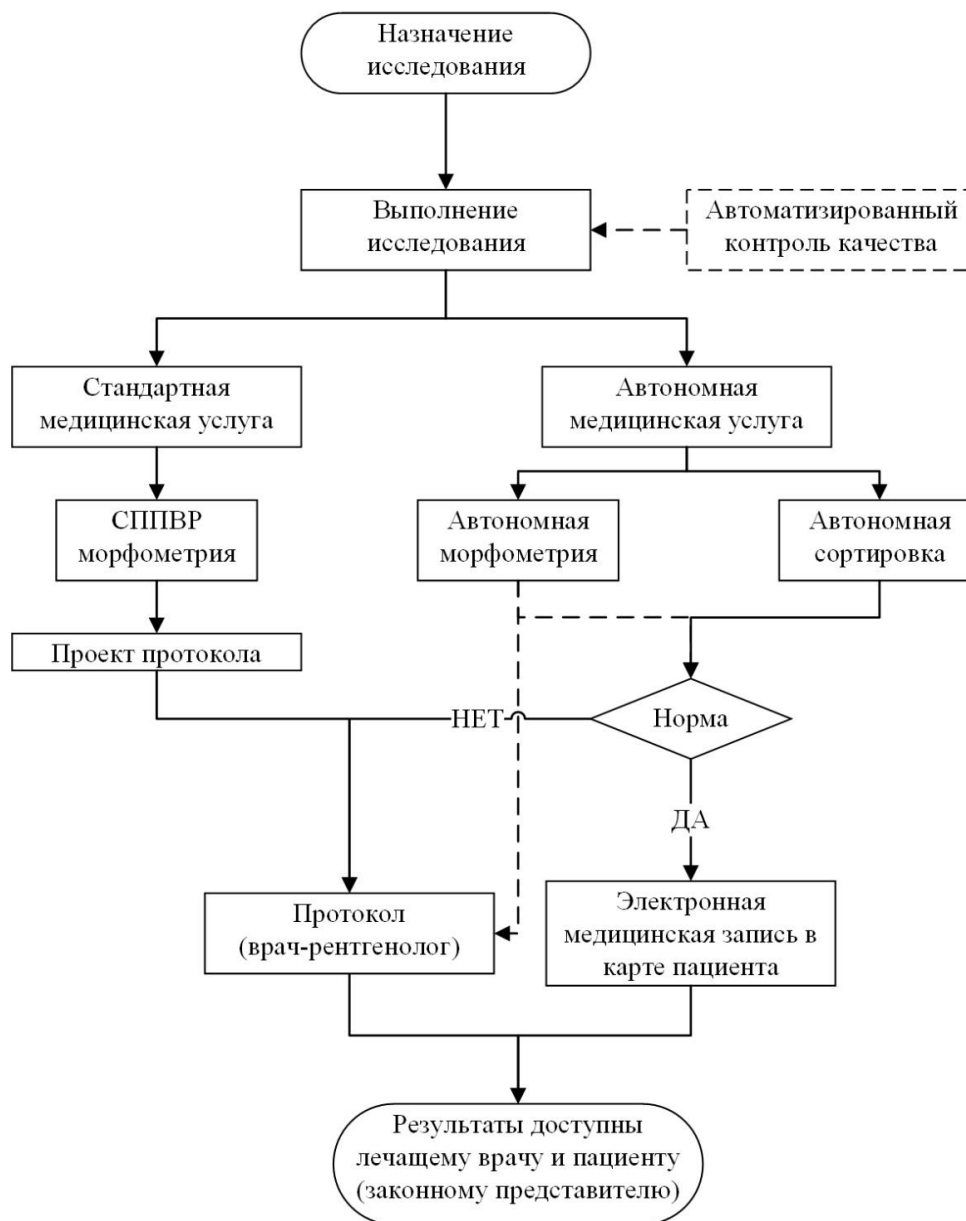


Рисунок 5.2 – Организационная технология реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта

работником, имеющим право на осуществление медицинской деятельности, по отношению к пациенту, затрагивающие физическое или психическое состояние человека и имеющие профилактическую, исследовательскую, диагностическую, лечебную, реабилитационную направленность виды медицинских обследований и (или) медицинских манипуляций, а также искусственное прерывание беременности (п. 5, ст. 2, Федерального закона от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации»).

Отметим, что детальная проработка юридических аспектов автономных медицинских услуг не входит в задачи диссертации. Скорее наше исследование создает научную основу для разработки и внедрения предложений по актуализации нормативно-правового регулирования охраны здоровья граждан.

В рамках стандартной медицинской услуги используются системы поддержки принятия врачебных решений на основе технологий ИИ. При этом, как было сказано выше, могут применяться инструменты морфометрии. В результате работы ИИ формируется проект протокола лучевого исследования. Далее врач-рентгенолог интерпретирует результаты лучевого исследования с обязательным применением данных от ИИ-сервисов. После обобщения всей информации формируется окончательный текст протокола, который оформляется и вносится в медицинскую документацию в установленном порядке.

В рамках автономной медицинской услуги могут выполняться морфометрия и сортировка.

Автономная сортировка допустима для результатов профилактических лучевых исследований. Как было показано выше, таковыми в настоящее время являются маммография и рентгенография (флюорография) органов грудной клетки. Максимальная чувствительность (1,0) соответствующих ИИ-сервисов гарантирует предельный минимум пропусков патологии, который при работе в реальных клинических условиях, как было показано выше, не превышает 0,08 %.

Сортировка предполагает разделение результатов профилактических лучевых исследований на две категории: «норму» и «не норму». Результаты категории «норма» сохраняются в формате электронной медицинской записи²¹ в карте пациента. Результаты категории «не норма» маршрутизируются для интерпретации и описания врачом-рентгенологом; по результатам оформляется протокол.

Результаты автономной морфометрии могут использоваться согласно двум сценариям:

²¹ 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».

– вносится в окончательный протокол лучевого исследования, который визируется подписью врача-рентгенолога, однако валидация результатов соответствующих измерений врачом не осуществляется;

– при отсутствии патологических отклонений результаты измерений сохраняются в формате электронной медицинской записи в карте пациента, при наличии патологических отклонений – маршрутизируются в виде проекта протокола для валидации врачом-рентгенологом.

Вне зависимости от вида итогового документа в информационной системе (протокол с подписью врача или медицинская электронная запись), результаты лучевого исследования становятся доступны лечащему врачу и пациенту либо законному представителю в установленном порядке.

Показателями результативности реализации организационных технологий являются количественные критерии (характеризуют объем и структуру разработки и внедрения медицинских изделий на основе ИИ) и качественные критерии (динамика показателей деятельности медицинских организаций, экономические, эпидемиологические и иные показатели). Соответствующие показатели могут изучаться применительно к конкретной медицинской организации, системе здравоохранения субъекта РФ или на общегосударственном уровне.

Предложенный подход и организационные технологии могут применяться в государственной, муниципальной и частной системах здравоохранения, на всех уровнях оказания медицинской помощи.

Организационная технология создания и внедрения технологий искусственного интеллекта в рентгенологии может, в том числе, учитываться и применяться при формировании политики развития рынка информационных технологий, цифрового здравоохранения, актуализации нормативно-правового обеспечения.

Залогом реализуемости идей, лежащих в ее основе, стала совокупность национальных стандартов, разработанных Подкомитетом 01 «Искусственный интеллект в здравоохранении» Технического комитета 164 «Искусственный

интеллект». Стандарты основаны на материалах комплексных научных исследований ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ», в том числе – на материалах диссертации.

Организационная технология реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта реализуется при оказании первичной медико-санитарной помощи, специализированной (в том числе высокотехнологичной), скорой (в том числе специализированной) и паллиативной помощи.

Вместе с тем, эта организационная технология – или точнее отдельные ее элементы – в большей мере предназначены для первичной медико-санитарной помощи, оказываемой в плановом порядке в амбулаторных условиях. К таковым элементам относится, прежде всего, автономная сортировка, представляющая собой автоматизацию производственных процессов массовых профилактических исследований с целью их оптимизации, экономии и перераспределения ресурсов, повышения охвата населения профилактическими мероприятиями.

Возможность реализации данной организационной технологии при оказании экстренной и неотложной помощи определяется сугубо техническими возможностями конкретной информационной системы в сфере здравоохранения и требованиями к скорости обработки результатов лучевых исследований. Такие требования формируются при закупке медицинских изделий на основе ИИ. Оптимально, если требования к скорости обработки устанавливаются централизованно и единообразно на уровне, как минимум, субъекта РФ.

Таким образом, подход к организации внедрения и применения ИИ, а также две организационные технологии для его практической реализации обеспечивают системность клинического целеполагания, разработки, клинической апробации, способов применения и оценки качества технологий искусственного интеллекта в рентгенологии.

Оценка эффективности организационных методологий. Научно обоснованные подход и организационные технологии нашли свое применение в практическом здравоохранении, что позволило оценить медицинскую,

социальную и экономическую эффективность. С методической точки зрения, решение такой задачи представляется достаточно сложным, так как эффекты автоматизации целесообразно изучать в контексте отдельной клинической задачи или в длительном временном промежутке, чтобы можно было выявить значимые колебания основных эпидемиологических и экономических показателей. Тем не менее, путем аналитической работы нами предложены пути решения.

Показателями медицинской эффективности – как степени достижения медицинского результата [88] – установили количество и динамику:

1. Случаев выполнения лучевых исследований, при проведении которых использованы технологии искусственного интеллекта.
2. Случаев, содержащих объективно выявленные при мероприятиях по контролю качества расхождения описаний результатов лучевых исследований. На данный показатель медицинской эффективности, безусловно, влияет значительное количество факторов. Однако, вклад организационной технологии обеспечения безопасности лучевых исследований на основе обязательного предварительного машинного анализа их результатов является очень значительным. Особенно это подчеркивает тот факт, что в изучаемый период никаких иных новых системных мероприятий по улучшению качества работы врачей-рентгенологов не проводилось.

Проанализированы данные о количестве лучевых исследований, результаты которых направлялись на автоматизированный анализ ИИ-сервисами в 2020–2022 гг. (2023 год исключен в связи с принципиальными изменениями правил маршрутизации из-за появления в системе обязательного медицинского медицинской услуги с применением ИИ). Соответствующие данные приведены в таблице 5.1.

Таблица 5.1 – Динамика случаев выполнения лучевых исследований, при проведении которых использованы технологии искусственного интеллекта

Год	Количество исследований, доступных для анализа ИИ	Количество исследований, направленных на ИИ-сервисы		Количество исследований, обработанных ИИ-сервисами	
		Абс.	Отн.,%	Абс.	Отн.,%*
2020	4445638	845020	19,0	1216032	-
2021	6823322	2500033	36,7	2681298	-
2022	7285445	3472429	47,7	3106648	-

* Расчет удельного веса неприменим, так как результаты одного исследования могли быть проанализированы несколькими ИИ-сервисами параллельно.

Отмечается ежегодное увеличение удельного веса случаев выполнения лучевых исследований, при проведении которых использованы технологии искусственного интеллекта: с 19,0 % в 2020 г. до 47,7 % в 2023 г. (то есть на 28,7 %). Такой результат оцениваем, как подтверждающий медицинскую эффективность организационных технологий создания и внедрения технологий искусственного интеллекта в рентгенологии, реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта.

Проанализированы данные о количестве мероприятий по внутреннему контролю качества (пересмотров случайной выборки результатов лучевых исследований уполномоченными врачами-экспертами), количестве выявленных клинически незначимых и клинически значимых расхождений (рис. 5.3).

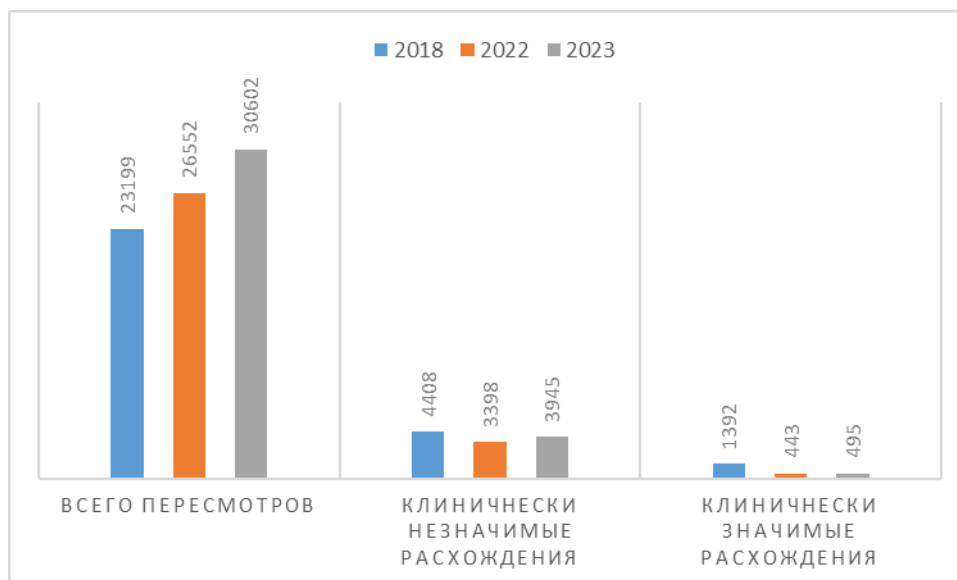


Рисунок 5.3 – Количество и результаты пересмотров случайной выборки результатов лучевых исследований уполномоченными врачами-экспертами, абс.

Первые исследования в области искусственного интеллекта, в той или иной форме, начались в ГБУЗ «НПКЦ ДиТ ДЗМ» в 2019 г., поэтому в качестве базового нами был выбран 2018 год. Количество проверок составило 23 199, из которых в 19,0 % случаев установлены клинически незначимых расхождений, а в 6,0 % – клинически значимые (соответствующие данные по этому году были ранее опубликованы [212]). В 2022 г. количество пересмотров составило 26 552, из которых в 13,0% случаев обнаружены клинически незначимые, а в 1,7 % – клинически значимые расхождения. В 2023 г. наблюдалась сходная картина: по результатам 30 602 пересмотров обнаружены клинически незначимые и значимые расхождения в 13,0 и 1,6 % случаев.

Таким образом, на фоне внедрения организационной технологии обеспечения безопасности лучевых исследований на основе обязательного предварительного машинного анализа их результатов отмечается снижение удельного веса случаев клинически значимых расхождений с 6,0 % до 1,6–1,7 %. Безусловно, выявленный эффект не связан исключительно с автоматизацией, но эмпирически вклад указанной организационной технологии оцениваем, как критично значимый.

Под социальным результатом (мерой социальной эффективности [88])

полагаем минимизацию предубеждений на фоне роста компетенций и осознанного применения технологий ИИ врачами-рентгенологами, то есть фактическую вовлеченность врачей. В рамках Московского Эксперимента маршрутизация результатов лучевых исследований на ИИ-сервисы осуществляется по определенной схеме в обязательном порядке. Независимо от мнения/отношения врача для каждого исследования в ЕРИС ЕМИАС появляется результат машинного анализа. Вместе с тем, существует параллельный механизм. В ЕРИС ЕМИАС реализована функция выборочной отправки: из специального каталога ИИ-сервисов врач может выбрать наиболее предпочтительный и отправить на анализ этим сервисом результат конкретного лучевого исследования. Используемость функции отражает добровольное, осознанное применение технологий ИИ врачом-рентгенологом, то есть позволяет объективно оценить вовлеченность.

Проанализирована динамика инициативной отправки результатов лучевых исследований на анализ ИИ-сервисами. В 2021 г. общее количество таких случаев составило 863, среднее значение в месяц было $172,6 \pm 146,5$. В 2022 г. общее число возросло в 5,5 раз до 4814; среднее составило $401,2 \pm 178,5$. В 2023 г. относительно исходного уровня рост составил 6,6 раза, общее количество инициативных отправок достигло 5719 (среднее – $519,9 \pm 421,1$). Различия средних значений достигали статистической значимости: при сравнении уровней 2021 г. и 2022 г. – t-критерий был $-3,428$, $p = 0,0024$; при сравнении уровней 2021 г. и 2023 г. – t-критерий был $-2,698$, $p = 0,0131$. Различия между 2022 и 2023 гг. были незначимыми. Можно утверждать, что по сравнению с 2021 г. уровень вовлеченности врачей-рентгенологов статистически достоверно ($p = 0,0024$, $p = 0,0131$) возрос в 5,5–6,6 раз в 2022 и 2023 гг. соответственно.

Такой результаты оцениваем, как подтверждающий медицинскую эффективность организационной технологии создания и внедрения технологий искусственного интеллекта в рентгенологии, реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта.

Под экономической эффективностью обычно понимают соотношение

полученных результатов и произведенных затрат [88]. Однако, предлагается расширить это понятие и рассмотреть его с позиции оценки развития экономики. Таковым развитием, в частности, можно считать появление новых отраслей, рынков товаров и услуг. Сказанное весьма показательно в отношении технологий искусственного интеллекта – создание нового рынка соответствующих решений в здравоохранении соответствует целям и задачам Национальной стратегии развития искусственного интеллекта в Российской Федерации до 2030 г. (утв. Указом Президента РФ от 10 октября 2019 г. № 490). Исходя из сказанного показателем экономической эффективности считаем динамику вывода на рынок медицинских изделий на основе технологий искусственного интеллекта.

В 2019 г. в России отсутствовали медицинские изделия с технологиями искусственного интеллекта, применимые в области рентгенологии. Первые две соответствующие разработки была допущены к обращению в соответствии с действующим законодательством в 2020 г. Далее, на фоне Московского Эксперимента и созданных условий для качественного развития продуктов на основе ИИ, в том числе в 2022–2023 гг., еще 14 разработок получили регистрационные удостоверения. В 2023 г. действие одного регистрационного удостоверения было приостановлено. Однако, очень важно подчеркнуть, что данный программный продукт на основе ИИ отличается крайне низким качеством, что полностью подтверждается результатами его работы в Московском Эксперименте. Соответствующие результаты доступны публично в рейтинге ИИ-сервисов (mosmed.ai).

Результат в виде появления нового рынка медицинских изделий считаем подтверждением медицинской эффективности организационной технологии создания и внедрения технологий искусственного интеллекта в рентгенологии.

Таким образом, медицинская эффективность разработанных организационных подходов и технологий состояла в увеличении удельного веса случаев выполнения лучевых исследований, при проведении которых использованы технологии искусственного интеллекта, на 28,7 % за период наблюдения (с 19,0 % в 2020 г. до 47,7 % в 2023 г.), а также в повышении качества

работы врачей-рентгенологов, которое объективно подтверждается снижением удельного веса случаев клинически значимых расхождений с 6,0 % до 1,6–1,7 %, выявляемых при внутреннем контроле качества медицинской помощи. Экономическая эффективность представленных в диссертации разработок подтверждается успешным формированием «с нуля» новой отрасли рынка медицинских изделий. В соответствии с действующим законодательством порядка 16 медицинских изделий на основе искусственного интеллекта для области рентгенологии допущены к обращению. Социальная эффективность проявилась статистически значимым ($p = 0,0024$, $p = 0,0131$) ростом вовлеченности врачей-рентгенологов в 5,5–6,6 раз в 2022 и 2023 гг. по сравнению с 2021 г.

ЗАКЛЮЧЕНИЕ

Современная лучевая диагностика как ключевой и неотъемлемый компонент современного здравоохранения характеризуется постоянным расширением парка цифрового оборудования, стремительным ростом количества проводимых исследований на фоне не менее выраженного роста потребности. Обеспечение доступности и качества лучевых исследований является ключевой проблемой, решению которой посвящены многие научные исследования. Особую сложность создает выраженный кадровый дефицит, так как физическое увеличение количества специалистов не может более соответствовать ни темпам роста парка, ни потребности в лучевых исследованиях. В условиях кадрового дефицита наиболее уязвимым становится профилактическое направление – не обеспечиваются должные охват и качество проводимых исследований (маммографии, флюорографии). Высокий уровень цифровизации лучевой диагностики создал основу для разработки и внедрения новых моделей организации медицинской помощи и широкой автоматизации как путей решения обозначенных проблем.

Технологии искусственного интеллекта находятся в фокусе внимания многочисленных авторов как в России, так и за рубежом. Вместе с тем при системном анализе источников четко выявляются принципиальные недостатки проведенных исследований. По-прежнему малое количество работ являются клиническими, подавляющее число публикаций имеет технический и математический характер.

Систематический обзор 38 метаанализов позволил нам выявить типовые методологические ошибки и недостатки: ретроспективный формат, несбалансированная выборка, отсутствие внешней валидации точности алгоритмов, средний или высокий риск предвзятости, некорректные дизайн и статистическая обработка, часто несоответствующие принципам доказательной медицины. Таким образом, в лучевой диагностике не предпринимались попытки системного внедрения ИИ, отсутствуют данные о точности и надежности

решений на основе искусственного интеллекта в реальных клинических условиях.

Сказанное обусловило цель исследования, состоящую в обеспечении возможности принципиального повышения производительности и качества лучевой диагностики путем научного обоснования подхода к организации внедрения и применения технологий искусственного интеллекта с целью автоматизации и повышения результативности производственных процессов в рентгенологии.

Для достижения цели решен ряд задач, прежде всего – изучена диагностическая точность технологий искусственного интеллекта при работе в условиях реальной клинической практики.

Отличительной чертой этого этапа исследования стал его многоцентровой и проспективный характер.

В период исследования при интерпретации и описании результатов лучевых исследований применялось программное обеспечение на основе технологий искусственного интеллекта (ИИ-сервисы), интегрированное в государственную информационную систему в сфере здравоохранения субъекта РФ – ЕРИС ЕМИАС.

При обращении пациента в МО ДЗМ осуществлялось стандартное обследование, в процессе которого лечащий врач устанавливал наличие показаний для лучевого исследования и формировал направление. Назначенное лучевое исследование выполнялось в установленном порядке, его результаты в формате DICOM автоматически сохранялись в ЕРИС ЕМИАС. Далее эти результаты маршрутизировались на определенный ИИ-сервис (в соответствии с модальностью и видом лучевого исследования), который выполнял автоматический их анализ. Результаты такого анализа возвращались в ЕРИС ЕМИАС и появлялись на автоматизированном рабочем месте врача-рентгенолога параллельно с исходными данными лучевого исследования. Результаты работы ИИ-сервиса представляли собой: дополнительную серию изображений с маркировкой патологических находок, дополнительными данными (например, результатами измерений) и обязательным информационным сообщением «Только

для исследовательских целей»; проект текстового описания результатов интерпретации данного лучевого исследования.

Врач-рентгенолог, проводящий интерпретацию и описание исследования в ЕРИС ЕМИАС, мог открыть дополнительную серию; изучить результаты работы программного обеспечения на основе технологий искусственного интеллекта; по своему усмотрению использовать, не использовать или частично использовать эти результаты при формировании протокола исследования. Исходно обращение к результатам работы ИИ-сервиса было добровольным, зависело от индивидуального профессионального решения врача.

Нами исследованы диагностическая точность и клиническая результативность 32 ИИ-сервисов для автоматизированного анализа результатов рентгенографии, флюорографии, маммографии, компьютерной и магнитно-резонансной томографии.

Все ИИ-сервисы, как было сказано выше, интегрировались в ЕРИС ЕМИАС по определенной процедуре; неотъемлемой ее частью было тестирование на эталонных наборах данных. Это ретроспективное тестирование по своей сути повторяло многочисленные публикации на тему ИИ в лучевой диагностике [175, 235, 239], но принципиально отличалось масштабом. В работах иных авторов обычно фигурируют от 1 до 4 алгоритмов ИИ, в то время как в диссертации их было 32. Другим принципиальным отличием стала унификация данных. Многолетние исследования ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ» позволили научно обосновать методологию создания и разметки эталонных наборов данных, валидированную в том числе независимыми авторами. Эта методология послужила одним из инструментов диссертационной работы. С ее помощью было создано необходимое количество эталонных наборов данных, включавших по 100 результатов исследований в разрезе модальностей и конкретной клинической задачи. Наконец, третьим отличием стал единый дизайн диагностического исследования (в соответствии с методологией STARD) и использование стандартизированного методически набора показателей с корректными способами их расчетов в полном соответствии с классическими

руководствами по биостатистике.

Ретроспективное тестирование на эталонных наборах данных представляло собой внешнюю валидацию ИИ-сервисов на новых данных, условно выражаясь «незнакомых» алгоритмам, то есть на данных, которые совершенно не были доступны разработчикам ранее и, как очевидно, не использовались ни при обучении, ни при внутреннем тестировании ИИ-сервисов. Внешняя валидация – важный этап создания ИИ для медицины, который крайне редко осуществляется на практике; о необходимости такой независимой проверке и о частом ее отсутствии как о методической ошибке сообщают авторы обобщающих исследований качества научных публикаций о ТИИ [192, 214, 224].

Сопоставление результатов ретроспективной проверки на новых данных с заявленными разработчиками показателями точности позволило выявить качество ИИ-сервисов с точки зрения воспроизводимости результатов их работы на новых данных. Однако это действие носило предварительный характер. Если ИИ-сервис преодолевал установленные базовые пороговые значения показателей точности, то он допускался к работе в реальных клинических условиях с потоком результатов лучевых исследований соответствующей модальности и вида.

Проспективное многоцентровое диагностическое исследование включало регулярные измерения точности и клинической результативности по унифицированной методологии.

Средствами описательной статистики нами обобщены показатели диагностической точности для каждой модальности в отдельности.

Среднее значение AUROC оказалось наиболее высоким для компьютерной томографии – $0,885 \pm 0,089$, медиана составила 0,908. Наименьшие значения отмечены для флюорографии: среднее – $0,809 \pm 0,076$, медиана – 0,798. Примечательно, что диапазон минимальных и максимальных значений был наиболее выражен для рентгенографии: от 0,369 до 0,977. Вместе с тем в половине случаев значение AUROC этой модальности попадало в диапазон 0,789–0,908, то есть было приемлемо высоким. Для остальных модальностей отмечено наличие минимальных значений AUROC в диапазоне 0,637–0,663.

Такой уровень говорит о крайне низкой диагностической ценности, приближающейся к «угадыванию». Вместе с тем максимальные значения AUROC составили 0,940–0,995, что говорит о достижении рядом ИИ-сервисов высокой диагностической ценности, приближающейся или даже идентичной врачу-рентгенологу при решении отдельных клинических задач. Медиана AUROC была наименьшей для флюорографии – 0,798, наибольшей для КТ – 0,908. Отметим, что также для КТ половина всех значений AUROC попадала в диапазон 0,823–0,956; его верхняя и нижняя границы были самыми высокими, по сравнению с иными модальностями. Базовый пороговый уровень каждого показателя точности составляет 0,81 [60]. Исходя из этого, выше порогового значения находились 50,0 % значений AUROC только для модальности «компьютерная томография». Среднее значение обобщенного AUROC для всех модальностей составило $0,846 \pm 0,099$, медианы – 0,855.

Среднее значение точности было наибольшим для КТ $0,865 \pm 0,115$ и МРТ $0,861 \pm 0,090$, наименьшим для маммографии – $0,723 \pm 0,136$ и флюорографии – $0,729 \pm 0,115$. Диапазоны значений точности для всех модальностей, за исключением МРТ, были довольно обширными. Минимальные значения колебались на уровне 0,375–0,567, максимальные – на уровне 0,863–0,990. Это подчеркивало разнообразие уровней качества реализации конкретных прикладных решений – ИИ-сервисов и вместе с тем говорило о в целом технически достигнутом уровне точности, сопоставимом или даже превосходящем врача-рентгенолога при решении отдельных клинических задач. Минимальное значение точности МРТ составило 0,721 и было принципиально более высоким, чем у других модальностей. Максимальное достигло 0,954 и также было относительно высоким. Выше базового порогового значения 0,81 находились 50,0 % значений точности только для модальности «компьютерная томография».

Среднее значение обобщенного показателя точности для всех модальностей составило $0,81 \pm 0,122$, медианы – 0,840.

Наибольшее среднее значение чувствительности вновь получено

для модальности «компьютерная томография» – $0,935 \pm 0,064$, наименьшее для маммографии – $0,87 \pm 0,11$.

Обращают на себя внимание максимальные значения этого показателя. У всех модальностей, за исключением рентгенографии, они достигали 1,0. Впрочем и у РГ оно было очень высоким и составляло 0,989. На этом фоне лишь у флюорографии и у МРТ минимальное значение превысило 0,7. У всех иных модальностей оно было ниже (хуже всего у рентгенографии – 0,611). Базовое пороговое значение 0,81 превысили медианы чувствительности всех модальностей. Вместе с тем только у ММГ медиана была ниже 0,9. Напротив, наибольшие значения получены для МРТ – 0,958 и КТ – 0,950.

Также у всех модальностей 50,0 % значений чувствительности находились выше базового порогового уровня, исключение составила только маммография – первый квартиль для этой модальности равнялся 0,792. Среднее значение обобщенной чувствительности для всех модальностей составило $0,907 \pm 0,080$, медиана – 0,926.

Среднее значение специфичности оказалось наибольшим для КТ – $0,810 \pm 0,165$, наименьшим для флюорографии – $0,655 \pm 0,152$. Особого внимания заслуживает факт того, что лишь для КТ и МРТ в пределах 95%-й доверительного интервала был преодолен базовый пороговый уровень. Для остальных модальностей верхняя граница доверительного интервала не превысила 0,78. Вместе с тем максимальные значения специфичности для всех модальностей были высоки – от 0,93 до 1,0. Но при этом катастрофически низкими были минимальные значения – от 0,083 до 0,480. Также лишь для КТ и МРТ медиана специфичности превысила базовый пороговый уровень в 0,81.

Среднее значение обобщенной специфичности для всех модальностей составило $0,746 \pm 0,177$, медианы – 0,775. Таким образом, только для этого показателя не был преодолен (или хотя бы включен в доверительный интервал) базовый пороговый уровень.

Перечисленные средние значения можно считать объективными параметрами уровня развития технологий ИИ для лучевой диагностики на момент

выполнения диссертации.

При совокупном анализе диагностической точности технологий ИИ в проспективном мультицентровом исследовании установлены средние значения обобщенных показателей: AUROC для всех модальностей – $0,846 \pm 0,099$ (медиана – 0,855), точность – $0,81 \pm 0,122$ (медиана – 0,840), чувствительность – $0,907 \pm 0,080$, (медиана – 0,926), специфичность – $0,746 \pm 0,177$ (медиана – 0,775).

На самом первом этапе исследования – информационно-аналитическом – нами выполнен систематический обзор 38 метаанализов, объединивших данные о применении и точности ТИИ в лучевой диагностике. Получены средние значения показателей точности по систематизированным литературным данным: площадь под характеристической кривой – 92,0 % (медиана – 93,8 %), чувствительность – 89,0 %, специфичность – 88,8 % (медиана – 89,5 %).

Ключевым отличием оригинальных результатов, полученных в условиях реальной клинической практики, стал принципиально более низкий уровень специфичности. Действительно, при работе с потоком исследований ИИ-сервисы демонстрируют специфичность ниже, чем при тестировании в «идеальных» условиях ретроспективной проверки на статичных наборах. Независимо от методологии и общего уровня качества каждый набор данных отличается неким уровнем обобщения и упрощения. За счет этого значительно упрощаются задачи автоматизированного анализа. В условиях реальной практики ИИ-сервис сталкивается с колоссальной вариабельностью патологии, анатомических особенностей, возрастных проявлений и т.д. Только на потоке данных появляется возможность объективно охарактеризовать уровень диагностической точности.

Далее изучили динамику показателей точности, в том числе по модальностям, а также сравнили точность ИИ для анализа результатов исследований различных модальностей. Для решения такой комплексной задачи использовано построение модели смешанных эффектов для многократных измерений (зависимая переменная – значение соответствующей диагностической характеристики, факторы – модальность (фиксированный эффект), ИИ-сервис (случайный эффект), квартал оценки (повторяющийся фактор)), а также

взаимодействие модальности с кварталом.

При обобщении анализа динамики и уровней показателей точности посредством указанной модели выраженного улучшения показателей диагностической точности ИИ по модальностям не отмечается. Вместе с тем четко проявляется тенденция «перенастройки» систем с целью повышения чувствительности за счет снижения специфичности: во временной динамике отмечается постепенное повышение уровня чувствительности на 0,013 единиц в квартал ($p = 0,001$) на фоне также достоверного снижения специфичности на 0,023 единиц в квартал ($p = 0,006$).

В разрезе модальностей наблюдалось преобладание точности технологий ИИ для анализа результатов компьютерной томографии. Однако во временной динамике выявлено значимое ухудшение показателя точности на 0,021 единицы в квартал ($p = 0,020$) для этой модальности.

В целом, полученные обобщенные результаты свидетельствовали о значительной вариативности и внушительных диапазонах колебания показателей точности. Поэтому был сделан переход от общего к частному и проведен детальный анализ точности каждого отдельного ИИ-сервиса.

В дизайне проспективного многоцентрового клинического исследования проанализирована диагностическая точность технологий искусственного интеллекта при решении ряда клинических задач, актуальных с точки зрения работы врача-рентгенолога.

Сравнительно изучены показатели AUROC, точности, чувствительности и специфичности, а также результаты оценки клинической результативности для 32 ИИ-сервисов, работавших на потоке исследований от 1 до 8 кварталов. Полученные результаты позволили нам утверждать следующее. Технологии ИИ могут и должны применяться для автоматизации решения клинических задач лучевой диагностики, однако уровни прикладной готовности достаточно сильно варьируются. Достигнутый уровень развития ТИИ позволяет использовать их для повышения производительности и качества работы врача-рентгенолога в виде:

1. Системы поддержки принятия врачебных решений при описании результатов: МРТ головного мозга с целью диагностики и оценки динамики рассеянного склероза; МРТ позвоночника с целью выявления дегенеративно-дистрофических изменений; КТ ОГК/ОБП с целью выявления новообразований надпочечников, компрессионных переломов позвонков и признаков снижения минеральной плотности костной ткани (остеопороза); РГ коленного сустава с целью стадирования деформирующего остеоартроза.

2. Инструмента морфометрии при описании результатов: КТ ОГК/ОБП с целью выявления аневризмы аорты; РГ стопы с целью выявления продольного плоскостопия.

3. Инструмента автономной сортировки при интерпретации результатов: профилактической РГ/ФЛГ; профилактической ММГ.

Выявлен ряд задач, для которых требуется существенное улучшение математических моделей, лежащих в основе соответствующих ИИ-сервисов.

Достаточно проблематично сопоставить полученные результаты с литературными данными. Исчезающее малое количество работ на тему ИИ в лучевой диагностике носит проспективный характер, а при использовании понятия «мультицентральной» обычно подразумевается лишь то, что набор данных для ретроспективного тестирования был сформирован на базе нескольких медицинских организаций. В основном проспективный дизайн встречается в исследованиях, посвященных радиомике и разработке предиктивных моделей (при этом размер выборки крайне редко превышает 1000 наблюдений) [159, 174, 175, 187, 235, 239].

Накопленный объем первичных данных, совершенно недоступный иным авторам, позволил нам перейти от декларативных утверждений к конкретным рекомендациям и методическим подходам.

В целом, можно констатировать принципиальное повышение точности и качества ИИ-сервисов, по сравнению с первым этапом эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе

здравоохранения города Москвы [61]. Действительно состоялся качественный переход, и в распоряжении специалистов лучевой диагностики появился целый ряд новых, достаточно точных и надежных инструментов.

Выявлен феномен снижения точности ИИ-сервисов при начале работы с потоком данных в реальных клинических условиях. В рамках нашего исследования переход ИИ-сервисов на работу с реальными клиническими данными состоялся 38 раз. При этом в 84,0 % (32) случаев зафиксировано снижение площади под характеристической кривой – интегрального показателя диагностической точности. Лишь в 16,0 % (6) случаев ИИ-сервисы продемонстрировали устойчивость или рост AUROC. Этот факт позволяет утверждать, что при разработке ИИ-сервиса его апробация в формате проспективного клинического исследования является обязательной. При этом такая апробация не может быть оформлена как клинические испытания в целях регистрации медицинского изделия. В результате проспективного клинического исследования должны появиться объективные оценки точности и результативности, должны быть предприняты усилия по доработке ИИ-сервиса с повторным, также проспективным, его тестированием. Лишь после этого целесообразно проводить клинические испытания в целях регистрации медицинского изделия уже как отдельное мероприятие.

Еще при планировании исследования нами эмпирически сформированы три основных способа применения ТИИ в контексте интерпретации и описаний результатов лучевых исследований: сортировка, система поддержки принятия врачебных решений, морфометрия.

Проведено исследование возможности масштабирования морфометрии как способа, тесно взаимодействующего с СППРВ, до самостоятельной автономной формы применения. Для оценки возможности автономной морфометрии выполнено экспериментальное исследование, заключающееся в математическом сравнении результатов автоматизированной морфометрии результатов рентгенографии стопы ($n = 100$) с целью выявления продольного плоскостопия и соответствующей разметки врачами-рентгенологами.

Использованы результаты работы ИИ-сервиса и 4 врачей-рентгенологов со стажем работы от 5 до 10 лет. Набор данных включал деперсонализированные результаты 100 рентгенографий, выполненных 100 различным пациентам, случайным образом отобранные в ЕРИС ЕМИАС. Для снижения субъективного фактора при проведении разметки врачами-рентгенологами каждое уникальное изображение повторялось 4 раза и перемешивалось случайным образом с остальными изображениями, т. е. врачи проводили пересмотр 400 изображений, из которых 100 было уникальными. Те же 100 уникальных изображений были направлены на анализ ИИ-сервису для морфометрии. Согласованность измерений, выполненных врачами-рентгенологами, зафиксирована в 82,6 % случаев, между врачами и морфометрическим ИИ-сервисом – в 68,8 % случаев. Однако минимальное среднеквадратическое отклонение между мнениями экспертов составило 0,60, максимальное отклонение – 1,76. Минимальное среднеквадратичное отклонение между экспертами и ИИ-сервисом составило 0,44, а максимальное – 1,33.

Таким образом средний уровень расхождения между результатами измерений врачами-рентгенологами и морфометрическим ИИ-сервисом ниже, чем средний уровень расхождения результатов измерений между врачами. По результатам проведенной оценки средних значений и среднеквадратических отклонений можно сделать вывод, что результаты измерений угла свода правой стопы, полученные автоматизировано, не отличаются от результатов врачей-рентгенологов. Потенциально ТИИ способны работать полностью автономно, проводя необходимые измерения (морфометрию) и создавая заключение о наличии и степени плоскостопия. Важным преимуществом использования данного подхода является объективизация измерений, поскольку, как было показано в исследовании, результат повторного измерения врача отличается от первоначального, тогда как результат измерения сервиса всегда одинаков и не зависит от количества повторений.

Полученные в ходе проспективного многоцентрового клинического исследования результаты позволили нам качественно пересмотреть трактовку

«сортировки» как одного из способов применения технологий ИИ при интерпретации и описаний результатов лучевых исследований.

Первоначально под сортировкой полагали выявление результатов исследований, содержащих патологические изменения, и их маркировку в рабочем списке врача-рентгенолога в программном обеспечении автоматизированного рабочего места.

Однако в ходе исследования установили, что наибольшее количество прикладных решений разработано для автоматизированного анализа результатов РГ/ФЛГ и профилактической ММГ. Выявлена достоверная их особенность: достижение стабильно высокого уровня чувствительности на фоне средней специфичности. Такое сочетание потенциально устраняет возможность ложноотрицательного результата, то есть ТИИ гарантированно не пропустит результаты исследования с патологическими изменениями. Вместе с тем точность дальнейшей дифференциации таких изменений остается средней. Исходя из этого был предложен переход от прежнего понимания «сортировки» к идее автономного ИИ при проведении профилактических исследований. Суть состоит в следующем: технологии ИИ с максимальной чувствительностью (1,0 (95%-й ДИ 1,0–1,0)) должны применяться для автономной сортировки результатов массовых профилактических исследований (РГ/ФЛГ и ММГ) на две категории: «норма» и «не норма». Во вторую категорию отнесли все случаи любых отклонений (патологические состояния, последствия перенесенных заболеваний и операций, возрастные и врожденные особенности и т.д.). При этом с учетом максимальной чувствительности ТИИ исследования категории «норма» не требуют описания врачом-рентгенологом, а сразу могут быть сохранены в карте пациента для дальнейшего использования врачом клинической специальности. Потенциально такой подход на десятки процентов уменьшит нагрузку на кадровый ресурс лучевой диагностики, позволит оптимизировать финансовую составляющую. Важным преимуществом подхода является возможность нарастить охват населения профилактическими мероприятиями без создания дополнительной нагрузки на ресурсы системы здравоохранения.

Отметим, что модель автономной сортировки была обоснована ранее в лабораторных условиях на эталонных наборах данных [18, 19]. В диссертации нами сделан следующий, принципиально более важный, шаг. Гипотеза о возможности автономной сортировке результатов профилактических лучевых исследований исследована в дизайне проспективного одноцентрового клинического исследования.

В исследование были включены по 3 ИИ-сервиса для РГ/ФЛГ и ММГ соответственно. Включенные ИИ-сервисы были настроены на чувствительность 1,0 (95%-й ДИ 1,0–1,0); значение специфичности не имело значения. В период исследования ИИ-сервисы осуществляли сортировку результатов лучевых исследований на две категории: «норма» и «не норма».

С применением инструмента для автоматизированного анализа текста протоколов проводилось сопоставление решений о категорировании ИИ-сервисов и врачей-рентгенологов. Клинические случаи расхождений направлялись на экспертный пересмотр квалифицированным врачом-рентгенологом с субспециализацией. По результатам пересмотра определялась правильность категорирования результатов данного исследования. Таким образом, референс-тестом №1 служил протокол, оформленный врачом-рентгенологом, а референс-тестом №2 – экспертный пересмотр квалифицированным врачом-рентгенологом с субспециализацией.

В проспективном формате за период исследования ИИ-сервисами проанализированы результаты 49 221 профилактических маммографий, выполненных в амбулаторных условиях в медицинских организациях Департамента здравоохранения г. Москвы.

К категории «норма» при автоматизированном анализе было отнесено 7 028 ММГ, что составило 14,3 %. К категории «не норма» – 42 193, или 85,7 % ММГ.

Полное согласие ИИ и врача-рентгенолога (референс-тест № 1) зафиксировано в 99,0 % (6 960) случаев, расхождения выявлены в 1,0 % (69) случаев (коэффициент каппа Коэна составил 0,994 (95%-й ДИ 0,992–0,996)).

Вторым этапом проведен пересмотр врачом-экспертом с субспециализацией

по маммографии (референс-тест № 2) результатов исследований ($n = 69$), в которых выявлены расхождения решения ИИ и врача-рентгенолога, проводившего первичное описание.

Решение ИИ «норма» было подтверждено в 41,2% (29) случаев, решение врача «не норма» – в 58,8 % (40). Таким образом, ИИ допустил неправильную сортировку в 40 случаях, из которых врачом-экспертом 80,0 % (32) были отнесены к категории BI-RADS-3, а лишь 17,5 % (7) – к категории BI-RADS-4 и 2,5 % (1) – к категории BI-RADS-5.

На этом этапе также согласованность всех решений была очень высока. Для решений врачей-рентгенологов, проводивших первичные описания результатов профилактической ММГ, и врачей-экспертов коэффициент каппа Коэна составил 0,99564 (95%-й ДИ 0,99397–0,99732), а для технологий ИИ и врачей-экспертов – 0,99649 (95%-й ДИ 0,99499–0,99799).

На 49 221 случаев проанализированный результат профилактической маммографии удельный вес ошибочной автономной сортировки с применением технологий ИИ составил лишь 0,08 % (40). Клинически значимое расхождение ИИ-сервисы допустили лишь в 8 случаях, что составило лишь 0,02%. В 99,92 % (49 181) случаев автономная сортировка была выполнена полностью корректно.

По литературным данным, удельный вес клинически значимых дефектов в работе врача-рентгенолога при интерпретации результатов профилактической ММГ составляет от 3,5 % до 30,0 % [148, 183, 169, 173, 206, 217]. По данным ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ», удельный вес дефектов в работе врача-рентгенолога при интерпретации результатов профилактической маммографии составляет 3,0–7,2 % [5, 146, 241].

Выявленный нами удельный вес дефектов в работе ИИ-сервисов, имеющих статус медицинского изделия и работающих на потоке данных в реальных клинических условиях, при автономной сортировке составляет 0,08 %. Удельный вес клинически значимых дефектов не превышает 0,02 %.

В проспективном формате за период исследования ИИ-сервисами проанализированы результаты 209 497 профилактических исследований органов

грудной клетки (рентгенографий и флюорографий), выполненных в амбулаторных условиях в медицинских организациях Департамента здравоохранения г. Москвы.

К категории «норма» при автоматизированном анализе было отнесено 117 041 РГ/ФЛГ, что составило 55,9 %; к категории «не норма» – 92 456, что составило 44,1 %.

Полное согласие ИИ и врача-рентгенолога (референс-тест №1) зафиксировано в 99,7 % (116 667) случаев, расхождения выявлены в 0,3 % (374) случаев (коэффициент каппа Коэна составил 0,988 (95%-й ДИ 0,987–0,990)).

На втором этапе проведен пересмотр врачом-экспертом с субспециализацией по торакальной радиологии (референс-тест №2) результатов исследований ($n = 374$), в которых выявлены расхождения решения ИИ и врача-рентгенолога, проводившего первичное описание.

Решение ИИ «норма» было подтверждено в 56,7 % (212) случаях, решение врача «не норма» – в 43,3 % (162). Таким образом, ИИ допустил неправильную сортировку в 162 случаях, из которых врачом-экспертом в 29,0 % (47) случаев определена клинически значимая патология, а в 71,0 % (115) – клинически незначимые отклонения от нормы. К последним относились кальцинаты, консолидированные переломы и иные состояния, непредусмотренные первоначальной клинической задачей (например, неосложненная диафрагмальная грыжа, фиброз, пневмосклероз, консолидированный перелом ребра и т.д.). В качестве клинически значимых расхождений выявлены: очаги – 76,6% (36), инфильтрация/консолидация – 8,5% (4), образование – 6,4% (3), диссеминация – 4,3% (2), пневмоторакс или расширение средостения – по 2,1% (по 1 случаю).

На этом этапе также согласованность всех решений была очень высока. Для решений врачей-рентгенологов, проводивших первичные описания результатов профилактической РГ/ФЛГ, и врачей-экспертов коэффициент каппа Коэна составил 0,99150 (95%-й ДИ 0,99055–0,99245), а для технологий ИИ и врачей-экспертов – 0,99680 (95%-й ДИ 0,99622–0,99739).

Из 209 497 проанализированных результатов профилактических лучевых

исследований органов грудной клетки удельный вес ошибочной автономной сортировки с применением технологий ИИ составил лишь 0,08 % (162), при чем клинически значимые расхождения выявлены только в 0,02 % (47) случаев. В 99,92 % (209 335) случаев автономная сортировка была выполнена полностью корректно.

Исходя из литературных данных и научно обоснованных средних показателей удельный вес дефектов в работе врача-рентгенолога при интерпретации результатов профилактической РГ/ФЛГ составляет 3,0–4,0 % [6, 146, 194, 195, 241]. Выявленный нами удельный вес дефектов в работе ИИ-сервисов, имеющих статус медицинского изделия и работающих на потоке данных в реальных клинических условиях, при автономной сортировке составляет 0,08 %.

Таким образом, в результате проспективного клинического исследования установлено, что согласованность решений ИИ-сервисов (медицинских изделий) и врачей-рентгенологов (как линейных сотрудников, так и экспертов) высока – капша Коэна $> 0,99$. Удельный вес дефектов, то есть ложных отнесений результатов исследований к категории «норма», составляет 0,08 % для всех проанализированных модальностей. Более того, удельный вес клинически значимых отнесений составляет всего лишь 0,02 %. Указанный уровень дефектов значительно ниже, чем у врача-рентгенолога, исходя из научно обоснованных усредненных показателей диагностической точности. На реальных клинических данных доказаны безопасность и целесообразность автономной сортировки результатов профилактических лучевых исследований с применением технологий искусственного интеллекта.

Исходя из проспективного мультицентрового анализа 32 ИИ-сервисов лидирующим по числу видов исследований является способ применения ТИИ в виде СППВР.

В научной литературе часто утверждается необходимость автоматизированной поддержки принятия решений врачом-рентгенологом [69, 78,

154, 160, 178, 179, 186, 191, 244], однако реальная значимость такой поддержки не обсуждается и не изучается.

В контексте автоматизации лучевой диагностики полагаем, что система поддержки принятия врачебных решений должна потенциально снижать частоту дефектов в работе врача-рентгенолога – прежде всего пропусков, а также гипердиагностики. В этом и заключается основная клиническая значимость таких решений. Для доказательства этого утверждения пошли от обратного. Изучена клиническая целесообразность систем поддержки принятия врачебных решений в дизайне обсервационного исследования. Из ЕРИС ЕМИАС отобраны результаты компьютерной томографии, первично описанные врачами-рентгенологами, проанализированные ТИИ и прошедшие пересмотр врачами-экспертами в рамках мероприятий по контролю качества. Включенные исследования проанализированы на основе следующего положения. Патологический признак был выявлен ИИ-сервисом непосредственно после выполнения исследования, а затем – врачом-экспертом в рамках планового контроля качества, но пропущен врачом-рентгенологом, выполнявшим первичное описание. Соответственно, такой признак послужил причиной клинически значимого расхождения при интерпретации результатов данного лучевого исследования.

В соответствии с критериями включения, выявлено 60 исследований с клинически значимыми расхождениями и релевантными результатами работы технологий ИИ, что составляет 10,2 % от числа случаев КТ органов грудной клетки с таким типом дефектов.

Определена следующая структура потенциально предотвратимых дефектов. В большинстве случаев зафиксирован пропуск признаков новообразований (как типичных, так и требующих дифференциальной диагностики именно онкологического характера) – 49,2 % (32), также значительный удельный вес составила гиподиагностика неспецифических воспалительных изменений в легких – 24,6 % (16). Принципиально ниже было количество пропусков признаков туберкулеза или иного специфического воспалительного процесса – 9,2

% (6) и патологии магистральных сосудов (аневризмы аорты, расширения легочного ствола) – 7,7 % (5). В одинаковом объеме обнаружены пропуски плеврального выпота и иной патологии (значимый коронарный кальциноз, компрессионный перелом позвонка) – по 4,6 % (3) каждая категория.

Все указанные патологические находки были верно обнаружены, идентифицированы и маркированы ИИ-сервисами. Игнорирование врачом-рентгенологом результатов работы искусственного интеллекта привело к формированию ложноотрицательного решения и по прошествии времени – к клинически значимому расхождению. Не было задачи проследить судьбу пациентов. Сам факт выявления потенциально предотвратимых дефектов при описании результатов лучевых исследований (особенно связанных с онкологической диагностикой) служит убедительным доказательством необходимости применения врачом-рентгенологом ИИ-сервисов как систем поддержки принятия врачебных решений.

Согласно литературным данным, основное внимание исследователей фокусируется на проблеме поддержки решений при назначении лучевых исследований. Для этого используются различные подходы и инструменты методического, справочного, организационного характера [139, 158, 205, 231]. Также часто изучается роль врачей-рентгенологов в совместном принятии решений в условиях консилиума (чаще онкологического) и в сложных клинических ситуациях [161, 164, 245]. В диссертации внимание было сконцентрировано на поддержке решений врача-рентгенолога при интерпретации и описании результатов лучевых исследований, направленной на профилактику дефектов и клинически значимых пропусков. Фактически это первое целевое обоснование необходимости СППВР на основе ТИИ для рентгенологии.

При этом результаты диссертационной работы носят проспективную направленность, то есть говорим о профилактике дефектов, в то время как опубликованные исследования посвящены лишь совершенствованию ретроспективного пересмотра. Например, ТИИ использовали для выявления

исследований с потенциально пропущенной патологией. При таком автоматизированном, подчеркнем – ретроспективном – аудите 2 573 результатов КТ-ангиографии выявлено 0,5 % исследований с гиподиагностикой, подтвержденной при пересмотре (коэффициент согласованности между ТИИ и врачами-экспертами составил 0,91) [236].

Таким образом, СППВР на основе ТИИ позволяют предотвратить клинически значимые дефекты (пропуск патологии) при интерпретации и описании результатов лучевых исследований. В структуре потенциально предотвратимых дефектов лидируют признаки новообразований (как типичных, так и требующих дифференциальной диагностики именно онкологического характера) – 49,2 % (32), неспецифических воспалительных изменений в легких – 24,6 % (16).

Продемонстрированную профилактику гиподиагностики (в том числе в 7,7 % за счет базовой автоматической морфометрии магистральных сосудов) следует считать основным доказательством клинической значимости СППВР на основе ИИ для врачей-рентгенологов.

В аспекте совершенствования методологий и нормативно-правового обеспечения в области контроля безопасности и качества медицинской помощи факт применения СППВР на основе ТИИ врачом-рентгенологом может рассматриваться как критерий качества оказания медицинской помощи, во всяком случае применительно к определенным нозологиям или клиническим ситуациям.

Результаты исследования обобщены в виде подхода к организации внедрения и применения технологий искусственного интеллекта с целью автоматизации и повышения результативности производственных процессов в рентгенологии. В целом, он призван обеспечить системную, многоуровневую и эффективную автоматизацию этапов производственных процессов медицинских организаций, связанных с выполнением, интерпретацией и описанием результатов лучевых исследований.

Многоуровневость подхода заключается в следующем. Первым уровнем решения проблемы применения ИИ в рентгенологии считаем адекватное

клиническое целеполагание и стандартизированные подходы к разработке. Вторым уровнем – обязательную апробацию в реальных клинических условиях с накоплением информации для доработки ИИ-сервиса, его максимальной адаптации для решения конкретной задачи в клиническом контексте. Третьим уровнем – унифицированные способы применения ИИ, для которых научно определен клинический контекст. Под последним понимаем совокупность модальностей лучевых исследований, групп нозологий, анатомические области и структуры. Например, для СППВР доказана возможность и целесообразность применения при интерпретации результатов МРТ, выполненной с целью диагностики демиелинизирующего заболевания – рассеянного склероза; для морфометрии – при интерпретации результатов КТ, для измерения диаметра отделов аорты с целью выявления аневризмы; для автономной сортировки – категорирования результатов профилактической маммографии и т.д. Четвертый уровень представляет собой «встраивание» унифицированных способов применения ИИ в стандартный процесс выполнения, интерпретации и описания результатов лучевого исследования.

Для реализации подхода на практике разработаны две организационные технологии: а) создания и внедрения технологий искусственного интеллекта в рентгенологии, б) реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта.

Первая технология обеспечивает реализацию I–III уровней подхода, она основана на унификации и алгоритмизации процессов разработки, обеспечения качества, а также способов применения технологий ИИ. Вторая – обеспечивает реализацию IV уровня подхода, она основана на унификации и алгоритмизации процессов проведения, интерпретации и описания результатов лучевых исследований с применением технологий ИИ.

В целом, подход и организационные технологии могут применяться в государственной, муниципальной и частной системах здравоохранения, на всех уровнях оказания медицинской помощи.

Организационная технология создания и внедрения технологий

искусственного интеллекта в рентгенологии может быть задействована, в том числе, при формировании политики развития рынка информационных технологий, цифрового здравоохранения, актуализации нормативно-правового обеспечения. Залогом реализуемости этого утверждения стала совокупность национальных стандартов, разработанных Подкомитетом 01 «Искусственный интеллект в здравоохранении» Технического комитета 164 «Искусственный интеллект». Стандарты основаны на материалах комплексных научных исследований ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ», в том числе – на материалах диссертации.

Организационная технология реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта, в целом, может быть реализована при оказании первичной медико-санитарной помощи, специализированной (в том числе высокотехнологичной), скорой (в том числе специализированной) и паллиативной помощи. Вместе с тем отдельные ее элементы в большей мере предназначены для первичной медико-санитарной помощи, оказываемой в плановом порядке в амбулаторных условиях. К таковым относится, прежде всего, автономная сортировка, представляющая собой автоматизацию производственных процессов массовых профилактических исследований с целью их оптимизации, экономии и перераспределения ресурсов, повышения охвата населения профилактическими мероприятиями.

Проведена оценка медицинской, социальной и экономической эффективности разработанных организационных методологий. Медицинская эффективность состояла в увеличении удельного веса случаев выполнения лучевых исследований, при проведении которых использованы технологии искусственного интеллекта, на 28,7% за период наблюдения (с 19,0 % в 2020 г. до 47,7 % в 2023 г.), а также в повышении качества работы врачей-рентгенологов, которое объективно подтверждается снижением удельного веса случаев клинически значимых расхождений с 6,0 % до 1,6–1,7 %, выявляемых при внутреннем контроле качества медицинской помощи. Экономическая эффективность подтвердилась успешным формированием «с нуля» новой отрасли

рынка медицинских изделий. В соответствии с действующим законодательством порядка 16 медицинских изделий на основе искусственного интеллекта для области рентгенологии допущены к обращению. Социальная эффективность проявилась статистически значимым ($p = 0,0024$, $p = 0,0131$) ростом вовлеченности врачей-рентгенологов в 5,5–6,6 раз в 2022 и 2023 гг. по сравнению с 2021 г. Мерой вовлеченности считали количество инициативных отправок врачами-рентгенологами результатов лучевых исследований на анализ ИИ-сервисами.

Таким образом, научно обоснованный подход и организационные технологии для его реализации обеспечивают системность клинического целеполагания, разработки, клинической апробации, способов применения и оценки качества технологий искусственного интеллекта в рентгенологии.

ВЫВОДЫ

В диссертации решена актуальная научная проблема обеспечения возможности принципиального повышения производительности и качества лучевой диагностики путем научного обоснования подхода к организации внедрения и применения технологий искусственного интеллекта с целью автоматизации и повышения результативности производственных процессов в рентгенологии.

1. В условиях реальной клинической практики средние значения точности технологий ИИ при анализе результатов лучевых исследований ($n = 6\,563\,753$) составляют: AUROC для всех модальностей – $0,846 \pm 0,099$, (медиана – $0,855$), точность – $0,81 \pm 0,122$ (медиана – $0,840$), чувствительность – $0,907 \pm 0,080$ (медиана – $0,926$), специфичность – $0,746 \pm 0,177$ (медиана – $0,775$). Во временной динамике выявлено постепенное повышение уровня чувствительности на $0,013$ единиц в квартал ($p = 0,001$) на фоне также достоверного снижения специфичности на $0,023$ единиц в квартал ($p = 0,006$). В разрезе модальностей выявлено преобладание точности ИИ для анализа результатов компьютерной томографии. В результате анализа точности отдельных ИИ-сервисов в контексте решения конкретной клинической задачи отмечается значительная вариативность диапазонов и динамики показателей точности, свидетельствующая о разной степени технической и методической готовности ИИ для практического применения.

2. При переходе ИИ-сервисов на работу с реальными клиническими данными в $84,0\%$ случаев происходит снижение площади под характеристической кривой (различия, как правило, достигают статистической значимости).

3. Путем проспективного анализа точности 32 ИИ-сервисов для анализа результатов исследований по модальностям и в контексте решения конкретной клинической задачи обоснованы основные способы автоматизированного анализа результатов лучевых исследований: поддержка принятия врачебных решений,

морфометрия (в том числе автономная), автономная сортировка. Доказана целесообразность применения отдельных способов, в частности: СППВР – для описания результатов компьютерной томографии, рентгенографии и магнитно-резонансной томографии с целью диагностики онкологических, демиелинизирующих, дегенеративно-дистрофических заболеваний, ряда острых состояний; инструмента морфометрии – для описания результатов компьютерной томографии и рентгенографии для выполнения измерений диаметров магистральных сосудов, объемов кровоизлияния или жидкости в полости, а также для выявления нарушений со стороны опорно-двигательной системы; автономной сортировки – для интерпретации результатов профилактической рентгенографии (флюорографии) органов грудной клетки и профилактической маммографии.

4. Клиническая значимость поддержки принятия решений врачом-рентгенологом с применением искусственного интеллекта состоит в предотвращении клинически значимых дефектов (пропусков патологии) при интерпретации и описании результатов лучевых исследований. В структуре потенциально предотвратимых дефектов лидируют признаки новообразований (49,2 %) и неспецифических воспалительных изменений (24,6 %). Разработана организационная технология обеспечения безопасности лучевых исследований на основе обязательного предварительного машинного анализа их результатов, позволяющая реализовать профилактику дефектов в работе врача-рентгенолога.

5. В ходе проведения экспериментального исследования установлено, что результаты автоматизированных измерений угла свода стопы значимо не отличаются от результатов измерений врачами-рентгенологами, тем самым доказана возможность автономной морфометрии результатов рентгенографии стопы с целью выявления продольного плоскостопия.

6. По итогам проспективного клинического исследования ($n = 258\ 718$) обоснована возможность автономной сортировки результатов массовых профилактических исследований. Согласованность решений ИИ-сервисов (медицинских изделий) и врачей-рентгенологов (проводящих первичное описание, экспертов) при сортировке высока – каппа Коэна $> 0,99$. Удельный вес

дефектов (ложных отнесений результатов исследований к категории «норма») в работе медицинских изделий на основе технологий ИИ составляет 0,08 %, а клинически значимых дефектов – 0,02 % как для маммографии, так и для рентгенографии/флюорографии; по этому показателю ИИ превосходит врача-рентгенолога.

7. Научно обоснованный подход к организации внедрения и применения искусственного интеллекта в рентгенологии, а также разработанные для его практической реализации организационные технологии позволяют получить медицинский, социальный и экономический эффект. Медицинская эффективность состоит в увеличении на 28,7 % случаев выполнения рентгенологических исследований, при проведении которых использованы технологии искусственного интеллекта, в снижении с 6,0 % до 1,6–1,7 % выявляемых при внутреннем контроле качества медицинской помощи случаев клинически значимых расхождений; экономическая эффективность состоит в формировании новой отрасли рынка (к обращению впервые допущено 16 медицинских изделий на основе искусственного интеллекта для области рентгенологии); социальная эффективность состоит в статистически значимом ($p = 0,0024$, $p = 0,0131$) росте вовлеченности врачей-рентгенологов в 5,5–6,6 раз в 2022 и 2023 гг. по сравнению с 2021 г.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

Органам исполнительной власти и регуляторам в сфере здравоохранения:

1. Принять во внимание организационные технологии создания и внедрения технологий искусственного интеллекта в рентгенологии, реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта.

2. Разработать и реализовать мероприятия по актуализации нормативно-правового обеспечения в сфере охраны здоровья граждан для обеспечения возможности практического внедрения автономной сортировки результатов массовых профилактических исследований, прежде всего – профилактической рентгенографии/флюорографии органов грудной клетки и профилактической маммографии.

3. При разработке требований к медицинским изделиям на основе ИИ:

– учитывать необходимость проспективного клинического исследования как обязательного этапа разработки;

– учитывать средние значения показателей диагностической точности ИИ-сервисов, полученные в условиях реальной клинической практики: AUROC – $0,846 \pm 0,099$ (медиана – 0,855), точность – $0,81 \pm 0,122$ (медиана – 0,840), чувствительность – $0,907 \pm 0,080$ (медиана – 0,926), специфичность – $0,746 \pm 0,177$ (медиана – 0,775).

4. При актуализации нормативно-правового обеспечения в сфере безопасности и качества медицинской помощи рассматривать факт применения врачом-рентгенологом системы поддержки принятия врачебных решений на основе ИИ как один из критериев качества медицинской помощи.

Медицинским организациям:

1. При внедрении медицинских изделий на основе ИИ:

– руководствоваться организационной технологией реализации медицинских услуг с применением рентгенологических методов и искусственного интеллекта;

– учитывать риск значимого падения точности ИИ при начале работы с потоком данных конкретной медицинской организации (в таких случаях необходимо взаимодействие с разработчиком для усовершенствования, перенастройки ИИ-сервиса).

2. Применять системы поддержки принятия врачебных решений на основе ИИ для описания результатов компьютерной томографии, рентгенографии и магнитно-резонансной томографии с целью диагностики онкологических (ЗНО молочной железы, надпочечников и т.д.), демиелинизирующих (рассеянный склероз и т.д.), дегенеративно-дистрофических заболеваний (деформирующий остеоартроз и т.д.), острых состояний (острого нарушения мозгового кровообращения, гидроторакса).

3. Применять инструменты морфометрии на основе ИИ для описания результатов компьютерной томографии и рентгенографии для выполнения измерений диаметров магистральных сосудов (аневризмы аорты и т.д.), объемов кровоизлияния (внутричерепная гематома и т.д.) или жидкости в полости (гидроторакс и т.д.), а также для выявления нарушений со стороны опорно-двигательной системы (продольного плоскостопия, остеопороза и т.д.).

4. Применять организационную технологию обеспечения безопасности лучевых исследований на основе обязательного предварительного машинного анализа их результатов с целью профилактики дефектов в работе врача-рентгенолога.

Разработчикам технологий искусственного интеллекта:

1. Руководствоваться организационной технологией создания и внедрения технологий искусственного интеллекта в рентгенологии.

2. Проводить проспективные клинические исследования ИИ-сервисов в качестве заключительного этапа разработки, предваряющего последующие действия по регистрации продукта как медицинского изделия.

СПИСОК СОКРАЩЕНИЙ

AUROC	– от англ. Area Under Receiver Operating Characteristic, площадь под характеристической кривой
ДИ	– доверительный интервал
ГБУЗ г. Москвы «НПКЦ ДиТ ДЗМ»	– государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения Москвы»
ЗНО	– злокачественное новообразование
ИИ	– искусственный интеллект
КТ	– компьютерная томография
ММГ	– маммография
МРТ	– магнитно-резонансная томография
ОБП	– органы брюшной полости
ОГК	– органы грудной клетки
ОНМК	– острое нарушение мозгового кровообращения
РГ	– рентгенография
РФ	– Российская Федерация
СПП(В)Р	– система поддержки принятия (врачебных) решений
ТИИ	– технологии искусственного интеллекта
ФЛГ	– флюорография

СПИСОК ЛИТЕРАТУРЫ

1. Алиев З. Ш., Гаврилов Г. В., Свистов Д. В. Применение систем искусственного интеллекта в диагностике идиопатической нормотензивной гидроцефалии // Российский нейрохирургический журнал имени профессора А. Л. Поленова. – 2022. – Т. 14, № S1. – С. 111–112.

2. Алифов Д. Г., Звезда С. А., Кельн А. А. [и др.]. Лучевая диагностика рака простаты на основе искусственного интеллекта и радиомного машинного обучения // Университетская медицина Урала. – 2021. – Т. 7, № 4 (27). – С. 48–50.

3. Аль-Хайдари А. А. М. Использование искусственного интеллекта в диагностике рака костей // Современная наука: актуальные вопросы, достижения и инновации: сборник статей XXVII Международной научно-практической конференции. – В 2 ч. – Пенза, 2022. – С. 38–41.

4. Андреев Д. В., Горьковой А. В., Дзагаштокова А. В. [и др.]. Единая диагностическая информационная система: использование результатов эндоскопии при интерпретации изображений компьютерной томографии шеи и краниофациальной области // Digital Diagnostics. – 2021. – Т. 2. – № S2. – С. 6–7.

5. Арзамасов К. М., Васильев Ю. А., Владзимирский А. В. [и др.]. Применение компьютерного зрения для профилактических исследований на примере маммографии // Профилактическая медицина. – 2023. – Т. 26. – № 6. – С. 117–123.

6. Арзамасов К. М., Семенов С. С., Кокина Д. Ю. [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.

7. Баранов И. А., Титова Л. А., Толстых Е. М. [и др.]. Перспективы внедрения программы искусственного интеллекта в процесс диагностики заболеваний легких // Системный анализ и управление в биомедицинских системах. – 2020. – Т. 19, № 3. – С. 165–170.

8. Барчук А. А., Подольский М. Д., Беляев А. М. [и др.]. Автоматизированная диагностика в популяционном скрининге рака легкого // Вопросы онкологии. – 2017. – Т. 63, № 2. – С. 215–220.
9. Бастанов А. Э., Зубаиров И. З., Ахметов И. В. Детекция признаков и диагностика ревматоидного артрита с использованием рентгенологического метода исследования и искусственного интеллекта // Перспективы развития экономики здоровья: сборник докладов II Всероссийской научно-практической конференции. – Уфа, 2020. – С. 38–41.
10. Белан И. Ю., Заборовский В. С., Лукашин А. А. Разработка системы инференса данных для платформы интеллектуальной диагностики рака легких // Математические методы в технике и технологиях – ММТТ. – 2020. – Т. 10. – С. 107–110.
11. Берген Т. А., Пухальский А. Н., Сеницын В. Е. [и др.]. Новые возможности в организации проведения лучевых исследований у онкологических пациентов // Вестник Росздравнадзора. – 2022. – № 6. – С. 49–56.
12. 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.
13. Богородская Е. М., Слогоцкая Л. В., Туктарова Л. М. [и др.]. Скрининг туберкулезной инфекции в группах риска у взрослого населения города Москвы // Туберкулез и болезни легких. – 2023. – Т. 101, № 4. – С. 13–21.
14. Борзов С. М., Карпов А. В., Потатуркин О. И. [и др.]. Применение нейронных сетей для дифференциальной диагностики легочных патологий по рентгенологическим изображениям // Автометрия. – 2022. – Т. 58, № 3. – С. 61–71.
15. Борисенко О. В., Коновалов В. К., Лазарев А. Ф. [и др.]. Возможности дифференциальной диагностики гистологических форм первичного рака легкого при мультиспиральной компьютерной томографии на основе искусственного интеллекта // Российский онкологический журнал. – 2019. – Т. 24, № 3–6. – С. 102–104.

16. Борисенко О. В., Коновалов В. К., Лазарев А. Ф. [и др.]. Дифференциальная диагностика первичного рака легкого при МСКТ-исследованиях на основе искусственного интеллекта // Евразийский онкологический журнал. – 2022. – Т. 10, № S2. – С. 989–990.

17. Бузаев И. В., Плечев В. В., Галимова Р. М. [и др.]. Развитие технологий искусственного интеллекта в онкологии и лучевой диагностике // Креативная хирургия и онкология. – 2018. – Т. 8, № 3. – С. 208–215.

18. Васильев Ю. А., Тыров И. А., Арзамасов К. М. [и др.]. Новая модель организации массовых профилактических исследований, основанная на автономном искусственном интеллекте для сортировки результатов флюорографии // Здоровье населения и среда обитания. – 2023. – № 11. – С. 23–32.

19. Васильев Ю. А., Тыров И. А., Владимировский А. В. [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.

20. Введение в медицинскую статистику с основами эпидемиологического анализа / под ред. Н. Д. Ющука, Н. Б. Найговзиной. – М. : ГЭОТАР-Медиа, 2021. – 192 с.

21. Ветшева Н. Н., Трофименко И. А., Морозов С. П. [и др.]. Повышение качества медицинской помощи за счет усовершенствования системы непрерывного медицинского образования // Медицинское образование и профессиональное развитие. – 2017. – № 2–3 (28–29). – С. 60–68.

22. Волосова А. В. Применение методов искусственного интеллекта для визуализации МРТ-изображений и ранней диагностики болезни Альцгеймера // Аспирант и соискатель. – 2019. – № 2 (110). – С. 124–126.

23. Выгоняйло В. Р., Михелев В. М. Повышение точности классификации рентгеновских снимков с использованием дообучения составной нейросети // Научный результат. Информационные технологии. – 2020. – Т. 5, № 1. – С. 20–26.

24. Газиева Т. В., Михайлов И. А., Мальков П. Г. [и др.]. Сравнение экономической эффективности проведения скрининга рака молочной железы маммографическим и ультразвуковым методами в системе здравоохранения Чеченской республики // Профилактическая медицина. – 2022. – Т. 25, № 5. – С. 52–60.

25. Ганцев Ш. Х., Франц М. В. Искусственный интеллект как инструмент поддержки в принятии решений по диагностике онкологических заболеваний // Медицинский вестник Башкортостана. – 2018. – Т. 13, № 4 (76). – С. 67–71.

26. Гиляревский С. Р., Гаврилов Д. В., Гусев А. В. Результаты ретроспективного анализа записей электронных амбулаторных медицинских карт пациентов с хронической сердечной недостаточностью: первый российский опыт // Российский кардиологический журнал. – 2021. – Т. 26, № 5. – С. 147–155.

27. Гиниятова Д. Х., Лапинский В. А. Детектирование новообразований на компьютерных томограммах с использованием нейронных сетей // Программные системы: теория и приложения. – 2022. – Т. 13, № 3 (54). – С. 61–79.

28. Глазнев М. Ю., Гусарова Н. Ф., Коцюба И. Ю. [и др.]. Сравнительный анализ методов классификации легочных узлов по изображениям компьютерной томографии // Оптический журнал. – 2017. – Т. 84, № 1. – С. 58–68.

29. Глушкова И. В., Кошкарров А. А., Мурашко Р. А. [и др.]. Региональная радиологическая информационная система Краснодарского края: организация работы референсного клинко-диагностического центра по патологии молочной железы // Врач и информационные технологии. – 2018. – № S1. – С. 18–27.

30. 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.

31. Гольдберг А. С. Организация служб лабораторной диагностики в России – текущий статус и перспективы развития // Бюллетень Национального научно-исследовательского института общественного здоровья имени Н. А. Семашко. – 2022. – № 1–2. – С. 26–35.

32. Гомболевский В. А., Чернина В. Ю., Блохин И. А. [и др.]. Основные достижения низкодозной компьютерной томографии в скрининге рака легкого // Туберкулез и болезни легких. – 2021. – Т. 99, № 1. – С. 61–70.

33. Гончарук Д. А., Велиев Е. И., Михайлов А. И. Прошлое, настоящее и будущее магнитно-резонансной томографии рака предстательной железы // Онкоурология. – 2021. – Т. 17, № 1. – С. 142–152.

34. 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.

35. Гусев А. В., Владзимирский А. В., Гавриленко Г. Г. Методический подход и рекомендации по научному описанию создания и валидации модели машинного обучения // Медицинские технологии. Оценка и выбор. – 2022. – № 3 (44). – С. 12–30.

36. Гусев А. В., Владзимирский А. В., Голубев Н. А. [и др.]. Информатизация здравоохранения Российской Федерации: история и результаты развития // Национальное здравоохранение. – 2021. – Т. 2, № 3. – С. 5–17

37. Гусев А. В., Гавриленко Г. Г., Гаврилов Д. В. Разработка модели машинного обучения для интерпретации результатов лабораторной диагностики с целью выявления подозрений на заболевания // Лабораторная служба. – 2022. – Т. 11, № 2. – С. 9–17.

38. Гусев А. В., Гаврилов Д. В., Новицкий Р. Э. [и др.]. Совершенствование возможностей оценки сердечно-сосудистого риска при помощи методов машинного обучения // Российский кардиологический журнал. – 2021. – Т. 26, № 12. – С. 171–180.

39. Гусев А. В., Добридюк С. Л. Искусственный интеллект в медицине и здравоохранении // Информационное общество. – 2017. – № 4–5. – С. 78–93.

40. Гусев А. В., Евгина С. А., Годков М. А. Искусственный интеллект в здравоохранении России. Роль лаборатории // Лабораторная служба. – 2022. – Т. 11, № 2. – С. 5–8.

41. Гусейнова С. Р., Завьялов Д. В. Результаты работы системы искусственного интеллекта CAD EYE производства Fujifilm при распознавании и дифференциальной диагностике новообразований толстой кишки в режиме реального времени в условиях эндоскопического отделения Ярославской областной онкологической больницы // Актуальные вопросы медицинской науки. – 2023. – № 1. – С. 229–230.

42. Данилов Г. В., Пронин И. Н., Королев В. В. [и др.]. Первые результаты неинвазивного типирования глиом головного мозга по данным магнитно-резонансной томографии с помощью машинного обучения // Вопросы нейрохирургии им. Н. Н. Бурденко. – 2022. – Т. 86, № 6. – С. 36–42.

43. Дин Н., Афанасьев Г. И. Состояние и перспективы применения искусственного интеллекта в визуализирующей диагностике заболеваний легких // E-Scio. – 2022. – № 4 (67). – С. 653–664.

44. Дорофеев Д. А., Казанова С. Ю., Мовсисян А. Б. [и др.]. Искусственный интеллект и нейросети в диагностике глаукомы // Национальный журнал глаукома. – 2023. – Т. 22, № 1. – С. 115–128.

45. Држевецкая К. С., Корженкова Г. П. Результаты двух лет маммографического скрининга на территории Калужской области // Вестник рентгенологии и радиологии. – 2022. – Т. 103, № 4–6. – С. 18–27.

46. Дрокин И. С., Еричева Е. В., Бухвалов О. Л. [и др.]. Опыт разработки и внедрения системы поиска онкологических образований с помощью искусственного интеллекта на примере рентгеновской компьютерной томографии легких // Врач и информационные технологии. – 2019. – № 3. – С. 48–57.

47. Думаев Р. И., Киряков И. М., Молодяков С. А. Особенности предобработки и сегментации изображений в задаче обнаружения COVID-19 по рентгеновским снимкам // Современная наука: актуальные проблемы теории и практики. Сер. Естественные и технические науки. – 2022. – № 9. – С. 88–95.

48. Елигулашвили Р. Р., Зароднюк И. В., Ачкасов С. И. [и др.]. Применение искусственного интеллекта в МРТ диагностике рака прямой кишки // Колопроктология. – 2022. – Т. 21, № 1 (79). – С. 26–36.

49. Заботнев М. С., Кулагин В. П. Методы и средства создания обучающих датасетов в задачах обработки медицинских изображений // Интеллектуальные информационные системы: теория и практика: сборник научных статей по материалам II Всероссийской конференции. – Курск, 2021. – С. 30–35.

50. Завадовский К. В., Веснина Ж. В., Анашбаев Ж. Ж. [и др.]. Современное состояние ядерной кардиологии в Российской Федерации // Российский кардиологический журнал. – 2022. – Т. 27, № 12. – С. 105–114.

51. Зеленина Л. И., Хаймина Л. Э., Деменкова Е. А. [и др.]. Сверточные нейронные сети в задаче классификации медицинских изображений // Современные наукоемкие технологии. – 2021. – № 9. – С. 68–73.

52. Зуков Р. А., Сафонцев И. П., Клименок М. П. [и др.]. Искусственный интеллект в диагностике рака легкого. Опыт Красноярского края // Вопросы онкологии. – 2023. – Т. 69, № 3S. – С. 94–95.

53. Иванова В. В. Исследование точности сегментации нижнечелюстного канала в зависимости от вида предварительной обработки изображений КТ // Наука настоящего и будущего. – 2023. – Т. 2. – С. 85–88.

54. Казакова В. А., Тюлякова С. А., Шивилов Е. В. [и др.]. Правовые основы применения технологий искусственного интеллекта в лучевой диагностике // Радиология – практика. – 2023. – № 2 (98). – С. 63–77.

55. Казакова В. А., Шивилов Е. В., Аничкина К. А. [и др.]. К вопросу о внедрении искусственного интеллекта в повседневную практику диагностики патологии молочной железы: все ли учтено? // Радиология – практика. – 2023. – № 3 (99). – С. 36–49.

56. Караваев А. Е. Интеллектуальная система распознавания туберкулеза с применением глубокого обучения // Отходы и ресурсы. – 2023. – Т. 10, № 1. – С. 1–13.

57. Кармазановский Г. Г., Кондратьев Е. В., Груздев И. С. [и др.]. Современная лучевая диагностика и интеллектуальные персонализированные технологии в гепатопанкреатологии // Вестник Российской академии медицинских наук. – 2022. – Т. 77, № 4. – С. 245–253.

58. Каталевская Е. А., Каталевский Д. Ю., Тюриков М. И. [и др.]. Перспективы использования искусственного интеллекта в диагностике и лечении заболеваний сетчатки. РМЖ // Клиническая офтальмология. – 2022. – Т. 22, № 1. – С. 36–43.

59. Кенц А. С., Хамад Ю. А., Симонов К. В. Апробация методики шиаплет-преобразования для визуализации патологических изменений легких на КТ-изображениях для диагностики COVID-19 // Медицина и высокие технологии. – 2021. – № 3. – С. 5–13.

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

61. Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента: монография / под ред. Ю. А. Васильева, А. В. Владзимирского. – М. : Издательские решения, 2022. – 388 с.

62. Костин К. А. Сегментация внутренних структур легких на данных компьютерной томографии с помощью методов глубинного обучения для пациентов с диссеминированным туберкулезом // Сборник избранных статей научной сессии ТУСУР. – 2019. – № 1–2. – С. 69–72.

63. Кошелев К. А., Бажанов Д. С., Белоусов Н. Н. [и др.]. Области современного применения и возможности искусственного интеллекта для 3D-визуализации в стоматологических исследованиях и практике // Проблемы стоматологии. – 2023. – Т. 19, № 2. – С. 5–10.

64. Кошкаров 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.

65. Кулаев К. И., Важенин А. В., Ростовцев Д. М. [и др.]. Искусственный интеллект в диагностике новообразований толстого кишечника – разработка,

внедрение технологии и первые результаты // Вопросы онкологии. – 2023. – Т. 69, № 2. – С. 292–299.

66. Куликов А. Н., Малахова Е. Ю., Мальцев Д. С. Искусственный интеллект и машинное обучение в диагностике центральной серозной хориоретинопатии на основании оптической когерентной томографии // Офтальмологические ведомости. – 2019. – Т. 12, № 1. – С. 13–20.

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

68. Курдюмов Д. А., Кашин А. В., Рябов Н. Ю. [и др.]. Опыт применения технологий искусственного интеллекта для развития профилактического здравоохранения на примере Кировской области // Менеджер здравоохранения. – 2023. – № 6. – С. 62–69.

69. Кушнир К. В. Управление качеством в современных отделениях лучевой диагностики // Медицинская визуализация. – 2015. – № 3. – С. 133–137.

70. Лантух З. А., Глигуров Ю. А., Солдатов И. В. [и др.]. Необоснованные направления на рентгенорадиологические исследования и их влияние на коллективную эффективную дозу пациентов в амбулаторных медицинских организациях // Радиационная гигиена. – 2023. – Т. 16, № 1. – С. 66–79.

71. Лебедев Г. С., Маслюков А. П., Шадеркин И. А. [и др.]. Глубокое машинное обучение (искусственный интеллект) в ультразвуковой диагностике // Журнал телемедицины и электронного здравоохранения. – 2020. – Т. 6, № 2. – С. 22–29.

72. Леин Г. А., Нечаева Н. С., Мамедова Г. М. К. [и др.]. Автоматизация анализа рентгенограмм позвоночника для объективизации оценки степени тяжести сколиотической деформации при идиопатическом сколиозе (предварительное сообщение) // Ортопедия, травматология и восстановительная хирургия детского возраста. – 2020. – Т. 8, № 3. – С. 317–326.

73. Ливзан М. А., Лялюкова Е. А., Петросян В. Ю. [и др.]. Скрининг злокачественных новообразований: состояние проблемы в России и мире // Экспериментальная и клиническая гастроэнтерология. – 2023. – № 3 (211). – С. 5–16.

74. Лисиенкова Л. Н., Митрофанова И. П. Перспективы использования искусственного интеллекта в диагностике COVID-19 // Качество. Инновации. Образование. – 2022. – № 2 (178). – С. 88–92.

75. Лыскова Ю. А., Сперанская А. А., Золотницкая В. П. [и др.]. Программы искусственного интеллекта лучевой диагностики в оценке нарушения кровообращения при внебольничной пневмонии до и во время пандемии COVID-19 // Регионарное кровообращение и микроциркуляция. – 2023. – Т. 22, № 1 (85). – С. 16–23.

76. Мазо М. Л., Блинов Д. С. Новая система компьютерного зрения на основе искусственного интеллекта как эффективный помощник для интерпретации рентгеновских маммограмм // VII Петербургский международный онкологический форум «Белые Ночи 2021»: материалы VII Петербургского международного онкологического форума. – СПб., 2021. – С. 311.

77. Макиев В. Г., Макиев Г. Г. Использование искусственного интеллекта для диагностики COVID-19 по данным компьютерной томографии // Актуальные научные исследования в современном мире. – 2020. – № 12–2 (68). – С. 73–81.

78. Мелдо А. А. Разработка и оценивание процессов системы менеджмента качества в условиях отделения лучевой диагностики бюджетного специализированного медицинского учреждения // Лучевая диагностика и терапия. – 2018. – № 1 (9). – С. 5–10.

79. Мелдо А. А., Уткин Л. В., Моисеенко В. М. Алгоритмы диагностики XXI века. Искусственный интеллект в распознавании рака легкого // Практическая онкология. – 2018. – Т. 19, № 3. – С. 292–298.

80. Мелдо А. А., Уткин Л. В., Трофимова Т. Н. [и др.]. Новые подходы к разработке алгоритмов искусственного интеллекта в диагностике рака легкого // Лучевая диагностика и терапия. – 2019. – № 1 (10). – С. 8–18.

81. 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.

82. Минязев Р. Ш., Румянцев А. А., Дыганов С. А. [и др.]. Анализ рентгеновских изображений для выявления патологий с использованием нейронных сетей // Известия Российской академии наук. Серия физическая. – 2018. – Т. 82, № 12. – С. 1685–1688.

83. Митькова М. Д., Швырев С. Л., Митьков В. В. Ультразвуковая диагностика в рутинной практике // Ультразвуковая и функциональная диагностика. – 2022. – № 1. – С. 44–50.

84. Михелев В. М., Мирошниченко А. С. Решение задачи классификации патологий головного мозга человека на снимках МРТ // Научный результат. Информационные технологии. – 2019. – Т. 4, № 2. – С. 43–52.

85. Моисеенко В. М., Мелдо А. А., Уткин Л. В. [и др.]. Автоматизированная система обнаружения объемных образований в легких как этап развития искусственного интеллекта в диагностике рака легкого // Лучевая диагностика и терапия. – 2018. – № 3 (9). – С. 62–68.

86. Нероев В. В., Брагин А. А., Зайцева О. В. Разработка прототипа сервиса для диагностики диабетической ретинопатии по снимкам глазного дна с использованием методов искусственного интеллекта // Национальное здравоохранение. – 2021. – Т. 2, № 2. – С. 64–72.

87. Никитаев В. Г., Тупицын Н. Н., Проничев А. Н. [и др.]. Технологии искусственного интеллекта в диагностике острых лимфобластных лейкозов и минимальной остаточной болезни // Медицинская техника. – 2020. – № 5 (323). – С. 42–44.

88. Орлов Е. М., Соколова О. Н. Категория эффективности в системе здравоохранения // Фундаментальные исследования. – 2010. – № 4. – С. 70–75.

89. Оценка качества рентгенорадиологических исследований / сост. С. П. Морозов, Н. Н. Ветшева, Н. В. Ледихова // Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 48. – М., 2019. – 47 с.

90. Павлов Н. А., Андрейченко А. Е., Владзимирский А. В. [и др.]. Эталонные медицинские датасеты (MosMedData) для независимой внешней оценки алгоритмов на основе искусственного интеллекта в диагностике // Digital Diagnostics. – 2021. – Т. 2, № 1. – С. 49–66.

91. Падалко М. А., Наумов А. М., Назариков С. И. [и др.]. Применение технологий искусственного интеллекта для диагностики туберкулеза и онкологических заболеваний // Туберкулез и болезни легких. – 2019. – Т. 97, № 11. – С. 62.

92. Парамзин Ф. Н., Какоткин В. В., Буркин Д. А. [и др.]. Радиомика и искусственный интеллект в дифференциальной диагностике опухолевых и неопухолевых заболеваний поджелудочной железы (обзор) // Хирургическая практика. – 2023. – Т. 8, № 1. – С. 53–65.

93. Паттохов А. Ш., Ходжибекова Ю. М., Ходжибеков М. Х. Выбор методов статистической обработки результатов радиомического анализа КТ-изображений опухолей головы и шеи // Медицинская радиология и радиационная безопасность. – 2023. – Т. 68, № 3. – С. 52–56.

94. Петри А., Сэбин К. Наглядная медицинская статистика / под ред. В. П. Леонова; пер. с англ. – М. : ГЭОТАР-Медиа, 2019. – 216 с.

95. Петрушин А. В., Коваленко А. В., Овсянникова А. В. Актуальные решения при диагностике заболеваний с использованием методов искусственного интеллекта и машинного обучения // Естественные и технические науки. – 2022. – № 6 (169). – С. 296–305.

96. Петрайкин А. В., Торощова Н. В., Никитинская О. А. [и др.]. Применение асинхронной количественной компьютерной томографии для оппортунистического скрининга остеопороза // Научно-практическая ревматология. – 2022. – Т. 60, № 3. – С. 360–368.

97. Пилюс П. С., Дрокин И. С., Баженова Д. А. [и др.]. Оценка перспектив использования технологий искусственного интеллекта для анализа КТ-изображений органов грудной клетки с целью выявления признаков злокачественных новообразований в легких // Медицинская визуализация. – 2023. – Т. 27, № 2. – С. 138–146.

98. Попова А. Ю., Водоватов А. В., Романович И. К. [и др.]. Влияние пандемии COVID-19 на структуру лучевой диагностики и коллективные дозы населения Российской Федерации при медицинском облучении в 2020 г. // Радиационная гигиена. – 2022. – Т. 15, № 3. – С. 6–39.

99. Пранович А. А., Исмаилов А. К., Карельская Н. А. [и др.]. Искусственный интеллект в диагностике и лечении мочекаменной болезни // Российский журнал телемедицины и электронного здравоохранения. – 2022. – Т. 8, № 1. – С. 42–57.

100. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения / под ред. В. З. Кучеренко. – М. : ГЭОТАР-Медиа, 2006. – 192 с.

101. Пугин К. В., Дорош М., Райковский Д. И. Реализация алгоритма автоматической сегментации аорты на КТ-снимках // Инновации. Наука. Образование. – 2022. – № 49. – С. 1196–1203.

102. Расмагина И. А., Бакулин И. Г., Машевский Г. А. [и др.]. Возможности искусственного интеллекта для клинико-лабораторной диагностики воспалительных заболеваний кишечника // Международная конференция по мягким вычислениям и измерениям. – 2023. – Т. 1. – С. 438–440.

103. Рашитов И. Н., Скородумова Е. А. Классификация внутричерепных кровоизлияний с помощью сверточных нейронных сетей // Телекоммуникации и информационные технологии. – 2023. – Т. 10, № 1. – С. 171–181.

104. Регламент подготовки наборов данных с описанием подходов к формированию репрезентативной выборки данных. Ч. 1: методические рекомендации / сост. С. П. Морозов, А. В. Владзимирский, А. Е. Андрейченко

[и др.] / Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 103. – М. : ГБУЗ «НПКЦ ДиТ ДЗМ», 2022. – 40 с.

105. Родионов Д. М., Карчков Д. А., Москаленко В. А. [и др.]. Диагностика синусового ритма и мерцательной аритмии средствами искусственного интеллекта // Проблемы информатики. – 2022. – № 1 (54). – С. 77–88.

106. 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. – С. 4–14.

107. Ройтберг П. Г., Блинов Д. С., Черемисин В. М. Технологии искусственного интеллекта в автоматизации выполнения стандартных задач врача-рентгенолога // Проблемы стандартизации в здравоохранении. – 2020. – № 9–10. – С. 29–33.

108. Румянцев П. О. Ранняя диагностика агрессивных форм рака: альянс специалистов, технологий и искусственного интеллекта // Лучевая диагностика и терапия. – 2023. – №2 (14). – С. 7–14.

109. Садулаева Т. А., Эдильгиреева Л. А., Бимурзаева М. Б. [и др.]. Использование искусственного интеллекта в цистоскопической диагностике рака мочевого пузыря // Онкоурология. – 2023. – Т. 19, № 2. – С. 146–152.

110. Сергеев В. Ю., Сергеев Ю. Ю., Тамразова О. Б. [и др.]. Исследование эффективности программ автоматизированной диагностики меланомы кожи с применением технологий искусственного интеллекта // Российский журнал кожных и венерических болезней. – 2020. – Т. 23, № 5. – С. 288–292.

111. Серговец А. А., Левин В. И., Борисов Д. Н. Современная функциональная диагностика и искусственный интеллект // Военно-медицинский журнал. – 2020. – Т. 341, № 2. – С. 40–45.

112. Серова Н. С., Пшеничникова Е. С., Шебунина А. Б. Сравнительный анализ применения методов лучевой диагностики в тактике ведения пациентов с новой коронавирусной инфекцией COVID-19 в разных странах мира //

Российский электронный журнал лучевой диагностики. – 2020. – Т. 10, № 4. – С. 8–20.

113. Скоробогач И. М., Хамидова Л. Т., Муслимов Р. Ш. [и др.]. Возможности машинного обучения в определении вариантов течения вирусной пневмонии, ассоциированной с COVID-19, на основании данных компьютерной томографии // Вестник медицинского института «РЕАВИЗ»: реабилитация, врач и здоровье. – 2023. – Т. 13, № 4. – С. 6–13.

114. Слуднова А. А., Шутько В. В., Гайдель А. В. [и др.]. Выявление патологических изменений в легких на основе совместного анализа радиологических отчетов и томографических изображений // Компьютерная оптика. – 2021. – Т. 45, № 2. – С. 261–266.

115. Смольникова У. А., Гаврилов П. В., Яблонский П. К. Диагностическая эффективность различных систем автоматического анализа рентгенограмм в выявлении периферических образований легких // Радиология – практика. – 2022. – № 3 (93). – С. 51–66.

116. Солодкий В. А., Каприн А. Д., Нуднов Н. В. [и др.]. Современные системы поддержки принятия врачебных решений на базе искусственного интеллекта для анализа цифровых маммографических изображений // Вестник рентгенологии и радиологии. – 2023. – Т. 104, № 2. – С. 151–162.

117. Солопова А. Е., Носова Ю. В., Бендженова Б. Б. Магнитно-резонансная томография при раке шейки матки: современные возможности радиомного анализа и перспективы развития методики // Акушерство, гинекология и репродукция. – 2023. – Т. 17, № 4. – С. 500–511.

118. Сперанская А. А. Роль искусственного интеллекта в оценке данных рентгеновской компьютерной томографии у пациентов с интерстициальными заболеваниями легких в условиях пандемии COVID-19 // Визуализация в медицине. – 2021. – Т. 3, № 3. – С. 3–12.

119. Сура М. В., Деркач Е. В. Формирование критериев оценки качества медицинской помощи при разработке клинических рекомендаций: проблемы и

пути их решения // Медицинские технологии. Оценка и выбор. – 2022. – № 4 (44). – С. 9–18.

120. Суровцев Е. Н., Капишников А. В., Колсанов А. В. Возможности радиомического анализа магнитно-резонансных томограмм в дифференциальной диагностике первичных внеозговых опухолей // Исследования и практика в медицине. – 2023. – Т. 10. № 2. – С. 50–61.

121. Тегин А. Д. Разработка программного модуля для определения наличия у человека легочных заболеваний с использованием нейронной сети // Молодой ученый. – 2023. – № 15 (462). – С. 10–12.

122. Терновой С. К., Серова Н. С., Беляев А. С. [и др.]. COVID-19: первые результаты лучевой диагностики в ответе на новый вызов // Российский электронный журнал лучевой диагностики. – 2020. – Т. 10, № 1. – С. 8–15.

123. Труфанов Г. Е., Ефимцев А. Ю. Технологии искусственного интеллекта в МР-нейровизуализации. Взгляд рентгенолога // Российский журнал персонализированной медицины. – 2023. – Т. 3, № 1. – С. 6–17.

124. Трухин А. А., Захарова С. М., Дунаев М. Е. [и др.]. Роль искусственного интеллекта в дифференциальной ультразвуковой диагностике узловых образований щитовидной железы // Клиническая и экспериментальная тиреоидология. – 2022. – Т. 18, № 2. – С. 32–38.

125. Тюльпин В. А., Уткин Л. В., Голубева И. Э. Разработка интеллектуальной подсистемы обнаружения и выделения новообразований в легком // Математические методы в технике и технологиях – ММТТ. – 2020. – Т. 10. – С. 120–123.

126. Тюрин И. Е. Лучевая диагностика в Российской Федерации // Онкологический журнал: лучевая диагностика, лучевая терапия. – 2018. – Т. 1, № 4. – С. 43–51.

127. Филиппова Ю. А., Ахвердиева Г. И. К. Возможности применения методов искусственного интеллекта в выявлении и оценке поражения легких у пациентов с пневмонией COVID-19 // E-Scio. – 2023. – № 5 (80). – С. 198–218.

128. Филь Т. С. Цифровое здравоохранение в Российской Федерации: текущий этап трансформации // Вестник Северо-Западного государственного медицинского университета им. И. И. Мечникова. – 2022. – Т. 14, № 1. – С. 39–50.

129. Харламова Н. В., Ясинский И. Ф., Ананьева М. А. [и др.]. Использование искусственного интеллекта для диагностики заболеваний и прогнозирования их исходов у новорожденных // Российский вестник перинатологии и педиатрии. – 2023. – Т. 68, № 4. – С. 108–114.

130. Цибин А. Лабораторная служба Москвы: под знаком централизации // Московская медицина. – 2016. – № S1 (12). – С. 33–36.

131. Цыбикова Э. Б., Мидоренко Д. А., Лапшина И. С., [и др.]. Организация скрининга в эпидемических туберкулезных очагах в субъектах Российской Федерации с низким уровнем заболеваемости туберкулезом // Социальные аспекты здоровья населения. – 2023. – Т. 69, № 2. – С. 1–31.

132. Черненький И. М., Черненький М. М., Фиев Д. Н. [и др.]. Сегментация почечных структур по изображениям контрастной компьютерной томографии с помощью сверточной нейронной сети // Сеченовский вестник. – 2023. – Т. 14, № 1. – С. 39–49.

133. Шамаев Д. М., Заяц В. В., Орлов С. Б. [и др.]. Синтетические датасеты: возможности развития продуктов медицинского искусственного интеллекта // Искусственный интеллект и принятие решений. – 2023. – № 1. – С. 100–107.

134. Шарифуллин Ф. А. К., Долотова Д. Д., Бармина Т. Г. [и др.]. Создание коллекции МСКТ-изображений и клинических данных при острых нарушениях мозгового кровообращения // Неотложная медицинская помощь. Журнал им. Н. В. Склифосовского. – 2020. – Т. 9, №2. – С. 231–237.

135. 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.

136. Шелехов П. В. Кадровая ситуация в лучевой диагностике // Современные проблемы здравоохранения и медицинской статистики. – 2019. – № 1. – С. 265–275.

137. Шелехов П. В. Эффективность использования оборудования лучевой диагностики в субъектах Российской Федерации // Менеджер здравоохранения. – 2017. – № 5. – С. 33–41.

138. Шелехов П. В., Омеляновский В. В. Анализ парка рентгенологического оборудования в Российской Федерации // Медицинские технологии. Оценка и выбор. – 2023. – № 3 (45). – С. 26–32.

139. Шулькин И. М., Владзимирский А. В. Управление на основе данных в лучевой диагностике: оценка результативности модели Единого радиологического информационного сервиса // Менеджер здравоохранения. – 2022. – № 7. – С. 68–80.

140. Abesi F., Jamali A. S., Zamani M. Accuracy of artificial intelligence in the detection and segmentation of oral and maxillofacial structures using cone-beam computed tomography images: a systematic review and meta-analysis // Pol J Radiol. – 2023. – Vol. 88. – P. e256-e263. – DOI: <https://doi.org/10.5114/pjr.2023.127624>.

141. Adamou A., Beltsios E. T., Bania A., et al. Artificial intelligence-driven ASPECTS for the detection of early stroke changes in non-contrast CT: a systematic review and meta-analysis // J NeuroIntervent Surg Epub. – 2022. – DOI: 10.1136/jnis-2022-019447.

142. Agarwal S., Wood D., Grzeda M., et al. Systematic Review of Artificial Intelligence for Abnormality Detection in High-volume Neuroimaging and Subgroup Meta-analysis for Intracranial Hemorrhage Detection // Clinical Neuroradiology. – 2023. – P. 1–14. – DOI: 10.1007/s00062-023-01291-1.

143. Aggarwal R., Sounderajah V., Martin G., et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis // NPJ Digital Medicine. – 2021. – Vol. 4. – P. 65. – DOI: 10.1038/s41746-021-00438-z.

144. Agrawal A., Khatri G. D., Khurana B., et al. A survey of ASER members on artificial intelligence in emergency radiology: trends, perceptions, and expectations //

Emerg Radiol. – 2023. – Vol. 30, №3. – P. 267–277. – DOI: 10.1007/s10140-023-02121-0.

145. Alsharif W., Qurashi A., Toonsi F., et al. A qualitative study to explore opinions of Saudi Arabian radiologists concerning AI-based applications and their impact on the future of the radiology // *BJR Open*. – 2022. – Vol. 4, №1. – P. 20210029. – DOI: 10.1259/bjro.20210029.

146. Arzamasov K., Vasilev Y., Vladzimirskyy A., et al. An International Non-Inferiority Study for the Benchmarking of AI for Routine Radiology Cases: Chest X-ray, Fluorography and Mammography // *Healthcare*. – 2023. – Vol. 11, №12. – P. 1684.

147. Badr F. F., Jadu F. M. Performance of artificial intelligence using oral and maxillofacial CBCT images: A systematic review and meta-analysis // *Niger J Clin Pract*. – 2022. – №25. – P. 1918–1927. – DOI: 10.4103/njcp.njcp_394_22.

148. Bae M. S., Moon W. K., Chang J. M., et al. Breast cancer detected with screening US: reasons for nondetection at mammography // *Radiology*. – 2014. – Vol. 270, №2. – P. 369–377. – DOI: 10.1148/radiol.13130724.

149. Bailey A. L., Ledebor N., Burnham C. D. Clinical Microbiology Is Growing Up: The Total Laboratory Automation Revolution // *Clin Chem*. – 2019. – Vol. 65, №5. – P. 634–643. – DOI: 10.1373/clinchem.2017.27452.

150. Bedrikovetski S., Dudi-Venkata N. N., et al. Artificial intelligence for the diagnosis of lymph node metastases in patients with abdominopelvic malignancy: A systematic review and meta-analysis // *Artificial Intelligence in Medicine*. – 2021. – Vol. 113. – P. 102022. – DOI: 10.1016/j.artmed.2021.102022.

151. Bedrikovetski S., Dudi-Venkata N. N., Kroon H. M., et al. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis // *BMC Cancer*. – 2021. – №21. – P. 1058. – DOI: 10.1186/s12885-021-08773-w.

152. Belue M. J., Yilmaz E. C., Daryanani A., et al. Current Status of Biparametric MRI in Prostate Cancer Diagnosis: Literature Analysis // *Life (Basel)*. – 2022. – Vol. 12, №6. – P. 804. – DOI: 10.3390/life12060804.

153. Bossuyt P. M., Reitsma J. B., Bruns D. E., et al. For the STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies // *Radiology*. – 2015. – Vol. 277, №3. – P. 826–832.

154. Bruno M. A. Extending the Scope of Quality and Safety in Radiology // *Radiology*. – 2022. – Vol. 302, №3. – P. 620–621. – DOI: 10.1148/radiol.2021212538.

155. Campello C. A., Castanha E. B., Vilardo M., et al. Machine learning for malignant versus benign focal liver lesions on US and CEUS: a meta-analysis // *Abdominal Radiology*. – 2023. – Vol. 48, №10. – P. 3114–3126. – DOI: 10.1007/s00261-023-03984-0.

156. Cao D. J., Hurrell C., Patlas M. N. Current Status of Burnout in Canadian Radiology // *Can Assoc Radiol J*. – 2023. – Vol. 74, №1. – P. 37–43. – DOI: 10.1177/08465371221117282.

157. Chamberlin J., Kocher M. R., Waltz J., et al. Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value // *BMC Med*. – 2021. – Vol. 19, № 1. – P. 55. – DOI: 10.1186/s12916-021-01928-3.

158. Chan S. S., Francavilla M. L., Iyer R. S., et al. Clinical decision support: the role of ACR Appropriateness Criteria // *Pediatr Radiol*. – 2019. – Vol. 49, № 4. – P. 479–485. – DOI: 10.1007/s00247-018-4298-2.

159. Chang Y. W., An J. K., Choi N., et al. Artificial Intelligence for Breast Cancer Screening in Mammography (AI-STREAM): A Prospective Multicenter Study Design in Korea Using AI-Based CADe/x // *J Breast Cancer*. – 2022. – Vol. 25, № 1. – P. 57–68. – DOI: 10.4048/jbc.2022.25.e4.

160. Chepelev L. L., Wang X., Gold B., Bonzel C. L., et al. Improved Appropriateness of Advanced Diagnostic Imaging After Implementation of Clinical Decision Support Mechanism // *J Digit Imaging*. 2021. – Vol. 34, № 2. – P. 397–403. – DOI: 10.1007/s10278-021-00433-6.

161. Chetlen A. L., Petscavage-Thomas J., Cherian R. A., et al. Collaborative Learning in Radiology: From Peer Review to Peer Learning and Peer Coaching // *Acad Radiol*. – 2020. – Vol. 27, №9. – P. 1261–1267. – DOI: 10.1016/j.acra.2019.09.021.

162. Chi J. M., Makaryus J. N., Rahmani N., et al. Coronary CT Calcium Score in Patients With Prior Nongated CT, Is it Necessary? // *Curr Probl Diagn Radiol.* – 2021. – Vol. 50, № 1. – P. 54–58. – DOI: 10.1067/j.cpradiol.2019.07.011.

163. Cho J., Lee S., Gu B. S., et al. The Impact of COVID-19 on the Use of Radiology Resources in a Tertiary Hospital // *J Korean Med Sci.* – 2020. – Vol. 35, №40. – P. e368. – DOI: 10.3346/jkms.2020.35.e368.

164. Cooper K., Heilbrun M. E., Gilyard S., et al. Shared Decision Making: Radiology's Role and Opportunities // *AJR Am J Roentgenol.* – 2020. – Vol. 214, № 1. – P. W62–W66. – DOI: 10.2214/AJR.19.21590.

165. Coppola F., Faggioni L., Regge D., et al. Artificial intelligence: radiologists' expectations and opinions gleaned from a nationwide online survey // *Radiol Med.* – 2021. – Vol. 126, № 1. – P. 63–71. – DOI: 10.1007/s11547-020-01205-y.

166. Da Silva R. B. Z., Fogliatto F. S., Krindges A., et al. Dynamic capacity allocation in a radiology service considering different types of patients, individual no-show probabilities, and overbooking // *BMC Health Serv Res.* – 2021. – Vol. 21, №1. – P. 968. – DOI: 10.1186/s12913-021-06918-y.

167. Din M., Agarwal S., Grzeda M., et al. Detection of cerebral aneurysms using artificial intelligence: a systematic review and meta-analysis // *J NeuroIntervent Surg.* – 2023. – №15. – P. 262–271. – DOI: 10.1136/jnis-2022-019456.

168. Dumitrescu E. A., Ungureanu B. S., Cazacu I. M., et al. Diagnostic Value of Artificial Intelligence-Assisted Endoscopic Ultrasound for Pancreatic Cancer: A Systematic Review and Meta-Analysis // *Diagnostics.* – 2022. – № 12. – P. 309.

169. Ekpo E. U., Alakhras M., Brennan P. Errors in Mammography Cannot be Solved through Technology Alone // *Asian Pac J Cancer Prev.* – 2018. – Vol. 19, № 2. – P. 291–301. – DOI: 10.22034/APJCP.2018.19.2.291.

170. European Society of Radiology (ESR). Current practical experience with artificial intelligence in clinical radiology: a survey of the European Society of Radiology // *Insights Imaging.* – 2022. – Vol. 13, №1. – P. 107. – DOI: 10.1186/s13244-022-01247-y.

171. European Society of Radiology (ESR). Impact of artificial intelligence on radiology: a EuroAIM survey among members of the European Society of Radiology // *Insights Imaging*. – 2019. – Vol. 10, №1. – P. 105. – DOI: 10.1186/s13244-019-0798-3.

172. Evangelista K., de Freitas Silva B. S., Yamamoto-Silva F. P., et al. Accuracy of artificial intelligence for tooth extraction decision-making in orthodontics: a systematic review and meta-analysis // *Clinical Oral Investigations*. – 2022. – № 26. – P. 6893–6905. – DOI: 10.1007/s00784-022-04742-0.

173. 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*. – 2013. – Vol. 8, № 5. – P. e64366. – DOI: 10.1371/journal.pone.0064366.

174. Feng L., Liu Z., Li C., et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study // *Lancet Digit Health*. – 2022. – Vol. 4, №1. – P. e8–e17. – DOI: 10.1016/S2589-7500(21)00215-6.

175. Freeman K., Geppert J., Stinton C., et al. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy // *BMJ*. – 2021. – №374. – P. n1872. – DOI: 10.1136/bmj.n1872.

176. Gampala S., Vankeshwaram V., Gadula S. S. P. Is Artificial Intelligence the New Friend for Radiologists? A Review Article // *Cureus*. – 2020. – Vol. 24, №12 (10). – P. e11137. – DOI: 10.7759/cureus.11137.

177. Gao L., Jiao T., Feng Q., et al. Application of artificial intelligence in diagnosis of osteoporosis using medical images: a systematic review and meta-analysis // *Osteoporosis International*. – Vol. 32, №7. – P. 1279–1286. – DOI: 10.1007/s00198-021-05887-6.

178. Golding L. P., Nicola G. N. Clinical Decision Support: The Law, the Future, and the Role for Radiologists // *Curr Probl Diagn Radiol*. – 2020. – Vol. 49, № 5. – P. 337–339. – DOI: 10.1067/j.cpradiol.2020.02.001.

179. Hendrix N., Lowry K. P., Elmore J. G., et al. Radiologist Preferences for Artificial Intelligence-Based Decision Support During Screening Mammography

Interpretation // J Am Coll Radiol. – 2022. – Vol. 19, №10. – P. 1098–1110. – DOI: 10.1016/j.jacr.2022.06.019.

180. Hickman S. E., Woitek R., Le E. Ph. V., et al. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis // Radiology. – 2022. – Vol. 302. – P. 88–104. – DOI: 10.1148/radiol.2021210391.

181. Hoffman R. M., Atallah R. P., Struble R. D., et al. Lung Cancer Screening with Low-Dose CT: a Meta-Analysis // J Gen Intern Med. – 2020. – Vol. 35, №10. – P. 3015–3025. – DOI: 10.1007/s11606-020-05951-7.

182. Hosny A., Parmar Ch., Quackenbush J., et al. Aerts, Artificial intelligence in radiology // Nature Reviews Cancer. – 2018. – Vol. 18. – P. 500–510. – URL: <https://www.nature.com/articles/s41568-018-0016-5> (дата обращения: 01.09.2023).

183. Hovda T., Larsen M., Romundstad L., et al. Breast cancer missed at screening; hindsight or mistakes? // Eur J Radiol. – 2023. – №165. – P. 110913. – DOI: 10.1016/j.ejrad.2023.110913.

184. International Prospective Register of Systematic Reviews. – URL: <https://www.crd.york.ac.uk/PROSPERO> (дата обращения: 01.09.2023).

185. Islam M., Poly T. N., Walther B. A., et al. Deep Learning for the Diagnosis of Esophageal Cancer in Endoscopic Images: A Systematic Review and Meta-Analysis // Cancers. – 2022. – Vol. 14, №23. – P. 5996. – DOI: 10.3390/cancers14235996.

186. Jabin M. S. R., Schultz T., Mandel C., et al. A Mixed-Methods Systematic Review of the Effectiveness and Experiences of Quality Improvement Interventions in Radiology // J Patient Saf. – 2022. – Vol. 1, №18. – P. e97–e107. – DOI: 10.1097/PTS.0000000000000709.

187. Jiang B., Li N., Shi X., et al. Deep Learning Reconstruction Shows Better Lung Nodule Detection for Ultra-Low-Dose Chest CT // Radiology. – 2022. – Vol. 303, № 1. – P. 202–212. – DOI: 10.1148/radiol.210551.

188. Jørgensen M. D., Antulov R., Hess S., et al. Convolutional neural network performance compared to radiologists in detecting intracranial hemorrhage from brain

computed tomography: A systematic review and meta-analysis // *European Journal of Radiology*. – 2022. – Vol. 146. – P. 110073. – DOI: 10.1016/j.ejrad.2021.110073.

189. Kapoor N., Lacson R., Khorasani R. Workflow Applications of Artificial Intelligence in Radiology and an Overview of Available Tools // *J Am Coll Radiol*. – 2020. – Vol. 17, № 11. – P. 1363–1370. – DOI: 10.1016/j.jacr.2020.08.016.

190. Keane P. A., Topol E. J. With an eye to AI and autonomous diagnosis // *NPJ Digital Med*. – 2018. – Aug 28. – № 1. – P. 40. – DOI: 10.1038/s41746-018-0048-y.

191. 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. – DOI: 10.1007/s00330-022-08784-6.

192. 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. 405–410. – DOI: 10.3348/kjr.2019.0025.

193. Kim H.Y., Cho S. J., Sunwoo L., et al. Classification of true progression after radiotherapy of brain metastasis on MRI using artificial intelligence: a systematic review and meta-analysis // *Neuro-Oncology Advances*. – 2021. – Vol. 3, №1. – P. 1–12. – DOI: 10.1093/oaajnl/vdab080.

194. Kligerman S. J., Mitchell J. W., Sechrist J. W., et al. Radiologist Performance in the Detection of Pulmonary Embolism: Features that Favor Correct Interpretation and Risk Factors for Errors // *J Thorac Imaging*. – 2018. – Vol. 33, № 6. – P. 350–357. – DOI: 10.1097/RTI.0000000000000361.

195. Kligerman S., Cai L., White C. S. The effect of computer-aided detection on radiologist performance in the detection of lung cancers previously missed on a chest radiograph // *J Thorac Imaging*. – 2013. – Vol. 28, № 4. – P. 244–252. – DOI: 10.1097/RTI.0b013e31826c29ec.

196. Komolafe T. E., Cao Yu., Nguchu B. A., et al. Diagnostic Test Accuracy of Deep Learning Detection of COVID-19: A Systematic Review and Meta-Analysis // *Academic Radiology*. – 2021. – Vol 28, № 11. – DOI: 10.1016/j.acra.2021.08.008.

197. Korevaar D. A., Gopalakrishna G., Cohen J. F., et al. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses // *Diagn Progn Res.* – 2019. – Vol. 3, № 22. – DOI: 10.1186/s41512-019-0069-2.

198. Kuo R. Y. L., Harrison C., Curran T.-A., et al. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis // *Radiology.* – 2022. – Vol. 304. – P. 50–62. – DOI: 10.1148/radiol.211785.

199. Lex J. R., Michele J. D., Koucheki R., et al. Artificial Intelligence for Hip Fracture Detection and Outcome Prediction A Systematic Review and Meta-analysis // *JAMA Network Open.* – 2023. – Vol. 6, № 3. – P. e233391. – DOI: 10.1001/jamanetworkopen.2023.3391.

200. Li D., Basilico R., Blanco A. et al. Emergency Radiology: Evolution, Current Status, and Future Directions // *Can Assoc Radiol J.* – 2022. – Vol. 73, № 4. – P. 697–703. – DOI: 10.1177/08465371221088924.

201. Li N., Tan F., Chen W., et al. National Lung Cancer Screening programme group. One-off low-dose CT for lung cancer screening in China: a multicentre, population-based, prospective cohort study // *Lancet Respir Med.* – 2022. – Vol. 10, № 4. – P. 378–391. – DOI: 10.1016/S2213-2600(21)00560-9.

202. Liu D. S., Abu-Shaban K., Halabi S. S., et al. Changes in Radiology Due to Artificial Intelligence That Can Attract Medical Students to the Specialty // *JMIR Med Educ.* – 2023. – № 9. – P. e43415. – DOI: 10.2196/43415.

203. Liu M., Wu J., Wang N., et al. The value of artificial intelligence in the diagnosis of lung cancer: A systematic review and meta-analysis // *PLoS One.* – 2023. – Vol. 18, № 3. – P. e0273445. – DOI: 10.1371/journal.pone.0273445.

204. Liu X., Rivera S. C., Moher D., et al. On behalf of the SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension // *Lancet Digital Health.* – 2020. – Vol. 2, № 10. – P. e537–e548.

205. Martinez G., Katz J. M., Pandya A. et al. Cost-Effectiveness Study of Initial Imaging Selection in Acute Ischemic Stroke Care // *J Am Coll Radiol.* – 2021. – Vol. 18, № 6. – P. 820–833. – DOI: 10.1016/j.jacr.2020.12.013.

206. Martiniussen M. A., Sagstad S., Larsen M., et al. Screen-detected and interval breast cancer after concordant and discordant interpretations in a population based screening program using independent double reading // *Eur Radiol.* – 2022. – Vol. 32, № 9. – P. 5974–5985. – DOI: 10.1007/s00330-022-08711-9.

207. Mazurowski M. A. Artificial Intelligence in Radiology: Some Ethical Considerations for Radiologists and Algorithm Developers // *Acad Radiol.* – 2020. – Vol. 27, №1. – P. 127–129. – DOI: 10.1016/j.acra.2019.04.024.

208. Menon N., Guidozi N., Chidambaram S., et al. Performance of radiomics-based artificial intelligence systems in the diagnosis and prediction of treatment response and survival in esophageal cancer: asystematic review and meta-analysis of diagnostic accuracy // *Diseases of the Esophagus.* – 2023. – Vol. 36, № 6. – P. doad034. – DOI: 10.1093/dote/doad034.

209. Michalopoulou E., Clauser P., Gilbert F. J., et al. A survey by the European Society of Breast Imaging on radiologists' preferences regarding quality assurance measures of image interpretation in screening and diagnostic mammography // *Eur Radiol.* – 2023. – Jul 22. – DOI: 10.1007/s00330-023-09973-7.

210. Milam M. E., Koo C. W. The current status and future of FDA-approved artificial intelligence tools in chest radiology in the United States // *Clin Radiol.* – 2023. – Vol. 78, № 2. – P. 115–122. – DOI: 10.1016/j.crad.2022.08.135.

211. Mohan G., Chattopadhyay S. Cost-effectiveness of Leveraging Social Determinants of Health to Improve Breast, Cervical, and Colorectal Cancer Screening: A Systematic Review // *JAMA Oncol.* – 2020. – Vol. 6, № 9. – P. 1434–1444. – DOI: 10.1001/jamaoncol.2020.1460.

212. Morozov S., Guseva E., Ledikhova N., Vladzimirskyy A., Safronov D. Telemedicine-based system for quality management and peer review in radiology // *Insights Imaging.* – 2018. – № 9 (3). – P. 337–341.

213. Nabizadeh F., Ramezannezhad E., Kargar A., et al. Diagnostic performance of artificial intelligence in multiple sclerosis: a systematic review and meta-analysis // *Neurological Sciences*. – 2023. – Vol. 44. – P. 499–517. – DOI: 10.1007/s10072-022-06460-7.

214. Nagendran M., Chen Y., Lovejoy C. A., et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies // *BMJ*. – 2020. – Vol. 25, № 368. – P. m689. – DOI: 10.1136/bmj.m689.

215. Naugler C., Church D. L. Automation and artificial intelligence in the clinical laboratory // *Crit Rev Clin Lab Sci*. – 2019. – Vol. 56, № 2. – P. 98–110. – DOI: 10.1080/10408363.2018.1561640.

216. Page M. J., McKenzie J. E., Bossuyt P. M., et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews // *BMJ*. – 2021. – №372. – P. n71. – DOI: 10.1136/bmj.n71.

217. Palazzetti V., Guidi F., Ottaviani L., et al. Analysis of mammographic diagnostic errors in breast clinic // *Radiol Med*. – 2016. – Vol. 121, № 11. – P. 828–833. – DOI: 10.1007/s11547-016-0655-0.

218. Pinnock H., Barwick M., Carpenter C. et al. Standards for Reporting Implementation Studies (StaRI) statement // *BMJ*. – 2017. – № 356. – P. i6795.

219. Poly T. N., Islam Md M., Li Y.-Ch. J., et al. Application of Artificial Intelligence for Screening COVID-19 Patients Using Digital Images: Meta-analysis // *JMIR Med Inform*. – 2021. – Vol. 9, № 4. – P. e21394. – URL: <https://medinform.jmir.org/2021/4/e21394> (дата обращения: 01.09.2023).

220. Potipimpanon P., Charakorn N., Hirunwiwatkul P. A comparison of artificial intelligence versus radiologists in the diagnosis of thyroid nodules using ultrasonography: a systematic review and meta-analysis // *European Archives of Oto-Rhino-Laryngology*. – 2022. – № 279. – P. 5363–5373. – DOI: 10.1007/s00405-022-07436-1.

221. QUADAS-2: University of Bristol. – URL: <https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2> (дата обращения: 01.09.2023).

222. Radiomics Quality Score – RQS 2.0. – URL: <https://www.radiomics.world/rqs2> (дата обращения: 01.09.2023).

223. Rescinito R., Ratti M., Payedimarri A. B., et al. Prediction Models for Intrauterine Growth Restriction Using Artificial Intelligence and Machine Learning: A Systematic Review and Meta-Analysis // *Healthcare*. – 2023. – № 11. – P. 1617. – DOI: 10.3390/healthcare11111617.

224. Rudolph J., Schachtner B., Fink N., et al. Clinically focused multi-cohort benchmarking as a tool for external validation of artificial intelligence algorithm performance in basic chest radiography analysis // *Sci Rep*. – 2022. – Vol. 12, № 1. – P. 12764. – DOI: 10.1038/s41598-022-16514-7.

225. Sadr S., Mohammad-Rahimi H., Motamedian S. R., et al. Deep Learning for Detection of Periapical Radiolucent Lesions: A Systematic Review and Meta-analysis of Diagnostic Test Accuracy // *Journal of Endodontics*. – Vol. 49, № 3. – P. 248–261. – P. e3. – DOI: 10.1016/j.joen.2022.12.007.

226. Schlemmer H. P., Bittencourt L. K., D'Anastasi M., et al. Global Challenges for Cancer Imaging // *J Glob Oncol*. – 2018. – Sep 4. – P. 1–10. – DOI: 10.1200/JGO.17.00036.

227. Schoonenboom J., Johnson R. B. How to Construct a Mixed Methods Research Design // *Kolner Z Soz Sozpsychol*. – 2017. – Vol. 69 (Suppl 2). – P. 107–131. – DOI: 10.1007/s11577-017-0454-1.

228. Shea B. J., Reeves B. C., Wells G., et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both // *BMJ*. – 2017. – № 358. – P. j4008. – DOI: 10.1136/bmj.j4008.

229. Soffer Sh., Klang E., Shimon O., et al. Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis // *Scientific Reports*. – 2021. – № 11. – P. 15814. – DOI: 10.1038/s41598-021-95249-3.

230. Sogani J., Allen Jr. B., Dreyer K., et al. Artificial intelligence in radiology: the ecosystem essential to improving patient care // *Clin Imaging*. – 2019. – Vol. 59, №1. – P. A3–A6. – DOI: 10.1016/j.clinimag.2019.08.001.

231. Subramaniam R. M., Kurth D. A., Waldrip C. A., et al. American College of Radiology Appropriateness Criteria: Advancing Evidence-Based Imaging Practice // *Semin Nucl Med*. – 2019. – Vol. 49, № 2. – P. 161–165. – DOI: 10.1053/j.semnuclmed.2018.11.011.

232. Sugibayashi T., Walston S. L., Matsumoto T., et al. Deep learning for pneumothorax diagnosis: a systematic review and meta-analysis // *Eur Respir Rev*. – 2023. – № 32. – P. 220259. – DOI: 10.1183/16000617.0259-2022.

233. Tækker M., Kristjánisdóttir B., Graumann O., et al. Diagnostic accuracy of low-dose and ultra-low-dose CT in detection of chest pathology: a systematic review // *Clin Imaging*. – 2021. – № 74. – P. 139–148. – DOI: 10.1016/j.clinimag.2020.12.041.

234. Taib B. G., Karwath A., Wensley K., et al. Artificial intelligence in the management and treatment of burns: A systematic review and meta-analyses // *Journal of Plastic, Reconstructive & Aesthetic Surgery*. – 2023. – № 77. – P. 133–161. – DOI: 10.1016/j.bjps.2022.11.049.

235. Tariq A., Purkayastha S., Padmanaban G. P., et al. Current Clinical Applications of Artificial Intelligence in Radiology and Their Best Supporting Evidence // *J Am Coll Radiol*. – 2020. – Vol. 17, №11. – P. 1371–1381. DOI: 10.1016/j.jacr.2020.08.018.

236. Thomas S. P., Fraum T., Ngo L., et al. Leveraging Artificial Intelligence to Enhance Peer Review: Missed Liver Lesions on Computed Tomographic Pulmonary Angiography // *J Am Coll Radiol*. – 2022. – Vol. 19, № 11. – P. 1286–1294. – DOI: 10.1016/j.jacr.2022.07.013.

237. Thong L. T., Chou H. Sh., Chew H. Sh. J., et al. Diagnostic test accuracy of artificial intelligence-based imaging for lung cancer screening: A systematic review and meta-analysis // *Lung Cancer*. – 2023. – № 176. – P. 4–13. – DOI: 10.1016/j.lungcan.2022.12.002.

238. Tricco A. C., Lillie E., Zarin W., et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation // *Ann Intern Med.* – 2018. – Vol. 169, № 7. – P. 467–473.

239. Van Leeuwen K. G., Schalekamp S., Rutten M. 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. – DOI: 10.1007/s00330-021-07892-z.

240. Vang S., Margolies L. R., Jandorf L. Mobile Mammography Participation Among Medically Underserved Women: A Systematic Review // *Prev Chronic Dis.* – 2018. – Vol. 15. – P. E140. – DOI: 10.5888/pcd15.180291.

241. Vasilev Y., Vladzimirskyy A., Omelyanskaya O., et al. AI-Based CXR First Reading: Current Limitations to Ensure Practical Value // *Diagnostics.* – 2023. – Vol. 13, № 8. – P. 1430.

242. Vasilev Y. A., Sergunova K. A., Bazhin A. V., et al. Chest MRI of patients with COVID-19 // *Magn Reson Imaging.* – 2021. – № 79. – P. 13–19. – DOI: 10.1016/j.mri.2021.03.005.

243. Wang Q., Ma J., Zhang L., et al. Diagnostic performance of corona virus disease 2019 chest computer tomography image recognition based on deep learning: Systematic review and meta-analysis // *Medicine.* – 2022. – № 101. – P. 42(e31346). – DOI: 10.1097/MD.00000000000031346.

244. Wintermark M., Willis M. H., Hom J., et al. Everything Every Radiologist Always Wanted (and Needs) to Know About Clinical Decision Support // *J Am Coll Radiol.* – 2020. – Vol. 17, № 5. – P. 568–573. – DOI: 10.1016/j.jacr.2020.03.016.

245. Woodhouse K. D., Tremont K., Vachani A., et al. A Review of Shared Decision-Making and Patient Decision Aids in Radiation Oncology // *J Cancer Educ.* – 2017. – Vol. 32, №2. – P. 238–245. – DOI: 10.1007/s13187-017-1169-8.

246. Xing W., Gao W., Lv X., et al. Artificial intelligence predicts lung cancer radiotherapy response: A meta-analysis // *Artificial Intelligence In Medicine.* – 2023. – № 142. – P. 102585. – DOI: 10.1016/j.artmed.2023.102585.

247. Xu H.-L., Gong T.-T., Liu F.-H., et al. Artificial intelligence performance in image-based ovarian cancer identification: A systematic review and meta-analysis // *EClinicalMedicine*. – 2022. – № 53. – P. 101662.

248. Xu Y., Ding W., Wang Y., et al. Comparison of diagnostic performance between convolutional neural networks and human endoscopists for diagnosis of colorectal polyp: A systematic review and meta-analysis // *PLoS One*. – 2021. – Vol. 16, № 2. – P. e0246892. – DOI: 10.1371/journal.pone.0246892.

249. Xue Y., Zhou Y., Wang T., et al. Accuracy of Ultrasound Diagnosis of Thyroid Nodules Based on Artificial Intelligence-Assisted Diagnostic Technology: A Systematic Review and Meta-Analysis // *International Journal of Endocrinology*. – 2022. – P. 9492056. – DOI: 10.1155/2022/9492056.

250. Yang Ch., Qin L.-H., Xie Y.-E., et al. Deep learning in CT image segmentation of cervical cancer: a systematic review and meta-analysis // *Radiat Oncol*. – 2022. – № 17. – P. 175. – DOI: 10.1186/s13014-022-02148-6.

251. Yang L., Ene I. C., Arabi Belaghi R., et al. Stakeholders' perspectives on the future of artificial intelligence in radiology: a scoping review // *Eur Radiol*. – 2022. – Vol. 32, № 3. – P. 1477–1495. – DOI: 10.1007/s00330-021-08214-z.

252. Yoon J. H., Strand Fr., Baltzer P. A. T., et al. Standalone AI for Breast Cancer Detection at Screening Digital Mammography and Digital Breast Tomosynthesis: A Systematic Review and Meta-Analysis // *Radiology*. – 2023. – Vol. 307, № 5. – P. e222639. – DOI: 10.1148/radiol.222639.

253. Zhang X., Yang Y., Shen Y.-W., et al. Diagnostic accuracy and potential covariates of artificial intelligence for diagnosing orthopedic fractures: a systematic literature review and meta-analysis // *European Radiology*. – 2022. – Vol. 32, № 10. – P. 7196–7216. – DOI: 10.1007/s00330-022-08956-4.

254. Zheng Zh., He B., Hu Y., et al. Diagnostic Accuracy of Deep Learning and Radiomics in Lung Cancer Staging: A Systematic Review and Meta-Analysis // *Front. Public Health*. – 2022. – № 10. – P. 938113. – DOI: 10.3389/fpubh.2022.938113.

АКТЫ ВНЕДРЕНИЯ РЕЗУЛЬТАТОВ ДИССЕРТАЦИОННОГО ИССЛЕДОВАНИЯ

ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
УЧРЕЖДЕНИЕ ЗДРАВООХРАНЕНИЯ
ГОРОДА МОСКВЫ
«ГОРОДСКАЯ ПОЛИКЛИНИКА №195
ДЕПАРТАМЕНТА ЗДРАВООХРАНЕНИЯ
ГОРОДА МОСКВЫ»
121614, г. Москва, ул. Крылатские холмы,
д. 51 Телефон/факс +7 (495) 122-02-21



УТВЕРЖДАЮ
ГБУЗ «ГП №195 ДЗМ»
Главный врач:
Лазарева Светлана Ильтеязревна
«25» сентября 2023 г.

А К Т

внедрения в практику результатов диссертационной работы Васильева Юрия Александровича
«**Иновационные диагностические и организационные технологии в рентгенологии**»,
представленной на соискание ученой степени доктора медицинских наук

Комиссия в составе: председателя - главного врача Лазаревой Светланы Ильтеязревной, членов комиссии – исполняющего обязанности заместителя главного врача по медицинской части Когут Оксаны Александровны, заместителя главного врача по клинико-экспертной работе Алехиной Светланы Павловны, заведующего консультативно-диагностическим отделением Лугинина Павла Леонидовича, заведующего отделением лучевой диагностики – врача ультразвуковой диагностики Коваленко Алексея Юрьевича, заведующего терапевтическим отделением Ильиной Ирины Викторовны подтверждает, что результаты диссертационной работы Васильева Юрия Александровича внедрены в практику работы отделения лучевой диагностики ГБУЗ «ГП № 195 ДЗМ» с октября 2022 года.

Результаты работы используются в практической деятельности медицинской организации ГБУЗ «ГП № 195 ДЗМ», отделения лучевой диагностики в виде:

1. Подходов к внедрению и обеспечению безопасности и качества технологий искусственного интеллекта при их прикладном применении.
2. Обоснованных способов использования технологий искусственного интеллекта в работе врачей-рентгенологов.
3. Подхода к профилактике гиподиагностики в работе врача-рентгенолога.

Внедрение результатов диссертационного исследования положительно сказалось на скорости и качестве работы врачей отделений лучевой диагностики. С позиции организации здравоохранения – позволило подойти к проблеме автоматизации работы диагностических служб системно и обосновано, избежав при этом ряда затруднений и барьеров.

С практической точки зрения результаты диссертационной работы использованы при проведении лучевых исследований 55 244 пациентов.

Председатель комиссии:

Главный врач _____ Лазарева Светлана Ильтеязревна

Члены комиссии:

Заместитель главного врача

по КЭР _____ Алехина С.П.

Заведующий филиалом №3

ГБУЗ «ГП №195 ДЗМ» _____ Когут О.А.

Заведующий КДО _____ Лугинин П.Л.

Заведующий отделением лучевой

диагностики-врач УЗИ _____ Коваленко А.Ю.

Заведующий терапевтическим отделением _____ Ильина И.В.

ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
УЧРЕЖДЕНИЕ ЗДРАВООХРАНЕНИЯ
ГОРОДА МОСКВЫ
«ГОРОДСКАЯ ПОЛИКЛИНИКА №220
ДЕПАРТАМЕНТА ЗДРАВООХРАНЕНИЯ
ГОРОДА МОСКВЫ»
123022, г. Москва, ул. Загоренова, д.27
Телефон/факс +7 (495) 122-02-21

УТВЕРЖДАЮ
ГБУЗ «ГП №220 ДЗМ»
заместитель главного врача
Карпов Сергей Сергеевич



«01» октября 2023 г.

А К Т

внедрения в практику результатов диссертационной работы Васильева Юрия Александровича
«Инновационные диагностические и организационные технологии в рентгенологии»,
представленной на соискание ученой степени доктора медицинских наук

Комиссия в составе: председателя - заместителя главного врача Карпова Сергея Сергеевича, членов комиссии – врача-методиста Назаркиной Валентины Александровны подтверждает, что результаты диссертационной работы Васильева Юрия Александровича внедрены в практику работы отделения рентгенологической диагностики ГБУЗ «ГП №220 ДЗМ» с декабря 2022 года.

Результаты диссертационного исследования используются в практической деятельности ГБУЗ "ГП № 220 ДЗМ», в том числе – отделения рентгенологической диагностики в виде:

1. Применения систем поддержки принятия врачебных решений на основе технологий искусственного интеллекта для описания результатов компьютерной томографии, рентгенографии и магнитно-резонансной томографии с целью диагностики онкологических, демиелинизирующих, дегенеративно-дистрофических заболеваний.
2. Применения технологий искусственного интеллекта для морфометрии (автоматизированных измерений) диаметров магистральных сосудов и степени деформации позвонков при интерпретации результатов компьютерной томографии органов грудной клетки.

Внедрение результатов диссертационного исследования по научно обоснованной оптимизации процессов внедрения и применению технологий искусственного интеллекта позволило положительно повлиять на качество и доступность лучевых исследований, а также повысило приверженность врачей-рентгенологов к использованию новых технологий ИИ в их практической деятельности.

Результаты диссертационной работы использованы при оказании медицинской помощи 46 498 пациентам.

Председатель комиссии:
заместитель главного врача _____ Карпов С.С.

Члены комиссии:

Врач-методист _____ Назаркина В.А.

ДЕПАРТАМЕНТ ЗДРАВООХРАНЕНИЯ ГОРОДА МОСКВЫ
Государственное бюджетное учреждение здравоохранения города Москвы
**«Научно-практический клинический центр диагностики
и телемедицинских технологий
Департамента здравоохранения города Москвы»**
(ГБУЗ «НПКЦ ДиТ ДЗМ»)
127051, г. Москва, ул. Петровка, д. 24, стр. 1
e-mail: npcmr@zdrav.mos.ru Тел.: +7 (495) 276 04 36
ОГРН 1037739481229 ИНН 7709064286 КПП 770701001

АКТ

внедрения результатов диссертационной работы Ю. А. Васильева на тему
«Инновационные диагностические и организационные технологии в
рентгенологии» на соискание ученой степени д.м.н. по специальностям
3.1.25. Лучевая диагностика; 3.3.9. Медицинская информатика в учебный
процесс Учебного Центра ГБУЗ «Научно-практический клинический центр
диагностики и телемедицины Департамента здравоохранения города
Москвы», 127051, г. Москва, ул. Петровка, д. 24, стр. 1.

В работе Васильева Ю. А. впервые показана возможность обеспечения
принципиального повышения производительности и качества лучевой
диагностики за счет научного обоснования концепции применения технологий
искусственного интеллекта.

Высокая практическая значимость работы состоит:

- в анализе и систематизации результатов научных исследований в
области применения технологий искусственного интеллекта в лучевой
диагностике;
- изучении диагностической точности технологий искусственного
интеллекта при работе в условиях реальной клинической практики;

- оценке клинической значимости поддержки принятия решений врачом-рентгенологом с применением технологий искусственного интеллекта;
- обоснования способов автоматического анализа результатов лучевых исследований в контексте решения конкретных клинических задач;
- обосновании концепции и возможности применения автономных технологий искусственного интеллекта в лучевой диагностике и его безопасного применения.

Практические рекомендации по применению технологий искусственного интеллекта в рентгенологии с 2022 года внедрены в учебный процесс Учебного центра ГБУЗ «НПКЦ ДиТ ДЗМ» в виде лекционного материала на курсах дополнительного профессионального образования, общедоступных вебинарах.

Заведующий Учебным Центром
ГБУЗ НПКЦ ДиТ ДЗМ,
к.м.н.

А. В. Бажин

Подпись
А. В. Бажина заверяю

Начальник отдела кадров
ГБУЗ НПКЦ ДиТ ДЗМ



О. А. Даниэль

