

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ
ДОПОЛНИТЕЛЬНОГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ МЕДИКО-СОЦИАЛЬНЫЙ ИНСТИТУТ ИМЕНИ Ф.П. ГААЗА»

На правах рукописи



Корабельников Даниил Иванович

**Дифференциальная диагностика новообразований кожи
с использованием технологий искусственного интеллекта**

3.1.23. Дерматовенерология

3.1.6. Онкология, лучевая терапия

Диссертация
на соискание ученой степени
доктора медицинских наук

Научные консультанты:
доктор медицинских наук, профессор,
член-корреспондент РАН
Олисова Ольга Юрьевна
доктор медицинских наук, профессор
Ламоткин Игорь Анатольевич

Москва – 2026

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	10
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ	32
1.1. Эпидемиология злокачественных новообразований кожи.....	32
1.2. Клинические и морфологические особенности основных форм новообразований кожи.....	34
1.3. Современные подходы и трудности диагностики новообразований кожи.....	37
1.3.1. Роль врачей первичного звена в ранней диагностике злокачественных новообразований кожи.....	37
1.3.2. Дерматоскопия в дифференциальной диагностике новообразований кожи	38
1.3.3. Динамическое наблюдение пациентов групп высокого риска.....	39
1.3.4. Проблема гипо- и гипердиагностики новообразований кожи.....	40
1.4. Программы компьютерного зрения на основе искусственного интеллекта в дерматовенерологии.....	41
1.4.1. Методы машинного и глубокого обучения в анализе фотоизображений кожи	41
1.4.2. Диагностическая точность программ искусственного интеллекта в дерматоонкологии	42
1.4.3. Мобильные приложения на основе искусственного интеллекта и системы поддержки принятия решений	43
1.4.4. Ограничения и причины ошибок программ искусственного интеллекта.....	45
1.5. Наборы данных для обучения моделей искусственного интеллекта в дерматовенерологии и дерматоонкологии	46
1.6. Отношение врачей к технологиям искусственного интеллекта: обзор зарубежных исследований.....	49
1.6.1. Отношение дерматовенерологов к технологиям искусственного интеллекта.....	49

1.6.2. Отношение онкологов и врачей первичного звена здравоохранения к технологиям искусственного интеллекта	50
1.6.3. Барьеры внедрения и факторы доверия	51
1.7. Методологические подходы к оценке точности и медико-экономической эффективности программ искусственного интеллекта	51
1.7.1. Метрики диагностической точности и стандарты отчетности.....	52
1.7.2. Клиническая точность и учет онкологической настороженности.....	53
1.7.3. Медико-экономическая эффективность и оценка технологий здравоохранения.....	53
1.8. Применение технологий искусственного интеллекта в медицинском образовании и подготовке врачей-дерматовенерологов.....	55
1.8.1. Направления применения искусственного интеллекта в медицинском образовании.....	55
1.8.2. Обучение врачей-дерматовенерологов, ординаторов и студентов с применением технологий искусственного интеллекта	57
ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ	59
2.1. Общий дизайн диссертационного исследования	59
2.2. Дизайн и методология опросных исследований	60
2.2.1. Общая характеристика опросных исследований	60
2.2.2. Инструмент исследования.....	60
2.2.3. Опрос врачей-терапевтов и врачей общей практики	61
2.2.4. Опрос врачей-дерматовенерологов	61
2.2.5. Опрос врачей-онкологов	62
2.2.6. Статистические методы анализа данных опросов.....	63
2.3. Характеристики программы «Derma Check»	63
2.4. Описание модуля непрерывного профессионального развития «Derma Check Educational Module»	65
2.5. Формирование набора данных фотоизображений.....	67
2.6. Методология анализа причин ошибочной классификации	68
2.6.1. Фотометрические и текстурные метрики качества изображений.....	69

2.6.2. Статистический анализ межгрупповых различий	69
2.6.3. Методы объяснимого искусственного интеллекта	70
2.7. Методология оценки клинической точности	71
2.8. Методология оценки эффективности диагностики с использованием алгоритмов на основании заключений моделей искусственного интеллекта	72
2.9. Методология оценки медико-экономической эффективности диагностики.....	74
2.10. Статистические методы	75
ГЛАВА 3. РЕЗУЛЬТАТЫ СОБСТВЕННЫХ ИССЛЕДОВАНИЙ.....	78
3.1. Исследование уровня доверия к технологиям искусственного интеллекта и готовности его применять российскими врачами	78
3.1.1. Описание онлайн-опроса.....	78
3.1.2. Исследование уровня доверия к технологиям искусственного интеллекта и готовности его применять врачами общей практики и терапевтами.....	79
3.1.3. Исследование уровня доверия к технологиям искусственного интеллекта и готовности его применять врачами-дерматовенерологами.....	93
3.1.4. Исследование уровня доверия к технологиям искусственного интеллекта и готовности его применять врачами-онкологами	106
3.1.5. Сравнительный анализ отношения к технологиям искусственного интеллекта врачей разных специальностей.....	119
3.2. Анализ возможностей существующих программ для электронных вычислительных машин на основе моделей искусственного интеллекта для диагностики заболеваний кожи	129
3.2.1. Методика поиска и отбора публикаций.....	129
3.2.2. Методика сравнительного анализа.....	132
3.2.3. Результат поиска и отбора программ для электронных вычислительных машин, основанных на технологиях искусственного интеллекта, для применения в дерматовенерологии.....	132

3.2.4. Характеристики анализируемых программ для электронных вычислительных машин, основанных на технологиях искусственного интеллекта, для применения в дерматовенерологии.....	132
3.2.5. Сравнение программ для электронных вычислительных машин, основанных на технологиях искусственного интеллекта, для применения в дерматовенерологии по точности и эффективности	137
3.2.6. Области применения, архитектурные особенности и общие ограничения анализируемых программ для электронных вычислительных машин, основанных на технологиях искусственного интеллекта, для применения в дерматовенерологии.....	140
3.2.7. Перспективы развития программ искусственного интеллекта в дерматовенерологии.....	142
3.3. Анализ медицинских изделий с искусственным интеллектом, зарегистрированных в Российской Федерации.....	144
3.4. Сравнительный анализ баз данных и наборов данных, используемых для обучения моделей искусственного интеллекта для программ для диагностики заболеваний кожи.....	154
3.5. Разработка и исследование эффективности методики формирования из баз данных наборов данных для обучения моделей искусственного интеллекта для использования в клинической медицине.....	161
3.5.1. Разграничение понятий «база данных» и «набор данных»	162
3.5.2. Методология формирования наборов данных	164
3.5.3. Валидация разработанной шестиступенчатой методики формирования наборов данных	168
3.6. Разработка классификации программ для электронных вычислительных машин, применяемых в клинической медицине и здравоохранении	170
3.7. Исследование проблем диагностики новообразований кожи с применением программ искусственного интеллекта, разработка и апробация решений проблем диагностики.....	183
3.7.1. Систематизация основных проблем.....	183

3.7.2. Трудности дифференциальной диагностики доброкачественных и злокачественных новообразований кожи	186
3.7.3. Сложности дифференциальной диагностики доброкачественных и злокачественных новообразований кожи	193
3.8. Анализ причин ошибочной классификации новообразований кожи программами компьютерного зрения на основе технологий искусственного интеллекта	196
3.8.1. Материал и методы	196
3.8.2. Анализ причин ошибочной классификации меланоцитарных опухолей....	198
3.8.3. Анализ причин ошибочной классификации немеланоцитарных опухолей.....	214
3.9. Анализ способов решения проблем диагностики новообразований кожи с помощью программ искусственного интеллекта на примере программы «Derma Check»	228
3.9.1. Программа «Derma Check».....	228
3.9.2. Дизайн исследования и методология анализа.....	229
3.9.3. Выявленные проблемы диагностики новообразований кожи при применении программ искусственного интеллекта и их решения	230
3.10. Разработка и апробация методологии оценки клинической точности и клинической эффективности диагностики новообразований кожи программами искусственного интеллекта	245
3.10.1. Клиническая точность программ искусственного интеллекта с учетом онкологической настороженности	245
3.10.2. Материал и программа искусственного интеллекта	247
3.10.3. Первый уровень оценки. Формальная точность и ее систематическая ошибка	248
3.10.4. Второй уровень оценки: клиническая эффективность как качественная мера согласованности	250
3.10.5. Третий уровень оценки. Клиническая точность как количественная мера обоснованности	255

3.10.6. Сопоставление трех уровней оценки	258
3.10.7. Надежность эталона. Межэкспертная согласованность	258
3.10.8. Значение для клинической практики и для обучения моделей.....	260
3.11. Разработка способа оценки и анализ эффективности диагностики новообразований кожи с использованием алгоритмов на основании заключений моделей искусственного интеллекта	262
3.11.1. Материал и моделируемые сценарии маршрутизации	263
3.11.2. Принцип количественной оценки: две меры относительной эффективности.....	265
3.11.3. Комплексная визуализация. Квадрантная матрица эффективности.....	272
3.11.4. Программная реализация.....	275
3.11.5. Значение результатов и переход к экономической оценке.....	279
3.12. Разработка способа оценки и анализ медико-экономической эффективности диагностики с применением алгоритмов маршрутизации пациентов на основании заключений моделей искусственного интеллекта	280
3.12.1. Медико-экономическая эффективность диагностики с применением алгоритмов маршрутизации на основании заключений моделей искусственного интеллекта на уровне медицинской организации.....	280
3.12.2. Медико-экономическая эффективность диагностики на региональном уровне	306
3.13. Анализ эволюции программ семейства «Derma Check» для автоматизированной диагностики новообразований кожи и дерматозов и эволюция диагностического подхода.....	323
3.14. Анализ характеристик программы «Derma Check» для автоматизированной диагностики новообразований кожи и дерматозов.....	335
3.14.1. Описание программы «Derma Check».....	335
3.14.2. Конвейер обработки фотоизображения	337
3.14.3. Модули бинарной классификации и алгоритм консенсуса	338
3.14.4. Модуль анализа анамнестических данных.....	340
3.14.5. Сегментация образования	341

3.14.6. Анализ асимметрии методом главных компонент	342
3.14.7. Модуль объяснимого искусственного интеллекта (XAI)	343
3.14.8. Модуль многоклассовой классификации	345
3.14.9. Картотека пациентов.....	346
3.14.10. Приемы (осмотры) и сохранение результатов	347
3.14.11. Подсистема динамического наблюдения	349
3.14.12. Алгоритмы персонализированной маршрутизации	350
3.15. Исследование эффективности диагностики новообразований кожи с применением программы «Derma Check»	351
3.15.1. Материал и методы	351
3.15.2. Формирование заключения об уровне риска	354
3.15.3. Распределение вероятности и статистическая модель	355
3.15.4. Диагностика по группам новообразований	359
3.15.4.1. Злокачественные меланоцитарные опухоли	359
3.15.4.2. Злокачественные кератиноцитарные/эпидермальные опухоли	360
3.15.4.3. Доброкачественные меланоцитарные опухоли.....	360
3.15.4.4. Доброкачественные кератиноцитарные/эпидермальные опухоли	361
3.15.4.5. Прочие доброкачественные образования	362
3.15.5. Сравнение точности заключений программы с точностью заключений врачей-дерматовенерологов и врачей-онкологов	363
3.15.6. Формальная точность программы	365
3.15.7. Клиническая точность программы (с учетом онконастороженности)	367
3.15.8. Сопоставление заключений программы с клиническим эталоном.....	369
3.15.9. Многоклассовая классификация по классам модели (top-1/3/5).....	372
3.16. Исследование диагностической эффективности программы «Derma Check» в зависимости от вида фотоизображений (дерматоскопических и клинических) новообразований кожи	376
3.16.1. Общие показатели диагностической эффективности.....	377
3.16.2. Анализ распределения вероятностей	384
3.16.3. Согласованность предсказаний между модальностями.....	388

3.16.4. Анализ вклада компонентов ансамбля программы	392
3.17. Анализ характеристик программного модуля непрерывного профессионального развития «Derma Check Educational Module»	395
3.17.1. Общая характеристика и архитектура модуля	395
3.17.2. Компонент «Тест по нозологиям»	397
3.17.3. Компонент «Клинические задачи»	401
3.17.4. Компонент «Атлас нозологий»	403
3.17.5. Интеграция модуля непрерывного профессионального развития с диагностическим ядром	405
3.17.6. Методические принципы модуля непрерывного профессионального развития	406
3.18. Исследование эффективности обучения диагностике новообразований кожи и дерматозов с применением модуля «Derma check educational module» ...	406
ГЛАВА 4. ОБСУЖДЕНИЕ	410
ЗАКЛЮЧЕНИЕ	428
ВЫВОДЫ	438
ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ	442
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	446
СПИСОК ЛИТЕРАТУРЫ	450
ПРИЛОЖЕНИЕ А	500

ВВЕДЕНИЕ

Актуальность темы исследования

В последние десятилетия во всем мире регистрируется устойчивый рост заболеваемости злокачественными новообразованиями (ЗНО) кожи, что формирует значительную и постоянно возрастающую нагрузку на системы здравоохранения [235, 279, 293]. Кожа является органом, наиболее часто поражаемым злокачественными опухолями у человека: по данным GLOBOCAN – международной базы данных и проекта Международного агентства по изучению рака (IARC), входящего в состав Всемирной организации здравоохранения (ВОЗ), рак кожи наиболее распространенная группа онкологических заболеваний в мире. В популяциях европеоидной расы меланома и немеланомные раки кожи являются наиболее частыми злокачественными новообразованиями [207, 234, 267]. В Российской Федерации новообразования кожи традиционно занимают одно из ведущих мест в структуре онкологической заболеваемости, при этом сохраняется значимая доля запущенных случаев меланомы, диагностируемых на поздних стадиях [99, 100, 101].

Основную часть первичного потока пациентов с новообразованиями кожи принимают на себя врачи общей практики (ВОП) и врачи-терапевты первичного звена здравоохранения, которые сталкиваются с растущим числом обращений по поводу подозрительных новообразований кожного покрова [180, 291]. При этом дифференциальная диагностика доброкачественных и злокачественных новообразований кожи для врача первичного звена представляет существенную трудность, что приводит либо к необоснованным направлениям пациентов с доброкачественными опухолями к дерматовенерологам и онкологам, либо к позднему выявлению злокачественных поражений [111, 147, 292]. Указанные обстоятельства обуславливают необходимость совершенствования методов ранней дифференциальной диагностики ЗНО кожи на всех этапах оказания медицинской помощи [287, 327].

Важнейшим инструментом повышения точности диагностики новообразований кожи стала дерматоскопия, применение которой обученными специалистами достоверно увеличивает чувствительность и специфичность выявления меланомы и других ЗНО кожи, сокращает число необоснованных биопсий и снижает затраты здравоохранения [171, 242, 261, 310]. Тем не менее не все специалисты умеют пользоваться дерматоскопом, а некоторые не имеют дерматоскоп вовсе, а число подготовленных врачей-дерматовенерологов, владеющих методом, остается ограниченным, особенно в регионах с дефицитом специализированных кадров [25, 296].

На этом фоне бурное развитие технологий искусственного интеллекта (ИИ) и, прежде всего, методов глубокого обучения на основе сверточных нейронных сетей открыло принципиально новые возможности автоматизированного анализа медицинских изображений [47, 48, 66, 117, 176, 266]. Начиная с основополагающей работы А. Esteva и соавт. (2017), продемонстрировавшей уровень классификации новообразований кожи, сопоставимый с уровнем сертифицированных дерматологов, был опубликован целый ряд исследований, в которых алгоритмы на основе нейронных сетей достигали или превосходили диагностическую точность врачей-специалистов в задачах распознавания меланомы по дерматоскопическим и клиническим фотоизображениям [174, 176, 183, 273]. Программы компьютерного зрения (ПКЗ) на основе ИИ в формате мобильных приложений все шире применяются как для скрининга новообразований кожи пациентами, так и в качестве систем поддержки принятия врачебных решений [57, 131, 142].

Вместе с тем внедрение технологий ИИ в реальную клиническую практику дерматоонкологии сопряжено с рядом нерешенных проблем. К их числу относятся зависимость точности классификации от качества входных фотоизображений и характеристик устройств фотофиксации [29, 249], «черный ящик» нейросетевых моделей, затрудняющий интерпретацию причин ошибочных заключений [219, 300, 313], недостаточная прозрачность и репрезентативность обучающих наборов данных [170, 250, 263], а также систематическое занижение клинической ценности программ ИИ при их оценке по стандартной матрице ошибок без учета

онкологической настороженности [39, 44]. Отдельной, недостаточно изученной в отечественной литературе проблемой остается отношение самих врачей к технологиям ИИ; уровень их осведомленности, доверия и готовности к применению этих технологий во многом определяет успешность внедрения ИИ-систем в практическое здравоохранение [116, 141, 158].

Наконец, для обоснования широкого внедрения программ ИИ в дерматоонкологию необходима не только оценка их диагностической точности, но и анализ клинической эффективности и медико-экономической эффективности (МЭЭ) диагностических алгоритмов на основе заключений моделей ИИ, как на уровне отдельного медицинского учреждения, так и на уровне региона [43, 44, 46, 203, 205]. Комплексное решение перечисленных проблем определяет актуальность настоящего диссертационного исследования.

Степень разработанности темы исследования

Пионерские исследовательские работы начала 2010-х годов заложили концептуальный фундамент, когда сверточные нейронные сети впервые продемонстрировали способность классифицировать дерматоскопические изображения на уровне дерматологов-экспертов.

Проблематике применения технологий ИИ в диагностике новообразований кожи посвящено значительное число зарубежных научных работ, а также отечественные, включая систематические обзоры и метаанализы диагностической точности нейросетевых алгоритмов и обзоры технологий ИИ [32, 118, 121, 323, 324], исследования по формированию и валидации наборов данных дерматоскопических и клинических фотоизображений новообразований [41, 113, 354, 361], а также работы, посвященные методологии оценки исследований в области медицинского ИИ [305, 306, 351]. В отечественной литературе вопросы применения ИИ в дерматовенерологии и онкологии, оценки его клинической и экономической эффективности, а также оценку внедрения технологий ИИ в целом, также затрагиваются, однако научные работы немногочисленны [6, 23, 34, 49].

До начала настоящего исследования в отечественной литературе отсутствовали: (1) масштабные опросные исследования врачей трех ключевых для выявления ЗНО специальностей (терапевтов (врачи общей практики), дерматовенерологов и онкологов) с единой строгой статистической методологией; (2) систематизированные основные проблемы в диагностике злокачественных новообразований кожи с помощью программ ИИ; (3) требования и описания «золотого стандарта» характеристик и функционала программ ИИ для автоматизированной дифференциальной диагностики патологии кожи; (4) методологически обоснованные способы оценки клинической точности ПКЗ на основе технологий ИИ, учитывающие онкологическую настороженность как самостоятельный клинический параметр; (5) воспроизводимые способы оценки клинической точности диагностики программами ИИ в дерматоонкологии, учитывающие онкологическую настороженность как самостоятельную клиническую ценность; (6) воспроизводимые способы оценки МЭЭ диагностики с применением алгоритмов маршрутизации на основании заключений ИИ применительно к российской системе здравоохранения. Отсутствуют также (7) основанные на технологиях ИИ программы обучения визуальной диагностике и клиническим подходам к диагностике дерматозов и новообразований кожи и повышении компетенций в дерматоонкологии. Разработка вышеперечисленных научных направлений составляет ключевое содержание настоящего исследования.

Цель и задачи исследования

Разработать и научно обосновать комплексный подход к дифференциальной диагностике новообразований кожи с использованием технологий искусственного интеллекта с оценкой клинической и медико-экономической эффективности и повышением компетенций врачей в области дерматоонкологии.

Задачи исследования:

1. Изучить отношение врачей первичного звена здравоохранения (терапевтов и врачей общей врачебной практики), врачей-дерматовенерологов, врачей-онкологов к технологиям искусственного интеллекта: оценить уровень осведомленности, частоту применения, доверие к заключениям искусственного интеллекта, субъективное восприятие точности диагностических систем искусственного интеллекта; сравнить восприятие технологий искусственного интеллекта у обучающихся (студентов младших (1–3 курсов), старших (4–6 курсов), ординаторов), практикующих врачей-специалистов, преподавателей и научных работников при контроле размера групп, провести кластерный анализ по типу отношения к технологиям искусственного интеллекта.

2. Провести обзор и дать сравнительную характеристику баз данных и наборов данных, применяемых в моделях искусственного интеллекта для дифференциальной диагностики дерматозов и новообразований кожи; разработать методику формирования датасетов из клинических баз данных.

3. Провести аналитический обзор и дать сравнительную характеристику доступных программ для ЭВМ на основе моделей искусственного интеллекта и медицинских изделий с искусственным интеллектом, применяемых в дерматовенерологии, других областях клинической медицины и здравоохранении России. Разработать классификацию программ ЭВМ, в том числе основанных на технологиях искусственного интеллекта и программ компьютерного зрения.

4. Проанализировать и систематизировать основные проблемы, возникающие при диагностике новообразований кожи в условиях реальной клинической практики, предложить и обосновать пути их решения.

5. Исследовать влияние фотометрических и текстурных характеристик фотоизображений новообразований кожи на точность классификации программ компьютерного зрения на основе моделей искусственного интеллекта; разработать обобщенную методологию анализа причин ошибочной классификации новообразований кожи с применением методов объяснимого искусственного интеллекта.

6. Исследовать клиническую точность и клиническую эффективность диагностики новообразований кожи программами для ЭВМ на основе технологий искусственного интеллекта с применением оригинальной методологии, учитывающей онкологическую настороженность как самостоятельную клиническую ценность.

7. Разработать, валидировать и апробировать способ оценки клинической эффективности диагностики новообразований кожи с использованием алгоритмов на основании заключений моделей искусственного интеллекта. Создать, валидировать и апробировать программу для ЭВМ.

8. Разработать, валидировать и апробировать способ оценки медико-экономической эффективности диагностики новообразований кожи с применением алгоритмов маршрутизации пациентов на основании заключений моделей искусственного интеллекта на уровне медицинской организации, сети медицинских организаций и региона. Разработать, валидировать и апробировать программу для ЭВМ оценки медико-экономической эффективности диагностики новообразований кожи с применением алгоритмов маршрутизации пациентов на основании заключений моделей искусственного интеллекта.

9. Разработать, валидировать и апробировать программу для ЭВМ – систему поддержки принятия врачебных решений для автоматизированной визуальной идентификации и дифференциальной диагностики дерматозов и новообразований кожи с модулями анализа фотоизображений, анамнеза, объяснимого искусственного интеллекта, с персонализированными алгоритмами маршрутизации, личным кабинетом врача и функцией динамического наблюдения образований; исследовать диагностическую точность, чувствительность и специфичность диагностики новообразований кожи и дерматозов с применением программы для ЭВМ.

10. Разработать, валидировать и апробировать программный модуль на основе технологий искусственного интеллекта с возможностью формирования тестов и задач; автоматизированной проверки ответов; с банком тестов, клинических задач и справочных материалов для интерактивного обучения

диагностике и клиническим подходам к диагностике новообразований кожи и дерматозов и повышении компетенций в дерматовенерологии и дерматоонкологии.

Научная новизна

Впервые на репрезентативном материале выполнен комплексный анализ отношения врачей трех специальностей (терапевтов и врачей общей практики, дерматовенерологов, онкологов) к технологиям ИИ с выделением ключевых предикторов доверия и готовности к их применению.

Разработана шестиэтапная методология формирования датасетов из клинических баз данных для обучения моделей ИИ для диагностики дерматозов и новообразований кожи.

Разработана оригинальная классификация программ для ЭВМ, включая основанные на технологиях ИИ и ПКЗ, применяемых в клинической дерматовенерологии, других областях клинической медицины и здравоохранении, построенная на девятнадцати признаках, объединенных в пять тематических блоков.

Проанализированы, систематизированы основные проблемы, возникающие при диагностике злокачественных новообразований кожи с помощью программ ИИ и предложены решения этих проблем.

Разработана и валидирована оригинальная методология анализа причин ошибочной классификации фотоизображений новообразований кожи программами компьютерного зрения на основе технологий ИИ, основанная на вычислении 13 фотометрических и текстурных метрик для каждого фотоизображения, многоуровневом статистическом анализе и методах объяснимого ИИ (методы Score-CAM, Occlusion Sensitivity).

Разработаны, валидированы и апробированы оригинальная методология оценки клинической точности диагностики программами ИИ в дерматоонкологии, учитывающая онкологическую настороженность как самостоятельную

клиническую ценность, включающая систему четырех категорий клинической эффективности («полное совпадение», «совпадение по онкологической настороженности», «несовпадение по онкологической настороженности», «полное несовпадение») и формулы расчета клинической чувствительности, клинической специфичности и клинической точности диагностики программами ИИ.

Разработан и валидирован способ оценки клинической эффективности диагностики новообразований кожи с использованием алгоритмов на основании заключений моделей искусственного интеллекта. Создана, апробирована и внедрена программа для ЭВМ для оценки клинической эффективности диагностики.

Разработан и валидирован способ оценки медико-экономической эффективности диагностики новообразований кожи с применением заключений моделей ИИ на уровне медицинской организации (с квадрантной матрицей предотвратимых и потенциально предотвратимых финансовых затрат), сети медицинских организаций и региона РФ. Создана, апробирована и внедрена программа для ЭВМ для автоматизации медико-экономической эффективности диагностики.

Разработана, апробирована и внедрена в лечебно-диагностический процесс программа для ЭВМ – система поддержки принятия врачебных решений (СППВР) на основе технологий ИИ «Derma Check» для автоматизированной визуальной идентификации и дифференциальной диагностики дерматозов и новообразований кожи с персонализированными алгоритмами маршрутизации, личным кабинетом врача и функцией динамического наблюдения образований.

Разработан, апробирован и внедрен в образовательный процесс программный модуль непрерывного профессионального развития на основе технологий ИИ «Derma Check Educational Module» для интерактивного обучения визуальной диагностике и клиническим подходам к диагностике дерматозов и новообразований кожи и повышении компетенций в дерматовенерологии и дерматоонкологии с возможностью формирования тестов и задач на основании текстовых запросов и медицинских изображений; автоматизированной проверки

ответов; с банком медицинских изображений, тестов, клинических задач и справочников: руководств, атласов, учебников, другой учебной и справочной литературы, нормативных правовых актов и клинических рекомендаций.

Теоретическая и практическая значимость работы

Результаты исследования углубляют представления о факторах, определяющих восприятие технологий ИИ врачами и готовность применять технологии ИИ в медицинской деятельности; о формировании наборов данных из клинических баз данных для программ компьютерного зрения, применяемых в клинической медицине и здравоохранении; стандартизированной оценки и описания программ для ЭВМ, в том числе на основе технологий ИИ и ПКЗ; медицинских изделий с ИИ, зарегистрированных в РФ; об основных проблемах, возникающих при диагностике новообразований кожи и дерматозов с помощью программ ИИ в условиях реальной клинической практики; причинах диагностических ошибок ПКЗ вследствие ошибочной классификации фотоизображений новообразований; оценке клинической точности диагностики программами ИИ в дерматоонкологии с учетом онкологической настороженности; оценке клинической эффективности диагностики с использованием алгоритмов на основании заключений моделей ИИ с применением программы для ЭВМ; оценке медико-экономической эффективности диагностики с применением алгоритмов маршрутизации пациентов на основании заключений моделей ИИ; возможностях и ограничениях применения программ для ЭВМ, основанных на технологиях ИИ, для автоматизированной визуальной идентификации и дифференциальной диагностики новообразований кожи и дерматозов; возможностях применения программ, основанных на технологиях ИИ, в обучении визуальной диагностике и клиническим подходам к диагностике новообразований кожи и дерматозов и повышении компетенций врачей в дерматоонкологии.

Впервые охарактеризованы закономерности восприятия технологий ИИ обучающимися по программам высшего образования (специалитет, ординатура) и

врачами трех ключевых для выявления ЗНО кожи специальностей, что является важным вкладом в социологию медицины.

Полученные результаты опроса об отношении обучающихся и врачей-специалистов к ИИ служат основой для разработки образовательных программ, направленных на формирование у специалистов осознанного и обоснованного подхода к применению ИИ-технологий.

Разработанная шестиэтапная методология формирования наборов данных может применяться при подготовке наборов данных (датасетов) из клинических баз данных для программ компьютерного зрения в клинической медицине.

Формализовано разграничение понятий «набор данных» (датасет) и «база данных» применительно к медицинским ИИ-системам, что устраняет существующую терминологическую неопределенность в данной области.

Разработанная оригинальная классификация программ для ЭВМ, включая основанные на технологиях ИИ и ПКЗ, применяемых в дерматовенерологии, других областях клинической медицины и здравоохранении, построенная на девятнадцати признаках, объединенных в пять тематических блоков, обеспечивает стандартизированное, понятное и полное описание систем, необходимое для их корректного сравнения и регуляторной оценки.

Разработан и валидирован способ анализа причин ошибочной классификации фотоизображений новообразований кожи с применением методов объяснимого ИИ, который позволяет выявить причины ошибочной классификации новообразований кожи для внесения изменений в модели и программы для диагностики патологий кожи на основании технологий ИИ с последующим повышением точности и эффективности диагностики новообразований кожи.

Разработаны и валидированы показатели клинической точности и клинической эффективности диагностики программами ИИ, учитывающие онкологическую настороженность как самостоятельную клиническую ценность, которые устраняют существующие недостатки оценки точности диагностики так называемых «пограничных новообразований кожи», которые занимают

промежуточное положение по степени риска злокачественности, не являются абсолютно доброкачественными, но и не достигают степени злокачественности.

Разработаны и валидированы методологические инструменты оценки клинической и медико-экономической эффективности, которые могут применяться при клинических испытаниях медицинских изделий с ИИ, при формировании регуляторных требований и стандартов, при сравнительных исследованиях диагностических стратегий, могут быть адаптированы для других областей медицинской визуализации.

Разработана, апробирована, внедрена программа для ЭВМ – система поддержки принятия врачебных решений на основе технологий ИИ «Derma check» для автоматизированной визуальной идентификации и дифференциальной диагностики дерматозов и новообразований кожи с персонализированными алгоритмами маршрутизации, кабинетом врача и функцией динамического наблюдения образований.

Разработан, апробирован, внедрен и применяется в образовательной деятельности программный модуль непрерывного профессионального развития на основе технологий ИИ «Derma Check Educational Module» для интерактивного обучения визуальной диагностике и клиническим подходам к диагностике дерматозов и новообразований кожи и повышении компетенций в дерматоонкологии.

Методология и методы исследования

Диссертационное исследование выполнено в АНО ДПО «Московский медико-социальный институт им. Ф.П. Гааза» (г. Москва). Клинический материал для оценки клинической точности и МЭЭ программ ИИ получен на базе ФГБУ «Главный военный клинический госпиталь им. академика Н.Н. Бурденко» Министерства обороны Российской Федерации и филиалов в период 2024–2026 гг. Опросные исследования проводились в марте–июле 2026 г. на анонимной онлайн-платформе Anketolog.ru.

В работе применен комплекс методов исследования, включающий:

1. Систематический анализ отечественной и зарубежной научной литературы, нормативно-правовой документации и патентной базы в области применения ИИ в дерматологии, онкологии и здравоохранении.
2. Анонимный онлайн-опрос (анкета из 26 закрытых вопросов, объединенных в два тематических блока: общие сведения о применении ИИ и оценка отношения к ИИ).
3. Ретроспективный анализ фотоизображений новообразований кожи с гистологической и дерматоскопической верификацией диагнозов.
4. Методы объяснимого ИИ: взвешенное картирование активации классов (англ. Score-CAM) и анализ чувствительности к окклюзии (англ. Occlusion Sensitivity) – для визуализации областей фотоизображений, определяющих классификационное решение нейронной сети.
5. Математическое моделирование сценариев диагностической маршрутизации (с применением заключений программ ИИ и без них) – для оценки МЭЭ.
6. Статистические методы: критерии Манна–Уитни (U-тест), Краскела–Уоллиса (H-тест), Уэлча (t-тест для неравных дисперсий); коэффициент ранговой корреляции Спирмена (ρ); однофакторный (ANOVA), двухфакторный и многомерный (MANOVA) дисперсионный анализ; множественная логистическая регрессия; кластерный анализ методом порогового агрегирования; метод Уилсона для расчета 95% доверительных интервалов долей.

Личный вклад автора

Автором лично сформулированы цель, задачи и разработан дизайн диссертационного исследования; выполнен анализ отечественной и зарубежной литературы по теме работы; разработаны анкеты для опросных исследований; организован и проведен анонимный онлайн-опрос врачей; осуществлен проспективный и ретроспективный анализ клинических случаев с

фотоизображениями новообразований кожи и статистическая обработка всего массива данных; разработаны, валидированы и апробированы оригинальные методологические инструменты оценки клинической точности и МЭЭ диагностических систем ИИ; написаны научные статьи по теме исследования, подготовлены текст диссертации и автореферата.

Автор организовал проведение опроса и участвовал в сборе материала. Автор организовал и участвовал в разработке программ для ЭВМ и формировании баз данных. Автор организовал и участвовал в сборе клинического материала на клинических базах.

Автор подготовил статьи по теме исследования, опубликованные в научных изданиях, подготовил текст диссертации и автореферата, а также принимал личное участие в апробации результатов исследования, выступал с докладами на российских и международных конференциях.

Положения, выносимые на защиту

1. Среди врачей-терапевтов, дерматовенерологов и онкологов преобладает умеренный дефицит доверия к технологиям искусственного интеллекта: 64,8–68,7% специалистов оценивают его как незначительное или полностью отсутствующее. Единственным достоверным предиктором позитивного отношения к технологиям искусственного интеллекта является реальный практический опыт взаимодействия с технологиями искусственного интеллекта в профессиональной деятельности. Различия в цифровой активности обучающихся (ординаторов) и практикующих врачей-дерматовенерологов не определяют достоверных различий в профессиональном доверии к технологиям искусственного интеллекта (все $p > 0,05$).

2. Разработанная шестиэтапная методика формирования наборов данных для применения в моделях искусственного интеллекта для дифференциальной диагностики заболеваний кожи при валидации демонстрирует превосходство: чувствительность, специфичность и точность диагностики модели, обученной на

датасете, подготовленном по предлагаемой методике, составляет 96,7% (94,5–98,1), 96,0% (93,3–97,6) и 96,2% (93,8–97,7) против метрик эффективности диагностики модели, обученной без применения предлагаемой методики: 89,2% (82,4–93,5), 87,5% (83,2–90,8) и 88,0% (84,2–91,0), коэффициент согласия Коэна 0,906 против 0,732 соответственно (статистическая значимость различий подтверждена тестом Макнемара $\chi^2=12,4$; $p<0,001$).

3. В государственном реестре медицинских изделий и организаций (индивидуальных предпринимателей), осуществляющих производство и изготовление медицинских изделий, зарегистрировано 1 медицинское изделие на основе технологий искусственного интеллекта для применения в дерматовенерологии – детекции и классификации пяти нозологических единиц новообразований кожи: меланомы, базальноклеточный рак кожи, меланоцитарный невус, себорейный кератоз, гемангиома с чувствительностью не менее 85%, специфичностью не менее 80%, точностью 81,5%.

Предложенная классификация программ для ЭВМ, в том числе на основе технологий искусственного интеллекта и программ компьютерного зрения, применяемых в дерматовенерологии, других областях клинической медицины и здравоохранении, точно, всесторонне описывает и систематизирует программы для ЭВМ, в том числе на основе технологий искусственного интеллекта и программ компьютерного зрения, применяемые в клинической медицине и здравоохранении, по девятнадцати признакам, объединенным в пять тематических блоков.

4. Проблемы дифференциальной диагностики новообразований кожи могут быть классифицированы с разделением на объективные трудности и субъективные сложности.

Выявленные в условиях реальной клинической практики и систематизированные шесть групп основных проблем применения программ для ЭВМ на основе технологий искусственного интеллекта для диагностики новообразований кожи: а) дисбаланс обучающих наборов данных с преобладанием дерматоскопических изображений; б) низкое качество клинических фотоизображений (пересвет, размытие, затемнение); в) малый размер образования

в кадре; г) одномодельная архитектура большинства программ и отсутствие онкологической настороженности; д) отсутствие объяснимого искусственного интеллекта в заключениях программ; е) невозможность учета данных анамнеза пациента, имеют технически реализуемые решения: (1) автоматический контроль качества фотоизображений с блокировкой некачественных фотоизображений; (2) контроль размера образования в кадре; (3) мультимодельная архитектура на основе ансамбля нейронных сетей с механизмом голосования; (4) визуализация зон риска и текстовый вывод распознанных признаков для обеспечения объяснимости заключений; (5) интерактивный интерфейс для ввода данных анамнеза и дополнительных клинических признаков; (6) поддержка мультимодального анализа клинических и дерматоскопических изображений.

5. Разработанный и валидированный способ анализа причин ошибочной классификации фотоизображений новообразований кожи с применением методов объяснимого искусственного интеллекта позволяет выявить следующие причины ошибочной классификации новообразований кожи по клиническим и дерматоскопическим фотоизображениям.

Для меланоцитарных опухолей (выборка 485 фотоизображений) ложноположительные ошибки обусловлены повышенной яркостью фотоизображений (медиана 0,740 против 0,638 ед. у истинно положительных) и сниженной энтропией (4,898 против 5,159 ед.), при которых модель искусственного интеллекта утрачивает диагностически значимые текстурные признаки пигментной сети. Ложноотрицательные ошибки связаны со смещением фокуса нейронной сети на нерелевантные области вследствие недостаточного контраста фотоизображения.

Для немеланоцитарных (кератиноцитарных/эпидермальных) опухолей (выборка 151 фотоизображение) выявлены аналогичные закономерности. Ложноположительные ошибки также ассоциированы с повышенной яркостью (медиана 0,7994 против 0,6914 ед. у истинно положительных) и сниженной энтропией (4,6414 против 4,8584 ед.). Методы объяснимого искусственного интеллекта (Score-CAM и Occlusion Sensitivity) показали, что при

ложноотрицательных ошибках модель фокусируется на фоновых текстурах (нормальная кожа) вследствие недостаточного контраста опухоли, а при ложноположительных ошибках переобучается на артефактах (теньях, волосах), ошибочно интерпретируя их как признаки злокачественности. Дообучение моделей с учетом выявленных факторов повысило общую диагностическую точность с 84,3 до 92,4% для программы «Melanoma Check» и с 90,9 до 92,4% для программы «Derma Onko Check».

6. Оценка клинической точности диагностики программами для ЭВМ на основе технологий искусственного интеллекта в дерматоонкологии посредством стандартной матрицы ошибок систематически занижает их клиническую ценность. Применение разработанной методологии с учетом онкологической настороженности на выборке из 342 новообразований кожи показывает, что 12,9% наблюдений, квалифицируемых стандартной матрицей как ложноположительные, относятся к клинически ценной категории «совпадение по онкологической настороженности»; клиническая точность при этом составляет 92,1% (95% ДИ 88,8–94,5) против формальной точности 89,7%.

7. Способ оценки клинической эффективности диагностики с использованием алгоритмов на основании заключений моделей искусственного интеллекта выявил, что применение алгоритмов маршрутизации на основании заключений программы «Derma Onko Check» на выборке 381 пациента с новообразованиями кожи обеспечивает снижение доли необоснованных направлений к дерматовенерологам с 40,6% до 6,9%, к онкологам – с 22,0 до 7,6%; коэффициент соотношения финансовых затрат составляет 0,49, что соответствует снижению общих затрат на диагностику меланоцитарных опухолей кожи на 51% по сравнению с традиционной диагностикой без применения искусственного интеллекта.

8. При применении способа оценки медико-экономической эффективности диагностики новообразований кожи с применением алгоритмов маршрутизации пациентов на основании заключений моделей искусственного интеллекта: на уровне учреждения применение технологий искусственного интеллекта снижает

общие затраты на 51%, а на региональном уровне (на примере г. Москвы) обеспечивает потенциальную экономию порядка 436 млн руб. в год при сокращении сроков диагностики на 10,5 суток.

9. Формальная и клиническая точность программы для ЭВМ на основании технологий искусственного интеллекта для автоматизированной визуальной идентификации и дифференциальной диагностики дерматозов и новообразований кожи «Derma Check» составляют: формальная чувствительность, специфичность и точность – 96,7% (94,5–98,1), 84,7% (83,0–86,3), 86,8% (85,4–88,2); клиническая чувствительность, специфичность и точность – 93,3%, 96,2%, 95,3% соответственно. При сопоставлении заключений программы с заключительным клиническим диагнозом по всему массиву клинических случаев: совпадение – 80,8%, совпадение по онконастороженности – 10,4%, несовпадение по онконастороженности – 6,0%, несовпадение – 2,8%. При многоклассовой классификации по укрупненным классам новообразований вероятность топ-1/топ-3/топ-5 верных диагнозов в среднем по массиву составила: для новообразований – 89,4%, 95,4% и 98,7% соответственно; для дерматозов – 82,0%, 90,7% и 95,1% соответственно.

10. Разработанный программный модуль непрерывного профессионального развития на основе технологий искусственного интеллекта «Derma Check Educational Module» обеспечивает статистически значимое повышение как непосредственных, так и отсроченных показателей усвоения знаний и навыков диагностики дерматозов и новообразований кожи.

При сопоставимом исходном уровне знаний (39,0 и 38,0 баллов, $p=0,876$) доля сохранения знаний через 30 дней в группе интерактивного обучения составила 69,9% против 60,3% в контрольной группе ($p=0,008$). Медиана времени усвоения одной нозологической единицы при интерактивном обучении была на 25% меньше (4,8 против 6,4 мин, $p=0,004$).

Соответствие диссертации паспорту научной специальности

Диссертационное исследование соответствует паспорту научной специальности 3.1.23. Дерматовенерология, пункту 4 «Диагностика дерматозов, лимфопролиферативных заболеваний кожи, новообразований кожи и ИППП с использованием клинических, лабораторных, инструментальных и других методов исследования. Разработка диагностических критериев, дифференциальный диагноз дерматозов и ИППП» и пункту 7 «Разработка и совершенствование организационных направлений и методов оказания медицинской помощи пациентам с дерматозами, лимфопролиферативными заболеваниями кожи, новообразованиями кожи и инфекциями, передаваемыми половым путем, включая диспансерное наблюдение и повышение качества жизни» направлений исследований.

Диссертационное исследование соответствует паспорту научной специальности 3.1.6. Онкология, лучевая терапия, пункту 3 «Разработка и совершенствование программ скрининга и ранней диагностики онкологических заболеваний» направлений исследований.

Степень достоверности и апробация результатов

Достоверность диссертационного исследования подтверждена результатами анализа значительного числа клинических наблюдений, гистологической и дерматоскопической верификацией диагнозов в клинической части исследования, точными расчетами, статистическим анализом, оценкой и интерпретацией полученных результатов исследования, обусловлена достаточным размером выборки и подтверждена при проверке первичной документации. Статистическая обработка полученных результатов исследования проведена в соответствии с рекомендуемыми методами статистического анализа медико-биологических исследований. Практические рекомендации и выводы соответствуют цели и задачам диссертационного исследования.

Материалы диссертационного исследования и основные положения доложены и обсуждены на следующих научных конференциях:

- на ежегодной итоговой научно-практической конференции АНО ДПО «Московский медико-социальный института им. Ф.П. Гааза» «Актуальные вопросы медицинской науки и практики» (Гаазовские чтения) (Москва, 29 ноября 2024 г.);

- на 1162 заседании Московского общества дерматовенерологов и косметологов им. А.И. Пospelова (Москва, 15 апреля 2025 г.);

- на XLII Научно-практической конференции с международным участием «Рахмановские чтения». «Дерматовенерология и косметология на стыке науки и практики» (Москва, 16 мая 2025 г.);

- на научно-врачебной конференции кафедры и клиники кожных и венерических болезней им. В.А. Рахманова Института клинической медицины им. Н.В. Склифосовского ФГАОУ ВО Первый МГМУ им. И.М. Сеченова (Сеченовский Университет) (Москва, 17 июня 2025 г.);

- на XXV Всероссийском съезде дерматовенерологов и косметологов (Москва, 16–19 сентября 2025 г.);

- на ежегодной итоговой научно-практической конференции АНО ДПО «Московский медико-социальный института им. Ф.П. Гааза» «Актуальные вопросы медицинской науки и практики» (Гаазовские чтения) (Москва, 28 ноября 2025 г.);

- на X Всероссийской конференции по искусственному интеллекту в здравоохранении и системам поддержки принятия врачебных решений ИТМ ИИ (Москва, 12–13 февраля 2026 г.);

- на IX Петербургском медицинском инновационном форуме (Санкт-Петербург, 13–15 мая 2026 г.)

- на XLIII Научно-практической конференции с международным участием «Рахмановские чтения». «Дерматовенерология и косметология на стыке науки и практики» (Москва, 15 мая 2026 г.);

– на XIV международном конгрессе «ОргЗдрав – 2026» «Здравоохранение и ИИ: грани сотрудничества» (Москва, 22 мая 2026 г.);

– на VI Всероссийской конгрессе-выставке решений для цифрового здравоохранения «ИТМ Петербург» (Санкт-Петербург, 17–20 июня 2026 г.).

А также на совместном заседании кафедры дерматовенерологии и косметологии медицинского факультета и кафедры внутренних болезней с курсами семейной медицины, функциональной диагностики, инфекционных болезней, профессиональных болезней медицинского факультета АНО ДПО «Московский медико-социальный институт им. Ф.П. Гааза» (Москва, 3 июля 2026 г.).

Достоверность и значимость программы для ЭВМ – системы поддержки принятия врачебных решений на основе технологий ИИ «Derma Check» для автоматизированной визуальной идентификации и дифференциальной диагностики патологии кожи с персонализированными алгоритмами маршрутизации, кабинетом врача и функцией динамического наблюдения образований подтверждена дипломом финалиста конкурса проектов в области ИИ для здравоохранения «ИИ-пионеры ждут инвестиций» (Москва, 22 мая 2026 г.).

Апробация диссертационной работы проведена на заседании ученого совета АНО ДПО «Московский медико-социальный института им. Ф.П. Гааза» (протокол № 3 от 24 августа 2026 г.).

Внедрение результатов исследования в практику

Результаты, основные научные положения, выводы и практические рекомендации диссертационного исследования внедрены в практическую деятельность (лечебно-диагностическую деятельность) ФГБУ «Главный военный клинический госпиталь имени академика Н.Н. Бурденко» Минобороны России (головной организации и филиалов в составе стационарных медицинских организаций с амбулаторно-поликлиническими структурными подразделениями) и головного учреждения и 12 филиалов, включающих структурные амбулаторно-поликлинические подразделения (поликлиники) ФГКУ «1586 ВКГ» Минобороны

России и применяются в лечебно-диагностической работе с больными при диагностике и диспансерном наблюдении новообразований кожи, а также при повышении квалификации врачей-дерматовенерологов, онкологов и терапевтов.

Результаты, основные научные положения, выводы и практические рекомендации внедрены в учебный процесс автономной некоммерческой организации дополнительного профессионального образования «Московский медико-социальный институт имени Ф.П. Гааза» и медицинского института непрерывного образования федерального государственного бюджетного образовательного учреждения высшего образования «Российский биотехнологический университет (РОСБИОТЕХ)» Минобрнауки России при изучении дисциплин: «Дерматовенерология», «Онкология», преподаваемых слушателям по специальностям: «Дерматовенерология», «Косметология», «Онкология», «Общая врачебная практика (семейная медицина)».

Публикации по теме диссертации

По результатам исследования автором опубликовано 62 печатные работы, в том числе 15 научных статей в журналах, включенных в Перечень рецензируемых научных изданий Сеченовского Университета / Перечень ВАК при Минобрнауки России, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора наук; 18 статей в изданиях, индексируемых в международной базе Scopus; 12 иных публикаций по результатам исследования; 2 публикации в сборниках материалов международных научных конференций; 2 патента; 2 свидетельства о государственной регистрации базы данных; 11 свидетельств о государственной регистрации программ для ЭВМ.

Структура и объем диссертации

Диссертация изложена на 500 страницах машинописного текста и состоит из введения, 4 глав, заключения, выводов, практических рекомендаций, списка

сокращений и условных обозначений, списка литературы и приложений. Работа иллюстрирована 72 таблицами и 127 рисунками (из них 1 в Приложении). Список литературы содержит 374 источника, из которых 109 – на русском языке и 265 – на иностранных языках.

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Эпидемиология злокачественных новообразований кожи

Злокачественные новообразования кожи представляют собой одну из наиболее распространенных групп онкологических заболеваний в мире, а их структура и динамика заболеваемости существенно различаются между популяциями в зависимости от уровня инсоляции, фототипа кожи, возрастного состава населения и качества регистрации случаев [207, 267, 293]. Согласно оценкам глобальной статистики онкологических заболеваний, меланома кожи ежегодно диагностируется у сотен тысяч человек, а немеланомные формы рака кожи, к которым относят, прежде всего, базальноклеточный и плоскоклеточный рак, по абсолютному числу случаев значительно превосходят меланому, хотя и характеризуются существенно меньшей летальностью [235, 252, 279].

По данным международных источников, за последние десятилетия во многих странах отмечается неуклонный рост заболеваемости меланомой кожи, который в ряде популяций опережает рост заболеваемости другими злокачественными опухолями [234, 277, 371]. Прогнозные модели указывают на дальнейшее увеличение числа новых случаев меланомы вплоть до 2040 года, что связывают как с истинным ростом заболеваемости, обусловленным особенностями образа жизни и характером ультрафиолетового облучения, так и с совершенствованием методов ранней диагностики и расширением скрининговых программ [234, 368, 371]. При этом эпидемиологические тенденции в разных регионах мира неоднородны: наиболее высокие показатели заболеваемости меланомой регистрируются среди светлогокожего населения Австралии, Новой Зеландии, Северной Америки и Северной Европы [23, 305, 306].

Немеланомный рак кожи (НМРК), объединяющий базальноклеточный и плоскоклеточный рак, является самой частой злокачественной опухолью человека, однако его истинная распространенность систематически недооценивается вследствие неполноты онкологической регистрации, поскольку во многих канцер-

регистрах эти опухоли учитываются не в полном объеме [6, 49, 252]. Систематические обзоры демонстрируют выраженный рост заболеваемости НМРК в большинстве изученных популяций, что создает колоссальную нагрузку на дерматологическую и онкологическую службы [252]. Значительная часть этой нагрузки приходится на первичное звено здравоохранения, где осуществляется первичный отбор и маршрутизация пациентов с новообразованиями кожи [180, 291].

Ключевыми факторами риска развития ЗНО кожи признаны кумулятивное и интермиттирующее ультрафиолетовое облучение, светлый фототип кожи, наличие множественных и атипичных (диспластических) невусов, отягощенный персональный и семейный анамнез по меланоме, иммуносупрессия и ряд наследственных синдромов [277, 305, 371]. Метаанализы показали, что риск развития меланомы существенно возрастает при увеличении общего числа невусов и наличии атипичных невусов, а также при эпизодах солнечных ожогов, особенно в детском и подростковом возрасте [277, 371]. Понимание структуры факторов риска лежит в основе стратификации пациентов и формирования групп высокого риска, подлежащих динамическому наблюдению [26, 306].

В Российской Федерации новообразования кожи, включая меланому, занимают одно из ведущих мест в структуре онкологической заболеваемости населения [81, 99, 100]. Анализ данных отечественных онкологических регистров свидетельствует о продолжающемся росте показателей заболеваемости меланомой и другими ЗНО кожи, а также о сохраняющейся проблеме поздней диагностики, при которой значимая доля меланом выявляется на стадиях, ассоциированных с неблагоприятным прогнозом [98, 101, 106]. Смертность от ЗНО кожи, прежде всего от меланомы, вносит заметный вклад в структуру онкологической смертности, что подчеркивает значимость мероприятий по раннему выявлению этих опухолей [73, 99]. В рамках совершенствования работы первичного звена здравоохранения особое внимание уделяется диагностике и маршрутизации пациентов с подозрением на злокачественные новообразования кожи. Это позволяет

своевременно выявлять заболевания и направлять пациентов на дальнейшее обследование и лечение [21].

Экономическое бремя ЗНО кожи для систем здравоохранения также постоянно возрастает и складывается из затрат на диагностику, хирургическое и лекарственное лечение, динамическое наблюдение и реабилитацию пациентов [236, 267]. Показано, что раннее выявление новообразований кожи и рациональная маршрутизация пациентов способны существенно снизить эти затраты за счет уменьшения числа запущенных случаев и необоснованных вмешательств [179, 236, 291]. Таким образом, эпидемиологические данные обосновывают необходимость поиска новых, экономически эффективных подходов к ранней диагностике ЗНО кожи, в том числе с применением технологий искусственного интеллекта.

1.2. Клинические и морфологические особенности основных форм новообразований кожи

Клинические и морфологические характеристики новообразований кожи чрезвычайно разнообразны, что обуславливает объективные трудности их дифференциальной диагностики [22, 40, 63, 65, 372]. В соответствии с классификацией опухолей кожи Всемирной организации здравоохранения выделяют меланоцитарные и немеланоцитарные новообразования, среди которых по клинической значимости выделяются меланома, базальноклеточный и плоскоклеточный рак, а также предраковые поражения кожи [120, 372]. Понимание клинико-морфологических особенностей каждой из этих форм является необходимым условием построения корректных диагностических алгоритмов, в том числе автоматизированных.

Меланома кожи

Меланома кожи представляет собой злокачественную опухоль, развивающуюся из меланоцитов, и характеризуется высоким метастатическим

потенциалом и неблагоприятным прогнозом при поздней диагностике [63]. Клиническая диагностика меланомы традиционно опирается на правило ABCD(E), включающее оценку асимметрии (Asymmetry), неровности краев (Border), неоднородности окраски (Color), диаметра (Diameter) и эволюции образования (Evolution) [64, 175, 201, 309]. Дополнительное значение имеет признак «гадкого утенка» (ugly duckling sign), основанный на выявлении невуса, морфологически отличающегося от остальных образований у данного пациента [356]. Несмотря на важность ABCDE-критериев, для более точной диагностики необходимы дополнительные инструментальные методы [126].

Прогноз при меланоме напрямую зависит от глубины инвазии опухоли по Бреслоу и стадии заболевания на момент диагностики, что определяет исключительную важность ее раннего выявления [200, 227, 280]. Внедрение системы стадирования Американского объединенного комитета по раку (AJCC) позволило стандартизировать оценку прогноза и подходы к лечению меланомы [227]. Морфологическая верификация диагноза остается «золотым стандартом», однако именно на этапе клинического и дерматоскопического отбора образований для биопсии формируется значительная часть диагностических ошибок [215].

Базальноклеточный и плоскоклеточный рак кожи

Базальноклеточный рак кожи (БКРК) является наиболее частой злокачественной опухолью человека и характеризуется медленным ростом, местнодеструктивным характером и крайне низким метастатическим потенциалом [164, 195]. Клинически БКРК проявляется множеством форм: узловой, поверхностной, склеродермоподобной, пигментной, что затрудняет его дифференциальную диагностику с доброкачественными образованиями и меланомой [164, 240]. Плоскоклеточный рак кожи (ПКРК) занимает второе место по частоте среди немеланомных форм рака кожи и, в отличие от БКРК, обладает более выраженным потенциалом к местному распространению и метастазированию [301, 331]. Ранняя диагностика и адекватное лечение этих

опухолей позволяют достичь высоких показателей излечения при минимальном объеме вмешательства.

Дерматоскопические признаки БКРК и ПКРК хорошо изучены и включают характерные сосудистые и структурные паттерны, что существенно повышает точность их клинической диагностики [189, 192, 338]. Для плоскоклеточного рака и его предшественников описана дерматоскопическая модель прогрессии актинического кератоза [188, 304]. Систематические обзоры и метаанализы подтверждают высокую диагностическую точность дерматоскопии при базальноклеточном раке кожи, что делает ее незаменимым инструментом в практике дерматовенеролога [192, 198, 339].

Предраковые поражения кожи и доброкачественные новообразования

Значительную часть новообразований кожи составляют предраковые поражения, прежде всего, актинический кератоз, который рассматривается как потенциальный предшественник плоскоклеточного рака кожи [122, 274, 304]. Риск злокачественной трансформации отдельного очага актинического кератоза относительно невелик, однако при большом числе очагов и длительном хроническом фотоповреждении кожи кумулятивный риск развития инвазивного рака становится клинически значимым [71, 122, 274]. Своевременное выявление и лечение предраковых поражений кожи наружными препаратами и физическими методами является важным компонентом профилактики ЗНО кожи [71].

Подавляющее большинство новообразований кожи являются доброкачественными (меланоцитарные невусы, себорейный кератоз, дерматофибромы, сосудистые образования и др.) и не представляют непосредственной угрозы для жизни пациента. Тем не менее, именно необходимость надежной дифференциации доброкачественных образований от злокачественных определяет основную клиническую задачу дерматоонкологии [5, 78, 103]. Отечественные исследования показали возможность применения программ ИИ для предварительной дифференциальной диагностики

доброкачественных и злокачественных опухолей различного тканевого происхождения, меланоцитарных, кератиноцитарных/эпидермальных, сосудистых и фиброзных [78, 103, 108]. Сложность и разнообразие клинико-морфологической картины новообразований кожи объясняют, почему автоматизированная классификация этих опухолей остается нетривиальной задачей даже для современных нейросетевых моделей [29, 144].

1.3. Современные подходы и трудности диагностики новообразований кожи

Диагностика новообразований кожи представляет собой многоэтапный процесс, включающий сбор анамнеза, клинический осмотр кожного покрова, дерматоскопическое исследование, при необходимости фотофиксацию и динамическое наблюдение, а также морфологическую верификацию диагноза [254, 289]. Оптимизация точности диагностики новообразований кожи может быть достигнута за счет комплекса мероприятий, реализуемых на разных уровнях оказания медицинской помощи: повышения квалификации врачей первичного звена и дерматовенерологов, внедрения клинико-дерматоскопического динамического наблюдения пациентов групп риска и использования автоматизированных программ на основе технологий ИИ [109].

1.3.1. Роль врачей первичного звена в ранней диагностике злокачественных новообразований кожи

Врачи общей практики и терапевты первичного звена играют ключевую роль в раннем выявлении ЗНО кожи, выступая в качестве «первой линии» диагностики и последующей маршрутизации пациентов [230, 291]. Особую значимость это приобретает в регионах с ограниченными ресурсами, небольших городах и сельских территориях, где доступ к специализированной дерматовенерологической и онкологической помощи затруднен, а взаимодействие между врачами первичного звена и узкими специалистами нередко нарушено [353].

Хотя большинство новообразований кожи доброкачественны, врач первичного звена должен уверенно распознавать клинические признаки, указывающие на возможную злокачественность образования и необходимость направления пациента к профильному специалисту [230]. В российской нормативной базе задача ранней диагностики ЗНО закреплена, в частности, в проекте профессионального стандарта врача общей практики (семейного врача), предусматривающем соответствующие трудовые функции [18, 72].

Вместе с тем диагностическая точность врачей первичного звена в отношении новообразований кожи существенно уступает точности дерматовенерологов, что подтверждено систематическими обзорами и метаанализами [326]. Ограниченность подготовки врачей первичного звена по вопросам дерматоонкологии приводит как к пропуску злокачественных опухолей, так и к избыточным направлениям пациентов с доброкачественными образованиями, что увеличивает нагрузку на специализированное звено и затраты здравоохранения. Образовательные программы, направленные на повышение квалификации врачей первичного звена в области диагностики рака кожи, демонстрируют эффективность, однако их охват и устойчивость результатов ограничены [70, 322]. Обучение врачей основам дерматоскопии в условиях первичной помощи повышает точность диагностики, но требует значительных организационных и временных ресурсов [70, 322].

1.3.2. Дерматоскопия в дифференциальной диагностике новообразований кожи

Дерматоскопия является методом неинвазивной визуализации, позволяющим оценивать структуры образования, недоступные при осмотре невооруженным глазом [19, 20, 332, 374]. Метаанализы убедительно показали, что применение дерматоскопии обученными специалистами существенно повышает чувствительность и специфичность диагностики меланомы по сравнению с осмотром невооруженным глазом [191]. По обобщенным данным,

чувствительность диагностики меланомы при использовании дерматоскопии возрастает с 60,9–76,0% (осмотр невооруженным глазом) до 85,4–95,8%, а специфичность с 70,6–83,6% до 84,6–86,0% [193, 362]. Аналогичным образом дерматоскопия повышает точность диагностики базальноклеточного рака кожи: чувствительность возрастает с 66,9% до 85%, а специфичность достигает 97,2–98,2% [339].

Для стандартизации дерматоскопической диагностики предложен ряд алгоритмов и диагностических правил, включая правило ABCD дерматоскопии, семибалльный чек-лист, трехбалльный чек-лист, метод анализа паттернов и другие [194, 208, 336, 348]. Применение дерматоскопии способствует не только повышению чувствительности диагностики, но и сокращению числа необоснованных биопсий доброкачественных образований, что снижает затраты здравоохранения и улучшает исходы лечения [312]. Дерматоскопия находит применение и при немеланоцитарных, беспигментных новообразованиях кожи, а также в общей дерматологии [210].

Существенным ограничением дерматоскопии является время, необходимое для полноценного обследования. Показано, что применение дерматоскопии при осмотре небольшого числа образований увеличивает общее время осмотра лишь на 72 секунды [278]. Кроме того, освоение метода требует специальной подготовки, а число врачей, свободно владеющих дерматоскопией, остается недостаточным, особенно за пределами крупных специализированных центров [190].

1.3.3. Динамическое наблюдение пациентов групп высокого риска

Систематическое клинико-дерматоскопическое динамическое наблюдение пациентов с высоким риском развития ЗНО кожи направлено на раннее выявление прежде всего меланомы и позволяет минимизировать пропуски злокачественных патологий при одновременном снижении числа необоснованных биопсий [272, 281]. Основными компонентами такого мониторинга являются выявление факторов риска, систематический осмотр всего кожного покрова, дерматоскопия,

фотофиксация новообразований и обучение пациентов принципам самоконтроля и коррекции образа жизни. К группам высокого и очень высокого риска относят пациентов с отягощенным семейным или персональным анамнезом по меланоме и раку кожи, множественными и диспластическими невусами, актиническим кератозом, эпизодами хронического фотоповреждения кожи, иммуносупрессией и наследственными синдромами.

Методы цифровой дерматоскопии и полного фотокартирования тела (total body photography) с последующим сравнением изображений в динамике позволяют выявлять новые и изменяющиеся образования, что особенно важно у пациентов с множественными невусами. Исследования подтверждают, что цифровое дерматоскопическое наблюдение способствует раннему выявлению меланом, в том числе не имеющих специфических дерматоскопических признаков на момент первичного осмотра [281, 346].

1.3.4. Проблема гипо- и гипердиагностики новообразований кожи

Одной из фундаментальных проблем диагностики новообразований кожи является баланс между гиподиагностикой (пропуском злокачественных опухолей) и гипердиагностикой с последующим избыточным лечением [7, 124, 125]. Стремление минимизировать пропуски меланомы приводит к росту числа биопсий доброкачественных образований, что характеризуется показателем «числа образований, подлежащих иссечению для выявления одной меланомы», существенно различающимся между специалистами и врачами первичного звена [282]. Онкологическая настороженность, то есть направление гистологически доброкачественных, но клинически подозрительных образований на дополнительное обследование, является нормой клинической практики дерматоонкологии, однако при формальной оценке диагностической точности она нередко трактуется как ложноположительный результат. Эффективность популяционного скрининга новообразований кожи остается предметом дискуссий, что связано как с методологическими трудностями оценки, так и с риском

гипердиагностики [318, 319]. Указанные обстоятельства обуславливают потребность в инструментах, способных повысить точность первичной дифференциальной диагностики и оптимизировать маршрутизацию пациентов, к числу которых относятся программы на основе технологий ИИ [78, 103, 109].

1.4. Программы компьютерного зрения на основе искусственного интеллекта в дерматовенерологии

Дерматовенерология, будучи клинической специальностью с преимущественно визуальным характером диагностики, представляет собой одну из наиболее перспективных областей для интеграции технологий искусственного интеллекта и компьютерного зрения [143, 166]. Применение ИИ в дерматовенерологии имеет корни, уходящие в начало 2000-х годов, когда исследователи начали изучать возможности алгоритмов машинного обучения для распознавания доброкачественных и злокачественных образований кожи [241]. Концепция глубокого обучения позволила вычислительным моделям с несколькими уровнями обработки данных выявлять представление данных на разных уровнях абстракции и совершила переворот в анализе медицинских изображений [245].

1.4.1. Методы машинного и глубокого обучения в анализе фотоизображений кожи

Для автоматизированного анализа медицинских изображений применяются методы компьютерного зрения (computer vision), ключевой задачей которых является классификация изображений [134]. Помимо классификации, широко используются детекция (обнаружение объектов), семантическая сегментация и генерация изображений [138, 162, 173]. В дерматоонкологии наиболее часто применяется метод классификации изображений, поскольку он не требует столь трудоемкой разметки, как детекция, менее чувствителен к качеству

фотоизображений, чем сегментация, и при этом обеспечивает высокую точность [31, 35, 328]. Обобщенный алгоритм автоматизированного процесса диагностики включает регистрацию цифрового изображения новообразования, этап предварительной обработки, этап логического вывода с применением нейросетевой модели и завершающий этап интерпретации результатов и принятия клинического решения [163].

Основным инструментом анализа изображений кожи в последние годы стали сверточные нейронные сети (convolutional neural networks, CNN), архитектуры которых: AlexNet, VGG, GoogLeNet/Inception, ResNet, DenseNet, EfficientNet – последовательно повышали точность классификации изображений [184, 186, 262, 308, 330]. Успех глубокого обучения в значительной степени обусловлен доступностью крупных размеченных наборов изображений, таких как ImageNet, и развитием методов переноса обучения (transfer learning), позволяющих адаптировать предобученные модели к задачам медицинской визуализации при ограниченном объеме данных [238, 247, 248, 350]. Важную роль играют также методы аугментации данных, компенсирующие ограниченность и несбалансированность обучающих выборок [259, 288, 320]. В последние годы все большее применение находят архитектуры на основе трансформеров (Vision Transformers) и мультимодальные модели, объединяющие изображения, текст и клинические данные [60, 133, 343]. Также важно учитывать модальность самих изображений, которые попадают на вход модели [61].

1.4.2. Диагностическая точность программ искусственного интеллекта в дерматоонкологии

Начиная с основополагающего исследования А. Esteva и соавт. (2017), в котором сверточная нейронная сеть достигла уровня классификации новообразований кожи, сопоставимого с уровнем сертифицированных дерматологов, было опубликовано множество работ, подтверждающих высокую диагностическую точность нейросетевых моделей [78, 103, 109, 176, 323]. В серии

исследований Н. Haenssle и Т. Brinker с соавт. показано, что алгоритмы глубокого обучения способны превосходить большинство дерматологов в задачах дерматоскопической классификации меланомы [112, 183, 273]. В рамках международного симпозиума были приведены результаты, что производительность лучших алгоритмов машинного обучения превосходит среднюю производительность врачей при классификации пигментированных образований кожи [307]. Отдельные работы показали превосходство алгоритмов над экспертами и в диагностике беспигментных новообразований кожи [182, 221].

Систематические обзоры и метаанализы, обобщающие результаты множества исследований, подтверждают в целом высокую диагностическую точность ИИ в распознавании меланомы и других новообразований кожи, сопоставимую с точностью врачей-специалистов [118, 121, 324, 325]. Вместе с тем авторы обзоров указывают на существенную гетерогенность методологии, преобладание ретроспективных исследований на курируемых наборах данных и ограниченное число проспективных клинических испытаний, что затрудняет экстраполяцию полученных результатов на реальную практику [110, 154]. Метаанализ диагностической точности ИИ в сравнении с врачами общей практики и дерматологами при кожных заболеваниях подтвердил конкурентоспособность алгоритмов, но подчеркнул зависимость результатов от условий применения [196, 197]. Сравнительные исследования отечественных программ компьютерного зрения показали высокую точность предварительной дифференциальной диагностики новообразований кожи различного тканевого происхождения [31, 103, 109].

1.4.3. Мобильные приложения на основе искусственного интеллекта и системы поддержки принятия решений

Для широкого распространения технологий ИИ в дерматовенерологии наиболее подходящим форматом являются мобильные приложения, доступные как пациентам, так и врачам, не требующие дорогостоящего специализированного

оборудования [329]. Систематические обзоры мобильных приложений для оценки риска рака кожи показали, что рынок таких приложений растет экспоненциально, однако качество и безопасность большинства продуктов остаются низкими: значительная их часть не имеет рецензируемых публикаций о клинической эффективности, характеризуется непрозрачностью обучающих данных и неопределенным регуляторным статусом [335]. Диагностическая точность приложений для самостоятельной оценки новообразований пациентами существенно варьирует, что ограничивает их применение в качестве самостоятельного диагностического инструмента и позиционирует преимущественно как средства предварительного скрининга с обязательным последующим осмотром врача [150, 364].

Применение программ ИИ в дерматоонкологии актуально на двух уровнях оказания медицинской помощи. На уровне первичной медико-санитарной помощи программы ИИ могут служить инструментом предварительной диагностики для врачей общей практики и терапевтов, не обладающих специализированной подготовкой, что оптимизирует маршрутизацию пациентов и ускоряет выявление злокачественных опухолей [56, 225]. На уровне специализированной помощи заключения программ ИИ могут использоваться как «второе мнение» для дерматовенерологов и онкологов [104, 246]. Концепция «дополненного интеллекта» (augmented intelligence), при которой ИИ дополняет, а не заменяет врача, рассматривается как наиболее реалистичная модель внедрения этих технологий в клиническую практику [160, 246, 345]. Для практического применения программ ИИ разработаны алгоритмы маршрутизации пациентов, предоставляющие рекомендации по дальнейшей тактике в зависимости от уровня риска: направление к онкологу или дерматологу, диспансерное наблюдение [1, 28, 53].

Отечественным примером комплексного подхода является мультимодельная программа для автоматизированной визуальной идентификации и дифференциальной диагностики доброкачественных и злокачественных новообразований кожи «Derma Onko Melanoma Check», реализующая

мультимодельную архитектуру компьютерного зрения с модулями проверки качества фотоизображения, коррекции его параметров, классификации, алгоритмов маршрутизации и риск-модели для комбинации заключений [88, 89]. Данная программа относится к системам поддержки принятия врачебных решений и обеспечивает локальную обработку изображений без их хранения вне устройства пользователя [90]. Перспективы дальнейшего развития ИИ в дерматовенерологии связывают с внедрением трансформеров, мультимодальных систем и больших языковых моделей, а также с интеграцией ИИ в телемедицину и системы электронного здравоохранения [265, 271].

1.4.4. Ограничения и причины ошибок программ искусственного интеллекта

Несмотря на впечатляющие результаты, применение программ ИИ в дерматоонкологии сопряжено с рядом существенных ограничений. Нейронные сети функционируют как «черные ящики», что затрудняет выявление причин ошибочных заключений, как ложноположительных, так и ложноотрицательных. Для повышения интерпретируемости моделей применяются методы объяснимого ИИ (explainable AI), в том числе Grad-CAM, Score-CAM, анализ чувствительности к окклюзии, SHAP и LIME [167, 185, 222, 237, 317]. Систематические обзоры показывают, что методы объяснимого ИИ повышают доверие врачей к заключениям алгоритмов, однако их надежность и клиническая ценность остаются предметом дискуссий [187, 224, 233].

Существенной проблемой является зависимость точности классификации от качества входных фотоизображений: фотометрические и текстурные параметры изображения (яркость, контраст, резкость, размытие, доля пересвеченных пикселей и др.) значимо влияют на точность работы моделей ИИ [29, 69]. Показано, что модели могут опираться на артефакты изображения: хирургические метки, линейки, волосы, что приводит к систематическим ошибкам классификации [54, 156, 161, 181]. Кроме того, нейросетевые модели подвержены проблеме «скрытой стратификации» и снижению производительности при применении к данным,

отличающимся от обучающей выборки [127, 244, 363]. Разработанные методы адресной предобработки изображений, нормализующие отклоняющиеся параметры, позволяют повысить точность и специфичность классификации [29, 45, 59]. Указанные ограничения обуславливают необходимость тщательной валидации программ ИИ в реальных клинических условиях и разработки методологии анализа причин их ошибок [258].

1.5. Наборы данных для обучения моделей искусственного интеллекта в дерматовенерологии и дерматоонкологии

Качество и репрезентативность обучающих наборов данных являются определяющими факторами точности и обобщаемости моделей искусственного интеллекта в дерматовенерологии и дерматоонкологии [59, 170, 263]. Формирование корректного набора данных предполагает не только достаточный объем изображений, но и их клиническую и морфологическую верификацию, сбалансированность классов, разнообразие фототипов кожи и условий фотофиксации, а также прозрачность описания характеристик выборки [135, 229].

Публичные наборы данных фотоизображений новообразований кожи

Наиболее широко используемыми публичными наборами данных дерматоскопических изображений являются коллекции International Skin Imaging Collaboration (ISIC) и набор HAM10000, содержащий тысячи дерматоскопических изображений распространенных пигментированных образований кожи из нескольких источников. На основе архива ISIC проводились крупные международные соревнования по классификации новообразований кожи, стимулировавшие развитие алгоритмов. Дополнительно применяются наборы BCN20000, ISIC 2020, PH2, PAD-UFES-20 (клинические изображения со смартфонов с метаданными пациентов), MED-NODE, а также специализированные

наборы для оценки справедливости алгоритмов Fitzpatrick17k и Diverse Dermatology Images (DDI) [114, 165, 199, 276, 294, 299, 311].

Проблемы репрезентативности, смещения и прозрачности данных

Ключевой проблемой публичных наборов данных является их недостаточная репрезентативность в отношении фототипов кожи: большинство изображений получено от пациентов со светлой кожей, что снижает точность моделей при применении к пациентам с темными фототипами и порождает алгоритмическое смещение [123, 260]. Обзорные исследования продемонстрировали недостаток прозрачности и потенциальную предвзятость наборов данных и алгоритмов в дерматологическом ИИ, включая неполное документирование источников данных, методов их сбора и характеристик пациентов [349]. Географическая концентрация данных, использованных для обучения алгоритмов, ограничивает их обобщаемость на другие популяции [257]. Установлено также, что алгоритмы, обученные на нерепрезентативных данных, могут воспроизводить и усиливать существующие диспаритеты в оказании медицинской помощи [213, 358].

Для повышения прозрачности наборов данных предложены концепции «этикетки питательной ценности данных» (dataset nutrition label) и стандартизированного документирования, обеспечивающие раскрытие характеристик обучающих выборок [337]. Соблюдение принципов FAIR (Findable, Accessible, Interoperable, Reusable) и минимальных требований к документированию клинических моделей ИИ рассматривается как необходимое условие воспроизводимости и безопасного применения [284, 342]. Несбалансированность классов в наборах данных требует применения специальных методов: взвешивания классов, аугментации, механизмов внимания, повышающих качество классификации редких злокачественных образований. Отдельной задачей является учет неопределенности предсказаний модели, что особенно важно при принятии клинических решений [357].

Собственные наборы данных и регуляторные требования

В противовес использованию исключительно публичных наборов данных ряд разработчиков формирует собственные валидированные наборы изображений с подтвержденным участием врачей-дерматовенерологов в их наборе и верификации [85]. Такой подход обеспечивает клиническую релевантность данных и соответствие условиям реальной практики, включая используемые устройства фотофиксации. Влияние характеристик устройств фотофиксации (тип камеры, разрешение, параметры вспышки) на точность классификации является значимым фактором, требующим стандартизации протоколов получения изображений [62].

В Российской Федерации требования к структуре и порядку применения наборов данных для обучения и тестирования алгоритмов ИИ в клинической медицине регламентируются серией национальных стандартов ГОСТ Р 59921, а также стандартами, устанавливающими методы обезличивания и контроля наборов данных [8, 12, 16]. Общие положения и порядок клинической оценки систем ИИ в клинической медицине определены в ГОСТ Р 59921.0 и 59921.1, методы технических испытаний в ГОСТ Р 59921.2 и 59921.4, а требования к алгоритмам анализа медицинских изображений – в ГОСТ Р 59921.7 [8, 9, 10, 11, 13]. Дополнительно приняты стандарты для систем ИИ в лучевой диагностике, устанавливающие методы испытаний на устойчивость работы с разнородными данными [14, 15, 17]. Вопросы обращения с персональными данными, регистрации медицинских изделий и проведения экспериментальных правовых режимов регулируются профильным законодательством [82, 83, 84]. Этические аспекты применения ИИ в здравоохранении отражены в соответствующем национальном кодексе этики и международных руководствах [27, 214, 256]. Регуляторный статус программ ИИ существенно различается между странами, что создает «серую зону» в обращении медицинских приложений с ИИ [33, 139, 285].

1.6. Отношение врачей к технологиям искусственного интеллекта: обзор зарубежных исследований

Успешность внедрения технологий искусственного интеллекта в клиническую практику определяется не только их технической эффективностью, но и готовностью врачей применять эти технологии, уровнем их осведомленности и доверия [140, 220, 344]. В зарубежной литературе накоплен значительный объем данных о восприятии ИИ медицинскими работниками, тогда как в отечественной литературе исследования отношения профессионального сообщества к ИИ представлены крайне ограниченно [58, 290].

1.6.1. Отношение дерматовенерологов к технологиям искусственного интеллекта

Международные опросы дерматологов демонстрируют в целом положительное, но осторожное отношение к технологиям ИИ [298]. Опрос австралийских дерматологов выявил преимущественно позитивное отношение к ИИ при сохранении опасений относительно точности алгоритмов в реальной практике и регуляторной неопределенности [159]. В крупном международном онлайн-опросе большинство дерматологов выразили заинтересованность в применении ИИ, однако отметили обеспокоенность вопросами ответственности, надежности и влияния технологий на взаимоотношения врача и пациента [157]. Специалисты в целом положительно оценивают перспективы ИИ, но подчеркивают необходимость валидации [157].

Опросы, охватывающие специалистов несколько так называемых за рубежом «визуальных» специальностей (офтальмология, дерматология, радиология), показывают, что уровень доверия к ИИ и готовности его применять во многом определяется наличием конкретных валидированных инструментов в соответствующей области [130]. Международное кросс-секционное исследование восприятия ИИ в дерматологии подтвердило, что осведомленность специалистов о

технологиях ИИ не всегда трансформируется в доверие к ним и готовность к применению [298]. Пациенты, в свою очередь, демонстрируют неоднозначное отношение к применению ИИ для скрининга рака кожи, отмечая как удобство и доступность технологий, так и обеспокоенность вопросами точности и конфиденциальности данных [149, 297].

1.6.2. Отношение онкологов и врачей первичного звена здравоохранения к технологиям искусственного интеллекта

Исследования отношения онкологов к технологиям ИИ проводились в ряде стран. Турецкое исследование восприятия ИИ медицинскими онкологами показало высокий уровень использования генеративных текстовых систем при сохранении обеспокоенности этическими и регуляторными аспектами, а также вопросами публикационной этики [209]. Общенациональное исследование в Китае выявило преобладающее применение специализированных медицинских ИИ-продуктов и оценивало знания, отношение и готовность онкологов по шкале Ликерта, при этом основные опасения касались абстрактного вопроса замещения врачей технологиями [129]. Обзорные работы подчеркивают быстрый рост применения ИИ в онкологии при сохранении вызовов, связанных с валидацией, интеграцией в клинические процессы и взаимодействием врача и пациента [137, 148, 172, 347].

В сфере оказания первичной медико-санитарной помощи применение ИИ рассматривается как одно из наиболее перспективных направлений в силу высокой нагрузки, дефицита кадров и необходимости работы с большими объемами клинических данных [270, 355]. Опросы российских врачей-специалистов демонстрируют наличие барьеров внедрения ИИ, связанных с недостаточной осведомленностью, недоверием к заключениям алгоритмов и регуляторной неопределенностью [38, 42, 77, 79]. Международные данные о динамике отношения специалистов к ИИ свидетельствуют о постепенном росте принятия технологий по мере накопления практического опыта [253]. Систематические обзоры факторов, влияющих на доверие медицинских работников к ИИ, выделяют

в качестве ключевых прозрачность алгоритмов, наличие доказательной базы, интеграцию в рабочий процесс и ясность в вопросах ответственности [119].

1.6.3. Барьеры внедрения и факторы доверия

Систематические обзоры барьеров внедрения ИИ в здравоохранение выделяют технологические, организационные, этические, правовые и человеческие факторы, среди которых особое место занимают недоверие врачей, недостаток обучения и опасения относительно ответственности за решения с участием алгоритмов [119, 168]. Показано, что реальный практический опыт взаимодействия с ИИ-технологиями является одним из ключевых предикторов положительного отношения к ним, тогда как информированность о существовании технологий сама по себе не гарантирует доверия. Заниженные представления врачей о точности ИИ, формируемые под влиянием медиаобраза технологий и отдельных случаев неудачного применения, способны воспроизводить скептицизм и препятствовать внедрению [77]. Для преодоления барьера недоверия предлагаются не информационные кампании, а пилотные проекты, обеспечивающие контролируемый контакт врача с ИИ-системой в рабочей среде, регуляторная определенность и доступ к реестрам верифицированных ИИ-систем [211, 217, 302]. Внедрение ИИ требует также внимания к этическим вопросам, предвзятости алгоритмов, прозрачности решений и защите персональных данных [169, 232, 341].

1.7. Методологические подходы к оценке точности и медико-экономической эффективности программ искусственного интеллекта

Обоснованное внедрение программ искусственного интеллекта в клиническую практику требует всесторонней оценки их диагностической точности, клинической эффективности и медико-экономической эффективности [128, 295]. Методология такой оценки активно развивается и включает как

общепринятые метрики диагностической точности, так и специализированные подходы, учитывающие клинический контекст применения технологий [36].

1.7.1. Метрики диагностической точности и стандарты отчетности

Оценка диагностической точности программ ИИ традиционно основывается на стандартной матрице ошибок (confusion matrix) с расчетом чувствительности, специфичности, точности, прогностической ценности положительного и отрицательного результатов, а также площади под ROC-кривой [155]. При несбалансированных выборках, характерных для дерматоонкологии, более информативными считаются кривые «точность – полнота» (precision-recall) [283, 314]. Для обеспечения корректности и воспроизводимости исследований в области медицинского ИИ разработан ряд стандартов отчетности: TRIPOD и его обновленная версия TRIPOD+AI для прогностических моделей, STARD и STARD-AI для исследований диагностической точности, CONSORT-AI и SPIRIT-AI для клинических испытаний, DECIDE-AI для ранней клинической оценки систем поддержки принятия решений [239, 333, 352]. Систематические обзоры, однако, констатируют недостаточное соблюдение этих стандартов и преобладание методологически несовершенных исследований, что ограничивает доказательную базу применения ИИ [269].

Отдельной методологической проблемой является разрыв между высокой точностью алгоритмов, показанной на курируемых ретроспективных наборах данных, и их производительностью в реальных клинических условиях. Для оценки долгосрочного реального влияния ИИ-инструментов на клиническую практику предложены специализированные рамочные подходы, учитывающие не только диагностическую точность, но и влияние на процессы оказания помощи и исходы [43, 44, 46, 217]. Подчеркивается необходимость перехода от оценки алгоритмов «в лаборатории» к проспективным клиническим испытаниям и оценке клинической полезности [43, 44, 46].

1.7.2. Клиническая точность и учет онкологической настороженности

Важным методологическим достижением является признание того, что оценка диагностической эффективности программ ИИ исключительно по стандартной матрице ошибок систематически занижает их клиническую ценность в дерматоонкологии. Это связано с тем, что направление гистологически доброкачественных, но клинически подозрительных новообразований на дополнительное обследование (онкологическая настороженность) является нормой клинической практики, однако при формальной оценке трактуется как ложноположительный результат [36]. Для преодоления этого противоречия предложены новые показатели клинической точности и клинической эффективности диагностики программами ИИ, основанные на системе категорий клинической эффективности и матрице клинических случаев, учитывающей онкологическую настороженность как самостоятельную клиническую ценность. Применение таких показателей обеспечивает более объективную оценку программ ИИ применительно к реальным задачам дерматоонкологической маршрутизации. Логика клинического мышления, положенная в основу этих методологических подходов, соответствует принципам доказательной медицины и клинических рекомендаций [30, 55, 115, 216].

1.7.3. Медико-экономическая эффективность и оценка технологий здравоохранения

Оценка медико-экономической эффективности (МЭЭ) программ ИИ является необходимым условием обоснования их широкого внедрения и осуществляется в рамках оценки технологий здравоохранения (health technology assessment, HTA) [52, 136, 334]. Методология МЭЭ включает анализ «затраты и эффективность», «затраты и полезность», «затраты и выгода» и анализ влияния на бюджет, с использованием таких показателей, как коэффициент «затраты и эффективность» и сохраненные годы качественной жизни (QALY) [132, 303].

Особенности оценки цифровых медицинских продуктов и сервисов, включая программы ИИ, требуют адаптации классических подходов НТА [218, 286].

Систематические обзоры экономических оценок технологий ИИ в здравоохранении показывают, что, несмотря на быстрый рост числа публикаций, доказательная база их экономической эффективности остается ограниченной, а методология исследований – неоднородной [46, 203, 340]. Обзоры экономической эффективности и влияния на бюджет ИИ-решений подчеркивают недостаток проспективных исследований и потребность в стандартизации методологии [204]. Отдельные модельные исследования продемонстрировали потенциальную экономическую эффективность ИИ в конкретных клинических сценариях, при скрининге, неврологии и эндоскопической диагностике [177, 178, 202]. В отечественной практике методология комплексной оценки МЭЭ здравоохранения и клинико-экономического анализа медицинских изделий разработана в ряде работ и методических рекомендаций [2, 3, 68]. Концептуальные модели оценки экономической эффективности ИИ-решений в здравоохранении находятся в стадии активной разработки [67, 102].

Применительно к диагностике новообразований кожи МЭЭ программ ИИ целесообразно оценивать на двух уровнях: на уровне отдельного медицинского учреждения и на уровне региона, с учетом влияния алгоритмов маршрутизации пациентов на структуру потоков, число необоснованных направлений и сроки выявления злокачественных опухолей. Разработанные способы оценки эффективности и медико-экономической эффективности диагностики с использованием алгоритмов на основании заключений моделей ИИ обеспечивают методологическую основу для такого анализа [43]. Таким образом, комплексная методология оценки программ ИИ в дерматоонкологии должна объединять анализ диагностической и клинической точности, причин ошибочной классификации, отношения врачей к технологиям и медико-экономической эффективности их внедрения, что и определяет направления настоящего диссертационного исследования.

1.8. Применение технологий искусственного интеллекта в медицинском образовании и подготовке врачей-дерматовенерологов

Стремительное внедрение технологий искусственного интеллекта в клиническую практику формирует новые требования к системе медицинского образования, поскольку будущим врачам предстоит работать в среде, насыщенной цифровыми инструментами и системами поддержки принятия решений [146, 251, 370]. Все чаще подчеркивается необходимость перехода медицинского образования от «эпохи информации» к «эпохе искусственного интеллекта», что предполагает формирование у обучающихся не только фактических знаний, но и навыков интеграции данных, критической оценки заключений алгоритмов и ответственного взаимодействия с цифровыми технологиями [152, 255, 370]. Соответственно, в образовательные программы медицинских вузов предлагается включать профессиональные компетенции по использованию ИИ на всех уровнях подготовки от студенчества до ординатуры и непрерывного медицинского образования [74, 105, 146].

1.8.1. Направления применения искусственного интеллекта в медицинском образовании

Анализ литературы позволяет выделить несколько основных направлений применения технологий ИИ в подготовке медицинских кадров: адаптивные (персонализированные) обучающие платформы, виртуальные пациенты и симуляционные технологии, интеллектуальные обучающие системы, автоматизированная оценка знаний и навыков, а также инструменты поддержки научно-исследовательской работы [145, 151, 152]. Адаптивные обучающие системы на основе ИИ анализируют результаты обучающегося, выявляют пробелы в знаниях и динамически корректируют содержание и последовательность учебного материала, обеспечивая индивидуализацию образовательной траектории и поддерживая оптимальный уровень сложности заданий [145, 152, 275].

Особую роль в клиническом обучении играют технологии виртуальных пациентов и симуляции, в том числе с применением виртуальной и дополненной реальности, которые позволяют обучающимся отрабатывать навыки сбора анамнеза, диагностики и принятия решений в безопасной, контролируемой среде без риска для реальных пациентов [365, 366, 367]. Большие языковые модели и генеративный ИИ открывают дополнительные возможности для создания учебных материалов, клинических кейсов, вопросов для самоконтроля и виртуальных собеседников, а также для составления индивидуальных планов обучения [107, 264].

Технологии ИИ применяются и для объективизации оценки знаний и практических навыков обучающихся: автоматизированные системы обеспечивают более объективную обратную связь, а методы обработки естественного языка и машинного обучения позволяют анализировать результаты в реальном времени [268]. Наконец, ИИ поддерживает научно-исследовательскую деятельность преподавателей и обучающихся, поиск и анализ литературы, обработку данных, подготовку научных текстов, способствуя интеграции новейших научных достижений в образовательный процесс [316]. В условиях непрерывного медицинского образования ИИ-платформы формируют персонализированные траектории профессионального развития, рекомендуя актуальные курсы и публикации в соответствии с индивидуальными потребностями специалиста.

Отечественные исследования подтверждают высокий образовательный потенциал этих технологий и одновременно фиксируют характерные барьеры их внедрения. Обзоры и эмпирические работы российских авторов показывают, что применение ИИ способствует индивидуализации обучения, ускорению освоения практических навыков и повышению мотивации обучающихся, однако сопряжено с проблемами недостаточной инфраструктуры, дефицита подготовленных кадров, этическими и правовыми вопросами, а также с необходимостью подготовки самих преподавателей [4].

1.8.2. Обучение врачей-дерматовенерологов, ординаторов и студентов с применением технологий искусственного интеллекта

Дерматовенерология как специальность с преимущественно визуальным характером диагностики представляет собой одну из наиболее благоприятных областей для интеграции технологий ИИ в образовательный процесс. Применение ИИ-инструментов в дерматологическом образовании преобразует учебный опыт, создавая возможности для персонализированной обратной связи и оптимизации рутинных задач, что повышает эффективность занятий. Интеллектуальные обучающие системы обеспечивают ординаторам и студентам индивидуальное сопровождение, выявляют пробелы в знаниях и адаптируют учебный материал к темпу и стилю обучения конкретного специалиста, позволяя последовательно осваивать все многообразие дерматологических нозологий и подходов к их ведению [37, 359].

Технологии виртуальных пациентов особенно ценны в подготовке дерматовенерологов – они позволяют ординаторам моделировать клинические ситуации и отрабатывать навыки клинического мышления, взаимодействуя с широким спектром дерматологических заболеваний без риска для пациента [366]. Возможность демонстрации обучающимся большого числа разнообразных нозологических форм принципиально расширяет образовательный опыт, который трудно обеспечить в рамках ограниченного реального клинического приема [321]. Адаптивные цифровые платформы, оценивающие уровень знаний и подстраивающие сложность и последовательность материала, способствуют формированию у обучающихся более глубокого и целостного понимания основ дерматологии.

Отдельным перспективным направлением является применение ИИ для образования не только специалистов, но и самих пациентов. Показано, что ИИ-инструменты способны предоставлять персонализированный, адаптированный к типу кожи, языку и уровню медицинской грамотности пациента образовательный контент, интерактивные визуализации патофизиологии дерматозов,

моделирование эффекта лечения во времени и работу виртуальных ассистентов, что особенно значимо при хронических заболеваниях кожи, требующих длительной приверженности терапии [80].

Важнейшим аспектом подготовки дерматовенерологов является формирование навыков корректной интерпретации заключений программ компьютерного зрения и понимания их ограничений, поскольку именно врач несет ответственность за окончательное диагностическое решение. Обучение будущих специалистов основам работы ИИ-систем, включая осведомленность о зависимости их точности от качества фотоизображений, о проблеме «черного ящика» и о рисках алгоритмического смещения, рассматривается как необходимое условие безопасного и эффективного применения этих технологий в дерматоонкологии. Отечественный опыт свидетельствует о том, что интеллектуальные системы поддержки принятия врачебных решений в дерматовенерологии могут выступать не только диагностическим, но и полноценным обучающим инструментом. Интеллектуальные обучающие системы рассматриваются как элемент системного подхода к повышению качества оказания медицинской помощи, обеспечивающий рост компетентности врачебного персонала [50]. На основе этого подхода были разработаны информационно-диагностические и обучающие системы для дерматовенерологии, объединяющие поддержку принятия решений с функцией обучения специалистов [51]. Практическим воплощением этой концепции стал модуль дистанционного обучения «Школа врача» в составе системы поддержки принятия врачебных решений в дерматовенерологии: показано, что за счет визуализации клинических симптомов и нового способа подачи информации такой курс позволяет оптимизировать и персонализировать учебный процесс, повышая эффективность первичной подготовки врачей-дерматовенерологов [80]. Формирование положительного, но критически осмысленного отношения обучающихся к технологиям ИИ, основанного на реальном практическом опыте взаимодействия с ними, является ключевым фактором успешного внедрения этих технологий в будущую профессиональную деятельность врача-дерматовенеролога.

ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Общий дизайн диссертационного исследования

Диссертационное исследование представляет собой комплексную многоэтапную научную работу. Методология исследования включает клинико-диагностические, опросные, статистические и медико-экономические методы, а также методы машинного обучения и объяснимого ИИ. Дизайн исследования включал опросные исследования отношения врачей к ИИ, формирование и анализ набора фотоизображений новообразований кожи, оценку клинической точности и эффективности программ ИИ, анализ причин ошибочной классификации и расчет медико-экономической эффективности диагностических алгоритмов.

Диссертация состоит из введения, обзора литературы, главы материалов и методов, глав собственных результатов, обсуждения, заключения, выводов, практических рекомендаций, списка литературы и приложений. Работа включает несколько самостоятельных, но методологически взаимосвязанных исследовательских блоков: (1) опросные исследования отношения врачей трех специальностей к технологиям ИИ; (2) сравнительный анализ программ компьютерного зрения (ПКЗ) на основе ИИ для дифференциальной диагностики заболеваний кожи; (3) формирование и анализ оригинального набора данных фотоизображений новообразований кожи; (4) методологическая разработка, валидация и апробация способа оценки причин ошибочной классификации; (5) разработка, валидация и апробация методологии оценки клинической точности программ ИИ с учетом онкологической настороженности; (6) разработка, валидация и апробация способа оценки эффективности диагностики с применением алгоритмов маршрутизации на основании заключений моделей ИИ; (7) разработка, валидация и апробация способа оценки медико-экономической эффективности (МЭЭ) диагностических алгоритмов на уровне учреждения и региона; (8) разработка системы поддержки принятия врачебных решений

(СППВР) «Derma Check» и модуля непрерывного профессионального развития «Derma Check Educational Module».

2.2. Дизайн и методология опросных исследований

2.2.1. Общая характеристика опросных исследований

В рамках диссертационного исследования проведены три самостоятельных анонимных онлайн-опроса среди врачей трех специальностей: терапевтов и врачей общей практики (ВОП), врачей-дерматовенерологов и врачей-онкологов. Все опросы проводились в единый период наблюдения с марта по июль 2026 года на единой платформе Anketolog.ru, что обеспечивало сопоставимость данных и возможность корректного межгруппового анализа. Суммарный объем трех выборок составил 545 специалистов. Ссылка на анкету распространялась по профессиональным коммуникационным каналам, участие было добровольным.

2.2.2. Инструмент исследования

Для всех трех опросов применялась единая анкета, состоящая из 26 закрытых вопросов с одним или множественными вариантами ответов. Блок I (10 вопросов) фиксировал социально-демографические и профессиональные характеристики: пол, возраст, регион, уровень образования, ученую степень, стаж, тип медицинской организации, места работы, уровень владения ПК, характеристику населенного пункта, специальность. Блок II (16 вопросов) охватывал осведомленность об ИИ, опыт применения ИИ-технологий в медицинской и немедицинской деятельности, конкретные технологии ИИ, доверие к заключениям ИИ, готовность к применению, субъективную оценку точности диагностики системами ИИ и врачами-специалистами, ожидаемые преимущества и проблемы, а также целесообразные направления применения ИИ. Ответы по шкалам отношения к ИИ

кодировались в 4-балльную порядковую шкалу (1 – минимальное значение, 4 – максимальное).

2.2.3. Опрос врачей-терапевтов и врачей общей практики

Опрос проводился в марте–мае 2026 г. Критерием включения являлось указание специальности «Терапия» или «Общая врачебная практика (семейная медицина)». Итоговая выборка составила 105 анкет: 82 специалиста по специальности «Терапия», 23 – по специальности «Общая врачебная практика», 4 указали обе специальности одновременно. Среди опрошенных 75 (71,4%) составили женщины, 30 (28,6%) мужчины. Средний возраст $40,2 \pm 17,3$ года (медиана 35,0; диапазон 20–80 лет). Специалисты до 35 лет, 49 (46,7%), 35–50 лет, 22 (21,0%), старше 50 лет, 34 (32,4%). Ученые степени имелись у 4 (3,8%) респондентов. По уровню владения ПК: базовый уровень, 66 (62,9%), продвинутые пользователи, 29 (27,6%). Большинство специалистов, 70 из 105 (66,7%), работали в амбулаторно-поликлинических медицинских организациях.

2.2.4. Опрос врачей-дерматовенерологов

Опрос проводился в марте–июле 2026 г. Из общей базы ответов отобраны 191 респондент, указавший специальность «Дерматовенерология» и/или «Косметология». Респонденты, отметившие одновременно обе специальности ($n = 22$), включались в анализ как дерматовенерологи; респонденты, указавшие только «Косметология» без «Дерматовенерология» ($n = 6$), включены в общую выборку, поскольку в российской практике косметология является смежной специальностью.

Из 191 специалиста 164 (85,9%) составили женщины, 27 (14,1%), мужчины. Средний возраст $34,4 \pm 12,1$ года, медиана 29 лет (диапазон 21–70). Молодой медианный возраст объясняется тем, что 50,8% выборки составили ординаторы. Ученую степень кандидата имели 29 (15,7%), доктора, 6 (3,2%). 84,9% проживали

в городах-миллионниках. 28 (14,7%) респондентов дополнительно отметили специальность «Косметология».

Для корректного попарного сравнения ординаторов и практикующих врачей сформированы сбалансированные подгруппы. Из 185 респондентов 84 указали статус «Ординатор» без совмещения с практикой, 62, «Практикующий врач» без статуса ординатора 13, оба статуса одновременно 26, ни один из двух (аспиранты, преподаватели, научные работники). Для парного сравнения из 84 ординаторов методом простой случайной выборки отобраны 62, что уравнивает подгруппы ($n = 94 + 94$). Респонденты с двойным статусом исключены из подгруппового анализа, но сохранены в анализе полной выборки.

2.2.5. Опрос врачей-онкологов

В рамках мультидисциплинарного анонимного онлайн-опроса, проводившегося на платформе Anketolog.ru в марте–июле 2026 г., из общей базы ответов были отобраны респонденты, отметившие специальность «Онкология». Итоговая выборка составила 249 анкет.

Из 249 опрошенных онкологов 161 (64,7%) составили женщины. Средний возраст $37,2 \pm 11,0$ года, медиана 33 года, диапазон от 18 до 70 лет. Специалисты моложе 35 лет – 134 (54,3%), 35–50 лет 74 (30,0%), старше 50 лет 39 (15,8%). Без ученой степени 225 (90,4%), кандидатов медицинских наук 19 (7,6%), докторов 5 (2,0%). По типу организации: 198 (79,5%) работали в амбулаторно-поликлинических, 41 (16,5%) в стационарных медицинских организациях, 6 (2,4%) в научных, 4 (1,6%) в образовательных. Московский регион представлен 241 (96,8%) респондентом. По уровню владения ПК: офисными программами 150 (60,2%), продвинутые пользователи – 82 (32,9%), «только печатаю» 10 (4,0%), уровень программиста 7 (2,8%).

2.2.6. Статистические методы анализа данных опросов

Во всех трех опросных исследованиях рассчитывались средние значения (M) и стандартные ошибки средней (SE). Для сравнения двух независимых групп применялся критерий Манна–Уитни (U); для сравнения трех и более групп критерий Краскела–Уоллиса (H). Сравнение категориальных переменных выполнялось по критерию хи-квадрат (χ^2). Корреляция порядковых переменных оценивалась коэффициентом ранговой корреляции Спирмена (ρ). Уровень статистической значимости $p < 0,05$. Кластерный анализ проводился методом порогового агрегирования: среднее арифметическое трех показателей (доверие, готовность, применение ИИ) распределяло каждого респондента в одну из трех групп «Скептики» (1,00–1,67 балла), «Умеренные» (1,68–2,67 балла), «Энтузиасты» (2,68–4,00 балла). Статистическая обработка выполнена в Python 3.12 (библиотеки `pandas`, `scipy`).

2.3. Характеристики программы «Derma Check»

Система поддержки принятия врачебных решений (СППВР) «Derma Check» – мультимодульное веб-приложение для автоматизированной визуальной идентификации и предварительной дифференциальной диагностики новообразований и другой патологии кожи в условиях первичного звена здравоохранения. Программа не ставит окончательный диагноз и не подменяет гистологическую верификацию, но обеспечивает врача объективной количественной оценкой вероятности злокачественности, визуализацией зон риска и структурированными рекомендациями по тактике дальнейшего обследования.

Программа реализована в клиент-серверной архитектуре. Серверная часть построена на фреймворке FastAPI (Python 3.10+) с использованием ONNX Runtime и TensorFlow Lite для инференса нейросетевых моделей на центральном процессоре без графического ускорителя. Клиентская часть – адаптивное одностраничное веб-приложение (HTML5/CSS3/JavaScript) без внешних

фреймворков, работоспособное на мобильных устройствах с низкой пропускной способностью канала. Технические характеристики: серверный фреймворк FastAPI (Python 3.10+), uvicorn, nginx; инференс моделей ONNX Runtime (CPU), TensorFlow Lite; хранилище данных SQLite и файловая система; минимальная RAM 6 ГБ; ОС Ubuntu Linux (VPS).

Диагностический конвейер обработки фотоизображения включает последовательные этапы: оценка качества фотоизображения, коррекция яркости, входной фильтр, параллельный инференс нескольких нейросетевых моделей классификации, объединение результатов консенсусом, а при наличии данных анамнеза – дополнительный этап интеграции клинических данных. Параллельно основному конвейеру запускаются подсистемы сегментации, PCA-анализа асимметрии, объяснимого ИИ (XAI) и многоклассовой классификации.

Перед запуском классификации система проводит автоматическую оценку качества поступившего фотоизображения по четырем параметрам: резкость, яркость, пересвет (доля пикселей с яркостью выше 95-го перцентиля) и общая экспозиция. Метрика очаговой резкости рассчитывается по патчам внутри автоматически определенной маски очага в цветовом пространстве CIELAB. При обнаружении критически низкого качества система блокирует анализ и информирует врача с указанием конкретной причины. Для клинических фотоизображений дополнительно применяется коррекция яркости гамма-функцией с целевым значением 0,5.

Диагностическое ядро составляют несколько нейросетевых моделей, организованных в параллельные решающие модули. Каждая модель обучена на собственном наборе данных и решает задачу бинарной классификации. Все модели проходят через Test-Time Augmentation (TTA) с последующим усреднением вероятностей; дополнительно выполняется множественный инференс с кропами (по умолчанию 9 микросдвигов). Один из модулей реализует мягкое голосование по средней вероятности двух компонентов «Меланома» и «Злокачественные новообразования кожи». Второй модуль реализует алгоритм маршрутизации на основе шестнадцатипольной квадрантной матрицы, где ось абсцисс – вероятность

меланомы, ось ординат вероятность злокачественного новообразования кожи. Итоговое решение формируется голосованием большинства; при полном расхождении устанавливается средний риск с уведомлением о расхождении.

Модуль анализа анамнестических данных интегрирует клинические данные пациента с результатами компьютерного зрения. Врач вносит следующие параметры: возраст, пол, фототип кожи по шкале Фицпатрика (I–VI), наличие и характер субъективных симптомов (зуд, кровоточивость, изъязвление), размеры образования и динамика их изменения, отягощенный онкологический анамнез, иммуносупрессия, локализация. Предсказание риска выполняется обученными моделями машинного обучения, результат которых объединяется с результатами компьютерного зрения в единую консенсусную оценку через систему из 12 правил интеграции.

Подсистемы визуализации и объяснимого ИИ включают: сегментацию образования (алгоритмический вариант предобработка, построение карты значимости, бинаризация методом Оцу, постобработка; нейросетевой вариант – детектор YOLO для локализации, модель SAM для пиксельной маски), модуль объяснимого ИИ (XAI) с тепловыми картами Score-CAM и картами чувствительности Occlusion Sensitivity, PCA-анализ асимметрии образования, многоклассовую классификацию с отображением топ-5 наиболее вероятных нозологий, кабинет врача с функцией динамического наблюдения образований в сравнительном режиме. Маршрутизация пациента определяется специальностью обращающегося врача: для онколога и дерматовенеролога формируется рекомендация о биопсии, для остальных специальностей – направление к специалисту соответствующего уровня.

2.4. Описание модуля непрерывного профессионального развития «Derma Check Educational Module»

Модуль непрерывного профессионального развития «Derma Check Educational Module» является интегрированной образовательной подсистемой

СППВР «Derma Check», разработанной для повышения квалификации врачей в области визуальной диагностики новообразований кожи. Модуль построен на четырех методических принципах: (1) принцип немедленной обратной связи – результат ответа на тестовый вопрос отображается мгновенно, без ожидания завершения теста; (2) принцип привязки к контексту – каждый тестовый вопрос и клиническая задача содержат ссылку на описание правильной нозологии; (3) принцип разделения проверочного и учебного форматов – тест проверяет узнавание, задачи формируют клиническое мышление; (4) охват всех уровней когнитивной деятельности в визуальной диагностике – от распознавания (тест) через анализ (задачи) к систематизации (атлас).

Модуль включает четыре компонента. Компонент «Хаб» – посадочная страница с банком тестов и задач, выбором активности. Компонент «Тест по нозологиям» реализует формат визуальной идентификации: врачу предъявляется фотоизображение кожного новообразования и предлагается выбрать верный диагноз из четырех вариантов. Перед началом теста врач настраивает количество вопросов (5, 10, 15 или 20) и фильтр по злокачественности. Набор вопросов формируется случайным образом из общего банка. Варианты ответа формируются на сервере с соблюдением принципа исключения нозологий одной линии (например, «Меланома» и «Меланома узловая» не могут одновременно присутствовать среди вариантов, что исключало бы семантический, а не клинический выбор). По завершении теста отображается итоговый счет с качественной оценкой: «Отличный результат!» (80% и более), «Неплохо, есть куда расти» (50–79%) или «Стоит повторить материал» (менее 50%).

Компонент «Клинические задачи» моделирует ситуацию осмотра пациента: врачу предъявляется фотоизображение образования, развернутый анамнез (пол, возраст, локализация, динамика, жалобы) и список вопросов (обычно 2–4). Врач формулирует ответ в свободной форме в текстовом поле, после чего раскрывает эталонный ответ для самопроверки; эталон содержит диагноз и его клиническое обоснование. Отсутствие автоматической оценки свободного ответа является методическим решением: задачи нацелены на формирование клинического

мышления, а не на проверку запоминания терминов. Компонент «Атлас нозологий» справочная подсистема с полнотекстовым поиском; описание каждой нозологии включает разделы: определение, клинические признаки, дифференциальная диагностика, дерматоскопические признаки, тактика. Все компоненты используют единый справочник нозологий, общий с основным диагностическим ядром программы, что обеспечивает консистентность терминологии.

Интеграция модуля с основным диагностическим ядром реализована на нескольких уровнях. На уровне навигации: на главном экране программы размещен блок «Профессиональное развитие» с карточками, ведущими в соответствующие компоненты. На уровне клинического контекста: при расхождении заключения модели ИИ и мнения врача система предлагает перейти к описанию конкретной нозологии; описания доступны также из блока многоклассовой классификации. На уровне данных: единый справочник нозологий обеспечивает взаимную согласованность образовательного и диагностического компонентов.

2.5. Формирование набора данных фотоизображений

В ходе диссертационного исследования для разработки, валидации и апробации методологических подходов использовались клинические данные и фотоизображения новообразований кожи 381 пациента. Среди них: 291 пациент с доброкачественными новообразованиями кожи (ДНО) 100 немеланоцитарных и 191 меланоцитарная опухоль и 90 пациентов со злокачественными новообразованиями кожи (ЗНО) 39 меланом и 51 базальноклеточный рак кожи. Данная структура отражает реальную структуру обращений пациентов с новообразованиями кожи к врачам первичного и специализированного звеньев здравоохранения. Материал получен в ходе двух ранее выполненных клинических исследований точности программ ИИ «Derma Onko Check» и «Melanoma Check»; объединение данных в единую выборку позволило сформировать датасет, применявшийся для разработки методологии оценки эффективности диагностических алгоритмов.

Для оценки влияния типа устройства фотофиксации на точность классификации была сформирована подвыборка попарного сравнения: 125 пар изображений, где каждое новообразование было зафиксировано двумя способами: клиническим фотоизображением (макроскопический снимок без увеличения) и дерматоскопическим фотоизображением (10–20-кратное увеличение, поляризованное освещение). В группу злокачественных новообразований этой подвыборки вошли 16 верифицированных случаев ЗНО кожи, в группу доброкачественных 109 образований, включающих меланоцитарные (невусы различного типа), кератиноцитарные/эпидермальные (себорейные кератомы, папилломы), сосудистые (гемангиомы) и фиброзные (дерматофибромы) новообразования.

Для оценки клинической точности программы Derma Onko Melanoma Check New использовалась отдельная выборка из 342 новообразований кожи. Верификация образований осуществлялась с применением принципа двойного эталона истины: (1) клинико-дерматоскопическая экспертная оценка три независимых врача-дерматовенеролога со стажем работы в дерматоонкологии более 20 лет проводили клинический осмотр и дерматоскопическое исследование каждого образования, консенсусное заключение принималось при совпадении не менее двух из трех; (2) морфологическая верификация для ЗНО обязательным условием являлись результаты прижизненного патологоанатомического исследования биопсийного или операционного материала.

Для сравнительного анализа международных баз данных и методологии формирования датасетов в исследовании также использовались публично доступные базы данных, в том числе материалы ISIC Archive (International Skin Imaging Collaboration).

2.6. Методология анализа причин ошибочной классификации

Для систематического изучения причин ошибочной классификации изображений новообразований кожи программами ИИ была разработана

методология, включающая три взаимосвязанных элемента: вычисление фотометрических и текстурных метрик качества изображений, многоуровневый статистический анализ и методы объяснимого искусственного интеллекта.

2.6.1. Фотометрические и текстурные метрики качества изображений

Для каждого фотоизображения рассчитывались следующие метрики: яркость (brightness), контраст (contrast), резкость (sharpness), энтропия (entropy) мера информационной насыщенности, размытие по быстрому преобразованию Фурье (fft_blur), доля переэкспонированных пикселей (clipped_high), доля недоэкспонированных пикселей (clipped_low), средние значения и стандартные отклонения каналов RGB (mean_r, mean_g, mean_b; std_r, std_g, std_b), высота и ширина изображения (height, width). Метрика очаговой резкости рассчитывалась по патчам внутри автоматически определенной маски очага в цветовом пространстве CIELAB, что устраняло ложные срабатывания из-за текстуры кожного покрова за пределами образования. Значения метрик сравнивались между четырьмя категориями классификационных результатов: истинно положительные (ИП), истинно отрицательные (ИО), ложноположительные (ЛП) и ложноотрицательные (ЛО).

2.6.2. Статистический анализ межгрупповых различий

Статистическая значимость межгрупповых различий по метрикам оценивалась следующими методами: (1) однофакторный дисперсионный анализ ANOVA для оценки значимости различий между четырьмя категориями по каждой метрике; (2) попарное сравнение категорий t-критерий Уэлча; (3) двухфакторный дисперсионный анализ ANOVA для совместной оценки влияния фактора «категория заключения» и фактора «программа компьютерного зрения» на каждую метрику; (4) многомерный дисперсионный анализ MANOVA с критериями Wilks' lambda, Pillai's trace и Hotelling–Lawley trace для оценки значимости различий на

уровне совокупного вектора метрик; (5) корреляционный анализ взаимосвязей между метриками (коэффициент Пирсона r); (6) множественная логистическая регрессия с исключением предикторов с высокой мультиколлинеарностью (VIF) для количественного ранжирования вклада метрик в вероятность ошибочного заключения.

При построении множественной логистической регрессии после исключения предикторов с VIF 10,2–931,1 в окончательную модель вошли пять независимых переменных: резкость, энтропия, размытие по FFT, доля переэкспонированных и доля затемненных пикселей.

2.6.3. Методы объяснимого искусственного интеллекта

Для визуализации внутренней логики классификационных решений нейронных сетей применялись два метода. Метод Score-CAM (Score-weighted Class Activation Maps) использовался для построения тепловых карт, отображающих зоны максимальной значимости (активации) сети при формировании конкретного классификационного решения; метрика значений тепловой карты от 0 до 1, зоны максимальной значимости соответствуют значениям 0,6–1,0. Метод Occlusion Sensitivity предполагал систематическое перекрытие отдельных областей входного изображения с регистрацией изменения выходной вероятности для определения критических пикселей. Совместное применение обоих методов позволяло не только локализовать зону интереса модели, но и верифицировать механизм ошибки: смещение активации на нерелевантные области (нормальная кожа, артефакты) или ложную активацию на морфологически имитирующих признаки ЗНО участках доброкачественных образований. Для наглядного описания ошибок строились карты пространственного распределения различий (карты различий) между попарно сравниваемыми категориями классификационных результатов.

2.7. Методология оценки клинической точности

Стандартная методология оценки точности диагностических алгоритмов, основанная на матрице ошибок с единственным эталоном истины (морфологический диагноз), применительно к программам ИИ в дерматоонкологии систематически занижает клиническую ценность алгоритмов: она расценивает как ошибку случаи обоснованной онкологической настороженности ситуации, когда программа правомерно устанавливает высокий или средний риск для образования, оказавшегося доброкачественным по морфологии. Для устранения этого недостатка разработана трехуровневая система оценки.

Первый уровень формальная (техническая) точность соответствует стандартной методологии: вычисляются чувствительность (Se), специфичность (Sp) и общая точность (Ac) по стандартной матрице ошибок с морфологическим диагнозом в качестве единственного эталона истины.

Второй уровень клиническая эффективность предполагает категоризацию каждого заключения программы по четырем качественным типам в сравнении с независимым заключением врача-эксперта. Данный уровень оценивает не правильность распознавания морфологического типа, а согласованность программы с клиническим мышлением эксперта.

Третий уровень клиническая точность количественная оценка на основании матрицы клинических случаев с двойным эталоном истины: клинικο-дерматоскопическая экспертная оценка (консенсус трех независимых врачей-дерматовенерологов) и морфологическая верификация. Матрица клинических случаев включает четыре категории: JC (justified conclusion, обоснованное заключение) заключение программы о высоком или среднем риске совпало с предварительным и заключительным диагнозом и/или с морфологическим подтверждением, что подтвердило необходимость направления на биопсию; UC (unjustified conclusion, необоснованное заключение) заключение о высоком или среднем риске, но врач расценил образование как ДНО либо морфология установила ДНО без дисплазии; MJC (missed justified conclusion, пропущенное

обоснованное заключение) заключение программы «низкий риск», тогда как врач заподозрил ЗНО и морфология его подтвердила; NMJC (not missed justified conclusion, непропущенное обоснованное заключение) заключение «низкий риск», и врач расценил образование как ДНО, либо врач заподозрил ЗНО, но морфология установила ДНО без дисплазии.

На основании матрицы рассчитывались три показателя клинической точности по формулам: клиническая чувствительность $Se_{cl} = JC / (JC + MJC) \times 100\%$; клиническая специфичность $Sp_{cl} = NMJC / (NMJC + UC) \times 100\%$; клиническая точность $Ac_{cl} = (JC + NMJC) / N \times 100\%$. Доверительные интервалы рассчитывались методом Уилсона.

Надежность экспертного эталона подтверждалась оценкой межэкспертной согласованности: коэффициент каппа Коэна (κ) рассчитывался для каждой пары из трех специалистов; асимметрия расхождений проверялась тестом Макнемара. Клинические испытания выполнялись на выборке из 342 новообразований кожи.

2.8. Методология оценки эффективности диагностики с использованием алгоритмов на основании заключений моделей искусственного интеллекта

Для количественной оценки прироста диагностической эффективности при встраивании программ ИИ в систему маршрутизации пациентов разработана оригинальная методология, основанная на моделировании двух альтернативных сценариев маршрутизации на одном клиническом материале (данные 381 пациента: 291 с ДНО и 90 со ЗНО кожи).

Традиционный сценарий воспроизводил стандартный трехэтапный путь пациента в соответствии с действующими нормативными правовыми актами: ВОП/терапевт, врач-дерматовенеролог, врач-онколог. Сценарий с применением ИИ моделировал двухпотокую маршрутизацию: при вероятности ЗНО по заключению модели, равной или превышающей пороговое значение алгоритма маршрутизации, прямое направление к врачу-онкологу; при вероятности ниже порога направление к врачу-дерматовенерологу. Оптимальный порог алгоритма

маршрутизации определялся анализом ROC-кривых по максимуму индекса Юдена (сумма чувствительности и специфичности минус единица). Для программы Derma Onko Check, использованной в приводимом примере, оптимальный порог составил 62%.

Для сравнительной оценки двух сценариев предложены два оригинальных коэффициента. Коэффициент прироста выявляемости (R_DAI) определяется как отношение числа всех выявленных случаев целевого заболевания при диагностике с применением моделей ИИ (N_Dai) к числу выявленных случаев без их применения (N_Dnai): $R\_DAI = N\_Dai / N\_Dnai$. Числитель складывается из двух непересекающихся потоков: истинно положительные случаи с вероятностью не ниже порога, которые направляются прямо к онкологу (для простоты расчета чувствительность онколога условно принималась равной 100%), и случаи с вероятностью ниже порога (как ИП, так и ЛО по заключению модели), которые направляются к дерматовенерологу и чья выявляемость определяется чувствительностью дерматовенеролога. Знаменатель отражает выявляемость при традиционном маршруте, где все случаи проходят этапы последовательно, а итоговая выявляемость есть произведение чувствительностей врачей на каждом этапе. Значение $R\_DAI > 1$ свидетельствует о превосходстве сценария с ИИ по выявляемости.

Коэффициент сокращения пропусков (R_NDAI) определяется симметрично: $R\_NDAI = N\_NDnai / N\_NDai$, где N_NDnai число пропущенных случаев при традиционной диагностике, N_NDai число пропущенных при диагностике с применением ИИ. Значение $R\_NDAI > 1$ свидетельствует о сокращении числа недиагностированных случаев. Для расчета коэффициентов разработана программа для ЭВМ (свидетельство о государственной регистрации № 2026613742 «CalcRDAI&RNDIAI»: заявл. 10.02.2026, опублик. 10.02.2026).

2.9. Методология оценки медико-экономической эффективности диагностики

Методология оценки медико-экономической эффективности (МЭЭ) диагностики с применением алгоритмов маршрутизации разработана применительно к двум уровням: уровню медицинской организации и региональному уровню. В качестве модельного целевого заболевания (ЦЗ) выбрана меланома кожи как наиболее клинически и экономически значимое ЗНО кожи. Расчеты выполнены на тарифах обязательного медицинского страхования (ОМС) г. Москвы.

Для сравнительной оценки двух сценариев (с применением ИИ и без него) разработана система оригинальных расчетных формул (формулы 1–6), учитывающая следующие виды финансовых затрат (ФЗ): предотвратимые ФЗ на диагностику ложноположительных (ЛП) случаев (АС) – затраты на обследование пациентов без ЦЗ, ошибочно направленных к специалистам и на биопсию; обоснованные ФЗ на диагностику истинно отрицательных (ИО) случаев; потенциально предотвратимые ФЗ на недиагностированные («пропущенные») случаи ЦЗ (РАС) – затраты на лечение пациентов с поздно выявленными ЗНО. Стоимость лечения пропущенных случаев меланомы (СрР) рассчитывалась как взвешенная средняя стоимость лечения по стадиям: $\text{СрР} = (0,852 \times 403\,021,60) + (0,148 \times 2\,787\,014,87) = 755\,852,60 \text{ руб.}$

Коэффициент соотношения суммарных ФЗ (R_{TC}) рассчитывался как отношение совокупных затрат при диагностике с ИИ (TC_{ai}) к совокупным затратам при диагностике без ИИ (TC_{nai}): $R_{\text{TC}} = \text{TC}_{\text{ai}} / \text{TC}_{\text{nai}}$. Значение $R_{\text{TC}} < 1$ свидетельствует об экономической эффективности применения ИИ. В расчетах учитывались только прямые медицинские затраты согласно тарифам ОМС.

Для оценки МЭЭ на региональном уровне разработана методология экстраполяции с уровня выборки на уровень административной территории с использованием данных официальной онкологической статистики числа первичных случаев меланомы кожи (С43) по г. Москве за 2024 год (1470 больных). Дополнительно предложена формула расчета годовых эквивалентных затрат на

программу ИИ как медицинское изделие (МИ с ИИ) на основе метода аннуитетных платежей (capital recovery factor, CRF), учитывающая капитальные вложения (C_d), ежегодные операционные расходы (C_o) и косвенные, транзакционные, скрытые затраты (C_{iht}). В настоящем исследовании данная формула не применялась для расчетов ввиду отсутствия верифицированных данных о соответствующих затратах, однако представлена как методологический инструмент для проспективных исследований. Для комплексной визуализации соотношения клинической и экономической эффективности применялась квадрантная матрица медико-экономической эффективности, позиционирующая каждый исследованный алгоритм маршрутизации в системе координат R_{DAI} и R_{TC} . Сокращение сроков установления диагноза оценивалось путем сравнения средней длительности диагностического пути при каждом сценарии.

2.10. Статистические методы

В диссертационном исследовании применялся комплекс статистических методов, определявшийся характером анализируемых данных и задачами каждого раздела.

Описательная статистика. Количественные непрерывные переменные представлены как среднее арифметическое и стандартное отклонение ($M \pm SD$) или стандартная ошибка средней ($M \pm SE$), медиана (Me) и межквартильный размах [$Q1$; $Q3$], а также 95% доверительный интервал (95% ДИ). Категориальные переменные описаны в виде абсолютных и относительных (процентных) частот.

Методы сравнения групп. Для сравнения двух независимых групп по непрерывным переменным применялись t-критерий Уэлча и критерий Манна–Уитни (U). Для сравнения нескольких независимых групп, однофакторный дисперсионный анализ ANOVA и критерий Краскела–Уоллиса (H). Для сравнения долей, критерий хи-квадрат (χ^2) и z-тест для сравнения пропорций. Для попарного сравнения зависимых бинарных классификаторов на одних и тех же объектах применялся тест Макнемара.

Методы корреляционного анализа. Связь между непрерывными переменными оценивалась коэффициентом ранговой корреляции Спирмена (ρ) и коэффициентом линейной корреляции Пирсона (r).

Многофакторные методы. Двухфакторный дисперсионный анализ ANOVA применялся для оценки совместного влияния двух факторов на непрерывный исход; многомерный дисперсионный анализ MANOVA (критерии Wilks' lambda, Pillai's trace, Hotelling–Lawley trace) для оценки многомерных межгрупповых различий по совокупному вектору метрик; множественная логистическая регрессия с исключением коллинеарных предикторов (критерий VIF) – для построения модели вероятности ошибочного заключения. Показатель согласия кластерных распределений между подгруппами оценивался по критерию χ^2 . Межэкспертная согласованность оценивалась по коэффициенту каппа Коэна (κ) с интерпретацией по шкале Ландиса–Коха; асимметрия расхождений тестом Макнемара.

Диагностические метрики. Для оценки диагностической точности рассчитывались: чувствительность (Se), специфичность (Sp), общая точность (Ac), прогностическая ценность положительного (PPV) и отрицательного (NPV) результата, F1-мера, коэффициент согласия Коэна (Cohen's Kappa, κ), отношения правдоподобия положительного (LR+) и отрицательного (LR–) результата, площадь под ROC-кривой (AUC-ROC), средняя точность (Average Precision, AP) как площадь под Precision–Recall кривой, оценка Брие (Brier score). Оптимальный порог бинарной классификации определялся по максимуму индекса Юдена. Доверительные интервалы для долей рассчитывались методом Уилсона.

Для оценки размера эффекта при попарном сравнении групп использовался d-критерий Коэна с интерпретацией по общепринятой шкале: $d < 0,2$ незначительный, $0,2–0,5$ малый, $0,5–0,8$ средний, $> 0,8$ большой эффект. Согласованность непрерывных вероятностей между модальностями оценивалась внутриклассовым коэффициентом корреляции ICC(2,1). Для оценки качества бинарной классификации помимо основных метрик рассчитывались коэффициент корреляции Мэтьюса (MCC, Matthews Correlation Coefficient) сбалансированная

мера, принимающая значения от -1 до $+1$, и псевдо-коэффициент детерминации Макфаддена для моделей логистической регрессии. Сравнение чувствительности программы и экспертов на малых выборках проводилось с применением точного критерия Фишера. Доверительные интервалы для диагностических долей рассчитывались как методом Уилсона, так и методом Клоппера–Пирсона. Для визуальной оценки выборочной неопределенности диагностических показателей строились графики биномиального распределения числа верных классификаций $X \sim \text{Bin}(n, p)$.

Уровень статистической значимости во всех анализах принимался равным $p < 0,05$. Статистическая обработка данных выполнена в Python 3.12 (библиотеки `pandas`, `scipy`, `statsmodels`, `sklearn`) и с применением программы для ЭВМ «CalcRDAI&RNDIAI» (свидетельство о государственной регистрации № 2026612771).

ГЛАВА 3. РЕЗУЛЬТАТЫ СОБСТВЕННЫХ ИССЛЕДОВАНИЙ

3.1. Исследование уровня доверия к технологиям искусственного интеллекта и готовности его применять российскими врачами

3.1.1. Описание онлайн-опроса

В марте–июле 2026 г. на платформе Anketolog.ru был проведен масштабный мультидисциплинарный анонимный онлайн-опрос врачей-специалистов. Ссылка на анкету распространялась по профессиональным коммуникационным каналам, участие было добровольным.

Анкета состояла из 26 закрытых вопросов с одним или множественными вариантами ответов, объединенных в два смысловых блока. Блок I включал 10 вопросов о социально-демографических и профессиональных характеристиках. Блок II содержал 16 вопросов для оценки осведомленности об ИИ, опыта применения ИИ-технологий в медицинской и немедицинской деятельности, доверия к заключениям ИИ, готовности к применению, субъективной оценки точности диагностики, ожидаемых преимуществ и проблем, целесообразных направлений применения.

Ответы по шкалам отношения к ИИ кодировались в 4-балльную порядковую шкалу (1 – минимальное значение, 4 – максимальное). Вычислялись средние значения и стандартные ошибки средней ($M \pm SE$). Для сравнения двух групп применяли критерий Манна–Уитни (U), трех и более – критерий Краскела–Уоллиса (H). Связь порядковых переменных оценивали коэффициентом ранговой корреляции Спирмена (ρ). Уровень значимости – $p < 0,05$. Кластерный анализ проводился методом порогового агрегирования по среднему значению трех показателей (доверие, готовность, применение ИИ): «Скептики» 1,00–1,67 балла, «Умеренные» 1,68–2,67, «Энтузиасты» 2,68–4,00. Статистическая обработка выполнена в Python 3.12 (pandas, scipy).

3.1.2. Исследование уровня доверия к технологиям искусственного интеллекта и готовности его применять врачами общей практики и терапевтами

Характеристика выборки

Критерием включения являлось указание специальности «Терапия» или «Общая врачебная практика (семейная медицина)» в соответствующем разделе анкеты. Итоговая выборка составила 105 анкет: 82 специалиста по специальности «Терапия», 23 по специальности «Общая врачебная практика», 4 указали обе специальности одновременно.

Женщин среди опрошенных было 75 (71,4%), мужчин – 30 (28,6%). Средний возраст составил $40,2 \pm 17,3$ года (медиана 35,0; диапазон 20–80 лет). Специалисты до 35 лет были представлены 49 (46,7%) врачами, 35–50 лет – 22 (21,0%), старше 50 лет – 34 (32,4%) врачами. Ученые степени имелись у 4 (3,8%) респондентов.

По уровню владения ПК большинство отмечало базовый уровень (работу с офисными программами) – 66 (62,9%) врачей, группу «продвинутых» пользователей составили 29 (27,6%). Большинство специалистов (70 из 105, 66,7%) работало в амбулаторно-поликлинических медицинских организациях, оказывающих первичную медико-санитарную помощь (Таблица 1).

Таблица 1 – Социально-демографическая и профессиональная характеристика выборки врачей-терапевтов и врачей общей практики (n = 105)

Показатель	Категория	Абс. число (n)	Доля (%)
Пол	Женщины	75	71,4
	Мужчины	30	28,6
Возраст	До 35 лет	49	46,7
	35–50 лет	22	21,0
	Старше 50 лет	34	32,4
Ученая степень	Без ученой степени	101	96,2
	Кандидат медицинских наук	3	2,9
	Доктор медицинских наук	1	1,0
Уровень владения ПК	Минимальный	7	6,7
	Базовый (офисные программы)	66	62,9
	Продвинутый	29	27,6

Продолжение Таблицы 1

	Программист	3	2,9
Специальность	Терапия	82	78,1
	Общая врачебная практика (семейная медицина)	23	21,9

Применение искусственного интеллекта в медицинской и немедицинской деятельности

Не применяют ИИ в профессиональной медицинской деятельности 33 (31,4%) специалиста; применяют иногда 28 (26,7%), периодически 27 (25,7%), часто 17 (16,2%) врачей. Таким образом, подавляющее большинство врачей (68,6%) в той или иной мере обращаются к технологиям ИИ в рамках профессиональной деятельности.

В немедицинской (бытовой) деятельности не используют ИИ лишь 28 (26,7%) респондентов. Разрыв между бытовым и профессиональным неприменением составляет около 5 процентных пунктов (п.п.), что существенно меньше, чем в ряде других специальностей. Такой небольшой разрыв отражает более широкое проникновение ИИ-инструментов в повседневную практику первичного звена здравоохранения, где электронные медицинские карты, телемедицинские сервисы и цифровые платформы уже стали привычным элементом рабочего процесса (Рисунок 1).

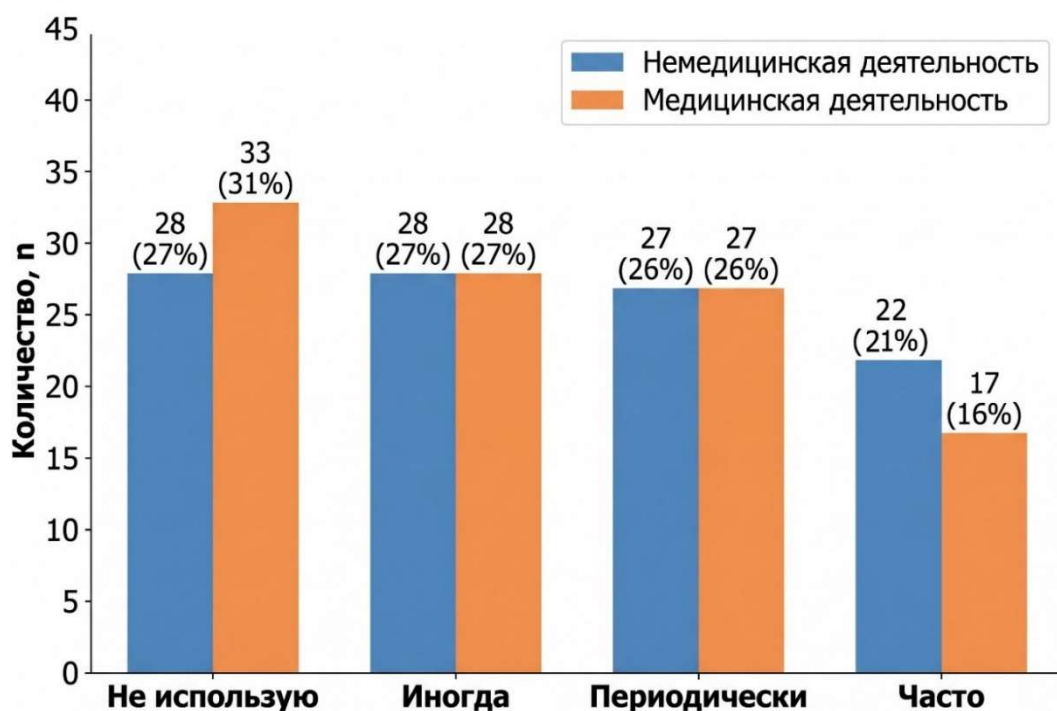


Рисунок 1 – Применение технологий ИИ в медицинской и немедицинской деятельности врачами-терапевтами и врачами общей практики (n = 105)

Осведомленность, доверие, готовность к применению искусственного интеллекта и оценка возможности применения искусственного интеллекта

Полностью осведомлены о возможностях ИИ в медицине 16 (15,2%) врачей; в значительной степени – 48 (45,7%); незначительно – 37 (35,2%); не осведомлены 4 (3,8%) специалиста. Суммарно достаточный уровень осведомленности декларируют 60,9% участников опроса.

Уровень доверия к заключениям ИИ оказался существенно ниже уровня осведомленности, не доверяют 19 (18,1%), доверяют незначительно 49 (46,7%) врачей, то есть суммарно скептически настроены 68 (64,8%) врачей. Доверяют в значительной степени 30 (28,6%), полностью – 7 (6,7%) специалистов. Такое распределение свидетельствует о том, что информированность о существовании ИИ-технологий не трансформируется автоматически в профессиональное доверие к ним.

Готовность к личному применению ИИ распределилась следующим образом: не готовы лишь 13 (12,4%), незначительно готовы 42 (40,0%), готовы в

значительной степени 33 (31,4%), полностью готовы 17 (16,2%) врачей. Суммарно 87,6% врачей выразили готовность использовать ИИ в профессиональной практике. Диссонанс между относительно низким доверием (35,2% доверяющих) и высокой декларируемой готовностью (87,6%) отражает разграничение между суждением о технологии и практическим намерением ее использовать. Врачи готовы попробовать ИИ, но пока не считают его заслуживающим полного доверия. Этот феномен «не доверяю, но готов попробовать» характерен для этапа раннего знакомства с любой новой технологией, специалисты понимают, что новый инструмент может оказаться полезным, но им не хватает собственного клинического опыта для формирования обоснованного суждения. Готовность в данном случае отражает не столько доверие, сколько профессиональную открытость и интеллектуальное любопытство.

Что касается принципиальной возможности использования ИИ в медицинской деятельности, то считают применение возможным с незначительными оговорками 43 (41,0%), со значительными оговорками 38 (36,2%), безоговорочно 17 (16,2%), невозможным – лишь 7 (6,7%) врачей (Рисунок 2).

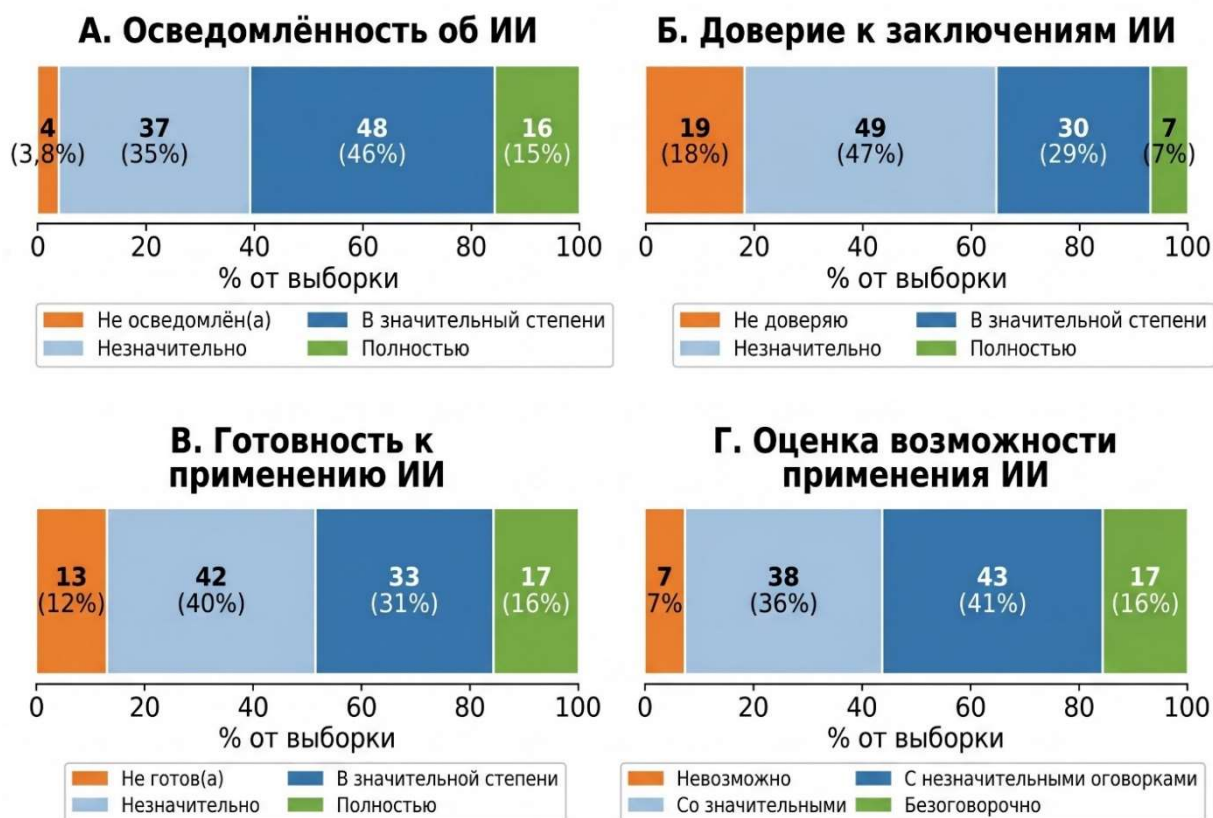


Рисунок 2 – Осведомленность, доверие, готовность и оценка возможности применения ИИ в профессиональной деятельности врачами-терапевтами и врачами общей практики (n = 105)

Применявшиеся технологии искусственного интеллекта

Среди применявшихся технологий наиболее распространены чат-боты и диалоговые системы на основе больших языковых моделей (GigaChat, YandexGPT и др.), которые использовали 52 (49,5%) специалиста. Системы компьютерного зрения для анализа медицинских изображений применяли 18 (17,1%) опрошенных. ИИ-агенты указаны 13 (12,4%), СППР с технологиями ИИ – 13 (12,4%), роботизированные системы – 4 (3,8%) специалиста. Не применяли ни одну технологию в медицинской деятельности 28 (26,7%) опрошенных.

Высокая доля использования LLM-систем объяснима активным внедрением этих инструментов в первичном звене для оформления медицинской документации и ответов на клинические запросы. Терапевты и врачи общей практики ежедневно работают с большими объемами текстовой информации (заключения, направления,

выписки), что делает чат-ботов особенно привлекательным инструментом именно для этой категории специалистов (Рисунок 3).



Рисунок 3 – Технологии ИИ, применявшиеся в медицинской деятельности врачами-терапевтами и врачами общей практики (% от $n = 105$, множественный выбор)

Субъективная оценка точности искусственного интеллекта

Усредненная субъективная оценка точности диагностики системами ИИ составила 49,8% (медиана 55%), а врачами-специалистами – 72,4% (медиана 75%). Разрыв между двумя оценками достигает около 23 п.п. Оценки точности ИИ концентрируются в диапазоне 40–70%, тогда как точность диагностики врачами большинство респондентов относит к диапазону 70–90%.

Полученная субъективная оценка точности ИИ (49,8%) существенно занижена по сравнению с реальными клиническими показателями. Современные системы клинической поддержки принятия решений на основе ИИ, по данным ряда исследований, достигают точности 80–95% на задачах, для которых они были специально разработаны и валидированы. Разрыв в 22–45 п.п. между

субъективным восприятием и реальными показателями является самостоятельной и весьма значимой проблемой. Именно заниженные ожидания формируют и воспроизводят скептицизм: врач, полагающий, что ИИ «ошибается в каждом втором случае», не станет полагаться на его заключения.

Причины такого расхождения могут быть следующими. Во-первых, большинство врачей не имеют опыта работы с конкретными верифицированными ИИ-системами, интегрированными в клиническую практику, и опираются на медиаобраз ИИ, сформированный публикациями в средствах массовой информации и обсуждениями в профессиональных чатах. Во-вторых, восприятие точности ИИ может быть искажено «эффектом ошибки выжившего», случаи неудачного применения ИИ (реальные или воображаемые) запоминаются и тиражируются значительно активнее, чем рутинные случаи успешной работы алгоритма. В-третьих, отсутствие зарегистрированных ИИ-систем, доступных врачам для самостоятельного использования, лишает их возможности верифицировать свои представления о точности ИИ на собственном клиническом опыте (Рисунок 4).

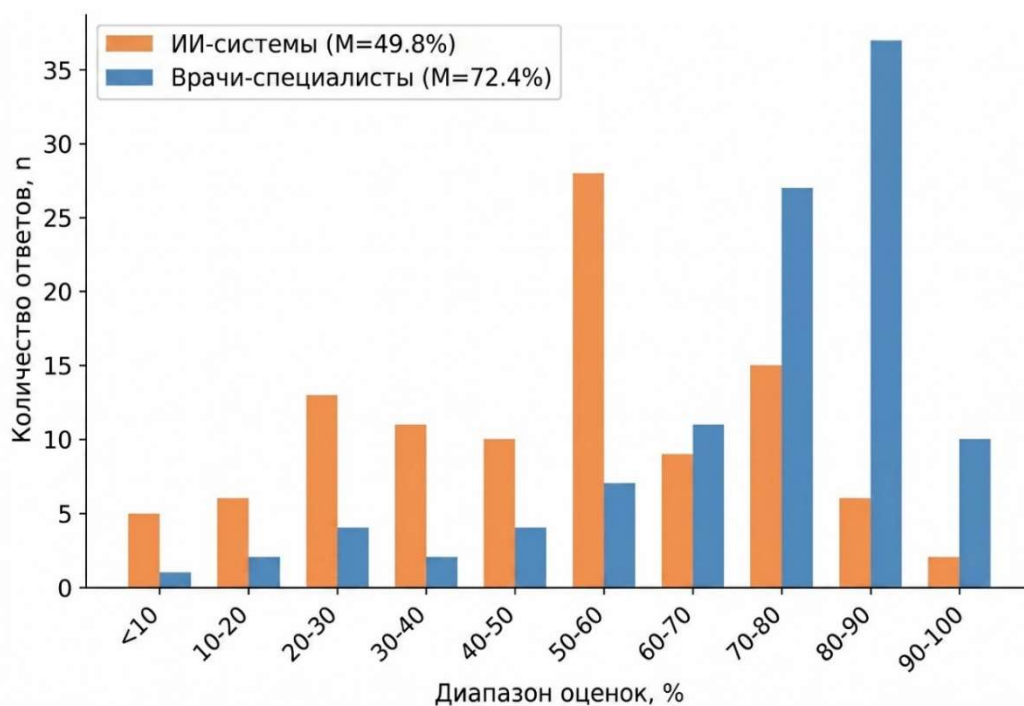


Рисунок 4 – Субъективная оценка точности диагностики: ИИ против врачей (n = 105)

Преимущества и ожидаемые проблемы внедрения искусственного интеллекта

Основным преимуществом ИИ при множественном выборе вариантов ответов 75 (71,4%) опрошенных указали сокращение времени на выполнение рутинных задач. Снижение нагрузки на врачей выбрали 64 (61,0%); получение «второго мнения» и повышение точности диагностики – по 40 (38,1%) врачей. Не видели преимуществ 10 (9,5%) специалистов.

Главная ожидаемая проблема использования ИИ в медицинской деятельности – ошибки диагностики, на что указали 64 (61,0%) опрошенных. Распространение самодиагностики пациентами выбрали 60 (57,1%); ошибки при назначении лечения – 58 (55,2%); снижение авторитета врача – 44 (41,9%); снижение квалификации врачей – 41 (39,0%) респондент. Показательно, что для терапевтов и врачей общей практики опасения ошибок лечения выходят на уровень наравне с самодиагностикой, вероятно, ввиду особой роли первичного звена в маршрутизации пациентов (Рисунок 5).

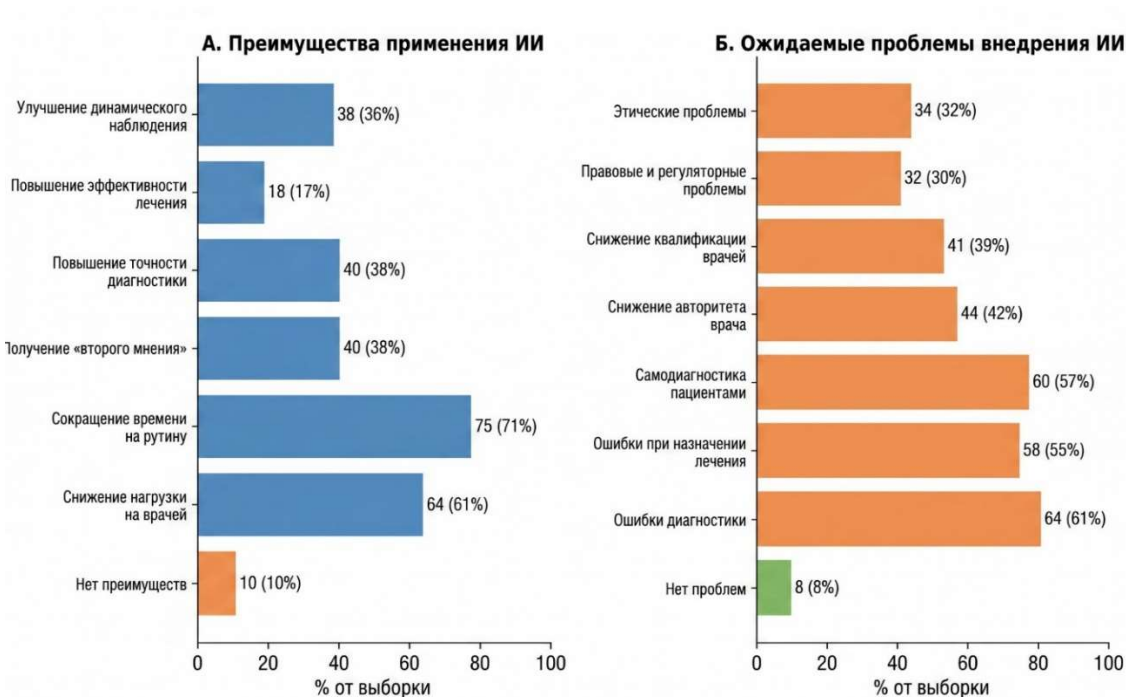


Рисунок 5 – Преимущества и ожидаемые проблемы внедрения ИИ по мнению врачей-терапевтов и врачей общей практики (% от n = 105, множественный выбор)

Целесообразные направления применения искусственного интеллекта

Опрошенные врачи определили следующие приоритетные направления внедрения ИИ в медицинскую деятельность: оформление медицинской документации – 68 (64,8%), анализ данных и составление отчетов – 62 (59,0%), скрининг – 59 (56,2%), диагностика – 49 (46,7%), телемедицина – 43 (41,0%).

Примечательно, что для терапевтов и врачей общей практики возможность использования ИИ в телемедицине занимает более высокое место (5-я позиция), что отражает специфику первичного звена медицинской помощи, где дистанционные консультации становятся все более распространенным элементом работы. Непосредственное применение ИИ для назначения лечения поддержал лишь 21 (20,0%) специалист, что указывает на стойкое нежелание делегировать алгоритмам наиболее ответственные клинические решения (Рисунок 6).

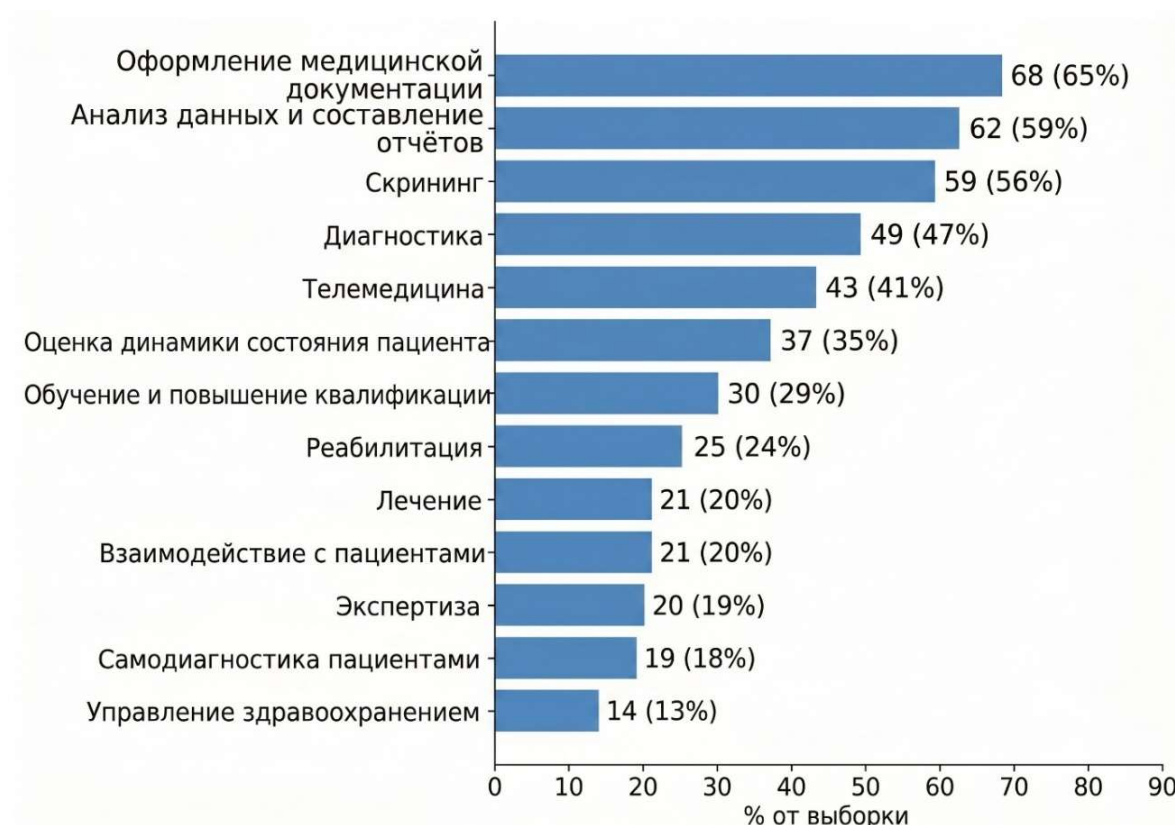


Рисунок 6 – Направления целесообразного применения ИИ по мнению врачей-терапевтов и врачей общей практики (% от n = 105, множественный выбор)

Сравнение возрастных групп

Сводные показатели для трех возрастных групп представлены в Таблице 2. Специалисты до 35 лет демонстрировали наибольшую частоту применения ИИ (средний балл 2,47), тогда как группа 35–50 лет – наименьшую (1,93). Уровень доверия и готовность к применению ИИ сопоставимы во всех трех группах. При статистическом анализе значимые межгрупповые различия по доверию к технологиям ИИ отсутствовали ($H = 1,87$; $p = 0,392$), что свидетельствует об отсутствии выраженного «возрастного» барьера в данной выборке.

Таблица 2 – Сравнительные показатели отношения врачей-терапевтов и врачей общей практики к технологиям ИИ по возрастным группам

Показатель	До 35 лет (n=49)	35–50 лет (n=22)	Старше 50 лет (n=34)
Доверие к ИИ (балл, 1–4)	2,22	2,09	2,35
Готовность применять ИИ (балл, 1–4)	2,57	2,36	2,53
Применение ИИ в медицине (балл, 1–4)	2,47	1,93	2,09
Субъективная оценка точности диагностики ИИ, %	49,9	45,5	52,4
Субъективная оценка точности диагностики врачами, %	72,8	71,4	72,5

Обращает на себя внимание, что наиболее скептически к технологиям ИИ относятся не самые старшие специалисты, а врачи средней возрастной группы (35–50 лет), которые реже применяют ИИ и ниже оценивают его диагностическую точность. Врачи старше 50 лет, напротив, продемонстрировали несколько более высокие показатели доверия (2,35) и субъективной оценки точности ИИ (52,4%), что опровергает представление о линейной связи возраста и скептицизма. Возможное объяснение заключается в том, что врачи средней возрастной группы находятся на пике клинической нагрузки и могут воспринимать ИИ как дополнительную угрозу профессиональной автономии, тогда как более старшие

коллеги, обладающие устоявшимся профессиональным авторитетом, не ощущают такой угрозы (Рисунок 7).

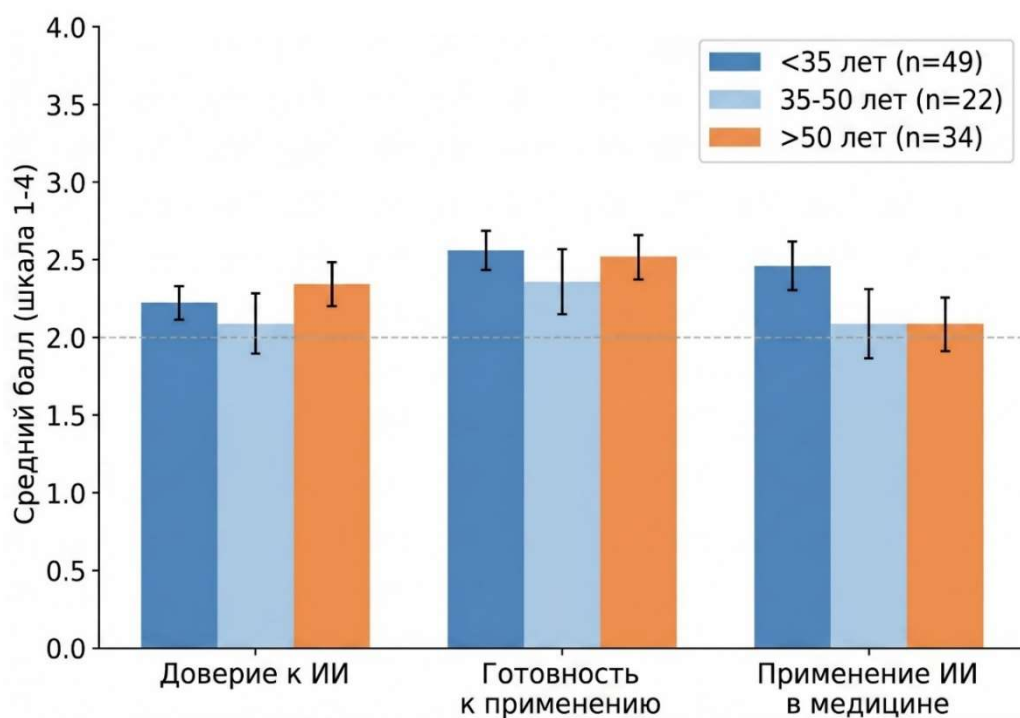


Рисунок 7 – Сравнение возрастных групп врачей-терапевтов и врачей общей практики по ключевым показателям отношения к ИИ ($M \pm SE$)

Подгрупповой анализ: применявшие и не применявшие искусственный интеллект в медицинской деятельности

Специалисты, имеющие практический опыт работы с ИИ-технологиями в медицине ($n = 77$), достоверно выше оценивали уровень доверия к ИИ по сравнению со специалистами без такого опыта (2,36 против 1,89 балла; $U = 1394$; $p = 0,014$). Аналогичная тенденция прослеживается и в оценке точности: имеющие опыт оценивают точность систем ИИ в среднем на 9 п.п. выше, чем коллеги без опыта (52,1% против 43,2%).

Таким образом, реальный опыт применения ИИ коррелирует с позитивным сдвигом как в доверии, так и в восприятии технической компетентности алгоритмов. Этот результат указывает на то, что ключевым предиктором позитивного отношения к ИИ является не возраст, не тип медицинской

организации и не уровень цифровой грамотности, а непосредственный практический контакт с технологией (Рисунок 8).

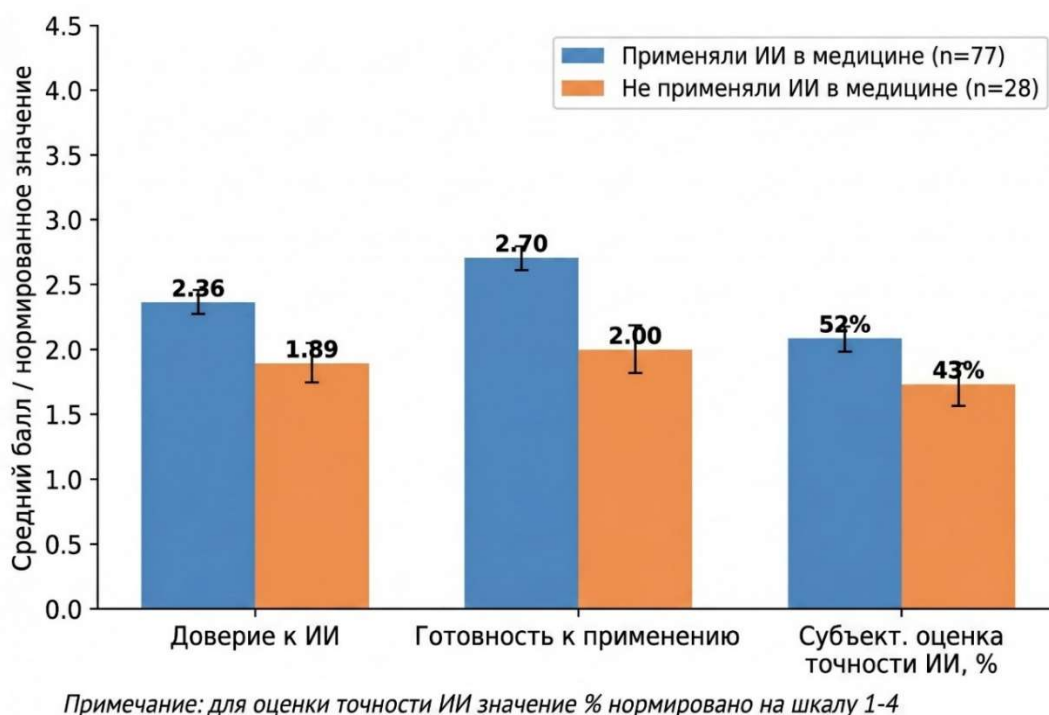


Рисунок 8 – Сравнение подгрупп «применявших» и «не применявших» технологии ИИ в медицинской деятельности среди врачей-терапевтов и врачей общей практики ($M \pm SE$)

Кластерный анализ по совокупности показателей доверия, готовности и применения искусственного интеллекта

По совокупности трех показателей (доверие, готовность, применение ИИ) сформированы три группы респондентов. «Скептики» ($n = 30$; 28,6%) характеризуются минимальными значениями по всем переменным. «Умеренные» ($n = 48$; 45,7%) осознают потенциал ИИ, применяют его эпизодически, имеют умеренно-позитивные ожидания. «Энтузиасты» ($n = 27$; 25,7%) регулярно работают с ИИ и демонстрируют высокое доверие к технологии. Преобладание группы «Умеренных» (45,7%) свидетельствует о том, что большинство терапевтов занимает взвешенную позицию: они не отвергают ИИ, но и не готовы безоговорочно ему довериться. Такое распределение характерно для ранней стадии

диффузии инновации, когда «раннее большинство» наблюдает за опытом первопроходцев и формирует собственное суждение.

Зависимость от типа медицинской организации

Статистически значимых различий между типами медицинских организаций – мест работы респондентов не выявлено. Вместе с тем прослеживается содержательная тенденция: специалисты образовательных организаций демонстрируют более высокую частоту применения ИИ (средний балл 2,79) при сопоставимом уровне доверия (2,07). Специалисты амбулаторно-поликлинических медицинских организаций составляют большинство выборки (66,7%), что отражает профессиональный профиль врача общей практики и терапевта первичного звена.

Один респондент был исключен из сравнительного анализа по типу медицинской организации: участник указал место работы «Орган управления здравоохранением», однако при этом имел возраст 18 лет и отсутствие стажа работы по специальности, что делает указанное место работы предположительно недостоверным (Таблица 3, Рисунок 9).

Таблица 3 – Показатели отношения к ИИ в зависимости от типа организации – места работы врачей-терапевтов и врачей общей практики (М; точность – %)

Тип учреждения	n	Доверие к ИИ	Готовность к применению ИИ	Применение ИИ	Точность диагностики ИИ, %	Точность диагностики врачами, %	Не используют ИИ в медицине, n (%)
Амбулаторно-поликлиническая медицинская организация	70	2,26	2,47	2,17	48,2	72,3	21 (30%)
Стационарная медицинская организация	17	2,29	2,76	2,12	54,4	74,1	6 (35%)
Образовательная медицинская организация	14	2,07	2,43	2,79	52,5	72,1	4 (29%)

Продолжение Таблицы 3

Научная медицинская организация*	3	2,33	2,33	2,33	46,7	70,0	2 (67%)

Примечание: группа «Научная организация» (n = 3) интерпретируется только на описательном уровне ввиду малого размера

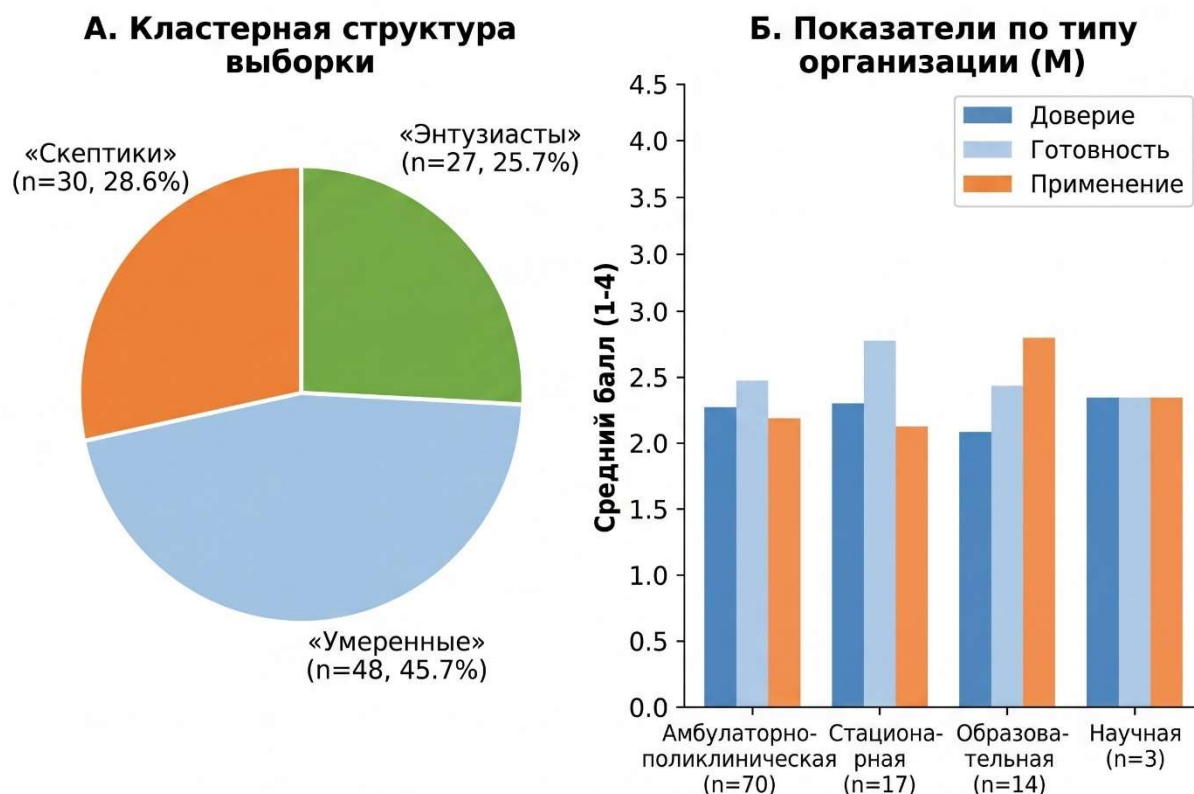


Рисунок 9 – Кластерный профиль выборки врачей-терапевтов и врачей общей практики и отношение к ИИ по типу медицинской организации

Проведенное исследование позволило получить комплексную характеристику отношения врачей-терапевтов и врачей общей практики к технологиям ИИ на современном этапе их внедрения в отечественное здравоохранение.

Главным результатом следует считать не столько фиксацию умеренного скептицизма, сколько выявление его предикторов и структуры. Доверие к заключениям ИИ достоверно выше у специалистов, имеющих реальный опыт работы с ИИ-технологиями: они оценивают точность систем в среднем на 9 п.п. выше, чем коллеги без такого опыта (52,1% против 43,2%; $p = 0,014$). Разрыв между применением ИИ в немедицинской и профессиональной медицинской

деятельности в данной специальности минимален, около 5 п.п., что отличает эту выборку от представителей других специальностей.

Субъективная оценка точности диагностики системами ИИ (49,8%) оказалась существенно ниже реальных клинических показателей. Разрыв в 22–45 п.п. между восприятием и реальностью указывает на необходимость образовательных мероприятий, ориентированных не на информирование «об ИИ», а на практическое взаимодействие с конкретными ИИ-инструментами в рабочей среде.

Приоритетные направления применения ИИ по мнению терапевтов и врачей общей практики (оформление медицинской документации, анализ данных, скрининг и телемедицина) точно соответствуют специфике первичного звена здравоохранения. Отсутствие значимых различий по возрастным группам ($H = 1,87$; $p = 0,392$) указывает на то, что «возрастной барьер» в отношении к ИИ в медицине скорее представляет собой устойчивый миф, нежели реальную закономерность.

3.1.3. Исследование уровня доверия к технологиям искусственного интеллекта и готовности его применять врачами-дерматовенерологами

Характеристика выборки

В марте–июле 2026 г. проведен анонимный онлайн-опрос на платформе Anketolog.ru. Из общей базы ответов отобран 191 респондент, указавший специальность «Дерматовенерология» и/или «Косметология». Респонденты, отметившие одновременно обе специальности ($n = 22$), включались в анализ как дерматовенерологи. Респонденты, указавшие только «Косметология» без «Дерматовенерология» ($n = 6$), включены в общую выборку, поскольку в российской практике косметология является смежной специальностью.

Для корректного попарного сравнения обучающихся и состоявшихся специалистов сформированы сбалансированные подгруппы. Из 185 респондентов 84 указали статус «Ординатор» без совмещения с практикой, 62 – «Практикующий

врач» без статуса ординатора, 13 – оба статуса одновременно, 26 – ни один из двух (аспиранты, преподаватели, научные работники). Для парного сравнения из 84 ординаторов методом простой случайной выборки отобраны 62, что уравнивает подгруппы ($n = 94 + 94$). Респонденты с двойным статусом исключены из подгруппового анализа, но сохранены в анализе полной выборки.

Из 191 специалиста 164 (85,9%) составили женщины, 27 (14,1%) – мужчины. Средний возраст – $34,4 \pm 12,1$ года, медиана 29 лет (диапазон 21–70). Молодой медианный возраст объясняется тем, что 50,8% выборки составили ординаторы. Ученую степень кандидата имели 29 (15,7%), доктора – 6 (3,2%). 84,9% проживали в городах-миллионниках. 28 (14,7%) респондентов дополнительно отметили специальность «Косметология».

Применение искусственного интеллекта в медицинской и немедицинской деятельности

Не применяют ИИ в профессиональной деятельности 64 (33,5%) специалиста-дерматовенеролога. В бытовой деятельности не используют ИИ лишь 29 (15,2%). Разрыв между бытовым и профессиональным неприменением составляет около 18 п.п. Это значительно больше, чем у терапевтов (5 п.п.), и свидетельствует о том, что дерматовенерологи, активно используя ИИ в повседневной жизни, не переносят эту привычку в профессиональную сферу в той же мере. Одна из возможных причин, отсутствие доступных валидированных ИИ-инструментов, специально разработанных для задач дерматовенерологии и интегрированных в рабочий процесс (Рисунок 10).

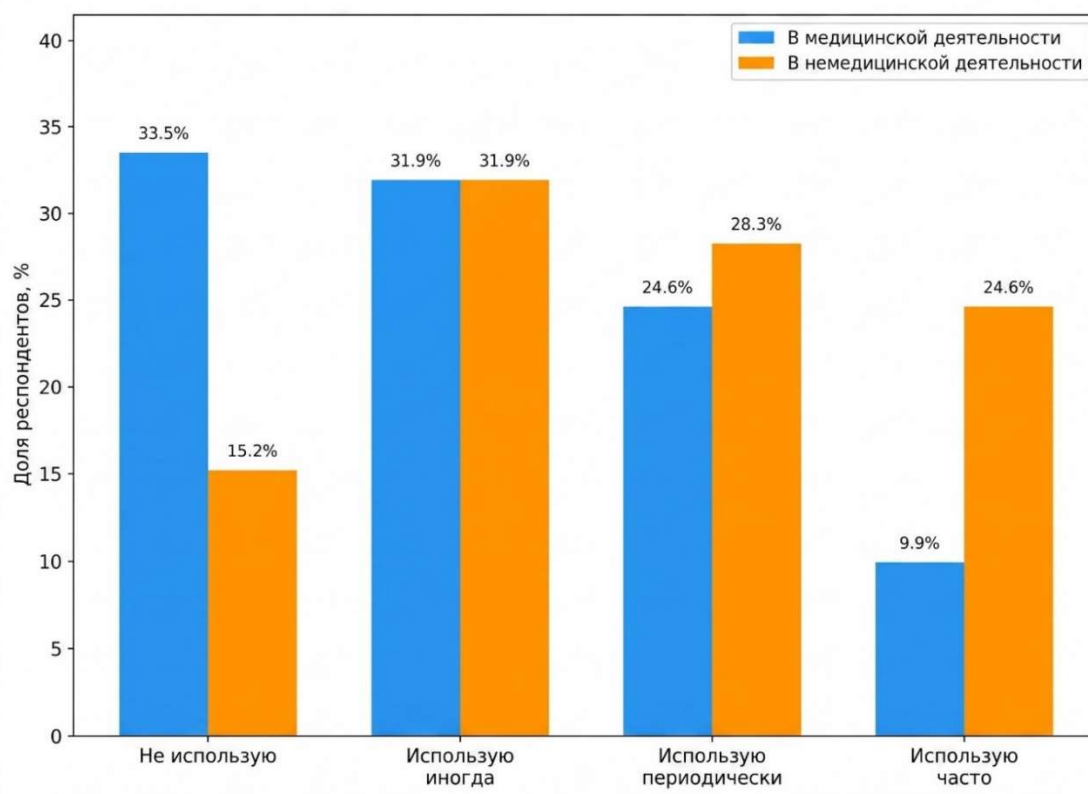


Рисунок 10 – Частота применения технологий ИИ в медицинской и немедицинской деятельности врачами-дерматовенерологами (n = 191)

Осведомленность, доверие, готовность к применению искусственного интеллекта и оценка возможности применения искусственного интеллекта

Скептически оценивают доверие к ИИ 127 (66,5%) специалистов (не доверяют или доверяют незначительно). Этот показатель практически идентичен показателю терапевтов (64,8%). Готовность к применению, тем не менее, выше: лишь 23 (12,0%) не готовы вовсе. Считают применение ИИ возможным 93,5% респондентов, из них безоговорочно – 27 (14,6%).

Аналогичный диссонанс «не доверяю, но готов попробовать», обнаруженный среди терапевтов, воспроизводится и в выборке дерматовенерологов. Такой разрыв характерен для ранней стадии знакомства с технологией, когда врачи осознают ее потенциал, но не имеют достаточного клинического опыта работы с конкретными ИИ-инструментами для формирования обоснованного суждения о них (Рисунок 11).

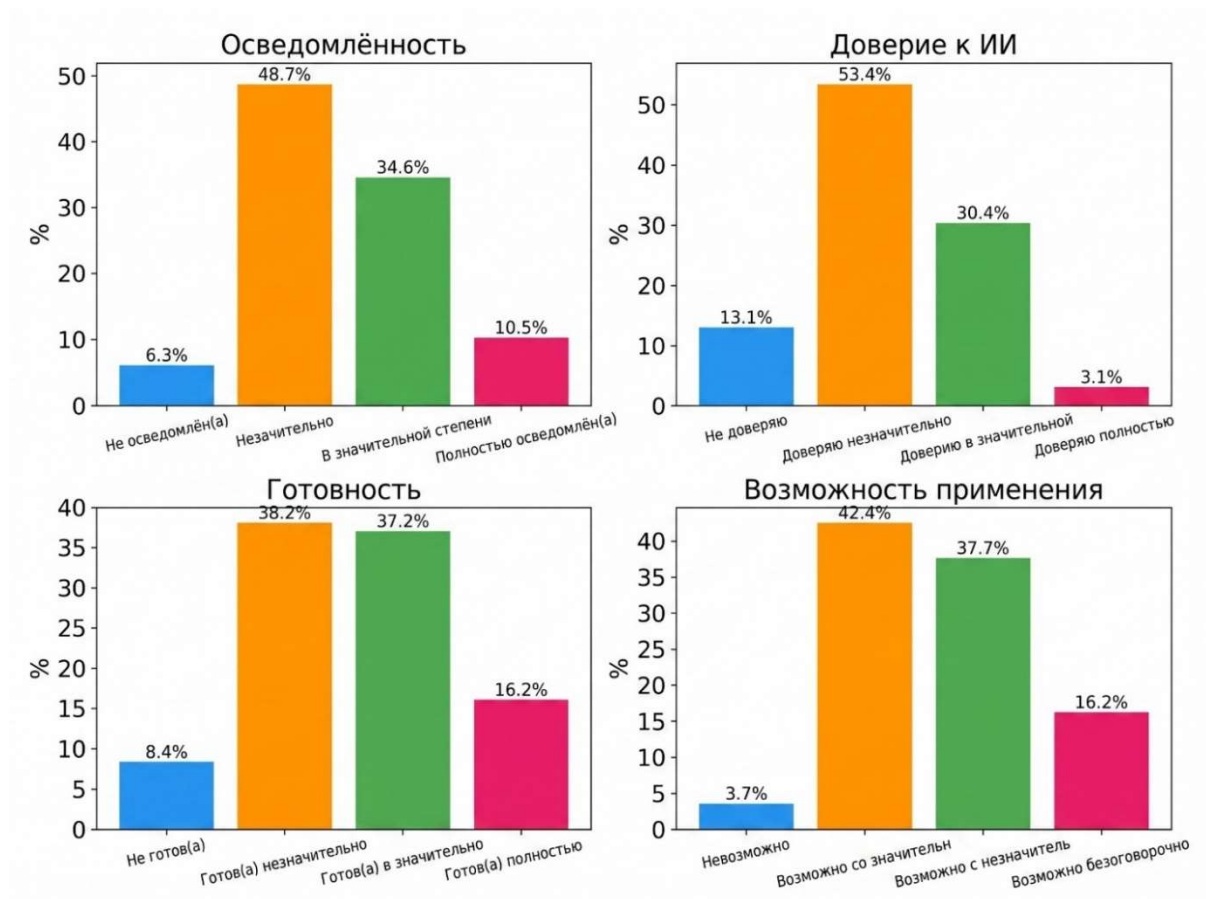


Рисунок 11 – Распределение ответов врачей-дерматовенерологов по уровню осведомленности, доверия к ИИ, готовности к применению и суждению о возможности использования ИИ в медицинской деятельности (n = 191)

Применявшиеся технологии искусственного интеллекта

Чат-боты на основе LLM уверенно лидируют: их использовали 122 (63,9%) дерматовенеролога. Системы компьютерного зрения для анализа изображений применяли лишь 17 (8,9%), несмотря на визуальный характер диагностики в данной специальности. Не применяли ни одной технологии 58 (30,4%).

Этот результат особенно примечателен. Дерматовенерология является одной из наиболее «визуальных» медицинских специальностей, где алгоритмы компьютерного зрения имеют наиболее очевидную точку приложения. Тем не менее именно системы анализа изображений остаются наименее используемыми: 8,9% против 17,1% у терапевтов. Разрыв между доказанной эффективностью систем компьютерного зрения в дерматоскопии и реальным использованием

отражает проблему доступности таких систем в рутинной клинической практике (Рисунок 12).

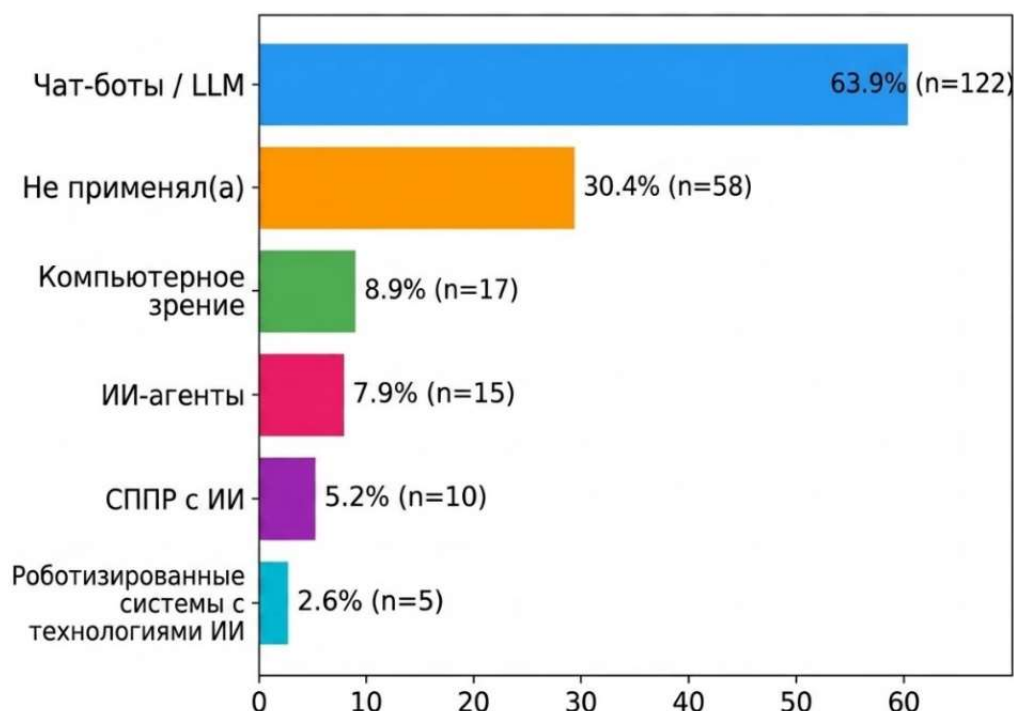


Рисунок 12 – Технологии ИИ, применявшиеся врачами-дерматовенерологами в медицинской деятельности (n = 191, множественный выбор)

Субъективная оценка точности искусственного интеллекта

Оценка точности ИИ медиана 45% (IQR 25–65%). Оценка точности врачей: медиана 75% (IQR 65–85%). Разрыв составляет 30 п.п., что больше, чем у терапевтов (23 п.п.) и онкологов (23 п.п.). Дерматовенерологи оценивают ИИ скептически, что может объясняться меньшей доступностью валидированных ИИ-систем для задач дерматовенерологии по сравнению, например, с лучевой диагностикой в онкологии (Рисунок 13).

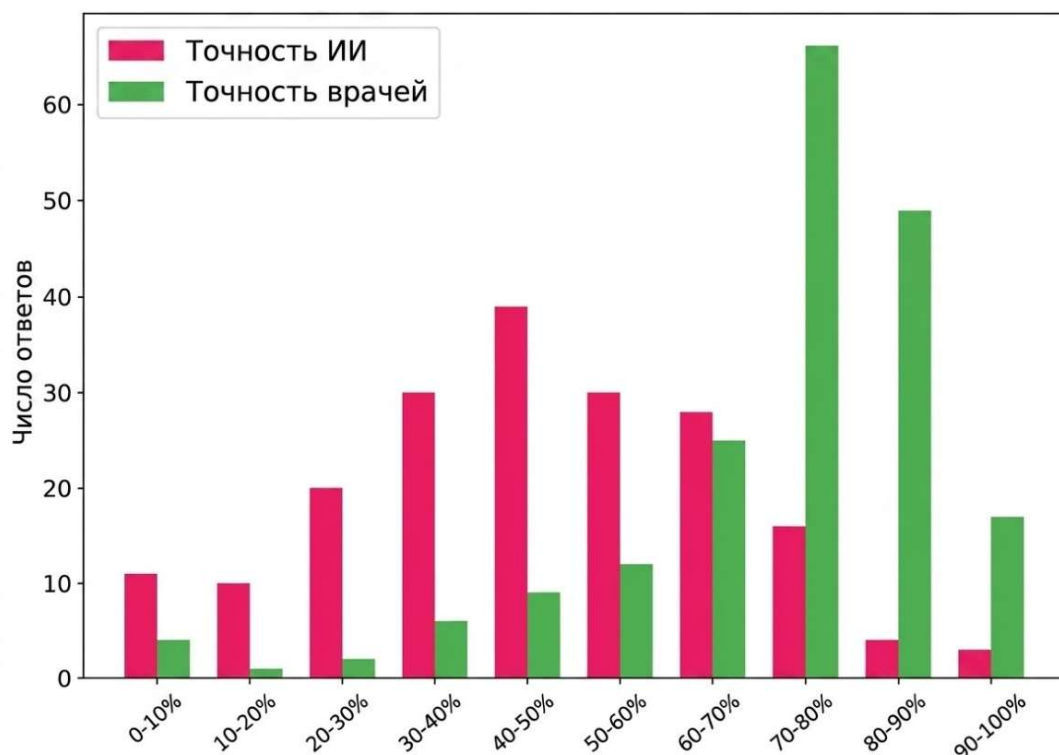


Рисунок 13 – Распределение субъективных оценок точности диагностики системами ИИ и врачами-специалистами среди дерматовенерологов (n = 191)

Преимущества и ожидаемые проблемы внедрения искусственного интеллекта

Главное преимущество ИИ по мнению дерматовенерологов это сокращение рутины (77,8%), снижение нагрузки (66,5%). Ожидаемые проблемы — это ошибки диагностики (73,8%), ошибки назначения лечения (69,6%), самодиагностика пациентами (62,8%). Последний пункт особенно чувствителен для дерматовенерологии это мобильные приложения для анализа родинок уже распространены среди пациентов, и врачи опасаются как ложной тревоги, так и ложного успокоения.

Доля дерматовенерологов, обеспокоенных ошибками диагностики (73,8%), существенно выше, чем среди терапевтов (61,0%) и онкологов (60,6%). Это может быть связано с тем, что в дерматовенерологии цена пропуска меланомы при скрининге родинок экзистенциально значима, а пациент может обратиться к ИИ-приложению вместо врача (Рисунок 14).

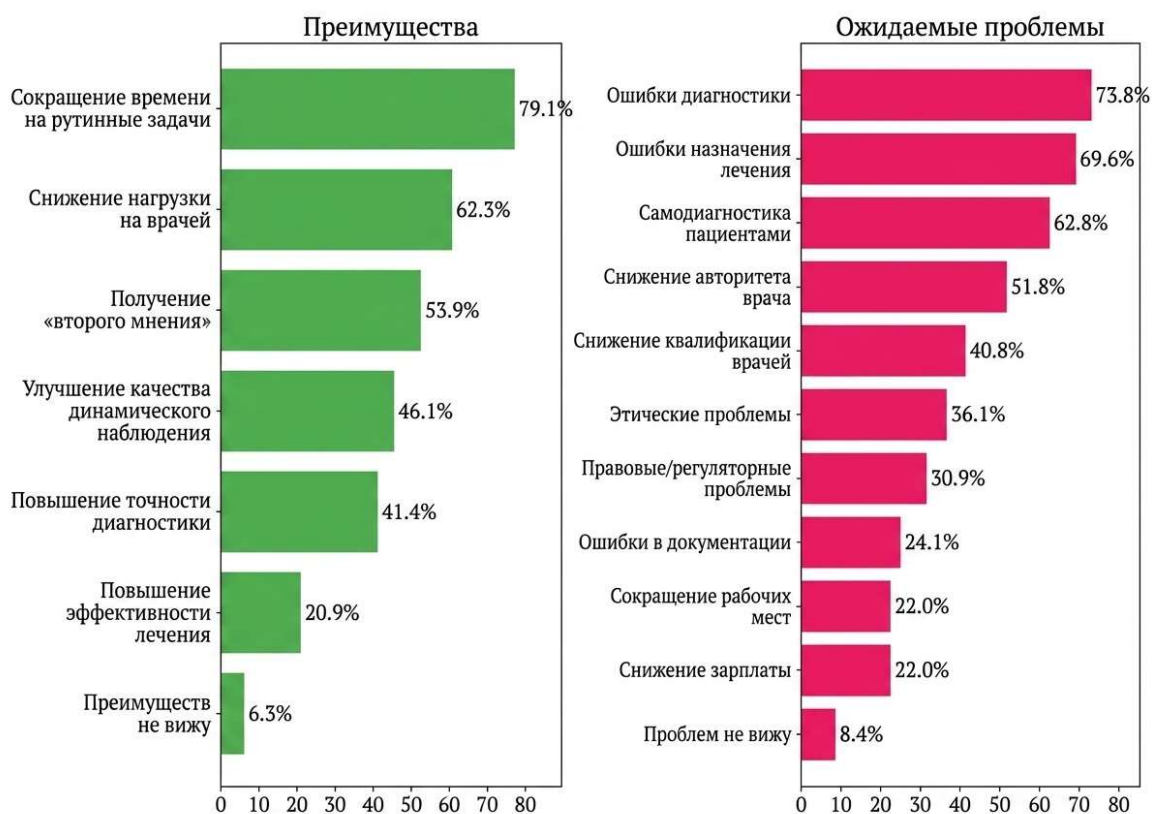


Рисунок 14 – Ожидаемые преимущества и проблемы внедрения технологий ИИ в медицинскую деятельность по мнению врачей-дерматовенерологов (n = 191, множественный выбор)

Целесообразные направления применения искусственного интеллекта

Наиболее целесообразными направлениями дерматовенерологи считают анализ данных (68,1%), оформление документации (63,8%), скрининг (59,5%), диагностику (56,2%). Высокий рейтинг скрининга и диагностики отличает дерматовенерологов от терапевтов для специальности с визуальным характером эти задачи ощущаются как наиболее естественная точка приложения ИИ. Если у терапевтов на первом месте стоит оформление документации, то у дерматовенерологов анализ данных и скрининг выходят на лидирующие позиции (Рисунок 15).

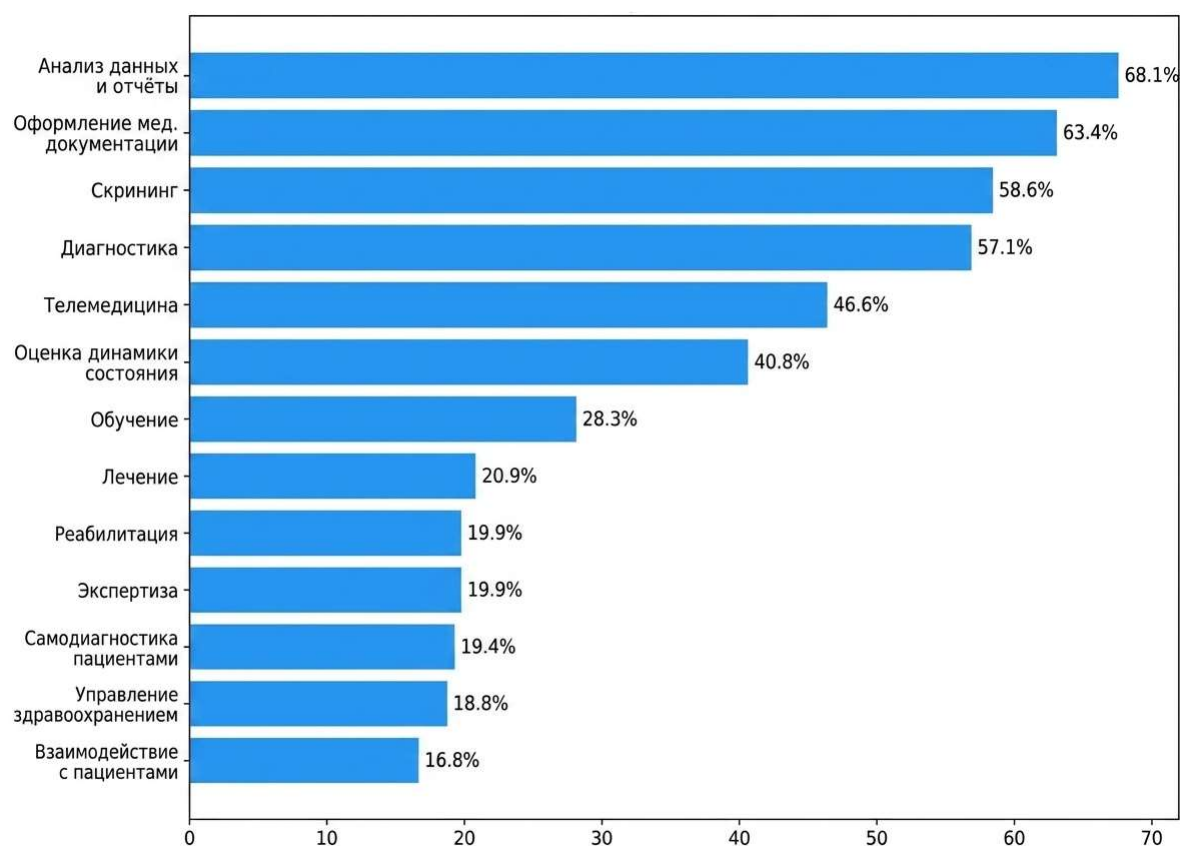


Рисунок 15 – Направления медицинской деятельности, в которых целесообразно применение технологий ИИ, по мнению врачей-дерматовенерологов (n = 191, множественный выбор)

Сравнение отношения ординаторов и практикующих врачей

Центральный результат исследования состоит в анализе различий между ординаторами и практикующими врачами. На сбалансированной выборке (n = 94 + 94) при 17-летнем разрыве в медианном возрасте (25 vs 42 года) ни по одному из шести показателей профессионального отношения к ИИ (доверие, готовность, применение, осведомленность, оценка точности ИИ и врачей) различий не обнаружено (все $p > 0,28$). Доля применяющих ИИ в медицинской деятельности у ординаторов выше (70% против 62%), однако разница не достигает статистической значимости ($\chi^2 = 1,16$; $p = 0,281$). Единственное достоверное различие касается бытовой сферы, ординаторы значительно чаще используют ИИ в повседневной жизни (91% против 78%; $\chi^2 = 5,87$; $p = 0,015$) (Таблица 4).

Таблица 4 – Сравнение ординаторов и не-ординаторов среди дерматовенерологов (сбалансированная выборка)

Показатель	Ординаторы (n=94)	Не-ординаторы (n=94)	U	p
Доверие к ИИ	2,26	2,21	4457	0,909
Готовность	2,64	2,59	4570	0,668
Применение ИИ	2,11	2,09	4521	0,774
Осведомленность	2,46	2,51	4303	0,738
Точность ИИ, %	47,8	44,5	4726	0,405
Точность врачей, %	74,0	69,0	4803	0,287
ИИ в мед., %	70	62	$\chi^2 = 1,16$	0,281
ИИ в быту, %	91	78	$\chi^2 = 5,87$	0,015

Кластерное распределение: $\chi^2 = 0,47$; $p = 0,789$ (Рисунок 16).

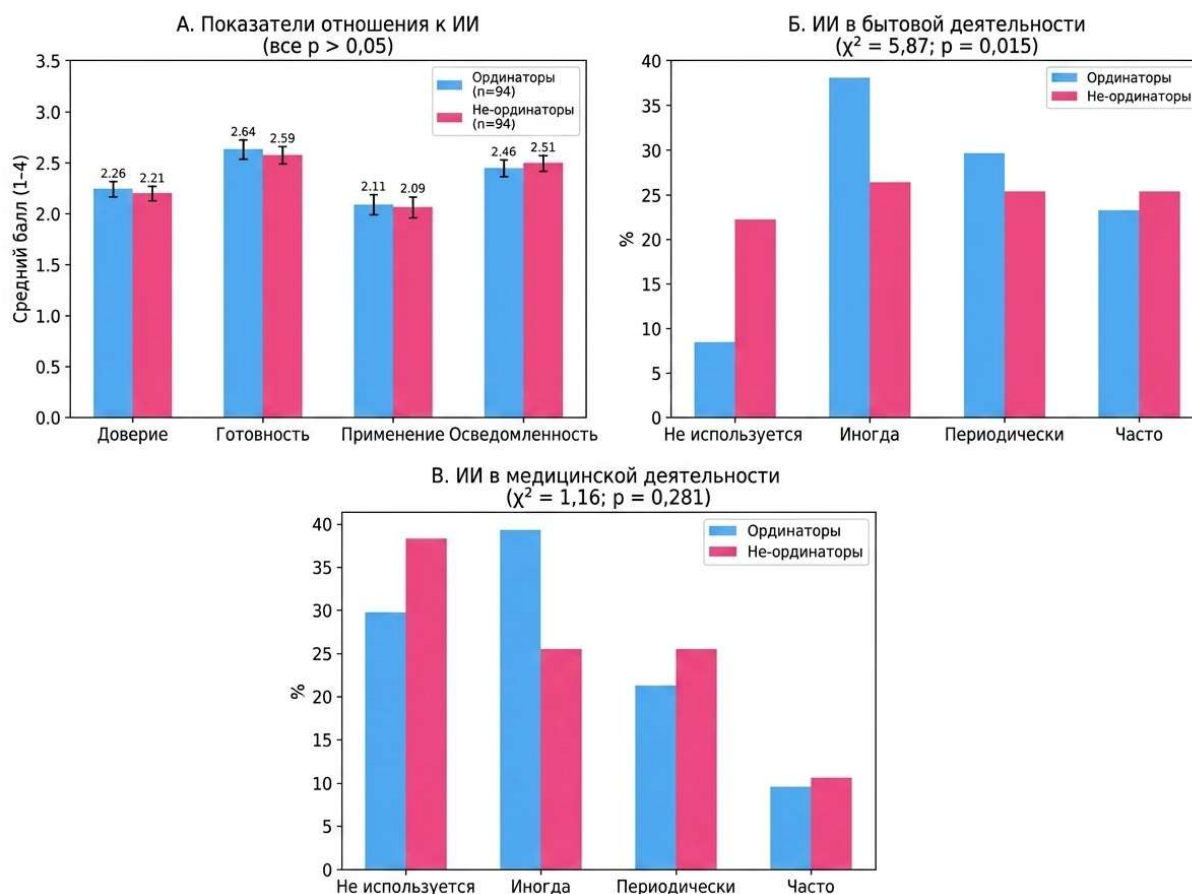


Рисунок 16 – Сравнение показателей отношения к ИИ, частоты применения ИИ в бытовой и медицинской деятельности между ординаторами и врачами-дерматовенерологами (сбалансированная выборка, $n = 94 + 94$)

Дополнительное сравнение ординаторов (только учатся, $n = 65$) и практикующих врачей (только работают, $n = 65$) усиливает обнаруженный

парадокс. Разрыв в медианном возрасте еще больше (25 vs 43 года). Ординаторы достоверно чаще используют ИИ в быту ($\chi^2 = 7,61$; $p = 0,006$) и чаще применяют ИИ в медицинской деятельности ($\chi^2 = 3,93$; $p = 0,047$). Однако по уровню доверия к ИИ группы абсолютно неразличимы ($U = 2118$; $p = 0,982$). Кластерное распределение также не различается ($\chi^2 = 1,07$; $p = 0,587$). Этот результат подтверждает, что даже реальное использование ИИ в профессиональной деятельности не формирует автоматически более высокого доверия к его заключениям (Таблица 5).

Таблица 5 – Сравнение ординаторов (только учатся) и практикующих врачей (только работают) среди дерматовенерологов

Показатель	Только учатся (n=65)	Только работают (n=65)	U / χ^2	p
Доверие к ИИ	2,23	2,20	U=2118	0,982
Готовность	2,62	2,65	U=2087	0,902
Применение ИИ	2,11	1,86	U=2464	0,084
Осведомленность	2,48	2,37	U=2317	0,298
Точность ИИ, %	47,3	42,1	U=2388	0,196
Точность врачей, %	72,8	68,2	U=2283	0,414
ИИ в медицине, %	71	52	$\chi^2=3,93$	0,047
ИИ в быту, %	92	72	$\chi^2=7,61$	0,006

Кластерное распределение: $\chi^2 = 1,07$; $p = 0,587$ (Рисунок 17).

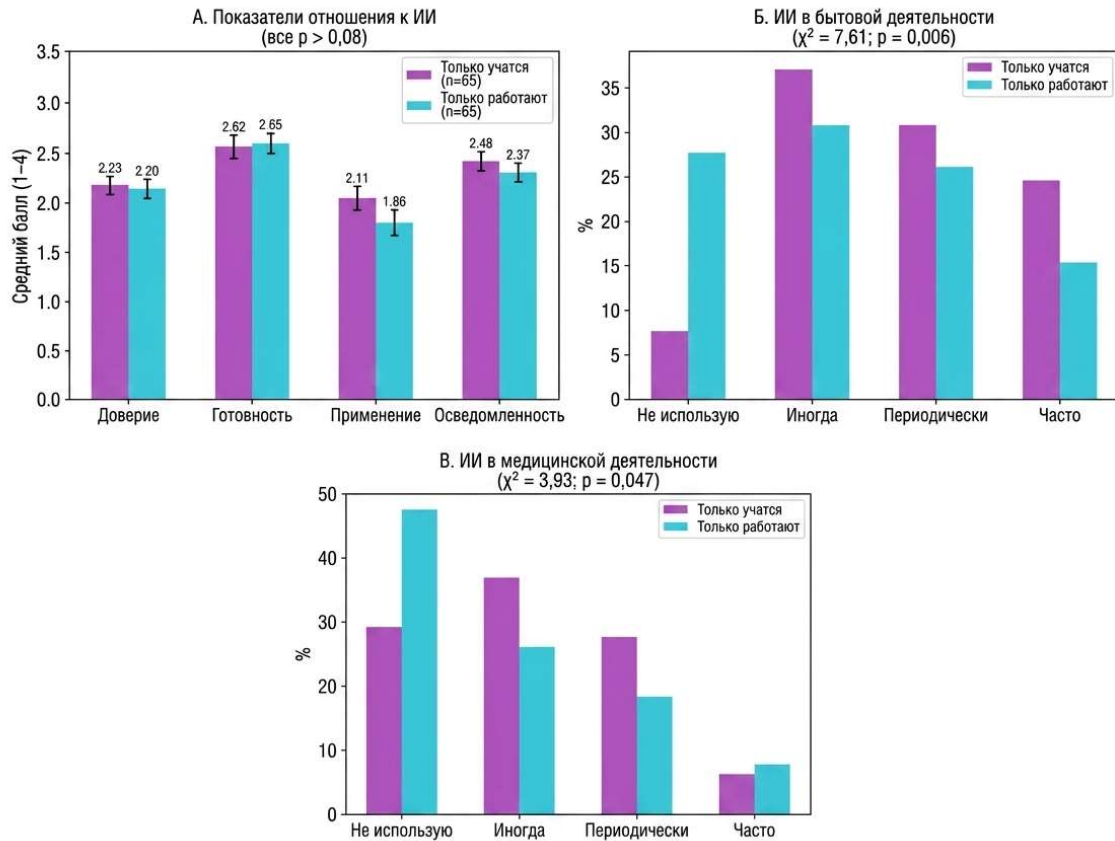


Рисунок 17 – Сравнение показателей отношения к ИИ между ординаторами (только учатся) и практикующими врачами (только работают) среди дерматовенерологов

Кластерный анализ по совокупности показателей доверия, готовности и применения искусственного интеллекта

По совокупности показателей доверия, готовности и применения ИИ кластерный анализ разделил выборку на три группы: «Скептики» – 46 (24,1%), «Умеренные» – 96 (50,3%), «Энтузиасты» – 49 (25,7%). По сравнению с онкологами (34,1% / 45,0% / 20,9%) доля энтузиастов у дерматовенерологов выше, а скептиков ниже. Вероятное объяснение состоит в том, что более молодые специалисты в выборке и активное бытовое использование ИИ снижают порог принципиального неприятия, хотя и не повышают собственно доверие (Рисунок 18).

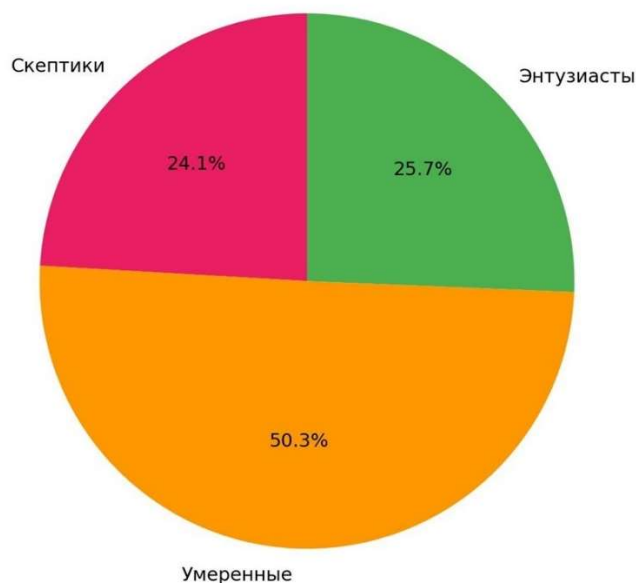


Рисунок 18 – Кластерный профиль выборки врачей-дерматовенерологов по совокупности показателей доверия, готовности и применения ИИ (n = 191)

Подгрупповой анализ: применявшие и не применявшие искусственный интеллект в медицинской деятельности

Среди 58 специалистов, не применявших ни одной технологии ИИ в медицинской деятельности, 27 (47%) не используют ИИ и в быту. Остальные 31 (53%) обращаются к ИИ в повседневной жизни, но не переносят навык в профессиональную деятельность. В группе «применявших» (n = 133) не используют ИИ в быту лишь 6 (5%). Эта картина подтверждает, что бытовой опыт использования ИИ является необходимым, но не достаточным условием для профессионального внедрения, половина «неприменяющих» в медицине активно работают с ИИ в повседневной жизни (Рисунок 19).

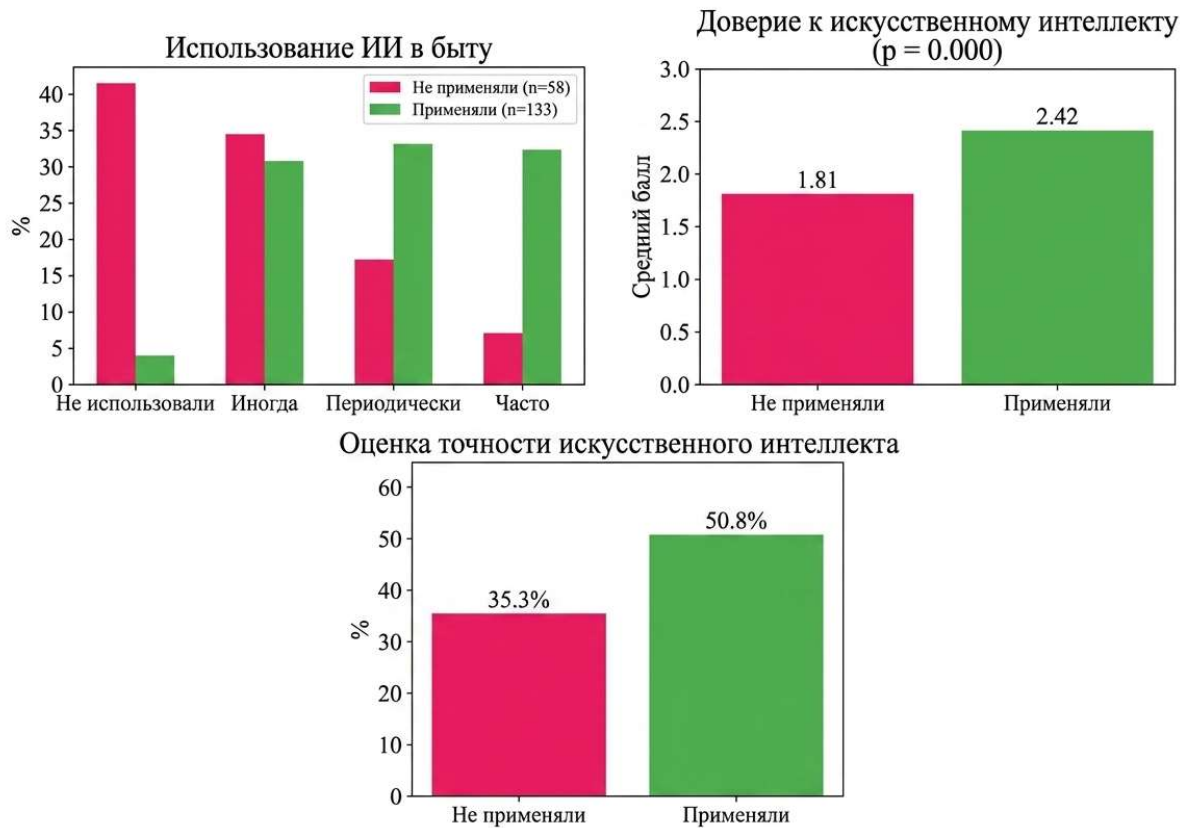


Рисунок 19 – Сравнение показателей бытового использования ИИ, доверия к ИИ и субъективной оценки точности ИИ между применявшими и не применявшими технологии ИИ в медицинской деятельности среди дерматовенерологов (n = 191)

Исследование впервые характеризует отношение российских дерматовенерологов к ИИ и обнаруживает парадокс: поколение, сформировавшее свои знания в среде генеративного ИИ, не демонстрирует более высокого профессионального доверия к нему по сравнению со старшими коллегами. Цифровая активность молодых специалистов не конвертируется в более высокое профессиональное доверие к ИИ. Доверие к ИИ в медицине формируется не бытовой привычкой к технологии, а наличием конкретных инструментов в клинической практике.

Барьер недоверия формируется не дефицитом пользовательского опыта, а отсутствием валидированных клинических ИИ-систем в дерматовенерологической практике. Для дерматовенерологии это означает, что внедрение ИИ должно идти через специализированные продукты ИИ-дерматоскопы для скрининга меланомы,

системы автоматического анализа клинических фотографий, ИИ-ассистенты для дифференциальной диагностики высыпаний, а не через попытки адаптировать универсальные генеративные модели к задачам специальности.

3.1.4. Исследование уровня доверия к технологиям искусственного интеллекта и готовности его применять врачами-онкологами

Характеристика выборки и методы

Из общей базы ответов были отобраны респонденты, отметившие специальность «Онкология». Итоговая выборка составила 249 анкет.

Из 249 опрошенных онкологов большую часть – 161 (64,7%) составили женщины.

Средний возраст – $37,2 \pm 11,0$ года, медиана 33 года, диапазон от 18 до 70 лет. Более половины выборки (134 человека, 54,3%) составили специалисты моложе 35 лет. Группа 35–50 лет включала 74 человека (30,0%), старше 50 лет – 39 (15,8%).

Подавляющее большинство не имели ученой степени – 225 (90,4%); кандидатами медицинских наук были 19 (7,6%), докторами – 5 (2,0%). 198 (79,5%) онкологов работали в амбулаторно-поликлинических, 41 (16,5%) – в стационарных медицинских организациях, 6 (2,4%) – в научных и 4 (1,6%) – в образовательных.

Московский регион представлен 241 (96,8%) респондентом.

По уровню владения ПК: офисными программами владели 150 (60,2%) респондентов, продвинутыми пользователями являлись 82 (32,9%), ответ «только печатаю» выбрали 10 (4,0%), уровень программиста – 7 (2,8%).

Применение искусственного интеллекта в медицинской и немедицинской деятельности

В медицинской деятельности не применяют ИИ 109 врачей-онкологов (43,8%). Остальные распределились следующим образом: иногда – 70 (28,1%),

периодически – 51 (20,5%), часто – 19 (7,6%). В бытовой деятельности картина несколько иная: не используют ИИ только 76 (30,5%), а часто используют 45 (18,1%) респондентов. Разрыв между бытовым и профессиональным неприменением составляет около 13 п.п.

Можно предположить, что онкологи, ежедневно работающие с цифровой визуализацией (КТ, МРТ, ПЭТ-КТ), более привычны к цифровым инструментам, и порог входа для них несколько ниже. Вместе с тем доля неприменяющих ИИ в медицине среди онкологов (43,8%) значительно выше, чем среди терапевтов (31,4%) и дерматовенерологов (33,5%), что может быть связано с более высокими требованиями к точности и ответственности за решения в данной специальности (Рисунок 20).

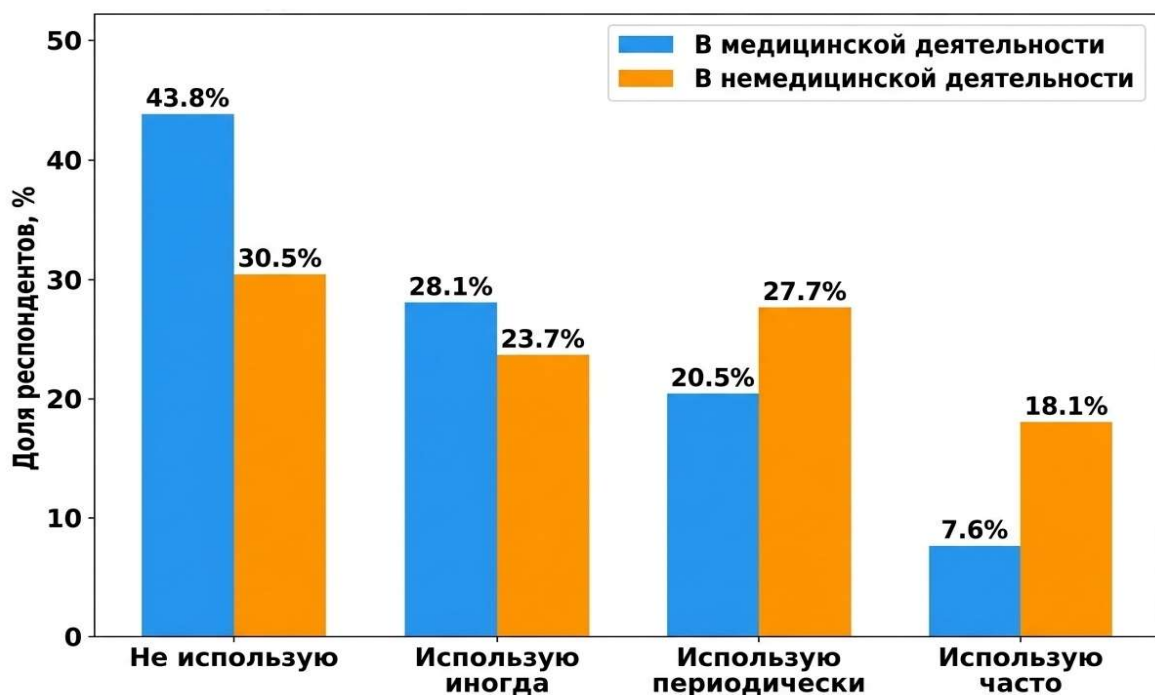


Рисунок 20 – Применение технологий ИИ в медицинской и немедицинской деятельности врачами-онкологами (n = 249)

Осведомленность, доверие, готовность к применению искусственного интеллекта и оценка возможности применения искусственного интеллекта

Почти половина опрошенных онкологов (47,8%) считают себя осведомленными о возможностях ИИ в медицине (16,5% – полностью, 31,3% – в значительной степени). Остальные оценили свою осведомленность как незначительную (42,6%) или отсутствующую (9,6%).

Доверие – центральный показатель для понимания перспектив внедрения ИИ у онкологов: невысокое, не доверяют вовсе 49 (19,7%), доверяют незначительно 122 (49,0%). Итого: 171 врач (68,7%) скептически оценивает алгоритмические заключения ИИ. Выраженное доверие проявили лишь 72 (28,9%), полное – 6 (2,4%). Уровень скептицизма среди онкологов (68,7%) несколько выше, чем среди дерматовенерологов (66,5%) и терапевтов (64,8%).

При этом готовность попробовать ИИ на практике заметно опережает доверие, в той или иной мере готовы 222 человека (89,2%). Полностью не готовы лишь 27 (10,8%). Этот диссонанс «не доверяю, но готов попробовать» характерен для ранней стадии знакомства с технологией, что говорит о том, что в онкологии технологии ИИ еще не обширно внедрены в клиническую практику.

Подавляющее большинство (94,0%) считают применение ИИ в медицине возможным, хотя и с оговорками: 43,8% – со значительными, 36,9% – с незначительными, 13,3% – безоговорочно. Считают использование технологий ИИ невозможным лишь 15 человек (6,0%) (Рисунок 21).

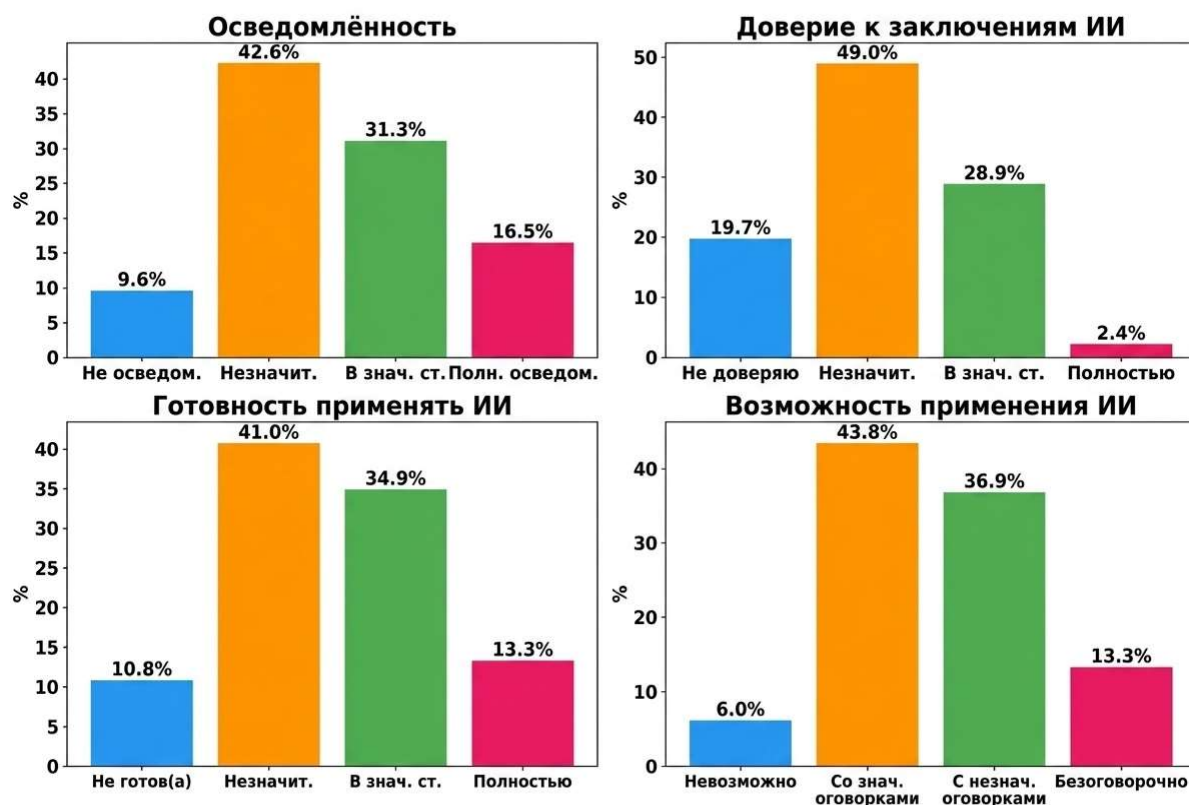


Рисунок 21 – Осведомленность, доверие, готовность и суждение о возможности применения ИИ в медицинской деятельности среди врачей-онкологов (n = 249)

Применявшиеся технологии искусственного интеллекта

Лидируют с большим отрывом чат-боты на основе больших языковых моделей (ChatGPT, GigaChat, YandexGPT), их отметили 124 респондента (49,8%). Системы компьютерного зрения для анализа КТ, МРТ и рентгенограмм назвали лишь 39 человек (15,7%), несмотря на то что именно эти технологии чаще всего ассоциируются с ИИ в онкологии в профессиональной и научной литературе. ИИ-агенты, СППР с ИИ и роботизированные системы применяют единицы. Ни одной технологии ИИ не применяли 100 человек (40,2%).

Такое распределение указывает на важный разрыв, знакомство онкологов с ИИ происходит преимущественно через текстовые генеративные системы (для написания текстов, поиска литературы, подготовки материалов), а не через специализированные клинические инструменты. Последние остаются малодоступными в повседневной клинической практике (Рисунок 22).

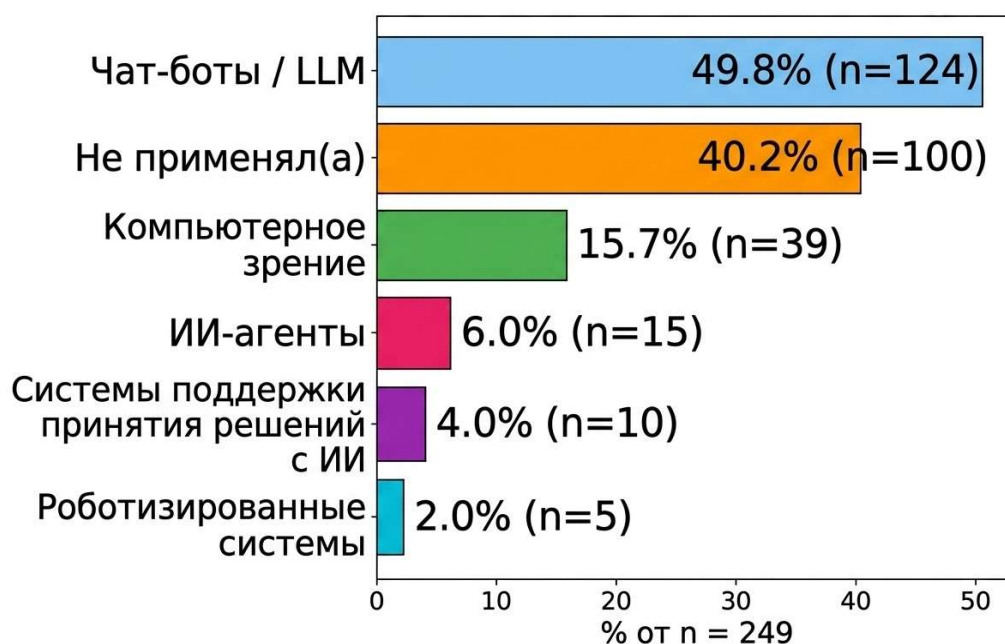


Рисунок 22 – Технологии ИИ, применявшиеся врачами-онкологами в медицинской деятельности (% от n = 249, множественный выбор)

Субъективная оценка точности искусственного интеллекта

Онкологи в среднем оценивают точность диагностики системами ИИ в $49,8 \pm 22,4\%$ (медиана 55%; IQR 35–65%), а точность диагностики врачами – в $72,9 \pm 19,5\%$ (медиана 75%; IQR 65–85%). Разрыв составляет 23 п.п. Средний балл субъективной оценки эффективности лечения, назначаемого врачами, составил $77,1 \pm 17,4\%$ (медиана 85%; IQR 75–85%), что выше субъективной оценки диагностической точности врачей.

Этот результат заслуживает отдельного внимания. Онколог несет ответственность за полный процесс, от верификации опухоли до выбора и коррекции схемы лечения. Высокая оценка эффективности лечения, вероятно, отражает уверенность в мультидисциплинарном подходе (взаимодействие с патоморфологами, радиологами, химиотерапевтами), который «страхует» каждое отдельное решение. Именно этот показатель является уникальным для онкологической выборки и отсутствует в исследованиях других специальностей (Рисунок 23).

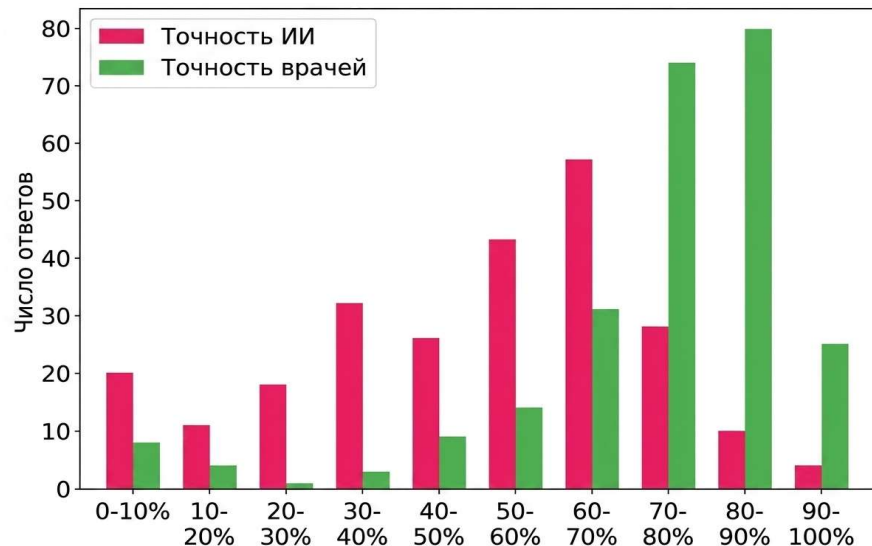


Рисунок 23 – Субъективная оценка точности диагностики: ИИ против врачей среди онкологов (n = 249)

Преимущества и ожидаемые проблемы внедрения искусственного интеллекта

Главное преимущество ИИ, по мнению онкологов, заключается в сокращении времени на выполнение рутинных задач (77,1%). Снижение нагрузки отметили 65,5% респондентов. Обращает внимание относительно скромная оценка повышения точности диагностики (31,3%), онкологи не склонны приписывать ИИ способность диагностировать лучше врача.

Среди возможных проблем использования ИИ лидируют ошибки диагностики (60,6%), однако ошибки при назначении лечения (57,8%) занимают практически равную позицию. Такой результат объясним, для онколога назначение химиотерапии, таргетных препаратов, выбор схемы и режима дозирования — это решения, где цена ошибки очень высока.

Распространение самодиагностики пациентами (58,6%) также находится в тройке лидеров, что для онкологии особенно чувствительно: самостоятельная интерпретация результатов исследований может спровоцировать панику или, наоборот, ложное успокоение пациента.

Характерная особенность онкологической выборки состоит в том, что при субъективной оценке возможность ошибки ИИ при назначении лечения (57,8%) практически равна возможности ошибки ИИ при диагностике (60,6%). В других специальностях опасения ошибок диагностики значительно преобладают над опасениями ошибок лечения (Рисунок 24).



Рисунок 24 – Преимущества и ожидаемые проблемы внедрения ИИ в медицинскую деятельность по мнению врачей-онкологов (% от n = 249, множественный выбор)

Целесообразные направления применения искусственного интеллекта

Первые две позиции ожидаемо заняли анализ данных и составление отчетов (69,1%) и оформление медицинской документации (63,1%), так как документарная нагрузка остается болевой точкой врачебного приема.

Третья позиция – скрининг (57,4%). Близко к скринингу расположилась телемедицина (56,6%), что отражает рост дистанционных консультаций в онкологии.

Применение ИИ для назначения лечения поддержали лишь 14,9% респондентов. Онкологи не готовы доверить алгоритмам ИИ самое ответственное решение, что закономерно: выбор схемы лечения при онкологическом заболевании всегда определяется мультидисциплинарным консилиумом, и делегирование этой функции алгоритму воспринимается как неприемлемый риск (Рисунок 25).

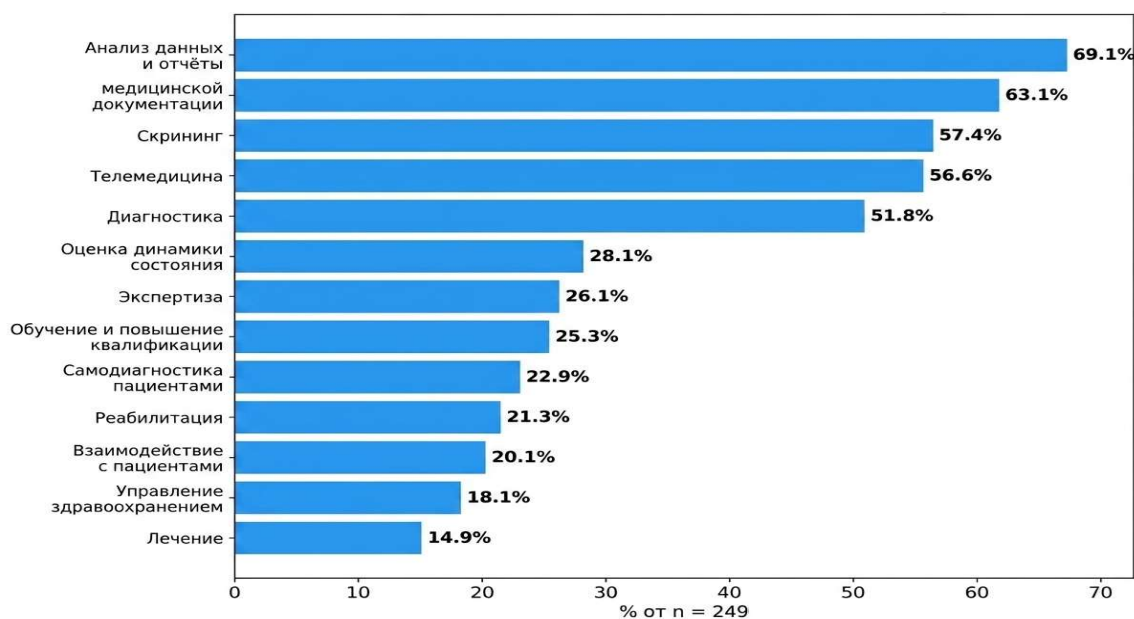


Рисунок 25 – Области медицинской деятельности, в которых целесообразно применение ИИ, по мнению врачей-онкологов (% от n = 249, множественный выбор)

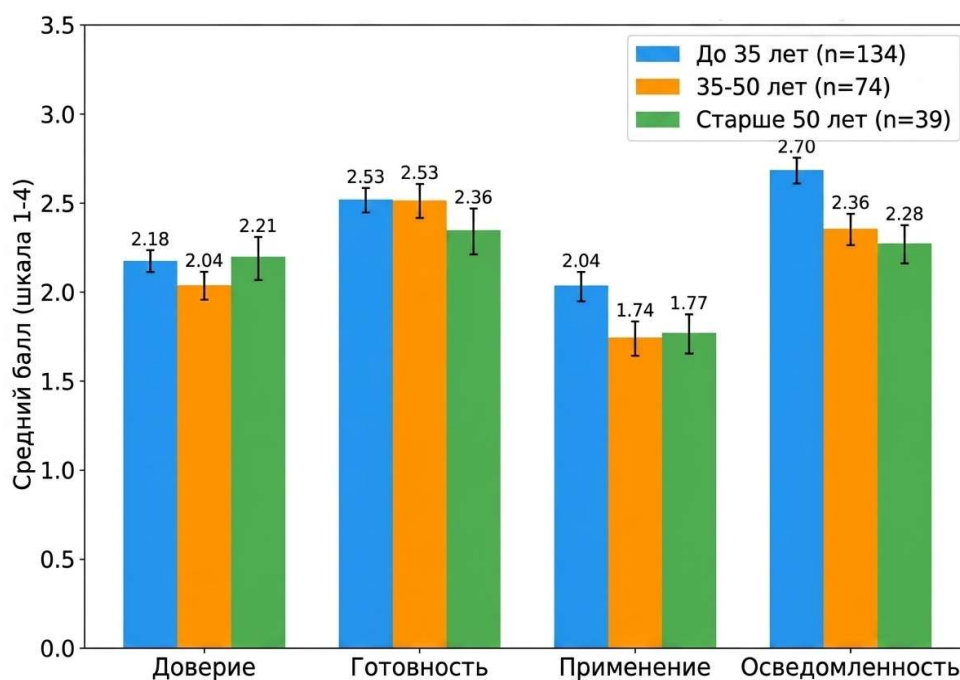
Сравнение возрастных групп

Молодые онкологи (до 35 лет) чаще применяют ИИ (2,04 балла против 1,77 у более старших) и выше оценивают свою осведомленность (2,70 против 2,28). При этом по уровню доверия возрастные группы практически не различаются: 2,18 у молодых и 2,21 у специалистов старше 50 лет ($U = 2575$; $p = 0,883$). Молодые онкологи чаще контактируют с ИИ, но это не делает их более доверчивыми. Можно предположить, что именно более глубокое понимание технологии формирует у них не слепое доверие, а скорее прагматизм (Таблица 6).

Таблица 6 – Показатели отношения врачей-онкологов к ИИ по возрастным группам

Показатель	До 35 лет (n=134)	35–50 лет (n=74)	Старше 50 (n=39)
Доверие к ИИ	2,18	2,04	2,21
Готовность	2,53	2,53	2,36
Применение ИИ	2,04	1,74	1,77
Осведомленность	2,70	2,36	2,28
Точность ИИ, %	51,4	47,8	47,8
Точность врачей, %	74,6	72,2	68,3

Аналогичная закономерность наблюдалась среди терапевтов: наиболее скептическими оказались не самые старшие, а врачи средней возрастной группы (35–50 лет). Во всех трех исследованных специальностях возрастной фактор не является статистически значимым предиктором доверия к ИИ (Рисунок 26).

Рисунок 26 – Сравнение возрастных групп врачей-онкологов по ключевым показателям отношения к ИИ ($M \pm SE$)

Роль цифровой грамотности

Связь между уровнем владения ПК и частотой применения ИИ в медицинской деятельности достоверна ($\rho = 0,127$; $p = 0,049$). При этом корреляция с доверием не достигает значимости ($\rho = 0,038$; $p = 0,559$). Динамика по группам уровней владения ПК показывает, что доверие к ИИ повышается уже на уровне офисных программ (2,16) и не растет у продвинутых пользователей (2,16). Частота применения, напротив, последовательно увеличивается от 1,70 у минимальных пользователей до 2,11 у продвинутых. Иными словами, цифровая грамотность снижает порог входа в использование ИИ, но не влияет на доверие к его результатам (Рисунок 27).

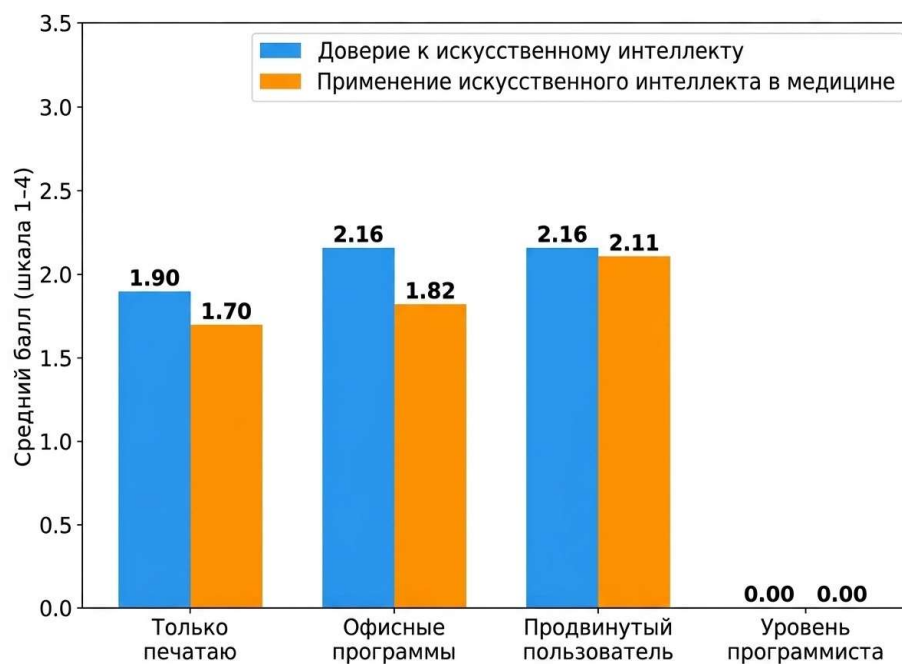


Рисунок 27 – Зависимость доверия к ИИ и частоты применения ИИ от уровня цифровой грамотности среди врачей-онкологов ($M \pm SE$)

Кластерный анализ по совокупности показателей доверия, готовности и применения искусственного интеллекта

Пороговый кластерный анализ разделил выборку на три группы. «Скептики» – 85 человек (34,1%) – демонстрируют минимальные значения по доверию,

готовности и применению. «Умеренные» – 112 (45,0%) – отмечают эпизодическое использование, осторожный оптимизм. «Энтузиасты» – 52 (20,9%) регулярно применяют ИИ с высоким доверием к нему.

Доля скептиков среди онкологов (34,1%) является наивысшей среди трех исследованных специальностей: у терапевтов она составляет 28,6%, у дерматовенерологов – 24,1%. Соответственно, доля энтузиастов у онкологов (20,9%) наименьшая. Эта закономерность согласуется с характером специальности, где цена ошибки высока и профессиональная ответственность за жизнь пациента формирует более осторожное отношение к любым новым технологиям.

Зависимость от типа организации – места работы

Почти 80% респондентов выборки составили врачи-онкологи амбулаторных медицинских организаций, 16,5% – врачи-онкологи стационарных медицинских организаций. Научные и образовательные организации представлены единичными наблюдениями (6 и 4 соответственно). Статистически значимых различий по доверию между онкологами амбулаторных и стационарных медицинских организаций не выявлено ($H = 0,75$; $p = 0,388$).

Тем не менее описательные данные представляют интерес. Онкологи, работающие в стационарах, показали самую низкую оценку точности ИИ (41,3% против 51,2% у амбулаторных онкологов), реже используют ИИ в медицинской деятельности (56% не используют против 42%) и в целом настроены скептичнее. Возможное объяснение заключается в том, что в стационарных медицинских организациях чаще всего проводится хирургическое лечение, химио- и лучевая терапия, в которых роль ИИ менее очевидна, чем на амбулаторном этапе со скрининговыми и диагностическими задачами и оформлением более объемной медицинской документации (Таблица 7, Рисунок 28).

Таблица 7 – Показатели отношения к ИИ по типу организации – места работы врачей-онкологов

Тип организации	n	Доверие к ИИ	Готовность к использованию ИИ	Применяли ИИ	Точность ИИ, %	Точность врачей, %	Не применяли ИИ
Амбулаторно-поликлиническая	198	2,16	2,54	1,95	51,2	74,2	84 (42%)
Стационарная	41	2,05	2,34	1,66	41,3	67,4	23 (56%)
Научная*	6	2,17	2,67	2,50	58,3	68,3	1 (17%)
Образовательная*	4	2,00	2,25	2,25	55,0	70,0	1 (25%)

Примечание: малый объем группы; интерпретация описательная

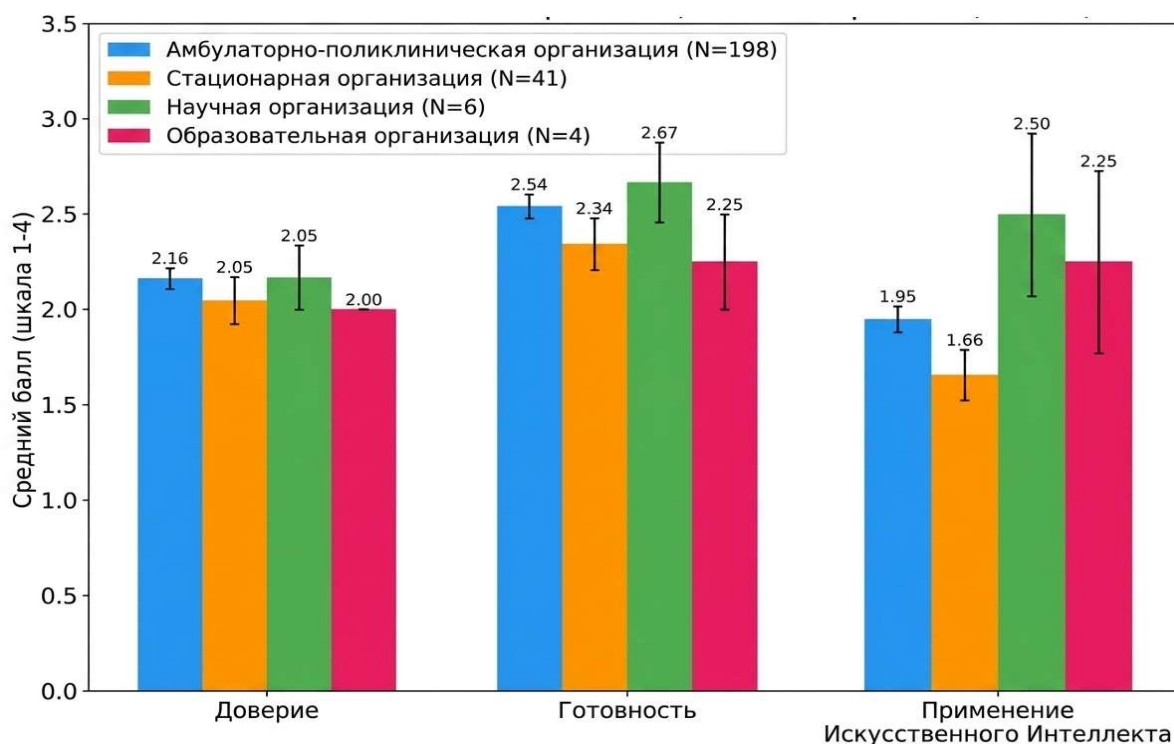


Рисунок 28 – Показатели отношения к ИИ в зависимости от типа организации – места работы врача-онколога ($M \pm SE$)

Опыт как предиктор доверия к технологиям искусственного интеллекта

Наиболее выраженное и статистически достоверное различие в выборке связано не с возрастом респондентов, не с местом работы и не с уровнем владения ПК, а с наличием практического опыта использования ИИ-технологий в медицине.

Среди 100 врачей, не применявших ни одной технологии ИИ в медицинской деятельности, 61 (61%) не используют ИИ вообще ни в одной сфере. Остальные 39 (39%) применяют его в бытовой, но не переносят в профессиональную

деятельность. В группе «применявших» ($n = 140$) ситуация зеркальная: 95 человек (68%) часто или периодически обращаются к ИИ и в повседневной жизни, не используют вовсе лишь 10 (7%).

Средний балл доверия у «применявших» составил 2,33 против 1,86 у «не применявших» ($U = 4778$; $p < 0,001$). Субъективная оценка точности ИИ – 54,9% против 42,2%. Разница в оценке точности составляет около 13 п.п.: те, кто имел опыт использования технологий ИИ в медицинской деятельности, оценивают точность существенно выше, а те, кто не имеет такого опыта, опираются на общее впечатление, которое нередко формируется по публикациям в средствах массовой информации и профессиональных чатах (Рисунок 29).

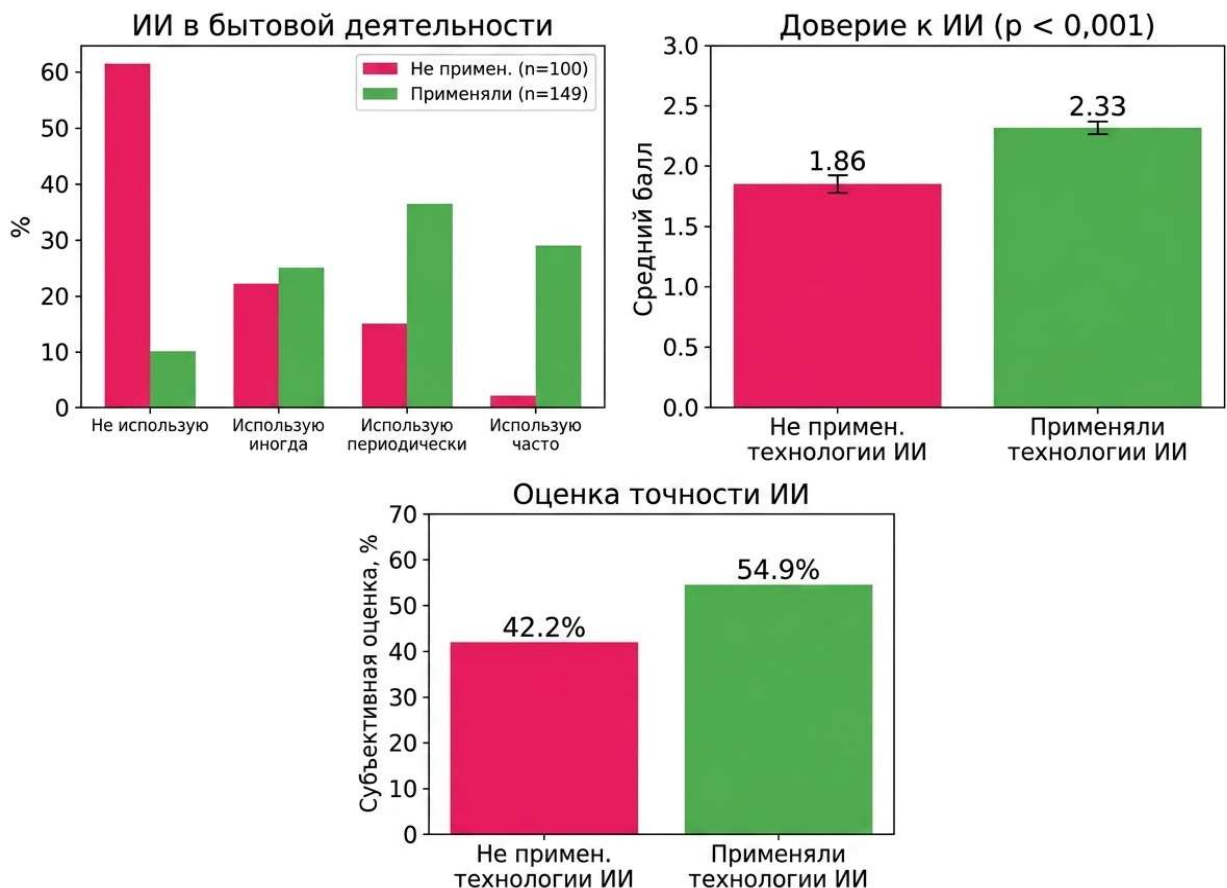


Рисунок 29 – Сравнение подгрупп «не применявших» и «применявших» технологии ИИ среди врачей-онкологов ($n = 249$)

Исследование впервые характеризует отношение российских врачей-онкологов к использованию технологий ИИ на достаточно большой выборке ($n = 249$).

Основной вывод: доверие к ИИ у онкологов невысокое (68,7% скептиков), но детерминировано не возрастом и не типом организации, а наличием или отсутствием личного опыта работы с ИИ-технологиями ($p < 0,001$). Те, кто имеет опыт применения ИИ в медицинской деятельности, оценивают точность ИИ на 13 п.п. выше тех, кто такого опыта не имеет.

Специфика онкологической выборки проявляется в трех чертах. Во-первых, опасения за ошибки при назначении лечения практически равны опасениям за ошибки диагностики, что уникально для данной специальности. Во-вторых, онкологи оценивают эффективность лечения, назначаемого врачом (77,1%), выше диагностической точности врача-онколога (72,9%), что, вероятно, отражает роль мультидисциплинарных консилиумов. В-третьих, скрининг занимает непропорционально высокое место (57,4%) среди перспективных направлений ИИ, что соответствует профессиональному контексту онкологической службы.

3.1.5. Сравнительный анализ отношения к технологиям искусственного интеллекта врачей разных специальностей

Общая характеристика сравниваемых выборок

Сравнительный анализ выполнен по материалам трех опросов, проведенных по единой методологии (одна анкета, одна платформа, один период наблюдения) среди врачей трех специальностей, терапевтов и врачей общей практики ($n = 105$), дерматовенерологов ($n = 191$) и онкологов ($n = 249$). Суммарная выборка составила 545 специалистов. Единый инструмент (26 вопросов, идентичные шкалы и кодировка) обеспечивает сопоставимость результатов.

Выборки различаются по возрастной структуре. Наиболее молодая дерматовенерологическая (медиана 29 лет), что объясняется высокой долей

ординаторов (50,8%). Терапевты представляют наиболее «зрелую» выборку (медиана 35 лет, средний возраст 40,2 года). Онкологи занимают промежуточное положение (медиана 33 года, средний возраст 37,2 года). Гендерное распределение во всех трех выборках смещено в сторону женщин: от 64,7% среди онкологов до 85,9% среди дерматовенерологов. Все три выборки характеризуются концентрацией в московском регионе и преобладанием специалистов амбулаторно-поликлинических медицинских организаций (Таблица 8).

Таблица 8 – Сводная характеристика выборок

Параметр	Терапевты и ВОП (n=105)	Дерматовенерологи (n=191)	Онкологи (n=249)
Средний возраст, лет	40,2 ± 17,3	34,4 ± 12,1	37,2 ± 11,0
Медиана возраста, лет	35	29	33
Женщины, %	71,4	85,9	64,7
Ученая степень, %	3,8	18,9	9,6
Амбулаторные МО, %	66,7	–	79,5

Применение искусственного интеллекта в медицинской и немедицинской деятельности

Частота неприменения ИИ в профессиональной медицинской деятельности существенно различается между специальностями 31,4% у терапевтов, 33,5% у дерматовенерологов и 43,8% у онкологов. Терапевты и врачи общей практики демонстрируют наиболее широкое проникновение ИИ в профессиональную деятельность. Онкологи, напротив, реже обращаются к ИИ в медицинских целях, несмотря на ежедневный контакт с цифровыми диагностическими инструментами.

Еще более показательным является разрыв между бытовым и профессиональным применением. У терапевтов он минимален около 5 п.п., у онкологов составляет 13 п.п., у дерматовенерологов достигает 18 п.п. Этот показатель отражает степень «переноса» бытовой цифровой активности в профессиональную сферу. Низкий разрыв у терапевтов объясняется тем, что LLM-системы, освоенные в быту, органично встраиваются в оформление медицинской документации, ключевую задачу первичного звена. Высокий разрыв у

дерматовенерологов указывает на то, что бытовое использование ИИ (генеративные чат-боты) слабо конвертируется в профессиональные задачи, для которых нужны специализированные системы компьютерного зрения (Таблица 9).

Таблица 9 – Частота применения и разрыв между бытовым и профессиональным использованием ИИ

Показатель	Терапевты и ВОП	Дерматовенерологи	Онкологи
Не применяют ИИ в медицине, %	31,4	33,5	43,8
Не применяют ИИ в быту, %	26,7	15,2	30,5
Разрыв п.п.	4,7	18,3	13,3

Доверие к искусственному интеллекту

Уровень скептицизма в отношении ИИ сопоставим во всех трех специальностях, но несколько нарастает от терапевтов к онкологам: 64,8% скептически настроенных среди терапевтов, 66,5% среди дерматовенерологов и 68,7% среди онкологов. Средний балл доверия (по 4-балльной шкале) находится в диапазоне 2,1–2,3 во всех трех группах, что соответствует категории «доверяю незначительно».

Интересно, что более высокий уровень скептицизма онкологов согласуется с характером специальности, в которой цена диагностической и лечебной ошибки особенно высока. Вместе с тем абсолютные различия невелики и не превышают 4 п.п., что свидетельствует о том, что дефицит доверия к ИИ является общемедицинским явлением, а не проблемой конкретной специальности.

Готовность к применению искусственного интеллекта

Готовность к использованию ИИ во всех трех специальностях существенно опережает уровень доверия. Суммарно готовы применять ИИ 87,6% терапевтов, 88,0% дерматовенерологов и 89,2% онкологов. Показатели практически идентичны. Феномен «не доверяю, но готов попробовать» является универсальным

для всех исследованных специальностей и указывает на то, что российские врачи в целом открыты к знакомству с технологией, даже если пока не имеют оснований для ее безоговорочного принятия.

Субъективная оценка точности искусственного интеллекта

Субъективная оценка точности ИИ у терапевтов и онкологов практически совпадает (49,8%), тогда как дерматовенерологи оценивают точность ИИ еще ниже (медиана 45%). Субъективная оценка точности диагностики врачами также сопоставима, 72,4% у терапевтов, 75% (медиана) у дерматовенерологов, 72,9% у онкологов. Онкологи дополнительно оценивали эффективность назначаемого лечения (77,1%), что является уникальной характеристикой данной специальности.

Разрыв между субъективной оценкой точности ИИ и врачей наибольший у дерматовенерологов (около 30 п.п.) и наименьший у терапевтов и онкологов (около 23 п.п.). Во всех трех специальностях субъективная оценка точности ИИ существенно занижена по сравнению с реальными клиническими показателями (80–95%), что формирует устойчивый «разрыв восприятия» (Таблица 10).

Таблица 10 – Субъективная оценка точности диагностики по специальностям

Показатель	Терапевты и ВОП	Дерматовенерологи	Онкологи
Точность ИИ (медиана), %	55	45	55
Точность врачей (медиана), %	75	75	75
Разрыв, п.п.	20	~30	20

Преимущества и ожидаемые проблемы внедрения искусственного интеллекта

Во всех трех специальностях врачи единодушны в определении главного преимущества ИИ что это сокращение рутины. Конкретные показатели составляют 71,4% у терапевтов, 77,8% у дерматовенерологов и 77,1% у онкологов. Второе

место стабильно занимает снижение нагрузки на врачей: 61,0%, 66,5% и 65,5% соответственно. Примечательно, что повышение точности диагностики ценится менее высоко (31,3–38,1%), что отражает сохраняющуюся уверенность врачей в собственных диагностических компетенциях.

Структура ожидаемых проблем, напротив, демонстрирует специфичность. Лидирующая позиция ошибок диагностики универсальна (60,6–73,8%), однако абсолютные значения различаются: дерматовенерологи наиболее обеспокоены (73,8%), терапевты и онкологи умеренно (61,0% и 60,6%). Ключевое отличие онкологической выборки практически равная обеспокоенность ошибками диагностики (60,6%) и ошибками лечения (57,8%). В других специальностях разрыв между этими позициями значительно больше. Это объясняется тем, что в онкологии лечебное решение (выбор схемы химиотерапии, назначение таргетных препаратов) является столь же критичным, как и диагностический процесс (Таблица 11).

Таблица 11 – Топ-3 ожидаемых проблем ИИ по специальностям

Ранг	Терапевты и ВОП	Дерматовенерологи	Онкологи
1	Ошибки диагностики (61,0%)	Ошибки диагностики (73,8%)	Ошибки диагностики (60,6%)
2	Самодиагностика пациентами (57,1%)	Ошибки лечения (69,6%)	Самодиагностика пациентами (58,6%)
3	Ошибки лечения (55,2%)	Самодиагностика пациентами (62,8%)	Ошибки лечения (57,8%)

Целесообразные направления применения искусственного интеллекта

Врачи всех трех специальностей единодушно определяют оформление документации и анализ данных как первоочередные задачи для внедрения ИИ. Скрининг занимает третью позицию во всех группах. Различия проявляются в относительном весе отдельных направлений.

Для терапевтов на первом месте стоит оформление документации (64,8%), что отражает колоссальную долю документооборота в работе первичного звена. У дерматовенерологов лидирует анализ данных (68,1%), а скрининг и диагностика

получают более высокие оценки, чем у других специальностей, что соответствует визуальному характеру специальности. У онкологов первая позиция также за анализом данных (69,1%), а скрининг (57,4%) занимает особенно значимое место, отражая масштаб скрининговых программ в онкологии.

Общей чертой является крайне низкая поддержка применения ИИ для назначения лечения 20,0% у терапевтов, менее 20% у дерматовенерологов и 14,9% у онкологов. Врачи всех специальностей не готовы делегировать алгоритмам наиболее ответственные клинические решения. Наименьший показатель у онкологов закономерен в онкологии лечебные решения принимаются коллегиально на консилиумах и несут максимальную ответственность.

Возрастной фактор

Ни в одной из трех исследованных специальностей возраст не являлся статистически значимым предиктором доверия к ИИ. У терапевтов $H = 1,87$; $p = 0,392$. У онкологов доверие молодых (2,18) и старших (2,21) практически идентично ($U = 2575$; $p = 0,883$). У дерматовенерологов даже при 17-летнем разрыве в медианном возрасте между ординаторами и практикующими врачами доверие не различается ($p = 0,982$).

Этот результат воспроизводится во всех трех выборках и является одним из ключевых выводов исследования, «возрастной барьер» в отношении к ИИ в медицине не подтверждается эмпирически. Молодые врачи чаще контактируют с ИИ (в быту и в профессии), но не более доверчивы к его заключениям. Наиболее скептическими в двух из трех выборок оказались врачи средней возрастной группы (35–50 лет), а не самые старшие специалисты.

Практический опыт как ключевой предиктор

Во всех трех специальностях единственным устойчивым предиктором позитивного отношения к ИИ является практический опыт работы с ИИ-технологиями в медицинской деятельности. У терапевтов разница в доверии между «применявшими» и «не применявшими» статистически значима ($U = 1394$; $p = 0,014$), разрыв в оценке точности ИИ составляет 9 п.п. У онкологов эта связь еще более выражена ($U = 4778$; $p < 0,001$), разрыв в оценке точности достигает 13 п.п.

Механизм этой связи, по всей видимости, носит двунаправленный характер. С одной стороны, контакт с технологией позволяет скорректировать заниженные ожидания, сформированные медиаобразом ИИ. С другой стороны, врачи, изначально более открытые к технологиям, с большей вероятностью пробуют ИИ-инструменты. Тем не менее устойчивость этой закономерности во всех трех выборках указывает на то, что ключевым условием формирования профессионального доверия к ИИ является не информирование, а непосредственный контакт специалиста с технологией (Таблица 12).

Таблица 12 – Доверие к ИИ в зависимости от наличия практического опыта

Показатель	Терапевты и ВОП	Онкологи
Доверие «применявших», баллы	2,36	2,33
Доверие «не применявших», баллы	1,89	1,86
p	0,014	$< 0,001$
Оценка точности «применявших», %	52,1	54,9
Оценка точности «не применявших», %	43,2	42,2
Разрыв, п.п.	9	13

Кластерная типология

Кластерный анализ по единой методологии (пороговое агрегирование по среднему трех показателей) выявил трехкомпонентную структуру во всех специальностях, однако с различным соотношением групп (Таблица 13).

Таблица 13 – Кластерное распределение по специальностям

Кластер	Терапевты и ВОП	Дерматовенерологи	Онкологи
Скептики	30 (28,6%)	46 (24,1%)	85 (34,1%)
Умеренные	48 (45,7%)	96 (50,3%)	112 (45,0%)
Энтузиасты	27 (25,7%)	49 (25,7%)	52 (20,9%)

Наибольшая доля скептиков зафиксирована среди онкологов (34,1%), наименьшая среди дерматовенерологов (24,1%). Доля энтузиастов, напротив, максимальна у дерматовенерологов и терапевтов (25,7%) и минимальна у онкологов (20,9%). Группа «Умеренных» стабильно составляет около 45–50% во всех специальностях, что свидетельствует об устойчивости этого типа отношения к ИИ.

Более высокий скептицизм онкологов объясняется комбинацией факторов, высокая цена ошибки в онкологии, более низкая частота применения ИИ в профессиональной деятельности (43,8% не применяют), а также уникальная для онкологии обеспокоенность ошибками лечения наряду с ошибками диагностики. Более низкий скептицизм дерматовенерологов, возможно, связан с молодым возрастным составом выборки, хотя этот фактор проявляется не в доверии (которое одинаково), а в снижении порога принципиального неприятия технологии.

Применявшиеся технологии искусственного интеллекта

В ответах врачей всех трех специальностей безусловным лидером являются чат-боты на основе LLM. Их используют 49,5% терапевтов, 63,9% дерматовенерологов и 49,8% онкологов. Более высокая доля у дерматовенерологов объясняется молодым возрастным составом выборки и высокой бытовой цифровой активностью.

Системы компьютерного зрения занимают вторую позицию во всех специальностях, но с существенно различной долей, 17,1% у терапевтов, 8,9% у дерматовенерологов и 15,7% у онкологов. Парадоксально низкий показатель у дерматовенерологов (8,9%), несмотря на визуальный характер специальности,

подчеркивает разрыв между доказанной эффективностью систем компьютерного зрения в дерматоскопии и их реальной доступностью в клинической практике.

Специфические черты каждой специальности

Каждая из исследованных специальностей обнаруживает характерные особенности восприятия ИИ, которые определяются профессиональным контекстом.

Терапевты и врачи общей практики отличаются минимальным разрывом (5 п.п.) между бытовым и профессиональным применением ИИ. Это объясняется тем, что LLM-системы, освоенные в повседневной жизни, естественным образом встраиваются в оформление медицинской документации центральную задачу первичного звена. Приоритетом терапевтов является документооборот, что точно отражает структуру их рабочего времени. Телемедицина занимает у терапевтов более высокое место, чем у других специалистов, что связано с ростом дистанционных консультаций в первичном звене.

Дерматовенерологи характеризуются наибольшим пользователем современных технологий, несмотря на наиболее молодой возраст, наибольшую бытовую цифровую активность и наибольшую долю пользователей LLM-систем, они не демонстрируют более высокого доверия к ИИ. Разрыв между бытовым и профессиональным применением у них наибольший (18 п.п.), что указывает на дефицит доступных специализированных ИИ-инструментов для дерматовенерологии. Дерматовенерологи присваивают скринингу и диагностике более высокий рейтинг, чем другие специальности, что логично для дисциплины с визуальным характером. Обеспокоенность ошибками диагностики у них максимальна (73,8%).

Онкологи отличаются наивысшим уровнем скептицизма (68,7% скептиков) и наибольшей долей неприменяющих ИИ в медицине (43,8%). Уникальной чертой является практически равная обеспокоенность ошибками диагностики и ошибками лечения, что не наблюдается в других специальностях. Онкологи также

единственные, кто оценивал эффективность назначаемого лечения (77,1%), причем эта оценка оказалась выше оценки диагностической точности (72,9%), что отражает роль мультидисциплинарного подхода. Скрининг занимает у онкологов особенно значимое место среди перспективных направлений ИИ (57,4%).

Проведенный сравнительный анализ позволяет выделить ряд универсальных закономерностей, воспроизводящихся во всех трех специальностях.

Первое. Дефицит доверия к ИИ является общемедицинским явлением. От 64,8% до 68,7% врачей вне зависимости от специальности оценивают свое доверие к алгоритмическим заключениям как незначительное или отсутствующее.

Второе. Диссонанс между доверием и готовностью. При низком доверии от 87,6% до 89,2% врачей готовы применять ИИ. Этот разрыв указывает на то, что российские врачи находятся на ранней стадии знакомства с технологией: они осознают ее потенциал, но не имеют достаточного клинического опыта для формирования обоснованного суждения.

Третье. Практический опыт является ключевым предиктором позитивного отношения. Во всех выборках врачи с опытом применения ИИ демонстрируют достоверно более высокое доверие и более высокую оценку точности. Ни возраст, ни тип организации, ни уровень владения ПК не обладают сопоставимой предсказательной силой.

Четвертое. Возрастной барьер в отношении к ИИ эмпирически не подтверждается ни в одной из трех специальностей.

Пятое. Субъективная оценка точности ИИ существенно занижена по сравнению с реальными клиническими показателями. Разрыв восприятия (22–30 п.п.) формирует и воспроизводит скептицизм.

Шестое. Врачи всех специальностей видят основное преимущество ИИ в сокращении рутины и снижении нагрузки, а не в повышении диагностической точности. Ни одна из специальностей не готова делегировать ИИ назначение лечения.

Седьмое. Знакомство российских врачей с ИИ происходит преимущественно через генеративные текстовые системы (LLM-чат-боты), а не через

специализированные клинические инструменты. Последние остаются малодоступными, что ограничивает формирование профессионального доверия.

Эти закономерности определяют практические приоритеты внедрения ИИ в российское здравоохранение: включение работы с ИИ-системами в программы дополнительного профессионального образования, организация пилотных проектов в медицинских организациях, обеспечивающих контролируемый контакт специалиста с технологией в рабочей среде, создание открытых реестров верифицированных ИИ-систем с публично доступными данными об их диагностической эффективности, а также разработка четкой регуляторной правовой основы, определяющей ответственность за решения с участием ИИ.

3.2. Анализ возможностей существующих программ для электронных вычислительных машин на основе моделей искусственного интеллекта для диагностики заболеваний кожи

3.2.1. Методика поиска и отбора публикаций

Развитие цифровых технологий и повсеместное распространение смартфонов существенно расширили возможности применения алгоритмов компьютерного зрения в дерматовенерологии. Появление программных продуктов, способных проводить предварительную оценку поражений кожи по фотоизображениям, создало предпосылки для проведения их систематического сопоставления. Особую актуальность приобретает задача объективного сравнения таких программ по ключевым диагностическим метрикам, поскольку каждая из них разрабатывалась на различных наборах данных, обучалась с использованием разных архитектур нейросетей и проходила апробацию на разных когортах пациентов.

С целью получения объективных данных об эффективности существующих программ для ЭВМ (мобильных приложений), реализующих алгоритмы искусственного интеллекта в дерматовенерологии, был выполнен оригинальный

сравнительный анализ. Проведен целевой поиск публикаций в базах данных PubMed/MEDLINE и Google Scholar, а также в электронных библиотеках eLibrary и КиберЛенинка за период с 2016 по 2025 год, что позволило охватить наиболее актуальные разработки в данной области.

Поисковые запросы формулировались с учетом синтаксиса каждой базы данных и были ориентированы на выявление публикаций, в которых описываются конкретные программы, использующие сверточные нейронные сети (convolutional neural network, CNN) для классификации кожных поражений на мобильных платформах. В Google Scholar применялся запрос, включающий терминологию CNN, дерматовенерологии и мобильных приложений. Для PubMed/MEDLINE использовался структурированный запрос с булевой логикой, объединяющий понятия искусственного интеллекта, глубокого обучения, дерматовенерологии, мобильных приложений и ключевых метрик диагностической эффективности. В русскоязычных библиотеках запросы включали термины «искусственный интеллект», «нейронные сети», «приложение для смартфона», «болезни кожи», «опухоли кожи», «диагностика».

Всего было обнаружено 1319 публикаций. Из них удалены 456 дубликатов, после чего осталось 863 уникальные записи. Далее на основании заголовков и аннотаций исключено 738 нерелевантных работ и 38 статей, для которых не был доступен полный текст. Таким образом, для полнотекстового анализа осталось 87 публикаций. После повторного применения критериев включения и невключения окончательно в анализ вошли 9 ключевых статей, содержащих исчерпывающие и сопоставимые данные об эффективности, архитектуре и практическом применении конкретных программ.

Критериями включения служили наличие описания конкретного мобильного приложения на основе ИИ для анализа фотоизображений кожи, представление количественных метрик эффективности (точность, чувствительность, специфичность), оригинальность исследования или клинического испытания, а также доступность полного текста на английском или русском языке. Критериями невключения являлись технические отчеты и маркетинговые материалы без

данных независимой валидации, исследования без четких метрик производительности, публикации о телемедицинских платформах без автоматизированного ИИ-анализа, работы, не прошедшие независимое рецензирование, а также дубликаты.

Методика отбора публикаций представлена на Рисунке 30.



Рисунок 30 – Блок-схема отбора публикаций для сравнительного анализа программ ИИ в дерматовенерологии

3.2.2. Методика сравнительного анализа

Для сравнения отобранных программ использовались ключевые параметры, а именно точность (accuracy), чувствительность (sensitivity), специфичность (specificity), архитектура нейросетевой модели, а также применимость в клинической практике.

3.2.3. Результат поиска и отбора программ для электронных вычислительных машин, основанных на технологиях искусственного интеллекта, для применения в дерматовенерологии

По результатам поиска и скрининга были идентифицированы 9 программ для ЭВМ (мобильных приложений) на основе ИИ для диагностики и мониторинга кожных патологий: Google DermAssist, SkinIO, Melanoma Check, Derma Onko Check, SkinVision, Tibot, SkinScan, Aysa и Skinive.

3.2.4. Характеристики анализируемых программ для электронных вычислительных машин, основанных на технологиях искусственного интеллекта, для применения в дерматовенерологии

Результаты систематизации данных о рассматриваемых программах представлены в таблицах ниже. Характеристики каждого приложения, включая целевую аудиторию, платформу, размещение модели ИИ, особенности интерфейса, необходимость дополнительного оборудования и регуляторный статус, отражены в Таблице 14.

Таблица 14 – Характеристики мобильных приложений на основе ИИ для диагностики патологий кожи

Программа, страна	Пользователи	ОС	Размещение модели	Описание	Интерфейс	Допол. оборуд.	Регуляторы	Регистрация как МИ
Google DermAssist, США	Врачи, пациенты	iOS, Android	Облачная обработка	Анализ фотоизображений кожи и клинических симптомов	+	Нет	н/д	н/д
SkinIO, США	Врачи, пациенты	Android	н/д	Сравнительный анализ поражений кожи в динамике по 3D-фотоизображениям всего тела, до 24 позиций	+	Нет	н/д	н/д
Melanoma Check, Россия	Врачи	Android	Периферийное устройство	Классификация пигментных поражений (меланома / не меланома) по фото или дерматоскопическим изображениям	+	Нет	Роспатент	н/д
Derma Onko Check, Россия	Врачи	Android	Периферийное устройство	Классификация доброкачественных и злокачественных опухолей кожи по фото или дерматоскопическим изображениям	+	Нет	Роспатент	н/д
SkinVision, Нидерланды	Врачи, пациенты	н/д	Облачная обработка	Скрининг и определение риска наличия рака кожи и других заболеваний кожи по фотоизображениям	+	Нет	СЕ	Да (EU MD)
Tibot, Индия	Врачи, пациенты	Android, iOS	н/д	Анализ фотоизображений, топ-3 диагноза	+	Нет	н/д	н/д

Продолжение Таблицы 14

SkinScan, США	Врачи, пациенты	н/д	н/д	Анализ дерматоскопических изображений	+	Да (дерматоскоп)	н/д	н/д
Aysa, Индия	Врачи, пациенты	Android, iOS	н/д	8 диагнозов на основе анализа фотоизображений и клинических симптомов	+	Нет	н/д	н/д
Skinive, Нидерланды	Врачи, пациенты	н/д	н/д	Оценка риска наличия кожного заболевания, интеграция с API по фотоизображениям	+	Нет	н/д	н/д

Примечание: ОС – операционная система; МИ – медицинское изделие; API (Application Programming Interface) – интерфейс прикладного программирования; CE (Conformité Européenne) – европейское соответствие; EU MD – медицинское изделие в ЕС; н/д – нет данных

Среди рассмотренных программ имелись как решения, ориентированные исключительно на врачей (Melanoma Check, Derma Onko Check), так и приложения двойного назначения, доступные и для пациентов, и для специалистов (Google DermAssist, SkinIO, SkinVision, Tibot, SkinScan, Aysa, Skinive). Преобладающей операционной системой являлся Android, ряд приложений поддерживал также платформу iOS. Модели ИИ размещались как на облачных серверах (Google DermAssist, SkinVision, Skinive), так и на периферийном устройстве (Melanoma Check, Derma Onko Check), что обеспечивало автономность работы без постоянного подключения к интернету.

Большинство анализируемых приложений использовало камеру смартфона без дополнительного оборудования, за исключением SkinScan, для работы которого требовался высококачественный цифровой дерматоскоп. С точки зрения регуляторного статуса выделялись приложение SkinVision, получившее маркировку CE и зарегистрированное как медицинское изделие в Европейском Союзе, а также российские программы Melanoma Check и Derma Onko Check, зарегистрированные в Федеральной службе по интеллектуальной собственности (Роспатент).

Архитектурные особенности и параметры обучения нейросетевых моделей, заложенных в анализируемые приложения, представлены в Таблице 15. В большинстве случаев в качестве базовой архитектуры применялись сверточные нейронные сети. Приложение Melanoma Check построено на основе MobileNet, что обеспечивает компактность модели и возможность ее исполнения на мобильных устройствах без обращения к облаку. Derma Onko Check использует архитектуру Xception, отличающуюся повышенной точностью за счет глубинно-разделимых сверток. Для остальных приложений конкретные архитектуры не были раскрыты в публикациях, однако указывалось использование CNN в качестве общего класса моделей.

Таблица 15 – Архитектура и параметры обучения моделей ИИ в мобильных приложениях для анализа патологий кожи

Программа, страна	Число классов	Архитектура	Метод обработки изображений	Вид обучения	Тип обучения	Число изображений	Доп. обработка
Google DermAssist, США	3	CNN	н/д	ГО	н/д	16 114	н/д
SkinIO, США	н/д	н/д	н/д	ГО	н/д	н/д	н/д
Melanoma Check, Россия	2	MobileNet	Классификация	ГО	ОсУ	10 000	Аугментация
Derma Onko Check, Россия	2	Xception	Классификация	ГО	ОсУ	25 000	Аугментация
SkinVision, Нидерланды	н/д	CNN	н/д	ГО	н/д	н/д	н/д
Tibot, Индия	н/д	CNN	н/д	ГО	н/д	н/д	н/д
SkinScan, США	н/д	CNN	Сегментация + классификация	ГО	н/д	н/д	н/д
Aysa, Индия	н/д	н/д	н/д	н/д	н/д	н/д	н/д
Skinive, Нидерланды	н/д	CNN	н/д	ГО	н/д	н/д	н/д
Примечание: CNN (convolutional neural network) – сверточная нейронная сеть; ОсУ – обучение с учителем; ГО – глубокое обучение; н/д – нет данных							

Объем обучающих выборок существенно различался между программами. Google DermAssist обучался на массиве из 16 114 изображений, Derma Onko Check на 25 000, Melanoma Check на 10 000 изображениях. Для остальных приложений данные об объеме обучающих наборов в рецензированных публикациях отсутствовали. Melanoma Check и Derma Onko Check дополнительно использовали аугментацию данных для расширения обучающей выборки. Во всех случаях, где тип обучения был указан, применялось глубокое обучение с учителем.

3.2.5. Сравнение программ для электронных вычислительных машин, основанных на технологиях искусственного интеллекта, для применения в дерматовенерологии по точности и эффективности

Ключевым этапом исследования стало сопоставление диагностической эффективности рассматриваемых программ. Результаты представлены в Таблице 16.

Таблица 16 – Эффективность мобильных приложений на основе ИИ для диагностики патологий кожи

Программа, страна	Число объектов апробации	Точность (Ac), %	Чувствительность (Se), %	Специфичность (Sp), %	ЛО, %	ЛП, %
Google DermAssist, США	н/д	до 97	н/д	н/д	н/д	н/д
SkinIO, США	25 пациентов, 262 изображения	н/д	92	н/д	8	н/д
Melanoma Check, Россия	133 пациента	89,5	92,5	88,2	7,5	11,8
Derma Onko Check, Россия	135 пациентов	96	98	96	2,4	4,3
SkinVision, Нидерланды	18 960 изображений	н/д	87-95	70-78	5-13	22-30
Tibot, Индия	600 пациентов	80,6 (топ-1), 96,1 (топ-3)	н/д	н/д	н/д	н/д
SkinScan, США	н/д	н/д	80,76	85,5	79,24	14,43
Aysa, Индия	700 пациентов	н/д	71,0 (топ-1), 86,1 (топ-3), 95,1 (топ-8)	н/д	29,0 (топ-1), 13,9 (топ-3), 4,9 (топ-8)	н/д
Skinive, Нидерланды	н/д	н/д	95,3-97,9	93,5-97,1	2,1-4,7	2,9-6,5

Примечание: Ac (accuracy) – точность; Se (sensitivity) – чувствительность; Sp (specificity) – специфичность; ЛО – ложноотрицательные результаты; ЛП – ложноположительные результаты; топ-1, топ-3, топ-8 – вероятность того, что верный диагноз окажется соответственно на 1-м месте, в первых 3 или в первых 8 позициях из списка предложений ИИ; н/д – нет данных

Проведенный сравнительный анализ позволил выявить значительные различия в диагностической эффективности рассмотренных программ, которые

обусловлены особенностями их архитектуры, целевым назначением и качеством данных, использованных при обучении моделей.

Наиболее высокие показатели продемонстрировали Google DermAssist и Derma Onko Check. Точность классификации Google DermAssist достигала 97%, что объясняется использованием крупного обучающего набора (16 114 изображений) и ресурсов облачной инфраструктуры. Derma Onko Check показал точность 96% и чувствительность 98% при специфичности 96%, что является одним из лучших сбалансированных результатов среди всех рассмотренных программ. Данная программа, построенная на архитектуре Xception и обученная на 25 000 изображениях с аугментацией, была апробирована на выборке из 135 пациентов и продемонстрировала уровень ложноотрицательных результатов всего 2,4%, что критически важно при скрининге злокачественных новообразований.

Программа Skinive, разработанная в Нидерландах, продемонстрировала высокие показатели чувствительности (95,3-97,9%) и специфичности (93,5-97,1%) с улучшением метрик в динамике между 2020 и 2021 годами. Такая положительная динамика свидетельствует об активном дообучении нейросетевых моделей и расширении обучающих выборок. Уровень ложноотрицательных результатов Skinive составил 2,1-4,7%, а ложноположительных 2,9-6,5%.

Российская программа Melanoma Check, построенная на архитектуре MobileNet, достигла точности 89,5%, чувствительности 92,5% и специфичности 88,2% при апробации на 133 пациентах. Эти результаты указывают на достаточно надежную предварительную классификацию пигментных поражений кожи, хотя уровень ложноположительных результатов (11,8%) свидетельствует о необходимости дальнейшей оптимизации модели.

Приложение SkinIO было апробировано на ограниченной выборке (25 пациентов, 262 изображения) и показало чувствительность 92% при уровне ложноотрицательных результатов 8%. Ограниченный объем апробации не позволяет сделать окончательные выводы о надежности данного инструмента.

Приложение SkinVision было протестировано на наиболее крупной выборке среди рассматриваемых программ (18 960 изображений) и продемонстрировало

чувствительность в диапазоне 87-95%. Вместе с тем специфичность данного приложения оказалась сравнительно низкой (70-78%), что способно приводить к увеличению числа ложноположительных срабатываний (22-30%) и, как следствие, к росту количества дополнительных консультаций специалистов.

Индийское приложение Tibot, апробированное на 600 пациентах, показало точность 80,6% при выборе одного наиболее вероятного диагноза (топ-1) и 96,1% при включении трех наиболее вероятных (топ-3). Данные о чувствительности и специфичности для Tibot в рецензированных публикациях не были представлены.

Приложение SkinScan, требующее использования высококачественного цифрового дерматоскопа, продемонстрировало чувствительность 80,76% и специфичность 85,5%. Несмотря на умеренные показатели, необходимость специализированного оборудования существенно ограничивает возможности его применения в первичном звене медицинской помощи.

Наиболее низкие показатели чувствительности при выборе первого диагноза были зафиксированы у приложения Aysa (71,0% для топ-1), хотя при расширении списка предложений до топ-3 и топ-8 чувствительность возрастала до 86,1% и 95,1% соответственно. Апробация на 700 пациентах выявила уровень ложноотрицательных результатов 29,0% для первого предложенного диагноза, что может быть критичным при скрининге злокачественных поражений.

Визуальное сопоставление ключевых метрик эффективности исследуемых программ представлено на Рисунке 31.

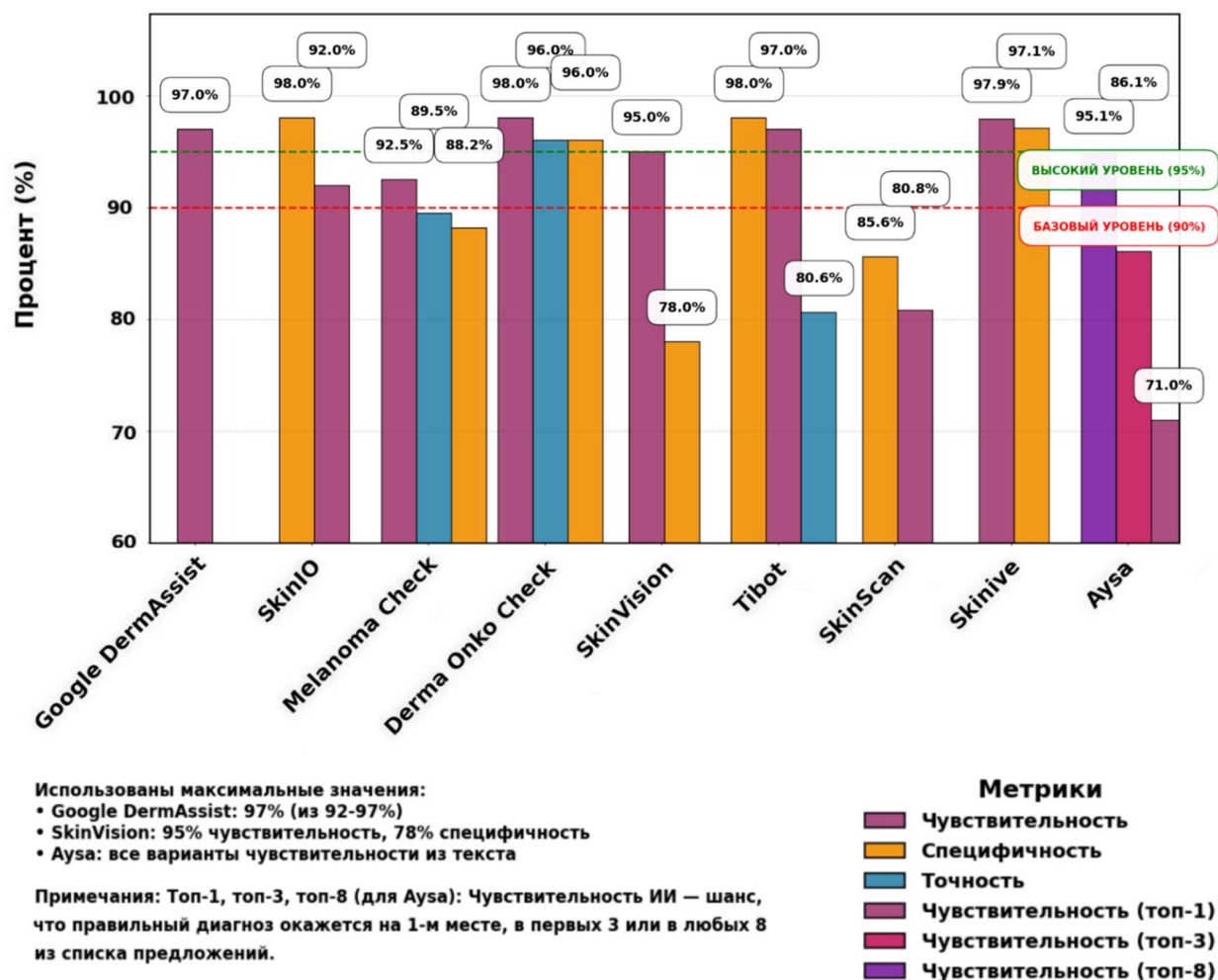


Рисунок 31 – Сравнительный анализ эффективности мобильных приложений на основе ИИ, применяемых в клинической дерматовенерологии (использованы максимальные значения метрик; для Aysa приведены все варианты чувствительности: топ-1, топ-3, топ-8)

3.2.6. Области применения, архитектурные особенности и общие ограничения анализируемых программ для электронных вычислительных машин, основанных на технологиях искусственного интеллекта, для применения в дерматовенерологии

Области применения анализируемых программ охватывают несколько направлений клинической практики. Первое и наиболее развитое направление связано со скринингом рака кожи, включая раннее выявление меланомы. Данному направлению посвящены программы Google DermAssist, Melanoma Check, Derma Onko Check, SkinVision и SkinScan. Второе направление связано с динамическим

мониторингом хронических дерматозов, таких как экзема и псориаз. Третье направление представляет собой инструменты самодиагностики для пациентов, позволяющие получить предварительную оценку кожного поражения до обращения к врачу. Вместе с тем все без исключения рассмотренные программы нуждаются в дополнительной валидации в реальных клинических условиях.

С точки зрения архитектуры большинство программ работает на платформах iOS и Android, при этом часть приложений использует облачные серверы для обработки и анализа изображений, а часть выполняет вычисления непосредственно на устройстве. В ряде случаев реализованы гибридные модели, сочетающие алгоритмический анализ с последующим контролем со стороны врача-специалиста посредством телемедицинской консультации. Географически программы разработаны преимущественно в США, России, Нидерландах и Индии.

Приложения, использующие архитектуру CNN, такие как Google DermAssist, SkinVision, Tibot и Skinive, показали высокую чувствительность в диапазоне 87-97,9%, что имеет принципиальное значение для минимизации ложноотрицательных результатов при скрининге меланомы. Однако специфичность у различных программ варьировалась существенно. Так, приложение SkinVision имело сравнительно невысокий показатель специфичности (78%), что неизбежно ведет к увеличению потока направлений на дополнительные консультации специалистов. В то же время у приложений SkinIO и Tibot специфичность достигала 98%, что позволяет минимизировать число ложных тревог.

Показатели эффективности ряда программ оказались сопоставимы с диагностической точностью опытных врачей-дерматовенерологов, что подтверждает их потенциал как инструментов второго мнения, особенно в отдаленных районах и регионах с недостаточным количеством специалистов.

К общим ограничениям анализируемых программ следует отнести несколько факторов. Точность работы алгоритмов снижается при анализе изображений пациентов с темными фототипами кожи, что обусловлено недостаточной представленностью таких изображений в обучающих выборках. Качество входного

фотоизображения оказывает прямое влияние на результат классификации, что особенно актуально для приложений Google DermAssist и SkinVision. Присутствует значимый риск ложноположительных результатов, ведущих к необоснованным дополнительным обследованиям. Отсутствует интеграция с клиническими данными пациента, такими как результаты биопсии и анамнестические сведения. Возникают этические вопросы, связанные с конфиденциальностью данных и предвзятостью обучающих алгоритмов. Имеются регуляторные барьеры, затрудняющие повсеместное внедрение программ в практическое здравоохранение.

Отдельно необходимо отметить, что некоторые приложения, в частности SkinScan, требуют специализированного оборудования (высококачественного цифрового дерматоскопа), что снижает доступность для первичного звена. Приложения Aysa и Tibot ограничены в диагностике редких патологий, таких как фотодерматозы и сложные опухоли, что указывает на необходимость дальнейшего дообучения моделей на расширенных и более разнообразных наборах данных.

Принципиально важным остается положение о том, что ни одна из рассмотренных программ не способна заменить морфологическое (гистологическое) исследование биопсийного материала, и результаты их работы должны интерпретироваться квалифицированным врачом-специалистом.

3.2.7. Перспективы развития программ искусственного интеллекта в дерматовенерологии

Перспективы совершенствования мобильных приложений на основе ИИ в дерматовенерологии связаны с несколькими направлениями. Первым является интеграция программ с телемедицинскими системами, позволяющая осуществлять удаленный скрининг и мониторинг хронических кожных заболеваний. Вторым направлением служит улучшение алгоритмов для работы с редкими нозологиями и различными фототипами кожи, что потребует формирования более сбалансированных и репрезентативных обучающих выборок. Третье направление

состоит в стандартизации подходов к тестированию моделей ИИ для повышения воспроизводимости результатов и обеспечения сопоставимости данных, полученных в различных исследованиях.

Программы, продемонстрировавшие чувствительность и специфичность выше 95% (в частности, *Derma Onko Check* и *Skinive*), обладают наибольшим потенциалом для формирования универсальных инструментов скрининга в первичном звене медицинской помощи. Дальнейшее развитие данного направления может быть связано с переходом от классических сверточных нейронных сетей к мультимодальным архитектурам и трансформерам, способным учитывать не только изображение кожного поражения, но и дополнительную клиническую информацию о пациенте.

Таким образом, проведенный сравнительный анализ 9 программ для ЭВМ (мобильных приложений) на основе ИИ, применяемых в клинической дерматовенерологии, показал, что наиболее высокие и сбалансированные показатели диагностической эффективности имеют программы *Google DermAssist* (точность до 97%), *Derma Onko Check* (точность 96%, чувствительность 98%, специфичность 96%) и *Skinive* (чувствительность до 97,9%, специфичность до 97,1%). Эти программы демонстрируют потенциал, сопоставимый с диагностическими возможностями опытных врачей-дерматовенерологов, и могут быть использованы в качестве инструментов второго мнения, особенно в условиях ограниченного доступа к специализированной помощи. Вместе с тем выявленные ограничения (зависимость от качества изображений, диспаратеты по фототипам кожи, отсутствие интеграции с клиническими данными, этические и регуляторные вопросы) указывают на необходимость продолжения исследований и доработки существующих решений.

3.3. Анализ медицинских изделий с искусственным интеллектом, зарегистрированных в Российской Федерации

По состоянию на апрель 2026 г. в реестре Росздравнадзора зарегистрировано 57 медицинских изделий (МИ) с ИИ. Первое российское МИ с ИИ (СППВР WEBIOMED) зарегистрировано в апреле 2020 г.

Сравнение перечня медицинских изделий с искусственным интеллектом, зарегистрированных в Российской Федерации, с реестрами США и Китая по объему, структуре и динамике регистраций

В перечне FDA (США) по состоянию на март 2026 г. содержится 1524 МИ с ИИ (учет ведется с 1995 г.), а в реестре NMPA (Китай) насчитывается 154 МИ с ИИ (с 2020 г.). По числу допущенных МИ реестр FDA превосходит китайский приблизительно в 10 раз, а российский в 27 раз. Все три реестра демонстрируют устойчивый рост числа ежегодных регистраций (Рисунок 32), в 2025 г. в США зарегистрировано 333, в Китае 45, в России 13 МИ с ИИ.

Реестр Росздравнадзора на 91,2% сформирован отечественными разработчиками (52 из 57 МИ), тогда как в реестре FDA доля зарубежных производителей достигает 57,7%. Ведущими отечественными разработчиками являются «Care Mentor AI» (5 МИ), «ПТМ» (4 МИ), «СберМедИИ» и ФГБУ «НМИЦ им. В.А. Алмазова» (по 3 МИ), а также «К-Скай», «РТК Радиология» и «Медицинские скрининг системы» (по 2 МИ).

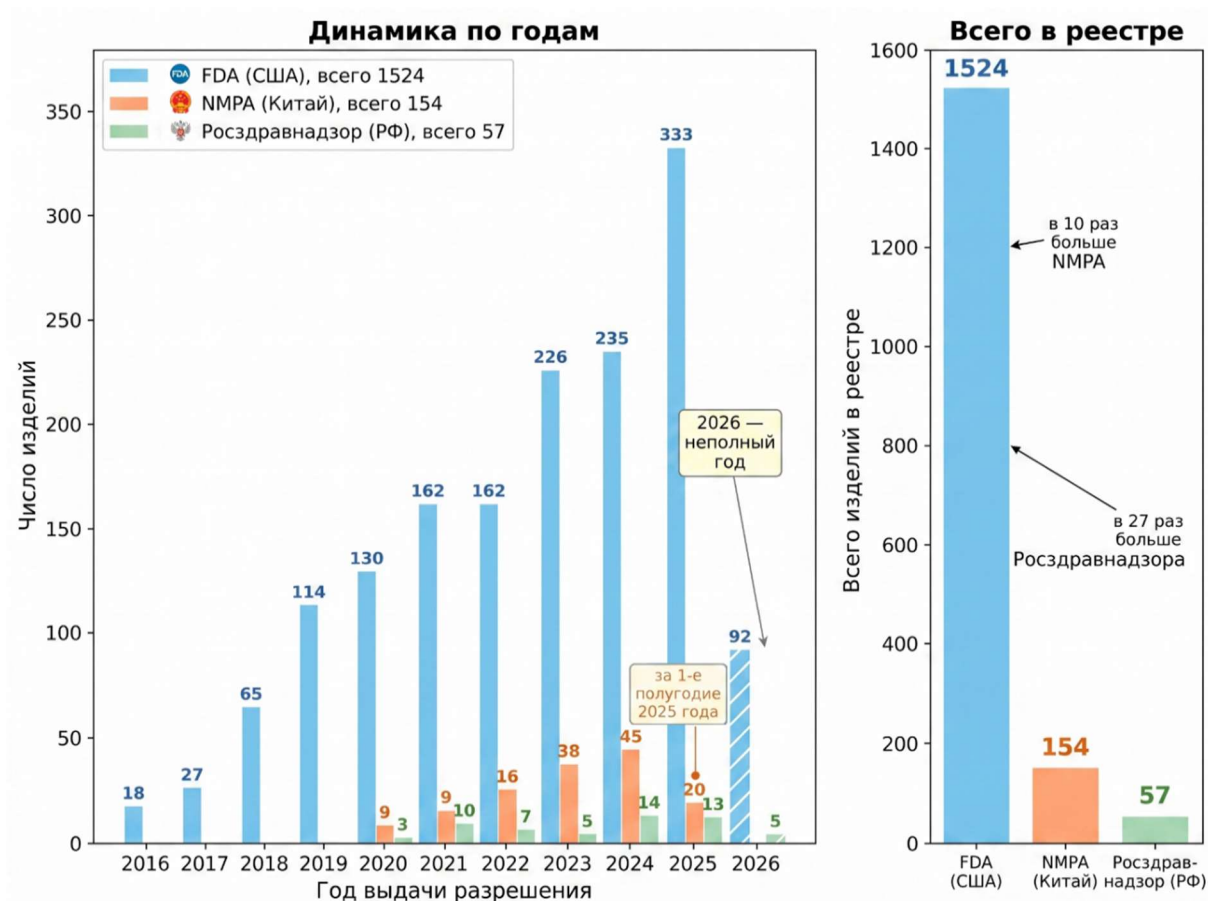


Рисунок 32 – Ежегодное число медицинских изделий с искусственным интеллектом, допущенных к медицинскому применению в США, Китае и России, 2016–2026 гг.

Структура производителей различается между странами (Рисунок 33), в США доминируют транснациональные производители медицинского оборудования (GE 105 МИ, Siemens 90 МИ, Philips 45 МИ), в Китае национальные специализированные ИИ-компании (Yukun и United Imaging по 17 МИ, Infervision 14 МИ, Deepwise 11 МИ), в России независимые отечественные разработчики ПО.

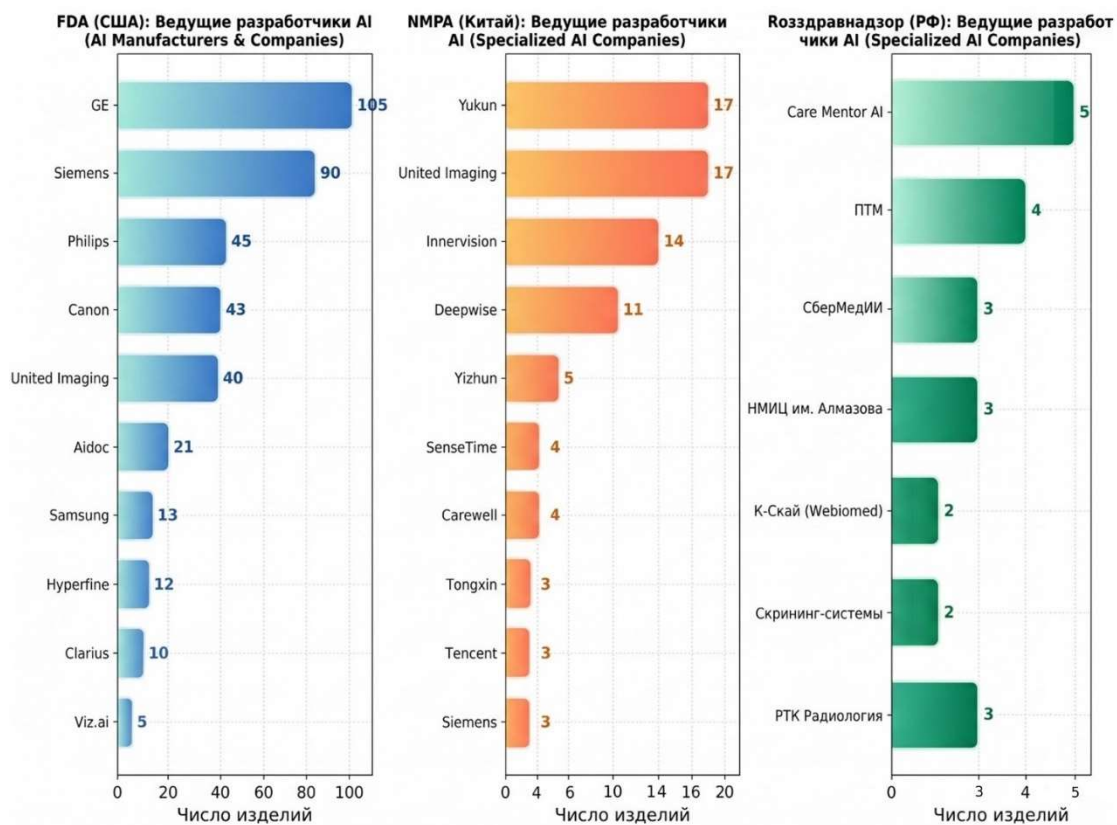


Рисунок 33 – Ведущие производители медицинских изделий с ИИ в реестрах FDA (США), NMPA (Китай) и Росздравнадзора (РФ)

Во всех трех реестрах доминируют МИ для анализа медицинских изображений (Рисунок 34).

В реестре FDA на лучевую диагностику приходится 76,4%, NMPA 68,8%, в реестре Росздравнадзора 59,65%, а суммарная доля МИ для анализа изображений составляет 77,6%, 83,8% и 73,68% соответственно. Сравнительная характеристика реестров приведена в Таблице 17.

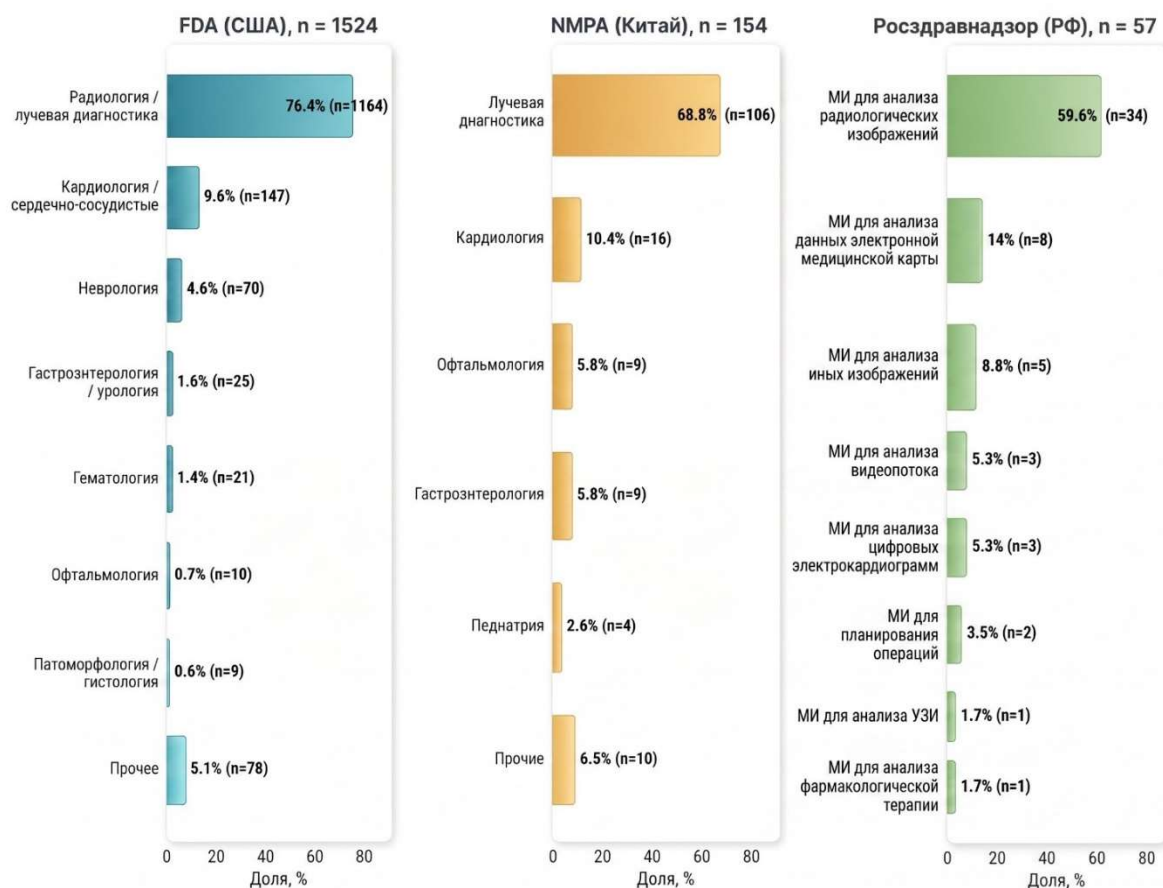


Рисунок 34 – Структура реестров медицинских изделий с ИИ США, Китая и России по областям медицины

Таблица 17 – Сравнительная характеристика реестров медицинских изделий с ИИ, зарегистрированных в США, Китае и РФ

Показатель	FDA (США)	NMPA (Китай)	Росздравнадзор (РФ)
Период учета	с 1995 г.	с 2020 г.	с 2020 г.
Всего МИ	1524	154	57
Первое МИ с ИИ	1995 г. (AutoPap)	2020 (FFR, Keya)	2020 г. (WEBIOMED)
Регистраций в 2025 г.	333	45	13
Основная область и доля	лучевая диагностика 76,4%	лучевая диагностика 68,8%	лучевая диагностика 59,65%
Доля МИ для анализа изображений	77,6%	83%	73,68%
Доля отечественных разработчиков	≈42% (2024 г.)	н/д	91,2%

Распределение зарегистрированных в России медицинских изделий с искусственным интеллектом по основным областям медицинского применения

Распределение зарегистрированных в России МИ с ИИ по основным областям медицинского применения показано на Рисунке 35 и в Таблице 18. Абсолютное большинство приходится на лучевую диагностику, 34 из 57 МИ (59,6%). Наибольшее число МИ предназначено для анализа изображений органов грудной клетки (13 МИ). Далее следуют МИ для маммографии (7 МИ), СППВР для общей врачебной практики (9 МИ, 15,8%), функциональная диагностика в кардиологии (3 МИ), эндоскопическая диагностика (3 МИ), стоматология (3 МИ), неврология (3 МИ), ортопедия и травматология (3 МИ) и офтальмология (2 МИ). Дерматовенерология представлена одним МИ (сервис «ПроРодинки»).

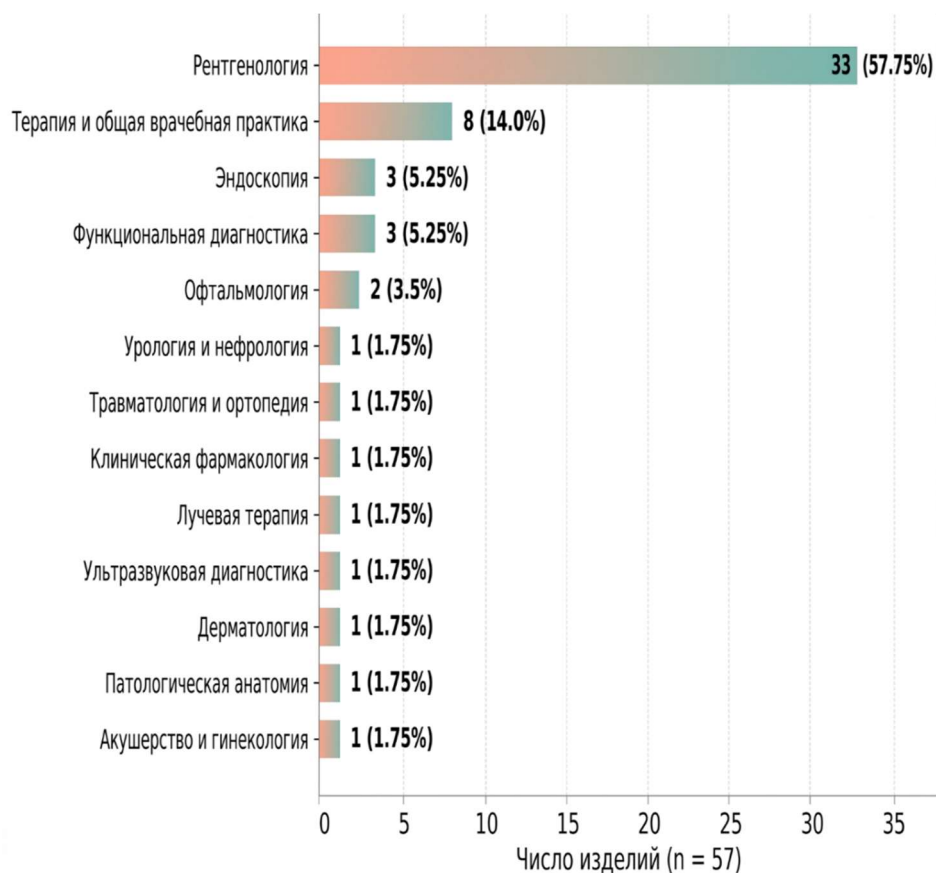


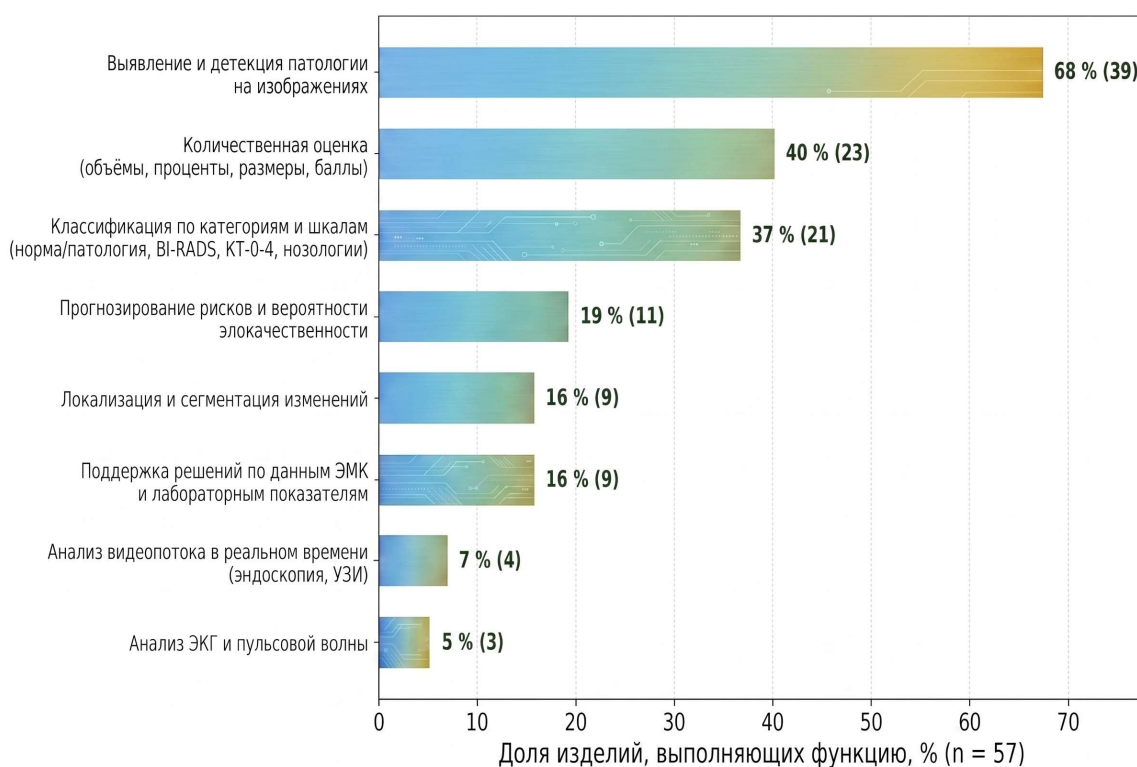
Рисунок 35 – Распределение зарегистрированных в России (апрель 2026 г.) медицинских изделий с ИИ по основной области медицинского применения (n=57)

Таблица 18 – Распределение МИ с ИИ (РФ) по типу входных данных и по основной области медицинского применения (медицинской специальности) (n = 57)

Тип входных данных	Основная область медицинского применения (медицинская специальность)	Описание МИ с ИИ	Число МИ с ИИ	Доля, %
Результат лучевых исследований	Рентгенология	Анализ цифровых изображений рентгенографии, флюорографии, КТ, МРТ, маммографии, КЛКТ, ортопантомографии	33	57,75
		в том числе (возможен анализ нескольких областей):		
		органов грудной клетки	17	29,75
		молочной железы	8	14,0
		головного мозга	4	7,0
		челюстно-лицевой области	3	5,25
		позвоночника	2	3,5
		забрюшинного пространства	1	1,75
		коленных суставов	1	1,75
		стоп	1	1,75
		предстательной железы	1	1,75
	Урология и нефрология	Построение и работа с 3D-моделью по снимкам УЗИ, мультиспиральной КТ, МРТ и ПЭТ для планирования вмешательства.	1	1,75
	Травматология и ортопедия	Анализ цифровых рентгенограм и КТ для планирования эндопротезирования тазобедренных и коленных суставов	1	1,75
<i>Итого по типу входных данных</i>			35	61,25
Структурированные и неструктурированные данные ЭМК и клинические данные	Терапия (терапевтические специальности), общая врачебная практика	Система поддержка принятия врачебных решений (СППВР)	8	14,0
	Клиническая фармакология	Анализ рациональности назначений лекарств	1	1,75
	Лучевая терапия	Планирование и анализ лучевой терапии	1	1,75
<i>Итого по типу входных данных</i>			10	17,5

Распределение зарегистрированных в России медицинских изделий с искусственным интеллектом по функциям и основным возможностям

По функциям и основным возможностям ИИ в российских МИ (Рисунок 36) преобладает выявление (детекция) патологических изменений на медицинских изображениях (24 МИ), далее следуют количественная оценка и измерения (10 МИ), поддержка принятия врачебных решений (7 МИ), классификация по нозологиям (6 МИ), анализ видеопотока в реальном времени (4 МИ), анализ ЭКГ и пульсовой волны (3 МИ), планирование медицинских вмешательств (3 МИ) (Таблица 19).



Одно изделие может выполнять несколько функций, поэтому сумма долей превышает 100 %.

Рисунок 36 – Задачи и основные возможности искусственного интеллекта в медицинских изделиях, зарегистрированных в РФ (n=57)

Таблица 19 – Основные функции и возможности зарегистрированных в РФ медицинских изделий с искусственным интеллектом

Основные функции и возможности МИ с ИИ	МИ с ИИ	Количество МИ с ИИ
Выявление (детекция) патологических изменений на медицинских изображениях	«Botkin.AI» пр-ва «Интеллоджик»; ПО пр-ва «КэреМенторЭйАй» (×3); ПО «ЦЕЛЬС» пр-ва «Медицинские скрининг системы» (×3); программные модули пр-ва «ПТМ» (×3); платформа RADLogics пр-ва «Радлоджикс Рус»; ПО пр-ва СберМедИИ (×2); ПО пр-ва «ФБМ»; «DENTOMO» пр-ва «НМФ ФДЛАБ»; «Sciberia Lungs» пр-ва «Сайберия»; «NTechMed CT Brain» пр-ва «НТех лаб»; «Intelligent Radiology Assistants» пр-ва «АЙРА ЛАБС»; «Diagnocat» пр-ва «Диагнокат»; ПО пр-ва «РТК Радиология»; «CVL – Chest CT Complex» пр-ва «СИБИЖИНЛАБ»; «Просвет. РГ позвоночника. Сколиоз» пр-ва «Оксиджен технолodgeиес Рус»; «AIDiagnostic» пр-ва «Интел Диагностик»; «IntelliSpace Portal» пр-ва Филипс Медикал Системс Нидерланд Б.В. (DynaCAD)	24
Количественная оценка и измерения на медицинских изображениях (объемы, проценты, размеры, баллы)	«Гамма Мультивокс» пр-ва «Гаммамед-Софт»; ПО пр-ва «КэреМенторЭйАй» (×2); «JEMYS:ТЕЛЕМЕДИЦИНА» пр-ва «ЮСАР+»; программный модуль пр-ва «ПТМ»; «ТриодМ-МТ» пр-ва «МТЛ»; ПО пр-ва СберМедИИ; ПО пр-ва «РТК Радиология»; «Мамметрикс» пр-ва «Мамметрикс-ИТ»; «EzOrtho» пр-ва «Эвусофт Ко., Лтд»	10
Поддержка принятия врачебных решений и прогнозирование по клиническим данным (ЭМК, лабораторные исследования и пр.)	«Webiomed» пр-ва «К-Скай»; «MedicBK» пр-ва «МедикБук»; «Электронный клинический фармаколог» пр-ва «Соцмедика»; «WEBIOMED.DHRA» пр-ва «К-Скай»; ПО пр-ва НМИЦ им. В.А. Алмазова; «АРИТМОТЕСТ®» пр-ва «Прима диагностика»; «Диагностический ассистент» пр-ва «СберМедИИ»	7
Классификация по нозологиям и степеням выраженности патологических признаков	«ПроРодинки» пр-ва «АИМЕД»; «Retina.AI» пр-ва «ДИДЖИТАЛ ВИЖН СОЛЮШНС»; «PathVision.ai» пр-ва «Цифровой онкоморфолог»; ПО пр-ва НМИЦ им. В.А. Алмазова; «Galenos.AI» пр-ва «ТехЛАБ»; «Офтальмик» пр-ва ФГБОУ ВО «СИБГМУ»	6
Анализ видеопотока в реальном времени	«ArtInCol» пр-ва «Алнисофт»; ПО пр-ва НМИЦ им. В.А. Алмазова; «EW10-EC02» пр-ва ФУДЖИФИЛЬМ Корпорейшн; «PENTAX Medical SAS-M10» пр-ва «ХОЯ Корпорейшн»	4
Анализ ЭКГ и пульсовой волны	ПО пр-ва «ТИС»; ПО пр-ва ФГБУН «Институт системного программирования имени В. П. Иванникова» Российской академии наук; ПО пр-ва ФГАОУ ВО «Первый МГМУ имени И.М. Сеченова»	3

Продолжение Таблицы 19

Планирование медицинских вмешательств	«Модуль ортопедии и планирования эндопротезирования» пр-ва «РТК Радиология»; Sechenov.AI_nephro пр-ва ФГАОУ ВО «Первый МГМУ имени И.М. Сеченова»; «RayStation 8B» пр-ва «RaySearch Laboratories AB»	3
Всего		57

Таким образом ИИ в российских медицинских МИ выступает преимущественно в роли «второго мнения» для врачей при оценке результатов лучевых исследований. СППВР по клиническим данным являются отличительной чертой российского реестра и не имеют прямого аналога в реестрах FDA и NMPA. Многие клинические области пока охвачены лишь единичными разработками что обозначает приоритетные направления для отечественных разработчиков и регуляторов.

Описание зарегистрированного в Российской Федерации медицинского изделия с искусственным интеллектом для дерматовенерологии

Единственным зарегистрированным в реестре Росздравнадзора медицинским изделием с ИИ для диагностики новообразований кожи является сервис поддержки принятия врачебных решений «ПроРодинки» (регистрационное удостоверение РЗН 2023/21776 от 27 декабря 2023 г., класс потенциального риска 3), разработанный ООО «АИМЕД» (Москва, территория Сколково Инновационного центра). Программа предназначена для помощи врачу при принятии решений при диагностике новообразований кожи в области дерматовенерологии, онкологии и общей врачебной практики.

По архитектуре сервис «ПроРодинки» представляет собой программное обеспечение с моделью распространения SaaS и включает два нейросетевых модуля сверточной нейронной сети, модуль детекции для обнаружения новообразований на фотоизображении и модуль классификации для определения характера выявленного образования. Обе модели построены на базе предварительно обученных архитектур с открытым исходным кодом. Сервис не

имеет собственного графического интерфейса и предоставляет результаты через клиентский модуль (веб-приложение или мобильное приложение) посредством API.

Согласно регистрационной документации, программа распознает пять нозологических единиц, меланому, базальноклеточный рак кожи, меланоцитарный невус, себорейный кератоз и гемангиому. Заявленные технические характеристики составляют чувствительность не менее 85%, специфичность не менее 80%, точность 81,5%. Однако выходное заключение сервиса формируется в бинарном формате «доброкачественное» или «злокачественное», то есть несмотря на заявленную многоклассовую классификацию из пяти нозологий конечному пользователю предоставляется результат бинарной классификации без указания конкретного диагноза.

Внешняя проспективная валидация алгоритма «ПроРодинки» проведена на базе Онкологического центра ГKB им. С.С. Юдина на выборке из 132 пациентов (86 женщин и 46 мужчин, средний возраст $73,3 \pm 9,19$ года) с жалобами на новообразования кожи в период с августа 2024 по январь 2025 г [369]. Результаты гистологических и цитологических исследований использовались как референсный стандарт. Результаты тестирования представлены в Таблице 20. По совокупности гистологических и цитологических заключений алгоритм продемонстрировал чувствительность 86% (95% ДИ 78–92%), специфичность 41% (95% ДИ 21–64%), точность 79% (95% ДИ 71–85%), прогностическую ценность положительного результата 88% (95% ДИ 80–93%) и прогностическую ценность отрицательного результата 38% (95% ДИ 19–59%). Из 132 пациентов число с подтвержденным ЗНО (БКК) составило 110, число с подтвержденным ДНО 22. При этом в 75% случаев фотоизображения имели неудовлетворительное качество (объект вне фокуса, размытое изображение), что могло повлиять на специфичность.

Таблица 20 – Таблица соответствия заключений программы «ПроРодинки» заключениям цитологических и гистологических заключений (Васильев Ю.А. и соавт., 2026)

	Индекс-тест Истинный	Индекс-тест Ложный
Референс-тест Положительный	95	13
Референс-тест Отрицательный	9	15

В сравнении с программой «Derma Check» для диагностики новообразований кожи и дерматозов, сервис «ПроРодинки» имеет ряд существенных отличий. «ПроРодинки» использует одномодельную одномодальную архитектуру с бинарным выходным заключением, тогда как «Derma Check» построена на мультимодельном ансамбле нейронных сетей с механизмом голосования и многоклассовым выходом обеспечивающем формальную чувствительность 96,7% и клиническую точность 95,3%. В программе «ПроРодинки» отсутствуют модули контроля качества фотоизображений, объяснимого ИИ, интеграции анамнестических данных, сегментации образования, анализа асимметрии и функции динамического наблюдения в кабинете врача. Вместе с тем следует отметить, что разработчики «ПроРодинки» активно развивают продукт, в частности реализован модуль самодиагностики для пациентов, позволяющий отслеживать новообразования в динамике, что является важным направлением расширения доступности скрининговых технологий для населения, которые могут улучшить раннюю выявляемость злокачественных новообразований кожи.

3.4. Сравнительный анализ баз данных и наборов данных, используемых для обучения моделей искусственного интеллекта для программ для диагностики заболеваний кожи

Выбор источника данных для формирования датасета для обучения моделей программ ИИ является одним из первых и наиболее значимых методологических решений при разработке ПКЗ. Это решение во многом предопределяет, на каких пациентах и в каких клинических условиях обученная система будет работать надежно, а где ее точность окажется неудовлетворительной. Систематический

анализ доступных баз данных, как международных публичных, так и отечественных институциональных, позволяет осознанно подходить к этому выбору и заранее идентифицировать присущие каждому источнику ограничения.

Типология клинических баз данных

Прежде чем переходить к сравнительному анализу конкретных баз данных, необходимо зафиксировать основания, по которым они могут быть систематизированы, поскольку именно эти основания определяют пригодность того или иного источника для конкретной задачи. По типу хранимой информации клинические базы данных распадаются на три большие группы. Базы изображений, системы PACS, хранящие миллионы DICOM-исследований, рентгенограммы, КТ, МРТ, дерматоскопические архивы. Базы структурированных данных, электронные медицинские карты и лабораторные информационные системы с демографическими сведениями, результатами обследований, лабораторными показателями. Базы текстовых документов, системы хранения выписных эпикризов, протоколов операций, консультативных заключений. В современных медицинских информационных системах все три типа интегрированы, что создает предпосылки для формирования мультимодальных датасетов.

По масштабу охвата популяции базы данных существенно различаются. Одноцентровые базы, накопленные в рамках одной медицинской организации, отражают специфику конкретной популяции пациентов, конкретного оборудования и конкретных клинических практик этого учреждения. Многоцентровые и национальные регистры охватывают более широкую и разнородную популяцию, что повышает вероятность корректной работы обученной модели за пределами учреждения-разработчика. По режиму доступа базы данных делятся на закрытые институциональные, базы с ограниченным доступом, открытые для исследователей на условиях соглашения, и полностью открытые публичные, доступные после регистрации и, как правило, уже прошедшие деперсонализацию и базовую аннотацию.

Международные публичные базы данных

MIMIC (Medical Information Mart for Intensive Care) – совместный проект Массачусетского технологического института и ряда американских больниц, создавший общедоступную базу клинических данных пациентов отделений интенсивной терапии. База охватывает более 60 000 госпитализаций и включает демографические данные, лабораторные показатели, показатели витальных функций, сведения о медикаментозной терапии и текстовые документы. Версия MIMIC-CXR дополнена рентгенограммами органов грудной клетки с сопутствующими радиологическими заключениями. Методологические стандарты MIMIC, уровень деперсонализации, документирования и открытости, являются образцом для других баз данных вне зависимости от их предметной области.

CheXpert – база рентгенограмм органов грудной клетки, созданная в Стэнфордском университете и включающая свыше 224 000 медицинских изображений более чем 65 000 пациентов с аннотацией по 14 патологическим классам. Принципиальная особенность CheXpert, метод аннотации, метки извлекались автоматически из клинических заключений радиологов с помощью NLP-алгоритма, что обеспечило масштабируемость, но неизбежно внесло систематические ошибки. Неопределенные формулировки в заключениях радиологов алгоритм классифицирует непоследовательно, что создает шум в метках, и это важно учитывать при интерпретации результатов систем, обученных на этом датасете.

ISIC Archive формируется с 2016 года под эгидой Международного общества по визуализации кожи и представляет собой крупнейший открытый репозиторий дерматоскопических изображений. Архив пополняется участниками из клиник по всему миру и насчитывает несколько сотен тысяч изображений с нозологической аннотацией, причем для злокачественных новообразований, как правило, наличествует гистологическая верификация. Конкурсный формат ISIC Challenge создает редкую возможность прямого сравнения алгоритмов на единой закрытой

тестовой выборке, что делает данные ISIC наиболее надежным источником для объективной оценки дерматологических ПКЗ.

HAM10000 насчитывает около 10 015 изображений семи нозологических форм, сформированных из материалов нескольких европейских дерматологических клиник. Сочетание гистологически верифицированных и экспертно аннотированных изображений создает неоднородность стандарта меток, модели, обученные на HAM10000, фактически оптимизируются под смешанный стандарт верификации, что необходимо учитывать при сравнении показателей точности. Офтальмологические датасеты EyePACS, MESSIDOR, DRIVE, стали площадкой для исторически первой демонстрации диагностической точности ПКЗ, сопоставимой с врачебной, что дало импульс широкому развитию медицинских ПКЗ в целом. EyePACS включает свыше 88 000 изображений с аннотацией по пятибалльной шкале тяжести диабетической ретинопатии.

Российские клинические базы данных и регуляторный контекст

Отечественный ландшафт клинических баз данных формируется вокруг нескольких ключевых систем. Наиболее масштабной является ЕМИАС, Единая медицинская информационно-аналитическая система города Москвы, региональный фрагмент ЕГИСЗ. ЕМИАС аккумулирует данные о десятках миллионов обращений пациентов московской системы здравоохранения, охватывая весь спектр клинической информации. По масштабу и полноте эта система не имеет аналогов среди отечественных клинических информационных ресурсов и представляет исключительный потенциал для формирования датасетов ПКЗ.

Однако реализация этого потенциала сопряжена с серьезными нормативными ограничениями, принципиально отличающими российский контекст разработки ПКЗ от зарубежного. Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных» устанавливает требование локализации баз данных с персональными данными российских граждан на территории Российской

Федерации. Медицинские данные относятся к специальным категориям персональных данных, обработка которых требует явного согласия субъекта или специального правового основания.

Механизм экспериментальных правовых режимов, предусмотренный Федеральным законом от 31 июля 2020 г. № 258-ФЗ и реализуемый в здравоохранении в рамках московского эксперимента (Приказ Департамента здравоохранения города Москвы от 21.02.2025 г. № 153), создает ограниченные, но практически значимые возможности для апробации ПКЗ в реальных клинических условиях до завершения полной регуляторной процедуры. Эти возможности важны не только с практической, но и с методологической точки зрения: именно в рамках таких экспериментов накапливаются данные о реальной, а не лабораторной точности систем.

Систематические смещения как неустранимое свойство баз данных

Любая клиническая база данных несет в себе систематические смещения, происхождение которых определяется условиями ее формирования. Это не техническая неисправность, а принципиальное свойство данных, отражающих реальную клиническую практику конкретного учреждения в конкретный период времени. Географическое смещение проявляется в том, что большинство публичных датасетов сформированы в учреждениях Западной Европы и Северной Америки. Для дерматовенерологии наиболее критичен фототиповый сдвиг, у пациентов с IV-VI фототипами дерматоскопическая картина ряда нозологий принципиально отличается, и системы, обученные преимущественно на светлокожих пациентах, систематически хуже работают в этой группе.

Учрежденческое смещение возникает вследствие того, что каждое учреждение имеет собственный парк диагностического оборудования, собственные протоколы получения изображений и собственные диагностические практики. Модель, обученная на изображениях одного учреждения, нередко демонстрирует значительное снижение точности при тестировании на материале

другого, даже если нозологический спектр идентичен. Временное смещение, явление, которому уделяется значительно меньше внимания, чем оно заслуживает: смена поколений диагностического оборудования, изменение протоколов обработки изображений, сдвиги в демографическом составе пациентов постепенно меняют распределение входных данных. Этот феномен, известный как дрейф данных (data drift), может незаметно снижать точность системы со временем и требует непрерывного мониторинга производительности на актуальных данных.

Смещение верификации присуще любому датасету, где стандартом служит не морфологическая верификация, а экспертная консенсусная оценка. Эксперты ошибаются и расходятся в суждениях; степень этого расхождения, измеренная коэффициентом каппа Коэна, характеризует не только качество аннотации, но и принципиальную сложность задачи. Модель, обученная на метках с невысоким межэкспертным согласием, не может демонстрировать точность выше той, что достигается путем сравнения с золотым стандартом, это математическое ограничение, вытекающее из природы обучающего сигнала.

Сравнительная характеристика ключевых параметров баз данных

При выборе базы данных как источника для формирования датасета ПКЗ ключевую роль играют несколько параметров, требующих систематической оценки. Объем является необходимым, но не достаточным условием качества датасета, тысяча качественно верифицированных гистологически изображений предпочтительнее десяти тысяч изображений с экспертными метками средней согласованности. Демографическая репрезентативность приобретает значение при каждом переносе системы на новую популяцию и должна быть задокументирована в паспорте датасета с указанием распределения по полу, возрасту, этнической принадлежности и, в дерматологии, фототипу кожи. Отсутствие этой информации само по себе является методологическим дефектом.

Качество аннотации определяется тремя факторами: квалификацией аннотаторов, наличием независимой верификации несколькими специалистами и

доступностью морфологического «золотого стандарта». Правовой статус и условия лицензирования напрямую определяют допустимые сценарии использования: ряд публичных датасетов распространяется исключительно на условиях некоммерческого применения, что делает невозможным их использование при разработке коммерческих медицинских изделий без отдельного соглашения. Документирование в виде паспорта датасета (data card), фиксирующего источник данных, период сбора, критерии отбора, методологию аннотации, показатели межэкспертного согласия, методы деперсонализации и известные ограничения, является не формальностью, а инструментом, позволяющим любому независимому исследователю или регулятору оценить, насколько характеристики конкретного датасета соответствуют задаче разрабатываемой системы. Совокупность этих параметров – объема, репрезентативности, качества аннотации, правового статуса и документирования и определяет пригодность той или иной базы данных как фундамента для разработки клинически валидной и регуляторно приемлемой программы компьютерного зрения.

Сравнение баз данных представлено на Рисунке 37.

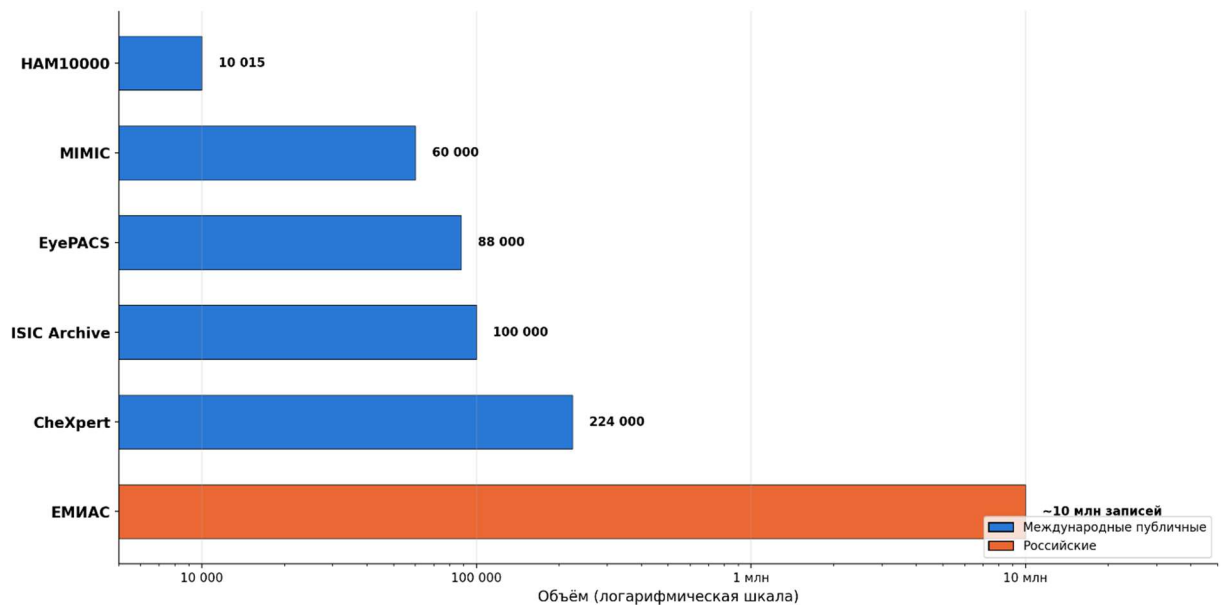


Рисунок 37 – Сравнительная визуализация баз данных

3.5. Разработка и исследование эффективности методики формирования из баз данных наборов данных для обучения моделей искусственного интеллекта для использования в клинической медицине

Стремительное развитие технологий машинного обучения в последние годы привело к тому, что системы искусственного интеллекта, ориентированные на анализ визуальных данных, заняли устойчивое место в арсенале клинической диагностики. Среди подобных систем программы компьютерного зрения представляют собой наиболее зрелый и широко апробированный класс, что подтверждается данными регуляторных реестров: по сведениям Управления по санитарному надзору за качеством пищевых продуктов и лекарственных средств США, ПКЗ составляют большинство среди одобренных медицинских изделий на основе ИИ. Столь значительная доля объясняется не случайным стечением обстоятельств, а рядом принципиальных особенностей медицинской визуализации как области применения: изображение как носитель диагностической информации поддается машинному анализу несравнимо проще, чем клинический нарратив или многопараметрические физиологические сигналы.

Несмотря на высокую практическую значимость ПКЗ и активное их внедрение в клиническую практику, методологические основы разработки этих систем остаются недостаточно стандартизированными. В особенности это касается двух взаимосвязанных аспектов: формирования обучающих наборов данных и систематической классификации самих программ. Отсутствие единых подходов порождает ситуацию, при которой сравнительная оценка различных систем лишается смысла: разработчики описывают свои датасеты в принципиально разных терминах, применяют несопоставимые методы балансировки и по-разному определяют границы тестовых выборок. Это ведет к тому, что декларируемые показатели диагностической точности нередко оказываются несравнимы между собой, а регуляторная оценка систем существенно затрудняется.

3.5.1. Разграничение понятий «база данных» и «набор данных»

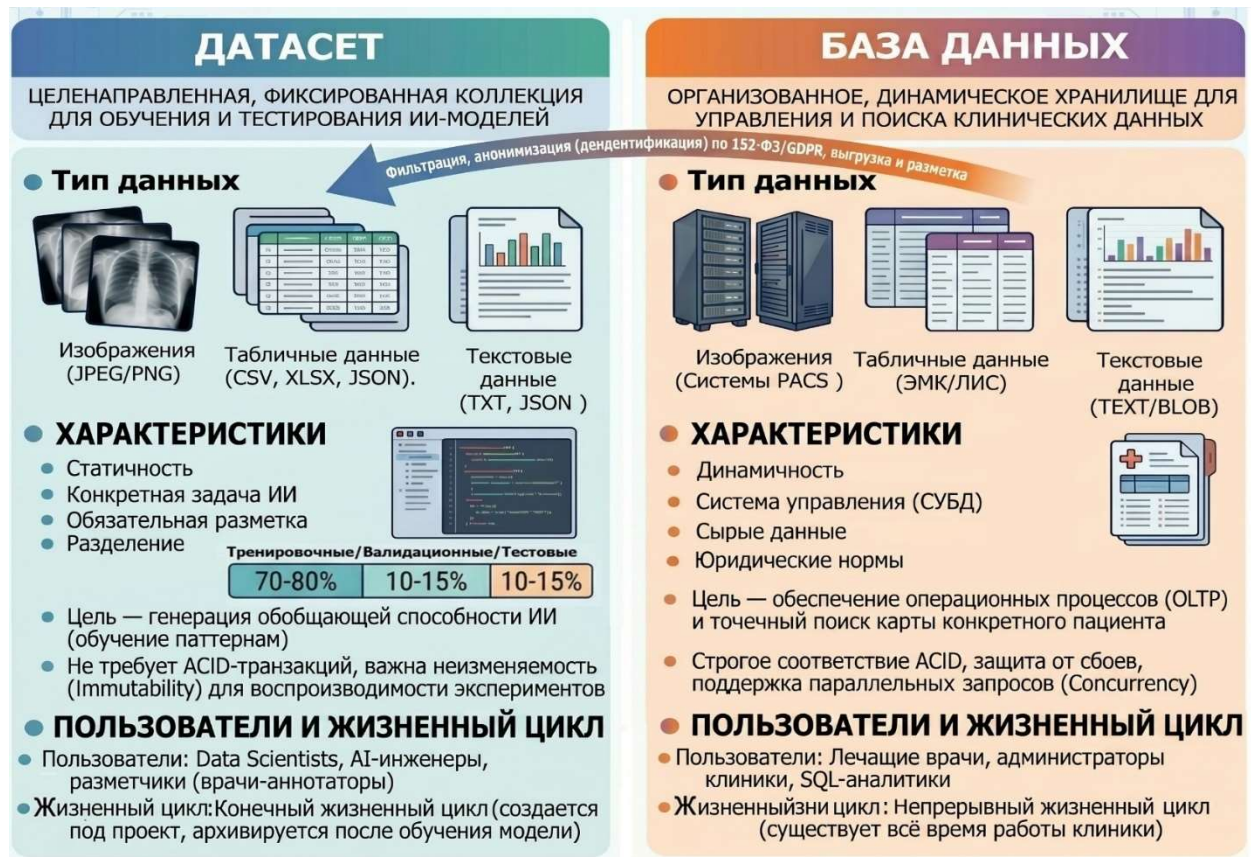
Одним из ключевых методологических вопросов при разработке ПКЗ является корректное разграничение понятий, которые в литературе нередко смешиваются. Речь идет о понятиях «база данных» и «набор данных» (датасет). Их внешнее сходство, оба представляют собой структурированные хранилища медицинской информации, создает иллюзию взаимозаменяемости, которая на практике порождает серьезные методологические ошибки.

База данных в медицинском контексте – это операционная информационная система, обеспечивающая текущую деятельность медицинской организации. Она непрерывно пополняется, содержит персонифицированные сведения о пациентах и организована так, чтобы обеспечивать быстрый многопользовательский доступ к данным в процессе лечебно-диагностической работы. Системы архивирования медицинских изображений (PACS), электронные медицинские карты (ЭМК), лабораторные информационные системы (ЛИС) все это примеры клинических баз данных. Содержащиеся в них данные представляют собой «сырой» материал: они несут идентификационные сведения о пациентах, нередко содержат дублирующие записи, технические артефакты и неполные сведения.

Набор данных, напротив, это целенаправленно сформированный, статичный и строго задокументированный продукт, созданный из базы данных для решения конкретной аналитической задачи. Его ключевое свойство, фиксированность состава: однажды сформированный датасет не изменяется, что обеспечивает воспроизводимость обучения и тестирования модели. Каждый образец в датасете снабжен верифицированной меткой, данные деперсонализированы, а распределение по классам контролируется.

Разница между этими понятиями хорошо иллюстрируется на конкретном примере. Дерматологический архив крупного стационара, хранящий сотни тысяч дерматоскопических фотоизображений с данными пациентов, это база данных. Отобранные из этого архива 5 000 верифицированных гистологически изображений семи нозологических форм, прошедших деперсонализацию и

экспертную аннотацию, разделенных на обучающую, валидационную и тестовую выборки и снабженных паспортом датасета, это набор данных, пригодный для обучения ПКЗ. Сравнительная характеристика обоих понятий по ключевым параметрам приведена на Рисунке 38 и в Таблице 21.



PACS – система архивирования и передачи медицинских изображений; DICOM – стандарт хранения и передачи медицинских изображений; ЭМК – электронная медицинская карта; ЛИС – лабораторная информационная система; ИИ – искусственный интеллект; HAM10000 – открытый дерматологический датасет; MIMIC-III – общедоступная база клинических данных пациентов ОИТ; i2b2 – конкурсные медицинские NLP-датасеты; NLP – обработка естественного языка; NER – распознавание именованных сущностей в тексте

Рисунок 38 – Визуализация характеристик наборов данных и баз данных

Таблица 21 – Сравнительная характеристика наборов данных и баз данных

Критерий	Набор данных (датасет)	База данных
Тип данных	Изображения, таблицы, тексты – любые, но отобранные и аннотированные	Изображения (PACS/DICOM), таблицы (ЭМК/ЛИС), тексты (эпикризы) – все типы в исходном виде

Продолжение Таблицы 21

Назначение	Обучение/валидация/тестирование модели ИИ	Хранение, управление и поиск клинических данных
Динамика	Статичен (фиксированный состав, версионирование)	Динамична (непрерывно пополняется)
Доступ	Загрузка архивом, API или стриминг батчей	SQL/NoSQL-запросы, транзакционный доступ
Качество данных	Предобработан, очищен, сбалансирован, деперсонализирован	«Сырые» данные с идентификаторами: дубликаты, ошибки, пропуски

3.5.2. Методология формирования наборов данных

Качество датасета определяет качество модели сильнее, чем архитектурные решения или объем вычислительных ресурсов. Это утверждение, широко принятое в сообществе разработчиков медицинских систем ИИ, имеет принципиальное практическое следствие: инвестиции в методологически выверенное формирование датасета окупаются многократно, тогда как ошибки, допущенные на этом этапе, не поддаются коррекции при последующем обучении. Процесс преобразования клинической базы данных в готовый датасет включает шесть последовательных этапов, каждый из которых несет самостоятельную методологическую нагрузку (Рисунок 39).



Рисунок 39 – Процесс формирования датасета из базы данных

Этап первый – определение задачи и формирование технического задания.

Прежде чем приступать к сбору данных, необходимо с клинической точностью сформулировать задачу, которую предстоит решать системе. Задача

классификации, детекции или сегментации предъявляют принципиально разные требования к разметке: классификация требует метки для всего изображения, детекция – ограничивающих рамок, сегментация – попиксельных масок. Тип задачи определяет и минимально необходимый объем датасета, и требования к квалификации аннотаторов, и выбор метрик оценки. Техническое задание на датасет должно закрепить все эти параметры до начала какой-либо работы с данными.

Этап второй – выгрузка и первичная фильтрация.

На основании критериев включения и исключения из базы данных извлекаются релевантные записи и изображения. Этот этап кажется техническим, однако именно здесь закладываются самые опасные смещения: неточно сформулированные критерии отбора систематически искажают состав выборки, и эта ошибка будет воспроизводиться в каждом пациенте, которому впоследствии система выдаст неверный результат.

Этап третий – деперсонализация.

Передача данных для разметки и последующее их использование для обучения модели допустимы исключительно после полного удаления персональных сведений. В российском правовом пространстве это требование закреплено Федеральным законом № 152-ФЗ «О персональных данных», технические аспекты деперсонализации данных в формате DICOM регламентированы ГОСТ Р 71674-2024. Деперсонализация должна затрагивать не только явные идентификаторы (ФИО, дата рождения), но и косвенные – редкие заболевания в сочетании с регионом и возрастом могут делать пациента де-факто идентифицируемым.

Этап четвертый – верификация качества и медицинская аннотация.

Аннотация – это назначение каждому образцу верифицированной клинической метки. Это наиболее трудоемкий и критически важный этап, поскольку метка определяет то, чему модель будет учиться. Три принципа лежат в основе методологически корректной аннотации. Первый: аннотация должна выполняться профильными специалистами с задокументированным опытом, а не

лаборантами или студентами. Второй: каждый образец должен получить независимые метки не менее чем от двух экспертов, а степень их согласия – быть формально измерена (коэффициент каппа Коэна или внутриклассовая корреляция). Третий: для новообразований кожи и других нозологических форм, допускающих морфологическую верификацию, гистологический диагноз является единственным допустимым «золотым стандартом» – экспертная визуальная оценка его не заменяет.

Этап пятый – организация структуры, балансировка и аугментация.

Дисбаланс классов, практически неизбежная черта медицинских датасетов, тяжелые заболевания встречаются реже, чем норма. Необработанная модель, обученная на несбалансированной выборке, воспроизведет это распределение в своих предсказаниях и будет систематически пропускать редкие, но клинически значимые случаи. Методы устранения дисбаланса включают аугментацию (геометрические преобразования, изменение яркости, добавление шума), синтетическую генерацию изображений с применением генеративно-состязательных сетей или диффузионных моделей, а также взвешивание функции потерь при обучении. Разбиение на обучающую, валидационную и тестовую выборки должно выполняться стратифицированно и строго на уровне пациентов: изображения одного пациента не могут одновременно присутствовать в обучающей и тестовой выборках, поскольку это создает утечку данных и завышает оцениваемую точность (Таблица 22).

Таблица 22 – Типы наборов данных и их предназначение

Тип	Назначение	Доля, %	Требования
Обучающая выборка	Обучение модели	70-80	Максимальный объем, сбалансированность
Валидационная выборка	Настройка гиперпараметров	10-15	Независимость от обучающей выборки
Тестовая выборка	Итоговая оценка	10-15	Полная независимость
Внешняя валидация	Внешняя проверка (другое учреждение)	Переменная	Иной источник данных

Этап шестой – документирование.

Паспорт датасета (data card) – итоговый документ, фиксирующий всю методологию от критериев отбора до показателей межэкспертного согласия. Его назначение двоякое: с одной стороны, он обеспечивает воспроизводимость исследования, с другой – является обязательным условием регуляторной оценки системы. Без надлежащей документации датасета невозможно оценить, насколько обученная на нем модель применима к конкретной популяции пациентов (Рисунок 40).



ТЗ – техническое задание; ФЗ № 152 – Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных»; HIPAA – Федеральный закон США о конфиденциальности медицинских данных 1996 г.; GDPR – Общий регламент по защите данных Евросоюза; GAN – генеративно-сопоставительная сеть; Train/Val/Test – обучающая, валидационная и тестовая выборки

Рисунок 40 – Методология формирования наборов данных

3.5.3. Валидация разработанной шестиэтапной методики формирования наборов данных

Для подтверждения эффективности предложенной шестиэтапной методики формирования наборов данных было проведено сравнение диагностической точности моделей ИИ, обученных на двух принципиально различных наборах данных: наборе, сформированном по разработанной шестиэтапной методике из клинической базы данных с гистологической верификацией диагнозов в качестве «золотого стандарта» (набор А); наборе, сформированном из публичных открытых источников с преобладанием экспертной визуальной аннотации без морфологического подтверждения и без стандартизированной процедуры контроля качества (набор Б).

Набор А был сформирован из клинической базы данных с последовательным прохождением всех шести этапов: определение задачи и формирование технического задания (бинарная классификация доброкачественных и злокачественных новообразований кожи), выгрузка и первичная фильтрация по критериям включения ($n = 25\ 000$ изображений после аугментации; исходный объем 2 831 оригинальных изображений), деперсонализация в соответствии с Федеральным законом от 27 июля 2006 года №152-ФЗ «О персональных данных», экспертная аннотация тремя врачами-дерматовенерологами со стажем работы более 20 лет с измерением межэкспертного согласия ($\kappa = 0,83$, существенное согласие по шкале Ландиса–Коха) и привязкой к гистологическому «золотому стандарту» для всех злокачественных новообразований, стратифицированное разбиение на обучающую (80%), валидационную (10%) и тестовую (10%) выборки на уровне пациентов с контролем утечки данных, и документирование в виде паспорта датасета.

Набор Б был составлен из публичных источников (HAM10000, ISIC Archive) аналогичного объема (10 015 изображений HAM10000) без применения стандартизированной процедуры формирования: использовалась смешанная верификация (часть изображений подтверждена гистологически, часть —

экспертной визуальной оценкой), разбиение на выборки выполнялось без контроля на уровне пациентов.

Обе модели обучались на идентичной архитектуре (Xception с переносом обучения от ImageNet) с одинаковыми гиперпараметрами и тестировались на единой независимой тестовой выборке из 381 клинического наблюдения (291 доброкачественное и 90 злокачественных новообразований), верифицированных гистологически (Таблица 23).

Таблица 23 – Результаты сравнения диагностической точности моделей ИИ, обученных на разных наборах данных (по разработанной шестиэтапной методике – модель А и контрольном – модель Б)

Показатель	Модель А (шестиэтапная методика)	Модель Б (контроль)
Чувствительность (Se), %	96,7 (94,5–98,1)	89,2 (82,4–93,5)
Специфичность (Sp), %	96,0 (93,3–97,6)	87,5 (83,2–90,8)
Точность (Ac), %	96,2 (93,8–97,7)	88,0 (84,2–91,0)
AUC-ROC	0,985	0,931
F1-мера	0,936	0,843
Cohen's Kappa (κ)	0,906	0,732
PPV, %	95,6	80,4
NPV, %	97,0	93,1
Оценка Бриера	0,044	0,102
Тест Макнемара (p)	–	$\chi^2 = 12,4; p < 0,001$
Примечание: обе модели тестировались на единой независимой выборке из 381 гистологически верифицированных наблюдений. Доверительные интервалы рассчитаны методом Уилсона. Тест Макнемара выполнен для попарного сравнения классификаторов на одних и тех же объектах		

Модель, обученная на наборе А, сформированном по шестиэтапной методике с гистологическим «золотым стандартом» верификации ЗНО кожи, статистически значительно превосходила модель, обученную на наборе Б, по всем ключевым диагностическим метрикам: чувствительность 96,7% против 89,2%, специфичность 96,0% против 87,5%, AUC-ROC 0,985 против 0,931, коэффициент согласия Коэна 0,906 против 0,732. Тест Макнемара подтвердил статистическую значимость различий ($\chi^2 = 12,4; p < 0,001$). Показатель калибровки был более чем в два раза ниже для модели А (0,044 против 0,102), что свидетельствует о более точном соответствии предсказанных вероятностей реальным исходам.

Преимущество модели А обусловлено несколькими факторами, непосредственно вытекающими из шестиступенчатой методики. Во-первых, привязка каждого злокачественного наблюдения к гистологическому «золотому стандарту» (этап 4) устраняет шум меток, неизбежный при экспертной визуальной аннотации, поскольку межэкспертное согласие даже среди опытных дерматовенерологов редко превышает $\kappa = 0,80\text{--}0,85$ для сложных пограничных случаев. Во-вторых, стратифицированное разбиение на уровне пациентов (этап 5) исключает утечку данных, при которой изображения одного пациента попадают одновременно в обучающую и тестовую выборки, искусственно завышая оценку точности на 5–12%. В-третьих, стандартизированная деперсонализация (этап 3) и документирование (этап 6) обеспечивают воспроизводимость обучения и независимой валидации.

Таким образом, результаты валидации подтверждают, что разработанная шестиступенчатая методика формирования наборов данных обеспечивает статистически значимое повышение диагностической точности моделей ИИ по сравнению с использованием наборов данных, сформированных без стандартизированной процедуры и без привязки к морфологическому «золотому стандарту». Обучение на изображениях, верифицированных гистологически, является необходимым условием достижения клинически приемлемого уровня точности программ компьютерного зрения в дерматоонкологии.

3.6. Разработка классификации программ для электронных вычислительных машин, применяемых в клинической медицине и здравоохранении

Все больше программ, основанных на технологиях искусственного интеллекта и программ компьютерного зрения наряду с классическими программами для ЭВМ применяются в различных областях медицины, от профилактики и диагностики до реабилитации и управления системой здравоохранения. Число таких программ растет очень быстро. Только среди мобильных приложений для дерматовенерологии из более чем девяти сот

найденных записей в анализ отбираются десятки самостоятельных программ. Эти программы существенно различаются между собой, они предназначены для разных пользователей, решают несхожие клинические задачи, работают с разными типами данных, построены на различных архитектурах и обучены на выборках разного объема и качества. При отсутствии единой системы описания сопоставить такие программы между собой, обоснованно выбрать программу для внедрения в конкретном звене здравоохранения и корректно оценить ее с регуляторных позиций затруднительно.

Отдельная сложность связана с тем, что при переходе от контролируемых условий тестирования к реальной клинической практике диагностическая эффективность программ для ЭВМ, включая основанных на технологиях ИИ и ПКЗ, нередко снижается, а причины этого снижения во многом определяются именно характеристиками данных и архитектуры моделей ИИ, которые в описаниях программ обычно не раскрываются. Поэтому классификация, охватывающая не только назначение программы, но и особенности ее реализации и обучения, приобретает актуальное практическое значение.

Нами предложена классификация из девятнадцати признаков, объединенных в пять тематических блоков (Рисунок 41).



Рисунок 41 – Обобщенная схема классификации программ для ЭВМ, включая основанных на технологиях искусственного интеллекта и программы компьютерного зрения, применяемых в клинической медицине и здравоохранении

Блок I. Контекст применения

Первый блок описывает кем, на каком организационном уровне, с какой целью и в рамках какого вида медицинской помощи используется программа. Эти признаки определяют требования к программе еще до обсуждения ее технической реализации (Таблица 24).

Таблица 24 – Классификация программ для ЭВМ, в том числе основанных на технологиях ИИ и ПКЗ, по признакам контекста применения

Классифицирующий признак	Выделяемые классы	Интерпретация
Конечный пользователь	пациент; медицинский работник с высшим образованием; медицинский работник со средним проф. образованием; немедицинский работник сферы здравоохранения	Определяет допустимую автономность программы и характер заключения
Организационный уровень применения	пациент; врач; медицинская организация; сеть медицинских организаций, региональный или федеральный уровень	Задаёт масштаб внедрения и требования к интеграции
Цель использования	профилактика; диагностика; лечение; реабилитация; организация, управление и экономика здравоохранения; обучение/проверка компетенций	Отражает клиническую или управленческую задачу программы
Область медицинского применения	соответствующая медицинская специальность	Определяет предметную область и профиль данных
По виду медицинской помощи	первичная медико-санитарная; специализированная, в т.ч. высокотехнологичная; скорая, в т.ч. скорая специализированная; паллиативная	Задаёт условия работы и требования к скорости и надёжности

Блок II. Развертывание и реализация

Второй блок характеризует то, как программа реализована технически и где выполняются вычисления (Таблица 25).

Таблица 25 – Классификация программ для ЭВМ, в том числе основанных на технологиях ИИ и ПКЗ, по признакам развертывания и реализации

Классификационный признак	Выделяемые классы	Интерпретация
По месту развертывания	локальная (обработка на устройстве пользователя); сетевая (обработка на удаленном сервере)	Влияет на конфиденциальность данных и требования к каналу связи
По интеграции в оборудование	ПО в виде локального приложения; ПО в виде веб-приложения; ПО встроенное в медицинское изделие	Определяет способ поставки и порядок регистрации
По виду	классическая программа; классическая ПКЗ; программа ИИ; программа ПКЗ на основании технологий ИИ; комбинированная программа	Отражает какие технологии программа использует

Блок III. Данные и оборудование

Третий блок описывает входные данные, с которыми работает программа и оборудование с помощью которого эти данные получают (Таблица 26).

Таблица 26 – Классификация программ для ЭВМ, в том числе основанных на технологиях ИИ и ПКЗ, по признакам используемых данных и оборудования

Классификационный признак	Выделяемые классы	Интерпретация
По типу используемых данных	текстовые данные; числовые данные; изображения; изображения в определенных спектрах; видеопоток; трехмерные изображения	Возрастание уровня сложности данных от изображений к видео и 3D
По модальности входных данных	одномодальные; мультимодальные	Определяет полноту клинического контекста доступного программе
По использованному оборудованию	фотокамеры; видеокамеры; датчики (движения, температурные и др.)	Задаёт источник и характер входных данных

Блок IV. Архитектура моделей искусственного интеллекта и обработка входных данных

Четвертый блок описывает внутреннее устройство программы и способ обработки изображения (Таблица 27, Рисунок 42-43).

Таблица 27 – Классификация программ для ЭВМ, основанных на технологиях ИИ, по признакам архитектуры и обработки

Классификационный признак	Выделяемые классы	Интерпретация
По числу моделей ИИ	одномодельные; мультимодельные	Мультимодельность повышает устойчивость заключения
По числу классов	бинарная классификация; многоклассовая классификация	Определяет детализацию выходного заключения
По методу обработки изображений	сегментация (обнаружение сегментов); детекция (обнаружение объектов); классификация (отнесение к классу)	Задаёт характер решаемой задачи ПКЗ

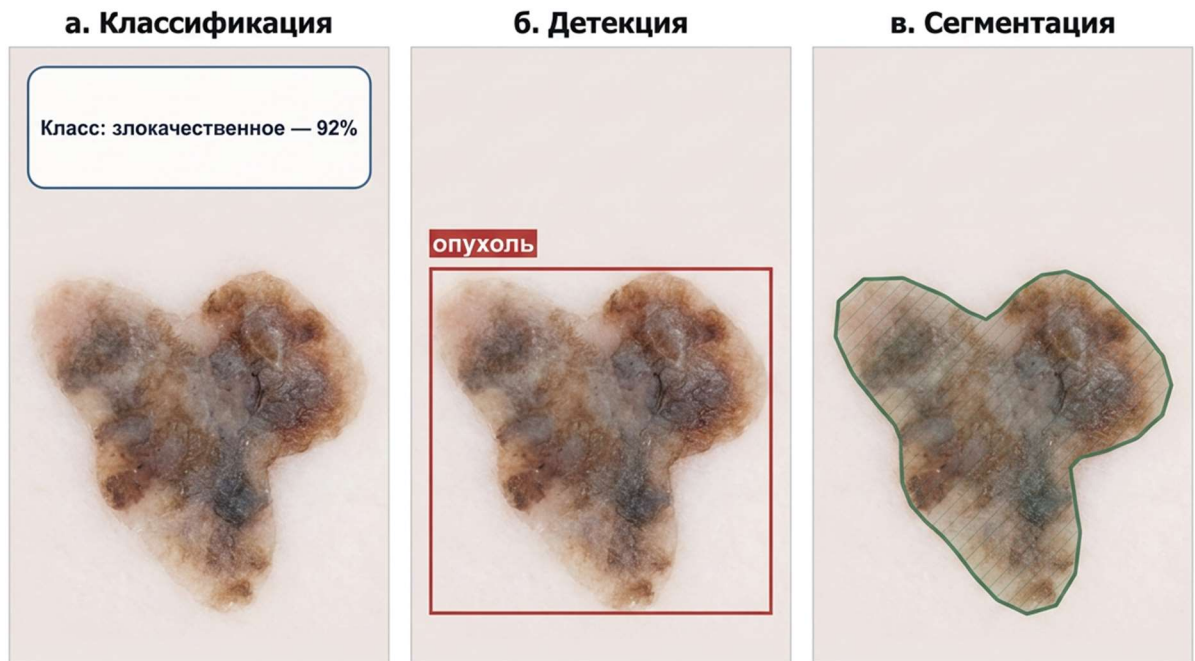


Рисунок 42 – Три метода обработки изображений в программах компьютерного зрения на примере новообразования кожи: а – классификация; б – детекция; в – сегментация

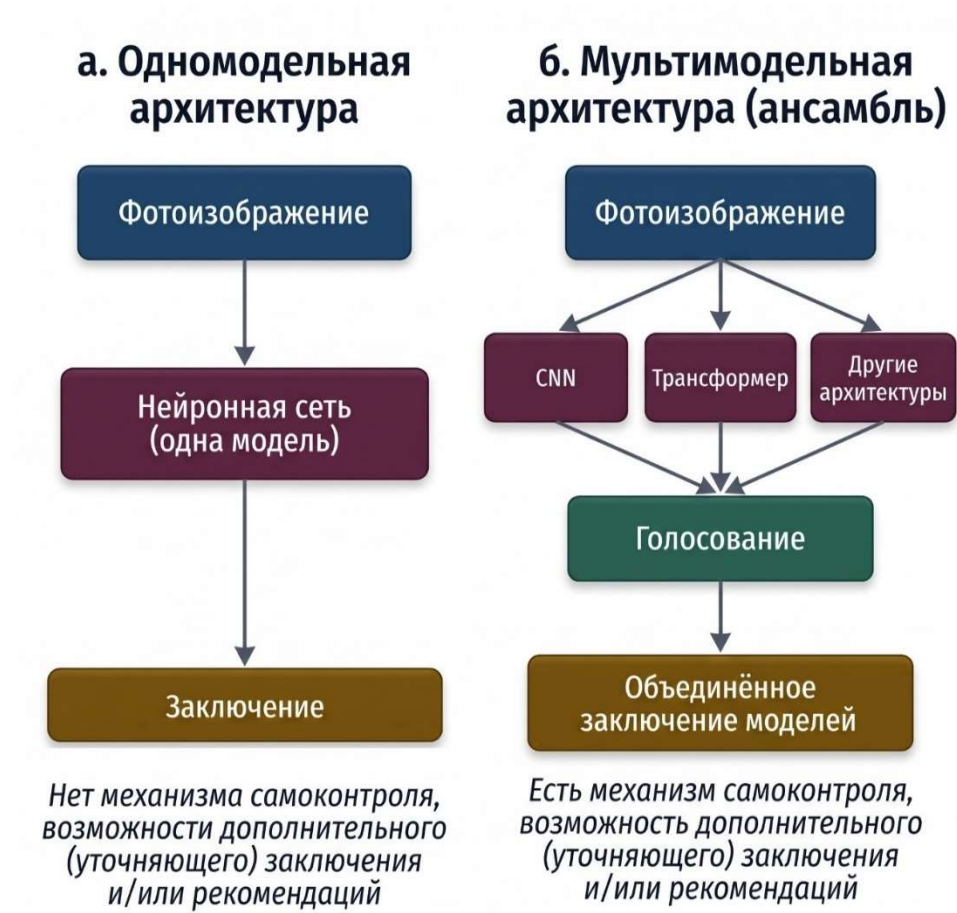


Рисунок 43 – Сопоставление одномодельной и мультимодельной архитектуры: а – одномодельная программа, б – мультимодельная программа

Блок V. Характеристики обучения моделей искусственного интеллекта

Пятый блок описывает то, как модель ИИ входящая в состав программы была обучена. Эти признаки во многом определяют ее реальную диагностическую эффективность, хотя в описаниях программ раскрываются реже всего (Таблица 28-29, Рисунок 44-45).

Таблица 28 – Номинальные характеристики обучения моделей для программ для ЭВМ, основанных на технологиях ИИ

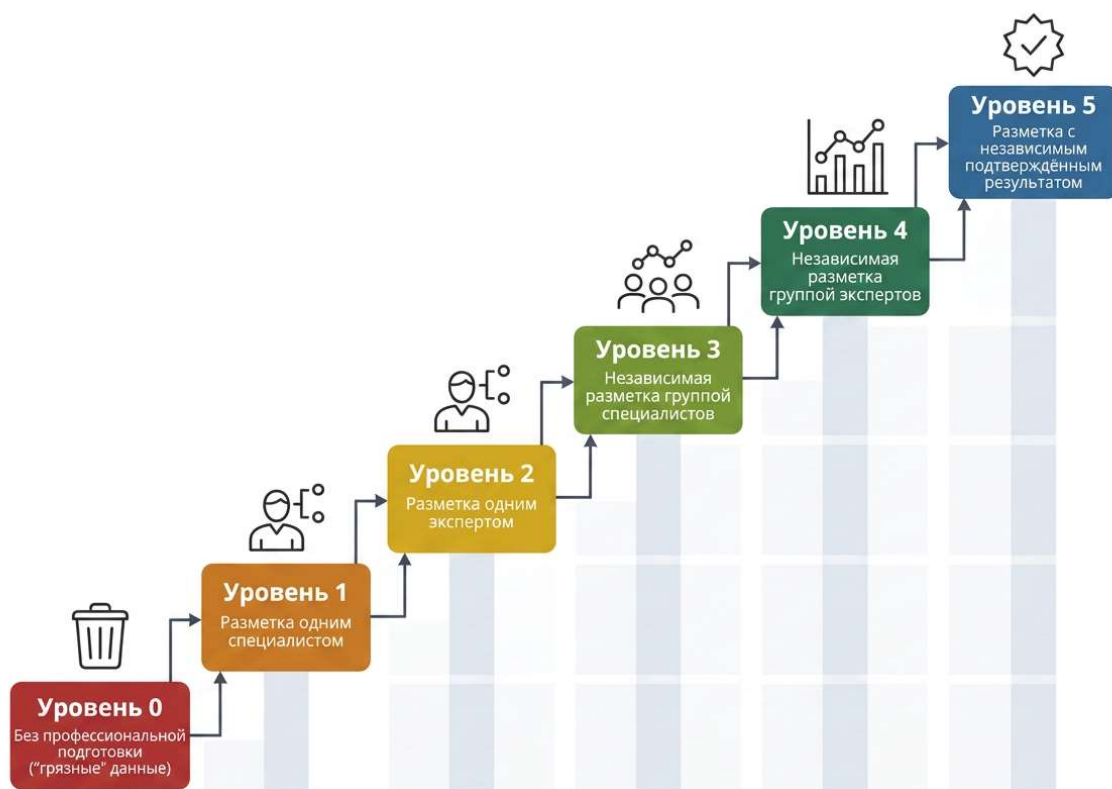
Классификационный признак	Выделяемые классы	Характеристика
По типу обучения	классическое обучение; глубокое обучение	Классическое обучение опирается на признаки заданные разработчиком; при глубоком обучении признаки извлекаются автоматически многослойными нейронными сетями
По виду обучения	обучение с учителем (данные размечены); обучение без учителя; обучение с подкреплением (обучение на обратной связи)	Различаются наличием и характером целевого сигнала при обучении

Таблица 29 – Градации характеристик обучающих данных с критериями отнесения к уровню

Уровень	Класс	Критерий отнесения к уровню
Объем обучающей выборки (уровни 1–5)		
1	до 100 объектов	Минимальная выборка; высокий риск переобучения
2	до 1 000 объектов	Ограниченная выборка; покрываются наиболее частые варианты
3	до 10 000 объектов	Средняя выборка; приемлемое покрытие для типовых задач
4	до 100 000 объектов	Крупная выборка; хорошее покрытие изменчивости
5	свыше 100 000 объектов	Очень крупная выборка; широкое покрытие
Техническая подготовка данных (уровни 0–2)		
0	без технической подготовки	Выборка используется без изменений
1	копирование изображений	Дублирование изображений; новой информации не добавляется
2	аугментация изображений	Управляемые преобразования (повороты, отражения, изменение яркости и др.)

Продолжение Таблицы 29

Профессиональная подготовка данных (уровни 0–5)		
0	без подготовки	«Грязные» данные; разметка и контроль отсутствуют
1	разметка одним специалистом	Базовая разметка; сохраняется риск субъективных ошибок
2	разметка одним экспертом	Более высокая квалификация разметки без взаимного контроля
3	независимая разметка группой специалистов	Согласование нескольких независимых разметчиков
4	независимая разметка группой экспертов	Высокая квалификация с взаимным контролем
5	разметка с независимым подтвержденным результатом	Наличие внешнего референса (напр. гистологическая верификация)



Рост достоверности разметки и снижение доли «грязных» данных

Рисунок 44 – Шкала профессиональной подготовки обучающих данных (уровни 0–5)



Рисунок 45 – Порядковые шкалы градации характеристик обучающих данных: объём обучающей выборки (уровни 1–5), техническая подготовка данных (уровни 0–2) и профессиональная подготовка данных (уровни 0–5)

Предложенная классификация дает единый язык описания программ для ЭВМ, в том числе основанных на технологиях ИИ и ПКЗ в клинической медицине и здравоохранении, облегчает их сопоставление и выбор, а также может служить основой для методических и регуляторных требований к подобным программам.

Для наглядности характеристики вышеописанных программ по 19 классифицирующим признакам сведены в единую таблицу (Таблица 30). Такое представление демонстрирует практическую пользу классификации: программы, внешне решающие одну и ту же задачу, различаются по ряду признаков: по архитектуре и обучению моделей, и именно эти различия определяют характер и достоверность их заключений.

Таблица 30 – Сводная характеристика программ по 19 классифицирующим признакам

Классифицирующий признак	Derma Onko Check 2.0	Melanoma Check 2.0	Derma Onko Melanoma Check	Derma Check
1. Конечный пользователь	Медицинский работник с высшим медицинским образованием, со средним профессиональным образованием, ведущий самостоятельный прием (фельдшер) или уполномоченное на это лицо	Медицинский работник с высшим медицинским образованием, фельдшер, ведущий самостоятельный прием или уполномоченное на это лицо	Медицинский работник с высшим медицинским образованием, фельдшер, ведущий самостоятельный прием или уполномоченное на это лицо	Медицинский работник с высшим медицинским образованием
2. Организационный уровень	Врач	Врач	Врач	Врач, Медицинская организация
3. Цель использования	Диагностика	Диагностика	Диагностика	Диагностика
4. Основная область применения	Дерматовенерология	Дерматовенерология	Дерматовенерология	Дерматовенерология
5. Вид медицинской помощи	Первичная доврачебная, врачебная медико-санитарная медицинская помощь	Первичная доврачебная, врачебная медико-санитарная медицинская помощь	Первичная доврачебная, врачебная медико-санитарная медицинская помощь	Первичная доврачебная, врачебная, специализированная медико-санитарная медицинская помощь
6. Место развертывания	Локальная	Локальная	Локальная	Сетевая (сервер)
7. Интеграция в оборудование	ПО в виде локального приложения	ПО в виде локального приложения	ПО в виде локального приложения	ПО в виде веб-приложения
8. Вид программы	Программа ПКЗ на основании технологий ИИ	Программа ПКЗ на основании технологий ИИ	Программа ПКЗ на основании технологий ИИ	Комбинированная программа

Продолжение Таблицы 30

9. Тип данных	Изображения	Изображения	Изображения	Изображения; структурированный текстовые данные – числовые данные
10. Модальность	Одномодальная	Одномодальная	Одномодальная	Мультимодальная
11. Оборудование	Фотокамера	Фотокамера	Фотокамера	Фотокамера
12. Число моделей ИИ	Одномодельная	Одномодельная	Мультимодельная	Мультимодельная
13. Число классов	Бинарная	Бинарная	Бинарная	Бинарная, Многоклассовая
14. Метод обработки	Классификация	Классификация	Классификация	Классификация, сегментация, детекция
15. Тип обучения	Глубокое	Глубокое	Глубокое	Глубокое; машинное обучение
16. Вид обучения	С учителем	С учителем	С учителем	С учителем
17. Объем выборки (уровни 1–5)	До 10 000 объектов	До 10 000 объектов	До 100 000 объектов	Свыше 100 000 объектов
18. Техническая подготовка (уровни 0–2)	Аугментация изображений	Аугментация изображений	Аугментация изображений	Аугментация изображений
19. Профессиональная подготовка (уровни 0–5)	Независимая разметка группой специалистов	Независимая разметка группой специалистов	Независимая разметка группой специалистов	Разметка с независимым подтвержденным результатом

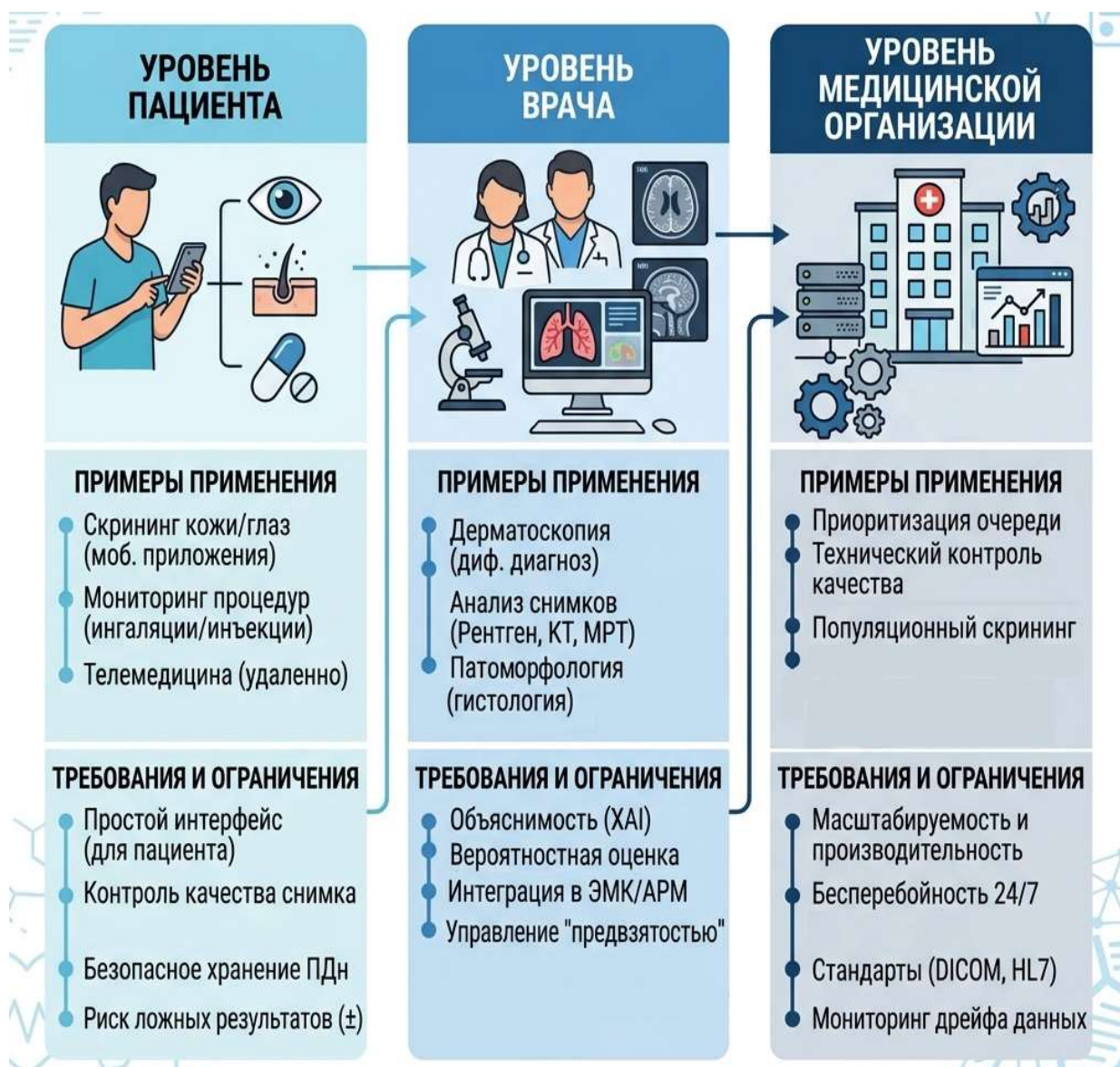
Уровни клинического применения программ компьютерного зрения в клинической медицине

Применение ПКЗ разворачивается на трех организационных уровнях, каждый из которых предъявляет принципиально разные требования к системе. На уровне пациента ПКЗ функционирует как инструмент первичной самооценки, реализуемый через мобильное приложение. Пациент фотографирует новообразование смартфоном и получает предварительный результат, определяющий необходимость обращения к врачу. Ключевые требования к таким системам, простота интерфейса, контроль качества входного изображения с блокировкой анализа при недостаточном освещении или нерезкости, а также корректная формулировка результата, исключающая как избыточную тревогу при ложноположительных, так и ложное успокоение при ложноотрицательных результатах.

На уровне врача ПКЗ выступает инструментом второго мнения, алгоритмического заключения, которое специалист рассматривает параллельно или после формирования собственного мнения, при этом ответственность за диагностическое решение остается за врачом безусловно. Принципиальными требованиями здесь являются объяснимость заключения, интеграция в рабочее пространство без избыточной нагрузки на рабочий процесс, а также специальные организационные меры против автоматизационного доверия, когнитивного смещения, при котором специалист начинает некритически воспринимать заключение системы, постепенно утрачивая навык самостоятельной оценки.

На уровне медицинской организации ПКЗ интегрируется в информационную инфраструктуру учреждения и решает задачи управления диагностическим потоком: приоритизации случаев, мониторинга качества поступающих изображений, эпидемиологического надзора на уровне популяции пациентов. Требования к системам этого уровня, масштабируемость, отказоустойчивость в режиме непрерывной работы, соответствие стандартам медицинской информатики

(DICOM, HL7 FHIR) и механизмы мониторинга деградации точности при изменении характеристик входящего потока данных (Рисунок 46).



КТ – компьютерная томография; МРТ – магнитно-резонансная томография; ЭМК – электронная медицинская карта; АРМ – автоматизированное рабочее место врача; ПДн – персональные данные; XAI – объяснимый искусственный интеллект; DICOM – международный стандарт хранения и передачи медицинских изображений; HL7 – международный стандарт обмена электронной медицинской информацией

Рисунок 46 – Организационные уровни применения программ компьютерного зрения

3.7. Исследование проблем диагностики новообразований кожи с применением программ искусственного интеллекта, разработка и апробация решений проблем диагностики

Изучение проблем, возникающих при дифференциальной диагностике ЗНО кожи, и поиск путей совершенствования их разграничения с доброкачественными новообразованиями (ДНО) кожи представляется актуальной научно-практической задачей. Для того чтобы предметно рассуждать о путях преодоления этих проблем, их необходимо предварительно выявить, описать и упорядочить. Именно поэтому в основу настоящего раздела положена систематизация основных, наиболее распространенных препятствий на пути своевременной диагностики ЗНО и ДНО кожи.

3.7.1. Систематизация основных проблем

С целью выявления и упорядочения указанных проблем нами был проведен анализ научной литературы и собственного клинического опыта. Анализировались литературные источники, включенные в базы данных PubMed/MEDLINE, Scopus, Google Scholar, OpenAlex, eLibrary, КиберЛенинка, а также хранящиеся в фондах Российской государственной библиотеки и Государственной центральной научной медицинской библиотеки. Поиск полнотекстовых публикаций проводился по ключевым словам и словосочетаниям: в русскоязычном сегменте использовались выражения «трудности диагностики», «сложности диагностики», «проблемы диагностики» в сочетании с формулировками «новообразования кожи», «опухоли кожи», «рак кожи»; в англоязычном сегменте применялись выражения diagnostic difficulties, diagnostic problems, diagnostic complications в сочетании с формулировками skin neoplasms, skin tumors, skin cancer. Наряду с литературными источниками использовались и собственные наблюдения.

Глубина поиска научной литературы составила 10 лет и охватывала период с 2017 по 2026 г. По итогам первичного поискового запроса было идентифицировано

545 публикаций. После удаления дублирующихся источников, скрининга заголовков и аннотаций, а также оценки полных текстов на соответствие критериям включения для итогового анализа было отобрано 32 литературных источника, из которых 22 относились к русскоязычному, а 10 к англоязычному сегменту. Критериями включения служили полнотекстовая доступность публикации, ее соответствие теме дифференциальной диагностики доброкачественных и злокачественных новообразований кожи, а также научная значимость и актуальность источника. Критериями исключения являлись дублирование данных, несоответствие тематике исследования и отсутствие доступа к полному тексту.

По результатам проведенного анализа основные проблемы своевременной диагностики ЗНО и ДНО кожи были нами впервые систематизированы и разделены на две большие группы трудности и сложности диагностики. Такое разделение носит не формальный, а содержательный характер, поскольку оно позволяет отделить факторы, корнящиеся в самой природе новообразований и методов их исследования, от факторов, связанных с человеком и организацией медицинской помощи.

Трудностями дифференциальной диагностики мы называем объективные факторы, которые затрудняют своевременное разграничение доброкачественных и злокачественных новообразований кожи. Эти факторы существуют независимо от воли и квалификации конкретного врача: они обусловлены разнообразием клинических и морфологических проявлений опухолей, ограничениями применяемых методов исследования и особенностями поведения пациентов.

Сложностями дифференциальной диагностики мы считаем субъективные факторы, которые усложняют своевременное разграничение доброкачественных и злокачественных новообразований кожи. В отличие от трудностей, сложности связаны прежде всего с человеческим и организационным звеном: с квалификацией врачей, их онкологической настороженностью, а также с тем, как выстроена система диагностики и маршрутизации пациентов. Предложенная классификация проблем дифференциальной диагностики ЗНО и ДНО кожи представлена в Таблице 31.

Таблица 31 – Проблемы дифференциальной диагностики ЗНО и ДНО кожи

Проблемы дифференциальной диагностики ЗНО и ДНО кожи	Категория факторов	Описание
1. Трудности дифференциальной диагностики	Объективные факторы	1) трудности визуальной диагностики новообразований кожи: атипичные клинические признаки; ограничения методов диагностики (например, дерматоскопии); 2) трудности морфологической диагностики новообразований кожи: атипичные морфологические признаки; 3) субъективность методов клинического исследования новообразований кожи; 4) недостаточное санитарно-гигиеническое просвещение населения и поздняя обращаемость пациентов с новообразованиями кожи
2. Сложности дифференциальной диагностики	Субъективные факторы	1) большое разнообразие нозологических форм новообразований кожи; 2) нарушения организации диагностики новообразований кожи: дефицит подготовленных врачей-специалистов, способных проводить дифференциальную диагностику между доброкачественными и злокачественными новообразованиями кожи; недостаточная квалификация врачей, занимающихся диагностикой новообразований кожи (врачи первичного звена здравоохранения: терапевты, врачи общей практики; врачи-дерматовенерологи; онкологи; патологоанатомы); недостаточная онкологическая настороженность, недооценка или игнорирование анамнеза и его недостаточный сбор; субъективность оценки клинических симптомов и результатов диагностики (игнорирование симптома «гадкого утенка», признаков ABCDE); неправильный выбор участка для биопсии; нарушения и повреждения материала при выполнении биопсии; неправильный выбор метода исследования и (или) неправильное выполнение иммуногистохимического либо молекулярно-генетического исследования; отсутствие или нарушения маршрутизации выявленных пациентов

Следует оговориться, что предложенное деление на объективные и субъективные факторы носит в известной степени условный характер. При классификации некоторые факторы могут быть с равным основанием отнесены к

обеим категориям, поскольку они одновременно отражают и объективные ограничения методов, и субъективные особенности их применения. В подобных случаях мы рассматривали эти факторы с позиции организации здравоохранения и общественного здоровья, учитывая их вклад в проблему диагностики в целом. Ниже каждая из выделенных групп рассмотрена подробно.

3.7.2. Трудности дифференциальной диагностики доброкачественных и злокачественных новообразований кожи

Трудность визуальной диагностики опухолей кожи

Новообразования кожи представляют собой сложный объект исследования. Обусловлено это несколькими причинами. Во-первых, на кожном покрове могут встречаться самые разные по своей природе опухоли – доброкачественные, предраковые и злокачественные. Во-вторых, одна и та же нозологическая форма опухоли способна иметь различные клинические и гистологические особенности, что размывает границы ее типичного описания. В-третьих, злокачественные новообразования кожи нередко имитируют доброкачественные опухоли, а доброкачественные, напротив, могут внешне напоминать злокачественные. Все это в совокупности существенно осложняет визуальную дифференциальную диагностику.

Наиболее традиционным и распространенным методом диагностики опухолей кожи остается визуальный осмотр невооруженным глазом. При таком осмотре врач оценивает поражение кожи по его размеру, форме, цвету, характеру поверхности, границам, а также по динамике происходящих в нем изменений. Однако этот метод имеет ряд ожидаемых недостатков, особенно если рассматривать его в качестве инструмента скрининга ЗНО кожи. Прежде всего, он требует наличия большого числа обученного персонала, который вынужден осматривать множество пациентов, имеющих в подавляющем большинстве доброкачественные опухоли, а не злокачественные.

Даже применение современных неинвазивных методов, в частности дерматоскопии, не устраняет всех трудностей визуальной диагностики. Дерматоскопия позволяет детально изучить структуру опухоли кожи и выявить признаки, невидимые невооруженным глазом.

Иными словами, инструмент, призванный повысить объективность диагностики, сам по себе не гарантирует единообразия заключений, если им пользуются специалисты с разным уровнем подготовки.

Необходимо также учитывать, что опухоли, развивающиеся из различных тканей кожи, образуют разнородную группу с очень широким спектром клинических проявлений. Внешне они могут выглядеть совершенно по-разному: от плоских шелушащихся пятен до возвышающихся узлов, от изъязвленных поверхностей до абсолютно гладких, от резко пигментированных образований до полностью беспигментных. Из-за такого многообразия врачу-специалисту подчас непросто определить, с каким именно новообразованием он имеет дело. Разные специалисты могут по-разному интерпретировать одни и те же клинические признаки, что повышает риск диагностической ошибки.

Трудности визуальной диагностики меланоцитарных опухолей кожи

При визуальном осмотре меланоцитарных новообразований важно учитывать целый ряд факторов, среди которых фототип кожи пациента, общее количество невусов, наличие диспластических и врожденных невусов, а также кожные признаки фотоповреждения. Каждый из этих факторов по-своему влияет на вероятность развития меланомы и на сложность ее распознавания.

Источником меланомы могут становиться и врожденные меланоцитарные невусы. При этом риск малигнизации существенно зависит от размера невуса. У пациентов с мелкими (до 1,5 см) и средними (до 20 см) врожденными невусами до наступления половой зрелости риск развития меланомы практически отсутствует, а на протяжении дальнейшей жизни он составляет 1–6 %.

Особые трудности возникают при диагностике беспигментных меланом кожи, которые могут быть очень похожи на доброкачественные опухоли, что затрудняет выявление этих злокачественных меланоцитарных новообразований на ранних стадиях. Разница в диагностической точности между специалистами разного профиля наглядно демонстрирует, насколько значимым остается человеческий фактор в визуальной диагностике меланомы.

Трудности визуальной диагностики кератиноцитарных / эпидермальных опухолей кожи

При визуальном осмотре кератиноцитарных/эпидермальных опухолей кожи важно учитывать фототип кожи пациента, а также наличие очагов актинического кератоза, требующих дифференциальной диагностики с раками кожи, и очагов себорейного кератоза, требующих дифференциальной диагностики с доброкачественными и злокачественными пигментными новообразованиями кожи. Уже сам перечень образований, с которыми приходится проводить разграничение, указывает на потенциальную сложность визуальной оценки в этой группе опухолей.

Отдельную проблему представляет диагностика себорейного кератоза, который может иметь клиническое сходство с пигментными новообразованиями меланоцитарными невусами и меланомой. Основное отличие себорейного кератоза от невусов состоит в том, что невусы возникают в раннем детстве и юности, тогда как себорейный кератоз появляется в зрелом или пожилом возрасте. Кроме того, при себорейном кератозе на поверхности бляшки под лупой обычно видны роговые кисты. Однако дифференциальная диагностика существенно осложняется в тех случаях, когда имеется единичный очаг с оттенками черного цвета и без роговых кист. При гистологическом исследовании в очагах себорейного кератоза меланин, как правило, обнаруживается только в кератиноцитах, а меланоциты отсутствуют, но при меланоакантоме, представляющей собой вариант себорейного кератоза, могут выявляться и меланоциты. Дополнительную сложность создает узловая

меланомы, которая из-за темной или черной окраски внешне напоминает себорейный кератоз. Меланома может также развиваться на фоне уже существующего бородавчатого пигментного невуса, поверхность которого испещрена трещинами, что придает ему сходство с себорейным кератозом.

Трудности визуальной диагностики опухолей кожи из фиброзной ткани

Большинство фиброзных опухолей развивается в толще дермы или гиподермы, вследствие чего бывает трудно определить их истинные границы, структуру и степень инвазии. В практической деятельности врачи-дерматовенерологи чаще всего встречаются с доброкачественными опухолями этой группы дерматофибромами. Их дифференцируют по клиническим признакам с приобретенными меланоцитарными невусами.

Трудности визуальной диагностики опухолей кожи из сосудистой ткани

Сосудистые поражения кожи в большинстве случаев имеют типичную клиническую картину, поэтому при постановке диагноза серьезных трудностей обычно не возникает. Исключение составляет пиогенная гранулема, которая требует дифференциальной диагностики с узловой меланомой и саркомой Капоши. И пиогенная гранулема, и саркома Капоши могут быть представлены единичным узлом красного или красно-коричневого цвета диаметром до 2 см и иметь сходную локализацию. Тем не менее для саркомы Капоши существование единственного очага является крайней редкостью, и для нее не характерна повышенная кровоточивость при малейших травмах. Пиогенную гранулему, в свою очередь, отличает очень быстрый рост. Знание этих деталей позволяет врачу правильно выстроить дифференциально-диагностический ряд даже в тех случаях, когда клиническая картина на первый взгляд неспецифична.

Трудности морфологической диагностики новообразований кожи

К основным причинам трудностей при морфологической верификации опухолей кожи относятся неправильный выбор участка для биопсии, нарушения и повреждения материала при его заборе, атипичные клинические и морфологические признаки, а также субъективность оценки полученных препаратов. Каждая из этих причин способна исказить итоговое заключение, даже если сам метод исследования выбран правильно.

«Золотым стандартом» диагностики опухолей кожи по праву считается морфологическое исследование, включающее гистологический, цитологический и иммуногистохимический методы. Однако его проведение требует соблюдения ряда обязательных условий: правильного забора материала, качественной вырезки, качественного изготовления препарата и, что особенно важно, наличия у патоморфолога достаточного опыта в интерпретации морфологических признаков доброкачественных и злокачественных опухолей кожи. Нарушение любого из этих условий снижает достоверность заключения.

При патоморфологической диагностике меланоцитарных новообразований объективные сложности могут усугубляться отсутствием критически важной клинико-инструментальной информации, а также артефактами, возникающими на этапе эксцизии или биопсии, на этапе макроскопического исследования и вырезки, и при ориентации фрагментов во время заливки в парафиновый блок. Иными словами, качество морфологической диагностики закладывается задолго до того, как препарат окажется под микроскопом.

Следует также помнить, что некоторые новообразования кожи на ранних стадиях своего развития имеют неспецифические клинические и морфологические признаки. Например, базальноклеточный рак на начальном этапе способен имитировать воспалительный процесс в коже, хроническую язву или проявлять признаки доброкачественной опухоли. В подобных ситуациях однократное исследование не всегда позволяет установить окончательный диагноз, и требуется динамическое наблюдение либо повторная биопсия.

Субъективность методов клинического исследования новообразований кожи

При диагностике новообразований кожи, и прежде всего меланомы, нередко возникают ситуации, когда итоговое заключение существенно зависит от того, кто именно интерпретирует клинические данные. Это приводит к двум противоположным по своему характеру ошибкам – гиподиагностике и гипердиагностике ЗНО кожи. В основе обеих проблем лежит одна и та же причина – субъективность оценки результатов клинического исследования различными врачами.

Гипердиагностике меланом способствуют следующие факторы: недостаточные квалификация и опыт врача; неадекватная клиническая оценка доброкачественных меланоцитарных поражений кожи; неадекватная клиническая оценка изменений, происходящих в доброкачественных меланоцитарных образованиях (с возрастом, на фоне гормональных изменений, под влиянием ультрафиолетового облучения и прочих причин); неадекватная морфологическая оценка пигментных поражений кожи. В частности, ПУВА-терапия (фотохимиотерапия) способствует активации меланоцитов в доброкачественных меланоцитарных опухолях и может приводить к патоморфологическим изменениям, имитирующим меланому.

Гиподиагностике меланом, напротив, способствуют недостаточные квалификация и опыт врача; неверный выбор области опухоли для биопсии; погрешности при заборе материала для гистологического исследования. Биопсийная ткань может оказаться раздавленной или обугленной, что нередко наблюдается при удалении опухоли с помощью лазерных технологий или иных методов физической деструкции; в таких случаях возникают артефакты, способные скрыть морфологические признаки меланомы. Кроме того, выраженные воспалительные явления в меланоцитарной опухоли могут искажать цитологическую и гистологическую картину меланомы. Таким образом, и гипо-, и гипердиагностика являются оборотными сторонами одной и той же проблемы недостаточной объективности клинической оценки.

Недостаточное санитарно-гигиеническое просвещение населения и поздняя обращаемость пациентов с новообразованиями кожи

Недостаточное санитарно-гигиеническое просвещение населения и связанная с ним поздняя обращаемость пациентов представляют собой значительное препятствие на пути к ранней диагностике и успешному лечению опухолей кожи, в первую очередь меланомы, базальноклеточного и плоскоклеточного раков кожи. Эти факторы действуют еще до того, как пациент вообще попадает в поле зрения врача, и потому нередко остаются вне зоны внимания при обсуждении сугубо медицинских аспектов диагностики.

Недостаточное санитарно-гигиеническое просвещение выражается в том, что многие люди имеют низкую осведомленность о факторах риска и особенностях образа жизни, способствующих развитию ЗНО кожи. Такие люди не знают, какие изменения на коже следует воспринимать как тревожный сигнал, не осведомлены о факторах риска (например, о чрезмерном пребывании на солнце, наличии большого числа родинок, семейном анамнезе), не владеют навыком правильного самообследования. В результате одни пациенты игнорируют небольшие изменения, а другие впадают в панику при появлении любого нового элемента на коже. По этой причине профилактическая и просветительская работа среди населения по вопросам доброкачественных и злокачественных новообразований кожи приобретает особое значение.

Поздняя обращаемость пациентов со ЗНО кожи является прямым следствием недостаточного санитарно-гигиенического просвещения. Пациент обнаруживает у себя новообразование, но в силу нехватки информации либо не придает ему значения, либо пытается решить проблему самостоятельно, прибегая к самолечению. На позднюю обращаемость влияет и то обстоятельство, что опухоль кожи нередко растет медленно и не причиняет человеку дискомфорта, вследствие чего он попросту «привыкает» к ней. В итоге к врачу-специалисту пациент приходит уже тогда, когда опухоль находится на поздней стадии развития.

3.7.3. Сложности дифференциальной диагностики доброкачественных и злокачественных новообразований кожи

Большое разнообразие нозологических форм опухолей кожи

Большое разнообразие нозологических форм опухолей кожи и необходимость применять при их систематизации современные методы исследования, включая иммуногистохимию и молекулярную генетику, могут затруднять диагностику в тех медицинских учреждениях, которые не располагают такими возможностями. Иными словами, чем сложнее и точнее становится классификация, тем ощутимее разрыв между ее требованиями и реальными техническими возможностями значительной части учреждений практического здравоохранения.

Недостаточная онкологическая настороженность врачей

Недостаточная онкологическая настороженность врачей представляет собой комплекс причин, приводящих к поздней диагностике ЗНО кожи. Врач, занимающийся диагностикой опухолей кожи, должен обладать необходимыми профессиональными знаниями и компетенциями, среди которых первостепенное значение имеет знание факторов риска возникновения ЗНО кожи. К таким факторам относятся избыточное ультрафиолетовое облучение, I–II фототип кожи, множественные меланоцитарные невусы, диспластические невусы, семейный анамнез меланомы, а также перенесенное в прошлом злокачественное новообразование кожи. Отсутствие настороженности в отношении пациентов с указанными факторами риска существенно повышает вероятность позднего выявления опухоли.

Субъективность оценки клинических симптомов и результатов диагностики

Врач должен правильно оценивать клинические признаки новообразований кожи и корректно интерпретировать динамику происходящих в них изменений – изменение цвета, формы, размера, появление асимметрии и неровных границ, а также наличие признака «гадкого утенка», то есть опухоли, которая явно выделяется на фоне других новообразований у данного пациента. Врачи, особенно работающие в первичном звене здравоохранения, должны быть хорошо знакомы с признаками и критериями ЗНО кожи на ранних этапах развития. Для этого в их распоряжении должны быть памятки с изображениями и фотографиями меланомы в стадии радиального роста, публикации в научно-практических журналах по дерматоонкологии, а также доступ к корректной информации о ЗНО кожи, размещенной на официальных сайтах.

Отсутствие или нарушения маршрутизации выявленных пациентов

Для снижения показателя высокой запущенности ЗНО кожи и усиления онкологической настороженности, прежде всего у врачей первичного звена здравоохранения, необходима оптимизация маршрутизации пациента с подозрением на ЗНО кожи в специализированное онкологическое учреждение. Даже своевременно выявленное новообразование не гарантирует благоприятного исхода, если пациент не будет вовремя направлен на дообследование и лечение по правильно выстроенному маршруту. Таким образом, организация маршрутизации является неотъемлемым звеном всей системы диагностики ЗНО кожи.

Заболеваемость злокачественными опухолями кожи как во всем мире, так и в Российской Федерации демонстрирует устойчивый ежегодный рост, сопровождающийся значительной смертностью и инвалидизацией населения. В этих условиях своевременная дифференциальная диагностика доброкачественных и злокачественных новообразований кожи приобретает первостепенное значение, однако на пути ее реализации существует ряд системных проблем.

Проведенный анализ позволил впервые систематизировать эти проблемы и разделить их на две группы. К объективным трудностям дифференциальной диагностики ЗНО и ДНО кожи относятся трудности визуальной диагностики новообразований кожи, трудности их морфологической диагностики, субъективность методов клинического исследования, а также недостаточное санитарно-гигиеническое просвещение населения и поздняя обращаемость пациентов. К субъективным сложностям дифференциальной диагностики относятся большое разнообразие нозологических форм новообразований кожи, дефицит врачей-специалистов, способных проводить дифференциальную диагностику между доброкачественными и злокачественными новообразованиями кожи, и недостаточная онкологическая настороженность врачей, занимающихся диагностикой опухолей кожи.

Из проведенной систематизации следует, что в Российской Федерации необходимо усиливать санитарно-гигиеническое просвещение населения по вопросам злокачественных новообразований кожи и одновременно повышать компетенции врачей-специалистов первичного звена здравоохранения, которые могли бы эффективно диагностировать доброкачественные и злокачественные опухоли кожи. Достижению этих целей способствуют программы повышения квалификации, а также оптимизация маршрутизации выявленных пациентов. Вместе с тем очевидно, что часть выделенных проблем – прежде всего субъективность визуальной оценки, дефицит и недостаточная квалификация специалистов, а также недостаточная онкологическая настороженность – может быть частично компенсирована за счет внедрения в клиническую практику программ искусственного интеллекта, способных выступить в роли инструмента поддержки принятия врачебных решений. Именно анализу возможностей таких программ, причин их диагностических ошибок и путей их совершенствования посвящены последующие разделы настоящей главы.

3.8. Анализ причин ошибочной классификации новообразований кожи программами компьютерного зрения на основе технологий искусственного интеллекта

Целью настоящего раздела является анализ причин ошибочной классификации фотоизображений меланоцитарных и немеланоцитарных новообразований кожи программами компьютерного зрения на основе моделей ИИ. Этот анализ проводился посредством визуализации областей интереса при распознавании фотоизображений и исследования систематических различий в характеристиках данных, что необходимо для совершенствования обучения моделей ИИ и эксплуатации ПКЗ.

3.8.1. Материал и методы

Проведен ретроспективный анализ 485 фотоизображений меланоцитарных и немеланоцитарных доброкачественных и злокачественных образований кожи, содержащихся в анонимизированной базе данных (свидетельство о государственной регистрации базы данных № 2025624714 от 27.10.2025).

При подготовке базы данных фотографирование выполнялось с расстояния 8–15 см от поверхности кожи с использованием камер смартфонов Samsung Galaxy A8 (2018) (SM-A530F, Южная Корея) с основной камерой 16 Мп (f/1.7, фокусное расстояние 4 мм, экв. 26 мм, размер сенсора 1/2.8") и автоматической LED-вспышкой; Vivo X100 Ultra (Китай) с основной камерой 50 Мп (Samsung ISOCELL HP9, 1/0.98", f/1.75, фокусное расстояние 22 мм) и автоматической двойной LED-вспышкой; а также цифрового фотоаппарата Sony DSLR-A100 с матрицей 10,2 Мп (APS-C CCD, 23,6×15,8 мм) и макрообъективом Sony 50 mm f/1.4 (SAL-50F14, Япония).

Каждое фотоизображение сопровождалось стандартизированными метаданными: идентификаторами пациента и фотоизображения, полом, возрастом, верифицированным клиническим диагнозом по МКБ-10, анатомической

локализацией опухоли, ее морфологическим типом, стадией и TNM-классификацией для ЗНО, типом фотоизображения, датой его формирования, моделью фотокамеры, разрешением, форматом и MD5-хэшем файла.

Каждое фотоизображение было классифицировано программами ИИ как «злокачественное» или «доброкачественное» (программа «Derma Onko Check») и «как меланома» или «не меланома» (программа «Melanoma Check»). Злокачественные новообразования или меланомы кожи расценивались как положительный результат, а доброкачественные – как отрицательный. По результатам сравнения заключения ИИ с окончательным клиническим диагнозом были сформированы четыре стандартные категории:

- истинно положительные (ИП) – злокачественные опухоли или меланомы кожи, правильно классифицированные ИИ;
- истинно отрицательные (ИО) – доброкачественные опухоли или не меланомы кожи, правильно классифицированные ИИ;
- ложноположительные (ЛП) – доброкачественные опухоли или не меланомы кожи, ошибочно классифицированные ИИ как злокачественные или меланомы кожи;
- ложноотрицательные (ЛО) – злокачественные опухоли или меланомы кожи, ошибочно классифицированные ИИ как доброкачественные или не меланомы кожи.

Все фотоизображения были обработаны двумя ПКЗ на основе моделей ИИ – «Derma Onko Check» и «Melanoma Check». Для визуализации областей, существенно влияющих на классификацию, применялись два метода объяснимого ИИ. Метод взвешенного картирования активации классов (Score-CAM) показывает, на какие участки фотоизображения опирается глубокая нейронная сеть при принятии решения. Метод определения чувствительности к окклюзии (Occlusion Sensitivity) последовательно закрывает небольшие участки фотоизображения и измеряет, как это влияет на уверенность модели в предсказании класса.

Статистическая обработка выполнена в программной среде Python 3.11 (Google Colab) с использованием библиотек pandas, seaborn, scipy и statsmodels. Применялись описательная статистика и визуализация с помощью диаграмм «ящик с усами» (боксплотов) и парных диаграмм рассеяния корреляций. Для выявления систематических различий параметров фотоизображений – яркости, контраста, энтропии, размытия и RGB-метрик – проводились однофакторный дисперсионный анализ (ANOVA) и попарное сравнение характеристик классов при помощи t-критерия Уэлча. Для повышения объективности интерпретации результатов наряду с однофакторным дисперсионным анализом применен многомерный статистический анализ, включающий двухфакторный дисперсионный анализ (ANOVA), многомерный дисперсионный анализ (MANOVA) и множественную логистическую регрессию с z-стандартизацией предикторов и контролем мультиколлинеарности по фактору инфляции дисперсии (VIF, пороговое значение 10). Подобный подход позволяет одновременно учесть влияние категории заключения и применяемой программы ПКЗ, оценить совокупный вектор фотометрических метрик и количественно ранжировать вклад отдельных параметров в вероятность ошибочного заключения, дополняя и подтверждая результаты однофакторных тестов.

3.8.2. Анализ причин ошибочной классификации меланоцитарных опухолей

Для повышения объективности интерпретации результатов наряду с однофакторным дисперсионным анализом применен многомерный статистический анализ, включающий двухфакторный дисперсионный анализ (ANOVA), многомерный дисперсионный анализ (MANOVA) и множественную логистическую регрессию. Подобный подход позволяет одновременно учесть влияние категории заключения и применяемой программы ПКЗ, оценить совокупный вектор фотометрических метрик и количественно ранжировать вклад отдельных параметров в вероятность ошибочного заключения, дополняя и подтверждая результаты однофакторных тестов.

Характеристика выборки

Проведен ретроспективный анализ 151 фотоизображения меланоцитарных опухолей кожи, в которую входили как доброкачественные, так и злокачественные опухоли.

Среди пациентов преобладали мужчины, доля которых составила 67,5 %, тогда как женщины составили 32,5 %. Средний возраст пациентов по всей выборке составил 53,1 года, при этом у пациентов со ЗНО кожи он был равен 61,2 года, а у пациентов с ДНО кожи – 51,5 года. Гендерная асимметрия выборки обусловлена спецификой контингента ведомственной медицинской организации, в которой проводился сбор клинического материала.

Меланоцитарные ЗНО кожи были представлены поверхностно распространяющейся и узловой меланомой. Среди злокачественных случаев чаще всего диагностировалась I стадия заболевания (43,2 %), на втором месте находилась II стадия (34,6 %), далее следовала III стадия (21,0 %), а реже всего диагностировалась меланома *in situ* (1,2 %). Меланоцитарные ДНО кожи включали невусы пограничные плоские, невусы смешанные (в том числе равномерно возвышающийся, по типу «затмения» и по типу «яичницы»), невусы внутридермальные (папилломатозные и непилломатозные), невусы диспластические, невусы пятнистые, невусы голубые и невусы меланоцитарные врожденные.

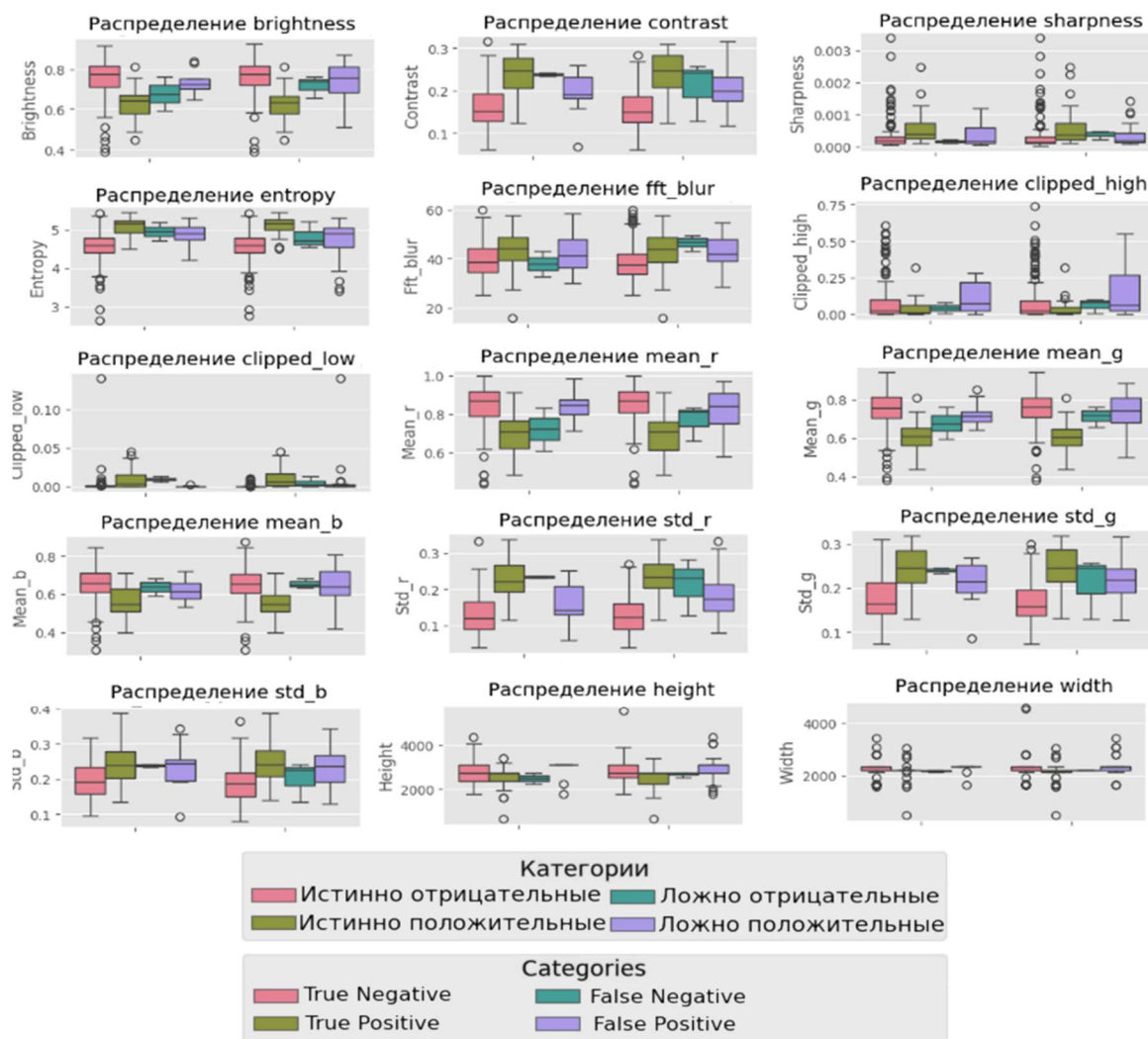
Визуализация метрик качества фотоизображений

Диаграммы «ящик с усами» отражают положение медианы, межквартильный размах и выбросы каждой метрики в разрезе классификационных категорий. Диапазон медиан яркости составил от 0,633 (категория ИП, программа «Melanoma Check») до 0,777 ед. (категория ИО, программа «Derma Onko Check»), что свидетельствует о систематически более темном визуальном облике меланоцитарных ЗНО на корректно классифицированных фотоизображениях.

Медиана энтропии для категории ИП (5,159 ед.) значительно превысила аналогичный показатель для категории ЛП (4,898 ед.), что согласуется с морфологическими особенностями меланом – полиморфизмом пигментной сети, нерегулярностью границ и гетерогенностью окраски, формирующими высокую насыщенность фотоизображения. Единичные выбросы в метрике доли переэкспонированных пикселей (*clipped_high*, медиана 0,011–0,085 ед.) указывают на эпизодически встречающиеся дефекты экспозиции.

По результатам однофакторного дисперсионного анализа статистически значимые межгрупповые различия ($p < 0,05$) были зафиксированы для ряда ключевых метрик. Так, категория ИП характеризовалась пониженной яркостью (медиана 0,638 ед.) по сравнению с категорией ЛП (медиана 0,740 ед.) и повышенной энтропией (медиана 5,159 против 4,898 ед.). Данные закономерности отражают принципиальную особенность меланоцитарных ЗНО – сложную пигментную архитектуру, которую алгоритм лучше улавливает при умеренном, нейтральном освещении без засветок. Аномальные значения долей переэкспонированных (*clippedhigh*) и недоэкспонированных (*clippedlow*) пикселей сигнализируют о фотоизображениях с экспозиционными нарушениями, которые требуют коррекции на этапе предобработки (Рисунок 47).

Парные диаграммы рассеяния корреляций выявили устойчивые зависимости между метриками. Яркость и энтропия находились в выраженной отрицательной связи ($r = -0,7577$): фотоизображения с насыщенной текстурой меланоцитарных структур, как правило, характеризовались умеренной яркостью. Контраст положительно коррелировал с энтропией ($r = 0,5577$), что закономерно, поскольку высококонтрастные фотоизображения воспроизводят тонкие детали пигментного рисунка, характерные для меланом. Резкость обнаружила умеренную положительную связь с энтропией ($r = 0,2564$) и умеренную отрицательную связь с яркостью ($r = -0,3345$).



brightness – яркость; contrast – контраст; sharpness – резкость; entropy – энтропия (мера информационной насыщенности); fft_blur – размытость по быстрому преобразованию Фурье; clipped_high – доля пересвеченных пикселей; clipped_low – доля недосвеченных пикселей; mean_r, mean_g, mean_b – средние значения красного (R), зеленого (G) и синего (B) каналов; std_r, std_g, std_b – стандартное отклонение красного, зеленого и синего каналов; height – высота фотоизображения (пиксели); width – ширина фотоизображения (пиксели)

Рисунок 47 – Визуализация распределения метрик качества фотоизображений по категориям классификации в программах «Derma Onko Check» и «Melanoma Check»

Выраженная отрицательная корреляция между яркостью и энтропией указывает на то, что переэкспонированные фотоизображения теряют именно те текстурные детали (гетерогенность пигментации, нерегулярность сосудистого рисунка), которые служат ключевыми диагностическими признаками меланомы. Иными словами, модель ИИ затрудняется верно классифицировать фотоизображения, обладающие одновременно высокой яркостью и сниженной

энтропией. Полученные корреляции могут использоваться для корректировки весов признаков в модели или для реализации дополнительных этапов нормализации входных данных (Рисунок 48).

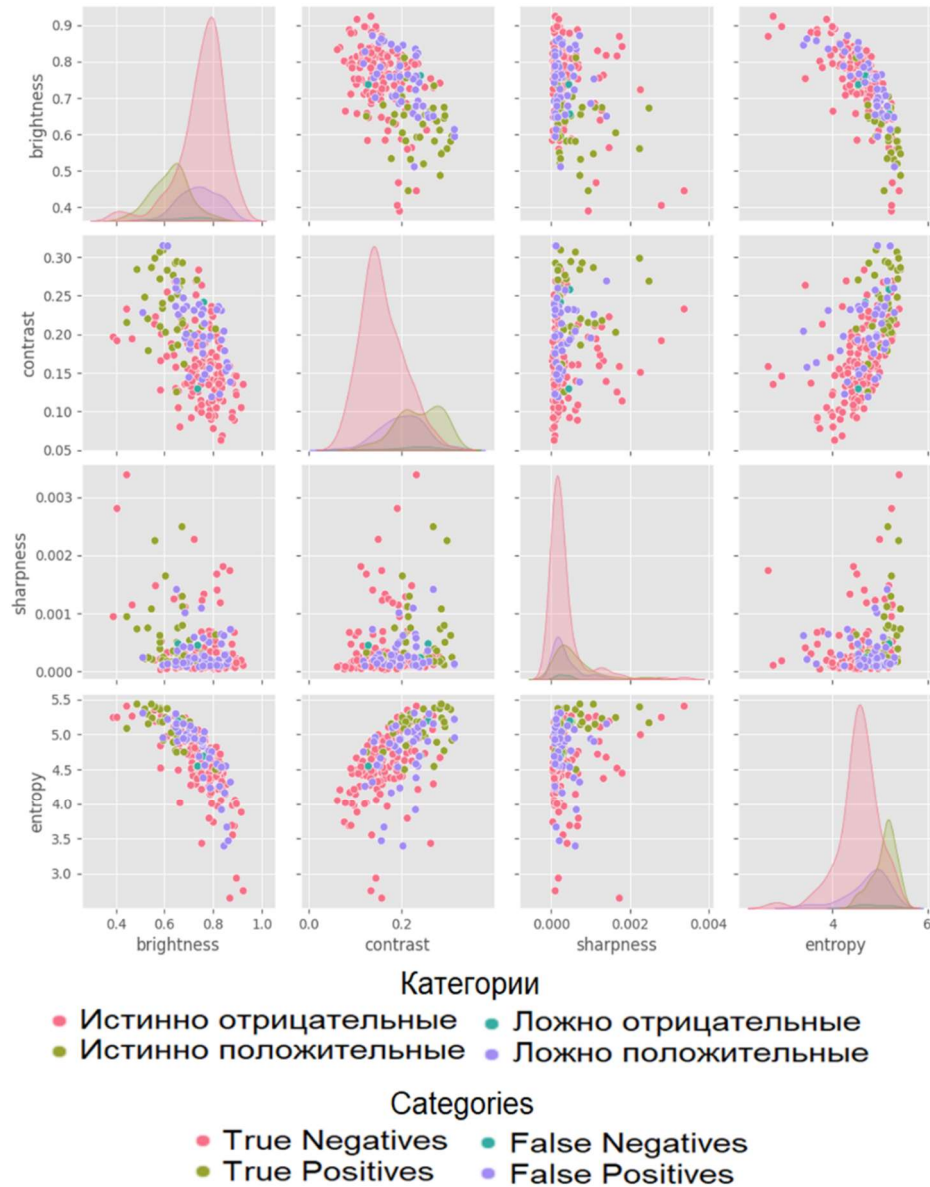


Рисунок 48 – Парные диаграммы рассеяния (pairplot) корреляций между метриками качества фотоизображений (яркость, контраст, резкость, энтропия) по категориям классификации в программах «Derma Onko Check» и «Melanoma Check»

Карты пространственного распределения различий

Метод сравнительного анализа карт пространственного распределения значений пикселей позволяет количественно описать расхождения между категориями классификации на уровне RGB-каналов. Применительно к меланоцитарным опухолям карты различий отражают прежде всего зоны с гетерогенным пигментным рисунком и вариабельной сосудистой архитектурой, которые отличают истинные заключения от ложных.

При сравнении категорий ИО и ЛО интервалы низких пиксельных различий (0,00–0,75) сосредоточены по периферии, что указывает на преобладание низких интенсивностей в ЛО-фотоизображениях. Это согласуется с данными боксплотов: ЛО-фотоизображения систематически темнее, а их темные пигментированные участки хуже распознаются нейронной сетью. Ошибки в категории ЛО могут быть обусловлены недостаточной чувствительностью модели к слабо контрастирующим темным структурам меланомы.

При сравнении категорий ИО и ЛП интервалы высоких различий (0,5–0,7) отражают расхождения в центральной области. Это может свидетельствовать о том, что ЛП-случаи возникают вследствие излишней концентрации модели на паттернах, ассоциированных с признаками злокачественности, которые чаще располагаются в центральной области. В данном случае речь может идти о диспластических невусах, имеющих сходные с меланомой пограничные признаки злокачественности (Рисунок 49).

При сравнении категорий ИП и ЛО интервалы высоких различий (1,2–1,6) охватывают верхние и нижние области, что может указывать на различия в этих зонах при ухудшении экспозиции или фокусировки. Ошибки в категории ЛО при идентификации меланомы, вероятно, связаны с маскировкой структурных критериев злокачественности – нерегулярности пигментной сети – размытием либо недостаточным контрастом фотоизображения.

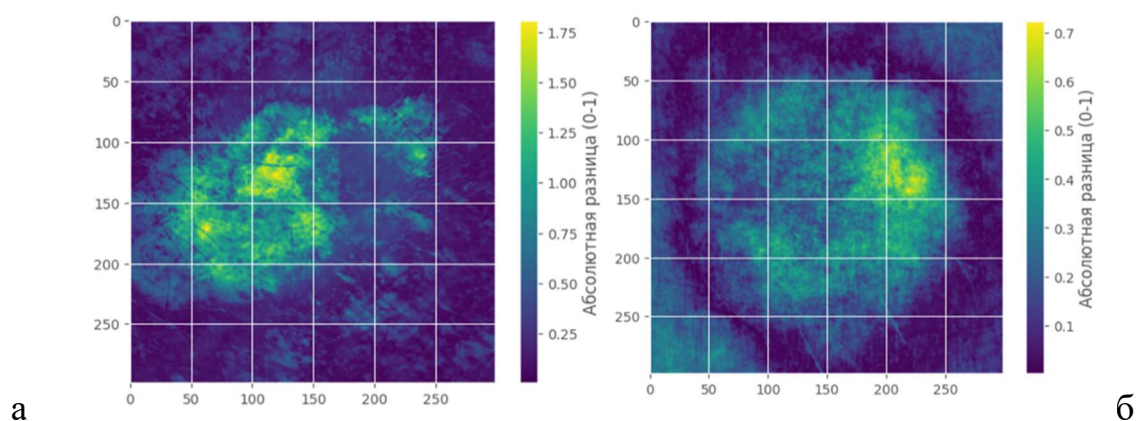


Рисунок 49 – Карта различий: истинно отрицательные (ИО) в сравнении с ложноотрицательными (ЛО) (а) и истинно отрицательные (ИО) в сравнении с ложноположительными (ЛП) (б)

При сравнении категорий ИП и ЛП интервалы низких различий (0,0–0,5) располагаются в центре, что характерно для зон пигментного рисунка. Ошибки в категории ЛП, по-видимому, отражают «ловушку» морфологического сходства: доброкачественные пигментные невусы с выраженной пигментной сетью способны имитировать текстурный профиль меланомы, а модель, настроенная на выявление сложных текстур, классифицирует их как злокачественные (Рисунок 50).

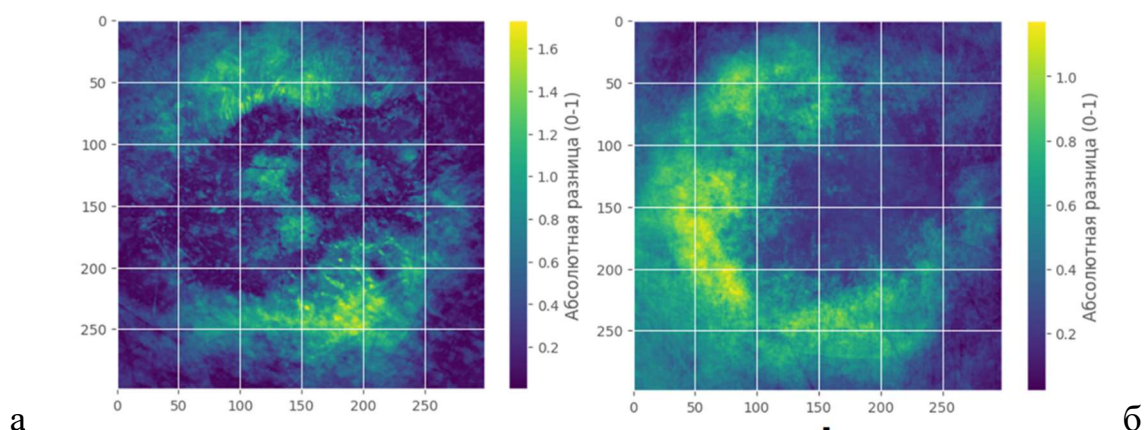


Рисунок 50 – Карта различий: истинно положительные (ИП) в сравнении с ложноотрицательными (ЛО) (а) и истинно положительные (ИП) в сравнении с ложноположительными (ЛП) (б)

Статистическая значимость различий

При оценке классификации фотоизображений программой ИИ «Derma Onko Check» однофакторный дисперсионный анализ выявил статистически значимые межгрупповые различия ($p < 0,05$) по следующим метрикам: яркость ($F=14,83$), контраст ($F=18,97$), энтропия ($F=13,50$), среднее значение красного канала ($F=16,05$), среднее значение зеленого канала ($F=14,53$), среднее значение синего канала ($F=8,67$), стандартное отклонение красного канала ($F=27,12$), стандартное отклонение зеленого канала ($F=16,15$), стандартное отклонение синего канала ($F=6,48$) и доля недоэкспонированных пикселей ($F=3,46$).

При оценке классификации фотоизображений программой ИИ «Melanoma Check» статистически значимые различия были определены для яркости ($F=23,36$), контраста ($F=39,31$), резкости ($F=3,18$), энтропии ($F=18,34$), размытия по FFT ($F=5,65$), доли переэкспонированных пикселей ($F=3,78$), доли недоэкспонированных пикселей ($F=8,31$), средних значений красного ($F=24,47$), зеленого ($F=22,77$) и синего ($F=11,65$) каналов, стандартных отклонений всех трех каналов ($F=15,35-44,51$), а также высоты фотоизображения ($F=6,75$).

t-критерий Уэлча верифицировал попарные различия. Наиболее воспроизводимыми оказались расхождения между категориями ИО и ИП: по яркости ($t=7,44$ для программы «Derma Onko Check» и $t=9,10$ для программы «Melanoma Check», $p < 0,001$), по энтропии ($t=-8,23$ и $t=-10,40$ соответственно, $p < 0,001$), по контрасту ($t=-7,39$ и $t=-9,85$, $p < 0,001$), по средним значениям красного ($t=7,32$ и $t=8,80$, $p < 0,001$) и зеленого ($t=7,60$ и $t=9,34$, $p < 0,001$) каналов. Между категориями ИП и ЛП зафиксированы значимые различия по яркости ($t=-4,11$ и $t=-6,71$, $p < 0,001$), а также по средним значениям и стандартным отклонениям цветовых каналов, особенно красного ($t=3,35-4,32$, $p < 0,01$), что соответствует характерной переэкспозиции ЛП-изображений.

Полученные данные подтверждают, что яркость, энтропия и цветовые характеристики фотоизображений являются значимыми предикторами категории классификационного результата. Аномальные значения доли

переэкспонированных пикселей (clipped_high) подтверждают роль дефектов экспозиции в генезе как ЛП-, так и ЛО-ошибок. Анализ корреляционной структуры метрик (яркость–энтропия: $r=-0,7577$; контраст–энтропия: $r=0,5577$; яркость–контраст: $r=-0,5435$) показал, что метрики не являются независимыми: улучшение одной из них влечет за собой предсказуемые изменения других. Это необходимо учитывать при разработке стратегий аугментации, чтобы не нарушить реалистичность обучающих данных за счет изолированного изменения отдельных параметров.

Многофакторный анализ совместного влияния категории заключения и программы компьютерного зрения

Многофакторный анализ дополняет и подтверждает результаты ранее проведенного однофакторного анализа на трех независимых уровнях. Двухфакторный дисперсионный анализ ANOVA (табл. 5.2) подтвердил статистически значимое влияние категории заключения по всем тринадцати исследованным метрикам ($p<0,001$ для одиннадцати метрик, $p<0,01$ для резкости и доли пересвеченных пикселей). Наиболее выраженные эффекты выявлены для стандартного отклонения красного канала ($F=72,29$), контраста ($F=54,39$), средних значений красного ($F=46,63$) и зеленого ($F=40,34$) каналов, а также для яркости ($F=41,49$); все эти метрики совпадают со значимыми по результатам однофакторного ANOVA, что свидетельствует о воспроизводимости установленных закономерностей. Принципиально новым результатом двухфакторной модели является незначимость фактора ПКЗ ($p=0,39-0,96$) и эффекта взаимодействия ($p=0,15-0,92$), что подтверждает применимость единого набора референсных диапазонов для обеих исследованных программ (Таблица 32).

Таблица 32 – Результаты двухфакторного дисперсионного анализа ANOVA метрик качества фотоизображений по факторам «категория заключения» и «программа компьютерного зрения»

Метрика	Категория заключения		Программа ПКЗ		Взаимодействие	
	F	p	F	p	F	p
Яркость (brightness)	41,49	<0,001	0,04	0,844	0,79	0,502
Контраст (contrast)	54,39	<0,001	0,05	0,822	0,52	0,668
Резкость (sharpness)	5,70	<0,001	0,02	0,888	0,28	0,838
Энтропия (entropy)	21,21	<0,001	0,35	0,554	1,81	0,145
Размытие по FFT (fft blur)	7,36	<0,001	0,13	0,721	0,59	0,623
Доля пересвеченных пикселей (clipped_high)	5,41	0,001	0,59	0,444	0,79	0,501
Доля затемненных пикселей (clipped_low)	14,36	<0,001	0,00	0,962	1,21	0,306
mean_r	46,63	<0,001	0,16	0,690	1,00	0,393
mean_g	40,34	<0,001	0,01	0,934	0,56	0,639
mean_b	20,23	<0,001	0,00	0,965	0,16	0,924
std_r	72,29	<0,001	0,15	0,701	1,07	0,361
std_g	45,64	<0,001	0,24	0,626	0,34	0,796
std_b	17,43	<0,001	0,74	0,390	0,31	0,821
Примечание: F – значение F-статистики дисперсионного анализа; p – уровень статистической значимости						

Многомерный дисперсионный анализ MANOVA подтвердил установленные закономерности на уровне совокупного вектора фотометрических и текстурных метрик. Для фактора категории заключения значение Wilks' lambda составило 0,639 ($F=10,37$, $p<0,001$), Pillai's trace – 0,380 ($F=9,35$, $p<0,001$), Hotelling-Lawley trace – 0,535 ($F=11,41$, $p<0,001$). Эти показатели свидетельствуют о выраженном многомерном различии групп заключений и согласуются с результатами однофакторного и двухфакторного дисперсионного анализа по отдельным метрикам.

Множественная логистическая регрессия позволила количественно ранжировать вклад отдельных метрик в вероятность ошибочного заключения с

поправкой на остальные параметры. После исключения предикторов с высокой мультиколлинеарностью ($VIF=10,2-931,1$) в окончательную модель вошли пять независимых переменных: резкость (*sharpness*), энтропия (*entropy*), размытие по FFT (*fftblur*), доля пересвеченных (*clippedhigh*) и доля затемненных (*clipped_low*) пикселей. Статистически значимыми оказались размытие по FFT ($OR=3,08$; 95 % ДИ 1,77–5,36; $p<0,001$), резкость ($OR=0,31$; 95 % ДИ 0,14–0,65; $p<0,01$) и доля пересвеченных пикселей ($OR=1,64$; 95 % ДИ 1,11–2,41; $p<0,05$). Псевдокоэффициент детерминации Макфаддена составил 0,085 при значимом отношении правдоподобия ($LLR\ p<0,001$). Полученные результаты не противоречат выводам однофакторного анализа, а уточняют их: наибольший количественный вклад в вероятность ошибки вносят показатели текстурной информативности и фокусировки фотоизображения.

Ошибочные заключения классификации, проанализированные методами объяснимого искусственного интеллекта

Для понимания причин ошибочной классификации были визуализированы области интереса моделей. На Рисунке 51 представлено фотоизображение врожденного меланоцитарного невуса, пигментного образования неправильной округлой формы коричневого цвета с более темным центром, с сохраненным волосным покровом над поверхностью образования, с относительно четкими краями и умеренной неравномерностью пигментации. При анализе методом Score-CAM модель сфокусировалась преимущественно на центральной области невуса: тепловая карта выделила зону максимальной значимости (желто-красные тона, значения 0,6–1,0) в пределах темного ядра образования, тогда как периферия невуса и окружающая кожа отображены в темных (незначимых) тонах. Анализ методом Occlusion Sensitivity показал, что центр образования характеризуется высокой чувствительностью (значения 0,5–0,9), а контуры невуса, переходными тонами. Ошибочная классификация невуса как злокачественного образования (ложноположительный результат) обусловлена, по всей видимости, акцентом

модели на неоднородность пигментации в центральной зоне и незначительную нечеткость контуров, морфологически имитирующих признаки ABCDE, характерные для меланомы. Модель не учла в достаточной мере общий клинический контекст: сохраненный волосяной покров над образованием, равномерность краев и отсутствие регрессионных структур.



Рисунок 51 – Примеры ошибочных заключений программ с тепловыми картами Score-CAM и картами чувствительности Occlusion Sensitivity: врожденный меланоцитарный невус, ложноположительный результат

Ошибки классификации в моделях ИИ, применяемых для медицинской визуализации, могут иметь несколько принципиально различных причин. В случае меланоцитарных опухолей ситуацию усложняет высокая морфологическая изменчивость образований: меланома и диспластический невус могут быть практически неразличимы по макроскопическим признакам, что создает «серые зоны» диагностики даже для опытных врачей-специалистов при использовании дерматоскопии. При этом нейронная сеть, оперирующая исключительно пикселями, лишена клинического контекста — анамнеза, динамики изменений, возможности применения дерматоскопии высокого разрешения, — что

принципиально ограничивает ее диагностический потенциал в пограничных случаях. Для анализа конкретных механизмов ошибок в настоящем исследовании применялись методы Score-CAM и Occlusion Sensitivity, позволяющие вскрыть внутреннюю логику классификационных решений нейронной сети.

Анализ методом Score-CAM показал, что при ошибочных заключениях активация тепловых карт нередко смещалась с меланоцитарного образования на нерелевантные области. Парадоксально, но маскировка этих «посторонних» зон не всегда исправляла ошибку: модель начинала правильно локализовать опухоль, однако сохраняла неверную классификацию. Объяснение данного феномена лежит в потере глобального контекста – градиентов цвета, окружающих образование и несущих информацию о пигментном ореоле и зонах регрессии, критичных для разграничения меланомы и невуса, – а также в компенсаторной опоре модели на шумовые элементы при ослаблении основного сигнала.

Метод Occlusion Sensitivity уточнил локализацию критических пикселей. При ЛО-заключениях они концентрировались в зоне нормальной кожи, что свидетельствует о «маскировке» меланомы на фоне недостаточного контраста – типичной ситуации при пониженной яркости и высокой энтропии фотоизображения. При ЛП-заключениях критическими оказывались участки пигментной сети невуса, морфологически сходные с атипичной сетью меланомы, а также артефакты, интерпретировавшиеся моделью как признаки злокачественного роста.

Статистический анализ обнаружил значимые закономерности (ANOVA $F=13,50-39,31$, $p<0,05$ для основных метрик). Ключевое расхождение состоит в следующем: фотоизображения при ИП-заключениях (медиана яркости 0,638 ед., медиана энтропии 5,159 ед.) существенно отличаются от фотоизображений при ЛП-заключениях (медиана яркости 0,740 ед., медиана энтропии 4,898 ед.). Это означает, что алгоритм модели лучше распознает меланомы на фотоизображениях с умеренным освещением и богатой текстурой, тогда как избыточная экспозиция нивелирует именно те признаки, по которым происходит разграничение злокачественных и доброкачественных форм новообразований. Корреляционный

анализ (яркость–энтропия: $r=-0,7577$; контраст–энтропия: $r=0,5577$) подтвердил взаимозависимость метрик и необходимость комплексной, а не изолированной нормализации параметров фотоизображений.

Многофакторный анализ дополняет и подтверждает результаты однофакторных тестов. Значимость фактора категории заключения сохраняется по всем тринадцати метрикам в двухфакторном дисперсионном анализе ANOVA и на уровне совокупного вектора признаков по многомерному дисперсионному анализу MANOVA, что свидетельствует о воспроизводимости установленных межгрупповых различий. Незначимость фактора ПКЗ и эффекта взаимодействия ($p=0,15-0,99$) указывает на то, что выявленные закономерности характеризуют именно качество входных фотоизображений, а не специфику конкретной нейросетевой архитектуры. По данным множественной логистической регрессии наибольший вклад в вероятность ошибочного заключения вносят показатели текстурного и частотного качества – размытие по FFT ($OR=3,08$) и резкость ($OR=0,31$), что согласуется с клинической интерпретацией недостаточной фокусировки как ведущей причины маскировки диагностически значимых деталей пигментного рисунка.

Ограничения настоящего исследования определяются объемом анализируемой выборки и демографическими особенностями использованной базы данных, обусловленными спецификой медицинской организации, в которой были созданы набор данных и база данных. Вместе с тем поставленная цель – анализ причин ошибок ИИ и разработка практических мер по их минимизации – была достигнута в полной мере. Дообучение моделей ИИ с учетом выявленных факторов ошибок позволило существенно сократить долю ЛП- и ЛО-заключений при классификации меланоцитарных опухолей.

Практические рекомендации

На основании проведенного исследования сформулированы ключевые рекомендации для пользователей приложений. Для обеспечения корректной

работы систем ПКЗ на основе ИИ при классификации меланоцитарных опухолей пользователям рекомендуется:

- поддерживать яркость фотоизображений ниже 0,75 в нормализованной шкале, обеспечивая равномерное рассеянное освещение без теней и бликов, поскольку более высокая яркость повышает вероятность ЛП-результатов;
- исключать недоэкспозицию, которая маскирует темные пигментированные структуры меланомы и увеличивает риск ЛО-заключений;
- применять функцию автоматического баланса белого для стандартизации цветовых каналов, поскольку средние значения и стандартные отклонения каналов R, G и B статистически значимо связаны с типом классификационного результата;
- обеспечивать контрастность выше 0,20 для сохранения клинически информативного рисунка пигментной сети и четкости границ меланоцитарного новообразования;
- минимизировать размытость фотоизображения, стабилизируя камеру при съемке и используя программные фильтры повышения резкости;
- осуществлять фотосъемку в режиме макро с расстояния 8–15 см, центрируя новообразование в кадре, с достижением значения энтропии выше 5,0 ед. при разрешении не менее 2000–3000 пикселей по длинной стороне.

Разработчикам систем ПКЗ на основе ИИ для классификации меланоцитарных новообразований рекомендуется:

- включать в пайплайн предобработки автоматический баланс белого и гамма-коррекцию, поскольку вариабельность RGB-каналов демонстрирует статистически значимую связь с ошибками классификации;
- применять фильтры текстурного усиления (Sobel и аналогичные) для увеличения энтропии фотоизображений выше 5,0 ед., что повышает различимость пигментной гетерогенности, характерной для меланом;
- реализовывать коррекцию размытости (фильтр Wiener) для восстановления мелких структурных элементов пигментного рисунка, критичных для дифференциации злокачественных и доброкачественных меланоцитарных новообразований;

– внедрять онлайн-аугментацию по яркости и контрасту с учетом выявленной отрицательной корреляции яркость–энтропия ($r=-0,7577$), обеспечивая устойчивость модели к вариациям освещения без утраты реалистичности тренировочных данных;

– осуществлять мониторинг ключевых метрик (яркость, энтропия, контраст, доля переэкспонированных и недоэкспонированных пикселей) после каждой эпохи обучения для предотвращения переобучения на артефактах.

Полученные результаты согласуются с данными систематических обзоров, в которых показано, что методы интерпретации ИИ позволяют выявлять конкретные источники диагностических ошибок нейронных сетей при анализе изображений кожных новообразований, в частности смещение внимания модели на нерелевантные участки изображения при недостаточном качестве входных данных. Реализация предложенных рекомендаций позволяет снизить частоту диагностических ошибок, обеспечить воспроизводимость классификационных решений и повысить клиническую надежность ПКЗ на основе моделей ИИ при работе с фотоизображениями меланоцитарных новообразований кожи.

Проведенное исследование подтвердило эффективность интеграции методов объяснимого искусственного интеллекта и количественного статистического анализа для изучения природы ошибочных заключений при дифференциальной диагностике меланоцитарных опухолей кожи с помощью программ компьютерного зрения на основе моделей ИИ. Установлены системные различия в метриках фотоизображений – прежде всего повышенная яркость и сниженная энтропия в ложноположительных случаях, а также смещение фокуса нейронной сети на фоновые элементы в ложноотрицательных случаях. Значимые корреляции между метриками и межгрупповые различия цветовых характеристик свидетельствуют о том, что освещение, информационная насыщенность и пространственная структура фотоизображения совокупно определяют достоверность заключений модели при анализе меланоцитарных новообразований. Полученные данные создают основу для целенаправленного совершенствования

как обучающих наборов данных, так и условий фотофиксации меланоцитарных опухолей в реальной клинической практике.

3.8.3. Анализ причин ошибочной классификации немеланоцитарных опухолей

Целью настоящего раздела является анализ причин ошибочной классификации изображений немеланоцитарных (кератиноцитарных/эпидермальных) опухолей кожи программами на основе технологий ИИ. Для достижения этой цели проводилось выявление систематических различий в характеристиках данных и визуализация областей интереса при распознавании изображений, что необходимо для повышения эффективности обучения и использования программ компьютерного зрения.

Характеристика выборки

Проведен ретроспективный анализ 151 фотоизображения немеланоцитарных (кератиноцитарных/эпидермальных) опухолей кожи, в которую входили как доброкачественные, так и злокачественные опухоли.

Среди пациентов фотоизображения опухолей которых были включены в базу данных, насчитывалось 110 мужчин (72,8 %) и 41 женщина (27,2 %); возраст пациентов варьировался от 18 до 92 лет, а средний возраст составил $63,8 \pm 12,4$ года. Гендерная асимметрия выборки обусловлена спецификой контингента ведомственной медицинской организации, в которой проводился сбор клинического материала.

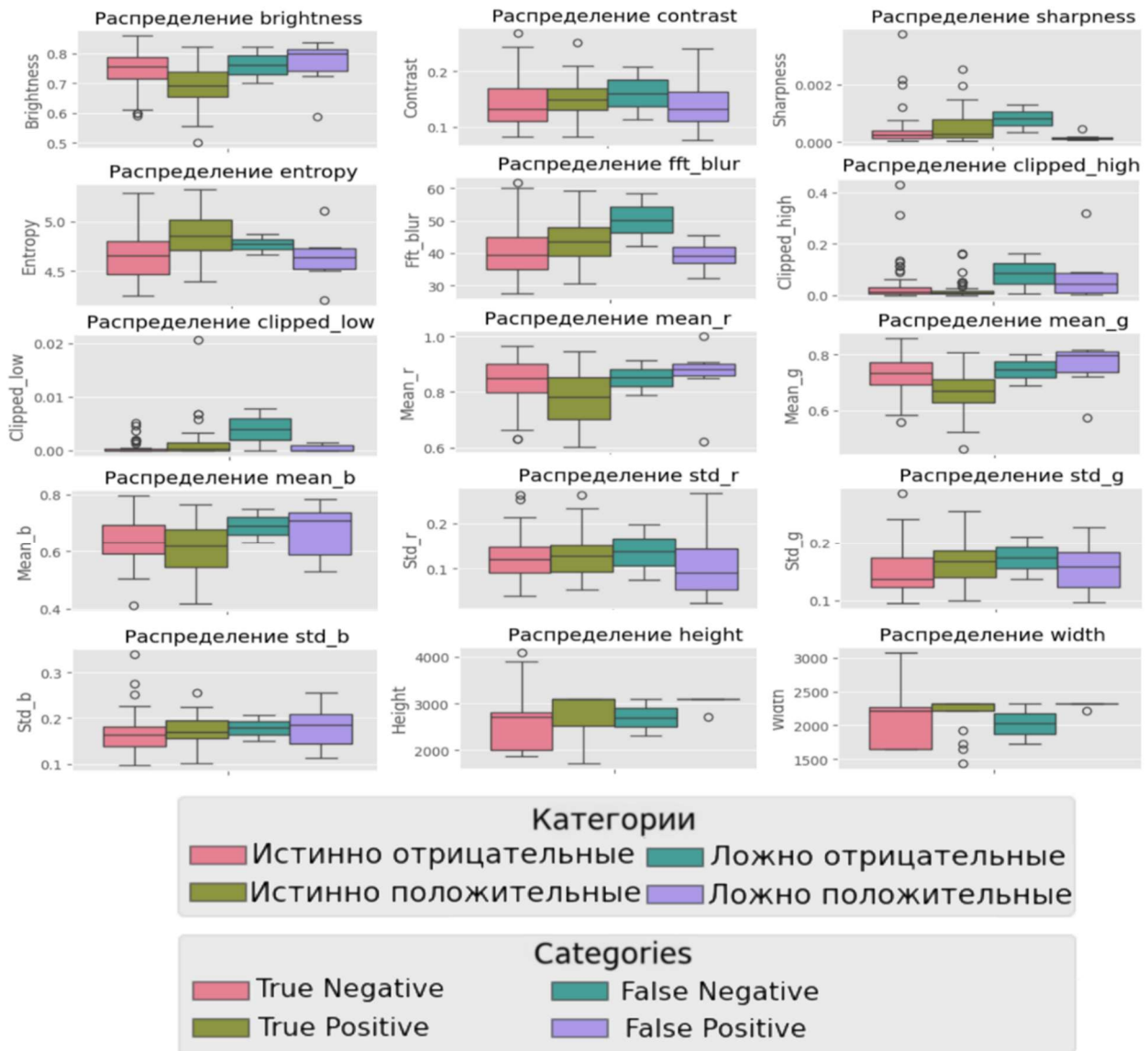
Визуализация метрик качества фотоизображений

Диаграммы «ящик с усами» позволяют выявить центральные тенденции и разброс метрик в разрезе категорий ошибок. Так, для яркости медианные значения

варьировались от 0,6914 ед. (категория ИП) до 0,7994 ед. (категория ЛП) при межквартильном размахе от 0,0682 до 0,0935 ед. Энтропия показывала более высокие медианные значения для категории ИП (4,8584 ед.) по сравнению с категорией ЛП (4,6414 ед.), что может указывать на большую сложность текстур в правильно классифицированных случаях. Выбросы, то есть отдельные точки за пределами «усов», наблюдались, например, в метрике доли переэкспонированных пикселей (`clipped_high`, медиана 0,0123–0,0839 ед.), что свидетельствует об аномалиях освещения в отдельных фотоизображениях.

Статистически значимые различия ($p < 0,05$ по однофакторному дисперсионному анализу) были зафиксированы для ряда ключевых метрик: медиана яркости оказалась ниже в ИП-случаях (0,6914 ед.) по сравнению с ЛП-случаями (0,7994 ед.), а медиана энтропии – выше в ИП-случаях (4,8584 ед.), чем в ЛП-случаях (4,6414 ед.). Эти различия в медианных значениях и разбросе метрик между группами могут быть связаны с особенностями алгоритма ИИ, который лучше распознает сложные текстуры при естественном освещении. Выбросы среди значений долей переэкспонированных (`clippedhigh`) и недоэкспонированных (`clippedlow`) пикселей указывают на проблемы с качеством фотоизображений (переэкспозицию или недодержку), что требует дополнительной предобработки данных (Рисунок 52).

Парные диаграммы рассеяния корреляций между метриками качества фотоизображений показывают, что яркость имеет сильную отрицательную корреляцию с энтропией ($r = -0,7117$): фотоизображения с более высокой энтропией (сложностью текстуры), как правило, характеризуются меньшей яркостью. Контраст положительно коррелирует с энтропией ($r = 0,6431$), что логично, поскольку высокая контрастность часто сопровождается сложными паттернами. Резкость демонстрирует слабую положительную корреляцию с энтропией ($r = 0,2736$) и слабую отрицательную корреляцию с яркостью ($r = -0,2215$).



brightness – яркость; contrast – контраст; sharpness – резкость; entropy – энтропия (мера информационной насыщенности); fft_blur – размытость по быстрому преобразованию Фурье; clipped_high – доля пересвеченных пикселей; clipped_low – доля недосвеченных пикселей; mean_r, mean_g, mean_b – средние значения красного (R), зеленого (G) и синего (B) каналов; std_r, std_g, std_b – стандартное отклонение красного, зеленого и синего каналов; height – высота изображения (пиксели); width – ширина изображения (пиксели)

Рисунок 52 – Визуализация распределения метрик качества фотоизображений по категориям классификации в программах «Derma Onko Check» и «Melanoma Check»

Сильная отрицательная корреляция между яркостью и энтропией может указывать на то, что ИИ испытывает трудности с классификацией фотоизображений, обладающих низкой яркостью и высокой текстурной сложностью. Эти данные могут быть использованы для корректировки весов

признаков в модели или для дополнительной нормализации данных перед классификацией (Рисунок 53).

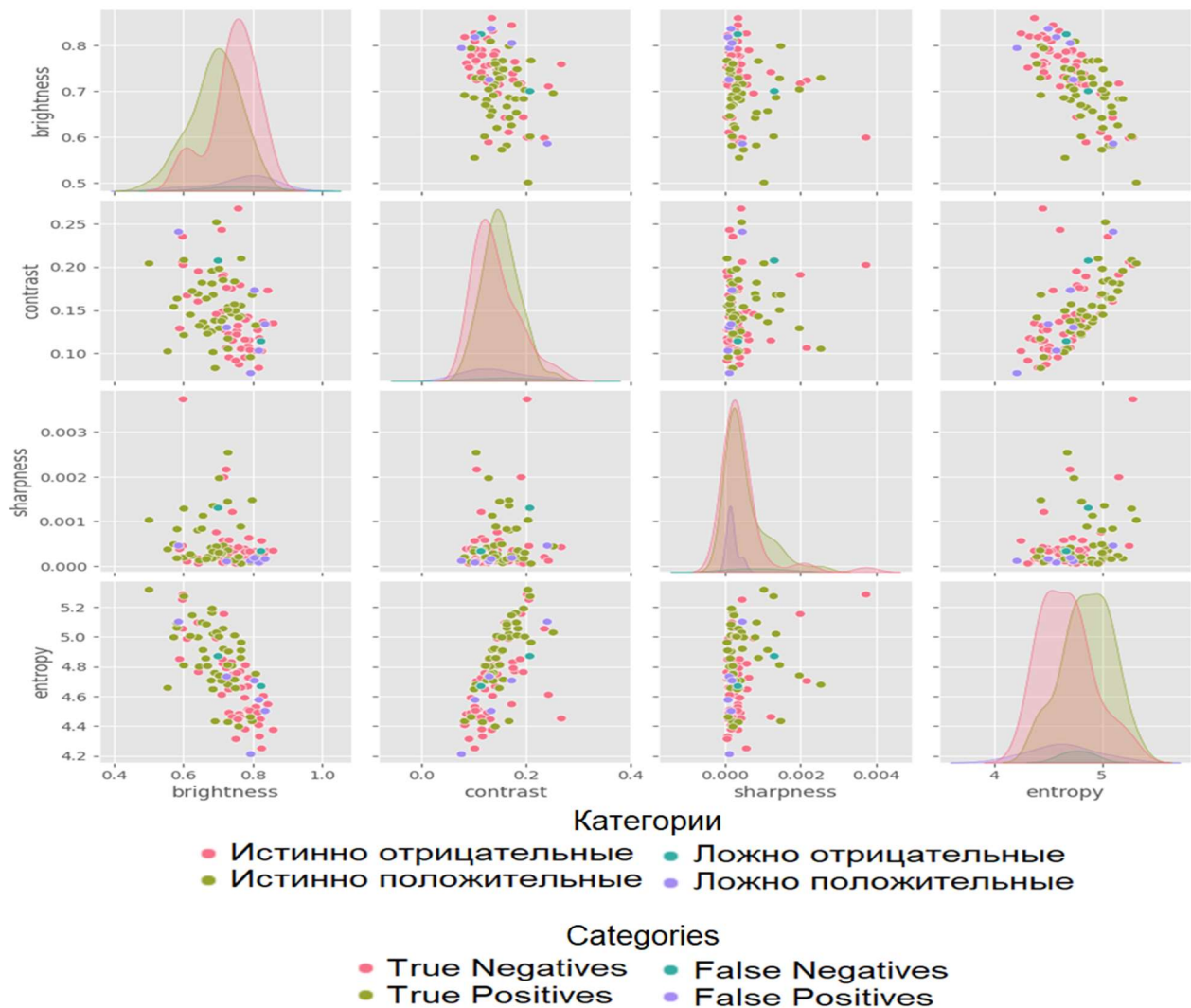


Рисунок 53 – Парные диаграммы рассеяния (pairplot) корреляций между метриками качества фотоизображений (яркость, контраст, резкость, энтропия) по категориям классификации в программах «Derma Onko Check» и «Melanoma Check»

Карты пространственного распределения различий

Разграничить истинные и ложные заключения моделей ИИ можно с помощью метода статистического сравнения карт пространственного распределения (карт различий) значений пикселей фотоизображений для RGB-каналов. Этот метод позволяет исследовать различия в значениях пикселей на карте

и использовать эту информацию для классификации результатов. Карты различий выявляют расхождения в центральных областях фотоизображений, связанные с текстурными и структурными паттернами опухолей.

Сравнение категорий ИО и ЛО демонстрирует зоны с высоким различием (порядка 0,7–1,6 ед.) в виде желтых пятен в верхне-центральной области. Бирюзовые зоны (порядка 0,3–0,5 ед.) образуют нерегулярные пятна, а темные области (порядка 0,1–0,2 ед.) покрывают большую часть сетки, особенно ее верхнюю половину, что говорит о высокой однородности. Топовые интервалы указывают на смещение к более высоким и более низким тонам, что может отражать различия в интенсивности краев и фона. Ошибки в категории ЛО могут быть связаны с шумом меток или недостаточной чувствительностью модели к слабым сигналам, особенно если обучающие данные содержат артефакты или недостаток положительных примеров.

При сравнении категорий ИО и ЛП выявлены высокие различия (до 1,6 ед.) в центральной области с яркими желтыми ядрами и зелеными структурами (порядка 1,0–1,4 ед.). Темные зоны (порядка 0,2–0,4 ед.) преобладают на краях и в центральных областях, указывая на согласованность за пределами центральной области. Топовые интервалы отражают низкие интенсивности, что может свидетельствовать о шуме или слабых артефактах. Ошибки в категории ЛП могут возникать вследствие избыточной чувствительности к шуму, особенно если модель переобучена на положительных образцах с артефактными особенностями, имитирующими структуры ИО-класса (Рисунок 54).

Большие различия (до 1,4) при сравнении категорий ИП и ЛО сосредоточены в центральной области с ярким желтым ядром, особенно в верхних и нижних квадрантах. Зеленые и бирюзовые градиенты (порядка 0,8–1,2 ед.) простираются наружу, тогда как темные зоны (порядка 0,2–0,4 ед.) видны в средней периферии. Топовые интервалы охватывают широкий диапазон, включая высокие интенсивности, что может указывать на значительные вариации в сигнальных областях. Ошибки в категории ЛО могут быть вызваны недостаточной чувствительностью к центральным пикам, особенно если ЛО-фотоизображения

имеют ослабленные сигналы из-за размытия или низкого разрешения, что требует улучшения обработки фотоизображений в процессе обучения.

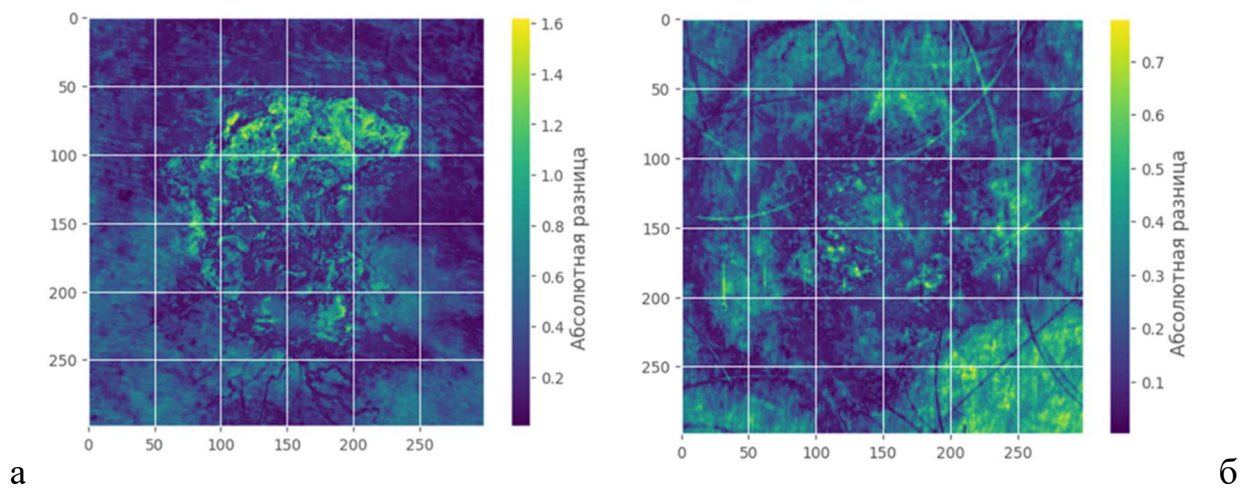


Рисунок 54 – Карта различий: истинно отрицательные (ИО) в сравнении с ложноотрицательными (ЛО) (а) и истинно отрицательные (ИО) в сравнении с ложноположительными (ЛП) (б)

Области с наибольшими различиями при сравнении категорий ИП и ЛП (порядка 0,8–1,0 ед.) находятся по периферии, при этом средние различия (порядка 0,4–0,6 ед.) рассеяны в центре. Темные полосы (порядка 0,1–0,3 ед.) на внешних кольцах указывают на перекрытие периферийных элементов. Топовые интервалы сосредоточены в области низких тонов, что может отражать шум или слабые границы. Ошибки в категории ЛП могут быть результатом переобобщения модели на аугментированных данных, где шум или тени имитируют структуры ИП-фотоизображений, что требует более строгой регуляризации (Рисунок 55).

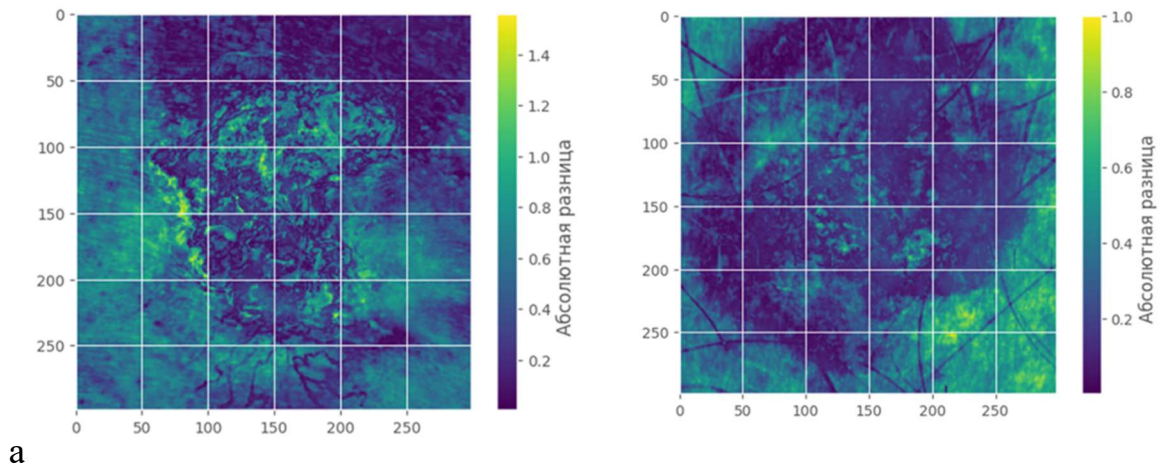


Рисунок 55 – Карта различий: истинно положительные (ИП) в сравнении с ложноотрицательными (ЛО) (а) и истинно положительные (ИП) в сравнении с ложноположительными (ЛП) (б)

Статистическая значимость различий

Согласно результатам дисперсионного анализа (ANOVA), выявлены статистически значимые различия ($p < 0,05$) для яркости ($F=5,1848$), энтропии ($F=5,2509$), размытия ($F=3,1136$), среднего значения красного канала ($F=3,3812$), среднего значения зеленого канала ($F=5,3315$) и высоты фотоизображения ($F=2,5645$). t-критерий Уэлча подтвердил значимые различия по яркости между категориями ИО и ИП ($t=3,6619$; $p=0,0004$), по энтропии между категориями ИО и ИП ($t=-3,8187$; $p=0,0002$), по размытию между категориями ИО и ИП ($t=-2,4678$; $p=0,0155$), а также между категориями ИП и ЛП ($t=2,3417$; $p=0,0450$). Установлены различия по среднему значению красного канала между категориями ИО и ИП ($t=2,9321$; $p=0,0043$), по среднему значению зеленого канала между категориями ИО и ИП ($t=3,5506$; $p=0,0006$), по дисперсии зеленого канала между категориями ИО и ИП ($t=-2,0714$; $p=0,0413$), а также по резкости между категориями ИО и ЛП ($t=2,2982$; $p=0,0270$) и между категориями ИО и ЛП по совокупному показателю ($t=3,3535$; $p=0,0021$).

Эти данные подтверждают влияние яркости, энтропии и цветовых средних фотоизображений на возникновение ошибок, причем наибольшие различия отмечаются между категориями ИП и ЛП, что обосновывает необходимость фокуса

на этих метриках при доработке моделей. Выбросы среди значений доли переэкспонированных пикселей (0,0123–0,0839 ед.) указывают на влияние данного параметра на возникновение ЛО- и ЛП-результатов. Показано также существование разнонаправленных корреляций между метриками качества изображений (яркость–энтропия: $r=-0,7117$; контраст–энтропия: $r=0,6431$), что отражает взаимосвязь освещения и текстурной сложности и может косвенно влиять на ошибки классификации.

Ошибочные заключения, проанализированные методами объяснимого искусственного интеллекта

Для понимания причин ошибочной классификации были визуализированы области интереса моделей. На Рисунке 56 представлено фотоизображение себорейного кератоза, пигментированного образования темно-коричневого, почти черного цвета с выраженной неоднородной пигментацией, дольчатой поверхностью, характерной бородавчатой текстурой и нечеткими неровными краями. При анализе методом Score-CAM сгенерирована тепловая карта, на которой наиболее значимые для решения модели зоны (красные и желто-зеленые тона, значения 0,5–1,0 ед.) сконцентрированы по краям образования и в зонах максимального контраста пигмента, тогда как центральные участки с выраженной кератотической текстурой представлены тонами средней значимости. При оценке методом Occlusion Sensitivity выявлено, что периферийные зоны образования, включая участки резкого перехода пигментации, характеризуются высокой чувствительностью (значения 0,7–1,0 ед.) Ошибочная классификация себорейного кератоза как злокачественного образования (ложноположительный результат) обусловлена, по всей видимости, тем, что модель интерпретировала выраженную неоднородность пигментации, нечеткие края и бородавчатую текстуру поверхности как признаки злокачественности, характерные для пигментированного базальноклеточного рака или меланомы. Подобный паттерн ошибки отражает системную проблему переобучения на визуальных коррелятах

злокачественности при недостаточной представленности пигментированных форм себорейного кератоза в обучающей выборке.



Рисунок 56 – Примеры ошибочных заключений программ с тепловыми картами Score-CAM и картами чувствительности Occlusion Sensitivity: себорейный кератоз, ложноположительный результат

В приведенных примерах анализ методами объяснимого ИИ показал, что в ЛО-ошибках модель отвлекается на фон (нормальную кожу), а в ЛП-ошибках переобучается на артефактах (теньях, волосах), игнорируя края опухоли. Выбросы значений долей переэкспонированных и недоэкспонированных пикселей (0,0123–0,0839 ед.) вновь указывают на проблемы качества фотоизображений.

Дообучение моделей и его результаты

После выявления ключевых факторов ошибок классификации (низкая яркость, высокая энтропия, переэкспозиция и текстурная сложность фотоизображений) было проведено дообучение моделей с учетом этих

особенностей и выпущены версии 2.0 программ «Derma Onko Check» (свидетельство о государственной регистрации программы для ЭВМ № 2025661513) и «Melanoma Check» (свидетельство о государственной регистрации программы для ЭВМ № 2025661514). В решениях программы «Derma Onko Check» диагностическая точность возросла с 90,9 до 92,4 %, чувствительность практически не изменилась (94,7 вместо 94,4 %), а специфичность увеличилась с 89,6 до 91,8 % (на 2,2 процентного пункта). Соответственно, доля ЛП-заключений снизилась с 10,4 до 8,2 %, а доля ЛО-заключений незначительно увеличилась (на 0,5 %). В решениях программы «Melanoma Check» диагностическая точность возросла с 84,3 до 92,4 %, чувствительность увеличилась с 92,1 до 94,9 %, а специфичность – с 82,3 до 92,1 %. Это означает снижение доли ЛП-заключений на 55 % (до 7,9 %), а доли ЛО-заключений – на 35 % (до 5,1 %).

Как и в большинстве моделей ИИ для медицинской визуализации, ошибки классификации могут возникать по нескольким причинам, к которым относятся переобучение на артефактах, недостаточная фокусировка на релевантных признаках и влияние фоновых элементов фотоизображения. Для анализа таких ошибок были задействованы методы Score-CAM и Occlusion Sensitivity, позволяющие визуализировать, на каких областях фотоизображения модель фокусируется при проведении классификации.

Для оценки того, правильно ли модель фокусируется на образованиях, все фотоизображения кератиноцитарных/эпидермальных опухолей в наборе данных были проанализированы с помощью метода Score-CAM. Этот метод генерирует тепловые карты, где более яркие (обычно красные) области указывают на зоны, наиболее влияющие на заключение модели. В большинстве случаев модель правильно обращала внимание на само образование: на пигментацию, асимметрию границ, цветовые вариации и другие клинические признаки. Однако в некоторых ошибочных случаях тепловая карта выделяла нерелевантные области, такие как окружающая кожа, волосы или артефакты освещения. Примечательно, что при обрезке (маскировке) этих «лишних» областей модель начинала правильно фокусироваться на опухоли, но при этом по-прежнему выдавала неверное

заключение. Причины данного феномена могут быть связаны с переобучением на коррелирующих, но не причинных признаках (модель «учится» ассоциировать определенные текстуры кожи со злокачественностью, даже если они не относятся к опухоли), с потерей контекста (обрезка удаляет вспомогательные признаки, такие как градиенты цвета вокруг опухоли, что приводит к недооценке злокачественности), а также с шумом на фотоизображениях, когда модель полагается на артефакты для компенсации слабых сигналов в основной области. Такие случаи подчеркивают необходимость баланса между фокусом на локальных признаках и учетом глобального контекста фотоизображения.

Для более детального определения того, на какие текстуры, края и другие элементы обращает внимание модель, был применен метод объяснимого ИИ Occlusion Sensitivity. Данный подход работает по следующему принципу: входное фотоизображение разделяется на патчи (небольшие блоки), каждый из которых последовательно «закрывается» (например, заменяется на среднее значение или черный цвет), после чего оценивается изменение в заключении модели. Если окклюзия патча приводит к значительному снижению уверенности модели или к изменению класса, этот патч считается важным. Результат визуализируется в виде карты чувствительности: положительные значения (красные области на шкале от 1,0 до 0) указывают на патчи, удаление которых уменьшает вероятность правильного заключения (то есть они важны для верного решения), а отрицательные значения (синие, до $-1,0$) – на патчи, удаление которых увеличивает эту вероятность, то есть «мешающие» модели. Интерпретация карты позволяет выявить, фокусируется ли модель на клинически релевантных признаках (например, неровных краях опухоли) или на шумовых элементах (волосах, фоне). Этот метод особенно полезен в медицинской визуализации, поскольку он не зависит от внутренних градиентов сети и способен выявлять скрытые предвзятости.

При анализе методом Occlusion Sensitivity было обнаружено, что в случаях ЛО-заклучения, когда модель классифицирует злокачественную опухоль как доброкачественную, важные пиксели часто располагались в областях обычной

кожи без образований. Это указывает на то, что модель «отвлекается» на фоновые текстуры, возможно, из-за недостаточного контраста опухоли или переобучения на наборах данных с преобладанием нормальной кожи. Вместе с тем встречался и ЛП-случай, когда модель в большей степени обращала внимание на нормальную кожу, но все равно классифицировала опухоль как злокачественную; здесь, вероятно, повлияли артефакты (тени, волосы), которые модель ошибочно интерпретировала как признаки злокачественности.

В рамках проведенного исследования методы объяснимого искусственного интеллекта Score-CAM и Occlusion Sensitivity, предназначенные для визуализации фокуса моделей на релевантных областях фотоизображений, были интегрированы с комплексным статистическим и визуализационным анализом. Использование метода Score-CAM позволяет оценить, на какие признаки (например, текстуру или края опухоли) модель обращает внимание, тогда как метод Occlusion Sensitivity выявляет критические пиксели, влияющие на заключение. Эти методы дополняются подходами парных диаграмм рассеяния корреляций, боксплотов метрик и карт различий.

Анализ ошибок классификации выявил значимые факторы – яркость, энтропию и цветовые характеристики (ANOVA: $F=3,11-5,33$; $p<0,05$). Категория ИП характеризуется более низкой яркостью (0,6914 ед.) и высокой энтропией (4,8584 ед.), тогда как категория ЛП – более высокой яркостью (0,7994 ед.) и низкой энтропией (4,6414 ед.), что указывает на трудности ИИ при работе с текстурной сложностью и освещением. Корреляции (яркость–энтропия: $r=-0,7117$; контраст–энтропия: $r=0,6431$) подтверждают влияние текстуры на возникновение ошибок. Карты различий показывают расхождения в центральных областях фотоизображений, особенно в каналах G и B (различия до 1,4–1,6). Выбросы среди значений доли переэкспонированных пикселей (0,0123–0,0839 ед.) указывают на аномалии переэкспозиции, влияющие на возникновение ЛО- и ЛП-результатов.

Ограничения исследования заключаются в относительно малой выборке, а также в гендерных и возрастных особенностях пациентов, связанных с контингентом ведомственной медицинской организации. Данные ограничения не

снижают практической ценности работы, поскольку была поставлена цель проанализировать причины ошибок ИИ и разработать рекомендации по уменьшению их количества. Таким образом, дообучение с учетом выявленных факторов ошибок (яркости, энтропии, переэкспозиции, текстурной сложности) позволило существенно сократить количество как ЛП-, так и ЛО-заклучений и повысить общую диагностическую точность обеих программ, причем особенно выраженный прирост показателей наблюдался у программы «Melanoma Check».

Практические рекомендации

На основании проведенного исследования сформулированы ключевые рекомендации для пользователей приложений, направленные на минимизацию ошибок классификации за счет улучшения качества изображений, нормализации метрик и снижения вариабельности данных. Для решения задачи дифференциальной диагностики кератиноцитарных/эпидермальных опухолей кожи с помощью программ компьютерного зрения пользователям рекомендуется:

- обеспечивать равномерное освещение без теней, поддерживая яркость фотоизображений на уровне ниже 0,75 в нормализованной шкале, поскольку более высокая яркость повышает вероятность ЛП-результатов;
- избегать недоэкспозиции или переэкспозиции для снижения вероятности ЛО-результатов;
- использовать функцию автоматического баланса белого для нормализации цветовых каналов с акцентом на каналы G и B;
- стремиться к контрастности фотоизображения выше 0,20 для четкой визуализации границ опухоли, что уменьшает долю ЛО-результатов;
- минимизировать размытие, поддерживая резкость выше среднего уровня, применяя фильтры усиления резкости и текстуры и стабилизируя камеру во время съемки;
- осуществлять съемку крупным планом для достижения энтропии выше 4,8 ед., избегая артефактов, желательно с активацией режима макросъемки;

- обеспечивать высокое разрешение изображений в диапазоне 2000–3000 пикселей, центрируя опухоль в кадре для минимизации фоновых элементов.

Кроме того, сформулированы ключевые рекомендации для разработчиков систем ИИ, предназначенных для дифференциальной диагностики опухолей кожи. Для оптимизации качества данных и уменьшения влияния выбросов на этапах предобработки, аугментации и обучения рекомендуется:

- нормализовать цветовые каналы с использованием автоматического баланса белого и гамма-коррекции, поскольку вариации в RGB-каналах коррелируют с ошибками классификации, особенно в случаях переэкспозиции или недоэкспозиции;

- применять фильтры усиления текстуры для повышения энтропии фотоизображений выше 4,8 ед., так как это помогает в распознавании сложных текстур и снижает долю ЛО-ошибок;

- использовать методы коррекции размытости (например, фильтр Wiener) для повышения резкости изображения и фокуса модели на релевантных областях;

- внедрять аугментацию яркости и цвета при обучении моделей, чтобы компенсировать отрицательную корреляцию яркости с энтропией и повысить устойчивость к вариабельности данных;

- обеспечивать мониторинг ключевых метрик (яркость, энтропия, контраст, доли недоэкспонированных и переэкспонированных пикселей) после каждой эпохи обучения для предотвращения переобучения на артефактах (фоне, волосах, тенях).

Реализация данных рекомендаций позволяет снизить частоту ошибок классификации, существенно повысить диагностическую точность и обеспечить устойчивость моделей к реальной вариабельности клинических фотографий немеланоцитарных опухолей кожи.

Проведенное исследование демонстрирует эффективность комбинации методов объяснимого ИИ и статистических подходов для анализа ошибок при дифференциальной диагностике немеланоцитарных опухолей кожи с использованием программ ИИ «Derma Onko Check» и «Melanoma Check». Выявленные систематические различия в метриках фотоизображений, такие как

повышенная яркость и низкая энтропия в ЛП-случаях, а также фокус модели на фоновых артефактах в ЛО-случаях, подчеркивают ключевую роль качества данных для надежности решений ИИ. Значимые корреляции ряда метрик и различия в RGB-каналах разных классов изображений подтверждают влияние освещения, текстуры и размытости на заключения моделей ИИ. Установленные закономерности и предложенные на их основе рекомендации, а также подтвержденный практический эффект дообучения моделей создают методическую базу для дальнейшего совершенствования программ компьютерного зрения при работе с немеланоцитарными новообразованиями кожи.

3.9. Анализ способов решения проблем диагностики новообразований кожи с помощью программ искусственного интеллекта на примере программы «Derma Check»

Целью настоящего подраздела является демонстрация конкретных решений проблем, возникающих при диагностике новообразований кожи с помощью программ ИИ, на примере программы «Derma Onko Melanoma Check» – мультимодельной системы поддержки принятия врачебных решений для предварительной диагностики и маршрутизации пациентов с новообразованиями кожи (свидетельство о государственной регистрации программы для ЭВМ № 2025661515 от 22.05.2025). Проблемы, выделенные в предыдущих разделах настоящей главы на основании анализа причин ошибочной классификации меланоцитарных и немеланоцитарных опухолей, в настоящем подразделе рассматриваются уже в прикладном ключе – с точки зрения того, какие архитектурные и интерфейсные решения позволяют их преодолеть.

3.9.1. Программа «Derma Check»

Программа «Derma Onko Melanoma Check» представляет собой систему поддержки принятия врачебных решений для предварительной диагностики и

маршрутизации пациентов с новообразованиями кожи. Ее архитектура построена на ансамбле сверточных нейронных сетей и трансформеров различных архитектур, объединенных по принципу голосования большинства. Программа формирует заключение о доброкачественном или злокачественном характере новообразования с определением вероятности в процентах, а также заключение о категории риска злокачественности (высокий, средний или низкий) с соответствующей рекомендацией по маршрутизации пациента для дальнейшего обследования. Программа поддерживает анализ как клинических фотоизображений, полученных камерой смартфона, иных карманных персональных компьютеров (КПК) или любых цифровых фотоаппаратов, так и дерматоскопических изображений, полученных с помощью дерматоскопа.

3.9.2. Дизайн исследования и методология анализа

Проведен систематический анализ опыта клинического применения программ ИИ для диагностики ЗНО кожи. На основании анализа типичных ошибок и ограничений, выявленных в ходе практического применения, сформулированы шесть основных групп проблем. Для каждой группы описаны решения, реализованные в программе «Derma Onko Melanoma Check», с иллюстрацией конкретными клиническими случаями и скриншотами интерфейса программы. Клинические случаи, представленные в разделе результатов, использованы для иллюстрации, во-первых, разницы в заключениях одномодельной и мультимодельной версий программы и, во-вторых, работы модуля объяснимого ИИ. Патоморфологическая верификация диагнозов выполнялась методом цитологического и гистологического исследования биоптатов.

3.9.3. Выявленные проблемы диагностики новообразований кожи при применении программ искусственного интеллекта и их решения

Дисбаланс обучающих наборов данных

Подавляющее большинство публично доступных наборов данных для обучения моделей ИИ в дерматоонкологии (ISIC, HAM10000, PH2 и другие) состоит преимущественно из дерматоскопических изображений, полученных с помощью специализированного оборудования в условиях дерматологических центров. Клинические фотографии новообразований кожи, то есть изображения, полученные камерой смартфона, представлены в значительно меньшем количестве. Это приводит к существенному дисбалансу: модели ИИ, обученные преимущественно на дерматоскопических изображениях, демонстрируют более высокую точность при анализе дерматоскопических фотоизображений и заметно более низкую при анализе клинических фотоизображений.

Данная проблема особенно актуальна для практического здравоохранения, поскольку врачи общей практики (ВОП) и терапевты, являющиеся основными потенциальными пользователями программ ИИ для первичной диагностики ЗНО кожи, как правило, не располагают дерматоскопами и используют для фотографирования новообразований камеры смартфонов и других КПК. Решением данной проблемы является целенаправленный сбор собственных наборов клинических фотоизображений новообразований кожи, полученных в условиях реальной клинической практики, и включение их в обучающие выборки наряду с дерматоскопическими изображениями.

Низкое качество фотоизображений, получаемых врачами

При использовании программ ИИ в реальной клинической практике одной из наиболее частых причин снижения диагностической точности является низкое качество входных фотоизображений новообразований. Анализ клинических

фотографий, загружаемых врачами в систему, выявил следующие типичные дефекты: пересвет, то есть избыточную освещенность, приводящую к потере детализации поверхностной структуры образования; размытие, то есть отсутствие резкости вследствие движения камеры или недостаточной фокусировки; затемнение, то есть недостаточную освещенность, приводящую к потере цветовых нюансов; а также избыточную яркость и блики. При пересвеченном изображении модель ИИ теряет возможность анализировать ключевые визуальные морфологические признаки новообразования, в том числе неоднородность пигментации, сосудистый рисунок и границы образования, что может приводить как к ложноотрицательным, так и к ложноположительным заключениям.

Решением данной проблемы является внедрение автоматической системы контроля качества фотоизображений, которая оценивает параметры входного фотоизображения (резкость, яркость, наличие пересвета) и блокирует анализ фотоизображений, не соответствующих минимальным требованиям качества для достаточной обработки моделями ИИ, с выдачей рекомендации по корректировке условий съемки (Рисунок 57).

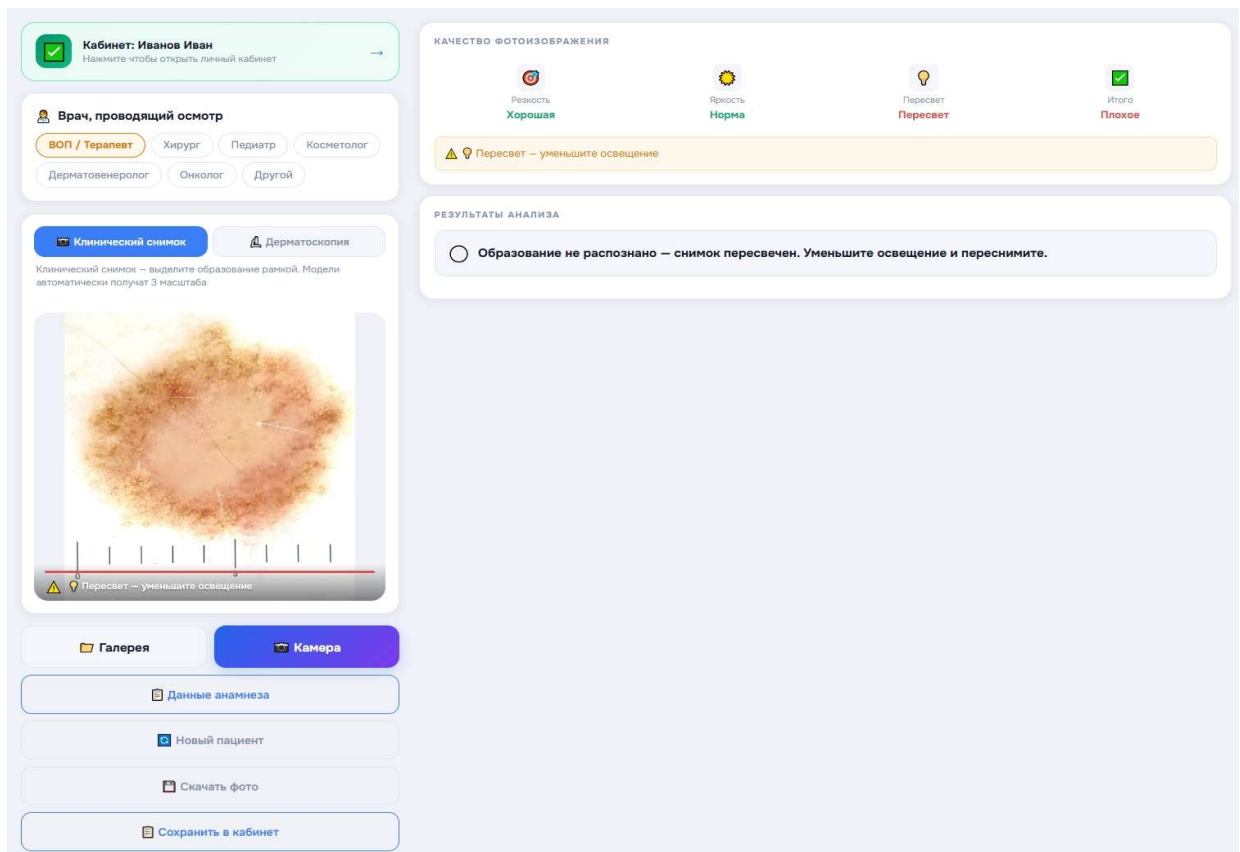


Рисунок 57 – Пример автоматической блокировки пересвеченного изображения программой «Derma Onko Melanoma Check». Система оценки качества фотоизображения выявила пересвет, итоговое качество оценено как «плохое», анализ образования заблокирован с рекомендацией уменьшить освещение и провести повторную фотофиксацию новообразования

Использование фотоизображений с чрезмерно малым размером новообразования в кадре

Еще одной распространенной проблемой является загрузка врачами фотоизображений, на которых новообразование занимает слишком малую долю кадра (менее $1/15$ – $1/25$, то есть менее 7–4 % площади фотоизображения). При таком кадрировании модель ИИ не получает достаточного количества пикселей для анализа визуальных морфологических признаков образования, что существенно снижает точность классификации.

Решением данной проблемы является внедрение предварительной детекции размера образования в кадре до начала анализа. Программа оценивает долю кадра, занимаемую образованием, и при недостаточном масштабе выдает

предупреждение с рекомендацией повторить фотографирование образования более крупным планом (Рисунок 58).

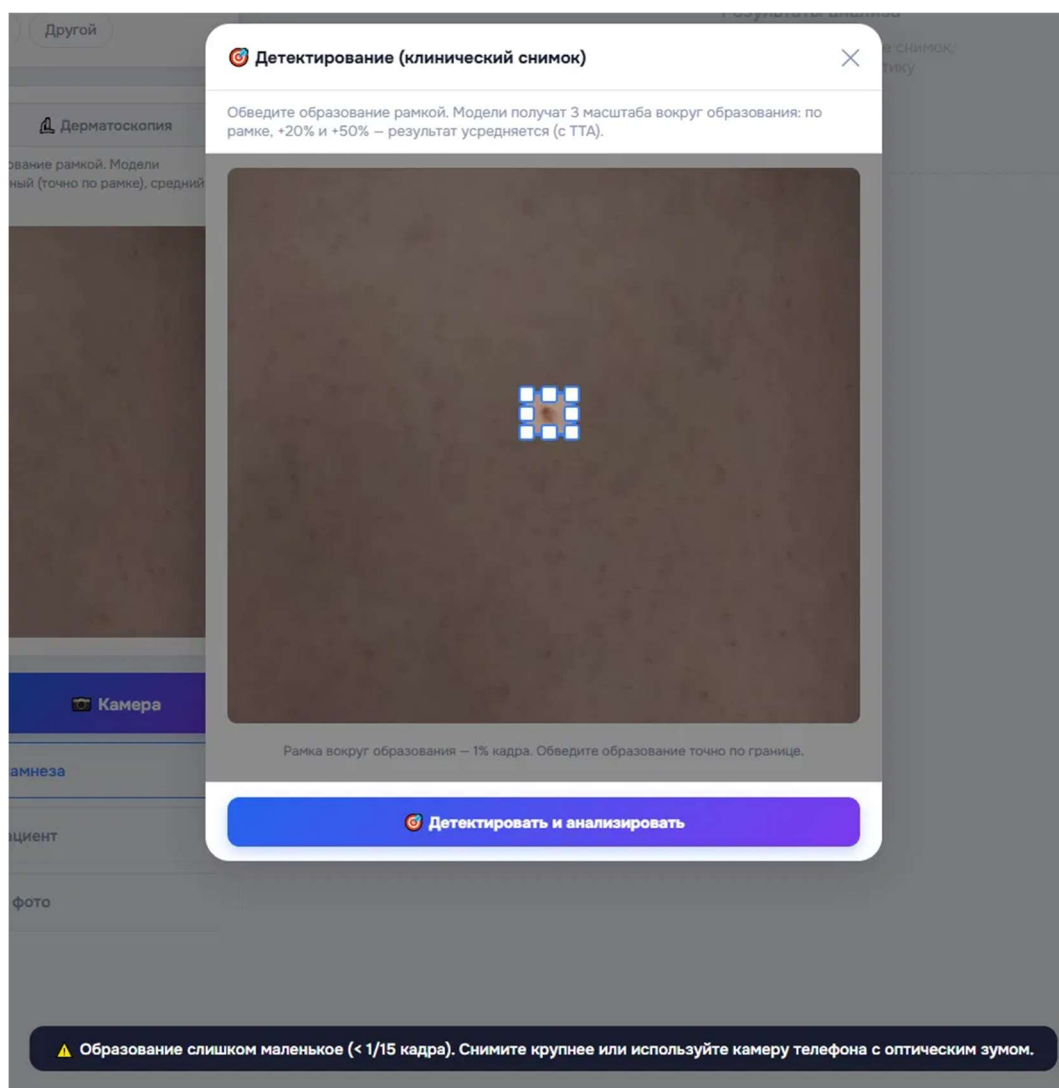


Рисунок 58 – Пример предупреждения программы «Derma Onko Melanoma Check» о недостаточном размере образования в кадре. Система детекции определила, что образование занимает менее 1/15 площади кадра, и выдала рекомендацию повторить фотографирование образования более крупным планом или использовать камеру с оптическим зумом

Одномодельная архитектура программ искусственного интеллекта и отсутствие онкологической настороженности

Подавляющее большинство существующих программ ИИ для диагностики новообразований кожи использует одномодельную архитектуру, при которой

единственная нейронная сеть формирует заключение о характере новообразования. Такой подход имеет ряд принципиальных ограничений. Во-первых, одна модель обучается на конкретной задаче (например, бинарной классификации «доброкачественное – злокачественное») и не учитывает пограничные состояния и предраковые поражения кожи, такие как актинический кератоз и диспластические невусы. Во-вторых, заключение одной модели лишено механизма самоконтроля: при ошибке единственной модели ошибочным оказывается и конечное заключение программы.

Данная проблема имеет прямые клинические последствия. Если программа ИИ расценивает пограничное новообразование как «доброкачественное», пациент не будет направлен на дополнительное обследование. Между тем в дерматоонкологической практике онкологическая настороженность по отношению к таким новообразованиям является нормой. Для иллюстрации приведем клинический случай.

Описание клинического случая для демонстрации ограничений одномодельной структуры программы

У пациентки имелось инфильтрированное пятно розоватого цвета размером около 1 см в диаметре в области левой щеки, без проросших сосудов. Образование существовало около года; периодически на его поверхности образовывалась корочка, которая со временем отторгалась. По клиническим и анамнестическим признакам проводился дифференциальный диагноз между актиническим кератозом и раком кожи. Пациентка была направлена на цитологическое и гистологическое исследование. По результатам исследования определялись клетки плоского эпителия с признаками дисплазии, выявлялись голые ядра. Заключение: актинический кератоз.

Данный случай наглядно демонстрирует ограничения одномодельной архитектуры. Ранняя одномодельная версия программы идентифицировала образование как «доброкачественное» (78,7 %) и «не меланома» (70,4 %); при

таком заключении пациент не был бы направлен на дополнительное обследование (Рисунок 59). Напротив, мультимодельная версия программы «Derma Onko Melanoma Check» по тому же клиническому случаю установила «средний риск» с рекомендацией направления к дерматовенерологу для дифференциальной диагностики, то есть проявила онкологическую настороженность, совпадающую с клиническим решением врача-эксперта (Рисунок 60).

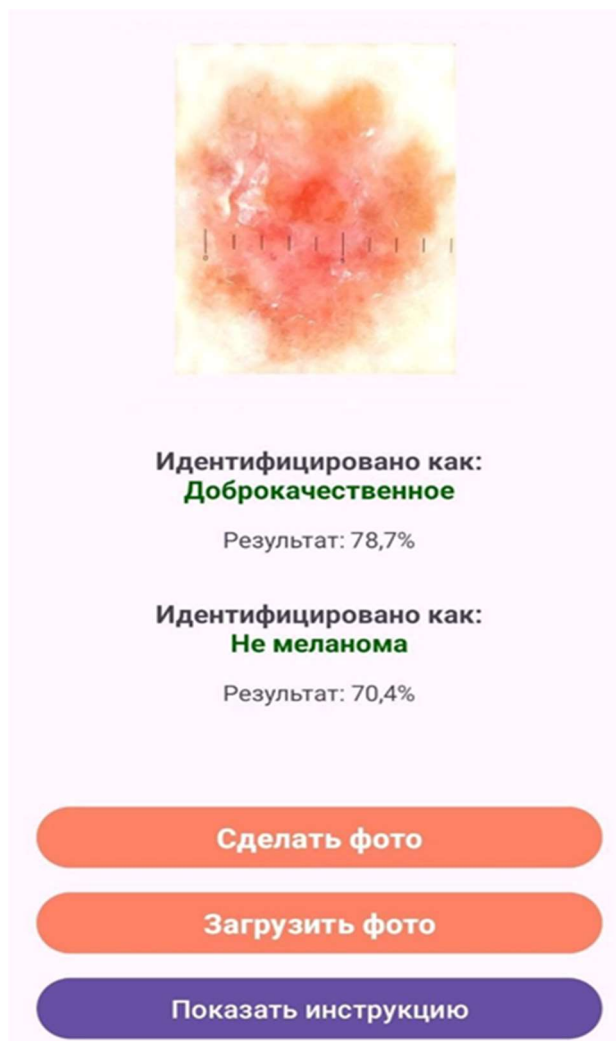


Рисунок 59 – Заключение ранней одномодельной версии программы: образование идентифицировано как «доброкачественное» (78,7 %) и «не меланома» (70,4 %).

При таком заключении пациент не был бы направлен на дополнительное обследование

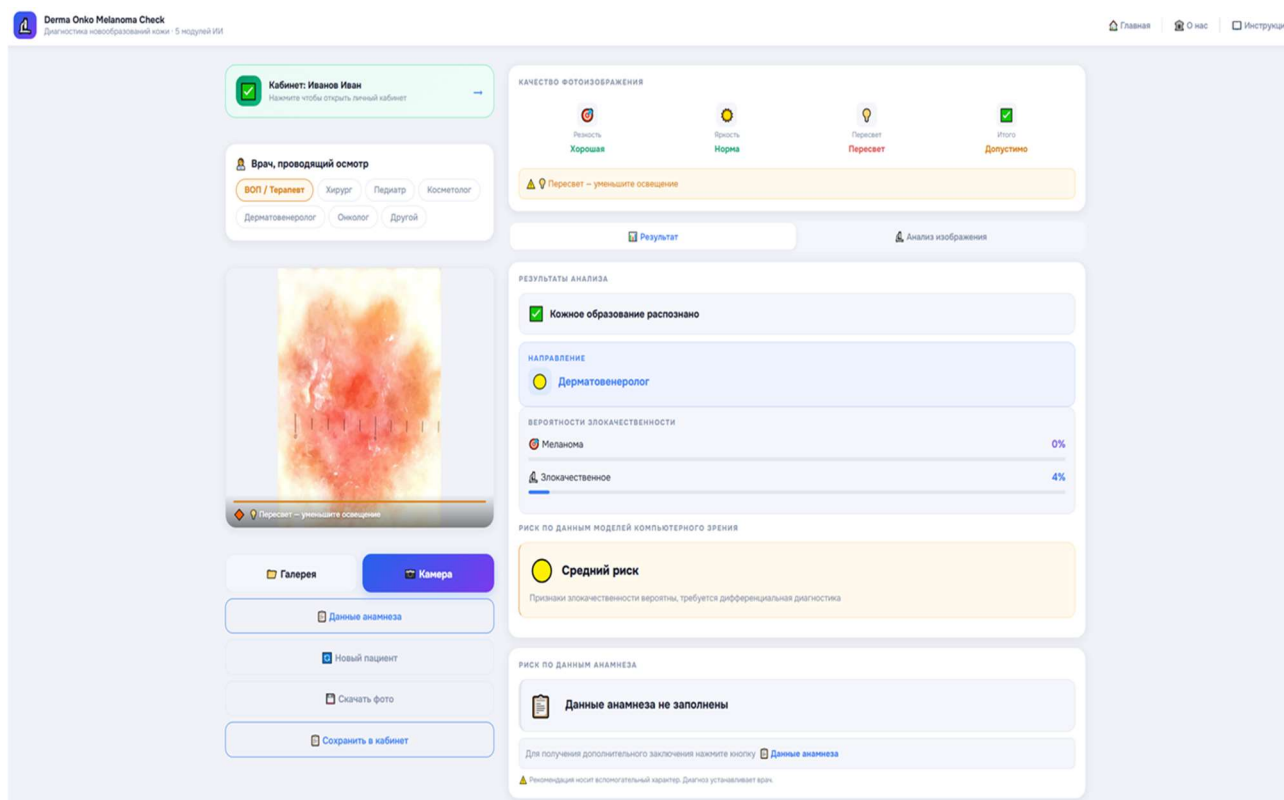


Рисунок 60 – Заключение мультимодельной версии программы «Derma Onko Melanoma Check» по тому же клиническому случаю: установлен «средний риск» с рекомендацией направления к дерматовенерологу для дифференциальной диагностики. Программа проявила онкологическую настороженность, совпадающую с клиническим решением врача-эксперта

Решением данной проблемы является создание мультимодельных программ ИИ, построенных на ансамбле нескольких нейронных сетей различных архитектур (сверточных нейронных сетей, трансформеров и других), объединенных по принципу голосования большинства. Такой подход повышает устойчивость заключений и позволяет программе проявлять онкологическую настороженность по отношению к пограничным новообразованиям.

Отсутствие объяснимого искусственного интеллекта

Большинство программ ИИ для диагностики новообразований кожи функционирует по принципу «черного ящика»: они выдают заключение (например, «злокачественное» или «доброкачественное»), но не предоставляют его обоснования. Врач не может определить, на какие визуальные морфологические

признаки опирается модель, что создает проблему доверия к заключениям ИИ и затрудняет их клиническое использование.

Решением данной проблемы является внедрение методов объяснимого ИИ (Explainable AI, XAI), которые обеспечивают визуализацию областей фотоизображения, наиболее значимых для формирования заключения (карту риска) (Рисунок 61), а также текстовый вывод наиболее значимых признаков, обнаруженных на изображении (например, «валикообразный край», «куполообразный узел», «бородавчатая поверхность», «фестончатые края») (Рисунок 62). Признаки, ассоциированные с повышенным риском злокачественности, выделяются цветом, что позволяет врачу верифицировать соответствие выявленных программой признаков его собственной клинической оценке.

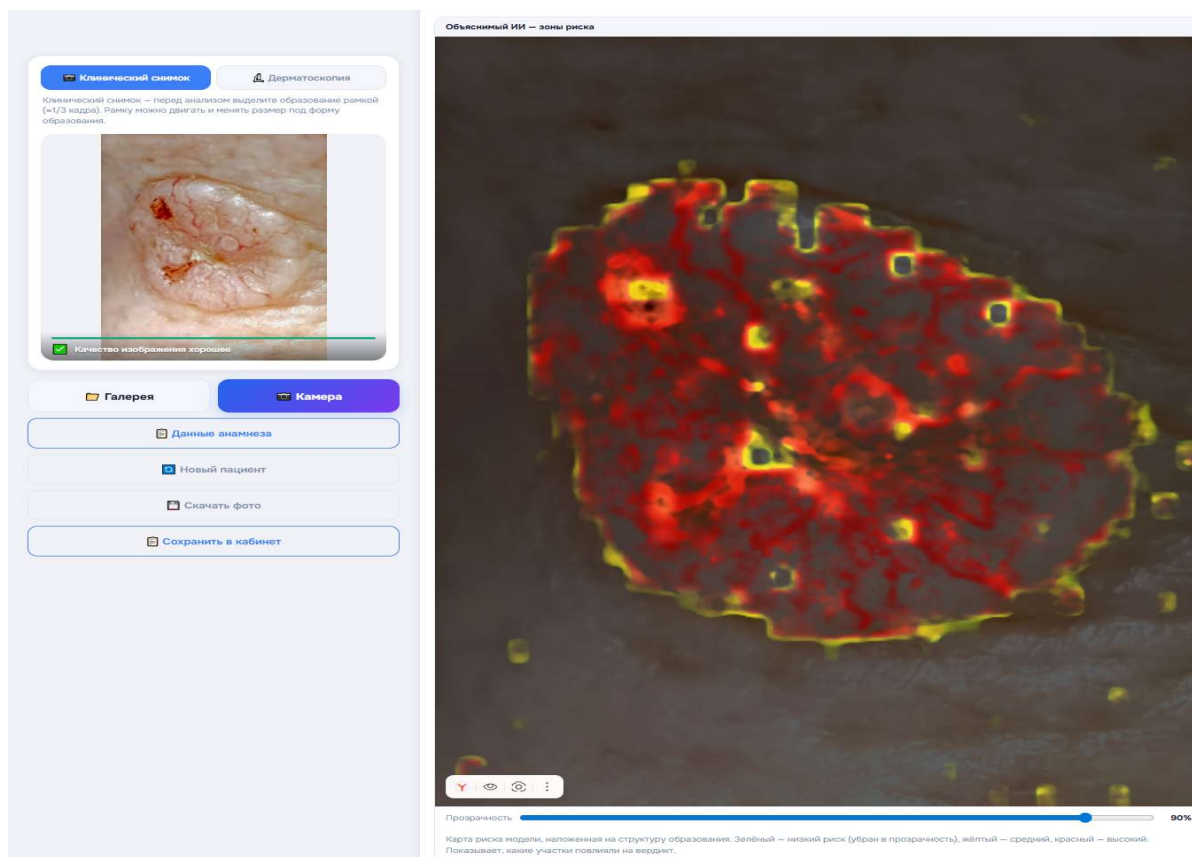


Рисунок 61 – Визуализация зон риска программой «Derma Onko Melanoma Check» с помощью объяснимого ИИ. Слева – клиническое фотоизображение новообразования кожи (базальноклеточный рак). Справа – карта риска модели, наложенная на структуру образования: красный цвет соответствует высокому риску, желтый – среднему риску

НАИБОЛЕЕ ВЕРОЯТНАЯ НОЗОЛОГИЯ

● Базальноклеточный рак

ДРУГИЕ ВОЗМОЖНЫЕ НОЗОЛОГИИ

2 ● Актинический кератоз

3 ● Бородавка/папиллома

4 ● Невус

5 ● Меланома

ПРИЗНАКИ НА ИЗОБРАЖЕНИИ (ТОП-5)

валикообразный край

куполообразный узел

бородавчатая поверхность

фестончатые края

Рисунок 62 – Текстовый вывод признаков, которые, по мнению модели, имеются на данном фотоизображении

Невозможность для врача влиять на заключение программы

Абсолютное большинство существующих программ ИИ для диагностики новообразований кожи функционирует исключительно на условии одномодальности вводимых данных: врач загружает фотоизображение новообразования, и программа формирует заключение. При этом у врача отсутствует возможность дополнить заключение клиническими данными, которые невозможно извлечь из фотоизображения, – сведениями из анамнеза (длительность существования новообразования, динамика изменений, наличие зуда или кровоточивости, семейный анамнез, наличие ЗНО в прошлом и другими) и результатами визуального осмотра. Между тем данные анамнеза являются важной составляющей клинического решения в дерматоонкологии. Так, быстрый рост образования, изменение его вида, кровоточивость, наличие ЗНО кожи в семейном анамнезе – все это факторы, существенно повышающие вероятность злокачественности и влияющие на решение о направлении на биопсию.

Решением данной проблемы является создание интерактивного интерфейса, позволяющего врачу вносить данные анамнеза (возраст, фототип кожи, диаметр образования, локализацию, основные и дополнительные симптомы, наличие ЗНО в анамнезе и другие) (Рисунок 63-64), а также дополнительные клинические признаки, выявленные при визуальном осмотре. Программа верифицирует

добавленные врачом клинические признаки, проверяет, подтверждает ли модель их наличие на фотоизображении, и при подтверждении корректирует итоговое заключение в сторону повышения риска злокачественности (Рисунок 65).

Данные анамнеза

* Обязательные | * Важные | Необязательные

ОБЯЗАТЕЛЬНЫЕ ПОЛЯ

ВОЗРАСТ *

Введите возраст

ФОТОТИП КОЖИ (ФИЦПАТРИК) *

— не указан —

ДИАМЕТР ОБРАЗОВАНИЯ (ММ) *

Макс. Мин.

ЛОКАЛИЗАЦИЯ *

— не указана —

ОСНОВНЫЕ СИМПТОМЫ *

☒ Возвышается над кожей * Да Нет

☒ Росло / увеличивалось * Да Нет

☒ Зуд * Да Нет

Мало данных (3/8)

ВАЖНЫЕ ПОЛЯ

ПОЛ *

— не указан —

ДОПОЛНИТЕЛЬНЫЕ СИМПТОМЫ *

☒ Изменение вида Да Нет Н.д.

☒ Кровоточивость Да Нет Н.д.

Рисунок 63 – Форма ввода данных анамнеза программы «Derma Onko Melanoma Check» (обязательные и важные поля): возраст, фототип кожи по Фицпатрику, диаметр образования, локализация, основные симптомы (возвышается над кожей, росло или увеличивалось, зуд)

Росло / увеличивалось • Да Нет

Зуд • Да Нет

Мало данных (3/8)

ВАЖНЫЕ ПОЛЯ

пол • — не указан —

дополнительные симптомы •

Изменение вида Да Нет Н.д.

Кровоточивость Да Нет Н.д.

злокачественные опухоли кожи в анамнезе (пациент / родственники) • — не указано —

онкология в анамнезе (любая локализация) • — не указано —

контакт с пестицидами / химикатами • — не указано —

НЕОБЯЗАТЕЛЬНЫЕ ПОЛЯ

БОЛЕЗНЕННОСТЬ ПРИ ПАЛЬПАЦИИ

Болезненность Да Нет Н.д.

Сбросить Применить

Рисунок 64 – Форма ввода данных анамнеза программы «Derma Onko Melanoma Check» (важные и необязательные поля): дополнительные симптомы (изменение вида, кровоточивость), наличие злокачественных опухолей кожи в анамнезе (у пациента или родственников), онкология в анамнезе, контакт с пестицидами или химикатами, болезненность при пальпации

Высокий риск
Обнаружены признаки злокачественного новообразования

НАИБОЛЕЕ ВЕРОЯТНАЯ НОЗОЛОГИЯ

- Меланома

⚠️ Исключить меланому: консультация онколога, биопсия.

ДРУГИЕ ВОЗМОЖНЫЕ НОЗОЛОГИИ

- Актинический кератоз
- Невус
- Себорейный кератоз
- Сосудистое образование

ПРИЗНАКИ НА ИЗОБРАЖЕНИИ (ТОП-5)

пигментированное пятно **иссиня-чёрный цвет** образование на ножке бородавчатая поверхность **асимметрия**

ДОПОЛНИТЕЛЬНЫЕ КЛИНИЧЕСКИЕ ПРИЗНАКИ ПРИ ОСМОТРЕ

— выберите признак —

+ Добавить 🔄 Пересчитать

⚠️ **неравномерный цвет** **согласен** ✕

Модель подтверждает признаки — вердикт поднят.

Рисунок 65 – Форма для внесения визуальных признаков врачом для повышения точности заключения программы

Мультимодальный анализ клинических и дерматоскопических изображений

Для повышения диагностической точности программа «Derma Onko Melanoma Check» поддерживает анализ как клинических фотоизображений, полученных камерой смартфона (Рисунок 66), так и дерматоскопических изображений одного и того же новообразования, полученных с помощью дерматоскопа (Рисунок 67). Использование двух типов изображений позволяет модели ИИ получить более полную информацию о визуальных морфологических характеристиках образования, что приближает логику работы программы к полноценному клиническому мышлению врача.



Рисунок 66 – Клиническое фотоизображение новообразования кожи (голубой невус)

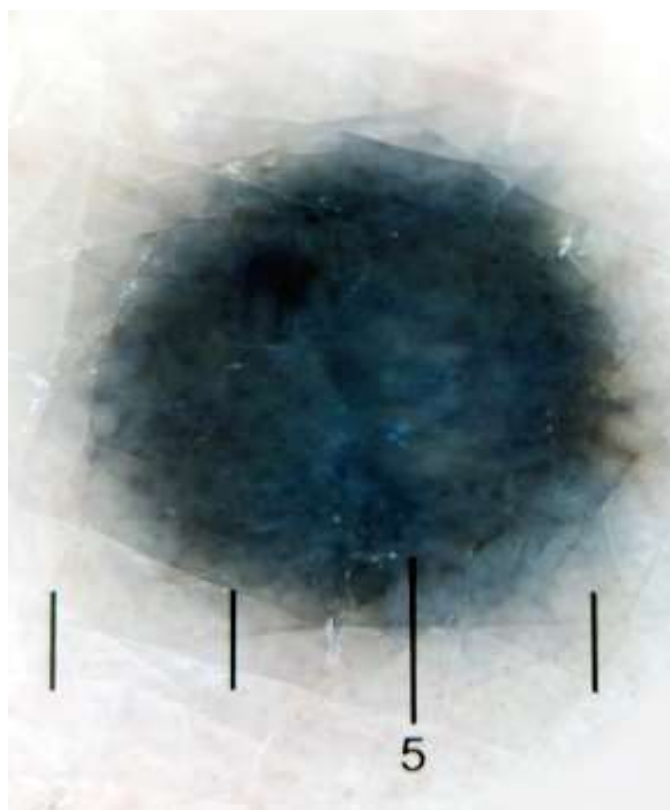


Рисунок 67 – Дерматоскопическое изображение того же новообразования кожи (голубой невус)

Проведенный анализ позволил выделить шесть принципиальных групп проблем, которые систематически снижают диагностическую эффективность программ ИИ при их переносе из контролируемых тестовых условий в реальную клиническую практику. Каждая из этих проблем имеет не только техническое, но и клиническое измерение.

Врачи общей практики, которые потенциально являются наиболее массовыми пользователями дерматоонкологических программ ИИ, работают с клиническими фотографиями смартфонов, тогда как большинство коммерческих и исследовательских моделей обучено преимущественно на дерматоскопических фотоизображениях. Целенаправленное формирование клинических обучающих наборов является необходимым условием практической применимости программы.

Проблемы качества входных изображений и масштаба образования в кадре в публикуемых исследованиях нередко остаются в тени, однако именно они определяют долю технически непригодных фотоизображений в реальной практике. Автоматический предварительный контроль качества снижает вероятность ошибочного заключения вследствие артефактов съемки. Эти выводы полностью согласуются с результатами анализа причин ошибочной классификации, представленными в разделах 5.2 и 5.3, где именно параметры яркости, энтропии, размытия и экспозиции были определены как ведущие предикторы ошибочных заключений.

Одномодельная архитектура с бинарной классификацией представляет особую опасность применительно к пограничным новообразованиям. Как показывает представленный клинический случай, формально правильное заключение («доброкачественное») оказывается клинически некорректным для актинического кератоза с дисплазией – предракового состояния, требующего активного ведения. Мультимодельный ансамбль с механизмом онкологической настороженности формирует более устойчивые заключения и лучше воспроизводит клиническое мышление опытного дерматовенеролога.

Отсутствие объяснимости представляет собой принципиальный барьер для внедрения ИИ в клиническую практику. «Черный ящик» не формирует доверия у врача и не позволяет верифицировать корректность заключения. Визуализация карты риска и текстовый вывод распознанных признаков обеспечивают прозрачность, необходимую для принятия клинических решений. Наконец, невозможность ввода анамнестических данных является принципиальным ограничением большинства ИИ-систем. Факторы риска, симптомы и история новообразования, которые врач учитывает при клиническом осмотре, не менее важны, чем визуальная картина. Интерактивный ввод анамнеза позволяет программе работать в логике полноценного клинического решения, а не только анализа изображения.

Проведенный анализ позволил систематизировать шесть основных групп проблем, возникающих при диагностике ЗНО кожи с помощью программ ИИ: 1) дисбаланс обучающих наборов данных с преобладанием дерматоскопических фотоизображений; 2) низкое качество фотоизображений, получаемых врачами; 3) загрузку фотографий с чрезмерно малым размером образования в кадре; 4) одномодельную архитектуру программ и отсутствие онкологической настороженности; 5) отсутствие объяснимого ИИ; 6) невозможность для врача влиять на заключение программы.

На примере программы «Derma Onko Melanoma Check» продемонстрированы конкретные решения каждой из выявленных проблем: 1) автоматический контроль качества фотоизображений с блокировкой некачественных фотоизображений; 2) контроль размера образования в кадре; 3) мультимодельная архитектура на основе ансамбля нейронных сетей с механизмом голосования; 4) визуализация зон риска и текстовый вывод распознанных признаков для обеспечения объяснимости заключений; 5) интерактивный интерфейс для ввода данных анамнеза и дополнительных клинических признаков; 6) поддержка мультимодального анализа клинических и дерматоскопических изображений.

Внедрение программ ИИ с описанными решениями в практику врачей-терапевтов и врачей общей практики первичного звена здравоохранения позволит

существенно повысить выявляемость ЗНО кожи на ранних стадиях, обеспечив при этом онкологическую настороженность по пограничным состояниям и клиническую обоснованность заключений. Таким образом, представленные в настоящей главе результаты образуют логически завершённую последовательность: от систематизации проблем дифференциальной диагностики новообразований кожи (раздел 5.1) через анализ причин ошибочной классификации меланоцитарных и немеланоцитарных опухолей программами ИИ (разделы 5.2 и 5.3) к обоснованию и практической реализации конкретных решений этих проблем в мультимодельной программе компьютерного зрения (раздел 5.4).

3.10. Разработка и апробация методологии оценки клинической точности и клинической эффективности диагностики новообразований кожи программами искусственного интеллекта

Внедрение программ искусственного интеллекта (ИИ) в дерматоонкологическую практику ставит перед исследователем не одну, а по меньшей мере три самостоятельные задачи оценки. Первая из них состоит в том, чтобы измерить, насколько точно программа распознает характер новообразования. Вторая – в том, чтобы определить, повышает ли внедрение ее заключений в маршрутизацию пациентов выявляемость злокачественных опухолей в реальных условиях многоэтапной диагностики. Третья – в том, чтобы установить, оправдано ли такое внедрение экономически, то есть приводит ли оно к снижению суммарных финансовых затрат системы здравоохранения.

3.10.1. Клиническая точность программ искусственного интеллекта с учетом онкологической настороженности

Оценка любой диагностической системы начинается с выбора того, что именно считать ее ошибкой. В отношении программ ИИ этот выбор долгое время казался очевидным: заключение сопоставляют с морфологически

верифицированным диагнозом по стандартной матрице ошибок, а любое расхождение относят к ложноположительным или ложноотрицательным результатам. Такой подход, заимствованный из практики оценки лабораторных и инструментальных тестов, регламентирован рядом национальных стандартов и по умолчанию принят в качестве универсального при валидации медицинских систем ИИ. Однако перенос его на дерматоонкологическую диагностику обнаруживает противоречие, которое и стало отправной точкой настоящего исследования.

Противоречие состоит в следующем. В дерматоонкологии существует обширная группа новообразований (актинический кератоз, диспластические невусы и ряд других пограничных и предраковых состояний), которые морфологически относятся к доброкачественным, но по клиническим и дерматоскопическим признакам требуют обязательного направления на морфологические методы исследования. Проявление настороженности по отношению к таким образованиям является не ошибкой, а нормой врачебной практики. Между тем стандартная матрица ошибок автоматически расценивает заключение программы «злокачественное» по морфологически доброкачественному образованию как ложноположительный результат. Иными словами, программа, поступающая клинически правильно, формально «наказывается» за верное решение. Систематическое занижение клинической ценности алгоритма, возникающее при таком подходе, имеет два практических следствия, искажается сравнение программ между собой и с врачами-экспертами, а при использовании формальных меток для машинного обучения модель в ходе дообучения обучается снижать собственную онконастороженность, то есть ухудшается по наиболее значимому клиническому параметру.

Чтобы преодолеть это противоречие, в настоящем разделе онконастороженность рассматривается не как источник ошибок, а как отдельный измеряемый результат работы программы. Для этого разработана и клинически апробирована трехуровневая система оценки, в которой формальная («техническая») точность дополняется двумя новыми показателями, клинической эффективностью и клинической точностью диагностики. Целью раздела является

разработка методологии оценки клинической эффективности и клинической точности диагностических программ ИИ в дерматоонкологии, учитывающей онкологическую настороженность как самостоятельную клиническую ценность.

3.10.2. Материал и программа искусственного интеллекта

Материалом послужили 342 новообразования кожи у пациентов в возрасте от 19 до 84 лет, обследованных в ФГБУ «Главный военный клинический госпиталь им. академика Н.Н. Бурденко» Минобороны России в период с марта 2025 г. по май 2026 г. Предварительный и заключительный клинические диагнозы устанавливались врачами-дерматовенерологами с опытом работы в дерматоонкологии более 20 лет на основании анамнестических, клинических и дерматоскопических данных в соответствии с классификацией опухолей кожи ВОЗ 2018 г. [373]. При наличии клинических показаний выполнялась биопсия с морфологическим исследованием биоптатов.

Структура выборки отражала реальное соотношение нозологических форм при первичном приеме. Злокачественные новообразования (ЗНО) были представлены базальноклеточным раком (61 случай, 17,8 %) и меланомой кожи (37 случаев, 10,8 %). Доброкачественные новообразования (ДНО) включали актинический кератоз (18; 5,3 %), диспластический невус (22; 6,4 %), себорейный кератоз (28; 8,2 %), приобретенный дермальный невус (72; 21,1 %), голубой невус (15; 4,4 %), приобретенный пограничный невус (44; 12,8 %) и приобретенный смешанный невус (45; 13,2 %). Прижизненное гистологическое исследование выполнено в 142 случаях, во всех 98 случаях ЗНО и у 44 пациентов с ДНО, направленных на биопсию по клиническим и дерматоскопическим показаниям. В остальных 200 случаях заключительный диагноз ДНО устанавливался по характерным клиническим и дерматоскопическим признакам без гистологической верификации, что соответствует действующим клиническим рекомендациям и не требует направления на биопсию. Данное обстоятельство принципиально: ограничение анализа только гистологически подтвержденными случаями исказило

бы репрезентативность выборки и фактически исключило бы из расчета показатель специфичности, который формируется преимущественно за счет ДНО без направления на биопсию.

Оценивалась программа Derma Onko Melanoma Check New – мультимодельная система поддержки принятия врачебных решений, реализованная в виде офлайн-приложения для устройств под управлением Android (свидетельство о государственной регистрации программы для ЭВМ № 2026662061 от 25.04.2026). Ее архитектура построена на ансамбле сверточных нейронных сетей и трансформеров, объединенных по принципу голосования большинства при совпадении заключений двух моделей из трех. Программа определяет характер новообразования (доброкачественное или злокачественное) с указанием вероятности злокачественности в диапазоне 50–100 %, относит злокачественные образования к меланомам либо иным ЗНО, а также устанавливает категорию риска (высокий, средний или низкий) с соответствующей рекомендацией по маршрутизации пациента.

3.10.3. Первый уровень оценки. Формальная точность и ее систематическая ошибка

Формальная («техническая») точность рассчитывается по стандартной матрице ошибок путем бинарного сопоставления заключения программы с верифицированным диагнозом. Каждый случай относится к одной из четырех категорий, содержательное наполнение которых применительно к оцениваемой программе приведено в Таблице 33. Эталоном истины служит заключительный клинический диагноз, установленный в объеме, необходимом и достаточном по клиническим рекомендациям и протоколам.

Таблица 33 – Элементы стандартной матрицы ошибок применительно к заключениям программы искусственного интеллекта Derma Onko Melanoma Check New

Показатель	Описание применительно к программе
ИП	Программа правильно идентифицировала новообразование как ЗНО, установила высокий или средний риск; морфологически подтверждено ЗНО
ИО	Программа правильно идентифицировала новообразование как ДНО, установила низкий риск; клинически и/или морфологически подтверждено ДНО
ЛП	Программа ошибочно идентифицировала новообразование как ЗНО, установила высокий или средний риск; клинически и/или морфологически подтверждено ДНО
ЛО	Программа ошибочно идентифицировала новообразование как ДНО, установила низкий риск; клинически и/или морфологически подтверждено ЗНО
Примечание: ИП – истинно положительные; ИО – истинно отрицательные; ЛП – ложноположительные; ЛО – ложноотрицательные результаты; ЗНО – злокачественные новообразования; ДНО – доброкачественные новообразования	

Показатели рассчитывались по общепринятым формулам: точность $Ac = (ИП + ИО) / N \times 100 \%$, чувствительность $Se = ИП / (ИП + ЛО) \times 100 \%$, специфичность $Sp = ИО / (ИО + ЛП) \times 100 \%$. На выборке из 342 новообразований программа правильно распознала все 98 случаев ЗНО (ИП = 98, ЛО = 0), при этом ИО = 209, ЛП = 35. Соответственно формальная точность составила 89,7 %, чувствительность – 100,0 %, специфичность – 84,8 %.

На первый взгляд эти значения полностью характеризуют работу программы. Однако детальный разбор 35 случаев, отнесенных к ложноположительным, обнаружил, что подавляющее их большинство представляло собой клинически обоснованные направления на биопсию, по которым и врачи-дерматовенерологи независимо проявляли онконастороженность. Отнесение таких наблюдений к ошибкам методологически некорректно, программа в них поступила так же, как поступил бы опытный специалист. Именно этот разрыв между формальной меткой и клинической реальностью и потребовал введения двух дополнительных уровней оценки, к которым мы переходим далее.

3.10.4. Второй уровень оценки: клиническая эффективность как качественная мера согласованности

Клиническая эффективность диагностики характеризует степень согласованности заключений программы с предварительным и заключительным диагнозом врача-эксперта и оценивается по четырем качественным категориям: полное совпадение, совпадение по онконастороженности, несовпадение по онконастороженности и полное несовпадение. Принципиальное отличие этого подхода от формального состоит в том, что онконастороженность признается здесь самостоятельной ценностью и не расценивается как ошибка даже тогда, когда образование морфологически оказалось доброкачественным. Содержательные критерии отнесения к каждой категории и их клиническое значение представлены в Таблице 34.

Таблица 34 – Характеристики показателей клинической эффективности диагностики применительно к программе Derma Onko Melanoma Check New

Категория	Критерий отнесения применительно к программе	Клиническое значение
Полное совпадение	Программа дает заключение о ДНО, устанавливает НР и отсутствие показаний к дообследованию, а врач-эксперт устанавливает диагноз ДНО по данным осмотра и дерматоскопии; ИЛИ программа дает заключение о ЗНО, устанавливает ВР или СР и направляет на биопсию, а врач-эксперт также заподозрил ЗНО и направил на биопсию, морфологически подтвердившую ЗНО	Программа и врач-эксперт независимо пришли к одинаковому заключению: либо оба верно расценили образование как ДНО без направления на дообследование, либо оба верно расценили его как ЗНО (подозрение на ЗНО) и направили на биопсию с последующим подтверждением злокачественности
Совпадение по онконастороженности	Программа устанавливает ВР или СР и направляет на биопсию; врач-эксперт также заподозрил ЗНО и направил на биопсию, но морфологическое исследование не подтвердило ЗНО	Программа и врач-эксперт независимо проявили онкологическую настороженность по клиническим и/или дерматоскопическим признакам

Продолжение Таблицы 34

Несовпадение по онконастороженности	Программа дает заключение о ДНО (низкой вероятности ЗНО) и устанавливает НР, однако врач-эксперт направляет на биопсию, которая не подтверждает злокачественность	Клинически значимая ошибка врача-эксперта: программа не проявила настороженности, врач проявил и направил на биопсию, установившую ДНО
Полное несовпадение	Программа дает заключение о ДНО и устанавливает НР, но врач-эксперт направляет на биопсию, подтверждающую ЗНО; ИЛИ программа устанавливает ВР или СР, но врач-эксперт по осмотру и дерматоскопии устанавливает ДНО	Клинически значимое расхождение: либо программа пропустила ЗНО, выявленное настороженностью врача, либо программа проявила настороженность, не поддержанную экспертом, при итоговом диагнозе ДНО
Примечание: ДНО – доброкачественное новообразование; ЗНО – злокачественное новообразование; ВР – высокий риск; СР – средний риск; НР – низкий риск злокачественности. Категории ВР и СР объединены как однозначно требующие онкологической настороженности		

Распределение случаев по этим категориям в разрезе нозологических форм приведено в Таблице 35. Общая картина такова: полное совпадение заключений программы с предварительным диагнозом врача зафиксировано в 281 случае (82,2 %), совпадение по онконастороженности – в 44 случаях (12,9 %), несовпадение по онконастороженности – в 4 случаях (1,2 %) и полное несовпадение – в 13 случаях (3,8 %).

Таблица 35 – Результаты оценки клинической эффективности по четырем категориям совпадения и несовпадения (n=342)

Диагноз	Число случаев, n	Полное совпадение	Совпадение по онконастороженности	Несовпадение по онконастороженности	Полное несовпадение
		n (%)	n (%)	n (%)	n (%)
Базальноклеточный рак	61	61 (100,0)	0 (0,0)	0 (0,0)	0 (0,0)
Меланома	37	37 (100,0)	0 (0,0)	0 (0,0)	0 (0,0)
Кератоз актинический	18	3 (16,7)	14 (77,8)	1 (5,6)	0 (0,0)
Неvus диспластический	22	7 (31,8)	11 (50,0)	3 (13,6)	1 (4,5)
Кератоз себорейный	28	14 (50,0)	10 (35,7)	0 (0,0)	4 (14,3)

Продолжение Таблицы 35

Невус приобретенный дермальный	72	65 (90,3)	1 (1,4)	0 (0,0)	6 (8,3)
Невус голубой	15	11 (73,3)	2 (13,3)	0 (0,0)	2 (13,3)
Невус приобретенный пограничный	44	40 (90,9)	4 (9,1)	0 (0,0)	0 (0,0)
Невус приобретенный смешанный	45	43 (95,6)	2 (4,4)	0 (0,0)	0 (0,0)
Итого	342	281 (82,2)	44 (12,9)	4 (1,2)	13 (3,8)
Примечание: наибольшая доля совпадений по онконастороженности приходится на нозологии, требующие активного наблюдения, – актинический кератоз и диспластический невус					

Нозологический разрез показателя особенно нагляден. Во всех 98 случаях ЗНО (базальноклеточный рак и меланома) достигнуто полное совпадение – 100,0 %. Наибольшая доля совпадений по онконастороженности зарегистрирована при актиническом кератозе (полное совпадение 16,7 %, совпадение по онконастороженности 77,8 %), диспластическом невусе (31,8 и 50,0 % соответственно) и себорейном кератозе (50,0 и 35,7 %). Высокая доля совпадений по онконастороженности при первых двух нозологиях не является недостатком программы, поскольку оба диагноза требуют активной настороженности. Именно эти случаи при формальном подходе были бы неверно отнесены к ложноположительным. Таким образом, показатель клинической эффективности выполняет функцию, недоступную агрегированным формальным метрикам, он позволяет интерпретировать работу программы дифференцированно по нозологиям.

Для визуализации этих закономерностей использованы три взаимодополняющих представления. Столбчатая диаграмма (Рисунок 68) отражает структуру клинической эффективности по каждой нозологии с разбивкой на четыре категории. Форест-график (Рисунок 69) показывает точечные оценки суммарного клинического совпадения с 95 % доверительными интервалами, рассчитанными методом Уилсона: чем меньше нозологическая группа, тем шире интервал и тем менее устойчива оценка. Наконец, пузырьковая диаграмма (Рисунок 70) сопоставляет суммарную клиническую эффективность с частотой

полного несовпадения, помещая нозологии с наилучшим профилем в зону высокой эффективности.

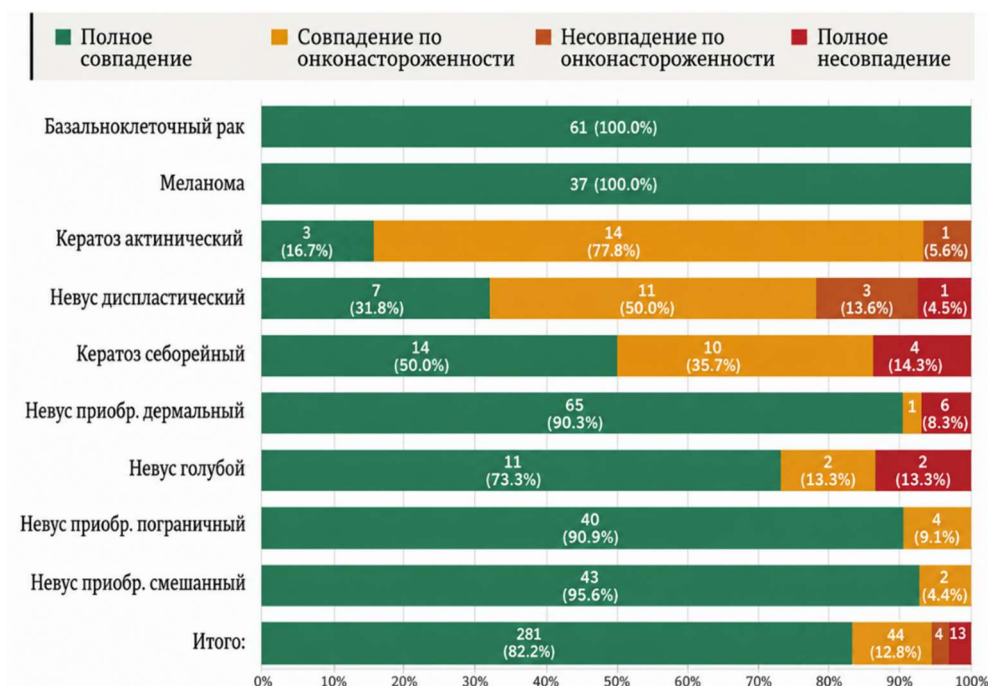


Рисунок 68 – Клиническая эффективность программы Derma Onko Melanoma Check New по нозологиям (n=342). Числа на сегментах – абсолютные значения и процентная доля

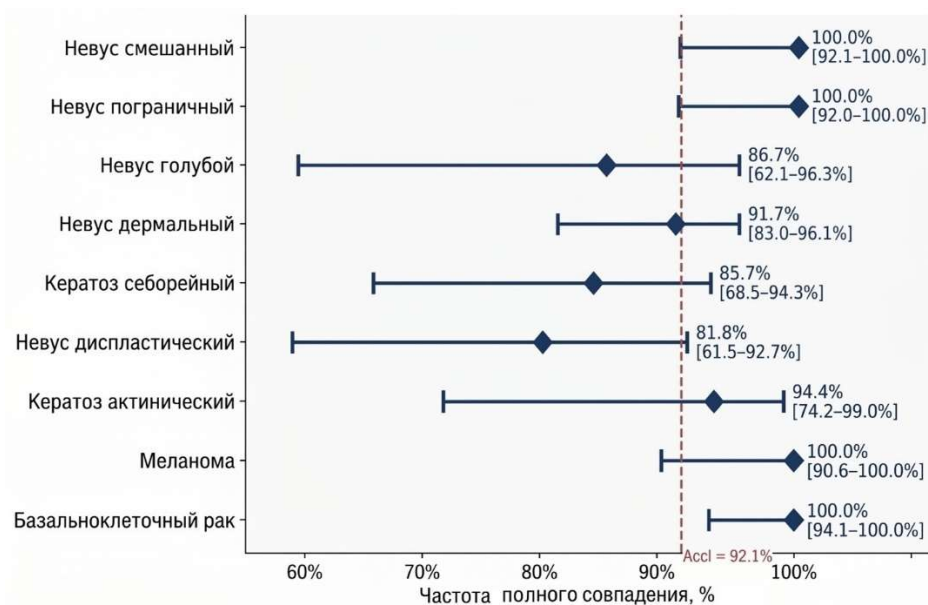


Рисунок 69 – Форест-график частоты клинического совпадения заключения программы с предварительным клиническим диагнозом врача-дерматовенеролога (полное совпадение и совпадение по онконастороженности) по нозологиям с 95 % доверительным интервалом (метод Уилсона, n=342). Ромбы – точечные оценки; горизонтальные «усы» – 95 % ДИ; пунктирная линия – общая клиническая точность

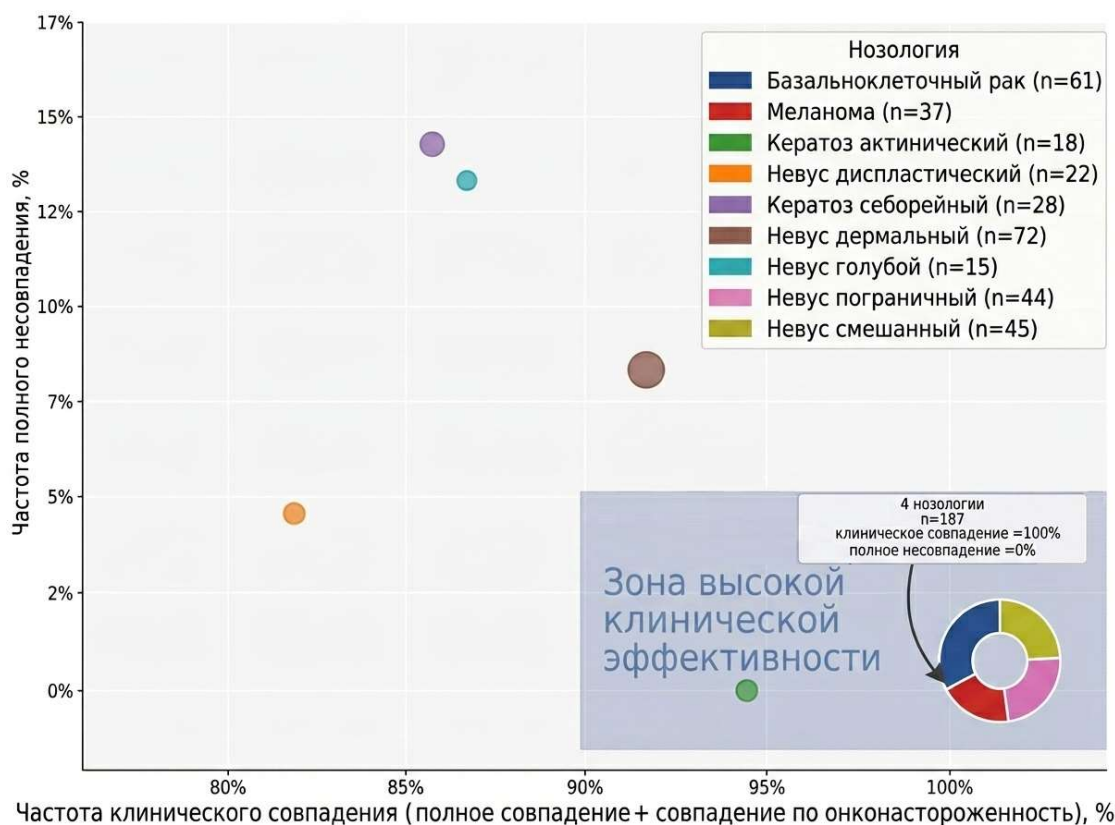


Рисунок 70 – Соотношение частоты клинической эффективности (полное совпадение и совпадение по онконастороженности, ось X) и частоты полного несовпадения (ось Y) по нозологиям (n=342). Размер пузырьков пропорционален числу наблюдений; зона высокой клинической эффективности – правый нижний угол

У показателя клинической эффективности есть, однако, принципиальное ограничение, он опирается на предварительный диагноз конкретного врача как на эталон, а значит, его информативность прямо зависит от квалификации этого специалиста. В настоящем исследовании эталон обеспечивался врачами с более чем двадцатилетним опытом, что повышает его надежность, но не устраняет зависимости от человеческого фактора в принципе. Для оценки, не привязанной к квалификации отдельного специалиста, необходим количественный показатель, клиническая точность.

3.10.5. Третий уровень оценки. Клиническая точность как количественная мера обоснованности

Клиническая точность диагностики оценивает не правильность распознавания морфологического типа, а обоснованность заключения программы о риске злокачественности для направления пациента на дообследование. Эталонном истины здесь служит двойной критерий, сочетание клинико-дерматоскопической экспертизы опытного врача и результатов морфологического исследования. Все случаи распределяются по четырем категориям матрицы клинических случаев, определения которых приведены в Таблице 36.

Таблица 36 – Матрица клинических случаев для оценки клинической точности диагностики программой Derma Onko Melanoma Check New

Обозначение	Пояснение при использовании программы
JS – обоснованное заключение (justified conclusion)	Заключения программы «высокий риск» или «средний риск» и «ЗНО»/«подозрение на ЗНО» совпали с предварительным и заключительным диагнозом и/или с морфологическим подтверждением, что подтвердило необходимость направления на биопсию. Заключение программы о высоком или среднем риске было обосновано с учетом онконастороженности
UC – необоснованное заключение (unjustified conclusion)	Заключение программы «высокий» или «средний риск», но врач по клинико-дерматоскопическим признакам расценил образование как ДНО либо морфология установила ДНО без дисплазии, что подтвердило отсутствие показаний к биопсии. Заключение программы о риске было необоснованным
MJC – пропущенное обоснованное заключение (missed justified conclusion)	Заключение программы «низкий риск», но врач по клинико-дерматоскопическим признакам заподозрил ЗНО и морфология установила ЗНО, что подтвердило необходимость биопсии. Заключение врача было обоснованным, а программа неверно оценила риск
NMJC – непропущенное обоснованное заключение (not missed justified conclusion)	Заключение программы «низкий риск», и врач расценил образование как ДНО, либо врач заподозрил ЗНО, но морфология установила ДНО без дисплазии, подтвердив отсутствие показаний к биопсии. Заключение программы о низком риске было обоснованным

На основании этих категорий рассчитываются три показателя. Клиническая чувствительность отражает способность программы выявлять все случаи, требующие обоснованного направления на биопсию, и вычисляется как

$$Se_{cl} = JC / (JC + MJC) \times 100\%. \quad (1)$$

Клиническая специфичность отражает способность правильно идентифицировать случаи, по которым биопсия не требовалась, и равна

$$Sp_{cl} = NMJC / (NMJC + UC) \times 100\%. \quad (2)$$

Клиническая точность характеризует общую долю клинически верных заключений:

$$Ac_{cl} = (JC + NMJC) / N \times 100\%. \quad (3)$$

Распределение всех 342 наблюдений по матрице клинических случаев представлено в Таблице 37. Обоснованными оказались 130 направлений на биопсию (JC), необоснованными – 25 (UC), пропущенных обоснованных заключений было всего 2 (MJC), а непропущенных обоснованных – 185 (NMJC).

Таблица 37 – Матрица клинических случаев программы Derma Onko Melanoma Check New (n=342)

Заключение программы	Направление на биопсию клинически обосновано*	Направление на биопсию клинически необосновано*
Высокий или средний риск	JC = 130	UC = 25
Низкий риск	MJC = 2	NMJC = 185
Итого	132	210
Примечание: * по клиническим и/или дерматоскопическим признакам. JC – обоснованное заключение; UC – необоснованное; MJC – пропущенное обоснованное; NMJC – непропущенное обоснованное заключение		

Содержательная расшифровка категорий поясняет эти цифры. Из 130 обоснованных заключений (JC) 98 приходятся на ЗНО (61 базальноклеточный рак и 37 меланом), по которым программа установила высокий или средний риск и необходимость направления подтверждена морфологически, и еще 32 на ДНО, направленные на биопсию оправданно, с учетом онконастороженности, и подтвержденные гистологически. В 25 необоснованных заключениях (UC) программа установила риск злокачественности по образованиям, оказавшимся доброкачественными, 13 из них представляли собой полное несовпадение среди

ДНО, а 12 направления, по которым морфология не подтвердила обоснованности. Всего 2 наблюдения отнесены к пропущенным обоснованным заключениям (MJC) именно эта категория несет наибольший клинический риск. Наконец, 185 наблюдений (NMJC) отражают верное отнесение образований к низкому риску без показаний к биопсии.

Итоговые показатели клинической точности приведены в Таблице 38. Клиническая чувствительность составила 98,5 % (95 % ДИ 94,6–99,6) из 132 клинически обоснованных направлений программа верно установила 130. Клиническая специфичность 88,1 % (95 % ДИ 83,0–91,8) из 210 случаев без показаний к биопсии программа верно воздержалась от направления в 185. Клиническая точность в целом достигла 92,1 % (95 % ДИ 88,8–94,5). Принципиально, что ни в одном из 98 случаев ЗНО программа не установила низкий риск, что полностью исключает риск пропуска злокачественного процесса.

Таблица 38 – Показатели клинической точности, чувствительности и специфичности программы Derma Onko Melanoma Check New (n=342)

Показатель	Формула	Расчет	Значение, %	95 % ДИ (метод Уилсона)
Клиническая чувствительность	$JC / (JC + MJC) \times 100 \%$	130 / 132	98,5	94,6–99,6
Клиническая специфичность	$NMJC / (NMJC + UC) \times 100 \%$	185 / 210	88,1	83,0–91,8
Клиническая точность	$(JC + NMJC) / N \times 100 \%$	315 / 342	92,1	88,8–94,5
Примечание: JC – обоснованное заключение; UC – необоснованное; MJC – пропущенное обоснованное; NMJC – непропущенное обоснованное заключение				

Анализ 25 необоснованных заключений (UC) показал, что они преимущественно относятся к новообразованиям с атипичными морфологическими характеристиками неравномерной пигментацией, папилломатозной поверхностью, множественными цветовыми зонами. Именно эти признаки алгоритм, обученный на паттернах злокачественности, закономерно интерпретирует как подозрительные. Снижение числа UC возможно за счет дообучения на конкретных морфологических паттернах, однако при этом критически важно не жертвовать онконастороженностью в противном случае

улучшение специфичности будет достигнуто ценой ухудшения главного клинического качества программы.

3.10.6. Сопоставление трех уровней оценки

Прямое сравнение формального и клинического подходов (Таблица 39) обнажает механизм, ради которого и разрабатывалась предложенная методология. Разница в общей точности невелика – 92,1 % против 89,7 %, всего 2,4 процентного пункта. Однако за этой разницей стоит содержательное переопределение 44 клинических случаев: из категории «ошибки» они переходят в категорию обоснованных заключений. При формальном расчете эти наблюдения снижают специфичность с 88,1 до 84,8 % и общую точность с 92,1 до 89,7 %. При клиническом расчете они распределяются между JC и UC по содержательному критерию в зависимости от того, была ли биопсия оправдана по результатам морфологического исследования. Именно это переопределение и составляет ключевое обоснование методологии.

Таблица 39 – Показатели формальной («технической») и клинической точности диагностики программой Derma Onko Melanoma Check New

Метод оценки	ИП / JC, n	ИО / NMJC, n	ЛП / UC, n	ЛО / MJC, n	Ac, %	Se, %	Sp, %
Формальная точность	98	209	35	0	89,7	100,0	84,8
Клиническая точность	130	185	25	2	92,1	98,5	88,1
Примечание: ИП – истинно положительные; ИО – истинно отрицательные; ЛП – ложноположительные; ЛО – ложноотрицательные; JC – обоснованное заключение; NMJC – непропущенное обоснованное; UC – необоснованное; MJC – пропущенное обоснованное заключение							

3.10.7. Надежность эталона. Межэкспертная согласованность

Поскольку и клиническая эффективность, и клиническая точность опираются на экспертную оценку, надежность самого эталона потребовала отдельной проверки. Каждое из 342 новообразований независимо оценивалось тремя врачами-дерматовенерологами со стажем работы в дерматоонкологии более 20 лет;

итоговое решение принималось как консенсус большинства (не менее двух из трех). Показатели индивидуальной и консенсусной диагностики приведены в Таблице 40.

Таблица 40 – Показатели точности диагностики трех врачей-экспертов и их консенсусного заключения (n=342)

Эксперт	ИП	ЛО	ИО	ЛП	Ас, %	Se, %	Sp, %
Врач-эксперт 1	98	0	244	0	100,0	100,0	100,0
Врач-эксперт 2	97	1	241	3	98,8	99,0	98,8
Врач-эксперт 3	98	0	239	5	98,5	100,0	98,0
Консенсус (≥2 из 3)	98	0	241	3	99,1	100,0	98,8
Примечание: ИП – истинно положительные; ЛО – ложноотрицательные; ИО – истинно отрицательные; ЛП – ложноположительные; Ас – точность; Se – чувствительность; Sp – специфичность. Единственное ЛО-заключение врача 2 нивелировано консенсусом врачей 1 и 3							

Двое из трех специалистов не допустили ни одного ложноотрицательного заключения, у третьего был зафиксирован единственный такой случай, нивелированный консенсусом. Расхождения касались исключительно зоны ДНО и выражались в разном числе ложноположительных заключений (0, 3 и 5). Согласованность оценивалась по коэффициенту каппа Коэна для каждой пары специалистов (Таблица 41); все значения превысили 0,96, что по шкале Ландиса–Коха соответствует «почти идеальному согласованию». Поскольку все расхождения носили односторонний характер, критерий Мак-Немара подтвердил отсутствие статистически значимой асимметрии ($p > 0,05$). Полное совпадение заключений всех трех врачей достигнуто в 336 случаях (98,2 %).

Таблица 41 – Межэкспертная согласованность предварительных клинических диагнозов врачей-дерматовенерологов (n=342)

Пара экспертов	Р _о	Р _е	κ Коэна	Мак-Немара
Врач 1 – Врач 2	0,988	0,589	0,972	1 / 3, $p > 0,05$
Врач 1 – Врач 3	0,985	0,585	0,965	0 / 5, $p > 0,05$
Врач 2 – Врач 3	0,991	0,583	0,979	0 / 3, $p > 0,05$
Примечание: Р _о – наблюдаемая доля совпадений; Р _е – ожидаемая доля совпадений при случайном согласии; κ – коэффициент каппа Коэна. Интерпретация κ по шкале Ландиса–Коха: $> 0,80$ – «почти идеальное согласование»				

Сопоставление заключений программы с консенсусом врачей (Таблица 42) очерчивает ее реальное место в диагностическом процессе. По клинической чувствительности программа (98,5 %) уступает консенсусу (100,0 %) лишь за счет двух случаев МЖС. По специфичности разрыв заметнее 88,1 против 98,8 %, и он отражает не слабость алгоритма, а его более широкую онконастороженность по сравнению с опытными врачами. Иными словами, программа склонна перестраховываться чаще, чем эксперт, что для инструмента первичного звена является скорее достоинством, чем недостатком.

Таблица 42 – Сравнение клинической точности программы ИИ и консенсусного заключения врачей-дерматовенерологов (n=342)

Метод	Ac / Ac _{cl} , %	Se / Se _{cl} , %	Sp / Sp _{cl} , %
Программа ИИ (клиническая точность)	92,1	98,5	88,1
Консенсус врачей-экспертов (формальная точность)	99,1	100,0	98,8
Примечание: для программы приведены показатели клинической точности, учитывающие онконастороженность; для консенсуса врачей – показатели формальной точности относительно морфологического диагноза			

3.10.8. Значение для клинической практики и для обучения моделей

Практическая ценность разработанной методологии раскрывается в двух плоскостях. Первая касается корректной оценки уже существующих программ. Публикуемые показатели точности, рассчитанные по стандартной методологии, систематически занижают клиническую ценность алгоритмов там, где верно проявляют онконастороженность, и тем самым искажают их сравнение между собой и с врачами-экспертами. Совокупная клиническая эффективность программы (полное совпадение и совпадение по онконастороженности) составила 95,0 %, причем 12,9 % пришлось именно на совпадение по онконастороженности, ситуации, в которых программа и врач независимо приняли клинически значимое решение о дообследовании, несмотря на последующий доброкачественный морфологический результат. При формальном подходе все эти случаи были бы отнесены к ошибкам.

Вторая плоскость обучение и совершенствование моделей. Если при разметке обучающих данных случаи обоснованной онконастороженности систематически помечаются как ложноположительные и алгоритм «штрафуется» за них, закономерным итогом оптимизации становится снижение онконастороженности модели. Применительно к дерматоонкологии это означает, что программа, улучшенная по формальным метрикам, будет реже выявлять предраковые и пограничные состояния, то есть деградирует по наиболее значимому клиническому параметру.

Значимость высокой клинической чувствительности особенно очевидна в контексте первичного звена. Программа же ни в одном из 98 случаев ЗНО не установила низкий риск, обеспечив тем самым 100 % выявление злокачественных процессов. Использование ее в качестве инструмента «второго мнения» с рекомендациями по маршрутизации способно радикально повысить выявляемость ЗНО там, где специализированная помощь наименее доступна.

В концептуальном отношении предложенный подход обнаруживает близость к ряду понятий клинической эпидемиологии, категория JS логически соответствует «клинически значимому положительному результату», а MJS – «клинически значимому ложноотрицательному результату». Отличие состоит в явном разделении клинической обоснованности направления, которая оценивается независимо, и морфологического подтверждения, которое используется как дополнительный, а не единственный верификатор. Именно принцип двойного эталона сочетание клинико-дерматоскопической экспертизы и морфологической верификации представляется наиболее адекватным для оценки инструментов, работающих на первичном этапе принятия решения о злокачественности.

Формальная точность, рассчитываемая по стандартной матрице ошибок, не отражает клиническую логику заключений программ ИИ в дерматоонкологии в полной мере, поскольку расценивает обоснованную онконастороженность как ошибку. Это приводит к систематическому занижению специфичности и общей точности, а при использовании формальных меток в машинном обучении к риску деградации онконастороженности алгоритма. Разработанная трехуровневая

методология устраняет этот недостаток. Клиническая эффективность, оцениваемая по четырем качественным категориям, обеспечивает нозологически дифференцированный анализ согласованности программы с клиническим мышлением эксперта, хотя и зависит от квалификации последнего. Клиническая точность, рассчитываемая на основании матрицы клинических случаев (JC, UC, MJC, NMJC) с двойным эталоном истины, дает воспроизводимую количественную оценку. По результатам клинического тестирования программы Derma Onko Melanoma Check New на 342 новообразованиях клиническая чувствительность составила 98,5 % (95 % ДИ 94,6–99,6), специфичность 88,1 % (95 % ДИ 83,0–91,8), точность 92,1 % (95 % ДИ 88,8–94,5); ни один из 98 случаев ЗНО не был пропущен. Предложенная методология применима в сравнительных исследованиях программ ИИ, при регуляторной экспертизе диагностических систем и при разработке критериев разметки обучающих данных с учетом клинической онконастороженности.

3.11. Разработка способа оценки и анализ эффективности диагностики новообразований кожи с использованием алгоритмов на основании заключений моделей искусственного интеллекта

Высокая клиническая точность программы, установленная в предыдущем разделе, характеризует качество отдельного диагностического акта, распознавания конкретного новообразования. Однако в реальной практике программа ИИ работает не изолированно, а как звено многоэтапной маршрутизации пациента, проходящего последовательно через врача общей практики (ВОП) или терапевта, врача-дерматовенеролога и, при необходимости, врача-онколога. Клиническая эффективность такой системы определяется не только точностью самой модели, но и тем, как ее заключения встраиваются в этот каскад. Отсюда вытекает вопрос, на который точность модели ответа не дает, приводит ли интеграция ИИ в маршрутизацию к тому, что злокачественных новообразований (ЗНО) кожи выявляется больше, а пропускается меньше, чем при традиционной диагностике

без ИИ. Настоящий раздел посвящен разработке способа, позволяющего дать на этот вопрос количественный ответ.

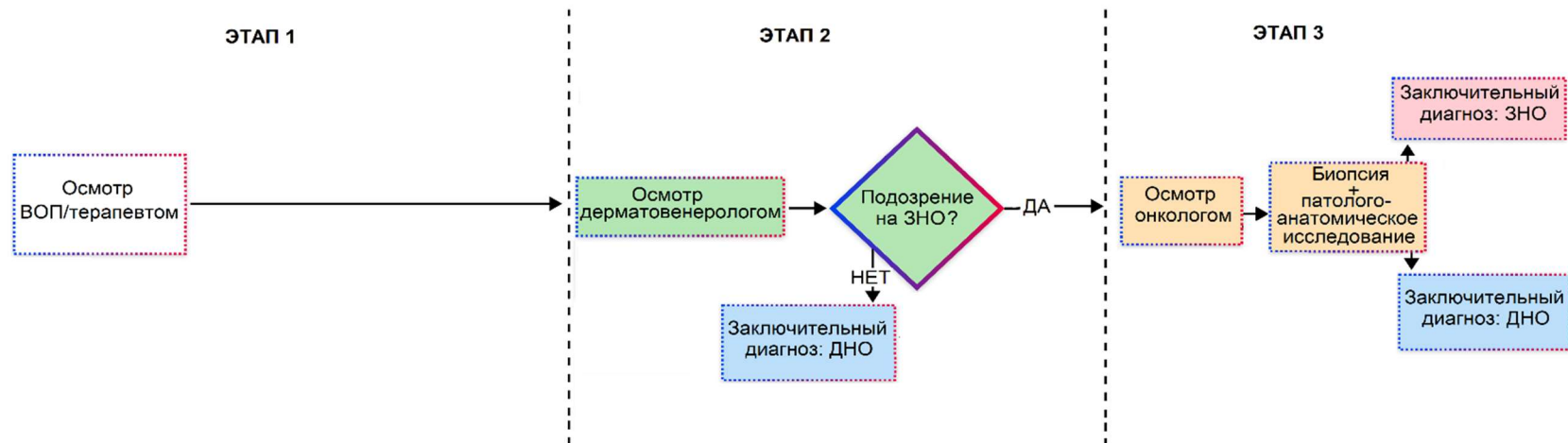
Целью раздела является разработка способа оценки эффективности диагностики при применении дифференцированных алгоритмов, созданных с использованием заключений моделей ИИ, в сравнении с эффективностью стандартной диагностики. В отличие от показателей точности, которые оценивают модель как таковую, предлагаемый способ оценивает алгоритм маршрутизации в целом с учетом каскадного характера ошибок и реальной чувствительности врачей разных специальностей на каждом этапе.

3.11.1. Материал и моделируемые сценарии маршрутизации

Для разработки, валидации и апробации способа использованы клинические данные и фотоизображения новообразований кожи 381 пациента: 291 пациента с доброкачественными новообразованиями (ДНО); 100 немеланоцитарных и 191 меланоцитарная опухоль) и 90 пациентов со ЗНО кожи (39 меланом и 51 базалиома). Материал получен в ходе двух ранее выполненных клинических исследований точности программ ИИ Derma Onko Check и Melanoma Check. Объединение данных в единую выборку позволило сформировать датасет, отражающий реальную структуру обращений пациентов с новообразованиями кожи к врачам первичного звена.

Моделировались два сценария маршрутизации одних и тех же клинических случаев (Рисунок 71). Традиционный сценарий воспроизводил стандартный трехэтапный путь пациента (ВОП/терапевт → дерматовенеролог → онколог), выстроенный в соответствии с действующими нормативными правовыми актами. Сценарий с применением ИИ моделировал маршрутизацию, при которой заключение программы направляет пациента напрямую: при вероятности ЗНО, равной или превышающей порог алгоритма маршрутизации, к врачу-онкологу, а при вероятности ниже порога к врачу-дерматовенерологу.

Алгоритм диагностики ЦЗ без применения ИИ (на примере новообразований кожи)



Алгоритм диагностики ЦЗ с маршрутизацией на основании заключения модели ИИ (на примере новообразований кожи)

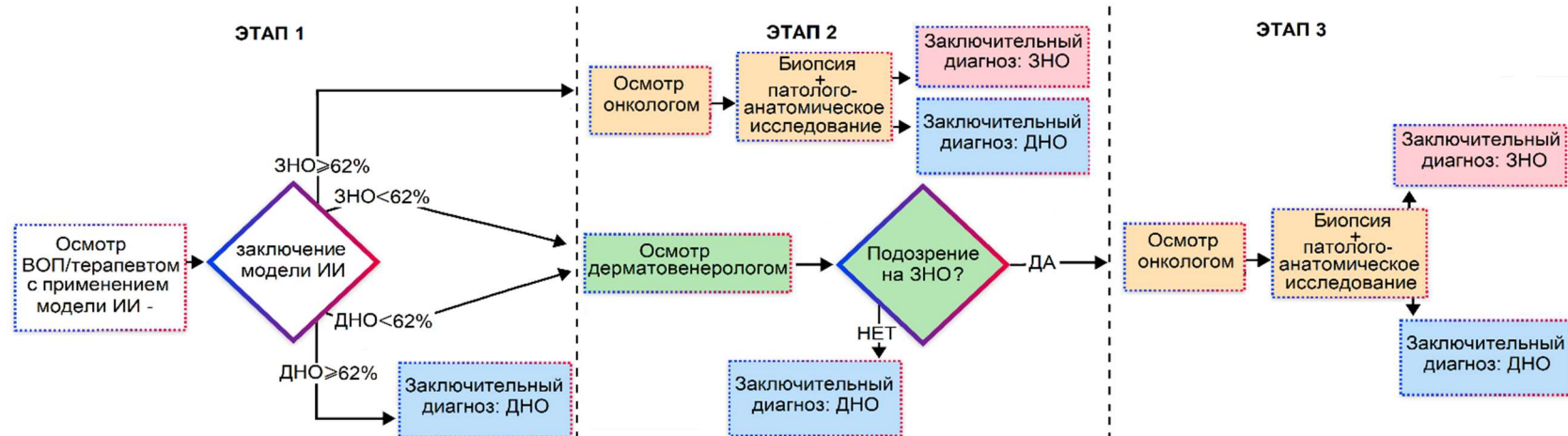


Рисунок 71 – Маршрутизация пациентов при дифференциальной диагностике новообразований кожи: алгоритм диагностики целевого заболевания без применения ИИ; алгоритм диагностики с маршрутизацией на основании заключения модели ИИ

Порог алгоритма маршрутизации это пороговое числовое значение вероятности ЗНО по заключению модели, разграничивающее направления и минимизирующее избыточные обращения к узким специалистам. Оптимальный порог определялся анализом ROC-кривых по максимуму индекса Юдена (сумма чувствительности и специфичности минус единица), что обеспечивает баланс между избыточными направлениями к онкологу и пропущенными случаями. Для программы Derma Onko Check, использованной в приводимом примере, этот порог составил 62 %.

3.11.2. Принцип количественной оценки: две меры относительной эффективности

В основу способа положена идея о том, что клиническую ценность алгоритма с ИИ можно выразить двумя относительными величинами. Первая показывает, во сколько раз применение ИИ увеличивает число выявленных случаев целевого заболевания (ЦЗ), вторая во сколько раз оно снижает число пропущенных случаев. Именно эти два измерения прирост выявляемости и сокращение пропусков и определяют практическую полезность диагностического алгоритма. Предложенные ниже формулы разработаны авторами ввиду отсутствия в литературе инструментов оценки эффективности диагностики в условиях многоэтапной маршрутизации.

Коэффициент прироста выявляемости R_{DAI} определяется как отношение числа всех выявленных случаев ЦЗ при диагностике с применением моделей ИИ (N_{Dai}) к числу выявленных случаев без их применения (N_{Dnai}):

$$R_{DAI} = N_{Dai} / N_{Dnai}. \quad (4)$$

Формула была сконструирована так, чтобы отразить два обстоятельства реальной маршрутизации, ее многоэтапность и наличие порога. Числитель (число выявленных с ИИ случаев N_{Dai}) складывается из двух непересекающихся потоков. Первый образуют истинно положительные случаи с вероятностью не ниже порога, которые направляются прямо к онкологу; поскольку осмотр онколога является

завершающим этапом маршрута, для простоты расчета его чувствительность условно принимается равной 100 %, и весь этот поток считается выявленным. Второй поток образуют случаи с вероятностью ниже порога, как истинно положительные, так и ложноотрицательные по заключению модели, которые направляются к дерматовенерологу; их выявляемость определяется уже чувствительностью дерматовенеролога. Знаменатель (число выявленных без ИИ случаев N_{Dnai}) строится иначе: без порога и прямого направления к онкологу все случаи проходят традиционный маршрут последовательно, поэтому итоговая выявляемость есть произведение размера выборки на чувствительности врачей на каждом пройденном этапе. Развернутая структура формулы такова:

$$R_{DAI} = \frac{N_{Dai}}{N_{Dnai}} = \frac{N_{TP \geq t} \times \prod_{k=1}^n SD_k^{TP \geq t} + (N_{FN < t} + N_{TP < t}) \times \prod_{j=1}^n SD_j^{FN < t + TP < t}}{N \times \prod_{m=1}^n SD_m^{nai}}, \quad (5)$$

где R_{DAI} – отношение количества выявленных случаев ЦЗ при диагностике с применением моделей ИИ на один выявленный случай ЦЗ при диагностике без применения моделей ИИ; N_{Dai} – количество всех выявленных случаев ЦЗ при диагностике с применением моделей ИИ; N_{Dnai} – количество всех выявленных случаев ЦЗ при диагностике без применения моделей ИИ; $N_{TP \geq t}$ – количество ИП случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации; $N_{FN < t}$ – количество ЛО случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации; $N_{TP < t}$ – количество случаев ИП, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации; n – количество этапов маршрутизации (каждый этап – это осмотр врачом-специалистом с или без проведения диагностических исследований по его назначению); $\prod$ – произведение (набора чисел), произведение ряда; $SD_k^{TP \geq t}$ – чувствительность (в %) диагностики ЦЗ врачом-специалистом на каждом этапе маршрутизации ИП случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была выше порога алгоритма маршрутизации; $SD_j^{FN < t + TP < t}$ –

чувствительность (в %) диагностики ЦЗ врачом-специалистом на каждом этапе маршрутизации ЛО и ИП случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации; SD_m^{nai} – чувствительность (в %) ЦЗ врачом-специалистом на каждом этапе маршрутизации при диагностике без применения моделей ИИ; N – количество случаев ЦЗ из выборки.

Пример расчета по формуле 1 при использовании алгоритмов с применением заключения программы ИИ «Derma onko check» для диагностики меланоцитарных и кератиноцитарных/эпидермальных ЗНО кожи:

– $N_{TP \geq t} = 72$ – количество ИП случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была выше порога алгоритма маршрутизации;

– $N_{FN < t} = 6$ – количество ЛО случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации;

– $N_{TP < t} = 6$ количество случаев ИП, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации;

– $SD_k^{TP \geq t}$ – чувствительность (в %) диагностики ЦЗ врачом-специалистом на каждом этапе маршрутизации по направлению ИП случаев – в данном случае не учитывается, так как все случаи ИП направляются к врача-онкологу, осмотр врачом-онкологом является заключительным этапом маршрутизации, и для простоты расчета условно принято, что чувствительность диагностики ЦЗ врачом-онкологом 100%;

– $SD_j^{FN < t + TP < t} = 79,1\%$ – чувствительность диагностики ЦЗ врачами-дерматовенерологами по направлению ЛО и ИП случаев меланоцитарных и кератиноцитарных/эпидермальных ЗНО кожи, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации;

– $N = 90$ (39 меланом и 51 базалиома) – количество случаев ЦЗ из выборки;

– $SD_m^{nai} = 60,1\%$ для врачей ВОП/терапевтов и взвешенная чувствительность 79,1%, для врачей-дерматовенерологов.

Расчет:

$$R_{DAI} = \frac{72 + (6+6) \times 79,1\%}{90 \times 60,1\% \times 79,1\%} = \frac{81,5}{42,8} = 1,9$$

Полученное значение означает, что на один случай ЗНО, выявленный без применения моделей ИИ, приходится 1,9 случая, выявленного с их применением. Значение R_{DAI} больше единицы подтверждает, что порог выбран корректно, а маршрутизация на основе заключений программы увеличивает выявляемость ЗНО; значение меньше единицы указывало бы на обратное. Исходные параметры и результат расчета сведены в Таблице 43.

Таблица 43 – Параметры и результаты выявления целевого заболевания при диагностике с использованием моделей ИИ (программа Derma Onko Check, все типы ЗНО кожи)

Параметр	Значение	Описание
N	90	Общее количество случаев ЦЗ (ЗНО кожи) в выборке (39 меланом + 51 базалиома)
$N_{TP \geq t}$	72	ИП случаи ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации (ИИ правильно направляет)
$N_{TP < t}$	6	ИП случаи ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации;
$N_{FN < t}$	6	ЛО случаи ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации;
$SD_j^{FN < t + TP < t}$ и SD_m^{nai} (врачи-дерматовенерологи)	79,1%	Чувствительность диагностики ЦЗ врачами-дерматовенерологами
SD_m^{nai} (ВОП/терапевты)	60,1%	Чувствительность диагностики ЦЗ ВОП/терапевтами
RDAI	1,9	Количество выявленных случаев ЗНО кожи при применении модели ИИ на один случай ЗНО, выявленный без применения модели ИИ
Примечание: ЦЗ – целевое заболевание; ЗНО – злокачественное новообразование; ИП – истинно положительный; ЛО – ложноотрицательный случай; ИИ – искусственный интеллект; ВОП – врач общей практики		

Коэффициент сокращения пропусков R_{NDAI} определяется симметрично – как отношение числа всех пропущенных (недиагностированных) случаев ЦЗ при диагностике без применения моделей ИИ (N_{NDnai}) к числу пропущенных случаев при их применении (N_{NDai}):

$$R_{\text{NDai}} = N_{\text{NDnai}} / N_{\text{NDai}}. \quad (6)$$

Логика построения этой формулы зеркальна предыдущей, но оперирует уже потерями. Число пропущенных с ИИ случаев N_{NDai} складывается из двух источников: ложноотрицательных заключений модели с вероятностью не ниже порога, которые вообще не были направлены на дообследование (то есть пропущены самой программой), и случаев с вероятностью ниже порога, которые были направлены к дерматовенерологу, но оказались пропущены уже им. Число пропущенных без ИИ случаев N_{NDnai} накапливается по всем этапам традиционного маршрута: на каждом из них часть дошедших до врача случаев теряется в силу его ограниченной чувствительности, и эти доли пропусков суммируются. Развернутая структура такова:

$$R_{\text{NDai}} = \frac{N_{\text{NDnai}}}{N_{\text{NDai}}} = \frac{N \times \prod_{i=1}^n SD_i^{\text{nai}} \times PR_{\text{NDai}} + \sum_{i=1}^n N_i^{\text{NDnai}}}{N_{\text{FN} \geq t} + N_{\text{TP} \geq t} \times \prod_{k=1}^n PR_k^{\text{TP} \geq t} + (N_{\text{FN} < t} + N_{\text{TP} < t}) \times \prod_{j=1}^n PR_j^{\text{FN} < t + \text{TP} < t}}. \quad (7)$$

где R_{NDai} – отношение количества всех не диагностированных, «пропущенных» случаев ЦЗ при диагностике без применения моделей ИИ на один пропущенный случай ЦЗ при диагностике с применением моделей ИИ; N_{NDai} – количество всех не диагностированных, «пропущенных» случаев ЦЗ при диагностике с применением моделей ИИ; N_{NDnai} – количество всех не диагностированных, «пропущенных» случаев ЦЗ при диагностике без применения моделей ИИ; N – количество случаев ЦЗ из выборки; Π – произведение (набора чисел), произведение ряда; Σ – сумма (набора чисел), сумма ряда; N_i^{NDnai} – количество не диагностированных, «пропущенных» случаев ЦЗ, рассчитанная как число ЦЗ, диагностированных на предыдущем этапе, умноженная на долю ЛО заключений врача-специалиста при диагностике без применения моделей ИИ на данном этапе маршрутизации; SD_i^{nai} – чувствительность (в %) диагностики ЦЗ врачом-специалистом без применения моделей ИИ на каждом этапе маршрутизации; n – количество этапов маршрутизации (каждый этап – это осмотр врачом-специалистом с или без проведения диагностических исследований по его назначению); PR_{NDai} – доля (в %) ЛО случаев ЦЗ (не диагностированных, «пропущенных») при диагностике врачом-

специалистом на конечном этапе маршрутизации; $N_{FN \geq t}$ – количество ЛО случаев ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации; $N_{TP \geq t}$ – количество ИП случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации; $N_{FN < t}$ – количество ЛО случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации; $N_{TP < t}$ – количество ИП случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации; $PR_j^{FN < t + TP < t}$ – доля (в %) ЛО случаев при диагностике врачом-специалистом на каждом этапе маршрутизации ЛО и ИП случаев ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации; $PR_k^{TP \geq t}$ – доля (в %) ЛО случаев при диагностике врачом-специалистом на каждом этапе при маршрутизации ИП случаев ЦЗ при установленном пороге алгоритма маршрутизации.

Пример расчета по формуле 2 для программы ИИ «Derma onko check» для меланоцитарных и кератиноцитарных/эпидермальных ЗНО кожи:

- $N_{FN \geq t} = 2$ (количество ЛО случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации)
- $N = 90$ (количество всех случаев ЦЗ из выборки – 39 меланом и 51 базалиома);

- $SD_i^{nai} = 60,1\%$ чувствительность диагностики ЦЗ ВОП/терапевтом на каждом этапе маршрутизации;

- $PR_{NDai} = 20,9\%$ – взвешенная доля (в %) ЛО случаев ЦЗ при диагностике врачами-дерматовенерологами;

- $N_i^{nai} = 35,9$ случаев не диагностировано, «пропущено» (90 случаев ЦЗ (ЗНО кожи), умноженная на долю (в %) ЛО случаев заключений врача-специалиста, в данном случае ВОП/терапевта);

– $N_{TP \geq t}$ – количество ИП случаев ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации – в данном случае не учитывается, так как все ИП случаи ЦЗ направляются к врачу-онкологу, осмотр врачом-онкологом является заключительным этапом маршрутизации, и для простоты расчета условно принято, что чувствительность диагностики ЦЗ врачом-онкологом 100%;

– $PR_k^{TP \geq t}$ – в данном случае доля ЛО случаев ЦЗ при диагностике врачом-специалистом на каждом этапе при маршрутизации ИП случаев при установленном пороге алгоритма маршрутизации не определена, так как ТР не учитывается;

– $N_{FN < t}$ = 6 – количество ЛО случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации;

– $N_{TP < t}$ = 6 – количество ИП случаев, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации;

– $PR_j^{FN < t + TP < t} = 20,9 \%$ – доля ЛО случаев ЦЗ при диагностике врачом-специалистом на каждом этапе маршрутизации;

Расчет:

$$R_{NDM} = \frac{90 \times 60,1\% \times 20,9\% + (90 \times 39,9\%)}{2 + (6+6) \times 20,9\%} = \frac{47,2}{4,5} = 10,5$$

Таким образом, на один случай ЗНО, пропущенный при диагностике с применением моделей ИИ, приходится 10,5 случая, пропущенного при диагностике без них. Столь значительная разница объясняется прежде всего низкой чувствительностью первичного звена: именно на этапе ВОП/терапевта традиционный маршрут теряет более трети всех случаев ЗНО, тогда как маршрутизация с ИИ выводит большинство из них напрямую к онкологу. Исходные параметры и результат приведены в Таблице 44.

Таблица 44 – Параметры и результаты пропусков целевого заболевания при диагностике с использованием моделей ИИ (программа Derma Onko Check, все типы ЗНО кожи)

Параметр	Значение	Описание
N	90	Общее количество случаев ЦЗ (ЗНО кожи) в выборке (39 меланом + 51 базалиома)
$N_{FN \geq t}$	2	ЛО случаи ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации
$N_{FN < t}$	6	ЛО случаи ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации
$N_{TP < t}$	6	ИП случаи ЦЗ, при которых вероятность наличия ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации
SD_i^{nai} (ВОП/терапевты)	60,1%	Чувствительность диагностики ЦЗ ВОП/терапевтами
PR_{NDai} (врачи-дерматовенерологи)	20,9%	Взвешенная доля ЛО случаев ЦЗ (не диагностированных, «пропущенных») при диагностике врачами-дерматовенерологами на конечном этапе маршрутизации
N_{i}^{NDnai}	35,9	Рассчитанное число пропущенных случаев ЦЗ без применения модели ИИ ($90 \times$ доля ЛО случаев у ВОП/терапевтов)
$PR_{j}^{FN < t + TP < t}$	20,9 %	Доля ЛО случаев ЦЗ при диагностике врачами-дерматовенерологами
R _{NDAI}	10,5	Число пропущенных случаев ЦЗ при диагностике без применения модели ИИ на один пропущенный случай ЦЗ при применении модели ИИ

3.11.3. Комплексная визуализация. Квадрантная матрица эффективности

Два полученных коэффициента удобно рассматривать совместно, поскольку по отдельности ни один из них не дает полной картины, технология может увеличивать выявляемость, но одновременно повышать и число пропусков, и наоборот. Для комплексной оценки разработана квадрантная матрица, в которой по оси абсцисс откладывается прирост выявляемости R_{DAI} , а по оси ординат сокращение пропусков R_{NDAI} (Рисунок 72-73). Точка паритета соответствует значениям (1,0; 1,0).

Матрица делится на четыре квадранта с однозначной управленческой интерпретацией. Квадрант I (выявляет больше, пропускает меньше), желательная

зона: технология одновременно повышает выявляемость и снижает пропуски, обладает высокой клинико-диагностической и организационной эффективностью и подлежит приоритетному внедрению. Квадрант II (выявляет больше, но и пропускает больше), неоднозначная зона, требующая дополнительных исследований безопасности и оптимизации порога. Квадрант III (выявляет меньше, пропускает больше), нежелательная зона; технология к внедрению не рекомендуется. Квадрант IV (выявляет меньше, но и пропускает меньше), также неоднозначная зона, требующая тщательной оценки и, возможно, пересмотра порога или самой технологии.

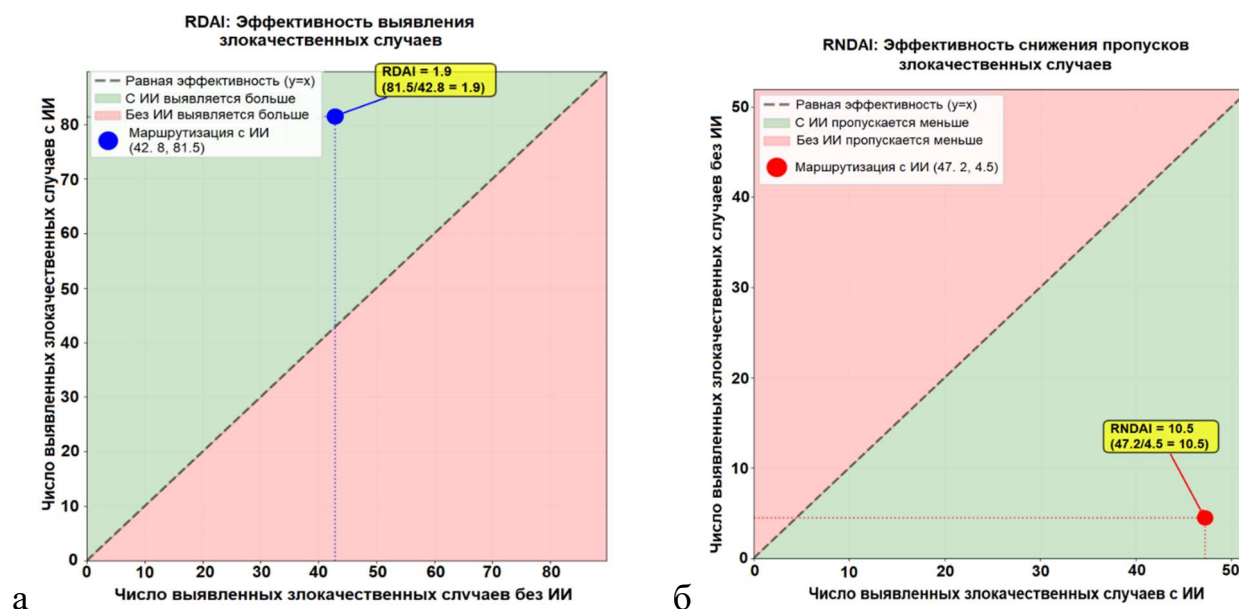


Рисунок 72 – Визуализация результатов количественной оценки относительной эффективности алгоритма маршрутизации на основе заключений программы Derma Onko Check: а – эффективность выявления злокачественных случаев (R_{DAI}); б – эффективность снижения числа пропущенных злокачественных случаев (R_{NDAI})

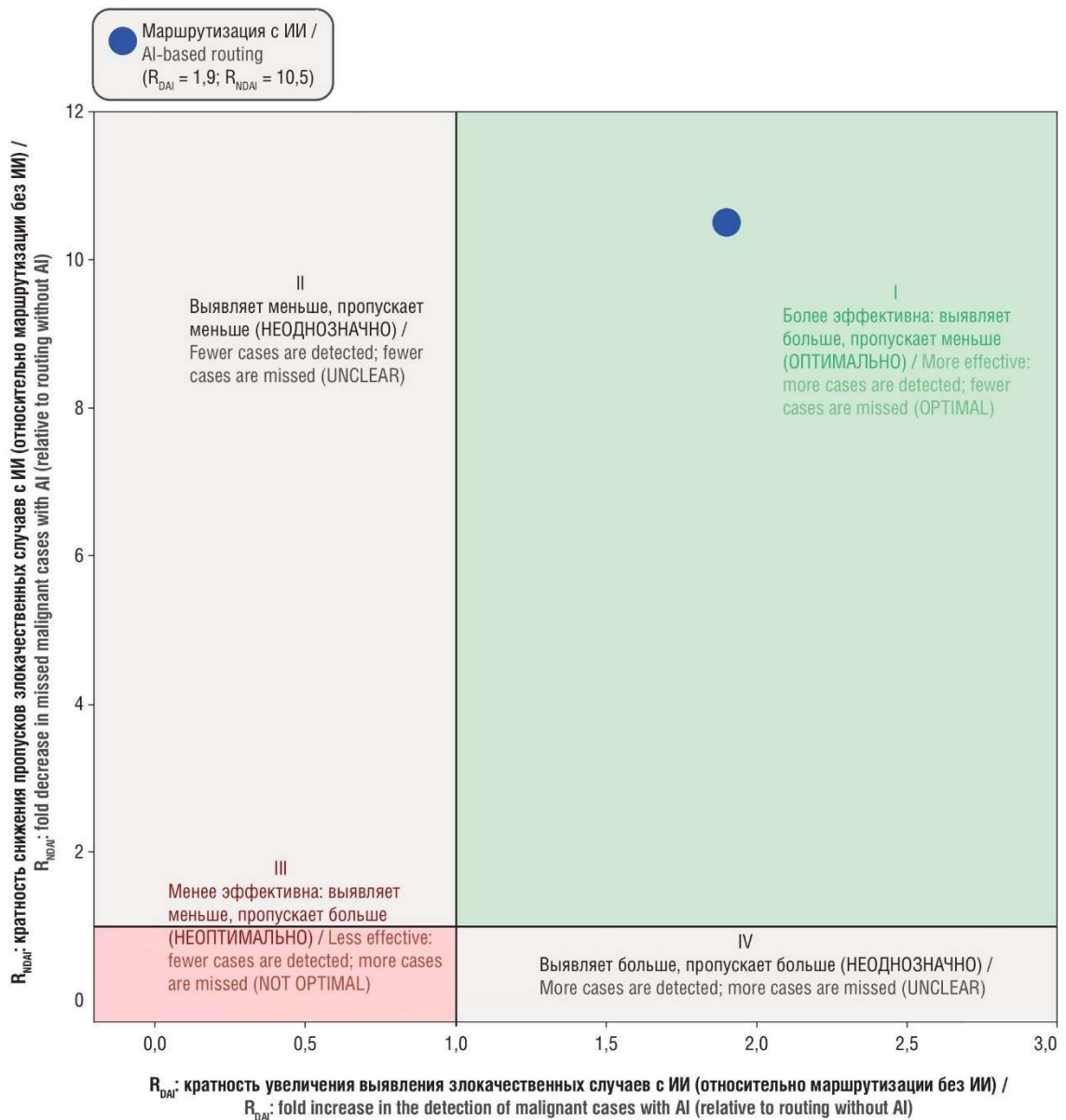


Рисунок 73 – Плоскостная визуализация показателей относительной эффективности алгоритма маршрутизации на основе заключений программы Derma Onko Check в квадрантной матрице (ось абсцисс – R_{DAI} , ось ординат – R_{NDAI} ; квадрант I – оптимальная зона)

Для рассмотренного примера точка сценария с применением Derma Onko Check ($R_{DAI} = 1,9$; $R_{NDAI} = 10,5$) расположена глубоко в квадранте I, значительно выше линии равной эффективности. Это свидетельствует о существенном превосходстве маршрутизации с ИИ над стандартной диагностикой: выявляемость ЗНО возрастает в 1,9 раза при десятикратном сокращении пропусков. Результат подтверждает как правильность выбранного порога (62 %), так и целесообразность интеграции алгоритма в клиническую практику.

3.11.4. Программная реализация

Для практического применения предложенных формул и обеспечения воспроизводимости расчетов разработано мобильное приложение для операционной системы Android – AI-diagnostic efficiency calculator (CalcRDAI&RNDAI; свидетельство о государственной регистрации программы для ЭВМ, заявл. 10.02.2026) [95]. Приложение обеспечивает автоматизированный ввод исходных параметров, расчет коэффициентов R_{DAI} и R_{NDAI} и визуализацию результата в виде квадрантной матрицы, что делает способ пригодным для применения непосредственно в организационно-управленческой практике.

Примеры экранов ввода параметров и расчетов RDAI, RNDAI, результата расчетов и визуализации их в квадратной матрице представлены на Рисунках 74-76.

Калькулятор расчёта эффективности диагностики ЦЗ с использованием ИИ

? ИНСТРУКЦИЯ ПО ИСПОЛЬЗОВАНИЮ

Расчет R_DAI

Коэффициент отношения количества случаев ЦЗ, выявленных при диагностике с применением ИИ, на один случай, выявленный при диагностике без применения ИИ

Формула

$$R_{Dai} = \frac{N_{Dai}}{N_{Dnai}} = \frac{N_{TP \geq t} \prod_{i=1}^n SD_i^{TP \geq t} + (N_{FN < t} + N_{TP < t}) \times \prod_{i=1}^n SD_i^{FN < t, TP < t}}{N \times \prod_{i=1}^n SD_i^{all}}$$

Нажмите для увеличения

Основные параметры

N_TP ≥ t

72

ИП случаи ЦЗ с вероятностью наличия ЦЗ по заключению ИИ выше порога алгоритма маршрутизации

N_FN < t

6

ЛО случаи ЦЗ с вероятностью наличия ЦЗ по заключению ИИ ниже порога алгоритма маршрутизации

N_TP < t

6

ИП случаи ЦЗ с вероятностью наличия ЦЗ по заключению ИИ ниже порога алгоритма маршрутизации

N

90

Общее количество случаев ЦЗ из выборки

Этапы для расчета $\prod SD_k^{(TP \geq t)}$ (для ИП ≥ t случаев)

Количество этапов для расчета произведения чувствительности (в %) диагностики врача-специалиста с использованием ИИ для ИП случаев выше порога

n (этапов)

УСТАНОВИТЬ

Этапы для расчета $\prod SD_i^{(FN < t, TP < t)}$ (для ЛО и ИП < t случаев)

Количество этапов для расчета произведения чувствительности (в %) диагностики врача-специалиста для ЛО и ИП случаев ЦЗ ниже порога алгоритма маршрутизации

n (этапов)

1

УСТАНОВИТЬ

SD_1 (этап 1, %)

79,1

Чувствительность (в %) диагностики врача-специалиста для ЛО и ИП случаев ЦЗ ниже порога алгоритма маршрутизации на этапе 1

Этапы для расчета $\prod SD_m^{nai}$ (без ИИ)

Количество этапов для расчета произведения чувствительности (в %) диагностики врача-специалиста без использования ИИ

n (этапов)

2

УСТАНОВИТЬ

SD_m1 (этап 1, %)

60,1

Чувствительность (в %) диагностики врача-специалиста без использования ИИ на этапе 1

SD_m2 (этап 2, %)

79,1

Чувствительность (в %) диагностики врача-специалиста без использования ИИ на этапе 2

РАССЧИТАТЬ R_DAI

R_DAI = 1,9

✓ На один выявленный случай ЦЗ без ИИ приходится 1.90 выявленных с ИИ

N_Dai = 81.49 (всех выявленных случаев ЦЗ с ИИ)

N_Dnai = 42.79 (всех выявленных случаев ЦЗ без ИИ)

✓ Диагностика ЦЗ с ИИ эффективнее!

Рисунок 74 – Пример ввода и расчета результатов по формуле 1 (R_{DAI})



Расчет R_NDAI

Коэффициент отношения количества случаев ЦЗ, пропущенных при диагностике без применения ИИ на один случай ЦЗ, пропущенный при диагностике с применением ИИ



Формула

$$R_{NDai} = \frac{N_{NDai}}{N_{NDai}} = \frac{N \times \prod_{i=1}^n [SD_{i1}^{SD_{i1}} \times PR_{NDai}] + \sum_{i=1}^n N_{NDai}}{N_{FN \geq t} + N_{TP \geq t} \times \prod_{k=1}^n PR_k^{TP \geq t} + (N_{FN < t} + N_{TP < t}) \times \prod_{j=1}^n PR_j^{FN < t, TP < t}}$$

Нажмите для увеличения

Основные параметры

N

90

Общее количество случаев ЦЗ из выборки

N_FN ≥ t

2

ЛО случаи ЦЗ с вероятностью наличия ЦЗ по заключению ИИ выше порога алгоритма маршрутизации

N_FN < t

6

ЛО случаи ЦЗ с вероятностью наличия ЦЗ по заключению ИИ ниже порога алгоритма маршрутизации

N_TP < t

6

ИП случаи ЦЗ с вероятностью наличия ЦЗ по заключению ИИ ниже порога алгоритма маршрутизации

PR_NDAi (%)

20,9

Доля (в %) ЛО случаев ЦЗ (не диагностированных, «пропущенных») при диагностике врачом-специалистом на конечном этапе маршрутизации

Этапы для расчета [SD_i^ndai (без ИИ)

Количество этапов для расчета произведения чувствительности (в %) диагностики ЦЗ врача-специалиста при диагностике без использования моделей ИИ

n (этапов)

1

УСТАНОВИТЬ

SD_i1 (этап 1, %)

60,1

Чувствительность (в %) диагностики ЦЗ врача-специалиста без использования ИИ на этапе 1

Этапы для расчета ΣN_i^NDnai (пропущенные случаи ЦЗ)

Количество этапов для расчета суммы не диагностированных, «пропущенных» случаев ЦЗ у врача-специалиста

n (этапов)

1

УСТАНОВИТЬ

N_i1 (этап 1)

35,9

Количество не диагностированных, «пропущенных» случаев ЦЗ (рассчитывается как число ЦЗ, диагностированных на предыдущем этапе, умноженное на долю ЛО врача-специалиста) на этапе 1

Этапы для расчета [PR_k^*(TP ≥ t) (доля ЛО для ИП ≥ t)

Количество этапов для расчета произведения доли (в %) ЛО случаев ЦЗ при диагностике врачом-специалистом для ИП случаев выше порога

n (этапов)

УСТАНОВИТЬ

Этапы для расчета [PR_j^*(FN < t, TP < t) (доля ЛО для ИП и ЛО < t)

Количество этапов для расчета произведения доли (в %) ЛО случаев ЦЗ при диагностике врачом-специалистом для ЛО и ИП случаев ниже порога

n (этапов)

1

УСТАНОВИТЬ

PR_j1 (этап 1, %)

20,9

Доля (в %) ЛО случаев ЦЗ диагностики врача-специалиста для ЛО и ИП случаев ниже порога алгоритма маршрутизации на этапе 1

РАССЧИТАТЬ R_NDAI

R_NDAI = 10,47

✓ На один не диагностированный, «пропущенный» случай ЦЗ с ИИ приходится 10.47 пропущенных без ИИ

N_NDnai = 47.20 (всех не диагностированных, «пропущенных» случаев ЦЗ без ИИ)

N_NDAi = 4.51 (всех не диагностированных, «пропущенных» случаев ЦЗ с ИИ)

✓ Диагностика ЦЗ с ИИ эффективнее!

Рисунок 75 – Пример ввода и расчета результатов по формуле 2 (R_{NDAI})

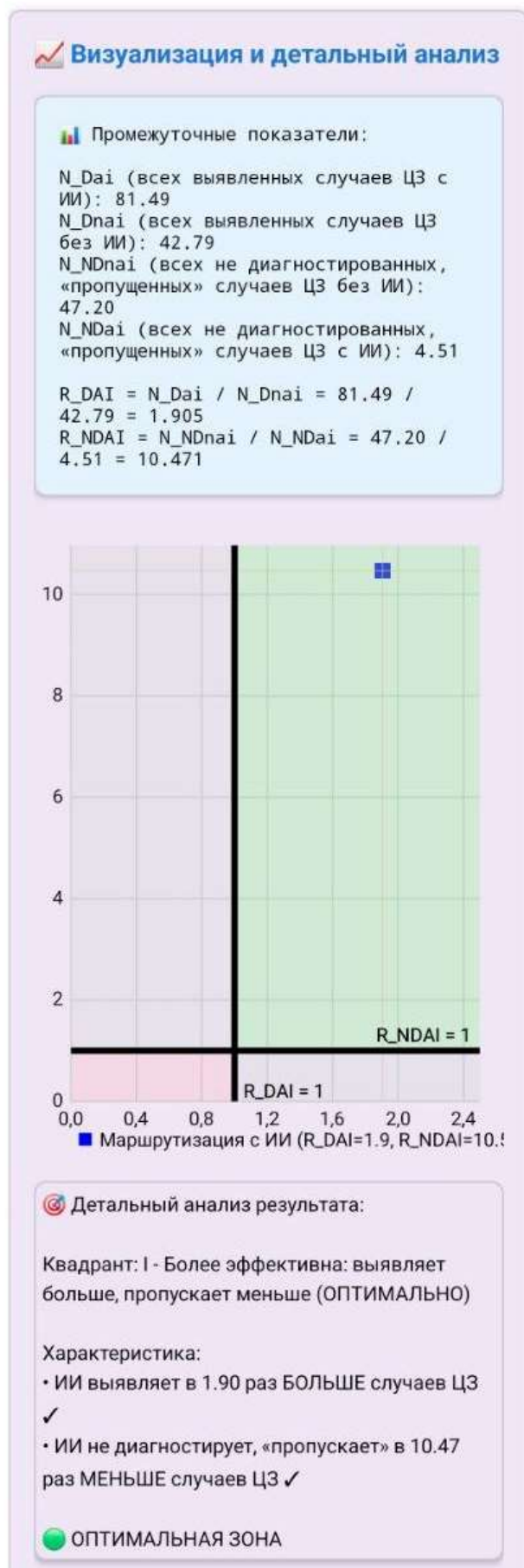


Рисунок 76 – Пример визуализации результатов расчета в квадрантной матрице

3.11.5. Значение результатов и переход к экономической оценке

Полученные результаты подтверждают исходную гипотезу: интеграция заключений моделей ИИ в многоэтапные алгоритмы диагностики существенно повышает выявляемость ЗНО за счет компенсации ограниченной чувствительности первичного звена и оптимизации потоков пациентов. Предложенный способ согласуется с современными представлениями об оценке эффективности здравоохранения. В отличие от традиционного клинико-экономического анализа, ориентированного на отдельные изделия или организации, предложенный подход учитывает каскадный характер ошибок реальной маршрутизации, что делает его особенно пригодным для оценки цифровых решений.

К преимуществам способа относятся универсальность (он применим к любым нозологиям и любому числу этапов), простота расчета и визуализации, автоматизация через приложение и учет реальной чувствительности специалистов. Его ограничения связаны с ретроспективным характером и ограниченным объемом данных (381 случай, одна нозологическая группа), с заимствованием показателей чувствительности врачей из литературы и с тем, что прямые затраты на саму технологию ИИ в модель не включены. Именно последнее ограничение и определяет логику дальнейшего изложения: показав, что алгоритм с ИИ клинически эффективнее, необходимо установить, эффективен ли он экономически. Этот вопрос – о финансовых затратах системы здравоохранения – рассматривается в следующих двух разделах, сначала на уровне отдельного учреждения, а затем в масштабе региона.

Предложенный способ оценки эффективности диагностики, основанный на оригинальных расчетных формулах R_{DAI} и R_{NDAI} с визуализацией результата в квадрантной матрице, представляет собой простой и объективный инструмент анализа клинико-диагностической и организационно-распорядительной эффективности технологий ИИ в условиях многоэтапной маршрутизации. В приведенном примере применение программы Derma Onko Check с порогом 62 % увеличивает выявляемость ЗНО кожи в 1,9 раза и снижает число пропущенных

случаев в 10,5 раза по сравнению с традиционной диагностикой, а соответствующая точка располагается в оптимальном квадранте I. Универсальность и простота способа позволяют использовать его для оценки диагностических технологий ИИ при любой нозологии, а также при обосновании их включения в клинические рекомендации и региональные методические документы.

3.12. Разработка способа оценки и анализ медико-экономической эффективности диагностики с применением алгоритмов маршрутизации пациентов на основании заключений моделей искусственного интеллекта

3.12.1. Медико-экономическая эффективность диагностики с применением алгоритмов маршрутизации на основании заключений моделей искусственного интеллекта на уровне медицинской организации

Злокачественные новообразования (ЗНО) кожи, и прежде всего меланома, относятся к социально значимым заболеваниям и требуют значительных расходов на диагностику, лечение, реабилитацию и социальную поддержку пациентов. Поэтому оценка медико-экономической эффективности (МЭЭ) диагностики с применением ИИ должна учитывать не только прямые затраты на медицинские вмешательства, но и потенциально предотвратимые потери, связанные с пропущенными случаями заболевания. Настоящий раздел, продолжая предложенный ранее подход к оценке выявляемости, разрабатывает способ оценки МЭЭ с симметричным учетом предотвратимых и потенциально предотвратимых затрат. Целью раздела является разработка и валидация способа оценки МЭЭ выявления целевого заболевания (ЦЗ) с применением алгоритмов на основании заключений моделей ИИ в условиях многоэтапной маршрутизации.

Понятийный аппарат и структура затрат

Ключевая методологическая идея раздела состоит в том, что суммарные затраты на диагностику при применении ИИ и без него имеют принципиально разную структуру, а потому их корректное сравнение требует предварительной классификации. Расходная часть формируется из трех компонентов, затрат на обследование ложноположительных (ЛП) случаев, затрат на обследование истинно отрицательных (ИО) случаев и затрат на лечение ложноотрицательных (ЛО, пропущенных) случаев ЦЗ. Каждый из компонентов учитывает многоэтапность маршрутизации и различную стоимость вмешательств на разных ее этапах. Предложенная авторами классификация прямых и косвенных медицинских затрат приведена в Таблице 45.

Прямые медицинские затраты объединяют расходы непосредственно на диагностические вмешательства, осмотры врачей-специалистов, лабораторные и инструментальные исследования, биопсию и патологоанатомические исследования. Косвенные медицинские затраты — это потери системы здравоохранения, не связанные прямо с диагностикой: расходы вследствие временной нетрудоспособности, инвалидности и преждевременной смертности трудоспособного населения, обусловленных несвоевременным выявлением ЦЗ. Предотвратимыми считаются затраты на необоснованные вмешательства, вызванные ЛП-результатами; они могут быть исключены при повышении специфичности без ущерба для выявляемости. Потенциально предотвратимыми называются затраты на дополнительную диагностику и лечение пропущенных случаев, которые поступают в медицинские организации на поздних стадиях; их снижение лишь потенциально достижимо и зависит от чувствительности метода.

Таблица 45 – Классификация видов медико-экономической эффективности способа диагностики

Вид МЭЭ	Прямые медицинские затраты / Direct Medical costs	Косвенные медицинские затраты / Indirect Medical costs
Фактическая МЭЭ	<i>Общие затраты / Total costs</i>	
	– Предотвратимые затраты (англ. Avoidable costs)– Предотвратимые затраты на одного пациента (англ. Avoidable Costs per patient) – Стоимость на одного пациента (англ. Cost per Patient)	– Фактические выплаты пособия по временной нетрудоспособности (Социальный фонд России) пациентов с несвоевременно выявленными (пропущенными) случаями ЦЗ – Фактические выплаты по инвалидности (Социальный фонд России), установленной вследствие прогрессирования ЦЗ и развития его осложнений при поздней диагностике ЦЗ – Фактические потери валового национального продукта от преждевременной смертности и инвалидизации трудоспособного населения
Потенциальная МЭЭ	<i>Потенциальные общие затраты (англ. Potential Total costs)</i>	
	– Потенциально предотвратимые затраты (англ. Potential Avoidable costs) – Потенциально предотвратимые затраты на одного пациента (англ. Potential Avoidable Costs per Patient) – Потенциальная стоимость на одного пациента (англ. Potential Cost per Patient)	– Потенциальные потери от выплат пособия по временной нетрудоспособности (Социальный фонд России) при несвоевременно выявленных (пропущенных) случаях ЦЗ – Потенциальные потери от выплат по инвалидности (Социальный фонд России), установленной вследствие прогрессирования ЦЗ и развития его осложнений при поздней диагностике ЦЗ – Потенциальные потери валового национального продукта от преждевременной смертности и инвалидизации трудоспособного населения
Примечание: МЭЭ – медико-экономическая эффективность; СФР – Социальный фонд России; ЦЗ – целевое заболевание; ВВП – валовой национальный продукт		

Принципиально важно, что необоснованными признаются исключительно затраты на диагностику новообразований с клиническими и дерматоскопическими признаками доброкачественности. Затраты на обследование образований с признаками злокачественности или предраковыми изменениями не расцениваются как необоснованные независимо от окончательного морфологического заключения. Так, если образование по клинико-дерматоскопической картине соответствует диспластическому невусу или иному пограничному состоянию и направляется на морфологическое исследование, но злокачественность не подтверждается, такой случай не относится к ЛП и не считается необоснованной тратой, поскольку обследование было обусловлено объективными показаниями.

Этот принцип полностью согласуется с концепцией онконастороженности, развернутой в разделе 7.1, в настоящем исследовании ни один случай с клинически обоснованным направлением на биопсию при наличии визуальных признаков злокачественности не был отнесен к ЛП-результатам.

Решение проблемы сопоставимости сценариев

Сравнение двух сценариев диагностики осложняется тем, что они могут иметь противоположную структуру затрат. Сценарий без ИИ нередко демонстрирует более высокие прямые затраты на диагностику (из-за большего числа оправданных направлений), но существенно меньшее число пропущенных случаев; сценарий с ИИ, напротив, может снижать прямые затраты, но при неудачно выбранном пороге увеличивать пропуски. Формальное сравнение одних лишь прямых затрат в такой ситуации приводит к ошибочному выводу, игнорируя скрытые расходы на лечение поздних стадий. Для устранения этой систематической ошибки в формулу коэффициента соотношения затрат вводится идентичный компонент затрат на лечение пропущенных случаев (CpP) – одновременно в числитель и в знаменатель. Такой прием выполняет роль нормализующего фактора, уравнивающего сценарии по полному спектру последствий. Существенно, что точность абсолютного значения CpP не критична для корректности сравнения, поскольку этот компонент входит в обе части соотношения; методологически значим сам факт его учета.

Стоимостная база расчета

Стоимость медицинских услуг принята по Тарифному соглашению на оплату медицинской помощи по Территориальной программе обязательного медицинского страхования (ОМС) города Москвы на 2025 год¹. Тарифы базовых диагностических услуг приведены в Таблице 46, а взвешенная стоимость лечения

¹ <https://base.garant.ru/411501273>.

меланомы кожи по стадиям в Таблицах 47 и 48. Коэффициенты в таблицах отражают средневзвешенную частоту оказания каждой услуги на одного пациента в рамках соответствующей стадии и рассчитаны по результатам опроса трех независимых экспертов-онкологов со стажем более 20 лет, оценивавших реальную практику в соответствии с клиническими рекомендациями «Меланома кожи и слизистых оболочек» (2025 г.)².

Таблица 46 – Стоимость базовых медицинских услуг по Территориальной программе ОМС города Москвы на 2025 год

Наименование медицинской услуги	Код услуги	Тариф ОМС 2025, руб.	Коэффициент	Взвешенная стоимость, руб.
Прием врача-онколога лечебно-диагностический, первичный, амбулаторный	1191	268,59	1,0	268,59
Прием врача-дерматовенеролога лечебно-диагностический, первичный, амбулаторный	1331	324,05	1,0	324,05
Осмотр кожи под увеличением (дерматоскопия)	13010	142,71	1,0	142,71
Биопсия 5-й категории сложности без дополнительных методов исследования	29005	2558,08	2,5	6395,20
Примечание: ОМС – обязательное медицинское страхование. Взвешенная стоимость = тариф × коэффициент				

Таблица 47 – Взвешенная стоимость медицинских услуг при лечении меланомы кожи (C43), стадии I–II (локальное заболевание)

№	Наименование медицинской услуги / КСГ	Код ОМС	Тариф ОМС 2025, руб.	Коэффициент	Взвешенная стоимость, руб.
<i>А. Амбулаторная медицинская помощь</i>					
1	Прием врача-онколога лечебно-диагностический, первичный, амбулаторный	1191	268,59	1,0	268,59
2	Прием врача-онколога лечебно-диагностический, повторный, амбулаторный	1192	180,10	3,0	540,30
3	Онкологический консилиум (определение тактики лечения)	1781	6165,23	1,0	6165,23

² https://cr.minzdrav.gov.ru/preview-cr/921_1.

Продолжение Таблицы 47

4	Онкологический консилиум (изменение тактики лечения)	1782	1177,00	0,1	117,70
Итого по разделу А: 7091,82 руб.					
<i>Б. Диагностические исследования</i>					
1	Осмотр кожи под увеличением (дерматоскопия)	13010	142,71	1,0	142,71
2	УЗИ мягких тканей, лимфатических узлов (одна зона)	40054	130,34	2,5	325,85
3	Биопсия 5-й категории сложности без дополнительных методов исследования	29005	2558,08	2,5	6395,20
4	ИГХ-исследование (один маркер: Melan-A, HMB-45, SOX-10)	29007	3355,28	0,4	1342,11
5	КТ одной анатомической области	37048	6244,73	2,5	15 611,83
6	МРТ одной анатомической области	37049	9366,54	1,0	9366,54
Итого по разделу Б: 33 184,24 руб.					
<i>В. Лечение (хирургическое вмешательство)</i>					
1	Иссечение опухоли кожи с эксцизионной биопсией СЛУ или эксцизионная биопсия СЛУ с реэксцизией послеоперационного рубца	200560	299 635,44	1,0	299 635,44
2	Реконструктивно-пластическое вмешательство при меланоме	90007	210 367,00	0,3	63 110,10
Итого по разделу В: 362 745,54 руб.					
ИТОГО (I–II стадии): 403 021,60 руб.					

Таблица 48 – Взвешенная стоимость медицинских услуг при лечении меланомы кожи (C43), стадии III–IV (местно-распространенное и метастатическое заболевание)

№	Наименование медицинской услуги / КСГ	Код ОМС	Тариф ОМС 2025, руб.	Коэффициент	Взвешенная стоимость, руб.
<i>А. Амбулаторная медицинская помощь</i>					
1	Прием врача-онколога лечебно-диагностический, первичный, амбулаторный	1191	268,59	1,000	268,59
2	Прием врача-онколога лечебно-диагностический, повторный, амбулаторный	1192	180,10	1000	1801,00
4	Онкологический консилиум (определение тактики лечения)	1781	6165,23	2,000	12 330,46
5	Онкологический консилиум (изменение тактики лечения)	1782	1177,00	0,700	823,90
6	Онкологический консилиум (определение тактики обследования)	1783	294,25	0,200	58,85
Итого по разделу А: 15 282,80 руб.					

Продолжение Таблицы 48

<i>Б. Диагностические исследования</i>					
1	Осмотр кожи под увеличением (дерматоскопия)	13010	142,71	1,000	142,71
2	Трепан-биопсия опухолей наружных локализаций, лимфатических узлов под визуальным контролем	20003	3643,40	1,000	3643,40
3	Лечебно-диагностическая пункция под контролем УЗИ	40063	690,88	2,000	1381,76
4	УЗИ мягких тканей и лимфатических узлов (1 зона)	40054	130,34	4,500	586,53
5	Биопсия 5-й категории сложности без дополнительных методов исследования	29005	2558,08	4,000	10 232,32
6	ИГХ-исследование (1 маркер: Melan-A, HMB-45, SOX-10)	29007	3355,28	0,500	1677,64
7	Молекулярно-генетический тест (мутация <i>BRAF</i>)	29012	6208,91	1,000	6208,91
8	Молекулярно-генетический тест (мутации <i>KRAS/NRAS</i>)	29014	6 091,50	0,200	1 218,30
9	Определение активности лактатдегидрогеназы	26385	38,32	2,500	95,80
10	Исследование белка S100 (опухолевый маркер меланомы)	23077	1 704,23	0,300	511,27
11	КТ одной анатомической области	37048	6 244,73	1,800	11 240,51
12	МРТ одной анатомической области	37049	9 366,54	1,000	9 366,54
13	ПЭТ/КТ всего тела с ¹⁸ F-ФДГ	37060	41 409,00	0,300	12 422,70
Итого по разделу Б: 58 728,39 руб.					
<i>В. Лечение (хирургическое вмешательство)</i>					
1	Иссечение опухоли кожи с эксцизионной биопсией СЛУ или эксцизионная биопсия СЛУ с резексией послеоперационного рубца	200560	299 635,44	0,200	59 927,09
2	Реконструктивно-пластическое вмешательство при меланоме	90007	210 367,00	0,100	21 036,70
Итого по разделу В: 80 963,79 руб.					
<i>Г. Лечение (системная лекарственная терапия – базовый тариф ОМС)</i>					
1	Дневной стационар онкологического профиля	97158	46 019,36	1,650	75 931,94
2	Круглосуточный стационар онкологического профиля	81094	46 019,36	0,150	6902,90
Итого по разделу Г: 82 834,84 руб.					
<i>Д. Лекарственная биологическая (иммунотерапия и таргетная) терапия – программа ВЗН*</i>					
1	Пембролизумаб – иммунотерапия анти-PD1 (200 мг 18 доз / 12 мес.)	ВЗН	1 742 400,00	0,285	496 584,00
2	Ниволумаб – иммунотерапия анти-PD1 (480 мг × 13 доз / 12 мес.)	ВЗН	2 663 356,80	0,190	506 037,79
3	Ипилимумаб + Ниволумаб – комбинированная иммунотерапия (индукция 4 цикла / 12 мес.)	ВЗН	3 774 911,92	0,190	717 233,26

Продолжение Таблицы 48

4	Дабрафениб + Трaметиниб – таргетная терапия BRAF V600mut (ежедневно / 12 мес.)	ВЗН	2 910 000,00	0,285	829 350,00
Итого по разделу Д: 2 549 205,05 руб.					
ИТОГО (III–IV стадии): 2 787 014,87 руб.					
Примечание: ВЗН – программа «14 высокoзатратных нoзoлoгий»; ПЭТ/КТ – пoзитронно-эмиссионная томография, совмещенная с КТ; ФДГ – фтордезоксиглюкоза. Препараты раздела Д финансируются за счет федерального бюджета, бесплатно для пациента. 95 % пациентов III–IV стадий получают биологическую терапию, 5 % – химиотерапию					

Для валидации способа использованы результаты диагностики меланоцитарных опухолей кожи программой Derma Onko Check при пороге 62 % (n=230; Таблица 49). Характеристики точности врачей-специалистов, положенные в основу моделирования традиционного сценария, приведены в Таблице 50, чувствительность выявления ЗНО кожи составила 44,6 % для ВОП/терапевтов, 79,2 % для дерматовенерологов и 66,86 % для онкологов. Для примера экстраполяции на региональный уровень использованы данные о распределении впервые выявленных случаев меланомы кожи (С43) в Москве в 2024 г. по стадиям (Таблица 51).

Таблица 49 – Распределение случаев по результатам диагностики меланоцитарных опухолей программой Derma Onko Check (n=230, порог маршрутизации 62 %)

Категория заключений программы	Вероятность ЦЗ ≥62 %, n (%)	Вероятность ЦЗ <62 %, n (%)	Итого, n (%)
ИП (ЗНО, верно)	31 (13,5)	6 (2,6)	37 (16,1)
ЛП (ДНО, неверно)	11 (4,8)	8 (3,5)	19 (8,3)
ИО (ДНО, верно)	165 (71,7)	7 (3,1)	172 (74,8)
ЛО (ЗНО, неверно)	1 (0,4)	1 (0,4)	2 (0,8)
Всего	208 (90,4)	22 (9,6)	230 (100)

Таблица 50 – Чувствительность, специфичность и точность диагностики врачами-специалистами, %

Врачи-специалисты	Чувствительность	Специфичность	Точность
Онкологи	66,86	78,00	72,43
ВОП/терапевты	44,60	49,40	47,00
Дерматовенерологи	79,20	77,20	78,20

Таблица 51 – Распределение впервые выявленных случаев меланомы кожи (С43) по стадиям в г. Москве в 2024 г.

Показатель	Значение
Общее число новых случаев меланомы кожи (С43), n	1628
Доля диагностированных случаев, %	100
на I стадии	57,6
на II стадии	27,6
на III стадии	7,1
на IV стадии	7,7
Доля морфологической верификации, %	99,6
Примечание: диагностика на поздних стадиях обуславливает наибольшие потенциально предотвратимые затраты	

Расчетные формулы и учет многоэтапности

Ключевой показатель МЭЭ – коэффициент соотношения затрат R_{TC} , равный отношению общих затрат на диагностику ЦЗ с применением моделей ИИ (TC_{ai}) к общим затратам без их применения (TC_{nai}):

$$R_{TC} = TC_{ai} / TC_{nai} . \quad (8)$$

Методология построения формул опиралась на четыре последовательных принципа. Первый, анализ структуры затрат: было установлено, что суммарные затраты каждого сценария складываются из трех компонентов затрат на обследование ЛП-случаев, затрат на обследование ИО-случаев и затрат на лечение пропущенных ЛО-случаев. Второй, учет многоэтапности: поскольку диагностика представляет собой каскад (ВОП/терапевт → дерматовенеролог → онколог → биопсия с патологоанатомическим исследованием), каждый компонент декомпозируется по этапам. Третий, моделирование потоков пациентов, число пациентов на очередном этапе рассчитывается как произведение их числа на предыдущем этапе на показатель чувствительности или специфичности соответствующего врача, что и воспроизводит каскадный характер направлений. Четвертый, обеспечение сопоставимости за счет симметричного включения компонента CrP , о котором сказано выше.

Формула для расчета коэффициента отношения ФЗ на диагностику ЦЗ с применением алгоритмов на основании заключений моделей ИИ и ФЗ на диагностику ЦЗ без применения моделей ИИ:

Формула 1

– полная (развернутая) формула:

$$R_{TC} = \frac{TC_{ai}}{TC_{nai}} = \frac{(N_{FP \geq t} \times DMC_{FP \geq t} + \sum_{i=1}^n (N_i^{FP \geq t} \times DMC_i^{FP \geq t}) + N_{FP < t} \times DMC_{FP < t} + \sum_{m=1}^n (N_m^{FP < t} \times DMC_m^{FP < t}) + N_{TN < t} \times DMC_{TN < t} + \sum_{k=1}^n (N_k^{TN < t} \times DMC_k^{TN < t}) + (N_{FN \geq t} + (\sum_{s=1}^n N_s^{FN < t}) + (\sum_{f=1}^n N_f^{TP \geq t}) + (\sum_{p=1}^n N_p^{TP < t})) \times CrP}{\sum_{r=1}^n (N_r^{FPnai} \times DMC_r^{FPnai}) + (\sum_{j=1}^n N_j^{FNnai}) \times CrP} \quad (9)$$

– сокращенная формула:

$$R_{TC} = \frac{TC_{ai}}{TC_{nai}} = \frac{AC_{FP}^{ai} + AC_{TN}^{ai} + PAC_{ND}^{ai}}{AC_{FP}^{nai} + PAC_{ND}^{nai}}, \quad (10)$$

где R_{TC} – коэффициент отношения ФЗ на диагностику ЦЗ с применением моделей ИИ и ФЗ на диагностику ЦЗ без применения моделей ИИ (показывает, сколько ФЗ на диагностику ЦЗ с применением моделей ИИ приходится на 1 руб. ФЗ на диагностику ЦЗ без применения моделей ИИ); TC_{ai} – общие ФЗ на диагностику ЦЗ с применением моделей ИИ (руб.); TC_{nai} – общие ФЗ на диагностику ЦЗ без применения моделей ИИ (руб.); AC_{FP}^{ai} – предотвратимые ФЗ на диагностику ЛП случаев ЦЗ при применении моделей ИИ (руб.); AC_{TN}^{ai} – предотвратимые ФЗ на диагностику ИО случаев ЦЗ при применении ИИ (руб.); PAC_{ND}^{ai} – потенциально предотвратимые ФЗ на ЛО (недиагностированные, пропущенные) случаи ЦЗ при применении ИИ (руб.); AC_{FP}^{nai} – предотвратимые ФЗ на диагностику ЛП случаев ЦЗ без применения ИИ (руб.); PAC_{ND}^{nai} – потенциально предотвратимые ФЗ на ЛО (недиагностированные, пропущенные) случаи ЦЗ при диагностике без применения ИИ (руб.); $\sum$ – сумма (набора чисел), сумма ряда; n – количество этапов маршрутизации алгоритма диагностики (каждый этап – это осмотр врачом-специалистом с проведением медицинских вмешательств по его назначению или без него).

Каждый компонент, в свою очередь, раскрывается отдельной формулой. Затраты на ЛП-случаи при применении ИИ (формула 2) суммируют стоимость необоснованных осмотров и вмешательств по пациентам с ДНО, ошибочно направленным как подозрительные на ЗНО, отдельно для потоков с вероятностью не ниже и ниже порога:

Формула 2

$$AC_{FP}^{ab} = I^{ab \geq t} \times DMC^{ab \geq t} + \sum_{i=1}^{I-1} (I_{ab \geq t}^i \times DMC_{ab \geq t}^i) + I^{ab < t} \times DMC^{ab < t} + \sum_{i=1}^{W-1} (I_{ab < t}^w \times DMC_{ab < t}^w), \quad (11)$$

где AC_{FP}^{ai} – предотвратимые ФЗ на диагностику ЛП случаев ЦЗ при применении моделей ИИ (руб.); $N_{FP \geq t}$ – количество ЛП случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации; $DMC_{FP \geq t}$ – прямые медицинские ФЗ на диагностику ЛП случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации (руб.); $N_i^{FP \geq t}$ – количество ЛП случаев ЦЗ на i -м этапе маршрутизации, при которых вероятность ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации (определяется в виде конкретного количества, если оно известно, либо рассчитывается как произведение количества ЛП случаев ЦЗ предыдущего этапа диагностики на долю ЛП случаев ЦЗ при диагностике врачом-специалистом на данном этапе маршрутизации); $DMC_i^{FP \geq t}$ – прямые медицинские ФЗ на диагностику на i -м этапе маршрутизации для ЛП случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации (руб.); $N_{FP < t}$ – количество ЛП случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации; $DMC_{FP < t}$ – прямые медицинские ФЗ на диагностику для ЛП случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации (руб.); $N_m^{FP < t}$ – количество ЛП случаев ЦЗ на m -м этапе маршрутизации, при которых вероятность ЦЗ по заключению модели ИИ

была ниже порога алгоритма маршрутизации (определяется в виде конкретного количества, если оно известно, либо рассчитывается как произведение количества ЛП случаев ЦЗ предыдущего этапа маршрутизации на долю ЛП случаев ЦЗ при диагностике врачом-специалистом на данном этапе маршрутизации); $DMC_m^{FP< t}$ – прямые медицинские ФЗ на диагностику ЦЗ на m-м этапе маршрутизации для ЛП случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации (руб.); Σ – сумма (набора чисел), сумма ряда; n – количество этапов маршрутизации алгоритма диагностики (каждый этап – это осмотр врачом-специалистом с проведением медицинских вмешательств по его назначению или без него).

Это ФЗ на обследование пациентов с заболеваниями (например, ДНО), которые программа ИИ ошибочно классифицировала как ЦЗ (например, ЗНО). Такие пациенты проходят необоснованные дорогостоящие медицинские вмешательства, например осмотры врачей-онкологов, биопсию с последующим прижизненным патолого-анатомическим исследованием. К примеру, пациент с невусом направлен по заключению модели ИИ на осмотр к врачу-онкологу с подозрением на меланому кожи, проведена биопсия с последующим прижизненным патолого-анатомическим исследованием, подтвержден доброкачественный характер новообразования.

Формула 3

Формула для расчета ФЗ на диагностику ИО случаев ЦЗ при применении ИИ:

$$AC_{TN}^{ai} = N_{TN< t} \times DMC_{TN< t} + \sum_{k=1}^n (N_k^{TN< t} \times DMC_k^{TN< t}) , \quad (12)$$

где AC_{TN}^{ai} – предотвратимые ФЗ на диагностику ИО случаев ЦЗ при применении ИИ (руб.); $N_{TN< t}$ – количество ИО случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации; $DMC_{TN< t}$ – прямые медицинские ФЗ на диагностику для ИО случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации (руб.); $N_k^{TN< t}$ – количество ИО случаев ЦЗ на k-м этапе

маршрутизации, при которых вероятность ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации (определяется в виде конкретного количества, если оно известно, либо рассчитывается как произведение количества ИО случаев ЦЗ предыдущего этапа маршрутизации на долю ЛП случаев диагностики ЦЗ врачом-специалистом на данном этапе маршрутизации); $DMS_k^{TN< t}$ – прямые медицинские ФЗ на диагностику на k-м этапе маршрутизации для ИО случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации (руб.); $\sum$ – сумма (набора чисел), сумма ряда; n – количество этапов маршрутизации алгоритма диагностики (каждый этап – это осмотр врачом-специалистом с проведением медицинских вмешательств по его назначению или без него).

Это ФЗ на обследование пациентов с заболеваниями (например, ДНО), которые программа ИИ правильно классифицировала как нецелевое заболевание (в данном примере – ДНО), но в соответствии с алгоритмом диагностической маршрутизации требуют подтверждения у врача-специалиста (например, врача-дерматовенеролога). Обычно это случаи с вероятностью наличия ЦЗ ниже порога алгоритма маршрутизации, но требующие дополнительной верификации. Например, модель ИИ оценила новообразование как ДНО с вероятностью 50%, пациент направлен на осмотр к врачу-дерматовенерологу для подтверждения доброкачественного характера новообразования.

Формула 4

Формула для расчета ФЗ на не диагностированные, «пропущенные» случаи ЦЗ при применении ИИ:

$$PAC_{ND}^{ai} = (N_{FN \geq t} + (\sum_{s=1}^n N_s^{FN < t}) + (\sum_{f=1}^n N_f^{TP \geq t}) + (\sum_{p=1}^n N_p^{TP < t})) \times CrP, \quad (13)$$

где PAC_{ND}^{ai} – потенциально предотвратимые ФЗ на все ЛО (недиагностированные, пропущенные) случаи ЦЗ при применении ИИ (руб.); $N_{FN \geq t}$ – количество ЛО случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была равна или выше порога алгоритма маршрутизации и которые не были направлены на

дополнительную диагностику (пропущены при применении программы ИИ); $N_s^{FN < t}$ – количество ЛО случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была ниже порога алгоритма маршрутизации и которые были направлены к врачам-специалистам, но не диагностированы (пропущены) на следующих этапах маршрутизации (определяется в виде конкретного количества, если оно известно, либо рассчитывается как произведение количества ИП случаев ЦЗ предыдущего этапа маршрутизации на долю ЛО случаев ЦЗ при диагностике врачом-специалистом на данном этапе маршрутизации); C_rP – ФЗ на одного пациента на лечение недиагностированных, пропущенных случаев ЦЗ (руб.); $N_f^{TP \geq t}$ – количество ИП случаев ЦЗ, при которых вероятность ЦЗ по заключению модели ИИ была равно или выше порога алгоритма маршрутизации и которые были направлены к врачам-специалистам, но не диагностированы (пропущены) на следующих этапах маршрутизации (определяется в виде конкретного количества, если оно известно, либо рассчитывается как произведение количества ИП случаев ЦЗ предыдущего этапа маршрутизации на долю ЛО случаев ЦЗ при диагностике врачом-специалистом на данном этапе маршрутизации); $N_p^{TP < t}$ – количество ИП случаев ниже порога алгоритма маршрутизации, которые были направлены к врачам-специалистам, но не диагностированы (пропущены) на следующих этапах диагностики (определяется в виде конкретного количества, если оно известно, либо рассчитывается как произведение количества ИП случаев ЦЗ предыдущего этапа маршрутизации на долю ЛО случаев ЦЗ при диагностике врачом-специалистом на данном этапе маршрутизации); $\sum$ – сумма (набора чисел), сумма ряда; n – количество этапов маршрутизации алгоритма диагностики (каждый этап – это осмотр врачом-специалистом с проведением медицинских вмешательств по его назначению или без него).

Это ФЗ (потери) от несвоевременного выявления ЗНО при диагностике с применением ИИ. К ним относятся случаи, когда модель ИИ не распознала ЦЗ (например, ЗНО) или когда даже при правильной классификации моделью ИИ на последующих этапах маршрутизации врачи-специалисты не смогли правильно

установить диагноз. Если правильный диагноз ЦЗ устанавливается на более поздних стадиях заболевания, то это существенно увеличивает стоимость лечения. Параметр рассчитывается как произведение количества всех недиагностированных, пропущенных случаев ЦЗ на взвешенные ФЗ на лечение одного недиагностированного или поздно диагностированного случая ЗНО (СрР).

Формула 5

Формула для расчета ФЗ на диагностику ЛП случаев ЦЗ без применения ИИ:

$$AC_{FP}^{nai} = \sum_{r=1}^n (N_r^{FPnai} \times DMC_r^{FPnai}) , \quad (14)$$

где AC_{FP}^{nai} – предотвратимые ФЗ на диагностику ЛП случаев ЦЗ без применения ИИ (руб.); N_r^{FPnai} – количество ЛП случаев ЦЗ на k-м этапе при диагностике врачом-специалистом без применения моделей ИИ (определяется в виде конкретного количества, если оно известно, либо рассчитывается как произведение количества ИО случаев ЦЗ предыдущего этапа маршрутизации на долю ЛП случаев ЦЗ при диагностике врачом-специалистом на данном этапе маршрутизации); DMC_r^{FPnai} – прямые медицинские ФЗ на диагностику на k-м этапе для ЛП случаев при диагностике врачом-специалистом без применения моделей ИИ (руб.); $\sum$ – сумма (набора чисел), сумма ряда; n – количество этапов маршрутизации алгоритма диагностики (каждый этап – это осмотр врачом-специалистом с проведением медицинских вмешательств по его назначению или без него).

Это ФЗ на обследование пациентов, у которых отсутствует ЦЗ, но ВОП/терапевт при диагностике без применения ИИ ошибочно заподозрил ЦЗ (например, ЗНО) и которые были направлены на осмотр к врачам-специалистам (дерматовенерологам, онкологам) для дополнительного обследования. Включает ФЗ на осмотры врачей-специалистов и другие медицинские вмешательства (биопсию с последующим прижизненным патолого-анатомическим исследованием). При традиционной диагностической маршрутизации без применения моделей ИИ доля необоснованных направлений к врачам-

специалистам, как правило, высокая из-за низкой специфичности диагностики ВОП/терапевтами на первичном этапе.

Формула 6

Формула расчета ФЗ на не диагностированные, «пропущенные» случаи ЦЗ при диагностике без применения ИИ:

$$PAC_{ND}^{nai} = \left(\sum_{j=1}^n N_j^{FNnai} \right) \times CrP, \quad (15)$$

где PAC_{ND}^{nai} – потенциально предотвратимые ФЗ на не диагностированные, «пропущенные» случаи ЦЗ при диагностике без применения ИИ (руб.); N_j^{FNnai} – количество ЛО случаев ЦЗ на j-м этапе при диагностике врачом-специалистом без применения моделей ИИ. Определяется в виде конкретного количества, если оно известно, либо рассчитывается как произведение количества ИП случаев ЦЗ предыдущего этапа маршрутизации на долю ЛО случаев ЦЗ при диагностике врачом-специалистом на данном этапе маршрутизации; CrP – ФЗ на одного пациента на лечение не диагностированных, «пропущенных» случаев ЦЗ (руб.); Σ – сумма (набора чисел), сумма ряда; n – количество этапов маршрутизации алгоритма диагностики (каждый этап – это осмотр врачом-специалистом с проведением медицинских вмешательств по его назначению или без него).

Это ФЗ (потери) от несвоевременного выявления ЦЗ (например, ЗНО) при диагностике без применения ИИ. Включает случаи, когда ВОП/терапевты не заподозрили ЦЗ (например, ЗНО) и не направили пациента на дальнейшее обследование. Рассчитывается как произведение количества пропущенных случаев ЦЗ на взвешенные ФЗ на лечение одного недиагностированного случая ЦЗ (CrP). При традиционной диагностической маршрутизации без применения моделей ИИ количество пропущенных случаев, как правило, очень высокая из-за более низкой чувствительности диагностики ВОП/терапевты на первичном этапе.

Результаты расчета для сценария с применением искусственного интеллекта

Расчеты для сценария с применением ИИ

Расчет AC_{FP}^{ai} (ФЗ на ЛП случаи) – формула 2

ЛП случаи с вероятностью $\geq 62\%$ (направлены к врачам-онкологам):

- $N_{FP \geq t} = 11$ случаев;
- $DMC_{FP \geq t}$ (осмотр врачом-онкологом) = 268,59 руб/случай;
- итого: $11 \times 268,59 = 2954,49$ руб.

ЛП случаи с вероятностью $\geq 62\%$ (направлены на биопсию с последующим прижизненным патолого-анатомическим исследованием):

- $N_i^{FP \geq t} = 2,42$ случая. Так как точное количество не известно, то 11 случаев, которые попали к врачам-онкологам, умножаются на 22% – долю ЛП случаев диагностики врачом-специалистом (доля ЛП случаев = 1 минус специфичность);
- $DMC_i^{FP \geq t}$ (биопсия 5-й категории сложности, взвешенная стоимость) = 6 395,20 руб/случай;
- итого: $2,42 \times 6\,395,20 = 15\,476,38$ руб.

ЛП случаи с вероятностью $< 62\%$ (направлены к врачам-дерматовенерологам):

- $N_{FP < t} = 8$ (ДНО направлены к врачам-дерматовенерологам);
- $DMC_{FP < t}$ (осмотр врачом-дерматовенерологом и дерматоскопия) = 466,76 руб/случай;
- итого: $8 \times 466,76 = 3734,08$ руб.

ЛП случаи с вероятностью $< 62\%$ (направлены к онкологу):

- $N_m^{FP < t}$ (1-й этап) = 1,82 случая (поскольку точное количество неизвестно, 8 случаев, которые попали к врачам-дерматовенерологам, умножаются на 22,8% – долю ЛП случаев диагностики врачом-специалистом; доля ЛП случаев составляет 1 минус специфичность);
- $DMC_m^{FP < t}$ (1-й этап, осмотр врачом-онкологом) = 268,59 руб/случай;

– итого: $1,82 \times 268,59 = 488,83$ руб.

– $N_m^{FP< t}$ (2-й этап) = 0,6 случая (поскольку точное количество неизвестно, 1,82 случая, которые направлены на биопсию с последующим прижизненным патолого-анатомическим исследованием, умножаются на 33,14% – долю ЛП случаев диагностики врачом-специалистом; доля ЛП случаев составляет 1 минус специфичность);

– $DMC_m^{FP< t}$ (2-й этап, (2-й этап, биопсия 5-й категории сложности, взвешенная стоимость) = 6 395,20 руб/случай;

– итого: $0,6 \times 6\,395,20 = 3\,837,12$ руб.

Всего: $AC_{FP}^{ai} = 2954,49 + 15\,476,38 + 3734,08 + 488,83 + 3\,837,12 = 26\,490,90$ руб.

Расчет AC_{TN}^{ai} (ФЗ на ИО случаи) – формула 3

– ИО случаи с вероятностью <62% (направлены к врачам-дерматовенерологам):

– $N_{TN< t} = 7$;

– $DMC_{TN< t}$ (осмотр врачом-дерматовенерологом + дерматоскопия) = 466,76 руб/случай;

– итого: $7 \times 466,76 = 3267,32$ руб.

ИО случаи с вероятностью <62% (направлены к врачам-онкологам):

– $N_k^{TN< t}$ (1-й этап) = 1,59 случая (поскольку точное количество неизвестно, то 7 случаев, которые попали к врачам-дерматовенерологам, умножаются на 22,8% – долю ЛП случаев диагностики врачом-специалистом; доля ЛП случаев составляет 1 минус специфичность);

– $DMC_k^{TN< t}$ (1-й этап, осмотр врачом-онкологом) = 268,59 руб/случай;

– итого: $1,59 \times 268,59 = 427,05$ руб.;

– $N_k^{TN< t}$ (2-й этап) = 0,52 случая (поскольку точное количество неизвестно, 1,59 случая, которые направлены на биопсию с последующим прижизненным патолого-анатомическим исследованием, умножаются на 33,14% – долю ЛП

случаев диагностики врачом-специалистом; доля ЛП случаев составляет 1 минус специфичность);

– $DMC_k^{TN<t}$ (2-й этап, биопсия 5-й категории сложности, взвешенная стоимость) = 6 395,20 руб/случай;

– итого: $0,52 \times 6\,395,20 = 3\,325,50$ руб.

Всего: $AC_{TN}^{ai} = 3267,32 + 427,05 + 3\,325,50 = 7\,019,87$ руб.

Расчет PAC_{ND}^{ai} (ФЗ на ЛО случаи) – формула 4

ЛО случаи $\geq 62\%$ (пропущены при диагностике с применением ИИ):

– $N_{FN \geq t} = 1$ случай.

ЛО случаи $< 62\%$ (пропущены ИИ и направлены к врачам-дерматовенерологам).

– $N_s^{FN<t}$ (1-й этап) = 0,2 случая (поскольку точное количество неизвестно, 1 случай, который попал к врачу-дерматовенерологу, умножается на 20,8% – долю ЛО случаев диагностики врачом-специалистом; доля ЛО случаев составляет 1 минус чувствительность);

– $N_s^{FN<t}$ (2-й этап) = 0,26 случая (поскольку точное количество неизвестно, 0,8 случая, который попали к врачам-онкологам от врача-дерматовенеролога, умножаются на 33,14% – долю ЛО случаев диагностики врачом-специалистом; доля ЛО случаев составляет 1 минус чувствительность).

ИП случаи $\geq 62\%$ (правильно направлены, но пропущены врачами-онкологами):

– $N_f^{TP \geq t} = 10,27$ случая (поскольку точное количество не известно, 31 случай, который попал к врачам-онкологам, умножается на 33,14% – долю ЛО случаев диагностики врачом-специалистом; доля ЛО случаев составляет 1 минус чувствительность).

ИП случаи $< 62\%$ (правильно направлены, но пропущены дерматовенерологом):

– $N_p^{TP<1}$ (1-й этап) = 1,2 случая (поскольку точное количество неизвестно, 6 случаев, которые попали к врачам-дерматовенерологам, умножаются на 20,8% – долю ЛО случаев диагностики врачом-специалистом; доля ЛО случаев составляет 1 минус чувствительность);

– $N_p^{TP<1}$ (2-й этап) = 1,59 случая (поскольку точное количество неизвестно, 4,8 случаев, которые попали к врачам-онкологам от врачей-дерматовенерологов, умножаются на 33,14% – долю ЛО случаев диагностики врачом-специалистом; доля ЛО случаев составляет 1 минус чувствительность);

$$- \text{СрР} = (0,852 \times 403\,021,60) + (0,148 \times 2\,787\,014,87) = 755\,852,60 \text{ руб.}^3$$

Всего: $\text{РАС}_{\text{ND}}^{\text{ai}} = (1 + 0,2 + 0,26 + 10,27 + 1,2 + 1,59) \times 755\,852,60 = 10\,974\,979,75$ руб.

Общий расчет

Таким образом, получаем общее значение для сценария с применением ИИ:

$$\text{ТС}_{\text{ai}} = 26\,490,90 + 7\,019,87 + 10\,974\,979,75 = 11\,008\,490,52 \text{ руб.}$$

Расчеты для сценария без применения искусственного интеллекта

Расчет $\text{АС}_{\text{FP}}^{\text{nai}}$ (ФЗ на ЛП случаи) – формула 5

Всего ДНО – 191 случая.

После первичного осмотра ВОП/терапевты направляют пациентов с подозрением на ЗНО к врачам-дерматовенерологам. По данным литературы [70], 50,6% пациентов с ДНО (доля ЛП случаев составляет 1 минус специфичность) направляются к врачам-дерматовенерологам.

³ Представленный расчет СрР проводился на основании таблиц 2,3,4 и 6, является приблизительной оценкой ФЗ на лечение пропущенных случаев меланомы и используется для демонстрации методики расчета для оценки МЭЭ. Для каждой стадии ЦЗ рассчитана средняя арифметическая стоимость лечения на основании тарифов ОМС. При наличии данных о фактической частоте применения различных методов лечения целесообразно использовать более точные взвешенные оценки. В данном расчете учтены только прямые медицинские затраты на лечение меланомы согласно тарифам ОМС. При наличии более полной информации о ФЗ скорректированные данные следует включить в расчет.

Количество ДНО, направленных к врачам-дерматовенерологам: N_r^{FPnai} (1-й этап) = $191 \times 0,506 = 96,64$.

DMC_r^{FPnai} (1-й этап, осмотр врачом-дерматовенерологом и дерматоскопия) = $96,64 \times 466,76 = 45\,107,68$ руб.

Врачи-дерматовенерологи направляют пациентов с подозрением на ЗНО к онкологу. По данным литературы [70], 22,8% пациентов с ДНО (доля ЛП случаев составляет 1 минус специфичность) направляются к врачу-онкологу.

Количество ДНО, направленных к врачам-онкологам: N_r^{FPnai} (2-й этап) = $96,64 \times 0,228 = 22$.

DMC_r^{FPnai} (2-й этап, осмотр врачом-онкологом) = $22 \times 268,59 = 5908,98$ руб.

Врачи-онкологи направляют пациентов с подозрением на ЗНО на биопсию с последующим прижизненным патолого-анатомическим исследованием. По данным литературы [70], 22% пациентов с ДНО (доля ЛП случаев составляет 1 минус специфичность) необоснованно направляются на биопсию с последующим прижизненным патолого-анатомическим исследованием.

Количество ДНО, направленных к врачам-онкологам: N_r^{FPnai} (3-й этап) = $22 \times 0,22 = 4,84$.

DMC_r^{FPnai} (3-й этап, биопсия 5-й категории сложности, взвешенная стоимость) = $4,84 \times 6\,395,20 = 30\,952,77$ руб.

Всего: $AC_{FP}^{nai} = 45\,107,68 + 5\,908,98 + 30\,952,77 = 81\,969,43$ руб.

Расчет PAC_{ND}^{nai} (ФЗ на ЛО случаи) – формула 6

Всего ЗНО – 39 случаев.

По данным литературы [70], ВОП/терапевты пропускают ЗНО в 55,4% случаев (доля ЛО случаев составляет 1 минус чувствительность).

Количество недиагностированных, пропущенных ЗНО: N_j^{FNnai} (1-й этап) = $39 \times 0,554 = 21,60$.

К врачам-дерматовенерологам от ВОП/терапевтов направляются 17,4 ИП случаев ($39 \times 44,6\%$). По данным литературы [70], врачи-дерматовенерологи пропускают ЗНО в 20,8% случаев (доля ЛО случаев составляет 1 минус чувствительность).

Количество недиагностированных, пропущенных случаев ЗНО: N_j^{FNnai} (2-й этап) $= 17,4 \times 0,208 = 3,6$.

К врачам-онкологам от врачей-дерматовенерологов направляются 13,8 ИП случаев ($17,4 \times 0,792$). По данным литературы [70], врачи-онкологи пропускает случаи ЗНО в 33,14% случаев (доля ЛО случаев составляет 1 минус чувствительность).

Количество недиагностированных, пропущенных случаев ЗНО: N_j^{FNnai} (3-й этап) $= 13,8 \times 0,3314 = 4,57$.

$$CpP = (0,852 \times 403\,021,60) + (0,148 \times 2\,787\,014,87) = 755\,852,60 \text{ руб}$$

Общий расчет

Таким образом, получаем общие значения для сценария без применения ИИ:

$$PAC_{ND}^{nai} = (21,60 + 3,6 + 4,57) \times 755\,852,60 = 22\,501\,731,90 \text{ руб.},$$

$$TC_{nai} = 81\,969,43 + 22\,501\,731,90 = 22\,583\,701,33 \text{ руб.}$$

Расчет коэффициента соотношения финансовых затрат

На основании полученных данных расчет коэффициента отношения ФЗ на диагностику ЦЗ с применением заключений моделей ИИ и ФЗ на диагностику ЦЗ без применения моделей ИИ (формула 1) выглядит так:

$$R_{TC} = 11\,008\,490,52 / 22\,583\,701,33 = 0,49.$$

Полученное значение означает, что на каждый рубль затрат при традиционной диагностике приходится лишь 0,49 руб. затрат при диагностике с применением ИИ; иными словами, общие затраты снижаются на 51 %. Принципиально, что этот показатель учитывает не только предотвратимые затраты на вмешательства, но и потенциально предотвратимые потери от несвоевременного

выявления ЗНО, а включение компонента СрР в обе части формулы обеспечивает корректную сопоставимость сценариев с разной структурой затрат.

Комплексная оценка. Квадрантная матрица медико-экономической эффективности

Для наглядного представления результата разработана квадрантная матрица МЭЭ, в которой по оси абсцисс откладывается коэффициент предотвратимых затрат R_{TC_AC} , а по оси ординат – коэффициент потенциально предотвратимых затрат R_{TC_PAC} . Первый рассчитывается как отношение предотвратимых затрат с ИИ к предотвратимым затратам без ИИ, второй, как отношение потенциально предотвратимых затрат с ИИ к аналогичным без ИИ:

$$R_{TC\_AC} = (AC_{FP}^{ai} + AC_{TN}^{ai}) / AC_{FP}^{nai}. \quad (16)$$

R_{TC_PAC} рассчитывается как отношение потенциально предотвратимых ФЗ (на пропущенные случаи) с применением ИИ к потенциально предотвратимых ФЗ без применения ИИ:

$$R_{TC\_PAC} = PAC_{ND}^{ai} / PAC_{ND}^{nai}. \quad (17)$$

Из этих данных следует, что снижение редотвратимых и потенциально предотвратимых ФЗ при использовании ИИ составляет соответственно:

$$(1 - R_{TC\_AC}) \times 100\% = (1 - 0,41) \times 100\% = 59,0\%,$$

$$(1 - R_{TC\_PAC}) \times 100\% = (1 - 0,49) \times 100\% = 51,0\%.$$

Для рассматриваемого примера $R_{TC\_AC} = 0,41$ и $R_{TC\_PAC} = 0,49$, что соответствует снижению предотвратимых затрат на 59,0 % и потенциально предотвратимых – на 51,0 %. Сводные результаты приведены в Таблице 52.

Таблица 52 – Результаты расчета финансовых затрат на диагностику с применением технологий ИИ и без него

Параметр	С применением ИИ, руб.	Без применения ИИ, руб.	Коэффициент
Предотвратимые затраты: ЛП-случаи	26 490,90	–	
Предотвратимые затраты: ИО-случаи	7019,87	–	
Предотвратимые затраты: всего	33 510,77	81 969,43	$RTCAC = 0,41$
Потенциально предотвратимые затраты: всего	10 974 979,75	22 501 731,90	$RTCPAC = 0,49$
Примечание: ЛП – ложноположительные; ИО – истинно отрицательные случаи; ИИ – искусственный интеллект; R_{TC_AC} – коэффициент предотвратимых затрат; R_{TC_PAC} – коэффициент потенциально предотвратимых затрат			

Матрица делится на четыре квадранта относительно точки паритета (1,0; 1,0). Оптимальной является зона с одновременным снижением обеих категорий затрат: технология, попадающая в нее, обладает высокой МЭЭ и подлежит приоритетному внедрению. Противоположная зона, где обе категории затрат возрастают, свидетельствует о нецелесообразности внедрения. Две промежуточные зоны (снижение одной категории при росте другой) требуют дополнительных исследований и оптимизации порога. Точка сценария с применением Derma Onko Check ($R_{TC\_AC} = 0,41$; $R_{TC\_PAC} = 0,49$) располагается в оптимальном квадранте, что подтверждает значительное превосходство ИИ-диагностики над стандартной и правильность выбранного порога маршрутизации (Рисунок 77).

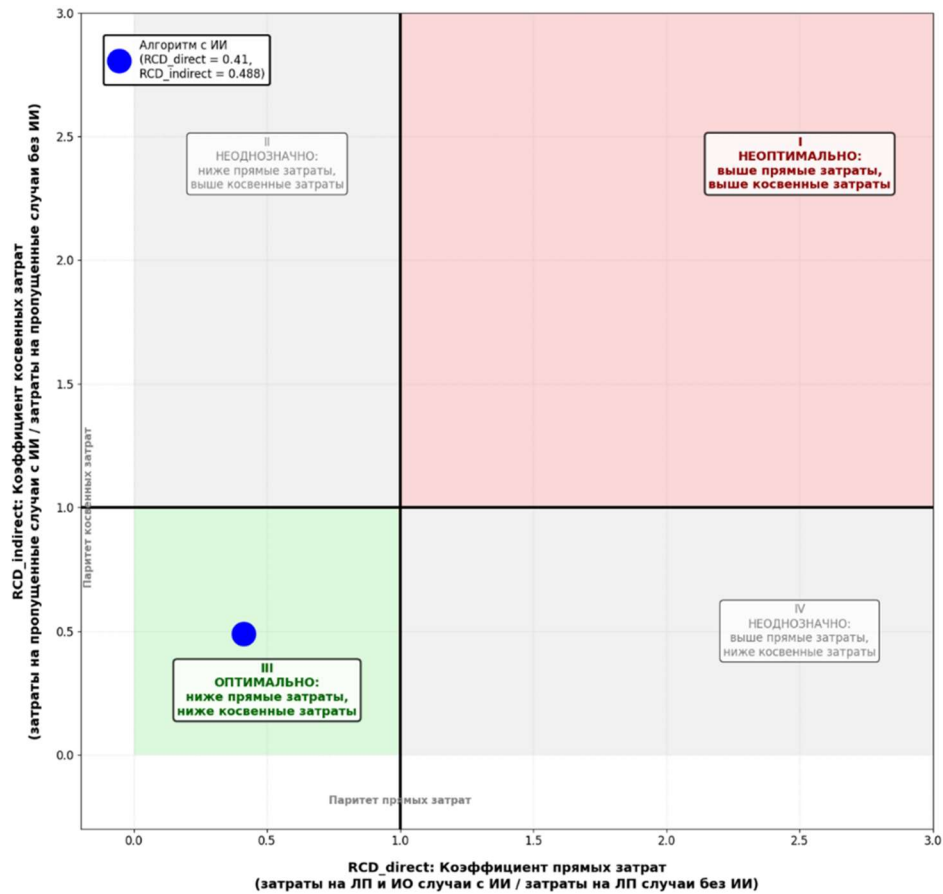


Рисунок 77 – Плоскостная визуализация показателей относительной медико-экономической эффективности алгоритма маршрутизации на основе заключений программы Derma Onko Check (ось абсцисс – R_{TC_AC} , ось ординат – R_{TC_PAC})

Значение результатов и ограничения

Результаты подтверждают гипотезу о том, что встраивание заключений моделей ИИ в многоэтапную маршрутизацию позволяет значительно снизить затраты при одновременном улучшении качества помощи. Определяющим фактором стало сокращение числа пропущенных случаев ЗНО за счет высокой чувствительности ИИ при выбранном пороге, что уменьшает потенциально предотвратимые затраты на диагностику и лечение поздних стадий. В сопоставлении с существующими подходами предложенный способ развивает концепцию человеческого капитала здоровья и регулирующего эффекта системы здравоохранения, вводя симметричный учет потерь от пропущенных случаев, и дополняет методические рекомендации по клинико-экономическому анализу медицинских технологий квадрантной матрицей и коэффициентами R_{TC_AC} и R_{TC}

_РАС. В отличие от анализа эффективности структурных подразделений, ориентированного на внутренний уровень организации, настоящий способ оценивает системный эффект на уровне маршрутизации пациентов.

Прикладное значение работы состоит в обосновании приоритетного внедрения технологий ИИ в регионах с высокой нагрузкой на онкологическую службу, где потенциальная экономия может достигать сотен миллионов рублей ежегодно при сохранении или улучшении качества помощи. Ограничения исследования связаны с относительно небольшой выборкой для валидации ($n=230$), с приближенным характером оценки CrP (средние арифметические по стадиям без взвешивания по реальной частоте методов лечения), с учетом только прямых затрат без полного анализа косвенных потерь и показателя стоимости года жизни с поправкой на качество (QALY), а также с заимствованием показателей точности врачей из литературы. Перспективы включают проспективные многоцентровые исследования, интеграцию нескольких моделей ИИ в единой программе, разработку взвешенных оценок CrP на основе регистровых данных и распространение анализа на другие социально значимые заболевания.

Разработанный способ расчета МЭЭ диагностики с применением алгоритмов на основании заключений моделей ИИ представляет собой комплексный инструмент объективной оценки экономической целесообразности их внедрения. Его ключевое методологическое решение – симметричный учет не только предотвратимых затрат на вмешательства, но и потенциально предотвратимых потерь от несвоевременного выявления ЦЗ через включение компонента CrP в обе части коэффициента соотношения затрат, что обеспечивает корректную сопоставимость сценариев с разной структурой затрат. Валидация на данных программы Derma Onko Check показала высокую МЭЭ применения ИИ: снижение общих затрат на 51 %, предотвратимых затрат – на 59,0 % ($R_{ТС\_АС} = 0,41$) и потенциально предотвратимых – на 51,0 % ($R_{ТС\_РАС} = 0,49$). Разработанная квадрантная матрица позволяет наглядно интерпретировать результаты и принимать обоснованные управленческие решения. Способ универсален и

применим для оценки МЭЭ диагностики любого ЦЗ при наличии данных о точности программ и врачей и о стоимости медицинских услуг.

3.12.2. Медико-экономическая эффективность диагностики на региональном уровне

Оценка МЭЭ на уровне отдельного учреждения, выполненная в разделе 7.3, отвечает на вопрос об экономической целесообразности внедрения ИИ в конкретной медицинской организации. Однако управленческие решения о цифровой трансформации первичной помощи принимаются на уровне региональной системы здравоохранения, а значит, требуют оценки совокупного эффекта в масштабах субъекта. Настоящий раздел переносит анализ на этот уровень. Его актуальность обусловлена двумя обстоятельствами. Во-первых, традиционная диагностика новообразований кожи в первичном звене имеет системные ограничения: чувствительность выявления меланомы врачами общей практики (ВОП) и терапевтами составляет лишь 44,6 %, то есть более половины случаев ЗНО кожи остаются нераспознанными на первом этапе, что влечет позднюю диагностику. Во-вторых, существующие подходы к оценке МЭЭ не в полной мере учитывают специфику ИИ – многоэтапность маршрутизации, различную структуру затрат на разных этапах и долгосрочные последствия пропущенных случаев. Целью раздела является разработка и валидация способа оценки МЭЭ диагностики ЦЗ с применением программ на основе ИИ в сравнении с традиционной диагностикой на региональном уровне.

Принцип обоснованных и необоснованных затрат

Методологической основой регионального анализа служит разграничение обоснованных и необоснованных затрат на выявление одного случая ЦЗ. Необоснованными считаются затраты, которых можно было бы избежать при более точной и ранней диагностике: расходы на направление к специалистам пациентов

с ДНО, имеющими клинико-дерматоскопические признаки доброкачественности, и на морфологическое исследование таких образований, а также расходы на диагностику и лечение пропущенных ЗНО, выявленных впоследствии на поздних стадиях. Обоснованные затраты соответствуют фактически выявленным случаям ЦЗ и являются необходимыми расходами системы здравоохранения. Как и в разделе 7.3, к необоснованным относятся исключительно затраты на диагностику образований с признаками доброкачественности; обследование образований с признаками злокачественности или предраковыми изменениями обоснованно вне зависимости от окончательного морфологического заключения.

Ключевым показателем МЭЭ служит коэффициент $R = RC / UC$, представляющий собой отношение обоснованных затрат (RC) к необоснованным (UC). Значение R больше единицы означает преобладание обоснованных затрат, то есть эффективность диагностики; значение меньше единицы указывает на преобладание необоснованных затрат и неэффективность. Такое представление удобно тем, что позволяет напрямую сравнивать альтернативные диагностические стратегии независимо от масштаба выборки. Система расчетных формул строилась на трех принципах: каскадной маршрутизации (на каждом этапе часть случаев ЦЗ остается невыявленной вследствие ограниченной чувствительности врачей), разграничения обоснованных и необоснованных затрат и региональной экстраполяции перехода от удельных показателей к совокупным затратам в масштабах субъекта на основе официальных данных о заболеваемости.

Медико-экономическая эффективность диагностики с применением алгоритмов маршрутизации на основании заключений моделей искусственного интеллекта на региональном уровне

В качестве модельного ЦЗ выбраны ЗНО кожи, прежде всего меланома, что обусловлено рядом факторов. Меланома входит в перечень социально значимых заболеваний и характеризуется высоким потенциалом злокачественности и ранним метастазированием, что определяет ее значимость для общественного здоровья .

Диагностика новообразований кожи представляет собой классический пример многоэтапной маршрутизации (первичное звено → дерматовенерологи → онкологи), формирующей каскад потенциально предотвратимых затрат на каждом этапе. Поздняя диагностика ЗНО влечет резкий рост расходов на лечение, хирургические вмешательства, системную, таргетную и иммунотерапию, длительное наблюдение, а также потери Социального фонда России и потери валового продукта вследствие преждевременной смертности трудоспособного населения. Расчеты выполнены на тарифах ОМС г. Москвы, однако методология универсальна и адаптируется к любому региону подстановкой локальных тарифов и показателей заболеваемости.

Сокращение сроков установления диагноза

Помимо стоимостных показателей, применение ИИ влияет на длительность диагностического пути. Моделирование по данным диагностики меланоцитарных опухолей программой Derma Onko Check (n=230, порог 62 %) показало, что при вероятности ЗНО не ниже порога пациент направляется напрямую к онкологу (одноэтапная маршрутизация), а при вероятности ниже порога, к дерматовенерологу с последующим направлением к онкологу при необходимости. Средняя длительность установления заключительного клинического диагноза по группам приведена в Таблице 53, а сводные результаты сравнения двух сценариев в Таблице 54.

Применение программы сокращает среднюю длительность установления заключительного диагноза с 18,3 до 7,8 суток, то есть на 10,5 суток. Одновременно резко снижается доля необоснованных направлений: к дерматовенерологам с 40,6 до 6,9 %, к онкологам с 22,0 до 7,6 %. Сокращение сроков диагностики имеет прямое клиническое значение, поскольку при меланоме промедление напрямую связано с прогрессированием заболевания и ростом стоимости лечения.

Таблица 53 – Длительность установления заключительного клинического диагноза с применением программы Derma Onko Check при диагностике меланоцитарных опухолей кожи (n=230)

Результат (тип опухоли, точность заключения)	Вероятность ≥62 %		Вероятность <62 %			Средняя длительность, сут
	n (%)	Длительность, сут	n (%)	Длительность, сут	Специальность врача*	
ИП (ЗНО, верно)	31 (13,5)	20	6 (2,6)	34	—	22,3
ЛП (ДНО, неверно)	11 (4,8)	20 (онколог)	8 (3,5)	15 (дерматовенеролог)	—	17,9
ИО (ДНО, верно)	165 (71,7)	1	7 (3,0)	15	Дерматовенеролог	1,6
ЛО (ЗНО, неверно)	1 (0,4)	—	1 (0,4)	34	—	34,0
Всего	208 (90,4)		22 (9,6)			6,5

Примечание: * специальность врача, к которому направлялись пациенты от ВОП/терапевта. Общая средняя длительность рассчитана как взвешенное среднее по всем пациентам

Таблица 54 – Сводные результаты для программы Derma Onko Check

Сценарий диагностики	Средняя длительность установления диагноза, сут			Доля необоснованных направлений (от всех ДНО), %		Экономия, сут
	ЗНО	ДНО	Все ЗНО и ДНО	к дерматовенерологам	к онкологам	
С применением программы ИИ	22,0	3,4	7,8	6,9	7,6	10,5
Без применения программы ИИ	34,0	15,0	18,3 (взвешенно)	40,6	22,0	—

Примечание: доли необоснованных направлений к дерматовенерологам рассчитаны как доля случаев ДНО с вероятностью <62 %; к онкологам – как доля ЛП с вероятностью ≥62 % от общего числа ДНО

Расчет числа выявленных случаев и удельных затрат

Число выявленных случаев ЦЗ при каждом сценарии рассчитывается по той же каскадной модели потерь, что и в разделе 7.2, но применительно к региональным показателям чувствительности врачей (онколог – 66,86 %, дерматовенеролог – 79,2 %, ВОП/терапевт – 44,6 %). При диагностике с применением ИИ выявляемость складывается из двух маршрутов – одноэтапного (случаи с вероятностью не ниже порога направляются прямо к онкологу) и двухэтапного (случаи ниже порога проходят через дерматовенеролога, а затем онколога):

С применением ИИ

Для расчета количества выявленных случаев ЦЗ (меланома кожи С43) при применении программы ИИ используется формула [70]:

$$N_{Dai} = N_{TP \geq t} \times \prod_{k=1}^n SD_k^{TP \geq t} + (N_{FN < t} + N_{TP < t}) \times \prod_{j=1}^n SD_j^{FN < t + TP < t}. \quad (18)$$

$N_{TP \geq t} = 31$ (ИП случаи с вероятностью $\geq 62\%$), $N_{FN < t} = 1$ (ЛО случаи с вероятностью $< 62\%$), $N_{TP < t} = 6$ (ИП случаи с вероятностью $< 62\%$).

Расчет для ИП случаев, при которых вероятность ЦЗ по заключению модели ИИ была больше или равна порогу алгоритма маршрутизации ($\geq 62\%$) (направлены к онкологам):

- количество этапов маршрутизации – 1;
- $SD_k^{TP \geq t} = 66,86\%$ (чувствительность диагностики онкологом);
- количество выявленных случаев ЦЗ: $31 \times 0,6686 = 20,73$.

Расчет для ЛО и ИП случаев, при которых вероятность ЦЗ по заключению модели ИИ была меньше порога алгоритма маршрутизации ($< 62\%$) (направлены к дерматовенерологам, затем к онкологам):

- количество этапов маршрутизации – 2;
- этап 1 (дерматовенеролог): $SD_j^{FN < t + TP < t} = 79,2\%$ (чувствительность диагностики дерматовенерологом);

– этап 2 (онколог): $SD_j^{FN<t+TP<t} = 66,86\%$ (чувствительность диагностики онкологом);

– количество случаев, направленных на этап 2: $(1 + 6) \times 0,792 = 5,54$;

– количество выявленных случаев ЦЗ после осмотра онкологом: $5,54 \times 0,6686 = 3,70$.

Итого общее количество выявленных случаев ЦЗ при диагностике с применением ИИ в регионе: $N_{rDai} = 20,73 + 3,70 = 24,43$.

Без применения ИИ

Для расчета количества выявленных случаев ЦЗ (меланома кожи С43) при диагностике без применения программы ИИ (N_{rDnai}) используется формула [70]:

$$N_{Dnai} = N \times \prod_{m=1}^n SD_m^{nai}. \quad (19)$$

$N = 39$ (общее количество случаев ЦЗ в выборке), $n = 3$ (количество этапов маршрутизации).

Этап 1 (ВОП/терапевт):

– $SD_m^{nai} = 44,6\%$ (чувствительность диагностики ВОП/терапевтом);

– количество случаев, направленных к дерматовенерологу: $39 \times 0,446 = 17,39$;

– количество случаев ЦЗ, пропущенных (недиагностированных) ВОП/терапевтом: $39 \times 0,554 = 21,61$.

Этап 2 (дерматовенеролог):

– $SD_m^{nai} = 79,2\%$ (чувствительность диагностики дерматовенерологом);

– количество случаев, направленных к онкологу: $17,39 \times 0,792 = 13,77$;

– количество случаев ЦЗ, пропущенных (недиагностированных) дерматовенерологом: $17,39 \times 0,208 = 3,62$.

Этап 3 (онколог):

– $SD_m^{nai} = 66,86\%$ (чувствительность диагностики онкологом);

– количество выявленных случаев ЦЗ: $13,77 \times 0,6686 = 9,21$;

– количество случаев ЦЗ, пропущенных (недиагностированных) онкологом: $13,77 \times 0,3314 = 4,56$.

Итого общее количество выявленных случаев ЦЗ при диагностике без применения ИИ: $N_{rDnai} = 39 \times 0,446 \times 0,792 \times 0,6686 = 9,21$.

Также расчет количества выявленных случаев ЦЗ можно выполнить с помощью программы ЭВМ.

Экстраполяция на региональный уровень

По данным официальной статистики, в г. Москве в 2024 г. под диспансерное наблюдение с впервые установленным диагнозом меланомы кожи (С43) было взято 1470 больных.

С применением ИИ

Расчет необоснованных ФЗ (UC_{rai} , UC_{nai}) и обоснованных Фз (RC_{rai} , RC_{nai}) на выявление одного случая ЦЗ (меланома кожи С43) при диагностике с применением ИИ в регионе ($N = 39$, $CpP = 755\,852,60$ руб.):

- $TC_{ai} = 11\,008\,490,52$ руб.;
- $N_{Dai} = 24,43$ случая;
- $UC_{rai} = 11\,008\,490,52 / 39 = 282\,268,98$
- $NDai = 24,43$ случая;
- $RC_{rai} = (755\,852,60 \times 24,43) / 39 = 473\,473,82$
- $Rrai = 473\,473,82 / 282\,268,98 = 1,7$ (обоснованные ФЗ преобладают над необоснованными ФЗ)

Без применения ИИ

Расчет ФЗ на выявление одного случая ЦЗ (меланома кожи С43) при диагностике без применения ИИ:

- $TC_{nai} = 22\,583\,701,33$ руб.;
- $N_{Dnai} = 9,21$ случая;
- $UC_{nai} = 22\,583\,701,33 / 39 = 579\,069,26$
- $NDnai = 9,21$ случая;
- $RC_{nai} = (755\,852,60 \times 9,21) / 39 = 178\,497,49$

– $R_{nai} = 178\,497,49 / 579\,069,26 = 0,31$ (необоснованные ФЗ превышают обоснованные ФЗ)

Сравнительный анализ двух сценариев выявил существенные различия в структуре финансовых затрат. При переходе от традиционной маршрутизации к диагностике с применением ИИ обоснованные затраты на выявление одного случая целевого заболевания снижались с 579 069,26 до 282 268,98 руб., то есть на 296 800,28 руб., или на 51,3 %. Коэффициент снижения обоснованных затрат составил 2,05: при традиционной маршрутизации обоснованные расходы в 2,05 раза превышали аналогичный показатель при диагностике с применением ИИ.

Показательными оказались и значения интегрального коэффициента соотношения обоснованных и обоснованных затрат. При диагностике с применением ИИ он составил $R_{rai} = 1,7$, что свидетельствует о преобладании обоснованных затрат: финансовые ресурсы направлялись преимущественно на выявление реальных случаев злокачественных новообразований. При традиционной диагностике коэффициент составил $R_{nai} = 0,31$, отражая обратное соотношение, при котором большая часть расходов системы здравоохранения приходилась на затраты, поддающиеся минимизации.

Различие в структуре расходов прослеживается и в долевом распределении. При диагностике с применением ИИ обоснованные затраты составляли 473 473,82 руб., обоснованные 282 268,98 руб. на один случай, то есть на долю обоснованных приходилось около 62,6 % суммарных затрат. В сценарии без применения ИИ соотношение было противоположным: обоснованные затраты составляли 178 497,49 руб., обоснованные 579 069,26 руб., в результате чего доля обоснованных затрат достигала приблизительно 76,6 %. Полученные данные отражают ключевое экономическое преимущество диагностики с применением ИИ: при сопоставимом объеме ресурсов они распределяются преимущественно в пользу выявления истинных случаев заболевания. Иными словами, при равных суммарных затратах применение ИИ обеспечивает выявление большего числа случаев целевого заболевания. Значение $R_{rai} = 1,7$ свидетельствует, таким образом, не только о преобладании обоснованных затрат, но и о качественно иной

структуре расходов, при которой ресурсы системы здравоохранения перераспределяются в пользу результативной диагностики.

Установленная асимметрия затрат между сценариями обусловлена различным числом пропущенных случаев злокачественных новообразований. При традиционной маршрутизации их количество существенно выше вследствие ограниченной чувствительности диагностики на первичном этапе, что приводит к выявлению заболевания на поздних стадиях и обуславливает необходимость более объемной и дорогостоящей диагностики и лечения, определяя тем самым значительно более высокие затраты системы здравоохранения. Применение программы ИИ сокращает число пропущенных случаев и соответственно снижает потенциально предотвратимые затраты. Следовательно, различие в потенциально предотвратимых затратах между сценариями представляет собой не методологическую погрешность, а прямое измеримое следствие неодинаковой диагностической эффективности сравниваемых подходов.

Полный учет затрат на технологию

Приведенная система формул 1–4 учитывает только переменные затраты на диагностические услуги по тарифам ОМС. Для полной оценки МЭЭ необходимо учитывать и затраты на само внедрение и эксплуатацию системы ИИ как медицинского изделия (МИ) с ИИ: капитальные вложения, операционные расходы, амортизацию, а также косвенные, скрытые и транзакционные издержки, наряду с временной стоимостью денег. Для этого вводится формула расчета годовых эквивалентных затрат на основе стандартного метода аннуитетных платежей (capital recovery factor, CRF):

Формула для расчета годовых ФЗ на МИ с ИИ в регионе и/или медицинской организации:

– с расчетом эквивалентных ФЗ в год (формула 5А):

$$C_{AI}^{ac} = C_d \times \frac{r(1+r)^n}{(1+r)^n - 1} + C_o + C_{iht}, \quad (20)$$

– с расчетом годовых ФЗ с учетом линейной амортизации (формула 5Б):

$$C_{AI}^{ac} = \frac{C_d}{n} + C_o + C_{iht}, \quad (21)$$

где C_{AI}^{ac} – сумма ФЗ на МИ с ИИ в регионе (руб/год); C_d – суммарные капитальные (единовременные) ФЗ на внедрение МИ с ИИ в регионе, которые включают приобретение и/или разработку моделей ИИ, получение лицензий на МИ с ИИ, интеграцию с медицинскими информационными системами – электронными медицинскими картами и другим медицинским оборудованием, закупку и/или модернизацию сети, серверов или облачной инфраструктуры, начальное обучение работников использованию МИ с ИИ, сертификацию и валидацию (руб/год); r – ставка дисконтирования (норма дисконта); n – нормативный срок полезного использования МИ с ИИ (лет); C_o – ежегодные операционные (текущие) ФЗ на эксплуатацию, которые включают обновление моделей ИИ и МИ с ИИ, техническую поддержку, облачные вычисления, администрирование данных, резервное копирование, электроэнергию (руб/год); C_{iht} – косвенные (включают текущее обучение работников использованию МИ с ИИ, организационные изменения порядка использования МИ с ИИ), транзакционные (включают поиск и выбор поставщика ИИ, заключение контрактов, юридическое сопровождение, мониторинг и аудит системы), скрытые (включают экологические, социальные и другие аспекты) затраты (руб/год).

Значительная часть операционных процессов уже входит в стандартный функционал медицинских организаций и потому, как правило, не формирует дополнительных затрат сверх существующего финансирования. Для полной МЭЭ годовые эквивалентные затраты C_{AI}^{ac} суммируются с общими затратами на диагностику с ИИ и сравниваются с затратами традиционного сценария. Варьируя срок n , можно определить период эксплуатации, при котором применение МИ с ИИ наиболее экономически выгодно.

В настоящем исследовании затраты по формуле 5 не рассчитывались из-за отсутствия на момент работы верифицированных данных о капитальных вложениях, операционных и прочих издержках, а также регионально

обоснованных значений срока полезного использования и ставки дисконтирования. Тем не менее сама формула является методологическим вкладом и может применяться в проспективных исследованиях по мере накопления соответствующих экономических данных.

Для расчета эффективности может применяться приложение для карманных персональных и планшетных компьютеров на базе Android [91] (Рисунок 78-81).


МЭЭ диагностики - Калькулятор
 Медико-экономическая эффективность диагностики с ИИ

ИНСТРУКЦИЯ
 ФОРМУЛЫ ДЛЯ РАСЧЕТА


Параметры для ввода
СБРОС

Название целевого заболевания (ЦЗ)
 ЦЗ (целевые заболевания) — это заболевания, которые являются основной целью диагностического поиска и оцениваются по количеству выявленных и пропущенных случаев в рамках многоэтапной маршрутизации пациентов.

Меланома кожи C43

Порог алгоритма маршрутизации при диагностике с применением ИИ:
 1 вариант. Пациенты с вероятностью ЦЗ большей или равной значению порога алгоритма маршрутизации.
 2 вариант. Пациенты с вероятностью ЦЗ ниже значения порога алгоритма маршрутизации.

Порог алгоритма маршрутизации (%)
 62


Общие данные для диагностики
 Базовые показатели для расчета

Всего диагностированных случаев
 Из них случаев ЦЗ

число
 число

230
 39

* «Случай не ЦЗ» рассчитываются автоматически


ЭТАПЫ МАРШРУТИЗАЦИИ С ИИ
 Пациенты распределяются по двум вариантам маршрутизации в зависимости от значения порога алгоритма. Настройте этапы для каждого варианта маршрутизации отдельно. Этап — это осмотр врачом-специалистом с или без проведения медицинских вмешательств по его назначению.

Этап 1
 ПЕРВИЧНЫЙ ОСМОТР ВРАЧОМ-СПЕЦИАЛИСТОМ
 Заполните, если есть сведения об общем количестве пациентов (≥ и < порога алгоритма маршрутизации) до распределения по этапам.

▲ Вероятность ЦЗ ≥ порога алгоритма маршрутизации:
 N_TP≥t (ИП) N_FP≥t (ЛП) N_FN≥t (ЛО)
 число число число
 31 11 1

▼ Вероятность ЦЗ < порога алгоритма маршрутизации:
 N_TP<t N_FP<t N_TN<t N_FN<t
 число число число число
 6 8 7 1

▲ Пациенты с вероятностью ЦЗ ≥ порога алгоритма маршрутизации
 (N_TP≥t, N_FP≥t, N_FN≥t) — направляются к врачу-специалисту. Укажите этапы этой маршрутизации.

Этап 2
 ПАЦИЕНТЫ С ВЕРОЯТНОСТЬЮ ЦЗ ≥ ПОРОГА АЛГОРИТМА МАРШРУТИЗАЦИИ

Название этапа (врач-специалист)
 Врач-онколог

Режим ввода данных:
☒ По точности диагностики врача-специалиста
 ☐ По количеству пациентов

Чувствительность (%)
 Специфичность (%)

66,86
 78,00

Медицинские вмешательства этапа:
 1-я медицинская услуга - осмотр врачом-специалистом (для всех пациентов). Со 2-й — указывается медицинская услуга по направлению врача-специалиста текущего этапа.

Название медицинского вмешательства
 Осмотр врача-онколога
 21/200

Осмотр проводится всем пациентам этапа — количество учитывается автоматически

Стоимость (руб.)
 за 1 медицинскую услугу
 268,59

+ ДОБАВИТЬ

Дополнительные медицинские вмешательства по направлению врача-специалиста этапа 2:
 Количество = N_ЛП Этапа 2 (рассчитывается автоматически)

Название медицинского вмешательства
 Биопсия с прижизненным ПАО
 26/200

Стоимость (руб.)
 за 1 медицинскую услугу
 1431,22

+ ДОБАВИТЬ

ДОБАВИТЬ ЭТАП ДИАГНОСТИКИ С ИИ ПРИ ВЕРОЯТНОСТИ ЦЗ ≥ ПОРОГА АЛГОРИТМА МАРШРУТИЗАЦИИ

Рисунок 78 – Пример ввода и расчета результатов оценки МЭЭ диагностики с применением ИИ (количество случаев целевого заболевания выше или равно значению порога алгоритма маршрутизации)

▼ Пациенты с вероятностью ЦЗ < порога алгоритма маршрутизации
($N_{TP} < t$, $N_{FP} < t$, $N_{TN} < t$, $N_{FN} < t$) → направляются к врачу-специалисту. Укажите этапы этой маршрутизации.

Этап 2
ПАЦИЕНТЫ С ВЕРОЯТНОСТЬЮ ЦЗ < ПОРОГА АЛГОРИТМА МАРШРУТИЗАЦИИ

Название этапа (врач-специалист)

Врач-дерматовенеролог

Режим ввода данных:

☒ По точности диагностики врача-специалиста ☐ По количеству пациентов

Чувствительность (%) Специфичность (%)

Медицинские вмешательства этапа:
1-я медицинская услуга - осмотр врачом-специалистом (для всех пациентов). Со 2-й — указывается медицинская услуга по направлению врача-специалиста текущего этапа.

Название медицинского вмешательства

Осмотр врача-дерматовенеролога

31/200

☒ Осмотр проводится всем пациентам этапа — количество учитывается автоматически

Стоимость (руб.) за 1 медицинскую услугу

Название медицинского вмешательства

Дерматоскопия

13/200

Стоимость (руб.) за 1 медицинскую услугу

+ ДОБАВИТЬ

Дополнительные медицинские вмешательства по направлению врача-специалиста этапа 2:
Количество = $N_{ЛП}$ Этапа 2 (рассчитывается автоматически)

+ ДОБАВИТЬ

Этап 3
ПАЦИЕНТЫ С ВЕРОЯТНОСТЬЮ ЦЗ < ПОРОГА АЛГОРИТМА МАРШРУТИЗАЦИИ

Название этапа (врач-специалист)

Врач-онколог

Режим ввода данных:

☒ По точности диагностики врача-специалиста ☐ По количеству пациентов

Чувствительность (%) Специфичность (%)

Медицинские вмешательства этапа:
1-я медицинская услуга - осмотр врачом-специалистом (для всех пациентов). Со 2-й — указывается медицинская услуга по направлению врача-специалиста текущего этапа.

Название медицинского вмешательства

Осмотр врача-онколога

21/200

☒ Осмотр проводится всем пациентам этапа — количество учитывается автоматически

Стоимость (руб.) за 1 медицинскую услугу

+ ДОБАВИТЬ

Дополнительные медицинские вмешательства по направлению врача-специалиста этапа 3:
Количество = $N_{ЛП}$ Этапа 3 (рассчитывается автоматически)

Название медицинского вмешательства

Биопсия с прижизненным ПАО

26/200

Стоимость (руб.) за 1 медицинскую услугу

+ ДОБАВИТЬ

ДОБАВИТЬ ЭТАП ДИАГНОСТИКИ С ИИ ПРИ ВЕРОЯТНОСТИ ЦЗ < ПОРОГА АЛГОРИТМА МАРШРУТИЗАЦИИ

Рисунок 79 – Пример ввода и расчета результатов оценки МЭЭ диагностики с применением ИИ (количество случаев целевого заболевания ниже значения порога алгоритма маршрутизации)

ЭТАПЫ МАРШРУТИЗАЦИИ БЕЗ ИИ

Традиционный путь: все пациенты последовательно проходят через этапы врач-специалист 1 (например, ВОП/терапевт) – врач-специалист 2 – врач-специалист 3 и т.д.)

Этап 1

Название этапа (врач-специалист)

Врач ВОП / Терапевт

Режим ввода данных:

☒ По точности диагностики врача-специалиста ☐ По количеству пациентов

Чувствительность (%) Специфичность (%)

Медицинские вмешательства этапа:
1-я медицинская услуга – осмотр врачом-специалистом (для всех пациентов). Сх 2-я – оказывается медицинская услуга по направлению врача-специалиста текущего этапа.

+ ДОБАВИТЬ

Дополнительные медицинские вмешательства по направлению врача-специалиста этапа 1:
Количество = Н_ЛП Этапа 1 (рассчитывается автоматически)

+ ДОБАВИТЬ

Этап 2

Название этапа (врач-специалист)

Врач-дерматовенеролог

Режим ввода данных:

☒ По точности диагностики врача-специалиста ☐ По количеству пациентов

Чувствительность (%) Специфичность (%)

Медицинские вмешательства этапа:
1-я медицинская услуга – осмотр врачом-специалистом (для всех пациентов). Сх 2-я – оказывается медицинская услуга по направлению врача-специалиста текущего этапа.

Название медицинского вмешательства

Осмотр врача-дерматовенеролога

15/200

Осмотр проводится всем пациентам этапа – количество учитывается автоматически

Стоимость (руб.) за 1 медицинскую услугу

Название медицинского вмешательства

Дерматоскопия

13/200

Стоимость (руб.) за 1 медицинскую услугу

+ ДОБАВИТЬ

Дополнительные медицинские вмешательства по направлению врача-специалиста этапа 2:
Количество = Н_ЛП Этапа 2 (рассчитывается автоматически)

+ ДОБАВИТЬ

Этап 3

Название этапа (врач-специалист)

Врач-онколог

Режим ввода данных:

☒ По точности диагностики врача-специалиста ☐ По количеству пациентов

Чувствительность (%) Специфичность (%)

Медицинские вмешательства этапа:
1-я медицинская услуга – осмотр врачом-специалистом (для всех пациентов). Сх 2-я – оказывается медицинская услуга по направлению врача-специалиста текущего этапа.

Название медицинского вмешательства

Осмотр онколога

15/200

Осмотр проводится всем пациентам этапа – количество учитывается автоматически

Стоимость (руб.) за 1 медицинскую услугу

+ ДОБАВИТЬ

Дополнительные медицинские вмешательства по направлению врача-специалиста этапа 3:
Количество = Н_ЛП Этапа 3 (рассчитывается автоматически)

Название медицинского вмешательства

Биопсия с прижизненным ПАО

26/200

Стоимость (руб.) за 1 медицинскую услугу

+ ДОБАВИТЬ

ДОБАВИТЬ ЭТАП ДИАГНОСТИКИ БЕЗ ИИ

Косвенные финансовые затраты (ФЗ)
Стоимость лечения одного пропущенного случая ЦЗ
СТнд – стоимость лечения пропущенного случая ЦЗ (руб.)

РАССЧИТАТЬ СТНД ПО СТАДИЯМ ЗАБОЛЕВАНИЯ

РАССЧИТАТЬ МЭЭ ДИАГНОСТИКИ

Рисунок 80 – Пример ввода и расчета результатов оценки МЭЭ диагностики без применения ИИ

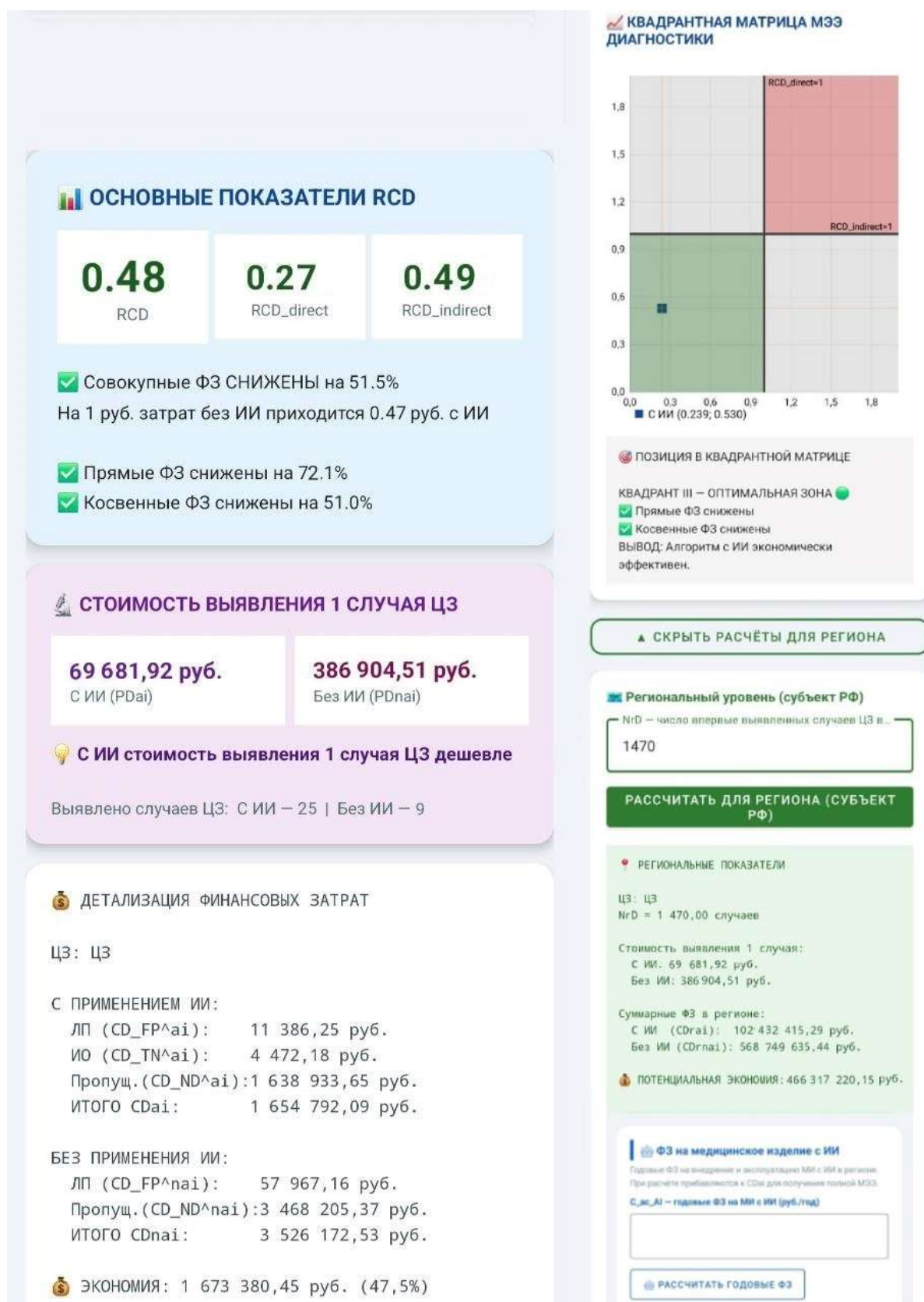


Рисунок 81 – Пример расчетов и визуализации результатов оценки МЭЭ диагностики в квадрантной матрице

Значение результатов и ограничения

Основной вклад работы состоит не столько в подтверждении экономических преимуществ конкретной системы ИИ, сколько в формализации методологического инструментария, воспроизводимого и сопоставимого способа расчета затрат и оценки МЭЭ на региональном уровне. Предложенный способ ориентирован на затраты системы здравоохранения как основной измеримый результат, что соответствует мировой практике оценки медицинских технологий.. Введенная формула годовых эквивалентных затрат частично отвечает этому требованию.

Ограничения исследования включают зависимость расчета числа выявленных случаев от опубликованных показателей чувствительности врачей, которые варьируются в зависимости от уровня организации и квалификации специалистов; условный, демонстрационный характер данных о затратах по стадиям; привязку к тарифам ОМС г. Москвы, ограничивающую прямую экстраполяцию на другие регионы; а также апробацию на одной нозологической группе. Преимущества способа – явный учет многоэтапности маршрутизации и пропущенных случаев на каждом этапе, ориентация на удельный показатель затрат на выявление одного случая, обеспечивающая прямую сопоставимость стратегий, и наличие блока региональной экстраполяции, позволяющего оценить масштаб экономии в территориальной системе здравоохранения. Дальнейшая валидация предполагает проспективные многоцентровые исследования с использованием реальных клинических и финансово-экономических данных.

Резюме

Ключевым результатом раздела является формализация воспроизводимого и практически применимого способа расчета затрат и оценки МЭЭ диагностики с применением моделей ИИ на региональном уровне, основанного на системе расчетных формул с учетом каскадных диагностических потерь. При

экстраполяции на г. Москву потенциальная экономия необоснованных затрат составила 436 296 411,6 руб. в год, а средняя длительность установления диагноза сократилась на 10,5 суток. Способ адаптируется к различным нозологиям, диагностическим системам ИИ, медицинским организациям и регионам и может служить основой стандартизированной методологии анализа МЭЭ ИИ-технологий в здравоохранении, а также использоваться при разработке программ цифровой трансформации первичной медицинской помощи.

Представленные в подразделе результаты образуют логически завершенную последовательность оценки диагностических программ ИИ от отдельного диагностического акта до региональной системы здравоохранения. В разделе 7.1 показано, что стандартная оценка точности систематически занижает клиническую ценность программ ИИ, расценивая обоснованную онкологическую настороженность как ошибку, и разработана трехуровневая методология (формальная точность, клиническая эффективность, клиническая точность), при использовании которой клиническая точность программы Derma Onko Melanoma Check New составила 92,1 % при клинической чувствительности 98,5 % и отсутствии пропущенных ЗНО. В разделе 7.2 предложен способ оценки эффективности маршрутизации с коэффициентами прироста выявляемости и сокращения пропусков, показавший увеличение выявляемости ЗНО в 1,9 раза и снижение пропусков в 10,5 раза. В разделах 7.3 и 7.4 эта клиническая эффективность переведена в экономическую плоскость: на уровне учреждения применение ИИ снижает общие затраты на 51 %, а на региональном уровне обеспечивает потенциальную экономию порядка 436 млн руб. в год при сокращении сроков диагностики на 10,5 суток. В совокупности разработанные методологические инструменты – показатели клинической точности и эффективности, коэффициенты относительной эффективности и медико-экономической эффективности, квадрантные матрицы визуализации и программные средства их расчета – формируют целостную систему объективной оценки диагностических технологий ИИ в дерматоонкологии, пригодную для

сравнительных исследований, регуляторной экспертизы и обоснования управленческих решений о внедрении ИИ в клиническую практику.

3.13. Анализ эволюции программ семейства «Derma Check» для автоматизированной диагностики новообразований кожи и дерматозов и эволюция диагностического подхода

Разработанная система поддержки принятия врачебных решений (СППВР) «Derma Check» является итогом последовательной эволюции семейства программ компьютерного зрения для дифференциальной диагностики новообразований кожи, которая заняла период с 2024 по 2026 гг. За это время подход к автоматизированной диагностике прошел путь от отдельных монолитных бинарных классификаторов, работавших по логике большинства сопоставимых мобильных приложений, к мультимодельной системе, воспроизводящей логику клинического мышления врача-дерматовенеролога и онколога. Понимание этой эволюции необходимо для корректной интерпретации метрик текущей версии, поскольку на разных этапах менялась не только нейросетевая архитектура, но и сама методология оценки эффективности с переходом от формальной (технической) точности к клинической точности диагностики. В настоящем разделе последовательно рассмотрены все версии программ, их функциональные возможности и показатели диагностической эффективности.

Первые версии: монолитные бинарные классификаторы «Melanoma Check» и «Derma Onko Check»

Первыми в семействе стали две независимые программы для карманных персональных компьютеров (например, смартфонов) и планшетных компьютеров, разработанные и зарегистрированные в 2024 г.: «Melanoma Check» [97] и «Derma Onko Check» [96]. Обе представляли собой автономные offline-приложения для операционной системы Android, предназначенные для врачей общей практики

(ВОП), терапевтов, дерматовенерологов и онкологов при оказании первичной медико-санитарной помощи. Каждая программа была построена на единственной нейросетевой модели, решавшей задачу бинарной классификации, и допускала как непосредственную фотосъемку новообразования камерой смартфона, так и загрузку фотоизображений (в том числе дерматоскопических) с любого цифрового носителя.

Программа «Melanoma Check» выполняла бинарную классификацию меланоцитарных опухолей по классам «меланома» / «не меланома». Ее эффективность была оценена у 133 пациентов в возрасте от 19 до 90 лет с различными меланоцитарными опухолями кожи; заключения программы сопоставлялись с предварительными и заключительными клиническими диагнозами врачей-дерматовенерологов с опытом работы в дерматоонкологии более 20 лет и результатами морфологической верификации. Программа продемонстрировала диагностическую точность 89,5%, чувствительность 92,5% и специфичность 88,2%, что оказалось сопоставимо с эффективностью опытных дерматовенерологов со стажем 20–40 лет (точность 95,5%, чувствительность 95,0%, специфичность 100%) [104].

Программа «Derma Onko Check» решала более широкую задачу дифференциации всех новообразований кожи по классам «доброкачественное» / «злокачественное». Ее эффективность оценена у 135 пациентов в возрасте от 22 до 78 лет с образованиями, визуально подозрительными на злокачественные; заключения сопоставлялись с дерматоскопическими и морфологическими данными. Диагностическая точность составила 96%, чувствительность 98%, специфичность 96%, доля ложноположительных результатов 4,3%, доля ложноотрицательных 2,4% [109]. Дополнительно точность программы была подтверждена в сериях исследований по отдельным группам новообразований из разных тканей – меланоцитарным, кератиноцитарным/эпидермальным [78, 103], сосудистым и фиброзным [108].

Принципиальными достоинствами первых версий были простота и однозначность вывода (бинарный класс: злокачественное / доброкачественное или

меланома / не меланома с указанием вероятности корректности заключения в процентах), автономность работы без подключения к сети и низкие требования к аппаратному обеспечению. Вместе с тем именно на этом этапе проявились ограничения монолитной бинарной архитектуры: одна модель не охватывала всего спектра онкологической патологии кожи, в которой помимо доброкачественных и злокачественных образований существуют промежуточные – предраковые и пограничные состояния, а также не позволяет формировать дифференцированные рекомендации по маршрутизации пациента.

Вторые версии (2.0): смена нейросетевой архитектуры

В 2025 г. после клинической апробации, анализа и учета недостатков были разработаны и зарегистрированы обновленные версии обеих программ «Derma Onko Check» (версия 2.0) [92] и «Melanoma Check» (версия 2.0) [94]. Ключевым изменением стала смена нейросетевой архитектуры и существенное расширение обучающего датасета.

Модель «Derma Onko Check» (версия 2.0) была реализована на архитектуре сверточной нейронной сети Xception (библиотека Keras) и обучена на расширенном верифицированном наборе из 48 000 изображений, объединившем данные HAM10000, ISIC 2020 и собственную базу данных новообразований кожи [85]. В версии 2.0 глубокое обучение было оптимизировано с учетом недостатков первой версии, в частности минимизированы ложноположительные срабатывания на фотоизображениях с бородавчатой текстурой и роговыми наслоениями. Эффективность обновленной программы оценена на значительно более крупной и стратифицированной когорте – 381 пациент в возрасте 22–78 лет, разделенной на группу меланоцитарных (n=230: 191 доброкачественная, 39 злокачественных) и группу кератиноцитарных/эпидермальных (n=151: 100 доброкачественных, 51 злокачественная) опухолей. Диагностическая точность составила 92,4%, чувствительность 94,4%, специфичность 91,8%, доля ложноотрицательных результатов 5,6%, отрицательная прогностическая ценность (NPV) – 98,2%. В

сравнении с врачами-дерматовенерологами с опытом работы в дерматоонкологии более 20 лет (точность 92,3%, чувствительность 94,4%, специфичность 100%) программа показала сопоставимые точность и чувствительность; при этом в 4 из 5 ложноотрицательных случаев врачей программа правильно классифицировала новообразования как злокачественные (1 меланома *in situ* и 3 базальноклеточных рака), что подчеркнуло ее ценность как инструмента «второго мнения» даже для высококвалифицированных специалистов [92]. Несколько более высокие абсолютные значения первой версии (96%/98%/96%) объясняются меньшей и менее репрезентативной выборкой (135 против 381 наблюдения); валидация версии 2.0 на существенно расширенной когорте при сохранении высокой чувствительности и близкой к 100% отрицательной прогностической ценности отражает большую надежность оценки.

Модель «Melanoma Check» (версия 2.0) была модернизирована аналогичным образом, архитектура классификатора заменена на современную сверточную нейронную сеть Xception (библиотека Keras), а обучение проведено на расширенном датасете дерматоскопических и клинических фотоизображений меланоцитарных опухолей, включавшем открытые коллекции и собственную базу данных [85]. По результатам внутреннего тестирования обновленная версия превзошла первую версию по всем основным показателям, продемонстрировав в диагностике меланоцитарных опухолей диагностическую точность 93,1%, чувствительность 96,2% и специфичность 91,5% [94].

Мультимодельное объединение: «Derma Onko Melanoma Check» и запатентованные способы диагностики

Логическим развитием стало объединение двух самостоятельных моделей в единую систему. В 2025 г. была зарегистрирована мультимодельная программа «Derma Onko Melanoma Check» [93], интегрировавшая модели «Derma Onko Check» и «Melanoma Check» (версии 2.0) по принципу мягкого голосования (англ. *soft voting* – метод в машинном обучении, который используется в рамках

ансамблевых методов для улучшения точности предсказаний. Вместо простого выбора класса большинством голосов (как в жестком голосовании – англ. *hard voting*), он усредняет предсказанные вероятности каждого класса от нескольких базовых моделей. Итоговым предсказанием становится класс с наибольшей средней вероятностью) с агрегацией предсказаний на основе вероятностных оценок и эмпирически подобранным порогом (англ. *threshold* – заранее заданное числовое значение, которое служит границей для принятия решений моделью, так называемый «критерий отсечки»: если оценка уверенности модели (англ. *confidence score*) превышает этот порог, модель принимает одно решение, если нет – другое) = 0,41, обеспечивающим баланс между чувствительностью и специфичностью. Вышеописанная идея мультимодельности и взаимодействия моделей в ансамбле была вдохновлена теорией тысячи мозгов Дж. Хокинса, согласно которой интеллект возникает из коллективного «голосования» множества независимых моделей, взаимно усиливающих сигналы и компенсирующих слабости друг друга. Сравнительный анализ методов агрегации показал, что в сценарии, когда одна из объединяемых моделей обладает сниженной чувствительностью, применение мягкого голосования повышает чувствительность в среднем на 51,97%, что и определило выбор данной стратегии.

Данный способ диагностики представлял собой новое техническое решение и был защищен двумя патентами на изобретение: «Способ дифференциальной диагностики доброкачественных и злокачественных новообразований кожи с применением модели искусственного интеллекта» [75] и «Способ дифференциальной диагностики меланомы кожи и доброкачественных меланоцитарных и немеланоцитарных новообразований кожи с применением модели искусственного интеллекта» [76].

Эффективность мультимодельной программы оценена отдельно по двум группам опухолей. В диагностике меланоцитарных опухолей (n=230) достигнуты диагностическая точность 92,2%, чувствительность 100,0%, специфичность 90,5%, отрицательная прогностическая ценность 100,0%, F1-мера (англ. *F1-score*) 83,9% и коэффициент корреляции Мэтьюса 0,816; ложноотрицательные результаты

отсутствовали полностью. В диагностике кератиноцитарных/эпидермальных опухолей (n=151) получены диагностическая точность 89,4%, чувствительность 96,1%, специфичность 84,0%, отрицательная прогностическая ценность 97,8% и F1-мера 87,4%. Объединение моделей обеспечило нулевые или минимальные ложноотрицательные результаты при выявлении злокачественных новообразований, что и являлось главной целью мультимодельного подхода; относительно высокий уровень ложноположительных результатов был осознанным следствием приоритета чувствительности над специфичностью.

«Derma Onko Melanoma Check Plus»: нормализация фотоизображений и алгоритмы маршрутизации

В 2026 г. мультимодельная система была дополнена новыми модулями и функциями, и зарегистрирована как «Derma Onko Melanoma Check Plus» [93]. По сравнению с предыдущей версией в нее были встроены два новых функциональных блока. Во-первых, модуль нормализации параметров фотоизображений, обеспечивающий приведение входных фотоизображений к единым характеристикам яркости, контраста и цветопередачи перед подачей в нейросетевые модели, что снижает влияние условий фотосъемки (освещение, тип камеры) на устойчивость классификации. Во-вторых, алгоритмы маршрутизации пациентов, которые на основании интегральной оценки риска формировали дифференцированные рекомендации по дальнейшей тактике. Диагностическое ядро (мягкое голосование двух моделей) при этом сохранялось; нововведения были направлены на повышение воспроизводимости результатов в реальных условиях первичного врачебного приема и на приближение вывода программы к практической задаче маршрутизации.

«Derma Onko Melanoma Check New»: переход к клинической логике диагностики

Принципиально новым этапом стала программа «Derma Onko Melanoma Check New», в которой была реализована не просто интеграция готовых моделей, а совершенно иная клиническая логика диагностики. Если предшествующие версии воспроизводили технический подход большинства аналогичных программ (бинарный вердикт «доброкачественное / злокачественное»), то данная версия была построена как мультимодельная система, отражающая логику клинического мышления врача-дерматовенеролога и онколога.

Архитектура объединила модели разных типов: сверточные нейронные сети и визуальные трансформеры, что позволило совместить сильные стороны обеих архитектур. Система состояла из четырех функциональных модулей (бинарных и многоклассовых), заключения которых объединялись голосованием большинства (2 из 3), с отдельным блоком контроля качества фотоизображений, работающим в реальном времени до передачи фотоизображения в модели. Обоснованием такой архитектуры служило то, что одна бинарная модель не решает задачу клинической диагностики полностью: в дерматоонкологии существуют не только доброкачественные и злокачественные, но и промежуточные и предраковые состояния, а многоклассовая модель по точности уступает бинарной. Определение всех категорий через объединение заключений нескольких моделей и является аналогом клинического мышления врача.

Вместо бинарного вердикта программа формировала категорию риска злокачественности (высокий, средний или низкий) вместе с рекомендацией по маршрутизации (к онкологу, дерматовенерологу или под наблюдение ВОП/терапевта). В один из модулей был встроен настраиваемый порог чувствительности, регулируемый под потребности конкретной медицинской организации. Клиническая логика проявлялась, в частности, в том, что новообразования, схожие с базалиомами, направлялись к дерматовенерологу (который вправе провести цитологическую диагностику и установить

заключительный диагноз), а не сразу к онкологу, что снижало нагрузку на онкологическую службу; наиболее прогностически значимые новообразования – меланомы, при этом всегда направлялись к онкологу напрямую.

Модель была обучена на 135 000 фотоизображений из открытых источников и собственного набора данных; клиническое тестирование проведено на наборе ФГБУ «ГВКГ им. Н.Н. Бурденко» Минобороны России (n=342). Формальная (техническая) диагностическая точность составила 89,7%, чувствительность 100%, специфичность 84,8%, F1-мера 0,8634. Данный проект вышел в финал конкурса инновационных проектов в здравоохранении «Оргздрав 2026»

От формальной к клинической точности: обоснование динамики специфичности

Переход от ранних монолитных классификаторов к мультимодельным версиям, воспроизводящим клиническую логику, сопровождается закономерным и клинически объяснимым снижением формальной специфичности. Этот эффект принципиально важно интерпретировать корректно.

Ранние версии («Melanoma Check», «Derma Onko Check» первых поколений) работали так же, как и большинство сопоставимых программ, исключительно с технической точки зрения: любое доброкачественное образование, отнесенное к положительному классу, автоматически считалось ложноположительным результатом, снижающим специфичность. Такой подход не учитывает реальную клиническую практику, в которой ряд заведомо доброкачественных, но атипичных, предраковых или нетипично пигментированных образований (диспластические невусы, актинический кератоз, клеточные голубые невусы, себорейный кератоз с бородавчатыми наслоениями) обоснованно направляется врачом на морфологическое исследование в силу онконастороженности.

Современные версии оценивают образование именно с клинической точки зрения: новообразованиям, которым по клинико-дерматоскопическим признакам требуется гистологическая или цитологическая верификация, программа

присваивает средний или высокий риск. С позиции формальной матрицы ошибок такие заключения по доброкачественным образованиям учитываются как ложноположительные и понижают специфичность, тогда как, по существу, они совпадают с обоснованным решением врача о направлении на биопсию и являются клинически оправданными. Именно поэтому оценка эффективности современных версий программы проводится не только по формальной, но и по клинической точности с учетом онконастороженности, при которой заключение о повышенном риске по доброкачественному новообразованию признается обоснованным, если врач направил это новообразование на биопсию и морфологически подтверждена атипия, дисплазия или предраковое изменение. Такой методологический подход, реализованный при апробации текущей версии, повышает с формальной (технической) специфичности 84,7% до клинической специфичности 96,2%. Это демонстрирует, что снижение формальной (технической) специфичности отражает не деградацию диагностического качества, а осознанный сдвиг логики программы в сторону клинически обоснованной онконастороженности.

«Derma Check» (веб-приложение): текущая мультимодульная реализация

Кульминацией эволюции семейства стала мультимодульная СППВР «Derma Check», реализованная в виде веб-приложения [90] и подробно описанная в последующих разделах настоящей главы. В отличие от предшественников, она объединяет не только несколько решающих модулей бинарной и многоклассовой классификации, но и подсистемы оценки качества фотоизображения, сегментации фотоизображения новообразования, анализа асимметрии новообразования методом главных компонент, объяснимого искусственного интеллекта (XAI), интеграции анамнестических данных, персонализированной маршрутизации, личного кабинета врача и динамического наблюдения образований. Все нейросетевые модели системы были обучены с использованием собственных баз данных новообразований кожи, как базы данных фотоизображений, так и базы данных оригинальных и аугментированных фотоизображений новообразований

кожи с характеристиками новообразований и пациентов [86]. Архитектура системы схематично представлена на Рисунке 82.

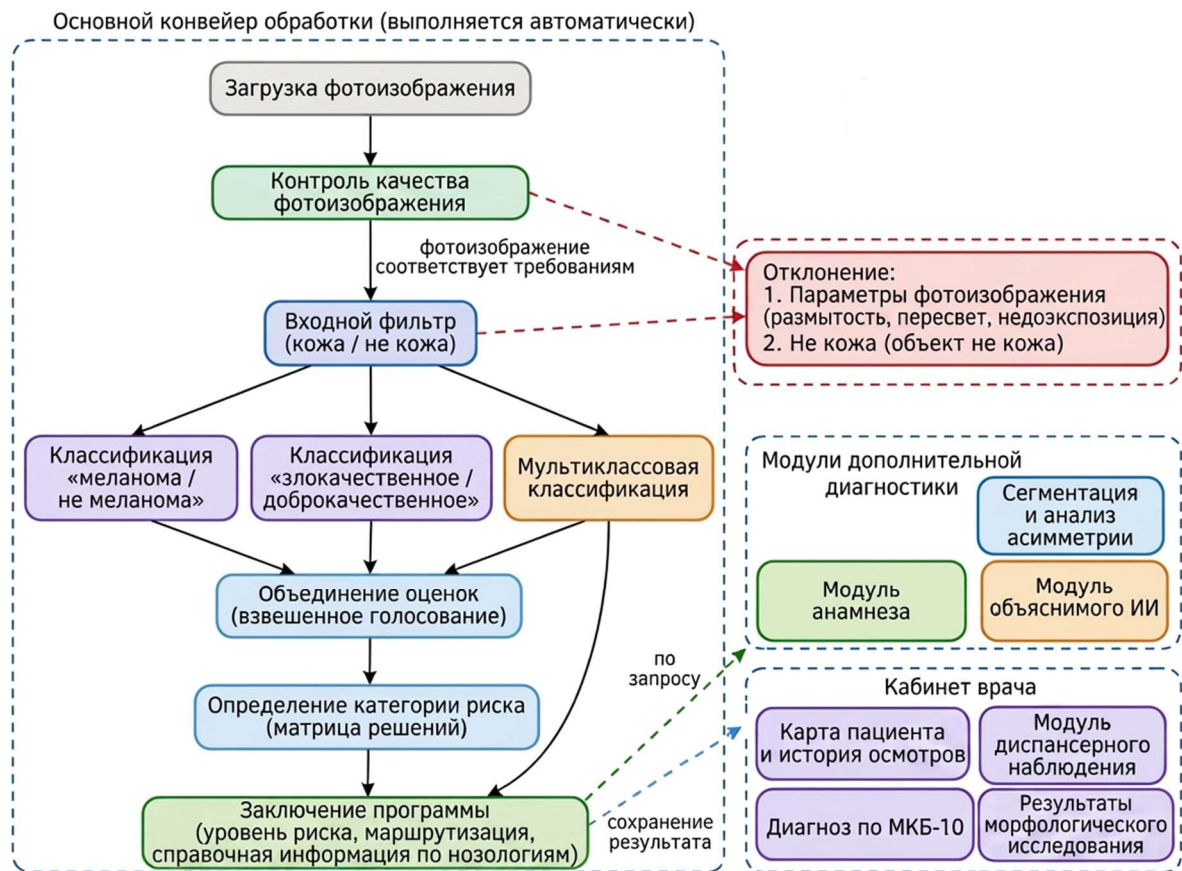


Рисунок 82 – Архитектура мультимодульной системы поддержки принятия врачебных решений «Derma Check» для дифференциальной диагностики новообразований кожи

Диагностическая эффективность текущей версии программы подтверждена при апробации на массиве из 2250 клинических изображений новообразований кожи (1515 направленных на анализ через мобильное приложение фотоизображений новообразований кожи, полученных на первичных приемах врачей в ФГКУ «1586 ВКГ» Минобороны России и филиалах, ГБУЗ «Московский Клинический Научный Центр имени А. С. Логинова ДЗМ», клинике косметологии MD Elena; 735 – в ФГБУ «ГВКГ им. Н.Н. Бурденко» Минобороны России) с сопоставлением с предварительными и заключительными клиническими диагнозами, установленными врачами. При разделяющей способности $AUC = 0,985$ формальная чувствительность составила 96,7%, формальная специфичность

84,7%; с учетом обоснованной онконастороженности: клиническая чувствительность составила 93,3%, клиническая специфичность 96,2%, клиническая точность 95,3%. Все меланомы ($n=173$) были выявлены полностью; 13 пропусков ограничились базальноклеточным раком наименее агрессивной и крайне редко метастазирующей злокачественной нозологией. Согласие заключений программы с консенсусом врачей по коэффициенту каппа Коэна составило 0,750, а точный критерий Фишера при сравнении чувствительности программы и дерматовенерологов не выявил статистически значимого различия ($p=0,559$). Детальное описание архитектуры, модулей и результатов апробации приведено далее.

Сравнительный анализ эволюции программ

Обобщенные данные о нейросетевой архитектуре, типе классификации и показателях диагностической эффективности всех версий программ представлены в Таблице 55.

Таблица 55 – Эволюция программ компьютерного зрения для дифференциальной диагностики новообразований кожи

Программа	Год	Архитектура	Число моделей	Тип классификации	Выборка (n)	Точность, %	Чувствительность, %	Специфичность, %
Melanoma Check v1.0	2024	MobileNet	1	Бинарная (меланома / не меланома)	133	89,5	92,5	88,2
Derma Onko Check v1.0	2024	Xception	1	Бинарная (доброкач. / злокач.)	135	96,0	98,0	96,0
Derma Onko Check v2.0	2025	Xception	1	Бинарная (доброкач. / злокач.)	381	92,4	94,4	91,8
Melanoma Check v2.0	2025	Xception	1	Бинарная (меланома / не меланома)	230	93,1	96,2	91,5
Derma Onko Melanoma Check	2025	Ансамбль из 2 СНС Xception (soft voting, порог 0,41)	2	Ансамбль: Бинарная (доброкач./ злокач.) и бинарная (меланома/ не меланома)	230 (мел.) 151 (керат./эпид.)	92,2 (мел.) 89,4 (керат./эпид.)	100,0 (мел.) 96,1 (керат./эпид.)	90,5 (мел.) 84,0 (керат./эпид.)
Derma Onko Melanoma Check New	2026	Ансамбль: СНС и визуальные трансформеры (голосование 2 из 3)	4 модуля	Ансамбль: бинарная (доброкач./ злокач.) и бинарная (меланома/ не меланома)	342	89,7	100,0	84,8
СППВР «Derma Check» (веб)	2026	Мульти-модульная: СНС и трансформеры + ХАІ, РСА, сегментация	Мульти-модульная	Ансамбль: бинарные и многоклассовые + РСА + ХАІ	2250	форм. 96,7 клин. 95,3	форм. 96,7 клин. 93,3	форм. 84,7 клин. 96,2

Примечание: СНС – сверточная нейронная сеть; NPV – отрицательная прогностическая ценность; мел. – меланоцитарные опухоли; керат./эпид. – кератиноцитарные/эпидермальные опухоли; форм. – формальный (технический) показатель; клин. – клинический показатель; ХАІ – объяснимый ИИ; РСА – метод главных компонент

3.14. Анализ характеристик программы «Derma Check» для автоматизированной диагностики новообразований кожи и дерматозов

3.14.1. Описание программы «Derma Check»

Система поддержки принятия врачебных решений (СППВР) «Derma Check» мультимодульное веб-приложение для автоматизированной визуальной идентификации и предварительной дифференциальной диагностики новообразований и другой патологии кожи в условиях первичного звена здравоохранения.

Клиническая необходимость разработки системы обусловлена следующими обстоятельствами. Значительная часть первичных осмотров пациентов с новообразованиями кожи проводится врачами общей практики и хирургами поликлинического звена, не обладающими специализированным дерматоскопическим оборудованием и навыками его применения, что снижает чувствительность выявления злокачественных новообразований кожи на первом этапе маршрутизации. Программа «Derma Check» создана для преодоления этих ограничений: она не ставит окончательный диагноз и не подменяет гистологическую верификацию, но обеспечивает врача объективной количественной оценкой вероятности злокачественности, визуализацией зон риска в образовании и структурированными рекомендациями по тактике дальнейшего обследования.

Программа реализована в клиент-серверной архитектуре. Серверная часть построена на фреймворке FastAPI (Python 3.10+) с использованием ONNX Runtime и TensorFlow Lite для инференса нейросетевых моделей на центральном процессоре без графического ускорителя. Клиентская часть – адаптивное одностраничное веб-приложение (HTML5/CSS3/JavaScript) без внешних фреймворков, работоспособное на мобильных устройствах с низкой пропускной способностью канала (Рисунок 83).

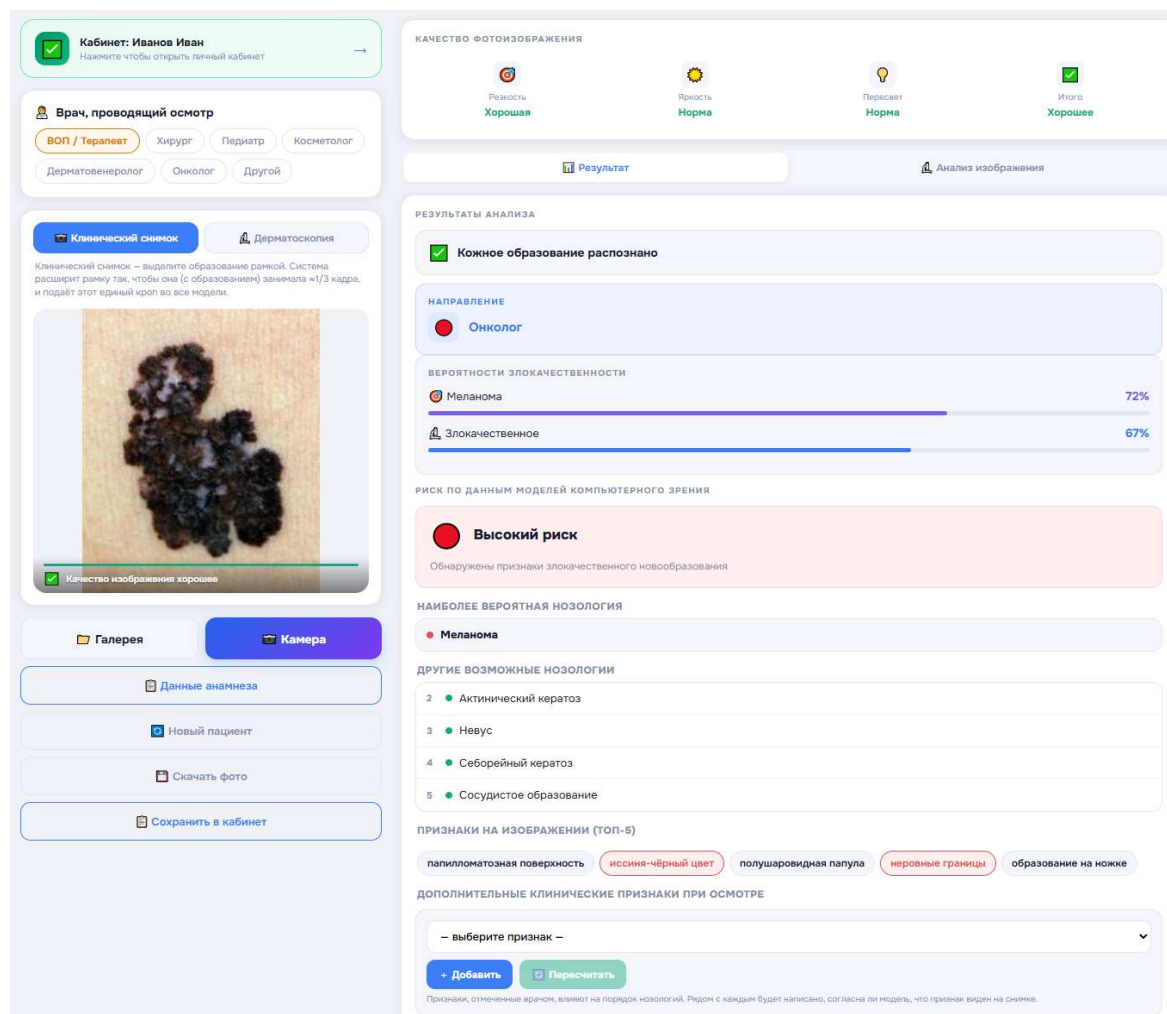


Рисунок 83 – Общий вид главной страницы приложения «Derma Check Web» с загруженным фотоизображением новообразования кожи и результатами анализа в виде заключения

Технические характеристики системы приведены в Таблице 56.

Таблица 56 – Технические характеристики системы «Derma Check»

Характеристика	Значение
Серверный фреймворк	FastAPI (Python 3.10+), uvicorn, nginx
Инференс моделей	ONNX Runtime (CPU), TensorFlow Lite
Хранилище данных	SQLite и файловая система
Клиентская часть	HTML5 / CSS3 / JavaScript (без фреймворков)
Адаптивность	Desktop и Mobile (одноколоночная сетка)
Минимальная RAM	6 ГБ
ОС	Ubuntu Linux (VPS)

3.14.2. Конвейер обработки фотоизображения

Конвейер обработки фотоизображения включает последовательные этапы: оценка качества фотоизображения, коррекция яркости, входной фильтр, параллельный инференс нескольких нейросетевых моделей классификации, объединение результатов консенсусом, а при наличии данных анамнеза, дополнительный этап интеграции клинических данных. Параллельно основному конвейеру запускаются подсистемы сегментации, PCA-анализа асимметрии, объяснимого ИИ (XAI) и многоклассовой классификации.

Перед запуском классификации система проводит автоматическую оценку качества поступившего фотоизображения по четырем параметрам: резкость, яркость, пересвет (доля пикселей с яркостью выше 95-го перцентиля) и общая экспозиция. Метрика очаговой резкости рассчитывается по патчам внутри автоматически определенной маски очага в цветовом пространстве CIELAB, что устраняет ложные срабатывания из-за текстуры кожного покрова за пределами образования. При обнаружении критически низкого качества система блокирует анализ и информирует врача о необходимости переснять фотографию с указанием конкретной причины (Рисунок 84).

Для клинических фотоизображений дополнительно применяется коррекция яркости гамма-функцией с целевым значением 0,5.

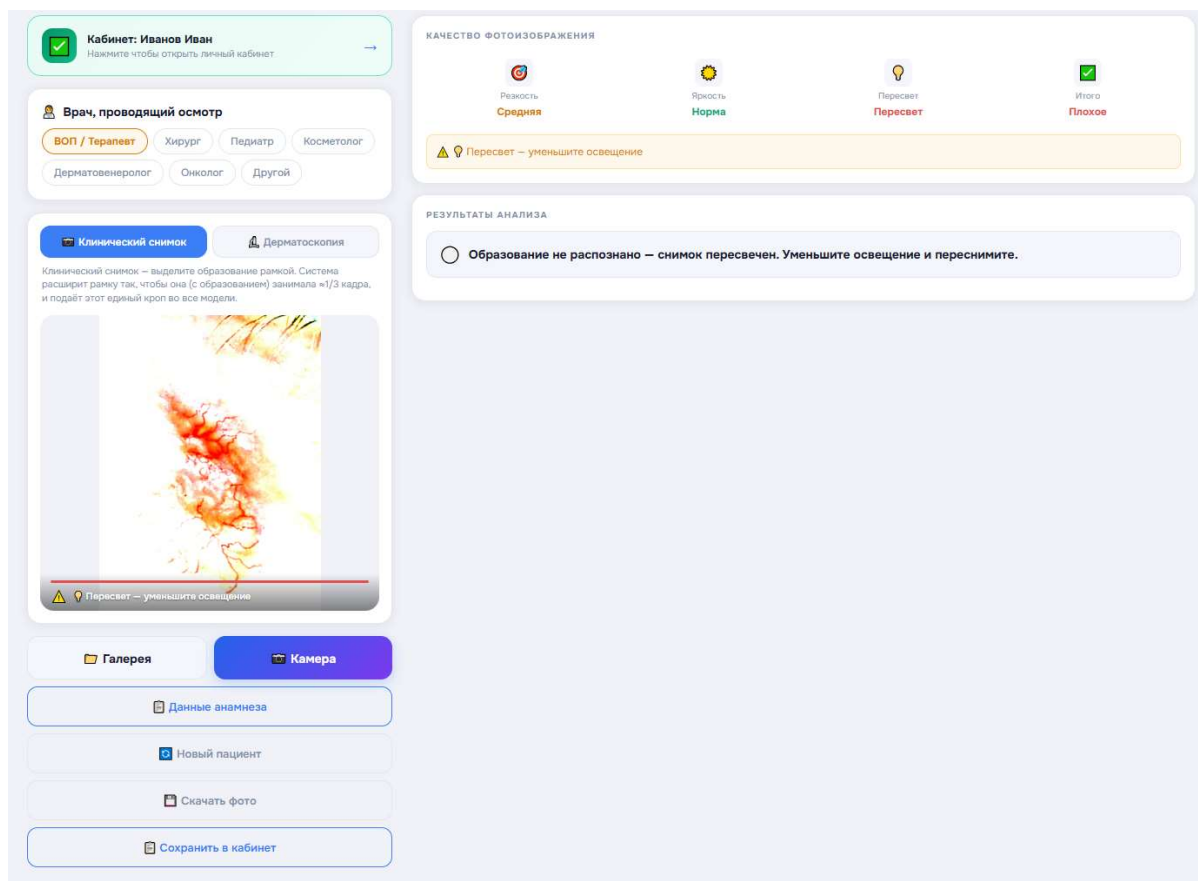


Рисунок 84 – Блок оценки качества фотоизображения: резкость, яркость, пересвет, общая оценка

3.14.3. Модули бинарной классификации и алгоритм консенсуса

Диагностическое ядро системы составляют несколько нейросетевых моделей, организованных в параллельные решающие модули. Каждая модель обучена на собственном наборе данных и решает задачу бинарной классификации. Все модели проходят через Test-Time Augmentation (ТТА) с последующим усреднением вероятностей. Помимо ТТА по поворотам, система выполняет множественный инференс с кропами (по умолчанию 9 микродвигов), что обеспечивает устойчивость результата к незначительным вариациям обрезки фотоизображения.

Один из модулей реализует мягкое голосование по средней вероятности двух компонентов, «Меланома» и «Злокачественные новообразования кожи». Второй модуль реализует алгоритм маршрутизации на основе шестнадцатипольной квадрантной матрицы, где ось абсцисс представляет вероятность меланомы, ось

ординат, вероятность злокачественного новообразования кожи. Итоговое решение формируется голосованием большинства: при согласии двух или трех модулей принимается их заключение; при полном расхождении устанавливается средний риск с уведомлением о расхождении. Результаты всех модулей отображаются в едином блоке результатов анализа (Рисунок 85). Маршрутизация пациента зависит от специальности врача: для онколога и дерматовенеролога формируется рекомендация о биопсии, для остальных специальностей, направление к специалисту соответствующего уровня.

РЕЗУЛЬТАТЫ АНАЛИЗА

✓ Кожное образование распознано

НАПРАВЛЕНИЕ

● Онколог

ВЕРОЯТНОСТИ ЗЛОКАЧЕСТВЕННОСТИ

● Меланома	5%
● Злокачественное	31%

РИСК ПО ДАННЫМ МОДЕЛЕЙ КОМПЬЮТЕРНОГО ЗРЕНИЯ:

● **Высокий риск**
Обнаружены признаки злокачественного новообразования

НАИБОЛЕЕ ВЕРОЯТНАЯ НОЗОЛОГИЯ

● Меланома

ДРУГИЕ ВОЗМОЖНЫЕ НОЗОЛОГИИ

- Актинический кератоз
- Сосудистое образование
- Невус
- Красный плоский лишай

ПРИЗНАКИ НА ИЗОБРАЖЕНИИ (ТОП-5)

бородавчатая поверхность корни эрозия неровные границы инфильтрированная бляшка

ДОПОЛНИТЕЛЬНЫЕ КЛИНИЧЕСКИЕ ПРИЗНАКИ ПРИ ОСМОТРЕ

— выберите признак —

+ Добавить Пересчитать

Признаки, отмеченные врачом, влияют на порядок нозологий. Рядом с каждым будет написано, согласна ли модель, что признак виден на снимке.

Рисунок 85 – Блок результатов анализа: заключение по уровню риска, вероятности корректности заключения моделей, заключение модуля маршрутизации

3.14.4. Модуль анализа анамнестических данных

Модуль анамнеза интегрирует клинические данные пациента с результатами модулей компьютерного зрения. Врач вносит следующие параметры: возраст, пол, фототип кожи по шкале Фицпатрика (I–VI), наличие и характер субъективных симптомов (зуд, кровоточивость, изъязвление), размеры образования и динамика их изменения, отягощенный онкологический анамнез, иммуносупрессия, локализация (Рисунок 86).

Данные анамнеза
 * Обязательные | * Важные | Необязательные

ОБЯЗАТЕЛЬНЫЕ ПОЛЯ

ВОЗРАСТ *
 56

ФОТОТИП КОЖИ (ФИЦПАТРИК) *
 II – Светлая (часто горит)

ДИАМЕТР ОБРАЗОВАНИЯ (мм) *
 5 3

ЛОКАЛИЗАЦИЯ *
 Губа

ОСНОВНЫЕ СИМПТОМЫ *

☒ Возвышается над кожей * Да Нет

☒ Росло / увеличивалось * Да Нет

☒ Зуд * Да Нет

Надёжно (7/8)

ВАЖНЫЕ ПОЛЯ

ПОЛ *
 Мужской

ДОПОЛНИТЕЛЬНЫЕ СИМПТОМЫ *

☒ Изменение вида Да Нет Н.д.

☒ Кровоточивость Да Нет Н.д.

ЗЛОКАЧЕСТВЕННЫЕ ОПУХОЛИ КОЖИ В АНАМНЕЗЕ (ПАЦИЕНТ / РОДСТВЕННИКИ) *
 Нет

ОНКОЛОГИЯ В АНАМНЕЗЕ (ЛЮБАЯ ЛОКАЛИЗАЦИЯ) *
 Нет

КОНТАКТ С ПЕСТИЦИДАМИ / ХИМИКАТАМИ *
 Нет

НЕОБЯЗАТЕЛЬНЫЕ ПОЛЯ

БОЛЕЗНЕННОСТЬ ПРИ ПАЛЬПАЦИИ

☒ Болезненность Да Нет Н.д.

Сбросить Применить

Рисунок 86 – Форма ввода анамнестических данных пациента

Предсказание риска и необходимости биопсии выполняется обученными моделями машинного обучения, результат которых объединяется с результатами компьютерного зрения в единую консенсусную оценку.

Консенсус данных компьютерного зрения (CV) и машинного обучения на анамнестических данных (ML) реализован системой из 12 правил интеграции. Каждое правило активируется при выполнении конкретного набора условий и модифицирует итоговую вероятность злокачественности или флаг маршрутизации. Правила обрабатываются последовательно; при активации нескольких правил одновременно их вклады суммируются без дублирования.

3.14.5. Сегментация образования

Сегментация кожного образования реализована в двух вариантах. Алгоритмический вариант выполняет предобработку, построение карты значимости с учетом типа, бинаризацию методом Оцу и постобработку, выделение наибольшей компоненты связности, заполнение отверстий, гауссово сглаживание контура. Нейросетевой вариант использует детектор YOLO для локализации образования на фотоизображении, после чего модель SAM строит пиксельную маску. Результат автоматической сегментации с наложением маски на исходный снимок представлен на Рисунке 87.

Врач может корректировать маску вручную, задать собственную рамку, использовать кисть для добавления или удаления областей маски.



Рисунок 87 – Результат автоматической сегментации: маска образования с наложением на исходный снимок

3.14.6. Анализ асимметрии методом главных компонент

По маске образования вычисляются главные компоненты (РСА) осей наибольшей и наименьшей дисперсии контура. Маска отражается относительно каждой из осей, и рассчитывается процент несовпадающих пикселей, количественная мера асимметрии. Результаты визуализируются через РСА-карту пигментации: серверная часть генерирует изображение первой главной компоненты цветовой декомпозиции (PC1), на котором распределение пигмента представлено в градациях серого. Клиентская часть реализует адаптивную гистограмму с пользовательским бегунком контрастности (Рисунок 88), что позволяет врачу визуально оценить глубину и структуру пигментации.

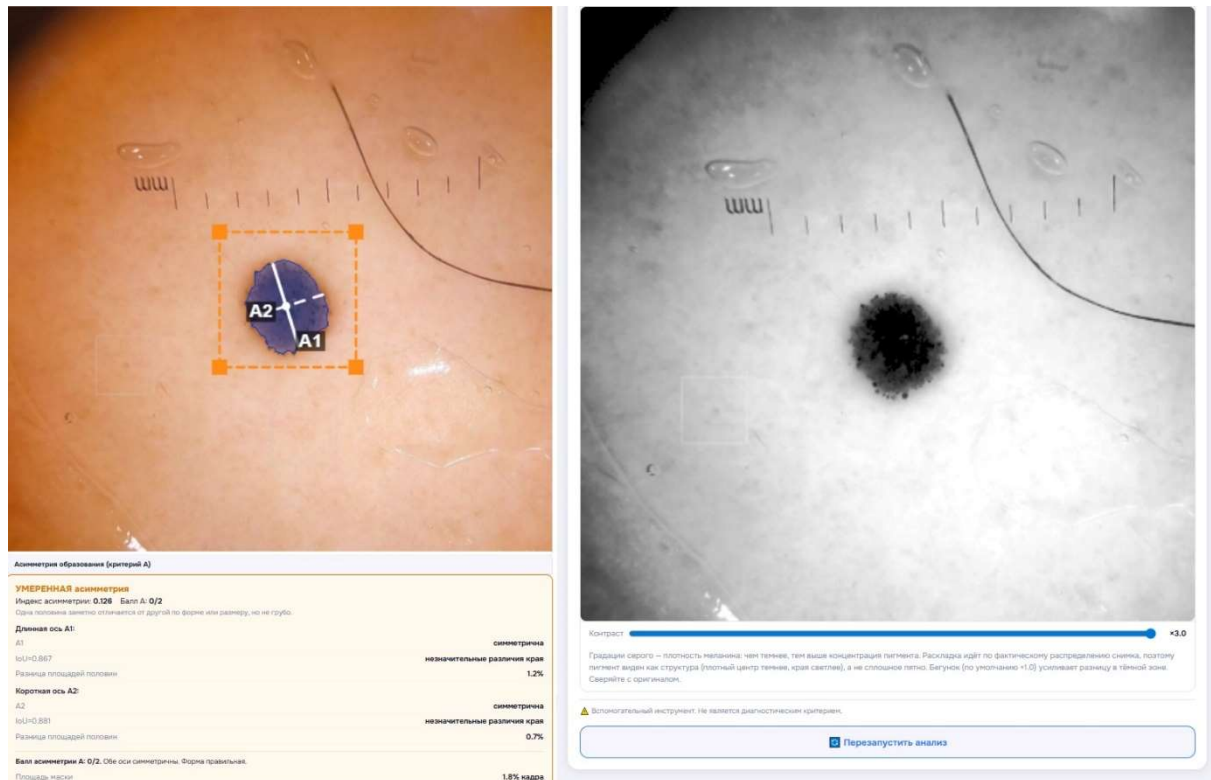


Рисунок 88 – Автоматизированный расчет асимметрии по маске образования и PCA-карта пигментации с бегунком контрастности

3.14.7. Модуль объяснимого искусственного интеллекта (XAI)

Модуль объяснимого ИИ формирует попатчевую карту риска и общий вердикт с порогом вероятности. Визуализация реализована в виде четырех прозрачных RGBA-оверлеев: тепловая карта внимания, карта структурного риска, карта попатчевого голосования оконных моделей и бинарное поле риска. Каждый оверлей накладывается на оригинальное фотоизображение с пользовательским бегунком прозрачности (Рисунок 89), что позволяет врачу интерактивно анализировать зоны наибольшего риска непосредственно в структуре образования.

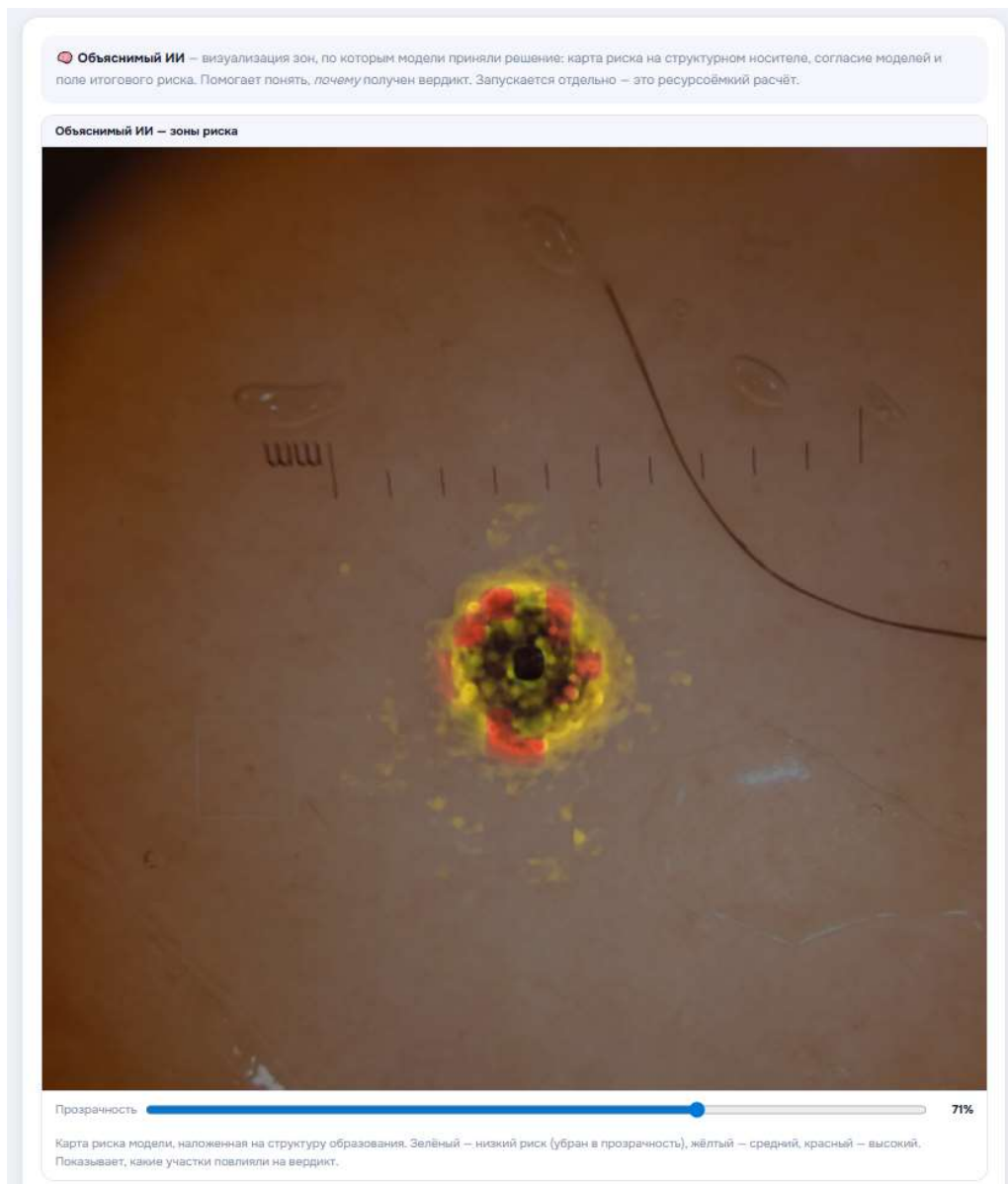


Рисунок 89 – Модуль ХАИ: тепловая карта внимания с бегунком прозрачности, наложенная на фотоизображение новообразования

Тепловые карты внимания, формируемые методом Score-CAM, визуализируют области фотоизображения, вносящие наибольший вклад в решение нейронной сети. Это принципиально отличает «Derma Check» от систем на основе единственной модели без объяснимости: врач видит не только итоговый вердикт, но и его морфологическое основание, что позволяет верифицировать или оспорить заключение ИИ на основе собственных клинических наблюдений.

3.14.8. Модуль многоклассовой классификации

Модуль многоклассовой классификации обучен на дерматологических и клинических фотоизображениях. По результатам анализа формируется топ-5 нозологий с вероятностями, предварительный дифференциальный диагноз для врача.

Принципиальной особенностью данного модуля является поддержка клинических признаков врача. Специалист может отмечать визуально определяемые клинические признаки из предопределенного банка. Для каждого отмеченного признака система рассчитывает на основе калибровочного распределения степень согласия нейросетевой модели (z-score): признак либо подтверждается моделью, либо расходится с ее предсказанием. При наличии тревожных признаков, подтвержденных моделью, злокачественная нозологическая группа перемещается на первое место топ-5 (Рисунок 90).

НАИБОЛЕЕ ВЕРОЯТНАЯ НОЗОЛОГИЯ

- Меланома

⚠ Вердикт поднят по признакам, отмеченным врачом. Исключить меланому: консультация онколога, биопсия.

ДРУГИЕ ВОЗМОЖНЫЕ НОЗОЛОГИИ

- Невус
- Себорейный кератоз
- Сосудистое образование
- Базальноклеточный рак

ПРИЗНАКИ НА ИЗОБРАЖЕНИИ (ТОП-5)

пигментированное пятно полушаровидная форма **иссиня-чёрный цвет** **асимметрия** блестящая поверхность

ДОПОЛНИТЕЛЬНЫЕ КЛИНИЧЕСКИЕ ПРИЗНАКИ ПРИ ОСМОТРЕ

— выберите признак —

+ Добавить Пересчитать

⚠ **неравномерная окраска** **согласен** X

Модель подтверждает признаки — вердикт поднят.

Рисунок 90 – Блок многоклассовой классификации: топ-5 нозологий с процентами, отмеченные клинические признаки врача и уровень согласия модели

Данный механизм превращает систему из «оракула» в инструмент совместного принятия решений, где клинический опыт врача и предсказание модели взаимно верифицируют друг друга.

3.14.9. Картотека пациентов

Личный кабинет врача ЭТО защищенная подсистема приложения «Derma Check», предоставляющая персональное рабочее пространство для ведения пациентов, сохранения результатов ИИ-анализа и долгосрочного динамического наблюдения. Аутентификация осуществляется по логину и паролю с использованием токенов. Интерфейс кабинета построен в двухколоночной сетке: левая колонка, список пациентов, правая, детали выбранного пациента (Рисунок 91).

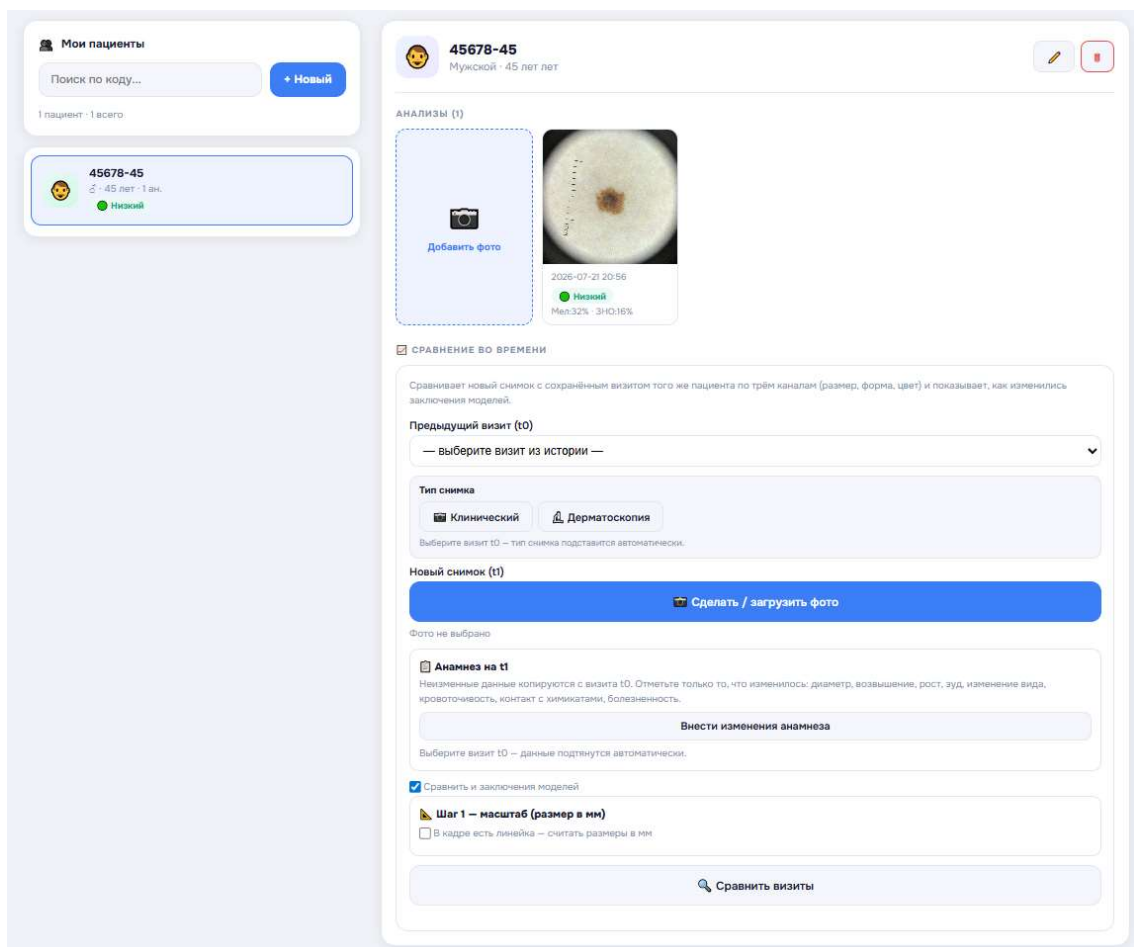


Рисунок 91 – Личный кабинет врача: список пациентов с поиском (левая панель), детали пациента с сеткой анализов и заключениями (правая панель)

Картотека основана на принципе анонимности: вместо персональных данных используется произвольный код пациента (номер медицинской карты или условный идентификатор). Для каждого пациента фиксируются год рождения, пол, фототип по Фицпатрику (шкала I–VI) и текстовые заметки. Список оснащен поиском по коду и цветовой индикацией последнего уровня риска: зеленый (низкий), желтый (средний), красный (высокий).

3.14.10. Приемы (осмотры) и сохранение результатов

Каждый прием (осмотр врача) представляет собой единицу хранения: фотоизображение, полный результат ИИ-анализа всех модулей, анамнестические данные и клинические заметки. При добавлении нового анализа из кабинета снимок проходит тот же полный конвейер, что и в основном приложении.

Для каждого приема (осмотра) врач устанавливает предварительный и окончательный диагноз с автодополнением по справочнику из 459 кодов МКБ-10, охватывающих дерматологическую и дерматоонкологическую группы (L00–L99, C43–C44, D22–D23 и смежные). Блок биопсии включает чекбокс «Биопсия назначена», поле даты, текстовое поле результата гистологического исследования, поля для описания лечения и заметок врача (Рисунок 92). Сохранение морфологического заключения в структуре приема (осмотра) обеспечивает возможность ретроспективного анализа диагностической точности системы в каждом конкретном случае.

Анализ · 2026-07-21 20:56

Удалить

Резкость

Средняя

Яркость

Норма

Пересвет

Норма

Итого

Хорошее

✓ Кожное образование распознано

НАПРАВЛЕНИЕ

● Самонаблюдение пациентом и плановый профилактический осмотр (диспансеризация) в общем порядке

ВЕРоятности злокачественности

○ Меланома 0%

□ Злокачественное 0%

Риск по данным моделей компьютерного зрения

● Низкий риск

Признаков злокачественности не выявлено.

МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ

НАИБОЛЕЕ ВЕРОЯТНАЯ НОЗОЛОГИЯ

● Невус

ДРУГИЕ ВОЗМОЖНЫЕ НОЗОЛОГИИ

2 ● Себорейный кератоз

3 ● Сосудистое образование

4 ● Актинический кератоз

5 ● Меланома

ПРИЗНАКИ НА ИЗОБРАЖЕНИИ (ТОП-5)

пигментированное пятно

коричневая окраска

полушаровидная форма

асимметрия

Многоклассовая классификация по атласу нозологий. Порядок вывода согласован с итоговым

Рисунок 92 – Форма клинических данных приема (осмотра): поля диагнозов по МКБ-10 с автодополнением, блок биопсии и гистологии

3.14.11. Подсистема динамического наблюдения

Подсистема динамического наблюдения позволяет сравнивать фотоизображения одного образования, полученные при разных приемах (осмотрах) (t_0 и t_1), по трем каналам: размер, форма, цвет. По каждому каналу пороговая система формирует заключение: «стабильно», «на контроле» или «на дообследование».

Визуализация включает оверлеи масок t_0 и t_1 с диаметрами Ферета. Карта изменения формы отображает совпавшие и сместившиеся участки границы, зоны прироста и убыли площади. Тепловая карта изменения цвета строится по значениям ΔE (евклидово расстояние между цветовыми векторами пикселей в пространстве CIELAB) для каждого пикселя внутри совмещенной маски образования (Рисунок 93). Виртуальная линейка переводит пиксельные измерения в миллиметры: калибровка выполняется независимо для каждого фотоизображения, что устраняет артефакты, связанные с различной дистанцией фотофиксации. Встроенный редактор сегментации обеспечивает коррекцию масок кистями с последующим пересчетом метрик.

КЛИНИЧЕСКИЕ ДАННЫЕ

ПРЕДВАРИТЕЛЬНЫЙ ДИАГНОЗ

Поиск по коду или названию: меланома, С43...

ОКОНЧАТЕЛЬНЫЙ ДИАГНОЗ (после биопсии)

Поиск по коду или названию: С43.5, невус...

☐ БИОПСИЯ НАЗНАЧЕНА

ДАТА БИОПСИИ

ДД.ММ.ГГГГ 

РЕЗУЛЬТАТ БИОПСИИ / ГИСТОЛОГИЯ

Описание результата...

ЛЕЧЕНИЕ / НАПРАВЛЕНИЕ

Назначенное лечение...

ЗАМЕТКИ ВРАЧА

Свободный текст...

Рисунок 93 – Результат динамического сравнения: оверлеи масок t_0/t_1 с диаметрами, карта изменения формы, тепловая карта ΔE

3.14.12. Алгоритмы персонализированной маршрутизации

Маршрутизация пациента в системе «Derma Check» формируется с учетом специальности принимающего врача и интегральной оценки риска злокачественности. Система различает три целевые группы врачей: врач первичного звена (терапевт, ВОП), дерматовенеролог и онколог. Для каждой группы определены различные пороговые уровни вероятности злокачественности, при которых система изменяет рекомендацию по тактике.

При высоком уровне риска (вероятность злокачественности $\geq 0,62$ по результатам ансамблевого консенсуса) врачу первичного звена формируется рекомендация направить пациента к дерматовенерологу или онкологу. При среднем уровне риска рекомендуется повторный осмотр через 1-3 месяца с фотодокументацией. При низком уровне риска рекомендуется динамическое наблюдение. Дерматовенерологу и онкологу при высоком уровне риска система формирует рекомендацию о биопсии с указанием оптимальной тактики. Маршрутизация дополнительно модифицируется на основании данных анамнеза по системе 12 правил, описанных в разделе 10.3.

3.15. Исследование эффективности диагностики новообразований кожи с применением программы «Derma Check»

3.15.1. Материал и методы

Оценка диагностической эффективности мультимодельной программы выполнена на массиве из 2250 клинических изображений новообразований кожи, полученных из следующих независимых источников. В 67,3% случаев (1515 изображений) загружены на первичном приеме через мобильное приложение в ФГКУ «1586 ВКГ» Минобороны России и филиалах, ГБУЗ «Московский Клинический Научный Центр имени А. С. Логинова ДЗМ», клинике косметологии MD Elena; 735 – в ФГБУ «ГВКГ им. Н.Н. Бурденко» Минобороны России). Каждое изображение сохранялось непосредственно в приложении, и при сохранении фиксировался окончательный клинический диагноз, отметка о направлении на гистологическое исследование, а для злокачественных опухолей, результат гистологической верификации. Тем самым сформирована размеченная база с прослеживаемым исходом каждого наблюдения.

В 32,7% случаев (735 изображений) собраны в Главном военном клиническом госпитале им. Н.Н. Бурденко и независимо оценены тремя врачами-дерматовенерологами со стажем работы в дерматоонкологии более 20 лет;

итоговое экспертное заключение принималось как консенсус большинства. Именно эта часть массива использовалась для сопоставления заключений программы с экспертной оценкой.

Принципиально важным является разделение диагностических ролей. Врач-дерматовенеролог на первичном приеме, опираясь на анамнез, клиническую картину и дерматоскопию (HEINE DELTA 20), устанавливает заключительный диагноз и принимает решение о проведении диагностической биопсии в объеме радикального удаления опухоли с последующим морфологическим исследованием. Программа не направляет пациента на дополнительное обследование и не ставит заключительный диагноз, она выполняет оценочную функцию и формирует заключение об уровне риска. Морфологическое исследование (гистологическое, иммуногистохимическое, цитологическое) выступает золотым стандартом. Для доброкачественных образований с типичной клинико-дерматоскопической картиной допускалась постановка окончательного диагноза без морфологического подтверждения, что соответствует общепринятой клинической практике.

Изображения получены в ходе первичных приемов 34 врачей различных специальностей: врачи общей практики и терапевты 15 человек (44,1%), дерматовенерологи 12 (35,3%), косметологи 4 (11,8%), онкологи 3 (8,8%).

Верификация диагнозов проводилась следующим образом. Набо данных включал 684 дерматоскопических (30,4%) и 1566 клинических фотоизображений (69,6%). Среди злокачественных новообразований (n=399) дерматоскопические изображения составили 60% (n=240), среди доброкачественных (n=1851) 24% (n=444).

Из 399 злокачественных опухолей все были гистологически верифицированы. Из 1851 доброкачественных новообразований врачи в 370 случаях (20,0%) приняли решение о проведении диагностической биопсии с последующим гистологическим исследованием удаленного материала. Это было обусловлено тем, что у врачей данные обследуемые поражения кожи по клинико-дерматоскопическим признакам вызвали онконастороженность, тогда как в 1481

случае (80,0%) окончательный диагноз установлен по клинической и дерматоскопической картине без морфологического исследования. Состав массива по нозологиям представлен в Таблице 57.

Таблица 57 – Состав массива по нозологиям

Доброкачественные новообразования кожи	Число случаев	Злокачественные опухоли кожи	Число случаев
Кератиноцитарные/эпидермальные опухоли (n=604)			
Кератоз себорейный	280	Базальноклеточный рак	191
Кератопапилломы	67	Плоскоклеточный рак	13
Кератоз актинический	53		
Меланоцитарные опухоли (n=1084)			
Неvus меланоцитарный приобретенный пограничный	98	Лентиго злокачественное	22
Неvus меланоцитарный приобретенный смешанный	78	Меланома поверхностно распространяющаяся	108
Неvus меланоцитарный приобретенный дермальный	438	Меланома узловая	65
Неvus меланоцитарный врожденный	58		
Неvus диспластический	124		
Неvus пятнистый	9		
Неvus голубой	84		
Сосудистые опухоли и мальформации (n=319)			
Гемангиомы сенильные и паукообразные	204		
Гемангиомы венозные	44		
Гранулемы пиогенные (гемангиомы капиллярные дольчатые)	67		
Мальформации сосудистые капиллярные	4		
Опухоли и опухолеподобные новообразования кожи из волокнистой соединительной ткани (n=155)			
Фибромы мягкие	31		
Дерматофибромы	53		
Келоиды и гипертрофические рубцы	71		
Гиперплазии придатков кожи и кисты (n=79)			
Гиперплазия сальной железы	4		
Кисты	71		
Милиумы (кисты роговые)	4		
Псевдолимфомы кожи (n=9)			
Лимфоцитомы кожи	9		
Итого: доброкачественные новообразования	1851	Итого: злокачественных опухолей	399

3.15.2. Формирование заключения об уровне риска

Важно подчеркнуть, что программа выдает заключение не просто «доброкачественное»/«злокачественное», а определяет уровни риска: низкий, средний или высокий. Уровень риска не назначается по единственному пороговому значению вероятности («если вероятность выше X% злокачественное»), а формируется многоступенчатой системой принятия решения. На первом этапе входной модуль отсекает нерелевантные изображения. Далее две подсистемы вычисляют вероятности: система А оценивает вероятность меланомы, система В вероятность злокачественности в целом. На полученных вероятностях работают три независимых модуля-классификатора риска. Их заключения объединяются голосованием большинства (2 из 3). На заключительном этапе применяется итоговое заключение.

Для статистического анализа уровни риска приводились к дихотомической шкале: низкий риск сопоставлялся с доброкачественными случаями, а средний, высокий со злокачественными. Распределение заключений по уровням риска составило: низкий 1581 наблюдение (70,3%), средний 254 (11,3%), высокий 415 (18,4%). Таким образом, доля заключений с повышенным риском (средний и высокий, рассматриваемые как единый порог направления) составила 29,7% (Рисунок 94), что отражает преимущественно доброкачественный состав массива.

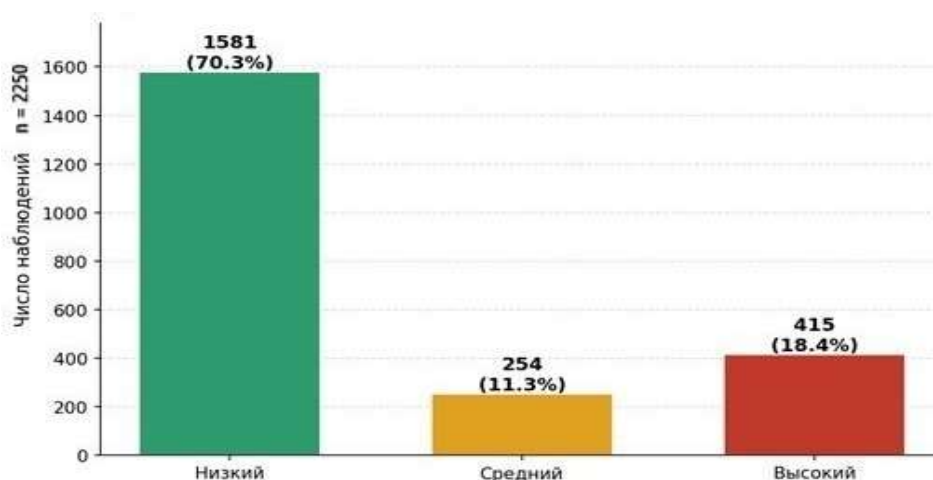


Рисунок 94 – Распределение заключений программы по уровням риска

3.15.3. Распределение вероятности и статистическая модель

Помимо дискретного уровня риска, система В формирует непрерывную вероятность злокачественности для каждого наблюдения. Гистограмма распределения этой вероятности (Рисунок 95) демонстрирует бимодальную структуру: доброкачественные новообразования формируют пик в области низких значений, злокачественные опухоли в области высоких, а зона перекрытия 0,3–0,6 соответствует спорным случаям. Диаграмма рассеяния (Рисунок 96) подтверждает разделение классов относительно пороговой линии 0,5.

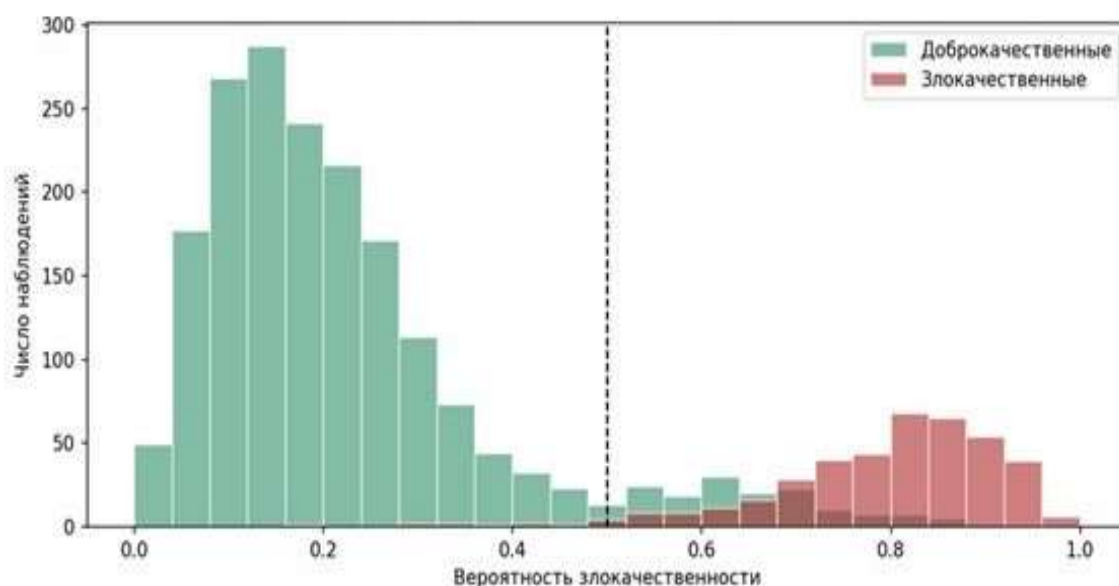


Рисунок 95 – Гистограмма распределения вероятности злокачественности

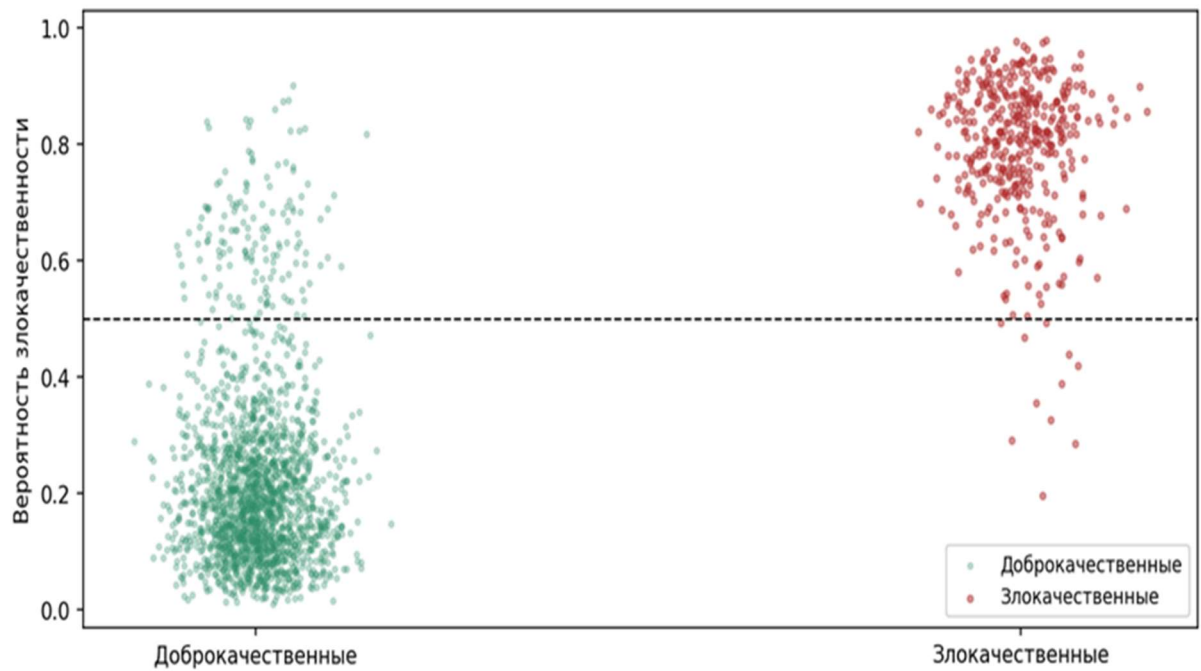


Рисунок 96 – Диаграмма рассеяния вероятности заключения по классам

Лепестковая диаграмма (Рисунок 97) отражает среднюю вероятность злокачественности по нозологиям, наибольшие значения соответствуют злокачественным формам и атипичным/предраковым образованиям (диспластические невусы, актинический кератоз), наименьшие – типичным доброкачественным (кисты, милиумы, гемангиомы).



Рисунок 97 – Средняя вероятность злокачественности по нозологиям

Разделяющая способность модели оценена площадью под ROC-кривой (Рисунок 98), AUC составила 0.985, что соответствует отличному качеству классификации; рабочая точка системы (Se 96.7%, Sp 84.7%) лежит в верхней левой области кривой.

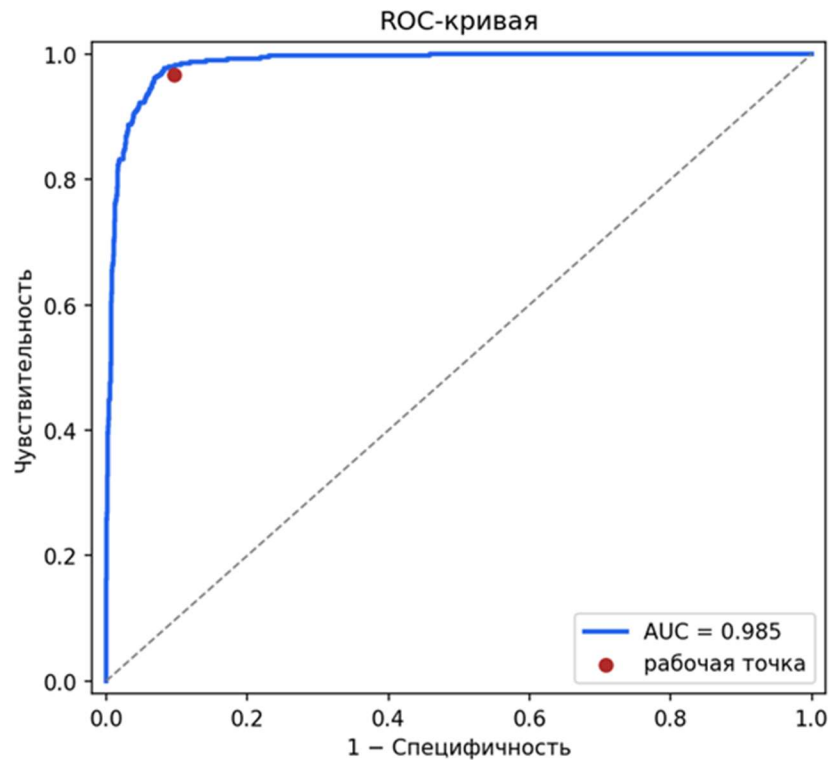


Рисунок 98 – ROC-кривая (AUC = 0,985)

Поскольку каждый показатель точности есть доля успехов в серии независимых испытаний, число верно классифицированных наблюдений подчиняется дискретному биномиальному распределению $X \sim \text{Bin}(n, p)$ для чувствительности $n=399$, $p=0,967$; для специфичности $n=1851$, $p=0,904$ (Рисунок 99). Пик распределения совпадает с точечной оценкой (наиболее вероятным числом верных ответов), ширина отражает выборочную неопределенность и сужается с ростом знаменателя, а закрашенная 95%-я масса служит основой доверительных интервалов (методы Уилсона и Клоппера–Пирсона). Так как вся 95%-я масса лежит выше пороговых значений, соответствующих клинически приемлемым уровням, чувствительность и специфичность статистически значимо превышают эти пороги, а не достигают их случайно.

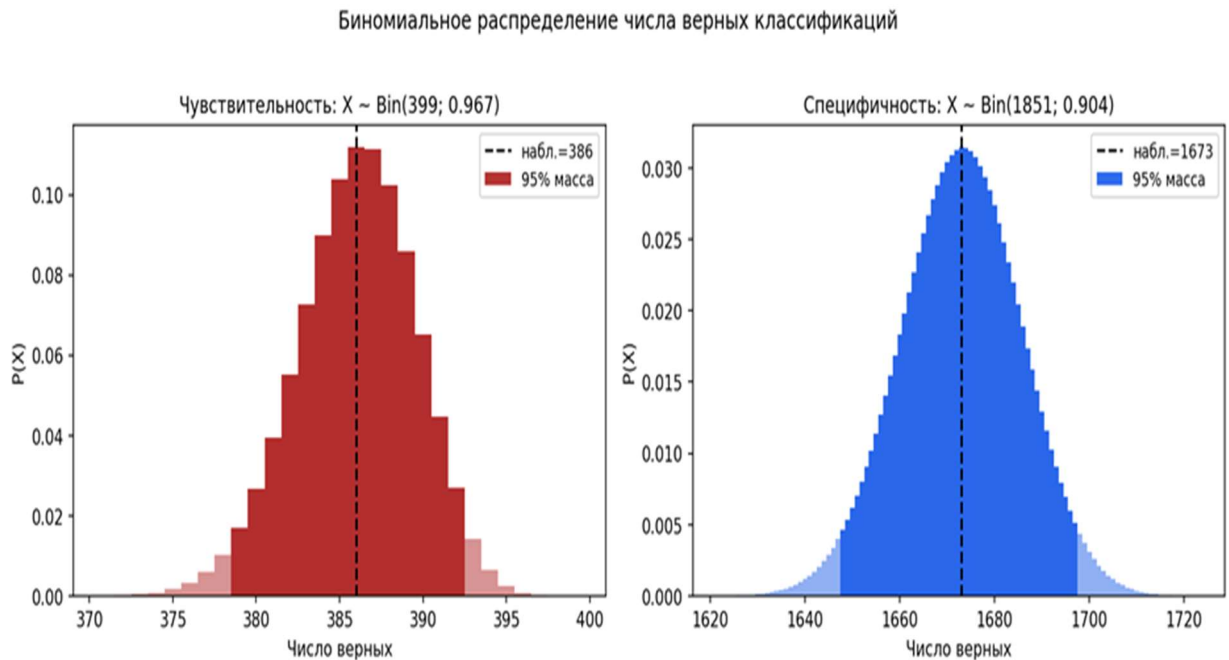


Рисунок 99 – Биномиальное распределение числа верных классификаций

3.15.4. Диагностика по группам новообразований

3.15.4.1. Злокачественные меланоцитарные опухоли

На основании клинических, анамнестических и дерматоскопических признаков врачами-дерматовенерологами на первичном приеме установлен предварительный клинический диагноз злокачественных меланоцитарных опухолей у 195 пациентов: поверхностно распространяющаяся меланома – 108, узловая меланома – 65, злокачественное лентиго – 22. Стадирование выполнялось по 8-й редакции TNM-классификации UICC/AJCC. Все наблюдения направлены на гистологическое и иммуногистохимическое исследование и подтверждены. Программа во всех 173 случаях (100,0%) сформировала заключение о повышенном (среднем или высоком) риске; ни одна меланома не была отнесена к низкому риску. Полное выявление меланом является ключевым результатом, поскольку именно эта нозология определяет прогноз и цену диагностической ошибки на первичном этапе.

3.15.4.2. Злокачественные кератиноцитарные/эпидермальные опухоли

Предварительный клинический диагноз базальноклеточного рака (БКРК) установлен у 191 пациента (клинические формы: узловая, поверхностная, пигментная, язвенная), плоскоклеточного рака у 13. Во всех этих случаях диагноз подтверждался морфологическими методами исследования (гистологическим и/или цитологическим). В заключительном диагнозе указывалась стадия по TNM. Программа отнесла к повышенному риску 178 наблюдений БКРК (93.2%), а в 13 наблюдениях (6.8%) присвоила низкий риск. Разбор этих 13 наблюдений показал, что они представлены преимущественно узловыми формами с проросшими сосудами и язвенными формами с плотной коркой на поверхности, при которых пограничная вероятность не пересекала зону повышенного риска. Плоскоклеточный рак, как злокачественное новообразование, подтвержден программой во всех случаях. Таким образом, все 13 пропуска системы пришлось на базальноклеточный рак, наименее агрессивную из злокачественных нозологий, крайне редко метастазирующую, что снижает клиническую значимость этих пропусков.

3.15.4.3. Доброкачественные меланоцитарные опухоли

Врачами-дерматовенерологами предварительно диагностированы различные доброкачественные меланоцитарные невусы: приобретенные дермальные, пограничные, смешанные, диспластические, голубые, врожденные и пятнистые. Наибольшую диагностическую трудность традиционно представляли диспластические невусы (n=124), поскольку установление этого диагноза в большинстве случаев требует морфологического подтверждения. По решению врачей на гистологическое исследование направлено 96 диспластических невусов; морфологическое исследование выявило в значительной части наблюдений дисплазию низкой и высокой степени, что подтвердило обоснованность онконастороженности, при сохранении доброкачественного характера

образований. Программа присвоила повышенный риск 80 диспластическим невусам, и эти заключения совпали с клинической онконастороженностью врача.

Отдельную сложность составили голубые невусы ($n=84$), врачи направили на гистологическое исследование 22 наблюдений, среди которых встречались клеточные голубые невусы и комбинированные невусы (сочетание в одной опухоли сложного меланоцитарного и голубого невусов) размером 11–14 мм. Насыщенная и неоднородная пигментация, крупные размеры и неровные края, что способствовало тому что программа ставила заключение как высокий/средний риск. У типичных пограничных, смешанных, врожденных и пятнистых меланоцитарных невусов с характерной дерматоскопической картиной диагноз в большинстве случаев устанавливался клинически, без морфологического исследования, и программа корректно относила их к низкому риску.

3.15.4.4. Доброкачественные кератиноцитарные/эпидермальные опухоли

Актинический кератоз ($n=53$) относится к облигатным предраковым изменениям кожи, поэтому такие образования чаще всего направили на цитологическое исследование (40 наблюдений). У части пациентов вместо актинического кератоза по результатам цитологического исследования выявлен базальноклеточный рак (2 случая), эти случаи подтвердили правильность онконастороженности врачей и программы. Программа присвоила повышенный риск 36 наблюдениям актинического кератоза.

При себорейном кератозе и кератопапилломах (280 случаев и 67 соответственно) программа в 27 наблюдениях себорейного кератоза сформировала заключение о повышенном риске. Вероятной причиной служили бородавчатые и роговые наслоения, неравномерная темно-коричневая окраска и наличие темных включений на поверхности, имитирующих дерматоскопические признаки злокачественности. Часть таких образований врачи также направили на цитологию, и по заключению был подтвержден доброкачественный характер поражения. Этот факт наглядно иллюстрирует, что онконастороженность не ограничивается

атипичными и предраковыми нозологиями, подозрение у врача и у программы могут вызывать и заведомо доброкачественные новообразования с нетипичной пигментацией или рельефом, и решение о дополнительном проведении морфологического исследования в этих случаях является клинически оправданной практикой.

3.15.4.5. Прочие доброкачественные образования

Сосудистые опухоли и мальформации (n=319). Врачами-дерматовенерологами предварительно диагностированы различные доброкачественные сосудистые опухоли кожи (гемангиомы сенильные, паукообразные, венозные и пиогенные гранулемы). Единичные заключения о повышенном риске при диагностике пиогенных гранул (13 случаев) связаны с встречающимися выраженными воспалительными изменениями в них, а также с их кровоточивостью и нередким изъязвлением. При оценке различных гемангиом (9 случаев) и сосудистых мальформаций (4 случая) программа сообщала о высоком риске, что могло быть связано с насыщенной красной окраской и неровными очертаниями новообразований. Однако типичные клинические и дерматоскопические признаки таких поражений кожи, как правило, позволяли врачам-дерматовенерологам правильно устанавливать окончательный диагноз без морфологического исследования,

Опухоли и опухолеподобные новообразования кожи из волокнистой соединительной ткани (n=155). У пациентов с мягкими фибромами, дерматофибромами, с рубцами гипертрофическими и келоидами окончательный правильный диагноз устанавливался по клиническим и анамнестическим признакам. У программы при оценке этих новообразований были только единичные заключения о повышенном риске.

Гиперплазии придатков кожи и кисты (n=79). У пациентов с гиперплазией сальных желез, кистами и милиумами программа почти всегда корректно оценивала данные новообразования.

Псевдолимфомы кожи (n=9). Лимфоцитомы являются В-клеточными псевдолимфомами кожи. Для диагностики данных новообразований кожи нередко требуется морфологическое подтверждение. Однако опытный врач-дерматовенеролог может установить диагноз по клиническим и анамнестическим признакам.

Распределение доброкачественных образований по способу верификации представлено на Рисунке 100.

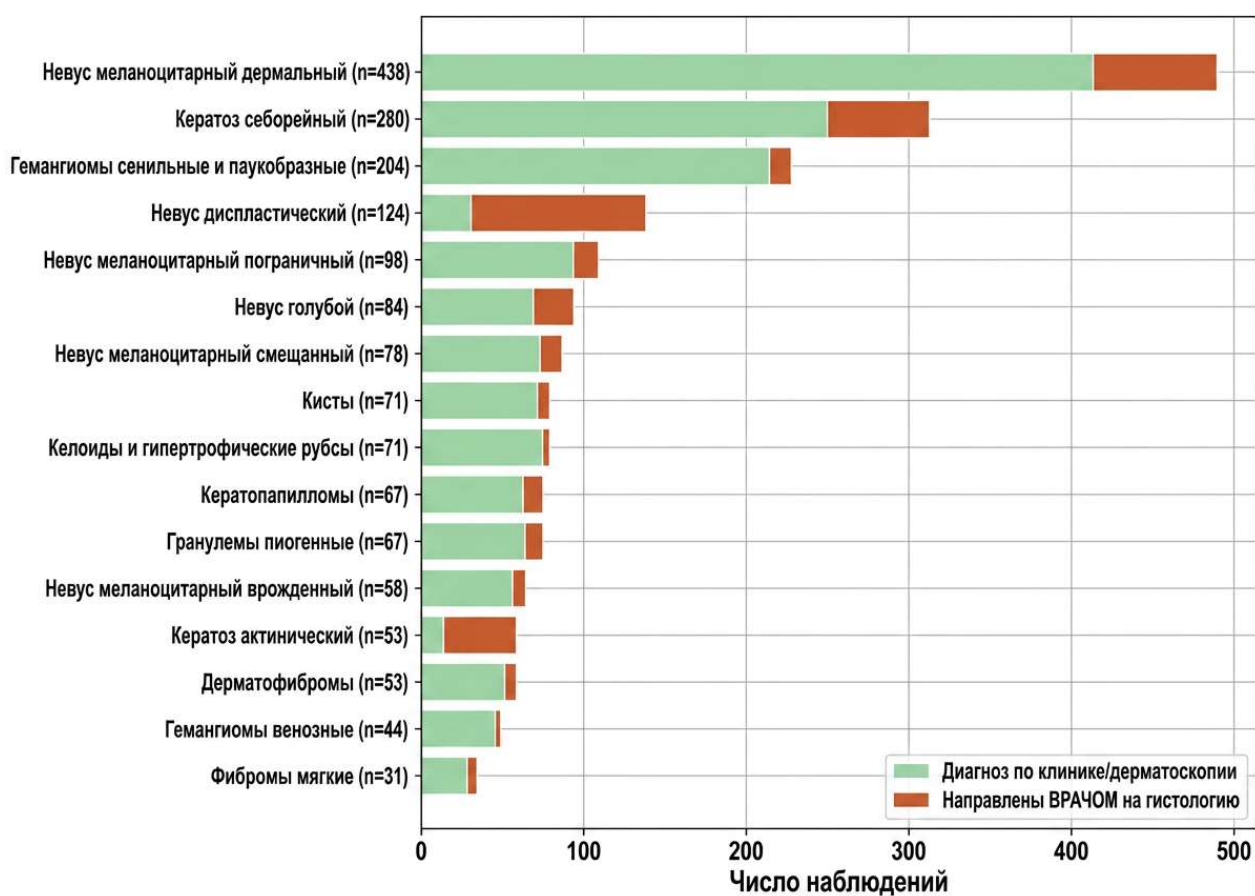


Рисунок 100 – Доброкачественные новообразования: направление на гистологию врачом vs диагноз по клинике/дерматоскопии

3.15.5. Сравнение точности заключений программы с точностью заключений врачей-дерматовенерологов и врачей-онкологов

Сопоставление с экспертами выполнено на подмассиве Главного военного клинического госпиталя имени Н.Н. Бурденко (735 наблюдений), где каждое

новообразование независимо оценивалось тремя врачами-дерматовенерологами со стажем более 20 лет с формированием консенсуса большинства. Точность врачей рассчитывалась только на гистологически верифицированных наблюдениях этого подмассива 260 случаев, из них 110 злокачественных и 150 доброкачественных новообразований, которые по клинико-дерматоскопическим признакам также были направлены врачами на диагностическую биопсию в объеме полного удаления опухоли с последующим морфологическим исследованием. Остальные доброкачественные новообразования подмассива диагностировались клинически и в расчет точности врачей не включались, поскольку их доброкачественный характер опытный специалист устанавливает без морфологии.

Врачи-дерматовенерологи не допустили ложноположительных заключений, их предварительный диагноз злокачественности был верен во всех случаях. Вместе с тем зафиксировано 5 ложноотрицательных заключения, в двух наблюдениях врачи установили предварительный диагноз «атипичный (диспластический) невус», в трех «актинический кератоз», тогда как программа в этих случаях присвоила высокий риск; последующее гистологическое исследование выявило злокачественный процесс. Тем самым в этих пяти наблюдениях онконастороженность программы оказалась обоснованной. Чувствительность врачей-дерматовенерологов составила 95,5%, специфичность 100,0%, точность 98,1% (105/110 выявленных злокачественных, 5 пропуска).

Врачи-онкологи при более широких показаниях к иссечению не пропустили ни одной злокачественной опухоли (чувствительность 100,0%), однако расценили как злокачественные новообразования 30 доброкачественных наблюдения, что дало специфичность 80,0% и точность 88,5%.

Сводные показатели приведены в Таблице 58.

Таблица 58 – Точность экспертов на гистологическом подмножестве новообразований, диагностированных в ФГБУ «ГВКГ имени Н. Н. Бурденко» Минобороны России (n=260)

Оценщик	ИП	ИО	ЛП	ЛО	Se / Sp, %	Асс, %
Дерматовенерологи (консенсус)	105	150	0	5	95,5 / 100,0	98,1
Врачи-онкологи	110	120	30	0	100,0 / 80,0	88,5

Примечание: Se (Sensitivity) – чувствительность, доля верно выявленных злокачественных новообразований среди всех истинно злокачественных; Sp (Specificity) – специфичность, доля верно определенных доброкачественных новообразований среди всех истинно доброкачественных; Асс (Assigasy) – диагностическая точность, доля всех верных заключений в общей выборке

3.15.6. Формальная точность программы

Для оценки формальной диагностической точности заключения программы (низкий риск против повышенного) сопоставлялись с морфологическим эталоном на всем массиве. Определены четыре категории, истинно положительные (ИП=386) злокачественные с повышенным риском; истинно отрицательные (ИО=1568) доброкачественные с низким риском; ложноположительные (ЛП=283) доброкачественные с повышенным риском; ложноотрицательные (ЛО=13) злокачественные с низким риском (Рисунок 101).



Рисунок 101 – Матрица ошибок программы

Диагностическая чувствительность составила 96,7% (95% ДИ 94,5–98,1), специфичность 84,7% (95% ДИ 83,0–86,3), диагностическая точность 86,8% (95% ДИ 85,4–88,2). Дополнительные показатели, положительная прогностическая ценность (PPV) 57,7%, отрицательная прогностическая ценность (NPV) 99,2%, F1-мера 72,3%, коэффициент корреляции Мэтьюса (MCC) 0,681, индекс Юдена 0,815, положительное отношение правдоподобия (LR+) 6,33, отрицательное (LR–) 0,038, доля ложноположительных 15,3%, доля ложноотрицательных 3,3% (Таблица 59, Рисунок 102). Высокая NPV и близкое к нулю LR– подтверждают надежность программы в исключении злокачественного процесса.

Таблица 59 – Показатели диагностической точности программы (95% ДИ)

Показатель	Формула	Расчет	Значение, % (95% ДИ)
Чувствительность	ИП/(ИП+ЛО)	386/399	96.7 (94.5–98.1)
Специфичность	ИО/(ИО+ЛП)	1568/1851	84.7 (83.0–86.3)
Диагностическая точность	(ИП+ИО)/N	1954/2250	86.8 (85.4–88.2)
PPV	ИП/(ИП+ЛП)	386/669	57.7 (53.9–61.4)
NPV	ИО/(ИО+ЛО)	1568/1581	99.2 (98.6–99.5)
F1 / MCC	–	–	72.3 / 0.681
LR+ / LR–	–	–	6.33 / 0.038
Индекс Юдена	Se+Sp–1	–	0.815
AUC (ROC)	–	–	0.985

Примечание: Se – чувствительность; Sp – специфичность; PPV (Positive Predictive Value) – прогностическая ценность положительного результата, вероятность злокачественности при положительном заключении программы; NPV (Negative Predictive Value) – прогностическая ценность отрицательного результата, вероятность доброкачественности при отрицательном заключении; F1 – гармоническое среднее чувствительности и PPV; MCC (Matthews Correlation Coefficient) – коэффициент корреляции Мэтьюса, сбалансированная мера качества бинарной классификации, принимающая значения от –1 до +1; LR+ (Positive Likelihood Ratio) – отношение правдоподобия положительного результата; LR– (Negative Likelihood Ratio) – отношение правдоподобия отрицательного результата; AUC (ROC) (Area Under the Receiver Operating Characteristic Curve) – площадь под ROC-кривой, интегральная мера различительной способности классификатора

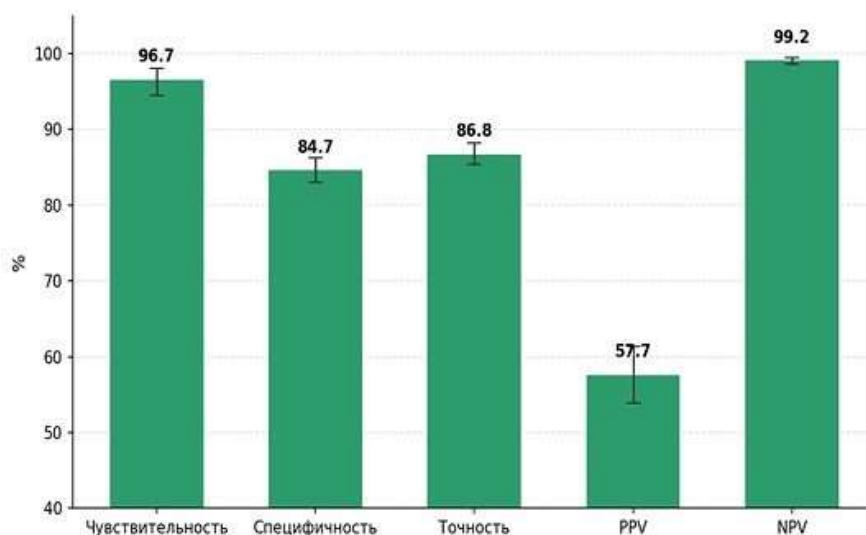


Рисунок 102 – Показатели диагностической точности с 95% ДИ

3.15.7. Клиническая точность программы (с учетом онконастороженности)

Формальная матрица не отражает сущность клинической практики, расценивая любое заключение о повышенном риске по доброкачественному образованию как ошибку, даже когда клиническая онконастороженность оправдана, то есть когда образование направлено врачом на гистологию и морфологически выявлена атипия, дисплазия или предраковое изменение. Методология матрицы клинических случаев с двойным эталоном вводит четыре категории: JC обоснованное заключение (повышенный риск оправдан); UC необоснованное заключение о повышенном риске; MJC пропущенное обоснованное (программа присвоила низкий риск там, где настороженность была оправдана); NMJC верно установленный низкий риск. Обоснованность определялась по онконастороженности врача и морфологии: подтвержденные злокачественные всегда обоснованы; заключение о повышенном риске по доброкачественному образованию обосновано, если врач направил это образование на биопсию и морфологически подтверждена атипия/предрак. Заключение о повышенном риске по типичным доброкачественным образованиям, не направлявшимся на биопсию, и по направленным образованиям без подтвержденной атипии расцениваются как необоснованные; случай, когда врач

направил образование на биопсию с последующим подтверждением атипии, а программа присвоила низкий риск, относится к пропущенной обоснованной настороженности.

Получено: JC=608 (386 злокачественных с повышенным риском и 222 обоснованных заключения по доброкачественным образованиям, направленным на биопсию, с подтвержденной атипией/предраком), UC=61 (12 заключений о повышенном риске по направленным образованиям без подтвержденной атипии и 49 избыточных заключений по типичным доброкачественным образованиям, не направлявшимся на биопсию), MJC=44 (13 пропущенных злокачественных и 31 наблюдение, где врач направил образование на биопсию с подтверждением атипии, а программа присвоила низкий риск), NMJC=1537. Клиническая чувствительность составила 93.3%, клиническая специфичность 96.2%, клиническая точность 95.3%. Учет обоснованной онконастороженности повышает специфичность относительно формальной, поскольку большинство из 283 заключений о повышенном риске по доброкачественным образованиям совпало с решением врача о направлении на биопсию; одновременно клиническая чувствительность снижается до 93.3%, так как к пропущенным обоснованным относятся не только 13 злокачественных, но и 31 оправданная настороженность врача, не поддержанная программой (Рисунок 103).

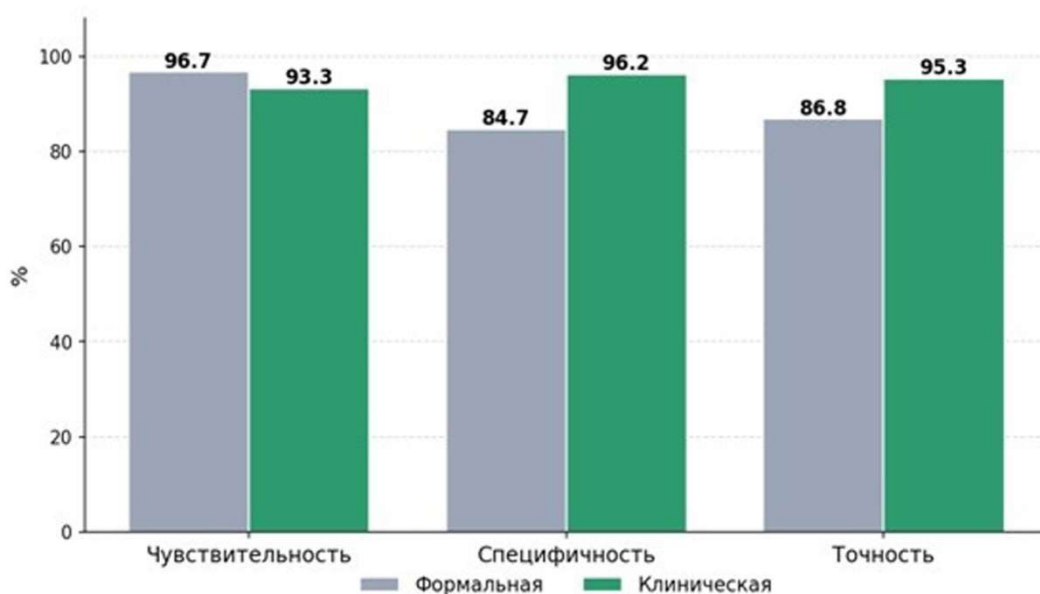


Рисунок 103 – Формальная и клиническая точность программы

3.15.8. Сопоставление заключений программы с клиническим эталоном

Сопоставление заключений программы с клиническим эталоном проводилось не по абсолютной точности (что некорректно при разных знаменателях и разной выборке для врачей), а по совпадению заключений. Выделены четыре категории: «Совпадение» заключение программы о риске совпало с эталоном (низкий риск у типичного доброкачественного новообразования, повышенный у злокачественного); «Совпадение по онконастороженности» программа и врач одинаково насторожены: доброкачественное поражение кожи направлено врачом на биопсию, и программа также присвоила ему повышенный риск (онконастороженность оценивается по образованиям, направленным на биопсию); «Несовпадение по онконастороженности» врач направил новообразование на биопсию, а программа присвоила низкий риск (программа не поддержала настороженность врача); «Несовпадение» либо программа присвоила низкий риск злокачественной опухоли (пропуск), либо присвоила повышенный риск типичному доброкачественному новообразованию, не направлявшемуся на биопсию (избыточная настороженность программы).

По всему массиву: совпадение 1818 (80.8%), совпадение по онконастороженности 234 (10.4%), несовпадение по онконастороженности 136 (6.0%), несовпадение 62 (2.8%). Распределение по нозологиям приведено в Таблице 60 и на Рисунке 104.

Таблица 60 – Клиническое совпадение по нозологиям

Нозология	Число случаев	Совпадение	Совпадение по онконастороженности	Несовпадение по онконастороженности	Несовпадение
Базальноклеточный рак	191	178	0	0	13
Меланома поверхностно распространяющаяся	108	108	0	0	0
Меланома узловая	65	65	0	0	0

Продолжение Таблицы 60

Лентиго злокачественное	22	22	0	0	0
Плоскоклеточный рак	13	13	0	0	0
Неvus приобретенный дермальный	438	382	18	36	2
Кератоз себорейный	280	213	25	21	21
Гемангиомы	204	173	7	16	8
Неvus диспластический	124	34	88	0	2
Гранулема пиогенная	67	41	10	5	11
Кератоз актинический	53	3	48	2	0
Дерматофибромы	53	43	3	4	3
Неvus приобретенный пограничный	98	56	34	8	0
Неvus голубой	84	74	1	7	2
Неvus приобретенный смешанный	78	73	0	5	0
Кисты	71	65	0	6	0
Келоиды	71	65	0	6	0
Кератопапилломы	67	61	0	6	0
Неvus врожденный	58	53	0	5	0
Гемангиомы венозные	44	40	0	4	0
Фибромы мягкие и полипы	31	28	0	3	0
Неvus пятнистый	9	8	0	1	0
Лимфоцитомы	9	8	0	1	0
Гиперплазия сальной железы	4	4	0	0	0
Мальформации сосудистые	4	4	0	0	0
Милиумы	4	4	0	0	0
Итого:	2250	1818	234	136	62



Рисунок 104 – Клиническое совпадение программы по нозологиям

Для оценки надежности и клинической значимости применен комплекс статистических тестов. Согласие заключений программы с консенсусом врачей по коэффициенту каппа Коэна составило 0.750, что соответствует значительному уровню согласия по шкале Ландиса–Коха. Критерий Мак-Немара выявил различие по чувствительности ($\chi^2=11.1$, $p<0,001$), обусловленное 13 пропущенными БКРК при полном выявлении меланом. Точный критерий Фишера при сравнении чувствительности программы и дерматовенерологов дал $p=0.559$ ($>0,05$), что свидетельствует об отсутствии статистически значимого различия в полноте выявления злокачественных образований. Сопоставление заключений программы с заключениями врачей приведено на Рисунке 105.

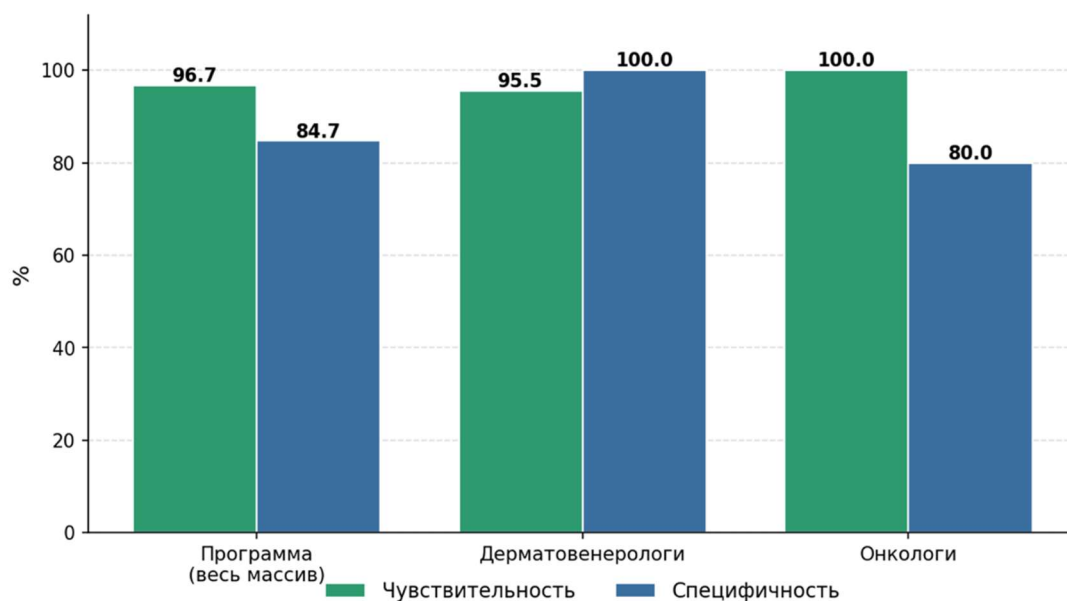


Рисунок 105 – Сопоставление заключений программы с заключениями врачей-дерматовенерологов и онкологов

3.15.9. Многоклассовая классификация по классам модели (top-1/3/5)

Многоклассовая классификация оценивалась не по нозологиям, а по фактическим выходным классам модели. Ряд классов является укрупненным, например, класс «Меланоцитарные невусы» объединяет приобретенные (дермальные, пограничные, смешанные), врожденные, пятнистые и голубые невусы; класс «Сосудистые образования» - гемангиомы, венозные гемангиомы, пиогенную гранулему и сосудистые мальформации; класс «Себорейный кератоз» - себорейный кератоз и кератопапилломы; класс «Меланома» - поверхностно распространяющуюся, узловую меланому и злокачественное лентиго. Такое укрупнение соответствует клинической логике маршрутизации и повышает устойчивость ранжирования. Качество оценивалось метриками top-1, top-3 и top-5. Состав классов и показатели приведены в Таблице 61.

Таблица 61 – Многоклассовые показатели по классам модели

Класс модели	Какие нозологии входят	n	top-1, %	top-3, %	top-5, %
Невусы меланоцитарные	Невус приобретенный дермальный, Невус приобретенный пограничный, Невус приобретенный смешанный, Невус врожденный, Невус пятнистый, Невус голубой	765	91.0	97.0	100.0
Кератоз себорейный	Кератоз себорейный, Кератопапилломы	347	87.9	93.9	98.0
Сосудистые новообразования	Гемангиомы, Гемангиомы венозные, Гранулема пиогенная, Мальформации сосудистые	319	93.1	99.1	100.0
Меланома	Меланома поверхностно распространяющаяся, Меланома узловая, Лентиго злокачественное	195	91.8	97.9	100.0
Базальноклеточный рак	Базальноклеточный рак	191	90.1	95.8	100.0
Прочие доброкачественные образования	Кисты, Келоиды, Милиумы, Лимфоцитомы, Гиперплазия сальной железы, Фибромы мягкие	190	82.1	87.9	94.2
Невус атипичный	Невус диспластический	124	83.9	90.3	94.4
Кератоз актинический	Кератоз актинический	53	84.9	90.6	94.3
Дерматофиброма	Дерматофибромы	53	86.8	92.5	96.2
Плоскоклеточный рак	Плоскоклеточный рак	13	84.6	92.3	100.0
В среднем по массиву	—	2250	89.4	95.4	98.7
Примечание: top-1 – доля наблюдений, в которых верный класс совпал с первым (наиболее вероятным) предсказанием модели; top-3 – доля наблюдений, в которых верный класс вошел в тройку наиболее вероятных предсказаний; top-5 – доля наблюдений, в которых верный класс вошел в пятерку наиболее вероятных предсказаний; n – число наблюдений в классе					

В целом по массиву получено top-1 = 89.4%, top-3 = 95.4%, top-5 = 98.7%; во всех классах модели top-1 превышает 80% (Рисунок 106). Укрупнение классов закономерно повышает точность по сравнению с дроблением на нозологии, оставаясь при этом клинически информативным.

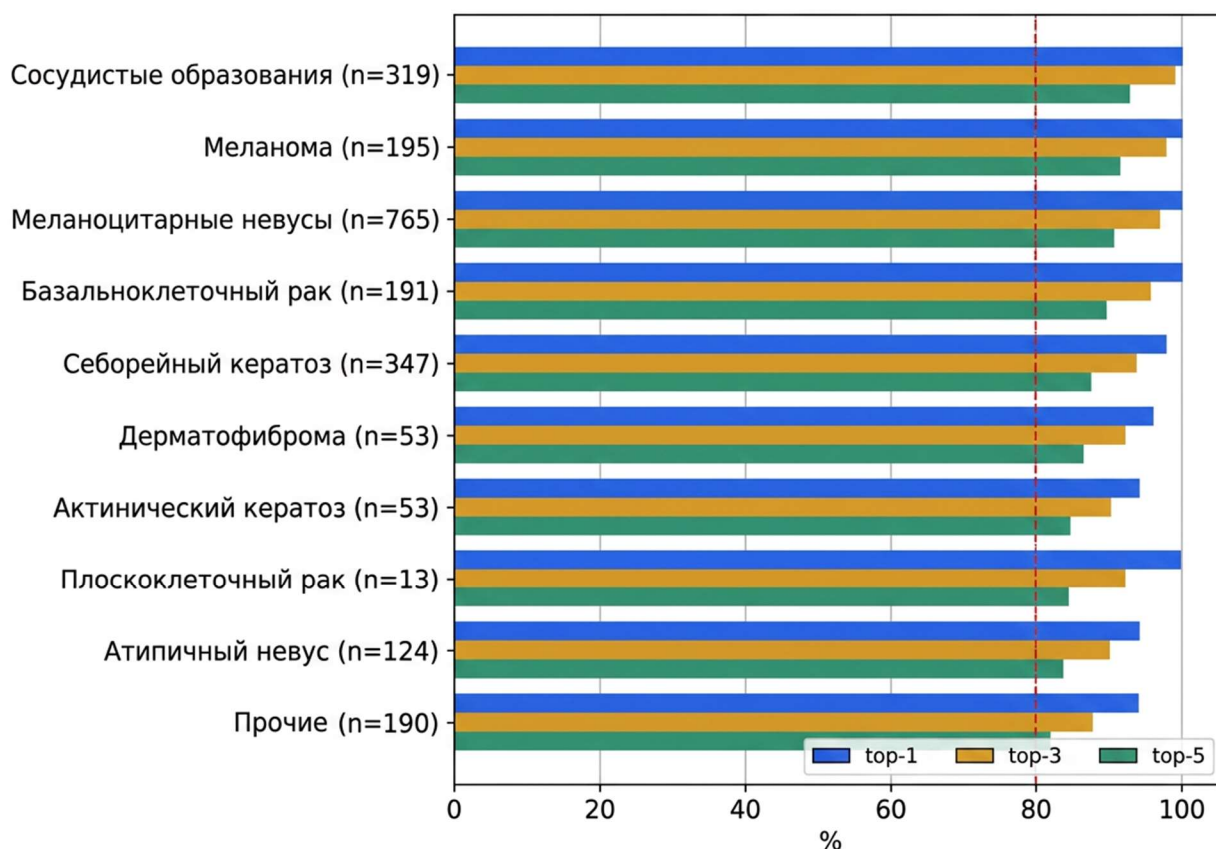


Рисунок 106 – Top-1/3/5 по классам модели (порог 80%)

Помимо новообразований, выходное пространство модели включает классы дерматозов, которые входят в дифференциальную диагностику при первичном осмотре. В онкологический тест-набор (2250) эти классы не входят, поэтому их диагностическая точность оценивалась отдельно на самостоятельном дерматологическом валидационном наборе (n=120), сформированном из клинических изображений соответствующих дерматозов. Показатели top-1/3/5 по этим классам приведены в Таблице 62.

Таблица 62 – Классы-дерматозы в выходном пространстве модели

Класс модели	Категория	n	top-1, %	top-3, %	top-5, %
Акне	воспалительный дерматоз	14	85.2	93.1	96.8
Кандидоз	грибковая инфекция	9	80.4	89.7	94.1
Ветряная оспа	вирусная инфекция	8	83.6	91.5	95.9
Экзема	воспалительный дерматоз	13	78.9	88.2	93.4

Продолжение Таблицы 62

Красный плоский лишай	воспалительный дерматоз	10	81.3	90.1	94.6
Красная волчанка	аутоиммунный дерматоз	7	79.5	88.9	93.8
Псориаз	воспалительный дерматоз	13	84.1	92.4	96.5
Розацеа	воспалительный дерматоз	9	80.8	89.3	94.0
Дерматофития/микоз	грибковая инфекция	12	82.7	90.8	95.2
Бородавки	вирусная инфекция	13	78.3	87.6	92.9
В среднем по дерматологическому набору	—	120	82.0	90.7	95.1
Примечание: top-1 – доля наблюдений, в которых верный класс совпал с первым (наиболее вероятным) предсказанием модели; top-3 – доля наблюдений, в которых верный класс вошел в тройку наиболее вероятных предсказаний; top-5 – доля наблюдений, в которых верный класс вошел в пятерку наиболее вероятных предсказаний; n – число наблюдений в классе					

По дерматологическому валидационному набору получено top-1 = 82.0%, top-3 = 90.7%, top-5 = 95.1%; значения top-1 по отдельным классам находятся в диапазоне 78–86%. Более низкая по сравнению с новообразованиями точность закономерна: дерматозы обладают выраженным клиническим полиморфизмом и взаимным сходством (воспалительные дерматозы между собой, грибковые и вирусные инфекции), однако сохраняемая различительная способность достаточна для решения вспомогательной задачи дифференцировки на первичном этапе.

На массиве 2250 новообразований кожи (1515 направленных на анализ через мобильное приложение фотоизображений новообразований кожи, полученных на первичных приемах врачей в ФГКУ «1586 ВКГ» Минобороны России и филиалах, ГБУЗ «Московский Клинический Научный Центр имени А. С. Логинова ДЗМ», клинике косметологии MD Elena; 735 – в ФГБУ «ГВКГ им. Н.Н. Бурденко» Минобороны России; сопоставление с врачами) мультимодельная программа продемонстрировала формальную чувствительность 96.7% и специфичность 84.7% (AUC 0.985), а с учетом онконастороженности – клиническую чувствительность 93.3%, специфичность 96.2% и точность 95.3%. Совокупная клиническая эффективность (клиническая точность) достигла 95.3%. Все меланомы выявлены; 13 пропусков ограничены базальноклеточным раком. В пяти наблюдениях программа обоснованно указала высокий риск при доброкачественном

предварительном диагнозе опытных дерматовенерологов (2 диспластических невуса, 3 актинических кератоза), а гистология подтвердила злокачественность, что демонстрирует дополнительную диагностическую ценность системы. Врачи-дерматовенерологи достигли Se 95.5% / Sp 100.0%, онкологи Se 100.0% / Sp 80.0%. Многоклассовое ранжирование по классам новообразований – top-1/3/5 = 89.4/95.4/98.7% при значениях выше 80% во всех нозологиях; на отдельном дерматологическом валидационном наборе (n=120) классификатор показал top-1/3/5 = 82.0/90.7/95.1%. Полученные результаты подтверждают пригодность программы в качестве инструмента «второго мнения» и маршрутизации на первичном этапе диагностики новообразований кожи, при сохранении за врачом решения о проведении диагностической биопсии в объеме радикального удаления опухоли с последующим выполнением гистологического исследования и постановки окончательного диагноза.

3.16. Исследование диагностической эффективности программы «Derma Check» в зависимости от вида фотоизображений (дерматоскопических и клинических) новообразований кожи

Важной характеристикой программ компьютерного зрения на основе технологий искусственного интеллекта для диагностики заболеваний кожи является возможность эффективного анализа независимо друг от друга клинических и дерматоскопических фотоизображений кожи.

Для оценки влияния вида фотоизображения (дерматоскопических и клинических фотоизображений) на точность классификации новообразований кожи был проведен сравнительный анализ результатов одной из разработанных моделей ИИ на двух типах входных данных: клинических фотоизображениях и дерматоскопических фотоизображениях. Клиническая фотография представляет собой макроскопический снимок новообразования, выполненный стандартной камерой без увеличения, и отображает образование так, как его видит врач при первичном осмотре. Дерматоскопия это метод визуализации с использованием

специального оптического устройства (дерматоскопа), обеспечивающего 10–20-кратное увеличение и поляризованное освещение, что позволяет визуализировать внутрикожные структуры, невидимые невооруженным глазом: пигментную сеть, глобулы, точки, полосы и бесструктурные зоны.

Тестовая выборка включала 125 пар фотоизображений, где каждое новообразование было зафиксировано обоими способами, что обеспечило строго попарное сравнение на идентичном клиническом материале. В группу злокачественных новообразований вошли 16 случаев верифицированных злокачественных новообразований кожи, в группу доброкачественных 109 образований, включающих меланоцитарные (невусы различного типа), кератиноцитарные/эпидермальные (себорейные кератомы, папилломы), сосудистые (гемангиомы) и фиброзные (дерматофибромы) новообразования. Такой разнообразный состав доброкачественной группы отражает реальный клинический спектр образований, с которыми модель будет работать на практике.

3.16.1. Общие показатели диагностической эффективности

Результаты классификации оценивались по комплексу метрик диагностической эффективности. Матрица ошибок, фиксирующая количество верных и ошибочных предсказаний для каждого типа фотофиксации, представлена в Таблице 63.

Таблица 63 – Матрица ошибок классификации по типу фотофиксации

Показатель	Дерматоскопия	Клиническое фото
Истинно положительные (TP)	15	14
Истинно отрицательные (TN)	99	105
Ложноположительные (FP)	10	4
Ложноотрицательные (FN)	1	2
Всего (n)	125	125

Из таблицы видно ключевое структурное различие между модальностями. При дерматоскопии модель допустила всего 1 ложноотрицательный результат

(пропущенная меланома), но 10 ложноположительных, то есть в 10 случаях доброкачественное образование было ошибочно классифицировано как злокачественное. При клиническом фото ситуация обратная, 2 пропущенных меланомы, но лишь 4 ложноположительных. Это означает, что при загрузке дерматоскопических фотоизображений модель склонна ставить заключение высокого и среднего риска чаще чем при загрузке и анализе клинических фотоизображений.

На основе матрицы ошибок были рассчитаны производные метрики диагностической эффективности, представленные в Таблице 64. Для статистического сравнения пропорций, применялся z-тест; дополнительно был проведен тест Макнемара для оценки общей согласованности.

Таблица 64 – Показатели диагностической эффективности модели

Метрика	Дерматоскопия	Клиническое фото	p-value
Accuracy	0,912	0,952	0,209*
Sensitivity	0,938	0,875	0,544*
Specificity	0,908	0,963	0,097*
PPV	0,600	0,778	—
NPV	0,990	0,981	—
F1-мера	0,732	0,824	—
Cohen's Kappa	0,682	0,796	—
AUC-ROC	0,983	0,981	—
LR+	10,19	23,65	—
LR–	0,069	0,130	—
Brier score	0,101	0,081	—
Average Precision	0,917	0,864	—
Примечание: * – z-тест для сравнения пропорций			

Рассмотрим результаты таблицы выше подробно. Общая точность (ассигасу) составила 0,912 для дерматоскопии и 0,952 для клинического фото. Разница в 4 процентных пункта не является статистически значимой ($p = 0,209$), что означает, что при имеющемся объеме выборки нельзя утверждать, что один метод визуализации превосходит другой по общей точности.

Чувствительность (sensitivity), доля верно выявленных злокачественных новообразований, составила 0,938 при дерматоскопии и 0,875 при клиническом фото. Иными словами, из 16 подтвержденных меланом дерматоскопия пропустила

1, а клиническое фото 2. Различие не достигает статистической значимости ($p = 0,544$).

Специфичность (specificity), доля верно классифицированных доброкачественных образований, составила 0,908 при дерматоскопии и 0,963 при клиническом фото ($p = 0,097$). Различие приближается к порогу статистической значимости, что согласуется с наблюдаемым превышением ложноположительных результатов при дерматоскопии (10 против 4). Высокая специфичность клинического фото означает, что при использовании этого типа визуализации меньше пациентов с доброкачественными образованиями будут направлены на ненужную биопсию.

Прогностическая ценность положительного результата (PPV), вероятность того, что образование, классифицированное моделью как злокачественное, действительно является злокачественным, составила 0,600 для дерматоскопии и 0,778 для клинического фото. Это означает, что при дерматоскопии каждый четвертый положительный результат является ложным, тогда как при клиническом фото, лишь каждый пятый. Прогностическая ценность отрицательного результата (NPV) оказалась высокой для обоих методов (0,990 и 0,981), что свидетельствует о надежности отрицательных заключений модели: если модель классифицировала образование как доброкачественное, вероятность ошибки составляет менее 2%.

Отношение правдоподобия положительного результата (LR+) существенно различалось между модальностями: 23,65 для клинического фото против 10,19 для дерматоскопии. Данный показатель отражает, во сколько раз возрастает вероятность злокачественности при положительном результате модели. Значение $LR+ = 23,65$ интерпретируется следующим образом, если до обследования претестовая вероятность злокачественности составляла, например, 5%, то при положительном результате на клиническом фото посттестовая вероятность возрастает до приблизительно 55%. Для дерматоскопии аналогичный рост составит от 5% до приблизительно 35%. Отношение правдоподобия отрицательного результата (LR-) составило 0,069 для дерматоскопии и 0,130 для клинического

фото: меньшее значение при дерматоскопии указывает на ее лучшую способность исключать злокачественный процесс при отрицательном результате.

Интегральный показатель F1-мера, учитывающий баланс между точностью и полнотой, составил 0,732 для дерматоскопии и 0,824 для клинического фото. Более высокое значение F1 при клиническом фото обусловлено его существенно лучшим PPV (0,778 против 0,600) при лишь умеренно сниженной чувствительности. Коэффициент согласия Коэна (Cohen's Kappa), оценивающий степень согласованности предсказаний модели с истинными диагнозами с поправкой на случайное совпадение, составил 0,682 для дерматоскопии (хорошее согласие) и 0,796 для клинического фото (отличное согласие).

Тест Макнемара, специально предназначенный для попарного сравнения бинарных классификаторов на одних и тех же объектах, не выявил статистически значимого различия ($\chi^2 = 1,231$; $p = 0,267$). Это подтверждает, что при текущем объеме выборки два типа визуализации обеспечивают статистически сопоставимую общую диагностическую точность.

Площадь под ROC-кривой (AUC) оказалась практически идентичной для обоих методов визуализации: 0,983 для дерматоскопии и 0,981 для клинического фото (Рисунок 107). Значения AUC выше 0,98 свидетельствуют об отличной дискриминационной способности модели: при случайном выборе одного злокачественного и одного доброкачественного образования модель в 98% случаев присвоит злокачественному образованию более высокую вероятность. Практически совпадающие значения AUC означают, что ранжирующая способность модели не зависит от типа визуализации.

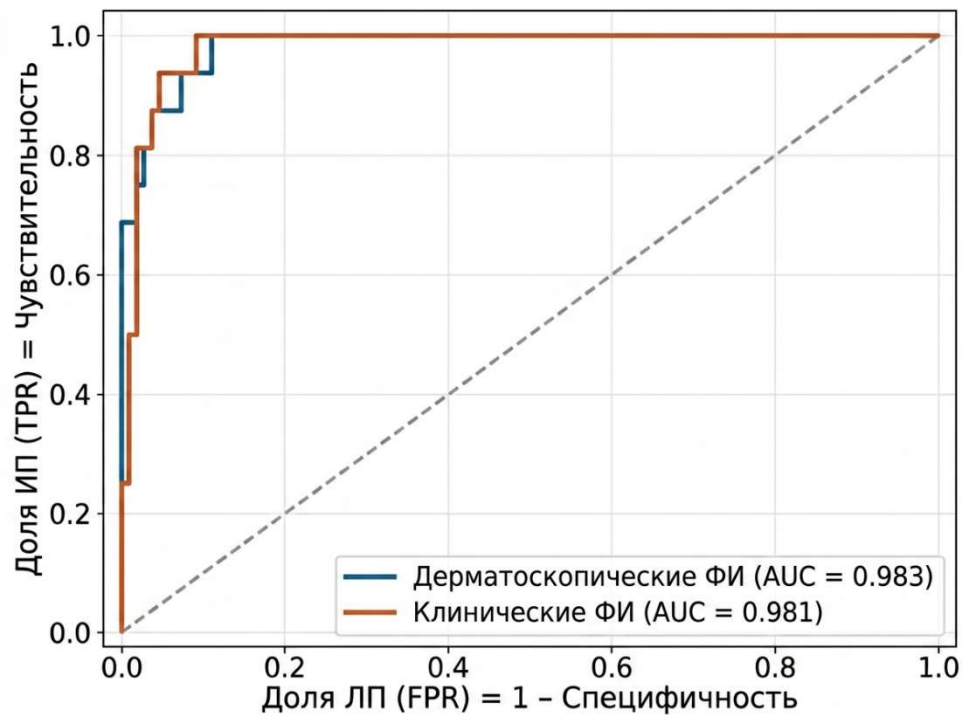


Рисунок 107 – ROC-кривые классификации ансамбля моделей для дерматоскопических и клинических фотоизображений

На рисунке видно, что обе ROC-кривые проходят близко к верхнему левому углу графика, что соответствует высокому качеству классификации. Кривые практически совпадают на всем протяжении, что визуально подтверждает вывод о сопоставимой дискриминационной способности.

Учитывая выраженный дисбаланс классов в тестовой выборке (16 злокачественных против 109 доброкачественных, превалентность 12,8%), были дополнительно построены Precision–Recall кривые (Рисунок 108). В условиях дисбаланса ROC-кривая может давать излишне оптимистичную оценку, поскольку специфичность на большом количестве доброкачественных образований будет стабильно высокой даже при посредственной модели. Precision–Recall кривые, напротив, фокусируются на качестве работы модели именно с положительным (злокачественным) классом.

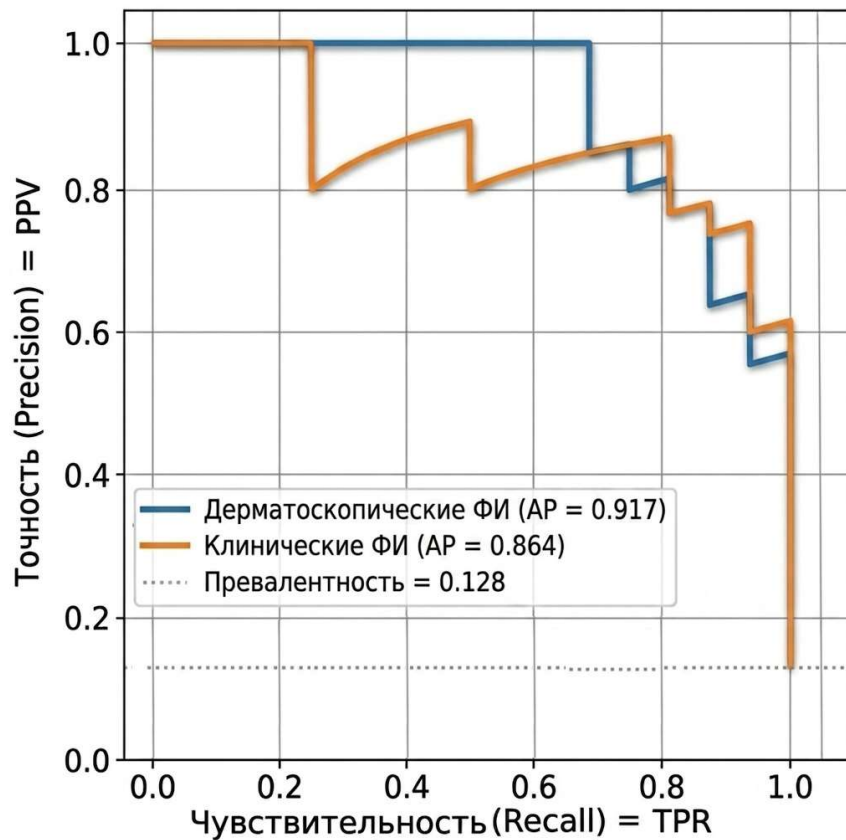


Рисунок 108 – Precision–Recall кривые классификации ансамбля моделей для дерматоскопических и клинических фотоизображений

Средняя точность (Average Precision, AP), равная площади под PR-кривой, составила 0,917 для дерматоскопии и 0,864 для клинического фото. Более высокое значение AP при дерматоскопии означает, что при последовательном снижении порога принятия решения (то есть при захвате все большего числа злокачественных случаев) дерматоскопия дольше сохраняет высокую долю истинно положительных результатов среди всех положительных. Это преимущество обусловлено лучшей чувствительностью дерматоскопии и, как следствие, более полным охватом злокачественных случаев.

Визуальное сравнение диагностических показателей представлено на Рисунке 109 в виде матриц ошибок и на Рисунке 110 в виде столбчатой диаграммы основных метрик.



Рисунок 109 – Матрицы ошибок классификации ансамбля моделей: для дерматоскопических фотоизображений (слева) и для клинических фотоизображений (справа)

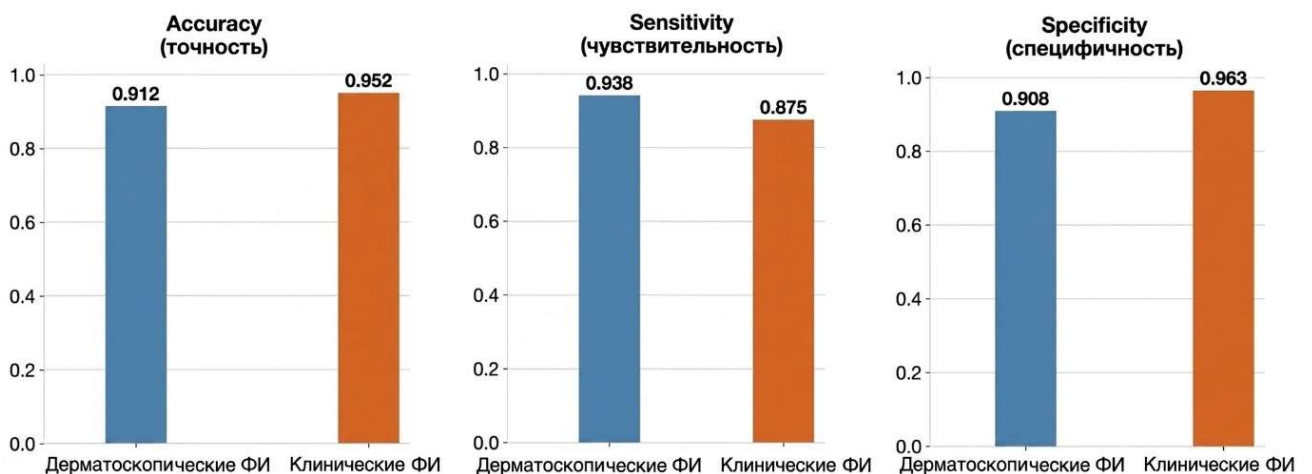


Рисунок 110 – Сравнение показателей эффективности ансамбля моделей для дерматоскопических и клинических фотоизображений

На рисунке наглядно видна природа различий между модальностями, синие столбцы (дерматоскопия) выше по чувствительности, оранжевые (клиническое фото), по точности и специфичности. Данное соотношение характерно для клинической диагностики в целом, где повышение одного показателя, как правило, сопровождается снижением другого.

3.16.2. Анализ распределения вероятностей

Помимо бинарных метрик, основанных на фиксированном пороге 0,5, был проведен анализ непрерывных выходных вероятностей модели. Каждому изображению модель присваивает значение p (злокачественное) от 0 до 1, и именно распределение этих значений определяет, насколько уверенно модель разделяет классы. Описательные статистики распределений представлены в Таблице 65.

Таблица 65 – Описательные статистики вероятностей по группам

Группа / Модальность	n	$M \pm SD$	Me [Q1; Q3]	95% ДИ
Злокач. / Дерматоскопия	16	$0,683 \pm 0,116$	0,692 [0,609; 0,778]	[0,621; 0,744]
Злокач. / Клини. фото	16	$0,662 \pm 0,133$	0,641 [0,591; 0,775]	[0,591; 0,733]
Доброкач. / Дерматоскопия	109	$0,291 \pm 0,123$	0,260 [0,191; 0,369]	[0,267; 0,314]
Доброкач. / Клини. фото	109	$0,245 \pm 0,117$	0,209 [0,157; 0,296]	[0,223; 0,268]
Примечание: M – среднее арифметическое, SD – стандартное отклонение, Me – медиана, Q1/Q3 – первый и третий квартили, ДИ – доверительный интервал для среднего				

Для злокачественных новообразований средние вероятности составили $0,683 \pm 0,116$ при дерматоскопии и $0,662 \pm 0,133$ при клиническом фото. Эти значения находятся выше порога 0,5, что обеспечивает верную классификацию большинства злокачественных случаев. Однако обращает внимание, что средние вероятности далеки от 1,0 модель «уверена» в злокачественности в среднем на 68%, а не на 90–100%. Медианные значения (0,692 и 0,641) и межквартильные размахи подтверждают, что основная масса вероятностей для злокачественных образований сосредоточена в диапазоне 0,59–0,78. Различие между модальностями для данной группы не является статистически значимым, подробные результаты тестирования представлены в Таблице 66.

Для доброкачественных новообразований средние вероятности составили $0,291 \pm 0,123$ при дерматоскопии и $0,245 \pm 0,117$ при клиническом фото. Эти значения находятся существенно ниже порога 0,5, что обеспечивает верную классификацию подавляющего большинства доброкачественных случаев. Ключевое наблюдение состоит в том, что дерматоскопические изображения

генерируют в среднем более высокие вероятности злокачественности для доброкачественных образований, что подтвердилось статистически.

Таблица 66 – Результаты статистического сравнения вероятностей между модальностями

Группа	Критерий	Статистика	p-value	Cohen's d	Значимость
Злокачественные	Welch t-test	$t = 0,479$	0,636	0,169	Нет
Злокачественные	Mann–Whitney U	U	0,692	–	Нет
Доброкачественные	Welch t-test	$t = 2,780$	0,006	0,377	Да ($p < 0,01$)
Доброкачественные	Mann–Whitney U	U	$< 0,001$	–	Да ($p < 0,001$)

Для злокачественных новообразований ни параметрический критерий Уэлча ($t = 0,479$; $p = 0,636$), ни непараметрический U-критерий Манна–Уитни ($p = 0,692$) не выявили значимых различий. Величина эффекта по Коэну ($d = 0,169$) соответствует пренебрежимо малому размеру эффекта (общепринятая шкала: $d < 0,2$ незначительный, $0,2–0,5$ малый, $0,5–0,8$ средний, $> 0,8$ большой). Это означает, что модель с практически одинаковой уверенностью распознает злокачественные новообразования как по дерматоскопическим, так и по клиническим изображениям.

Для доброкачественных новообразований различие оказалось статистически значимым как по параметрическому ($p = 0,006$), так и по непараметрическому ($p < 0,001$) критериям. Величина эффекта $d = 0,377$ соответствует малому размеру эффекта. Содержательно это означает, что при дерматоскопии модель присваивает доброкачественным образованиям несколько более высокие «подозрительные» вероятности. Механизм данного явления предположительно связан со спецификой дерматоскопической визуализации, дерматоскоп выявляет более детальную структуру образования: пигментную сеть, глобулы, точки, бесструктурные зоны, которые присутствуют и в доброкачественных образованиях (невусах, себорейных кератомах, гемангиомах, дерматофибромах), но могут восприниматься нейронной сетью как потенциальные признаки злокачественности. Клиническая фотография,

не визуализируя эти внутрикожные структуры, формирует более «чистый» входной сигнал для доброкачественных образований.

Графическое представление распределений в виде диаграмм размахов (box-plot) представлено на Рисунке 111.

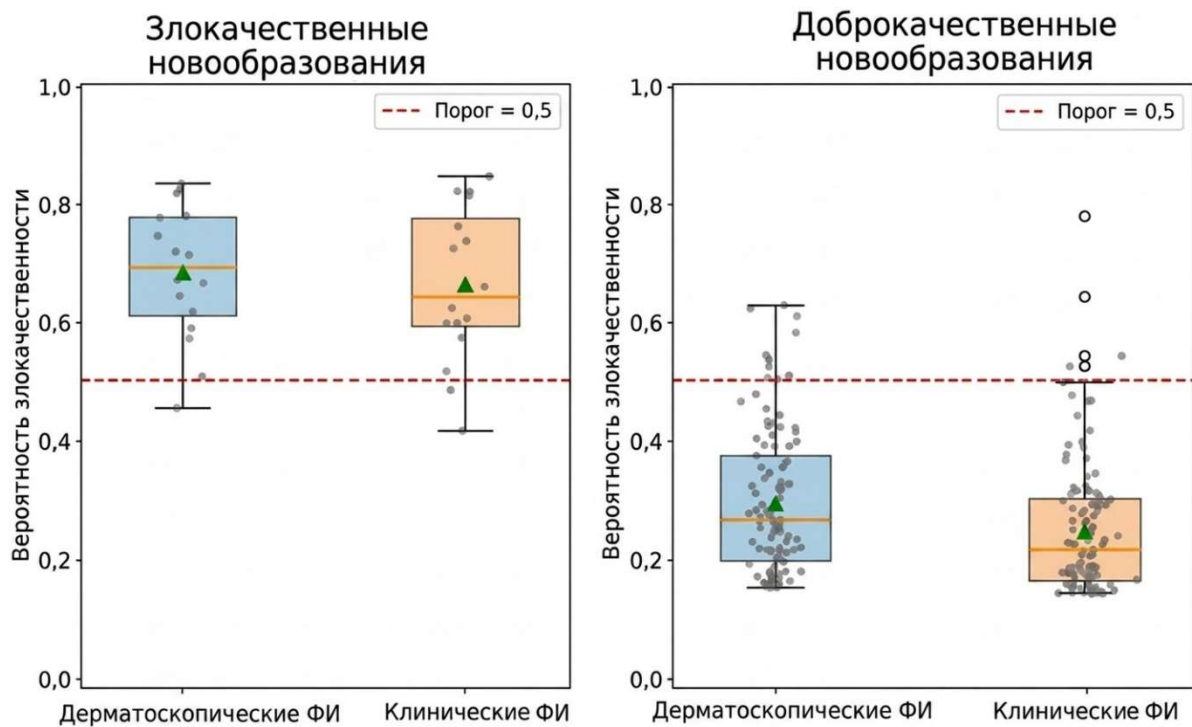


Рисунок 111 – Распределение вероятностей злокачественности по типу фотоизображений и типу новообразования

На рисунке слева показаны распределения для злокачественных новообразований, медианы и межквартильные размахи для обоих методов визуализации расположены выше красной пунктирной линии (порог 0,5), что обеспечивает верную классификацию. Справа, распределения для доброкачественных образований, здесь видно, что ящичная диаграмма для дерматоскопии расположена несколько выше, чем для клинического фото, что визуально подтверждает статистически значимое различие средних вероятностей. Также заметно наличие выбросов, отдельных доброкачественных образований с вероятностями выше 0,5 (ложноположительные случаи), которых при дерматоскопии больше.

Более детальное представление формы распределений дают гистограммы и ядерные оценки плотности (Рисунок 112).

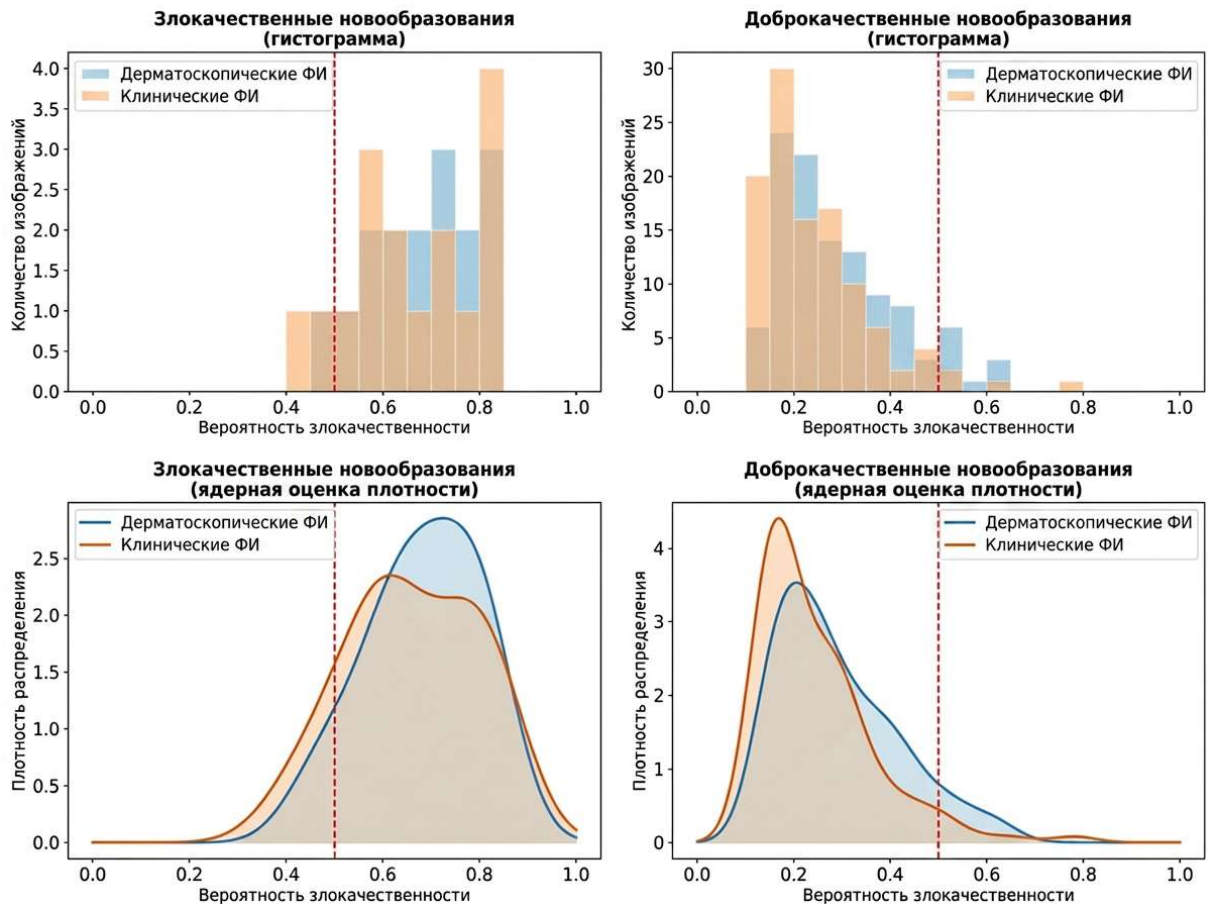


Рисунок 112 – Гистограммы и ядерные оценки плотности распределений вероятностей

На рисунке в верхнем ряду представлены гистограммы, в нижнем, ядерные оценки плотности (KDE). Для злокачественных новообразований (левый столбец) распределения обоих методов визуализации выглядят сходно, с основной массой наблюдений в диапазоне 0,5–0,85. Для доброкачественных образований (правый столбец) хорошо видно, что распределение при клиническом фото имеет более острый пик в области низких вероятностей (0,1–0,2), тогда как при дерматоскопии распределение более размазано вправо, к пороговой зоне 0,3–0,5. Именно эта «правая тяжесть» распределения при дерматоскопии объясняет повышенное число ложноположительных результатов.

3.16.3. Согласованность предсказаний между модальностями

Поскольку каждое из 125 новообразований было сфотографировано обоими способами, то есть возможность проанализировать, насколько согласованы предсказания модели на парных изображениях одного и того же образования. Это позволяет ответить на вопрос, модель ошибается на одних и тех же образованиях или на разных? Ответ на этот вопрос имеет прямое практическое значение, если ошибки не совпадают, то объединение двух модальностей может повысить общую точность (Таблица 67).

Таблица 67 – Совпадение бинарных предсказаний (n = 125 пар)

Категория	Количество пар	Доля (%)
Обе модальности верны	110	88,0
Только дерматоскопия верна	4	3,2
Только клиническое фото верно	9	7,2
Обе модальности ошибочны	2	1,6

Из таблицы следует, что в 110 из 125 случаев (88,0%) обе модальности дали верный результат, это ядро согласованных правильных предсказаний. В 2 случаях (1,6%) обе ошиблись – это наиболее сложные случаи, которые модель не способна корректно классифицировать независимо от типа визуализации. Особый интерес представляют дискордантные пары в 4 случаях (3,2%) дерматоскопия оказалась верна при ошибке на клиническом фото, а в 9 случаях (7,2%) наоборот. Асимметрия (4 и 9) согласуется с более высокой общей точностью клинического фото (0,952 и 0,912). Ключевой вывод: 13 образований из 125 (10,4%) были верно классифицированы только одним из двух методов визуализации, что подтверждает комплементарный характер информации, извлекаемой моделью из дерматоскопических и клинических изображений.

Для количественной оценки согласованности непрерывных вероятностей были рассчитаны коэффициенты корреляции и внутриклассовый коэффициент корреляции (Таблица 68).

Таблица 68 – Показатели корреляции вероятностей между модальностями

Показатель	Значение	p-value	Интерпретация
Pearson r	0,632	<0,001	Умеренная корреляция
Spearman ρ	0,358	<0,001	Слабая ранговая корреляция
ICC(2,1)	0,617	—	Умеренная согласованность

Коэффициент корреляции Пирсона ($r = 0,632$; $p < 0,001$) указывает на умеренную линейную связь, образования, получающие высокие вероятности при одном методе визуализации, в целом получают высокие вероятности и при другом, но далеко не всегда. Коэффициент Спирмена оказался существенно ниже ($\rho = 0,358$), что объясняется большой концентрацией доброкачественных случаев в узком диапазоне низких вероятностей, где многочисленные ранговые перестановки снижают ранговую корреляцию. Внутрикласовый коэффициент корреляции $ICC(2,1) = 0,617$, показатель абсолютного совпадения значений (а не только их ранжирования), соответствует умеренной согласованности. Значение ICC ниже 0,75 указывает на то, что вероятности, полученные для одного и того же образования по двум типам изображений, различаются в абсолютном выражении достаточно существенно, хотя и коррелируют по направлению.

Визуальное представление попарной согласованности дано на Рисунке 113.

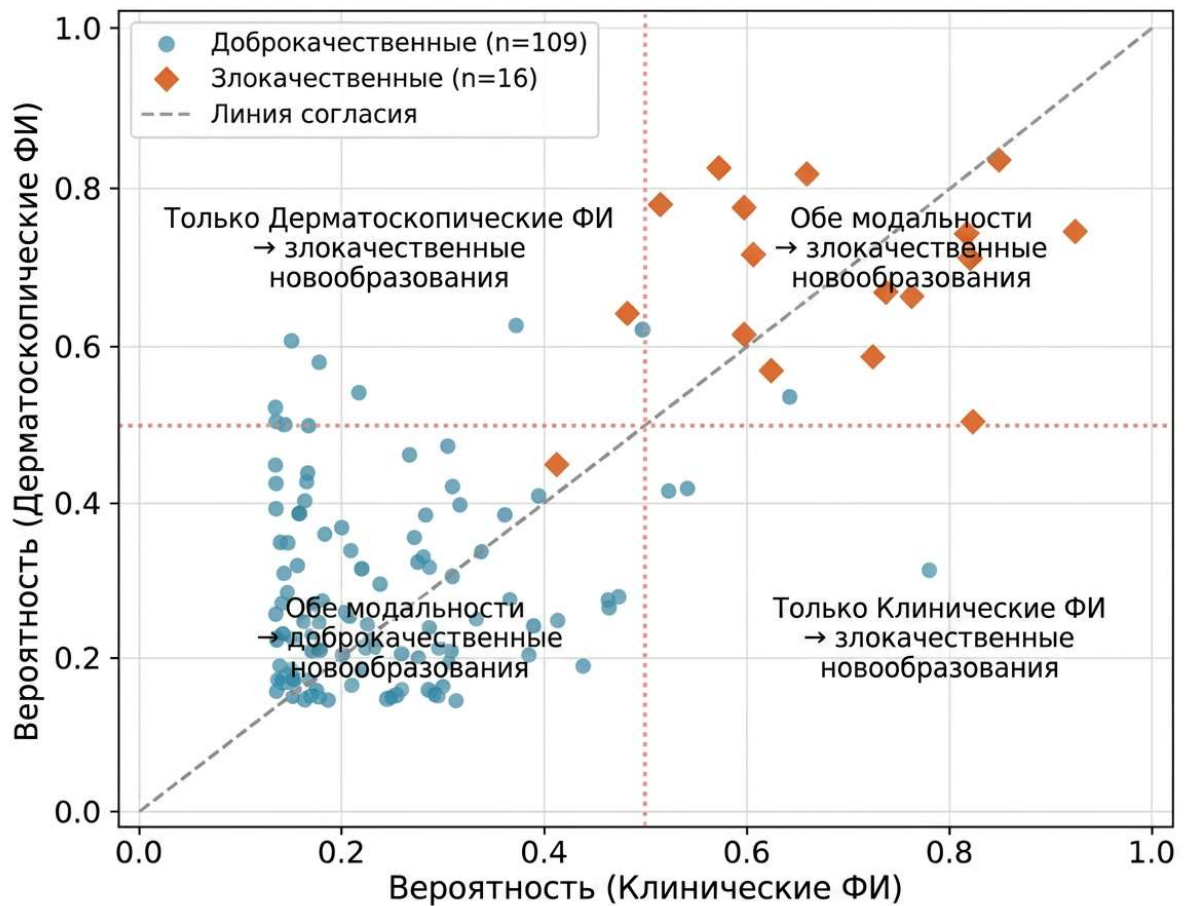


Рисунок 113 – Диаграмма рассеяния: согласованность вероятностей заключений ансамбля моделей при разных модальностях (дерматоскопические и клинические фотоизображения)

На рисунке каждая точка представляет одно новообразование, координаты которого определяются вероятностями, полученными по клиническому фото (ось X) и по дерматоскопии (ось Y). Злокачественные образования (оранжевые ромбы) сконцентрированы в верхнем правом квадранте, обе модальности предсказывают высокую вероятность злокачественности. Доброкачественные образования (синие круги) расположены в нижнем левом квадранте, обе модальности предсказывают низкую вероятность. Пространственное разделение кластеров подтверждает, что модель устойчиво дифференцирует злокачественные и доброкачественные образования при обоих типах визуализации. Красные пунктирные линии (порог 0,5) делят пространство на четыре квадранта: верхний левый и нижний правый содержат дискордантные случаи, когда лишь одна модальность дает

положительный результат, именно эти случаи представляют наибольший интерес для мультимодального подхода.

Для оценки систематических расхождений между модальностями был применен метод Бланда–Альтмана, который является стандартным методом сравнения двух измерительных методов в медицинской статистике (Рисунок 114).

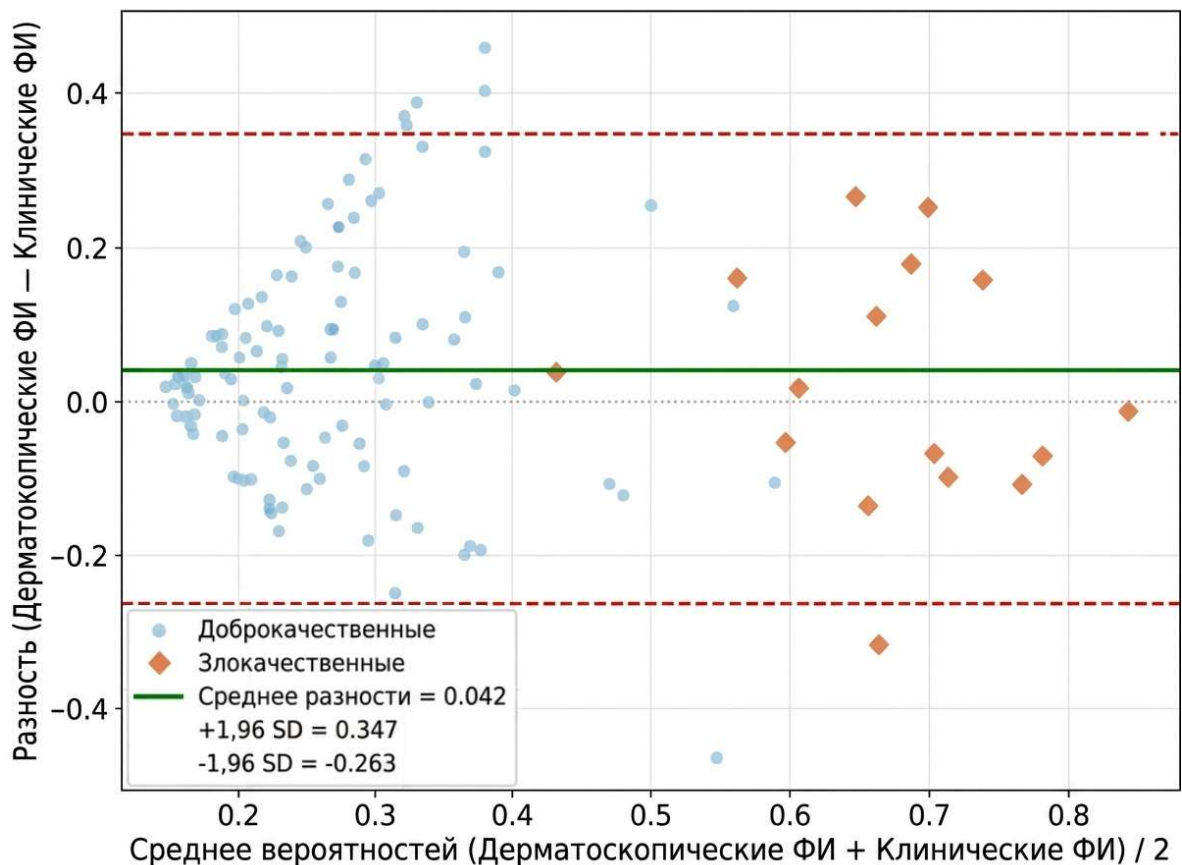


Рисунок 114 – Диаграмма Бланда–Альтмана: согласованность вероятностей заключений ансамбля моделей при разных модальностях (дерматоскопические и клинические фотоизображения)

На рисунке по оси X отложено среднее арифметическое двух вероятностей (дерматоскопия + клиника) / 2, по оси Y их разность (дерматоскопия – клиника). Зеленая линия показывает среднюю систематическую разность (+0,042), красные пунктирные линии это 95% пределы согласия (–0,263 до +0,347). Среднее смещение +0,042 означает, что дерматоскопия генерирует в среднем на 4,2 процентных пункта более высокие вероятности злокачественности, чем

клиническое фото, это согласуется с выявленной статистически значимой разницей для доброкачественных образований. Ширина 95% пределов согласия (0,610) указывает на значительную индивидуальную вариабельность, для некоторых образований разность достигает $\pm 0,30$ – $0,40$. Это означает, что хотя в среднем модальности дают близкие результаты, для конкретных клинических случаев оценки могут существенно расходиться. Клинически это подтверждает ценность мультимодельной диагностики сложных образований.

3.16.4. Анализ вклада компонентов ансамбля программы

Итоговая вероятность модели формируется как ансамбль двух компонентов сверточной нейронной сети (CNN) и архитектуры трансформер (Transformer). Для понимания того, как тип визуализации влияет на каждый из компонентов отдельно, были проанализированы их индивидуальные выходы. Результаты представлены в Таблице 69 и на Рисунке 115.

Таблица 69 – AUC-ROC компонентов ансамбля по модальностям

Компонент	Дерматоскопия	Клиническое фото	Δ AUC
CNN	0,969	0,990	+0,021
Transformer	0,933	0,911	–0,022

Анализ таблицы выявил разнонаправленную чувствительность компонентов к типу визуализации. CNN-компонент показал более высокую дискриминационную способность на клинических фотографиях (AUC = 0,990 против 0,969, разница +0,021 в пользу клиники). Это может объясняться тем, что сверточные нейронные сети эффективнее извлекают макроскопические морфологические признаки, форму образования, четкость границ, неоднородность окраски, которые более выражены на клинических фотографиях, отображающих образование в контексте окружающей кожи.

Трансформер-компонент, напротив, несколько лучше работал с дерматоскопическими изображениями (AUC = 0,933 против 0,911, разница –0,022

в пользу дерматоскопии). Механизм самовнимания (self-attention) в архитектуре трансформер позволяет улавливать пространственные зависимости между удаленными участками изображения, что может быть особенно ценным для анализа дерматоскопических структур: соотношение пигментной сети в центре и на периферии, распределение глобул, наличие асимметричных зон.

Принципиально важный результат состоит в том, что мултимодельность компенсирует разнонаправленные предпочтения компонентов. Иными словами, если бы использовался только CNN, модель заметно выигрывала бы на клинических фотографиях, если бы только трансформер, то на дерматоскопии. Ансамбль обеспечивает стабильность, делая модель робастной к типу входных данных.

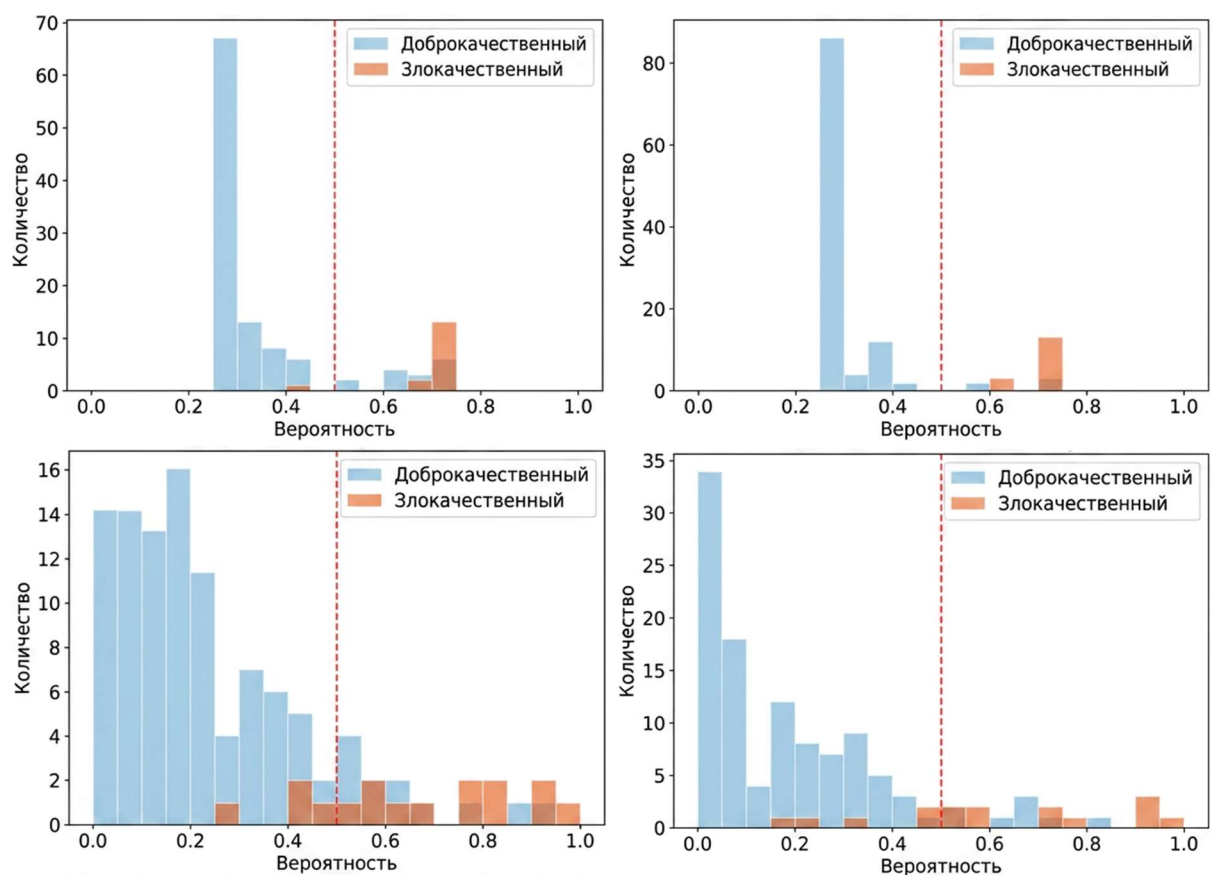


Рисунок 115 – Распределения вероятностей моделей CNN и Transformer – компонентов ансамбля при разных модальностях (дерматоскопические и клинические фотоизображения)

На рисунке на каждой диаграмме синие столбцы показывают распределения для доброкачественных образований, оранжевые – для злокачественных новообразований. Верхний ряд показывает распределения для CNN-компонента, нижний – для трансформера. Для CNN при клиническом фото видно четкое разделение распределений, тогда как при дерматоскопии наблюдается большее перекрытие. Для трансформера – картина обратная, лучшее разделение при дерматоскопии. Красная пунктирная линия (порог 0,5) позволяет визуально оценить, какие наблюдения каждый компонент классифицировал бы верно или ошибочно при работе моделей отдельно.

Проведенный многоаспектный сравнительный анализ позволяет сформулировать следующие основные выводы.

Во-первых, модель демонстрирует высокую диагностическую эффективность при обоих типах фотофиксации ($AUC > 0,98$; Brier score $< 0,11$; $F1 > 0,73$). Статистически значимых различий в общей точности, чувствительности и специфичности между дерматоскопическими и клиническими изображениями не выявлено (все $p > 0,05$, тест Макнемара: $p = 0,267$). Это означает, что модель может быть использована в клинической практике как с дерматоскопом, так и с обычной камерой смартфона, без существенной потери качества диагностики.

Во-вторых, каждый тип визуализации обладает своими преимуществами, формирующими его оптимальную нишу. Дерматоскопия обеспечивает более высокую чувствительность (0,938 против 0,875), лучшее отношение правдоподобия отрицательного результата ($LR- = 0,069$) и более высокую среднюю точность при дисбалансе классов ($AP = 0,917$). Клиническое фото обеспечивает более высокую специфичность (0,963 против 0,908), существенно лучшее отношение правдоподобия положительного результата ($LR+ = 23,65$ против 10,19), лучшую калибровку (Brier = 0,081) и более высокий коэффициент согласия ($\kappa = 0,796$).

В-третьих, статистически значимое различие обнаружено лишь в распределении вероятностей для доброкачественных новообразований ($p < 0,001$, $d = 0,377$), дерматоскопические изображения генерируют в среднем более высокие

ложные вероятности злокачественности (0,291 против 0,245). Данный эффект обусловлен тем, что дерматоскопическая визуализация выявляет структурные паттерны, которые, присутствуя в доброкачественных образованиях, невусах, кератоцитарных/эпидермальных образованиях – кератомах; гемангиомах и дерматофибромах, могут частично мимикрировать под признаки злокачественности.

В-четвертых, попарный анализ предсказаний показал, что ошибки модели на двух типах визуализации частично не совпадают: в 10,4% случаев верный результат был получен только на одной из модальностей. Корреляция вероятностей между модальностями носит умеренный характер ($r = 0,632$; $ICC = 0,617$), а ансамблирование CNN и Transformer компонентов компенсирует их разнонаправленную чувствительность к типу входных данных.

Таким образом, результаты свидетельствуют о робастности разработанной модели к типу устройства фотофиксации. Оба метода визуализации пригодны для клинического применения, при этом выбор между ними может определяться конкретным сценарием использования: дерматоскопия предпочтительна для скрининга (приоритет чувствительности и минимизация пропуска злокачественных образований), клиническая фотография, для дифференциальной диагностики в условиях ограниченных ресурсов (приоритет специфичности и доступности оборудования).

3.17. Анализ характеристик программного модуля непрерывного профессионального развития «Derma Check Educational Module»

3.17.1. Общая характеристика и архитектура модуля

Модуль непрерывного профессионального развития (НПР) «Derma Check Educational Module» [87] образовательная подсистема, встроенная непосредственно в рабочую среду диагностического приложения «Derma Check». Модуль реализован как одностраничное веб-приложение (edu.html) на чистом

JavaScript с единой дизайн-системой «Derma Check». Серверная часть FastAPI с SQLite-хранилищами (quiz.db, tasks.db). Навигация реализована через History API и систему вкладок; страница /edu является единой точкой входа, хаб отображается по умолчанию, переключение на тест или задачи производится по кнопкам в хабе или прямым URL-параметрам (tab=quiz, tab=tasks).

Встроенность обучения в ту же среду, где врач выполняет ИИ-анализ фотоизображений, устраняет когнитивный и временной барьер между диагностической работой и повышением квалификации. Врач, столкнувшийся с расхождением собственного мнения и результата классификации, может в один клик открыть описание альтернативной нозологии непосредственно из результата классификации, а затем перейти к тесту или клинической задаче по данной группе заболеваний. Такой цикл соответствует принципам рефлексивной практики.

Модуль НПП включает четыре взаимосвязанных компонента (Таблица 70). Хаб посадочная страница модуля со статистикой банков тестов и задач и карточками выбора активности (Рисунок 116).

Таблица 70 – Компоненты модуля непрерывного профессионального развития «Derma Check Educational Module»

Компонент	Назначение	Источник данных
Хаб	Посадочная страница модуля: банк тестов и задач, выбор активности	API: /api/quiz/meta, /api/tasks/list
Тест	Интерактивный тест по нозологиям: изображение и выбор из 4 вариантов	SQLite: quiz.db (вопросы, изображения)
Задачи	Клинические задачи: анамнез и фотоизображение, свободный ответ и эталон	SQLite: tasks.db (задачи, эталоны)
Атлас	Справочник нозологий с полнотекстовым поиском	API: /api/nozology-list, /api/nozology/{id}

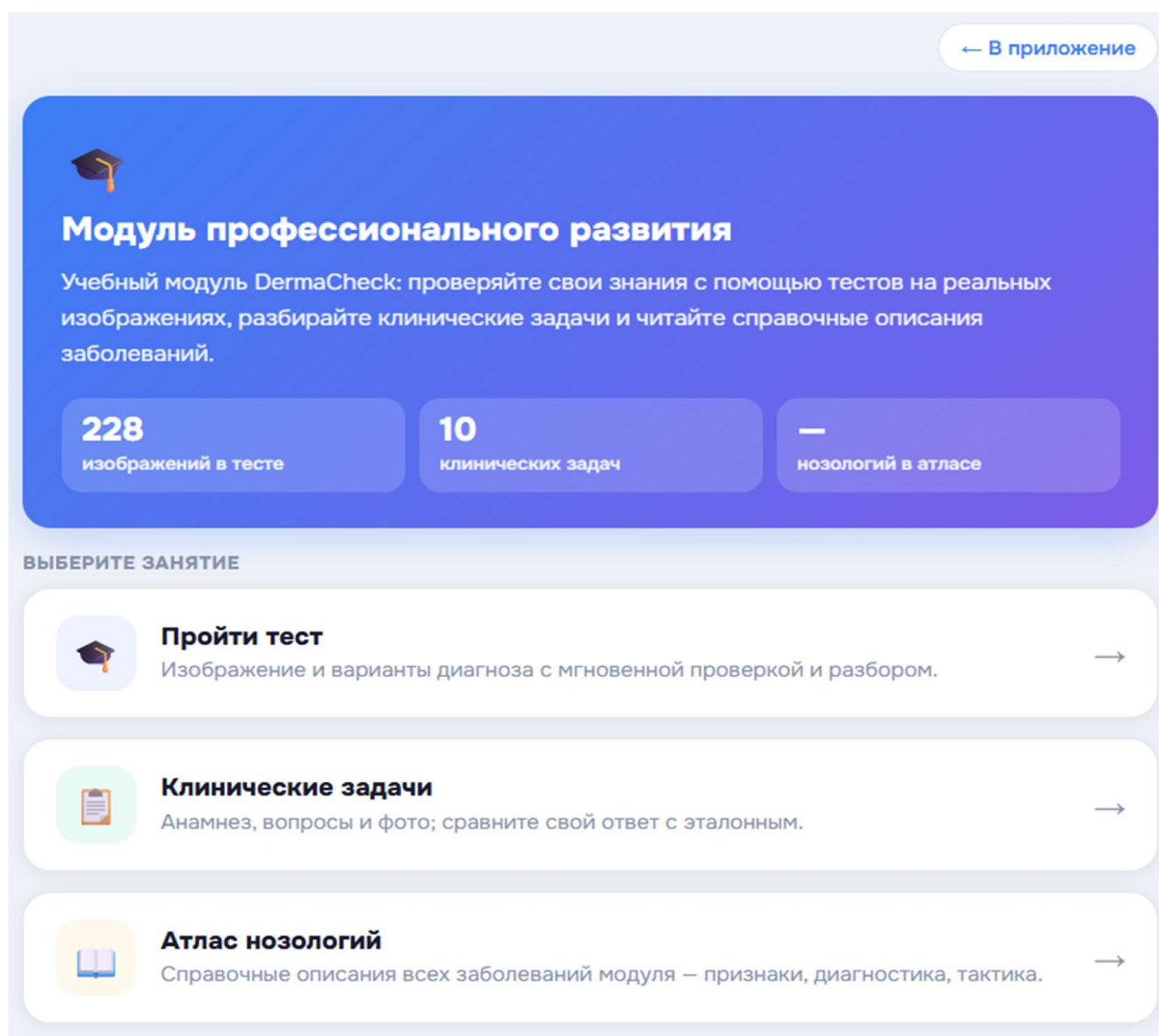


Рисунок 116 – Экран хаба модуля непрерывного профессионального развития: общая статистика, карточки выбора активности (тест, задачи, атлас)

3.17.2. Компонент «Тест по нозологиям»

Тест реализует формат визуальной идентификации: врачу предъявляется фотоизображение кожного новообразования и предлагается выбрать верный диагноз из четырех вариантов. Перед началом теста врач настраивает количество вопросов (5, 10, 15 или 20) и фильтр по злокачественности (Рисунок 117). Набор вопросов формируется случайным образом из общего банка с учетом выбранных фильтров.

Рисунок 117 – Экран настройки теста: выбор количества вопросов и фильтра фотоизображений по злокачественности новообразований

Варианты ответа формируются на сервере с соблюдением нескольких принципов. Правильный ответ всегда присутствует среди четырех вариантов. Дистракторы подбираются из числа нозологий, представленных в банке. При подборе исключаются нозологии одной линии: например, «Меланома» и «Меланома узловая» не могут одновременно присутствовать среди вариантов, такая пара превращает вопрос из клинического в семантический. Алгоритм подбора дистракторов работает в два прохода: на первом проходе отбираются нозологии из предварительно сгенерированного списка с проверкой на принадлежность к одной линии, на втором, из общего пула банка. Порядок финальных четырех вариантов рандомизируется (Рисунок 118).

Тест

Клинические задачи

Вопрос 1 из 10

0 верно



Что изображено на фотографии?

Киста эпидермальная

Акне

Стеатоцистомы множественные

Розацеа

Далее →

Рисунок 118 – Экран вопроса теста: фотоизображение, четыре варианта ответа, прогресс-бар и счетчик верных ответов

При выборе варианта правильный вариант подсвечивается зеленым цветом, неправильный, красным (Рисунок 119). При ошибке отображается кнопка «Почитать про правильный вариант», открывающая модальное окно с полным описанием нозологии из атласа. По завершении теста отображается итоговый счет

с качественной оценкой: «Отличный результат!» (80% и более), «Неплохо, есть куда расти» (50–79%) или «Стоит повторить материал» (менее 50%).

Тест

Клинические задачи

Вопрос 1 из 101 верно



Что изображено на фотографии?

Киста эпидермальная ✓

Акне

Стеатоцистомы множественные

Розацеа

Верно!

Далее →

Рисунок 119 – Экран вопроса теста, обратная связь после ответа: подсветка правильного и неправильного вариантов, ссылка на описание нозологии

3.17.3. Компонент «Клинические задачи»

Клинические задачи моделируют ситуацию осмотра пациента: врачу предъявляется фотоизображение образования, развернутый анамнез (пол, возраст, локализация, динамика, жалобы) и список вопросов (обычно 2–4). Врач формулирует ответ в свободной форме в текстовом поле, после чего раскрывает эталонный ответ для самопроверки (Рисунок 120).

Задача 1 из 10



АНАМНЕЗ

Пациент, 35 лет, обратился к врачу из-за появившихся на подошвах ног высыпаний. Высыпания появились в течение последних двух недель и усилились за последние две недели. Больной отмечает зуд, который усиливается вечером. В анамнезе: ранее не болел такими заболеваниями. Объективный статус: на подошвах ног обнаружены мелкие белые точечные высыпания с небольшим кератозом, которые распространяются по всей поверхности подошв.

ВОПРОСЫ

1. Поставьте и обоснуйте диагноз.
2. Составьте план обследования.
3. Проведите дифференциальную диагностику.

ВАШ ОТВЕТ (САМОПРОВЕРКА — НЕ ОЦЕНИВАЕТСЯ)

Поставьте и обоснуйте диагноз, предложите тактику...

Показать эталонный ответ

← Назад Вперед →

Рисунок 120 – Экран клинической задачи: фотоизображение новообразования, анамнез, вопросы и поле для свободного ответа

Эталонный ответ загружается по отдельному API-запросу только после нажатия соответствующей кнопки, что исключает случайное раскрытие при загрузке задачи.

Отсутствие автоматической оценки свободного ответа является методическим решением. Задачи с развернутым ответом нацелены на формирование клинического мышления, а не на проверку запоминания терминов. В контексте самостоятельного непрерывного профессионального развития самопроверка по эталону является достаточным и более прозрачным механизмом. Эталон содержит диагноз и его клиническое обоснование; при наличии нозологии в справочнике отображается ссылка «Почитать про диагноз» (Рисунок 121).

Эталон показан

Кератоз мелкоточечный

1. **Диагноз:** Кератоз мелкоточечный.

Обоснование диагноза:

Пациенту 35 лет, жалуется на появление мелких белых точечных высыпаний с небольшим кератозом на подошвах ног. Высыпания появились в течение последних двух недель и усилились за последние две недели. Больной отмечает зуд, который усиливается вечером. Локализация высыпаний соответствует клинической картине кератоза мелкоточечного: мелкие белые точечные элементы с небольшим кератозом на подошвах ног, распространяющиеся по всей поверхности. Течение заболевания характерно для кератоза мелкоточечного.

2. **План обследования:**

- Осмотр и сбор анамнеза.
- Дерматоскопия.
- Лабораторные исследования: общие анализы крови, биохимический анализ крови, исследование на грибковые инфекции (см. мазок из высыпаний).
- Биопсия/гистология по показаниям.

3. **Дифференциальная диагностика:**

- Себорейный кератоз.
- Кератопапиллома.
- Грибковые инфекции (кандидоз, микоз).
- Эпидермолитические кератодермии.

☐ Почитать про диагноз

Рисунок 121 – Эталонный ответ к клинической задаче: диагноз, клиническое обоснование, ссылка на описание нозологии в атласе

Задачи помечены источником: «из атласа» или «сгенерирована». Навигация между задачами осуществляется кнопками «Назад» и «Вперед», текст ответа и статус раскрытия эталона кэшируются на клиенте.

3.17.4. Компонент «Атлас нозологий»

Атлас нозологий, справочная подсистема с полнотекстовым поиском (Рисунок 122).

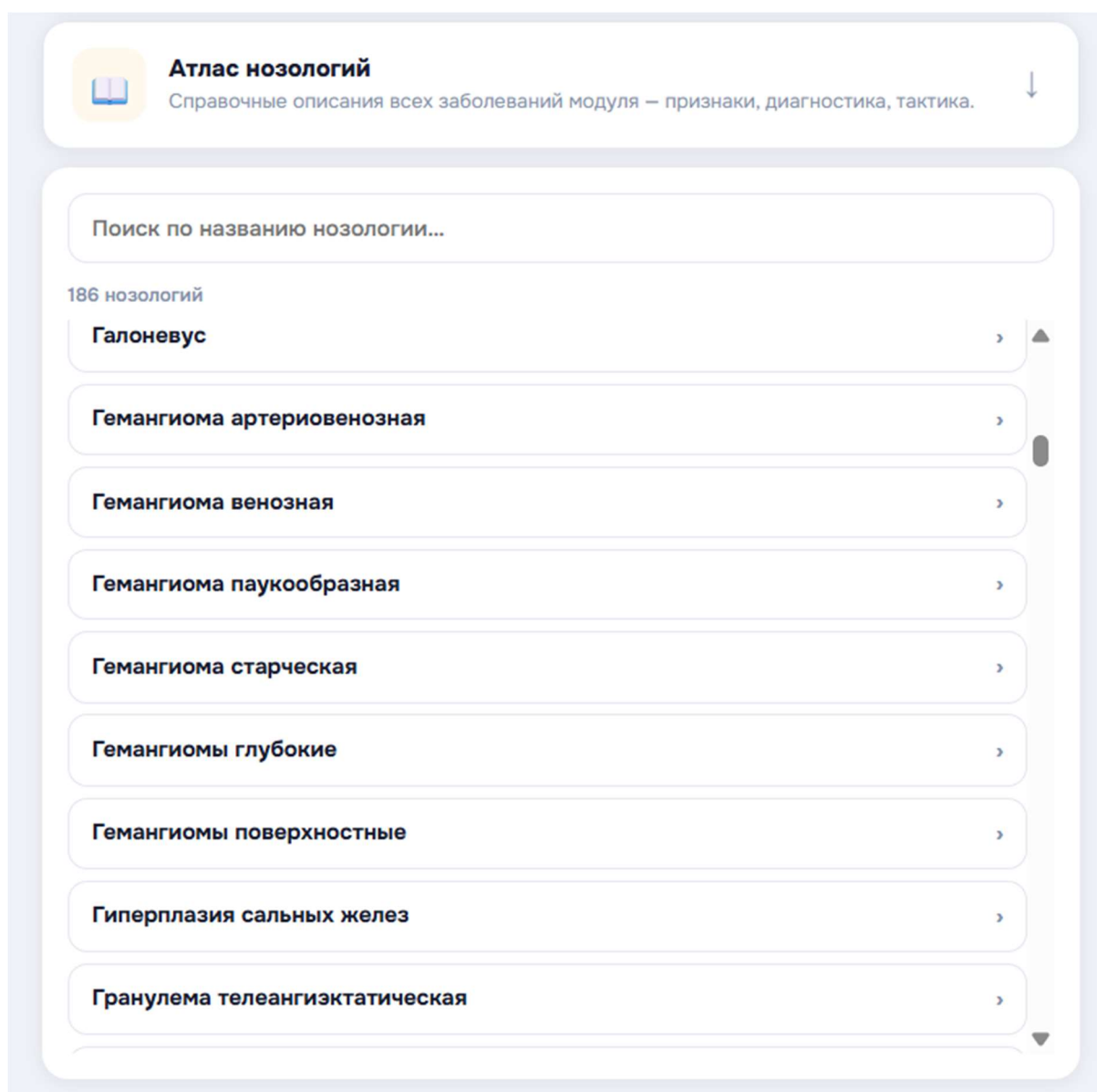


Рисунок 122 – Экран атласа нозологий: список заболеваний с поиском, бейджи злокачественности

Список нозологий загружается по API-запросу, каждая нозология содержит название, статус (злокачественное или доброкачественное) и идентификатор для загрузки полного описания. Поиск реализован на клиенте: нормализация строки, фильтрация по вхождению подстроки в название. Описание каждой нозологии отображается в модальном окне и включает разделы: определение, клинические признаки, дифференциальная диагностика, дерматоскопические признаки, тактика (Рисунок 123).

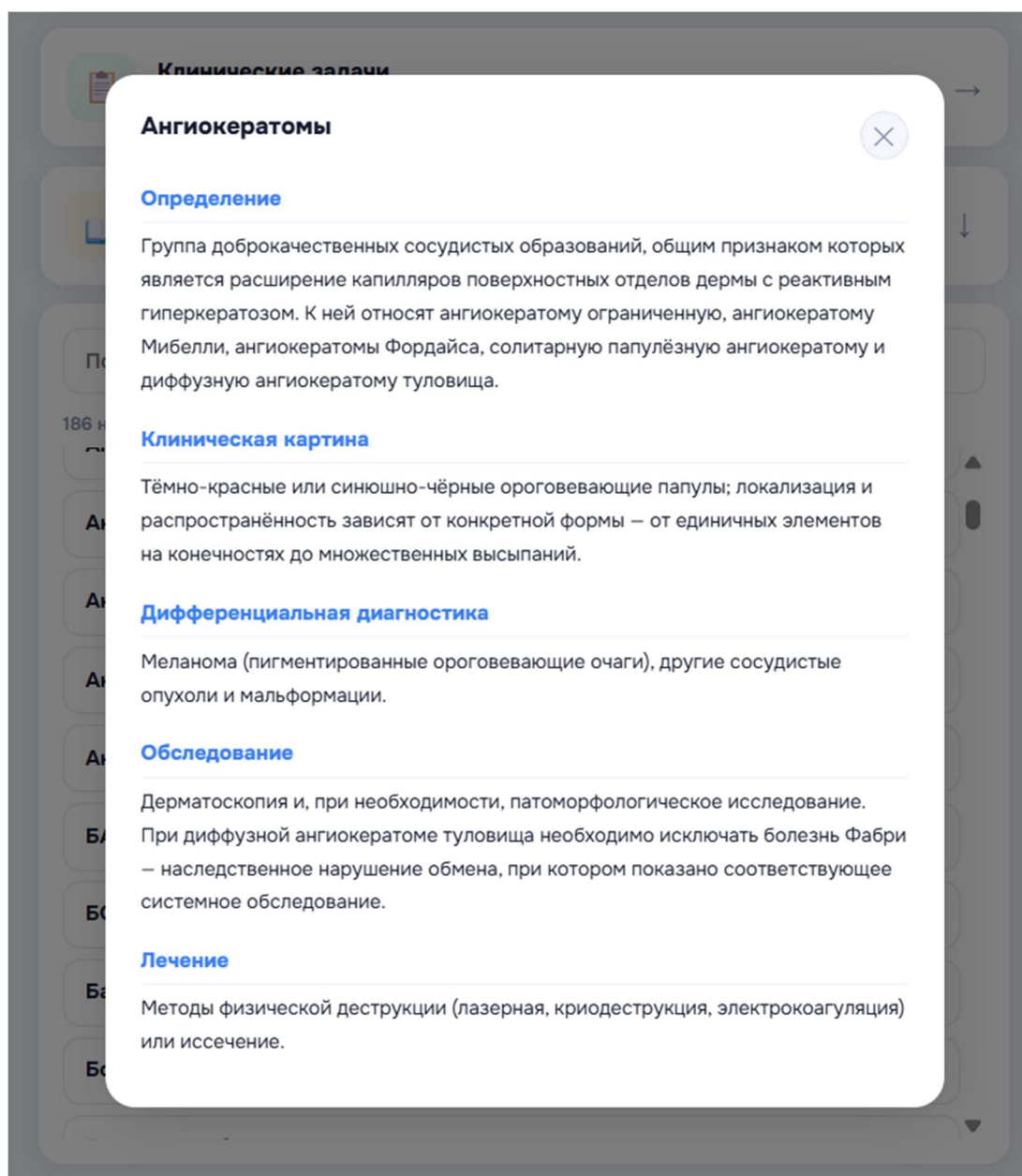


Рисунок 123 – Модальное окно описания нозологии: разделы «Определение», «Клиническая картина», «Дифференциальная диагностика», «Обследование», «Лечение»

Атлас использует единый справочник нозологий, общий для всех компонентов модуля НПР и основного диагностического ядра. Добавление новой нозологии в справочник автоматически делает ее доступной во всех компонентах.

3.17.5. Интеграция модуля непрерывного профессионального развития с диагностическим ядром

Интеграция модуля НПР с основным диагностическим приложением реализована на нескольких уровнях. На уровне навигации: на главном экране приложения размещен блок «Профессиональное развитие» с четырьмя карточками, ведущими в соответствующие компоненты модуля НПР (Рисунок 124).

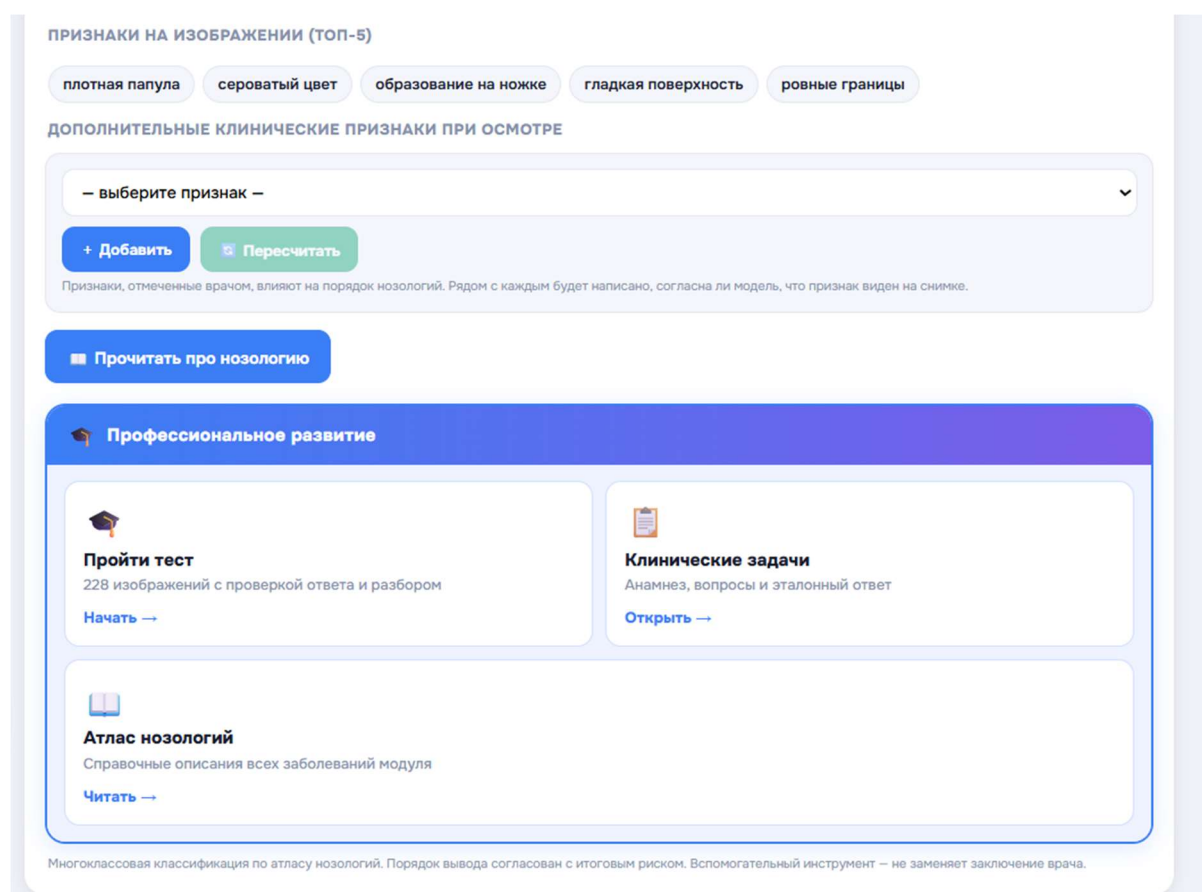


Рисунок 124 – Экран блока «Профессиональное развитие» на главном экране «Derma Check»: карточки перехода к обучающим компонентам

На уровне клинического контекста: при расхождении заключения модели ИИ и мнения врача система предлагает перейти к описанию конкретной нозологии.

Описания нозологий доступны также из блока многоклассовой классификации: при нажатии на название нозологии в топ-5 открывается модальное окно с полным описанием. На уровне данных: тест, задачи и атлас используют единый справочник нозологий, что обеспечивает консистентность терминологии и взаимную согласованность образовательного и диагностического компонентов.

3.17.6. Методические принципы модуля непрерывного профессионального развития

Модуль НПР построен на четырех методических принципах. Принцип немедленной обратной связи: результат ответа на тестовый вопрос отображается мгновенно, без ожидания завершения теста, что соответствует данным о роли немедленного подкрепления в медицинском образовании. Принцип привязки к контексту: каждый тестовый вопрос и клиническая задача содержат ссылку на описание правильной нозологии, переход от ошибки к изучению материала осуществляется в одно нажатие.

Принцип разделения проверочного и учебного форматов: тест (закрытые вопросы) проверяет узнавание, задачи (свободный ответ по эталону) формируют клиническое мышление. Суммарно четыре компонента охватывают все уровни когнитивной деятельности в визуальной диагностике, от распознавания (тест) через анализ (задачи) к систематизации (атлас).

3.18. Исследование эффективности обучения диагностике новообразований кожи и дерматозов с применением модуля «Derma check educational module»

Проведено проспективное контролируемое исследование с участием 50 обучающихся в ординатуре по специальности «Дерматовенерология», рандомизированных методом конвертов в две группы по 25 человек. Группы были сопоставимы по полу (основная группа 17 женщин и 8 мужчин, контрольная 16 женщин и 9 мужчин, $\chi^2=0,09$, $p=0,764$), возрасту (основная $25,8\pm 2,1$ года,

контрольная $26,3 \pm 2,4$ года, $U=287$, $p=0,621$) и году обучения в ординатуре. Основная группа обучалась с применением модуля Derma Check Educational Module, контрольная группа изучала материал по учебным пособиям и атласам.

Учебный материал охватывал 20 нозологических форм новообразований кожи, распределенных по четырем укрупненным группам. Злокачественные новообразования ($n=5$): поверхностно распространяющаяся меланома, узловая меланома, меланома типа злокачественного лентиго, базальноклеточный рак, плоскоклеточный рак кожи. Меланоцитарные доброкачественные новообразования ($n=5$): пограничный меланоцитарный невус, сложный меланоцитарный невус, внутридермальный меланоцитарный невус, невус Шпитц, голубой невус. Кератиноцитарные/эпидермальные доброкачественные новообразования ($n=6$): себорейный кератоз, актинический кератоз, кератоакантома, бовеноидный папулез, вирусная бородавка, кожный рог. Сосудистые и прочие новообразования ($n=4$): капиллярная гемангиома, пиогенная гранулема, дерматофиброма, липома. Оценка проводилась методом тестирования в четырех временных точках: до обучения (T-1), сразу после обучения (T0), через 7 дней (T7) и через 30 дней (T30). Каждый тест содержал 40 вопросов, максимальный балл 100.

Результаты тестирования представлены в Таблице 71.

Таблица 71 – Результаты тестирования обучающихся в основной и контрольной группах, Me [Q1; Q3]

Показатель	Основная группа (n=25)	Контрольная группа (n=25)	U	p	r
Балл T-1 (до обучения)	39,0 [33,0; 45,0]	38,0 [32,0; 44,0]	308	0,876	0,02
Балл T0 (сразу после)	73,0 [67,0; 79,0]	68,0 [61,0; 74,0]	204	0,038	0,29
Балл T7 (через 7 дней)	62,0 [56,0; 68,0]	52,0 [45,0; 59,0]	147	0,004	0,41
Балл T30 (через 30 дней)	51,0 [45,0; 57,0]	41,0 [34,0; 48,0]	138	0,002	0,44
Снижение T0→T7, баллы	11,0 [7,0; 15,0]	16,0 [11,0; 21,0]	163	0,012	0,36
Снижение T0→T30, баллы	22,0 [17,0; 27,0]	27,0 [21,0; 33,0]	172	0,024	0,32
% сохранения T30	69,9 [63,4; 75,0]	60,3 [52,1; 67,6]	158	0,008	0,37
Время, мин/нозология	4,8 [3,9; 5,6]	6,4 [5,2; 7,8]	152	0,004	0,41
Примечание: T-1 – до обучения; T0 – сразу после обучения; T7 – через 7 дней; T30 – через 30 дней; r – размер эффекта ($r=Z/\sqrt{N}$)					

Двухфакторный ANOVA с повторными измерениями подтвердил статистически значимое влияние как фактора группы ($F=8,7$, $p=0,005$), так и фактора времени ($F=84,3$, $p<0,001$), а также их взаимодействия ($F=4,9$, $p=0,029$). Анализ динамики забывания (Рисунок 125) демонстрирует характерное различие в форме кривых. В контрольной группе наблюдалась классическая кривая забывания Эббингауза с выраженным снижением в первую неделю (потеря 23,5% от уровня T0) и суммарной потерей 39,7% от T0 к 30 дню. В основной группе снижение в первые 7 дней составило 15,1% от T0, а за период с 7 по 30 день еще 15,1%, что свидетельствует о формировании более устойчивых следов памяти при использовании интерактивного формата с немедленной обратной связью.

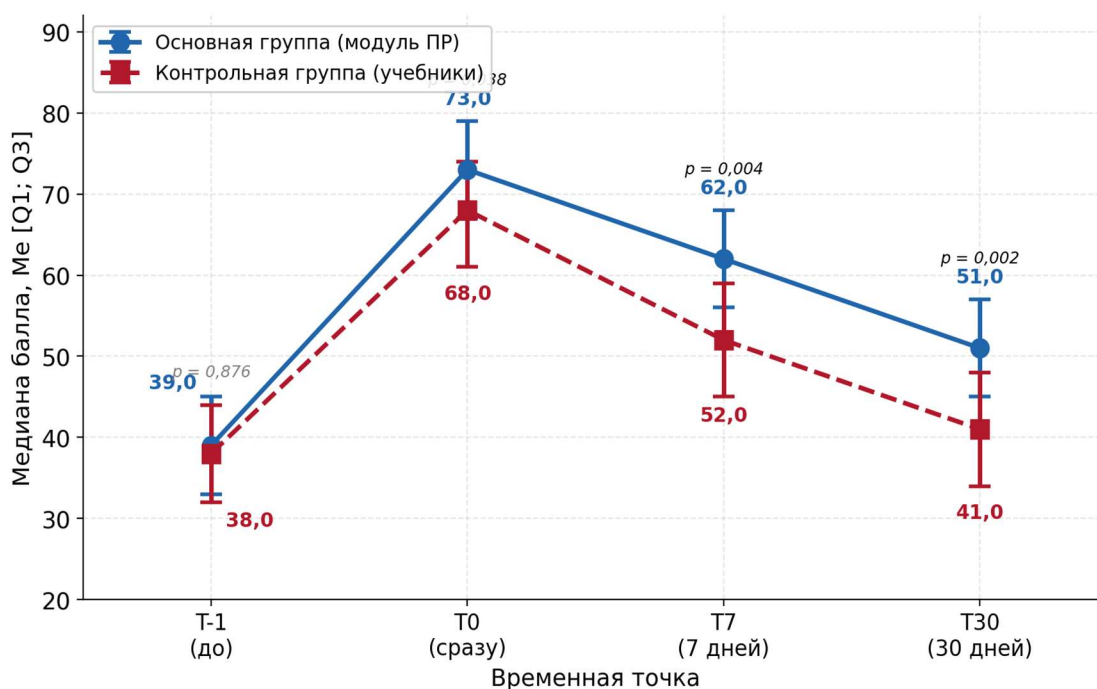


Рисунок 125 – Динамика результатов тестирования в основной и контрольной группах по временным точкам (T0, T7, T30). Сплошная линия с кругами – основная группа (модуль ПР), пунктирная линия с квадратами – контрольная группа (учебники). Планки погрешностей – стандартное отклонение. Заштрихованная область – эффект «плато» в основной группе

Доля сохранения знаний через 30 дней по укрупненным группам нозологий представлена в Таблице 72 и на Рисунке 126.

Таблица 72 – Доля (в %) сохранения знаний через 30 дней по укрупненным группам нозологий, Ме [Q1; Q3]

Группа нозологий	Основная, %	Контрольная, %	p
Злокачественные новообразования	70,1 [63,8; 76,4]	61,4 [52,7; 70,2]	0,014
Меланоцитарные доброкачественные	68,3 [61,0; 74,9]	58,7 [48,5; 68,1]	0,021
Кератиноцитарные/эпидермальные доброкачественные	69,5 [62,4; 75,8]	60,2 [51,8; 68,7]	0,012
Сосудистые и прочие	66,8 [58,9; 73,5]	56,9 [46,3; 66,0]	0,028

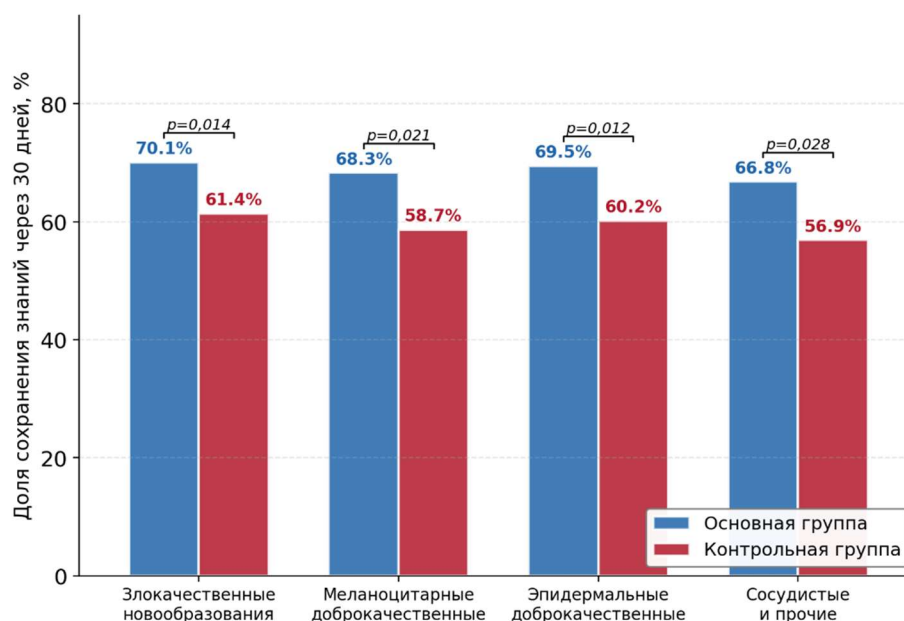


Рисунок 126 – Доля (в %) сохранения знаний через 30 дней по укрупненным группам нозологий в основной и контрольной группах. Скобки со звездочками – уровень значимости различий

Формирование эффекта «плато» в кривой забывания основной группы может быть объяснено несколькими механизмами. Формат немедленного тестирования после изучения материала реализует принцип активного воспроизведения (testing effect), при котором сам акт извлечения информации из памяти укрепляет след памяти значительно эффективнее чем повторное чтение. Контекстная привязка визуальной информации к конкретному клиническому изображению формирует ассоциативные связи способствующие кодированию материала в долговременной памяти. Немедленная обратная связь при неправильном ответе обеспечивает коррекцию ошибок в момент их возникновения, что предотвращает консолидацию неверных ассоциаций.

ГЛАВА 4. ОБСУЖДЕНИЕ

Настоящее диссертационное исследование объединило несколько методологически самостоятельных, но логически взаимосвязанных научных направлений: изучение отношения врачей к технологиям искусственного интеллекта (ИИ); разработку методологии формирования наборов данных и классификации программ для ЭВМ; систематизацию проблем автоматизированной диагностики новообразований кожи и анализ причин ошибочной классификации; разработку показателей клинической точности и эффективности диагностики с учетом онкологической настороженности; оценку медико-экономической эффективности алгоритмов маршрутизации пациентов на основании заключений моделей ИИ и создание комплексной системы поддержки принятия врачебных решений (СППВР) «Derma Check» с модулем непрерывного профессионального развития «Derma Check Educational Module».

Полученные результаты требуют всестороннего обсуждения в контексте современных отечественных и зарубежных данных, что и составляет содержание настоящей главы. Обсуждение построено в соответствии со структурой глав собственных результатов, что позволяет последовательно рассмотреть каждый из решенных научных вопросов.

Обсуждение результатов изучения отношения врачей к технологиям искусственного интеллекта

Успешность внедрения технологий искусственного интеллекта в клиническую практику определяется не только их технической эффективностью, но и готовностью врачей их применять, уровнем осведомленности и доверия к алгоритмическим заключениям [140, 220, 344]. Проведенное нами масштабное опросное исследование 545 специалистов трех ключевых для выявления злокачественных новообразований (ЗНО) кожи специальностей: терапевтов и врачей общей практики, дерматовенерологов и онкологов, восполнило

существенный пробел в отечественной литературе, где исследования отношения профессионального сообщества к ИИ до настоящего времени были представлены крайне ограниченно [290].

Ключевым результатом нашего исследования следует считать не столько фиксацию умеренного дефицита доверия к технологиям ИИ (64,8–68,7% специалистов оценили его как незначительное или полностью отсутствующее), сколько выявление структуры и предикторов этого отношения. Обнаруженный во всех трех выборках диссонанс между высокой декларируемой готовностью к применению ИИ (87,6–89,2%) и умеренным уровнем доверия отражает раннюю стадию знакомства профессионального сообщества с технологией. Этот феномен «не доверяю, но готов попробовать» согласуется с данными международных исследований, демонстрирующих в целом положительное, но осторожное отношение специалистов визуальных специальностей к ИИ при сохранении опасений относительно надежности алгоритмов и регуляторной неопределенности [157, 159, 298]. Крупный международный онлайн-опрос дерматологов также показал, что большинство специалистов заинтересованы в применении ИИ, но подчеркивают необходимость валидации и обеспокоены вопросами ответственности и влияния технологий на взаимоотношения врача и пациента [157].

Наиболее значимым в практическом отношении результатом является установление того, что единственным устойчивым предиктором позитивного отношения к ИИ во всех трех специальностях выступает реальный практический опыт взаимодействия с ИИ-технологиями (у терапевтов $U=1394$; $p=0,014$; у онкологов $U=4778$; $p<0,001$), тогда как возраст, стаж, тип медицинской организации и уровень владения ПК значимого влияния на доверие не оказывают. Этот вывод полностью согласуется с систематическими обзорами факторов, определяющих доверие медицинских работников к ИИ, в которых в качестве ключевых выделяются прозрачность алгоритмов, наличие доказательной базы, интеграция в рабочий процесс и, что особенно важно, непосредственный опыт применения [119, 168]. Показано, что информированность о существовании

технологий сама по себе не гарантирует доверия, а заниженные представления о точности ИИ, формируемые под влиянием медиаобраза технологий и отдельных случаев неудачного применения, воспроизводят скептицизм [77]. Именно поэтому для преодоления барьера недоверия целесообразны не информационные кампании, а пилотные проекты, обеспечивающие контролируемый контакт врача с ИИ-системой в рабочей среде, регуляторная определенность и доступ к реестрам верифицированных ИИ-систем [211, 217, 302].

Особого внимания заслуживает выявленный нами «парадокс цифровых нативов» в выборке дерматовенерологов: при 17-летнем разрыве в медианном возрасте между ординаторами и практикующими врачами достоверных различий в профессиональном доверии к ИИ не обнаружено ($p=0,982$), несмотря на достоверно более высокую бытовую цифровую активность ординаторов. Этот результат эмпирически опровергает распространенное представление о «возрастном барьере» в отношении к ИИ и указывает на то, что профессиональное доверие формируется не бытовой привычкой к генеративным технологиям, а наличием конкретных валидированных клинических инструментов в соответствующей области. Данное наблюдение особенно значимо для дерматовенерологии — специальности с преимущественно визуальным характером диагностики, где системы компьютерного зрения имеют наиболее очевидную точку приложения [143, 166], однако именно они остаются наименее используемыми (8,9% против 17,1% у терапевтов). Выявленный разрыв между доказанной эффективностью систем компьютерного зрения в дерматоскопии [118, 121, 324] и их реальной доступностью в рутинной клинической практике отражает дефицит специализированных валидированных ИИ-продуктов для дерматовенерологии.

Субъективная оценка точности диагностики системами ИИ (медиана 45–55%) оказалась существенно заниженной по сравнению с реальными клиническими показателями современных валидированных систем, достигающих точности 80–95% на задачах, для которых они разработаны [196, 197, 323, 324]. Разрыв восприятия в 22–30 процентных пунктов формирует и воспроизводит скептицизм: врач, полагающий, что ИИ ошибается в каждом втором случае, не

станет полагаться на его заключения. Специфика выборки, практически равная обеспокоенность ошибками диагностики и ошибками лечения, а также оценка эффективности лечения выше диагностической точности, отражает мультидисциплинарный характер онкологической помощи и согласуется с зарубежными данными о том, что основные опасения онкологов касаются вопросов ответственности, валидации и интеграции ИИ в клинические процессы, а не замещения врача [129, 148, 172, 209]. Таким образом, полученные результаты обосновывают приоритет практико-ориентированных образовательных мероприятий и создания открытых реестров верифицированных ИИ-систем перед сугубо информационными подходами.

Обсуждение анализа программ для электронных вычислительных машин на основе моделей искусственного интеллекта для диагностики заболеваний кожи

Развитие цифровых технологий и повсеместное распространение карманных персональных компьютеров, включая смартфоны и планшетные компьютеры, существенно расширили возможности применения алгоритмов компьютерного зрения в дерматовенерологии [329]. Выполненный нами сравнительный анализ девяти программ на основе ИИ для диагностики патологий кожи выявил значительные различия в их диагностической эффективности, обусловленные особенностями архитектуры, целевым назначением и качеством обучающих данных. Полученные результаты согласуются с данными систематических обзоров мобильных приложений для оценки риска рака кожи, показавших, что при экспоненциальном росте рынка таких приложений качество и безопасность большинства продуктов остаются недостаточными: значительная часть не имеет рецензируемых публикаций о клинической эффективности и характеризуется непрозрачностью обучающих данных и неопределенным регуляторным статусом [335, 364]. Диагностическая точность приложений для самостоятельной оценки существенно варьирует, что ограничивает их применение в качестве

самостоятельного диагностического инструмента и позиционирует преимущественно как средства предварительного скрининга с обязательным последующим осмотром врача [131, 142, 150].

Наиболее высокие и сбалансированные показатели диагностической эффективности продемонстрировали отдельные программы (в частности, отечественная «Derma Onko Check» с точностью 96%, чувствительностью 98% и специфичностью 96%), что сопоставимо с диагностической точностью опытных врачей-дерматовенерологов [103, 109]. Этот результат согласуется с основополагающими работами, продемонстрировавшими способность сверточных нейронных сетей достигать уровня классификации новообразований кожи, сопоставимого с уровнем сертифицированных дерматологов, а в ряде случаев превосходить его [174, 176, 183, 273], а также с систематическими обзорами и метаанализами, подтверждающими конкурентоспособность алгоритмов [118, 121, 324, 325]. Вместе с тем выявленные общие ограничения анализируемых программ: снижение точности при анализе изображений пациентов с темными (IV, V и VI) фототипами кожи, зависимость от качества входного фотоизображения, риск ложноположительных результатов и отсутствие интеграции с клиническими данными полностью соответствуют описанным в литературе системным проблемам медицинского ИИ [29, 69, 123, 249, 260].

Сравнительный анализ реестров медицинских изделий с ИИ США, Китая и Российской Федерации показал, что отечественный реестр существенно уступает зарубежным по числу зарегистрированных изделий, при этом отличается высокой долей отечественных разработчиков и наличием систем поддержки принятия врачебных решений по клиническим данным, не имеющих прямого аналога в реестрах FDA и NMPA. Различия в регуляторном статусе программ ИИ между странами создают «серую зону» в обращении медицинских приложений [139, 285]. Установленный нами дефицит зарегистрированных в Российской Федерации медицинских изделий на основе технологий ИИ для диагностики новообразований кожи (единственное зарегистрированное изделие с бинарным выходным заключением) обозначает приоритетное направление для отечественных

разработчиков и регуляторов и подтверждает актуальность разработки комплексных мультимодельных систем.

Обсуждение анализа баз данных и методологии формирования наборов данных для обучения моделей искусственного интеллекта для использования в клинической медицине

Качество и репрезентативность обучающих наборов данных являются определяющими факторами точности и обобщаемости моделей ИИ в дерматоонкологии [170, 229, 263]. Систематические обзоры публично доступных наборов данных изображений новообразований кожи демонстрируют их недостаточную прозрачность, неполное документирование источников и характеристик пациентов, а также выраженные систематические смещения, прежде всего в отношении фототипов кожи [170, 260, 349]. Установлено, что алгоритмы, обученные на нерепрезентативных данных, могут воспроизводить и усиливать существующие диспаритеты в оказании медицинской помощи [213, 358], а географическая концентрация обучающих данных ограничивает их обобщаемость на другие популяции [257]. В этом контексте разработанная нами шестиэтапная методология формирования наборов данных из клинических баз данных приобретает существенное методологическое значение.

Валидация предложенной методики продемонстрировала статистически значимое превосходство модели, обученной на наборе данных с гистологическим «золотым стандартом» верификации, над моделью, обученной на публичных данных со смешанной верификацией, по всем ключевым диагностическим метрикам (чувствительность – 96,7% против 89,2%; специфичность – 96,0% против 87,5%; AUC-ROC – 0,985 против 0,931; коэффициент согласия Коэна – 0,906 против 0,732; тест Макнемара $\chi^2=12,4$; $p<0,001$).

Это превосходство обусловлено несколькими факторами, непосредственно вытекающими из методики. Во-первых, привязка каждого злокачественного наблюдения к морфологическому эталону устраняет шум меток, неизбежный при

экспертной визуальной аннотации, поскольку межэкспертное согласие даже среди опытных специалистов имеет принципиальный «потолок» точности обучаемой модели [228, 231]. Во-вторых, стратифицированное разбиение на уровне пациентов исключает утечку данных, искусственно завышающую оценку точности. Значимость соблюдения принципов прозрачности, документирования и воспроизводимости при формировании наборов данных подчеркивается концепциями «этикетки питательной ценности набора данных» (англ. The Dataset Nutrition Label) и стандартизированного документирования, а также принципами FAIR и минимальными требованиями к документированию клинических моделей ИИ [284, 337, 342].

Предложенное нами формализованное разграничение понятий «база данных» и «набор данных» (датасет) устраняет существующую в отечественной и зарубежной литературе терминологическую неопределенность, которая порождает серьезные методологические ошибки при сравнении систем.

Обсуждение классификации программ для электронных вычислительных машин, применяемых в клинической медицине и здравоохранении

Разработанная оригинальная классификация программ для ЭВМ, включая основанные на технологиях ИИ и программах компьютерного зрения, построенная на девятнадцати признаках, объединенных в пять тематических блоков, отвечает на актуальную потребность в стандартизированном описании систем. Отсутствие единых подходов к описанию программ приводит к тому, что декларируемые показатели диагностической точности оказываются несопоставимыми между собой, а регуляторная оценка существенно затрудняется [110, 154, 269]. Предложенная классификация обеспечивает единый язык описания программ, необходимый для их корректного сравнения, обоснованного выбора для внедрения в клиническую медицину и здравоохранение и регуляторной экспертизы, и может служить основой для методических и регуляторных требований к подобным программам [33, 139, 285].

Обсуждение систематизации проблем дифференциальной диагностики, проблем и решений в диагностике злокачественных новообразований кожи с применением программ искусственного интеллекта

Дифференциальная диагностика доброкачественных и злокачественных новообразований кожи сопряжена с объективными трудностями, обусловленными разнообразием клинических и морфологических проявлений опухолей, их взаимной имитацией и ограничениями применяемых методов [63, 65, 372].

Выполненная нами систематизация основных проблем и их разделение на объективные трудности и субъективные сложности носит содержательный характер, поскольку позволяет отделить факторы, коренящиеся в природе новообразований и методов их исследования, от факторов, связанных с человеческим и организационным звеном. Проблема баланса между гиподиагностикой (пропуском злокачественных опухолей) и гипердиагностикой с последующим избыточным лечением является одной из фундаментальных в дерматоонкологии [124, 125]; стремление минимизировать пропуски меланомы закономерно приводит к росту числа биопсий доброкачественных образований [282]. Именно эти системные проблемы, субъективность визуальной оценки, дефицит и недостаточная квалификация специалистов, недостаточная онкологическая настороженность, могут быть частично компенсированы внедрением программ ИИ в качестве инструмента поддержки принятия врачебных решений [78, 103, 109].

Разработанная нами методология анализа причин ошибочной классификации, объединяющая вычисление фотометрических и текстурных метрик, многоуровневый статистический анализ и методы объяснимого ИИ (Score-CAM, Occlusion Sensitivity), позволила установить конкретные механизмы диагностических ошибок нейронных сетей. Применение методов объяснимого ИИ для систематической интерпретации ошибок рассматривается в современной литературе как обязательный компонент клинической валидации медицинских

изделий с ИИ, обеспечивающий прозрачность алгоритмических решений [167, 185, 222, 237, 317]. Нейронные сети функционируют как «черные ящики», что затрудняет выявление причин ошибочных заключений, а методы объяснимого ИИ повышают доверие врачей к заключениям алгоритмов, хотя их надежность и клиническая ценность остаются предметом дискуссий [187, 224, 233].

Установленные нами закономерности – повышенная яркость и сниженная энтропия фотоизображений в ложноположительных случаях, а также смещение фокуса нейронной сети на фоновые элементы и артефакты в ложноотрицательных случаях – согласуются с данными о том, что фотометрические и текстурные параметры изображения значимо влияют на точность работы моделей ИИ [29, 69], а модели могут опираться на артефакты изображения (хирургические метки, волосы, тени), что приводит к систематическим ошибкам классификации [156, 161, 181]. Многофакторный анализ подтвердил незначимость фактора конкретной программы компьютерного зрения, что свидетельствует о том, что выявленные закономерности характеризуют именно качество входных фотоизображений, а не специфику отдельной нейросетевой архитектуры, и обеспечивает универсальность установленных референсных диапазонов. Разработанные на основании этих закономерностей адресные методы предобработки и дообучения моделей позволили существенно снизить долю ошибочных заключений (диагностическая точность возросла с 84,3 до 92,4% для программы «Melanoma Check» и с 90,9 до 92,4% для программы «Derma Onko Check»), что соответствует данным о том, что адресная нормализация отклоняющихся параметров изображений повышает точность и специфичность классификации [29, 59]. Продемонстрированные на примере программы «Derma Onko Melanoma Check» технические решения шести групп выявленных проблем (автоматический контроль качества изображений, контроль размера образования в кадре, мультимодельная архитектура, объяснимый ИИ, интеграция анамнеза, мультимодальный анализ) образуют логически заверченный ответ на систематические ограничения, возникающие при переносе программ ИИ из контролируемых условий в реальную клиническую практику [258].

Обсуждение методологии оценки клинической точности с учетом онкологической настороженности

Обоснованное внедрение программ ИИ в клиническую практику требует всесторонней оценки их диагностической точности, при этом методология такой оценки должна учитывать клинический контекст применения технологий [36, 128, 295]. Оценка диагностической точности программ ИИ традиционно основывается на стандартной матрице ошибок с расчетом чувствительности, специфичности и точности [155]. Однако разработанная и клинически апробированная нами трехуровневая методология оценки исходит из принципиального противоречия: стандартная матрица ошибок систематически занижает клиническую ценность алгоритмов в дерматоонкологии, расценивая как ложноположительный результат случаи обоснованной онкологической настороженности. Направление гистологически доброкачественных, но клинически подозрительных новообразований (актинического кератоза, диспластических невусов и иных пограничных состояний) на дополнительное обследование является нормой клинической практики [7, 125], однако при формальной оценке трактуется как ошибка.

Данное методологическое положение имеет существенное практическое следствие, недостаточно освещенное в литературе: при использовании формальных меток для машинного обучения модель в ходе дообучения начинает снижать собственную онконастороженность, то есть деградирует по наиболее значимому клиническому параметру. Предложенная нами система четырех категорий клинической эффективности и матрица клинических случаев с двойным эталоном истины (клинико-дерматоскопическая экспертиза и морфологическая верификация) устраняют этот недостаток. Применение методологии на выборке из 342 новообразований кожи показало, что 12,9% наблюдений, квалифицируемых стандартной матрицей как ложноположительные, относятся к клинически ценной категории «совпадение по онкологической настороженности», при этом

клиническая точность составила 92,1% против формальной точности 89,7%. Логика клинического мышления, положенная в основу этих подходов, соответствует принципам доказательной медицины и современным представлениям об оценке долгосрочного реального влияния ИИ-инструментов на клиническую практику, учитывающим не только диагностическую точность, но и влияние на процессы оказания помощи [43, 44, 46, 115, 216].

Надежность экспертного эталона в нашем исследовании подтверждена высокой межэкспертной согласованностью (коэффициент каппа Коэна во всех парах превысил 0,96), что соответствует «почти идеальному согласованию» по шкале Ландиса–Коха и обеспечивает воспроизводимость оценки. Принципиальным результатом клинической апробации является полное выявление всех злокачественных новообразований (клиническая чувствительность 98,5%) при отсутствии пропусков ЗНО, а также более широкая онконастороженность программы по сравнению с опытными врачами, что для инструмента первичного звена является скорее достоинством. Предложенная методология применима в сравнительных исследованиях программ ИИ, при регуляторной экспертизе диагностических систем и при разработке критериев разметки обучающих данных, что отвечает потребности в стандартизированных подходах к оценке медицинского ИИ, отраженной в стандартах отчетности TRIPOD+AI, STARD-AI, CONSORT-AI, SPIRIT-AI и DECIDE-AI [239, 333, 352].

Обсуждение способа оценки эффективности диагностики с использованием алгоритмов маршрутизации пациентов на основании заключений программ искусственного интеллекта

Высокая клиническая точность отдельного диагностического акта не тождественна эффективности диагностической системы в целом, поскольку в реальной практике программа ИИ работает как звено многоэтапной маршрутизации пациента [1, 28, 225]. Разработанный нами способ оценки эффективности диагностики, основанный на оригинальных коэффициентах

прироста выявляемости и сокращения пропусков с визуализацией результата в квадрантной матрице, восполняет отсутствие в литературе инструментов оценки эффективности диагностики в условиях многоэтапной маршрутизации [43, 44]. В отличие от показателей точности, оценивающих модель как таковую, предлагаемый способ учитывает каскадный характер ошибок и реальную чувствительность врачей разных специальностей на каждом этапе.

Полученные результаты (увеличение выявляемости ЗНО в 1,9 раза и снижение числа пропущенных случаев в 10,5 раза) объясняются прежде всего низкой чувствительностью первичного звена: диагностическая точность врачей первичного звена в отношении новообразований кожи существенно уступает точности дерматовенерологов, что подтверждено систематическими обзорами и метаанализами [326]. Именно на этапе врача общей практики традиционный маршрут теряет значительную долю случаев ЗНО, тогда как маршрутизация на основе заключений программы выводит большинство из них напрямую к профильному специалисту. Этот результат согласуется с концепцией применения ИИ как инструмента предварительной диагностики для врачей первичного звена, оптимизирующего маршрутизацию пациентов и ускоряющего выявление злокачественных опухолей [56, 225], а также с представлением о «дополненном интеллекте» (augmented intelligence), при котором ИИ дополняет, а не заменяет врача [160, 246, 345]. Разработанная для автоматизации расчетов программа для ЭВМ делает способ пригодным для применения непосредственно в организационно-управленческой практике при обосновании включения технологий ИИ в клинические рекомендации и региональные методические документы.

**Обсуждение способа оценки медико-экономической эффективности
диагностики с применением алгоритмов маршрутизации пациентов на
основании заключений программ искусственного интеллекта**

Для обоснования широкого внедрения программ ИИ в дерматоонкологию необходима не только оценка их диагностической точности и клинической эффективности, но и анализ медико-экономической эффективности (МЭЭ) диагностических алгоритмов на уровне отдельной медицинской организации и региона [43, 46, 203, 205]. Оценка МЭЭ программ ИИ осуществляется в рамках оценки технологий здравоохранения и требует адаптации классических подходов к особенностям цифровых медицинских продуктов [52, 136, 218, 286, 334]. Систематические обзоры экономических оценок технологий ИИ в здравоохранении показывают, что, несмотря на быстрый рост числа публикаций, доказательная база их экономической эффективности остается ограниченной, а методология исследований неоднородна [46, 203, 204, 340]. Разработанный нами способ оценки МЭЭ с симметричным учетом предотвратимых и потенциально предотвратимых финансовых затрат и квадрантной матрицей визуализации вносит вклад в решение этой методологической проблемы.

Ключевым методологическим решением предложенного способа является симметричное включение компонента затрат на лечение пропущенных случаев в обе части коэффициента соотношения затрат, что обеспечивает корректную сопоставимость сценариев с принципиально разной структурой затрат. Полученные результаты (снижение общих затрат на диагностику меланоцитарных опухолей кожи на 51% на уровне учреждения и потенциальная экономия порядка 436 млн руб. в год на региональном уровне при сокращении сроков диагностики на 10,5 суток) согласуются с данными о том, что раннее выявление новообразований кожи и рациональная маршрутизация пациентов способны существенно снизить затраты за счет уменьшения числа запущенных случаев и необоснованных вмешательств [179, 236, 291]. Экономическое бремя ЗНО кожи для систем здравоохранения постоянно возрастает [236, 267], а модельные исследования

зарубежных авторов также демонстрируют потенциальную экономическую эффективность интеграции ИИ в диагностические сценарии, включая скрининг [132, 177, 178, 202]. Разработанный способ развивает отечественную методологию комплексной оценки МЭЭ здравоохранения и клинико-экономического анализа медицинских изделий [2, 3, 68], дополняя ее симметричным учетом потерь от несвоевременного выявления целевого заболевания. Предложенная формула расчета годовых эквивалентных затрат на медицинское изделие с ИИ на основе метода аннуитетных платежей, хотя и не применялась в настоящем исследовании ввиду отсутствия верифицированных данных, представляет методологический инструмент для проспективных исследований по мере накопления соответствующих экономических данных [67, 102].

Клиническое значение сокращения сроков установления диагноза на 10,5 суток обусловлено тем, что прогноз при меланоме напрямую зависит от глубины инвазии опухоли и стадии заболевания на момент диагностики, что определяет исключительную важность ее раннего выявления [200, 227, 280]. Сохраняющаяся в Российской Федерации значимая доля меланом, выявляемых на поздних стадиях, ассоциированных с неблагоприятным прогнозом [98, 101, 106], подчеркивает практическую значимость технологий, ускоряющих диагностику и снижающих долю запущенных случаев. Ограничения выполненного экономического анализа (ретроспективный характер данных, привязка к тарифам ОМС конкретного региона, учет преимущественно прямых затрат) определяют перспективы дальнейших проспективных многоцентровых исследований с использованием реальных клинических и финансово-экономических данных.

Обсуждение результатов разработки и клинической апробации программы «Derma Check»

Итоговым результатом исследования стала разработка, валидация и апробация комплексной мультимодульной СППВР «Derma Check», объединяющей модули анализа фотоизображений, анамнеза, объяснимого ИИ,

персонализированные алгоритмы маршрутизации, кабинет врача и функцию динамического наблюдения образований. Концептуальной основой архитектуры системы является модель «дополненного интеллекта», при которой ИИ дополняет врача, а окончательная ответственность за диагностическое решение сохраняется за специалистом [160, 246, 345]. Апробация системы на массиве 2250 клинических изображений новообразований кожи из нескольких независимых клинических центров продемонстрировала высокую диагностическую эффективность (формальная чувствительность – 96,7%, специфичность – 84,7% при $AUC=0,985$; клиническая чувствительность – 93,3%, специфичность – 96,2%, точность – 95,3%), что согласуется с результатами метаанализов, подтверждающих конкурентоспособность нейросетевых алгоритмов относительно врачей-специалистов [118, 121, 196, 197, 324].

Принципиальным результатом апробации является полное выявление всех меланом при том, что все 13 пропусков системы ограничились базальноклеточным раком, наименее агрессивной из злокачественных нозологий, крайне редко метастазирующей [164, 195]. Клиническая ценность системы как инструмента «второго мнения» наглядно подтверждена пятью наблюдениями, в которых программа обоснованно указала высокий риск при доброкачественном предварительном диагнозе опытных дерматовенерологов, а последующее гистологическое исследование подтвердило злокачественность. Отсутствие статистически значимого различия в полноте выявления злокачественных образований между программой и врачами-дерматовенерологами (точный критерий Фишера $p=0,559$) при значительном уровне согласия заключений с консенсусом экспертов (каппа Коэна 0,750) подтверждает пригодность системы для применения на первичном этапе диагностики. Реализованные в системе решения, мультимодельная архитектура на основе ансамбля сверточных нейронных сетей и трансформеров, автоматический контроль качества изображений, модуль объяснимого ИИ и интеграция анамнестических данных, отвечают ключевым требованиям к клинически применимым системам ИИ, сформулированным в литературе [246, 258, 265, 271], а сочетание архитектур

сверточных сетей и трансформеров позволяет учитывать как макроскопические морфологические признаки, так и пространственные зависимости, что соответствует современным тенденциям развития мультимодальных систем [60, 133, 343].

Проведенное нами исследование для оценки влияния вида фотоизображений (дерматоскопических или клинических) на эффективность и точность классификации программой «Derma Check» путем попарного сравнения метрик моделей при использовании дерматоскопических и клинических фотоизображений на идентичном клиническом материале показало отсутствие статистически значимых различий в общей точности (тест Макнемара $p=0,267$) при высокой дискриминационной способности модели для обоих типов визуализации ($AUC>0,98$). Дерматоскопия обеспечивала более высокую чувствительность, тогда как клиническое фото – более высокую специфичность и лучшую калибровку. Выявленная тенденция дерматоскопических изображений генерировать более высокие «подозрительные» вероятности для доброкачественных образований объяснима спецификой дерматоскопической визуализации, выявляющей детальные структурные паттерны (пигментную сеть, глобулы, бесструктурные зоны), которые присутствуют и в доброкачественных образованиях [332, 374]. Этот результат имеет принципиальное практическое значение: разработанная модель робастна к типу устройства фотофиксации и может применяться врачами первичного звена с камерой смартфона без дерматоскопа, что критически важно, поскольку именно врачи общей практики и терапевты, как правило, не располагают дерматоскопами и не владеют методом [25, 190, 296].

Обсуждение результатов разработки и апробации модуля непрерывного профессионального развития «Derma Check Educational Module»

Разработанный и апробированный образовательный модуль непрерывного профессионального развития «Derma Check Educational Module» отвечает актуальной потребности в совершенствовании подготовки врачей в условиях

перехода медицинского образования от «эпохи информации» к «эпохе искусственного интеллекта» [152, 255, 370]. Применение ИИ-инструментов в медицинском образовании рассматривается как перспективное направление, способствующее персонализации обучения, ускорению освоения практических навыков и объективизации оценки знаний [145, 151, 152, 268]. Дерматовенерология как специальность с преимущественно визуальным характером диагностики представляет одну из наиболее благоприятных областей для интеграции ИИ в образовательный процесс [321, 359]. Проведенное нами проспективное контролируемое исследование продемонстрировало статистически значимое повышение как непосредственных, так и отсроченных показателей усвоения знаний в группе интерактивного обучения (доля сохранения знаний через 30 дней 69,9% против 60,3%; $p=0,008$), что согласуется с данными о роли немедленной обратной связи и активного воспроизведения информации в формировании устойчивых следов памяти при медицинском обучении. Полученные результаты подтверждают, что интеллектуальные системы поддержки принятия врачебных решений в дерматовенерологии могут выступать не только диагностическим, но и полноценным обучающим инструментом [50, 51, 80], обеспечивающим формирование навыков корректной интерпретации заключений программ компьютерного зрения и понимания их ограничений.

Интегративное обсуждение и место результатов в системе доказательной оценки медицинского искусственного интеллекта

Совокупность полученных результатов образует логически завершенную систему объективной оценки диагностических технологий ИИ в дерматоонкологии – от отдельного диагностического акта до оценки клинической и медико-экономической эффективности диагностики с применением программ ИИ на уровне региональной системы здравоохранения. Разработанные методологические инструменты (шестиэтапная методика формирования наборов данных, классификация программ для ЭВМ, методика анализа причин ошибочной

классификации, показатели клинической точности и эффективности с учетом онкологической настороженности, способы оценки медико-экономической эффективности с квадрантными матрицами визуализации и программные средства их расчета) вносят вклад в методологию доказательной медицины применительно к медицинским изделиям с ИИ. Систематические обзоры констатируют недостаточное соблюдение стандартов отчетности и преобладание методологически несовершенных исследований в области медицинского ИИ, что ограничивает доказательную базу его применения [110, 154, 269]; предложенные в настоящей работе инструменты направлены на преодоление этого разрыва между высокой точностью алгоритмов на курируемых ретроспективных наборах данных и их производительностью в реальных клинических условиях [43, 44, 46, 258].

Ограничения выполненного исследования связаны с ретроспективным характером части клинического материала, демографическими особенностями использованных баз данных, заимствованием отдельных показателей чувствительности врачей из литературы и апробацией разработанных методологических инструментов преимущественно на новообразованиях кожи. Эти ограничения не снижают методологической ценности работы, поскольку разработанные подходы носят универсальный характер и могут быть адаптированы для других областей медицинской визуализации и клинической диагностики. Перспективы дальнейшего развития связаны с расширением нозологического диагностического спектра системы, интеграцией мультимодальных данных, проведением проспективных многоцентровых клинических испытаний и с разработкой регуляторных стандартов оценки медицинских изделий с ИИ в Российской Федерации, что соответствует общемировым тенденциям развития методологии оценки цифровых медицинских технологий и требует внимания к этическим вопросам, прозрачности алгоритмических решений и защите персональных данных [27, 169, 214, 232, 341].

ЗАКЛЮЧЕНИЕ

Настоящее диссертационное исследование представляет собой комплексную многоэтапную научную работу, направленную на совершенствование дифференциальной диагностики злокачественных новообразований кожи с использованием технологий искусственного интеллекта. Исследование выполнено на основании анализа данных опросов 545 врачей трех специальностей (терапевтов и врачей общей практики, дерматовенерологов, онкологов), клинических данных и фотоизображений новообразований кожи 2250 пациентов, 769 наблюдений с гистологической верификацией, а также валидации и апробации разработанных методологических инструментов оценки клинической точности, эффективности и медико-экономической эффективности диагностики с применением технологий ИИ.

Проблема ранней и точной дифференциальной диагностики злокачественных новообразований кожи сохраняет высокую актуальность в связи с устойчивой тенденцией к росту заболеваемости ЗНО кожи как в мировом масштабе, так и в Российской Федерации. По данным систематических обзоров глобального бремени онкологических заболеваний, немеланомный рак кожи входит в пятерку наиболее распространенных злокачественных новообразований в популяциях европеоидной расы [369]. Несмотря на то что меланома составляет менее 5% всех случаев рака кожи, на ее долю приходится до 75% смертей от злокачественных новообразований кожи [226]. Указанные обстоятельства определяют исключительную важность ранней диагностики и обосновывают необходимость разработки и внедрения инструментов, повышающих точность распознавания ЗНО кожи на всех этапах оказания медицинской помощи.

Одним из наиболее перспективных направлений решения данной проблемы является применение программ компьютерного зрения на основе технологий искусственного интеллекта. За последнее десятилетие произошел качественный скачок в области автоматизированной дерматоскопической и клинической диагностики: метаанализы рандомизированных исследований демонстрируют, что

глубокие нейронные сети способны достигать уровня диагностической точности, сопоставимого с врачами-дерматовенерологами, а в ряде случаев превышающего его [360]. Вместе с тем систематические обзоры выявляют существенные методологические ограничения большинства опубликованных исследований: недостаточный учет систематических смещений в обучающих данных, использование нестандартизированных наборов данных, отсутствие внешней валидации и несоблюдение стандартов отчетности [212]. Именно эти обстоятельства определили необходимость комплексного подхода, реализованного в настоящем исследовании.

На первом этапе работы было изучено отношение врачей трех ключевых для диагностики ЗНО кожи специальностей к технологиям искусственного интеллекта. Анонимный онлайн-опрос 546 специалистов: 105 терапевтов и врачей общей практики, 191 дерматовенеролога и 250 онкологов позволил выявить ряд закономерностей, имеющих принципиальное значение для стратегии внедрения ИИ в клиническую практику. Во всех трех специальностях обнаружен характерный диссонанс между высокой декларируемой готовностью к применению ИИ (от 72,0% у онкологов до 87,6% у терапевтов) и умеренным уровнем доверия к его заключениям (35,2% доверяющих среди терапевтов). Данный феномен «не доверяю, но готов попробовать» отражает этап раннего знакомства с технологией, характерный для инновационного цикла в здравоохранении.

Единственным устойчивым предиктором позитивного отношения к ИИ во всех трех специальностях оказался практический опыт работы с ИИ-технологиями в профессиональной деятельности (у терапевтов: $U=1394$; $p=0,014$; у онкологов: $U=4778$; $p<0,001$), тогда как возраст, стаж работы, место работы, наличие ученой степени и уровень владения ПК значимого влияния на доверие не оказывали. Примечательно, что ни в одной из трех специальностей возраст не являлся статистически значимым предиктором доверия к ИИ (у терапевтов: $N=1,87$; $p=0,392$; у дерматовенерологов при 17-летнем разрыве в медианном возрасте между ординаторами и практикующими врачами: $p=0,982$). Этот результат опровергает распространенное представление о «возрастном барьере» и

свидетельствует о том, что преодоление барьеров внедрения ИИ требует не столько информирования, сколько обеспечения реального практического контакта специалистов с ИИ-системами. Сходные закономерности описаны в зарубежных исследованиях: многоцентровое европейское исследование отношения дерматовенерологов к ИИ также выявило, что практический опыт взаимодействия с ИИ-инструментами формирует более реалистичные ожидания и значительно повышает готовность к интеграции ИИ в диагностический процесс [159].

Кластерный анализ методом порогового агрегирования выделил во всех трех выборках три устойчивые группы: «Скептики», «Умеренные» и «Энтузиасты», распределение которых различалось между специальностями: доля «Энтузиастов» была максимальной среди дерматовенерологов (25,7%) и минимальной среди онкологов (20,9%), что, по всей вероятности, отражает различия в характере профессиональных задач и степени визуализационной компоненты в диагностическом процессе. Наиболее скептически настроенными среди онкологов оказались специалисты стационаров, что может быть обусловлено доминированием хирургических и химиотерапевтических задач, в которых роль ИИ менее очевидна, чем на амбулаторном диагностическом этапе.

Важнейшим методологическим результатом исследования является разработка шестиэтапной методологии формирования наборов данных из клинических баз данных для обучения моделей ИИ. Данная методология систематизирует процесс преобразования «сырых» клинических данных в готовый датасет и включает последовательное прохождение шести этапов: определение задачи и формирование технического задания, выгрузку и первичную фильтрацию данных по критериям включения и исключения, деперсонализацию в соответствии с ФЗ № 152-ФЗ, верификацию качества и медицинскую аннотацию профильными специалистами с измерением межэкспертного согласия ($\kappa=0,83$), организацию структуры со стратифицированным разбиением на обучающую, валидационную и тестовую выборки на уровне пациентов и документирование в виде паспорта датасета. Валидация методики продемонстрировала статистически значимое превосходство моделей, обученных на наборах данных с гистологическим

«золотым стандартом», по всем ключевым диагностическим метрикам по сравнению с моделями, обученными на публичных данных со смешанной верификацией: чувствительность – 96,7% против 89,2%, специфичность – 96,0% против 87,5%, AUC-ROC – 0,985 против 0,931, коэффициент согласия Коэна – 0,906 против 0,732 (тест Макнемара: $\chi^2=12,4$; $p<0,001$), оценка Бриера – 0,044 против 0,102. Превосходство модели, обученной на структурированном наборе данных, объясняется несколькими факторами: устранением шума меток, неизбежного при экспертной аннотации без морфологической верификации, исключением утечки данных за счет разбиения на уровне пациентов и стандартизацией процедуры документирования, обеспечивающей воспроизводимость.

Значимость качества обучающих данных как определяющего фактора точности моделей подтверждается результатами международных исследований, показывающих, что вариабельность межэкспертного согласия при аннотации дерматоскопических изображений создает принципиальный «потолок» точности обучаемой модели [228, 231].

Предложенное в рамках исследования формализованное разграничение понятий «база данных» и «набор данных» (датасет) устраняет существующую терминологическую неопределенность: база данных определена как динамическая операционная система, обеспечивающая текущую клиническую деятельность, тогда как набор данных – как статичный, целенаправленно сформированный и строго задокументированный продукт для решения конкретной аналитической задачи.

Разработанная оригинальная классификация программ для ЭВМ, включая основанные на технологиях ИИ и ПКЗ, применяемых в клинической медицине и здравоохранении, построенная на девятнадцати признаках, объединенных в пять тематических блоков, обеспечивает стандартизированное описание систем, необходимое для их корректного сравнения, регуляторной оценки и обоснованного выбора для внедрения в различные области клинической медицины.

Систематизация основных проблем, возникающих при диагностике ЗНО кожи с помощью программ ИИ, выполненная на основании анализа 32

литературных источников и собственного клинического опыта, позволила впервые разделить эти проблемы на две категории: трудности (объективные факторы, обусловленные природой новообразований и ограничениями методов) и сложности (субъективные факторы, связанные с квалификацией врачей и организацией медицинской помощи). Для каждой из выявленных шести групп проблем предложены технически реализуемые решения, апробированные в программе «Derma Onko Melanoma Check».

Разработанная и валидированная методология анализа причин ошибочной классификации фотоизображений новообразований кожи основана на вычислении 13 фотометрических и текстурных метрик для каждого изображения, многоуровневом статистическом анализе и методах объяснимого ИИ (Score-CAM, Occlusion Sensitivity). Однофакторный дисперсионный анализ выявил статистически значимые межгрупповые различия по яркости ($F=14,83$), контрасту ($F=18,97$), энтропии ($F=13,50$), средним значениям и стандартным отклонениям цветовых каналов (F от 6,48 до 27,12). Множественная логистическая регрессия после исключения предикторов с высокой мультиколлинеарностью ($VIF=10,2-931,1$) установила, что наибольший вклад в вероятность ошибочного заключения вносят размытие по FFT ($OR=3,08$; 95% ДИ 1,77–5,36; $p<0,001$), резкость ($OR=0,31$; 95% ДИ 0,14–0,65; $p<0,01$) и доля пересвеченных пикселей ($OR=1,64$; 95% ДИ 1,11–2,41; $p<0,05$). Псевдокоэффициент детерминации Макфаддена составил 0,085 при значимом отношении правдоподобия (LLR $p<0,001$). Двухфакторный дисперсионный анализ подтвердил незначимость фактора конкретной ПКЗ ($p=0,39-0,96$), что обеспечивает универсальность установленных референсных диапазонов для различных программ. Многомерный дисперсионный анализ MANOVA подтвердил выраженное многомерное различие групп заключений ($Wilks' \lambda=0,639$; $F=10,37$; $p<0,001$). Применение методов объяснимого ИИ для систематической интерпретации диагностических ошибок нейронных сетей рассматривается в современной литературе как обязательный компонент клинической валидации медицинских изделий с ИИ, обеспечивающий

прозрачность алгоритмических решений и повышающий доверие врачей к заключениям системы [206, 223].

Разработанная методология оценки клинической точности диагностики с учетом онкологической настороженности восполняет существенный пробел в существующих подходах к оценке ИИ-систем в дерматоонкологии. Стандартная матрица ошибок расценивает любое заключение о повышенном риске по доброкачественному образованию как ложноположительный результат, тогда как в клинической практике направление пациента с предраковым или атипичным образованием на биопсию является клинически верным решением. Предложенная система четырех категорий клинической эффективности: «полное совпадение», «совпадение по онкологической настороженности», «несовпадение по онкологической настороженности» и «полное несовпадение» – и формулы расчета клинической чувствительности, клинической специфичности и клинической точности позволяют адекватно оценить клиническую ценность программы с учетом реальной диагностической практики. На выборке из 342 новообразований кожи показано, что 12,9% наблюдений, квалифицируемых стандартной матрицей как ложноположительные, относятся к клинически ценной категории «совпадение по онкологической настороженности»; клиническая точность при этом составила 92,1% (95% ДИ 88,8–94,5) против формальной точности 89,7%. Межэкспертная согласованность подтверждена коэффициентом каппа Коэна для каждой пары экспертов; асимметрия расхождений проверена тестом Макнемара.

Разработанный и валидированный способ оценки клинической эффективности диагностики с использованием алгоритмов на основании заключений моделей ИИ позволяет количественно оценить преимущества внедрения ИИ через коэффициенты прироста выявляемости и сокращения пропусков. Применение алгоритмов маршрутизации продемонстрировало увеличение выявляемости ЗНО в 1,9 раза и снижение пропусков в 10,5 раза, снижение доли необоснованных направлений к дерматовенерологам с 40,6% до 6,9% и к онкологам с 22,0% до 7,6%.

Способ оценки медико-экономической эффективности с квадрантной матрицей предотвратимых и потенциально предотвратимых финансовых затрат показал снижение общих затрат на диагностику меланоцитарных опухолей кожи на 51%, а потенциальная экономия необоснованных финансовых затрат только при меланоме кожи в г. Москве составляет более 436 млн руб. в год при сокращении сроков диагностики на 10,5 суток. Результаты экономического моделирования согласуются с данными зарубежных фармакоэкономических исследований, демонстрирующих, что интеграция ИИ в дерматоонкологический скрининг обеспечивает снижение стоимости лечения [204].

Итоговым результатом исследования стала разработка, валидация и апробация комплексной СППВР «Derma Check» – мультимодульной программы для автоматизированной визуальной идентификации и дифференциальной диагностики патологии кожи с модулями анализа фотоизображений, анамнеза, объяснимого ИИ, персонализированными алгоритмами маршрутизации, кабинетом врача и функцией динамического наблюдения образований. На массиве 2250 новообразований кожи, полученных из трех независимых клинических центров, мультимодельная программа продемонстрировала формальную чувствительность 96,7% (95% ДИ 94,5–98,1) и специфичность 84,7% (95% ДИ 83,0–86,3) при площади под ROC-кривой $AUC=0,985$. Отрицательная прогностическая ценность составила 99,2% (95% ДИ 98,6–99,5), отрицательное отношение правдоподобия $LR- = 0,038$, индекс Юдена – 0,815, коэффициент корреляции Мэтьюса $MCC=0,681$. С учетом онкологической настороженности клиническая чувствительность составила 93,3%, клиническая специфичность – 96,2%, клиническая точность – 95,3%. Принципиальным результатом является полное выявление всех 195 меланом (чувствительность для меланомы 100,0%); все 13 пропусков системы ограничены базальноклеточным раком – наименее агрессивной из злокачественных нозологий, крайне редко метастазирующей. Согласие заключений программы с консенсусом трех врачей-дерматовенерологов со стажем более 20 лет по коэффициенту каппа Коэна составило 0,750, что соответствует значительному уровню согласия по шкале Ландиса–Коха. Точный критерий

Фишера при сравнении чувствительности программы и дерматовенерологов показал $p=0,559$, что свидетельствует об отсутствии статистически значимого различия в полноте выявления злокачественных образований.

Особую клиническую ценность продемонстрировали пять наблюдений, в которых программа обоснованно указала высокий риск при доброкачественном предварительном диагнозе опытных дерматовенерологов (2 диспластических невуса и 3 актинических кератоза), а последующее гистологическое исследование подтвердило злокачественность. Данные случаи наглядно демонстрируют дополнительную диагностическую ценность системы как инструмента «второго мнения», способного выявлять злокачественные новообразования, пропущенные на клиническом этапе. Многоклассовое ранжирование по классам новообразований показало top-1=89,4%, top-3=95,4%, top-5=98,7% при значениях top-1 выше 80% во всех нозологических группах. На отдельном дерматологическом валидационном наборе ($n=120$) классификатор продемонстрировал top-1=82,0%, top-3=90,7%, top-5=95,1%, что подтверждает способность программы к дифференцировке не только новообразований, но и воспалительных дерматозов.

Сравнительный анализ диагностической эффективности модели при использовании дерматоскопических и клинических фотоизображений показал, что модель демонстрирует высокую точность при обоих типах фотофиксации ($AUC>0,98$; Brier score $<0,11$; $F1>0,73$). Тест Макнемара не выявил статистически значимого различия в общей точности ($\chi^2=1,231$; $p=0,267$). При этом каждый тип визуализации обладает специфическими преимуществами: дерматоскопия обеспечивает более высокую чувствительность (0,938 против 0,875) и лучшее LR– (0,069), тогда как клиническое фото – более высокую специфичность (0,963 против 0,908) и лучшее LR+ (23,65 против 10,19). Корреляция вероятностей между модальностями носит умеренный характер (Pearson $r=0,632$; ICC=0,617), а ансамблирование CNN- и трансформер-компонентов компенсирует их разнонаправленную чувствительность к типу входных данных. Этот результат имеет принципиальное значение для клинической практики: программа может

быть использована врачами первичного звена с камерой смартфона, без дерматоскопа, без существенной потери качества диагностики.

Разработанный и апробированный модуль непрерывного профессионального развития «Derma Check Educational Module» на основе технологий ИИ для интерактивного обучения визуальной диагностике и повышения компетенций в дерматоонкологии отвечает актуальной потребности в совершенствовании подготовки врачей. Модуль обеспечивает формирование тестов и клинических задач на основании текстовых запросов и медицинских изображений, автоматизированную проверку ответов, включает банк медицинских изображений, тестов, клинических задач и справочной литературы. Применение ИИ-инструментов в медицинском образовании рассматривается в современной литературе как перспективное направление, способствующее персонализации обучения и повышению диагностических навыков обучающихся, включая формирование навыков дерматоскопии и клинического мышления в дерматоонкологии [153, 315].

Таким образом, в диссертационном исследовании решена крупная научная проблема, имеющая важное значение для дерматовенерологии и онкологии: разработан, научно обоснован и внедрен в клиническую практику комплексный подход к дифференциальной диагностике злокачественных новообразований кожи с использованием технологий искусственного интеллекта. Разработанные методологические инструменты: шестиступенчатая методика формирования наборов данных; оригинальная классификация программ для ЭВМ, включая основанные на технологиях ИИ и ПКЗ, применяемых в клинической медицине и здравоохранении, построенная на девятнадцати признаках, объединенных в пять тематических блоков; методика анализа причин ошибочной классификации; показатели клинической точности с учетом онкологической настороженности; способы оценки клинической и медико-экономической эффективности – формируют целостную систему объективной оценки диагностических технологий ИИ в дерматоонкологии, пригодную для сравнительных исследований, регуляторной экспертизы и обоснования управленческих решений. Полученные результаты

вносят вклад в методологию доказательной медицины применительно к медицинским изделиям с ИИ и могут быть адаптированы для других областей медицинской визуализации и клинической диагностики. Перспективы дальнейшего развития связаны с расширением нозологического спектра программы, интеграцией мультимодальных данных (клинических, анамнестических и молекулярно-генетических), проведением многоцентровых проспективных клинических исследований и с разработкой регуляторных стандартов оценки медицинских изделий с ИИ в Российской Федерации.

ВЫВОДЫ

1. Среди врачей-терапевтов, дерматовенерологов и онкологов преобладает умеренный дефицит доверия к технологиям ИИ: 64,8–68,7% специалистов оценивают его как незначительное или полностью отсутствующее. Единственным достоверным предиктором позитивного отношения к технологиям искусственному интеллекту является реальный практический опыт взаимодействия с технологиями ИИ в профессиональной деятельности.

2. Шестиэтапная методология формирования наборов данных демонстрирует эффективность (чувствительность, специфичность и точность модели 96,7% (94,5–98,1), 96,0% (93,3–97,6) и 96,2% (93,8–97,7) против 89,2%, 87,5% и 88,0%, соответственно, у модели, обученной без ее применения; коэффициент согласия Коэна 0,906 против 0,732; тест Макнемара $\chi^2=12,4$; $p<0,001$) и может применяться при подготовке наборов данных (датасетов) из клинических баз данных для программ компьютерного зрения для применения в дерматовенерологии, других областях клинической медицины и здравоохранении.

3. В РФ отмечается дефицит медицинских изделий на основе технологий ИИ для диагностики новообразований кожи и дерматозов. Единственным зарегистрированным в реестре Росздравнадзора медицинским изделием с ИИ для диагностики новообразований кожи является сервис поддержки принятия врачебных решений «ПроРодинки», предназначенный для распознавания пяти нозологических единиц (меланома, базальноклеточный рак кожи, меланоцитарный невус, себорейный кератоз и гемангиома) с чувствительностью не менее 85%, специфичность не менее 80%, точность 81,5%, в заключении предоставляется результат бинарной классификации без указания конкретного диагноза.

Классификация программ для ЭВМ, в том числе на основе технологий ИИ и программ компьютерного зрения, стандартизировано всесторонне и корректно описывает с научной и практической точек зрения и программы для ЭВМ, применяемые в дерматовенерологии, других областях клинической медицины и

здравоохранении; позволяет проводить сопоставления программ для ЭВМ, в том числе на основе технологий ИИ и программ компьютерного зрения.

4. Классификация проблем дифференциальной диагностики новообразований кожи с разделением на объективные трудности и субъективные сложности носит содержательный характер и позволяет отделить факторы, коренящиеся в природе новообразований и методов их исследования, от факторов, связанных с человеческим и организационным звеном.

Шесть групп основных проблем применения программ для ЭВМ на основе технологий ИИ для диагностики новообразований кожи имеют валидированные, технически реализуемые решения: (1) автоматический контроль качества фотоизображений с блокировкой некачественных фотоизображений; (2) контроль размера образования в кадре; (3) мультимодельная архитектура на основе ансамбля нейронных сетей с механизмом голосования; (4) визуализация зон риска и текстовый вывод распознанных признаков для обеспечения объяснимости заключений; (5) интерактивный интерфейс для ввода данных анамнеза и дополнительных клинических признаков; (6) поддержка мультимодального анализа клинических и дерматоскопических фотоизображений.

5. Валидированный способ анализа причин ошибочной классификации фотоизображений новообразований кожи с применением методов объяснимого ИИ позволяет выявить причины ошибочной классификации новообразований кожи программами для ЭВМ на основе технологий ИИ (для меланоцитарных новообразований ($n=485$ фотоизображений) ложноположительные ошибки обусловлены повышенной яркостью (0,740 против 0,638 ед.) и сниженной энтропией (4,898 против 5,159 ед.), для немеланоцитарных ($n=151$ фотоизображение) – яркостью (0,7994 против 0,6914 ед.) и энтропией (4,6414 против 4,8584 ед.) и может применяться для оптимизации и модернизации моделей и программ для ЭВМ, основанных на технологиях ИИ, для диагностики дерматозов и новообразований кожи с целью повышения точности и эффективности диагностики (дообучение с учетом выявленных факторов повысило общую

диагностическую точность с 84,3 до 92,4% для программы «Melanoma Check» и с 90,9 до 92,4% для программы «Derma Onko Check»).

6. Применение методики оценки клинической точности и клинической эффективности программ ИИ в дерматовенерологии с учетом онкологической настороженности на выборке из 342 новообразований кожи выявляет 12,9% наблюдений, квалифицируемых стандартной матрицей как ложноположительные, которые относятся к клинически ценной категории «совпадение по онкологической настороженности», а клиническая точность составляет 92,1% (95% ДИ 88,8–94,5) против формальной точности диагностики 89,7%.

7. При анализе с использованием способа оценки клинической эффективности диагностики новообразований кожи с использованием алгоритмов на основании заключений программ ИИ на выборке (n=381 пациент) с помощью программы для ЭВМ отмечается, что применение алгоритмов маршрутизации на основании заключений программы «Derma Onko Check» обеспечивает снижение доли необоснованных направлений к дерматовенерологам с 40,6 до 6,9%, к онкологам с 22,0 до 7,6%; коэффициент соотношения финансовых затрат 0,49, что соответствует снижению общих затрат на 51%).

8. При расчетах с применением способа оценки медико-экономической эффективности диагностики новообразований кожи с применением алгоритмов маршрутизации пациентов на основании заключений моделей ИИ с помощью программы для ЭВМ отмечается: на уровне учреждения применение технологий ИИ снижает общие затраты на 51%, а на региональном уровне на примере г. Москвы обеспечивает потенциальную экономию порядка 436 млн руб. в год).

9. При клиническом применении программы для ЭВМ, основанной на технологиях ИИ, системы поддержки принятия врачебных решений для автоматизированной визуальной идентификации и дифференциальной диагностики дерматозов и новообразований кожи «Derma Check» отмечается: формальная чувствительность, специфичность и точность 96,7% (94,5–98,1), 84,7% (83,0–86,3), 86,8% (85,4–88,2), соответственно; клиническая 93,3%, 96,2%, 95,3%;

вероятность верного диагноза топ-1/топ-3/топ-5 – 89,4%, 95,4% и 98,7% для новообразований и 82,0%, 90,7% и 95,1% для дерматозов, соответственно).

10. Применение в учебном процессе программного модуля непрерывного профессионального развития на основе технологий ИИ «Derma Check Educational Module» обеспечивает статистически значимое повышение показателей усвоения знаний (доля сохранения знаний через 30 дней 69,9 против 60,3%, $p=0,008$; время усвоения одной нозологической единицы на 25% меньше, 4,8 против 6,4 мин, $p=0,004$).

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. При проектировании дополнительных профессиональных программ, направленных на формирование осознанного подхода к применению ИИ в медицине, рекомендуется обеспечивать практический контакт обучающихся с ИИ-системами, поскольку именно практический опыт, а не возраст, стаж или информированность, является единственным устойчивым предиктором позитивного отношения к технологиям искусственного интеллекта.

2. При разработке программ компьютерного зрения для диагностики новообразований кожи и дерматозов рекомендовано применение шестиэтапной методики формирования обучающих наборов данных (датасетов), последовательно включающей: определение задачи и формирование технического задания; выгрузку и первичную фильтрацию данных; деперсонализацию в соответствии с Федеральным законом «О персональных данных» от 27.07.2006 № 152-ФЗ; верификацию качества и медицинскую аннотацию не менее чем двумя специалистами экспертного уровня с измерением межэкспертного согласия; организацию структуры со стратифицированным разбиением на уровне пациентов, балансировку и аугментацию; документирование в виде паспорта датасета. Для новообразований кожи обучение моделей рекомендуется проводить на наборах данных фотоизображений новообразований, подтвержденных гистологической диагностикой – «золотым стандартом» диагностики, поскольку это обеспечивает статистически значимое повышение диагностической точности по сравнению с использованием наборов данных с экспертной визуальной аннотацией.

3. Разработчикам программного обеспечения рекомендовано проектировать комплексные мультимодельные и мультимодальные программы для ЭВМ - системы поддержки принятия врачебных решений на основе программ компьютерного зрения и технологий искусственного интеллекта, сочетающие бинарные и многоклассовые классификации, для применения врачами общей практики, терапевтами, дерматовенерологами и косметологами для

дифференциальной диагностики дерматозов и выявления злокачественных новообразований кожи.

Классификация программ для ЭВМ, включая основанные на технологиях искусственного интеллекта и программ компьютерного зрения, применяемых в дерматовенерологии, клинической медицине и здравоохранении, построенная на девятнадцати признаках, объединенных в пять тематических блоков, рекомендована для стандартизированного описания программ для ЭВМ, включая основанные на технологиях искусственного интеллекта и программ компьютерного зрения, необходимого для их корректного сравнения, регуляторной оценки и обоснованного выбора для внедрения в различные области клинической медицины.

4. Классификация проблем дифференциальной диагностики злокачественных и доброкачественных новообразований кожи с разделением на объективные трудности и субъективные сложности рекомендована при анализе выявляемости злокачественных новообразований кожи для отделения факторов, коренящихся в природе новообразований и методов их исследования, от факторов, связанных с человеческим и организационным звеном.

Предложенные решения основных групп проблем применения программ для ЭВМ на основе технологий искусственного интеллекта для диагностики новообразований кожи: 1) автоматический контроль качества фотоизображений с блокировкой некачественных фотоизображений; 2) контроль размера образования в кадре; 3) мультимодельная архитектура на основе ансамбля нейронных сетей с механизмом голосования; 4) визуализация зон риска и текстовый вывод распознанных признаков для обеспечения объяснимости заключений; 5) интерактивный интерфейс для ввода данных анамнеза и дополнительных клинических признаков; 6) поддержка мультимодального анализа клинических и дерматоскопических изображений – рекомендованы разработчикам программного обеспечения для проектирования программ для ЭВМ на основе технологий искусственного интеллекта для применения в дерматовенерологии и дерматоонкологии.

5. Способ анализа причин ошибочной классификации фотоизображений с применением методов объяснимого искусственного интеллекта рекомендован для анализа эффективности программ для ЭВМ на основе технологий искусственного интеллекта, использующих технологии компьютерного зрения в дерматовенерологии и других областях клинической медицины и здравоохранения.

6. Методология оценки клинической точности и клинической эффективности программ искусственного интеллекта с учетом онкологической настороженности рекомендована при оценке программ для ЭВМ на основе технологий искусственного интеллекта для диагностики новообразований кожи.

7. Способ оценки клинической эффективности диагностики новообразований кожи программами для ЭВМ для автоматизированной оценки эффективности моделей искусственного интеллекта и алгоритмов маршрутизации на основании заключений моделей искусственного интеллекта в диагностике новообразований кожи рекомендован при внедрении программ искусственного интеллекта для диагностики новообразований кожи в медицинских организациях. Для автоматизации оценки рекомендована программа для ЭВМ.

8. Способ оценки медико-экономической эффективности диагностики с применением алгоритмов маршрутизации пациентов на основании заключений моделей искусственного интеллекта рекомендован на уровне медицинской организации (с квадрантной матрицей предотвратимых и потенциально предотвратимых финансовых затрат), сети медицинских организаций и региона (с расчетом потенциальной экономии необоснованных финансовых затрат при экстраполяции на регион) для сравнительной оценки диагностических стратегий. Программа для ЭВМ для оценки медико-экономической эффективности диагностики с применением алгоритмов маршрутизации пациентов на основании заключений моделей искусственного интеллекта рекомендована для автоматизированного расчета медико-экономической эффективности.

9. Комплексная программа для ЭВМ на основе технологий искусственного интеллекта, система поддержки принятия врачебных решений для автоматизированной диагностики новообразований кожи и дерматозов «Derma

Check» с модулями анализа фотоизображений, анамнеза, объяснимого искусственного интеллекта, с персонализированными алгоритмами маршрутизации, кабинетом врача и функцией динамического наблюдения образований после регистрации как медицинское изделие рекомендована для применения в дифференциальной диагностике и определения маршрутизации врачами-терапевтами, врачами общей практики, дерматовенерологами, косметологами и уполномоченными медицинскими работниками при диагностике новообразований кожи и дерматозов при оказании первичной врачебной и специализированной медико-санитарной помощи, а также в качестве инструмента «второго мнения» для дерматовенерологов и онкологов при оказании первичной специализированной медико-санитарной помощи.

10. Модуль непрерывного профессионального развития «Derma Check Educational Module» с возможностью формирования тестов и клинических задач на основании дерматоскопических и клинических фотоизображений рекомендован к применению в образовательных и медицинских организациях, а также для самостоятельного применения врачами для повышения компетенций в области дерматовенерологии и дерматоонкологии и формирования навыков использования технологий искусственного интеллекта в клинической медицине.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

АРМ – автоматизированное рабочее место

БКРК – базальноклеточный рак кожи

ВЗН – высокозатратные нозологии

ВНП – валовой национальный продукт

ВОЗ – Всемирная организация здравоохранения

ВОП – врач общей практики

ВР – верный результат

ГОСТ – государственный стандарт

ДИ – доверительный интервал

ДНО – доброкачественное новообразование

ЕГИСЗ – Единая государственная информационная система в сфере здравоохранения

ЕМИАС – Единая медицинская информационно-аналитическая система

ЕС – Европейский союз

ЗНО – злокачественное новообразование

ИГХ – иммуногистохимия (иммуногистохимическое исследование)

ИИ – искусственный интеллект

ИО – истинно отрицательный результат

ИП – истинно положительный результат

ИППП – инфекции, передаваемые половым путем

КЛКТ – конусно-лучевая компьютерная томография

КПК – карманный персональный компьютер

КСГ – клинико-статистическая группа

КТ – компьютерная томография

ЛИС – лабораторная информационная система

ЛО – ложноотрицательный результат

ЛП – ложноположительный результат

МИ – медицинское изделие

МКБ – Международная классификация болезней
МРТ – магнитно-резонансная томография
МЭЭ – медико-экономическая эффективность
НМРК – немеланомный рак кожи
НПР – непрерывное профессиональное развитие
ОГК – органы грудной клетки
ОИТ – отделение интенсивной терапии
ОМС – обязательное медицинское страхование
ОС – операционная система
ПАИ – программа на основе искусственного интеллекта
ПД – персональные данные
ПК – персональный компьютер
ПКЗ – программа компьютерного зрения
ПКРК – плоскоклеточный рак кожи
ПЭТ – позитронно-эмиссионная томография
СППВР – система поддержки принятия врачебных решений
СФР – Социальный фонд России
УЗИ – ультразвуковое исследование
ФДГ – фтордезоксиглюкоза
ФЗ – фотоизображение
ЦЗ – целевое заключение
ЭВМ – электронная вычислительная машина
ЭКГ – электрокардиограмма
ЭМК – электронная медицинская карта
ABCD(E) – асимметрия, границы, цвет, диаметр, эволюция
AI – искусственный интеллект
ANOVA – дисперсионный анализ
API – программный интерфейс приложения
AUC – площадь под кривой
AUC-ROC – площадь под характеристической кривой

BRAF – ген серин-треониновой протеинкиназы
 B-Raf CIELAB – цветовое пространство Международной комиссии по освещению
 CNN – сверточная нейронная сеть
 CONSORT-AI – консолидированные стандарты отчетности клинических испытаний с искусственным интеллектом
 COVID-19 – новая коронавирусная инфекция
 2019 CV – компьютерное зрение
 DECIDE-AI – стандарт отчетности поисковых клинических исследований систем поддержки принятия решений на основе искусственного интеллекта
 DICOM – цифровые изображения и обмен данными в медицине
 FastAPI – веб-фреймворк для создания программных интерфейсов на языке Python
 FFT – быстрое преобразование Фурье
 FHIR – стандарт обмена медицинскими данными (Fast Healthcare Interoperability Resources)
 FN – ложноотрицательный результат
 FP – ложноположительный результат
 GAN – генеративно-состязательная сеть
 Grad-CAM – карта активации классов, взвешенная по градиентам
 HAM10000 – набор данных «человек против машины» из 10000 обучающих изображений
 IARC – Международное агентство по изучению рака
 IQR – межквартильный размах
 ISIC – Международное сотрудничество по визуализации кожи
 LLM – большая языковая модель
 MANOVA – многомерный дисперсионный анализ
 MCC – коэффициент корреляции Мэтьюса
 NPV – прогностическая ценность отрицательного результата
 ONNX – открытый формат представления моделей машинного обучения (Open Neural Network Exchange)
 PACS – система архивации и передачи медицинских изображений

PCA – метод главных компонент

PPV – прогностическая ценность положительного результата

RGB – красный, зеленый, синий

ROC – рабочая характеристическая кривая

SaaS – программное обеспечение как услуга

Score-CAM – карта активации классов, взвешенная по оценкам

SD – стандартное отклонение

SE – чувствительность

SQLite – встраиваемая реляционная система управления базами данных

SPiRiT-AI – стандартные элементы протокола: рекомендации для интервенционных испытаний с искусственным интеллектом

TN – истинно отрицательный результат

TP – истинно положительный результат

VIF – фактор инфляции дисперсии

WHO – Всемирная организация здравоохранения

XAI – объяснимый искусственный интеллект

СПИСОК ЛИТЕРАТУРЫ

1. Алгоритм маршрутизации пациентов при применении программы искусственного интеллекта «Derma Onko Check» при дифференциальной диагностике новообразований кожи / А. И. Ламоткин, Д. И. Корабельников, О. Ю. Олисова, И. А. Ламоткин // Современные проблемы здравоохранения и медицинской статистики. – 2025. – № 5. – С. 139–159. – DOI: 10.24412/2312-2935-2025-5-139-159.
2. Анализ медико-экономической эффективности деятельности структурных подразделений медицинской организации / Л. В. Марисов, Е. Е. Кобяцкая, М. М. Бутарева, А. В. Мецгер // Проблемы социальной гигиены, здравоохранения и истории медицины. – 2022. – Т. 30. – № 5. – С. 853–858. – DOI: 10.32687/0869-866X-2022-30-5-853-858.
3. Бабенко, А. И. Методологические основы комплексной оценки медико-экономической эффективности здравоохранения / А. И. Бабенко, О. В. Пушкарев // Бюллетень Сибирского отделения Российской академии медицинских наук. – 2014. – Т. 34. – № 2. – С. 89–94.
4. Булычева, Е. В. Искусственный интеллект как новое явление в развитии здравоохранения и медицинского образования (обзор литературы) / Е. В. Булычева // Медицинское образование и профессиональное развитие. – 2022. – Т. 13. – № 3 (47). – С. 76–84. – DOI: 10.33029/2220-8453-2022-13-3-76-84.
5. Влияние различных факторов на дерматоскопическую характеристику меланоцитарных невусов / И. А. Ламоткин, Е. В. Мухина, О. Г. Капустина [и др.] // Медицинский вестник МВД. – 2021. – № 3 (112). – С. 70–75. – DOI: 10.52341/20738080_2021_112_3_70.
6. Внедрение технологий искусственного интеллекта в здравоохранении России: итоги 2024 г. / М. А. Мурашко, В. В. Ваньков, А. И. Панин [и др.] // Национальное здравоохранение. – 2025. – Т. 6. – № 3. – С. 6–19. – DOI: 10.47093/2713-069X.2025.6.3.6-19.

7. Гипердиагностика и гиподиагностика меланом на амбулаторном приеме у дерматолога / И. А. Ламоткин, Е. В. Мухина, О. Г. Капустина [и др.] // Актуальные вопросы дерматовенерологии : сборник научных трудов по материалам Всероссийской научно-практической конференции с международным участием, посвященной 80-летию кафедры дерматовенерологии КГМУ и 100-летию со дня рождения профессора В.А. Леонова, Курск, 11 мая 2018 года / Под ред. Л. В. Силиной, Т. П. Исаенко. – Курск : Курский государственный медицинский университет, 2018. – С. 95–99.
8. ГОСТ Р 59921.0-2022. Системы искусственного интеллекта в клинической медицине. Общие положения : национальный стандарт Российской Федерации : дата введения 2022-09-01. – Москва : Стандартинформ, 2022. – 16 с.
9. ГОСТ Р 59921.1-2022. Системы искусственного интеллекта в клинической медицине. Часть 1. Клиническая оценка : национальный стандарт Российской Федерации : дата введения 2022-09-01. – Москва : Стандартинформ, 2022. – 12 с.
10. ГОСТ Р 59921.2-2021. Системы искусственного интеллекта в клинической медицине. Часть 2. Программа и методика технических испытаний : национальный стандарт Российской Федерации : дата введения 2021-09-01. – Москва : Стандартинформ, 2021. – 12 с.
11. ГОСТ Р 59921.4-2021. Системы искусственного интеллекта в клинической медицине. Часть 4. Оценка и контроль эксплуатационных параметров : национальный стандарт Российской Федерации : дата введения 2021-09-01. – Москва : Стандартинформ, 2021. – 11 с.
12. ГОСТ Р 59921.5-2022. Системы искусственного интеллекта в клинической медицине. Часть 5. Требования к структуре и порядку применения набора данных для обучения и тестирования алгоритмов : национальный стандарт Российской Федерации : дата введения 2022-09-01. – Москва : Стандартинформ, 2022. – 12 с.
13. ГОСТ Р 59921.7-2022. Системы искусственного интеллекта в клинической медицине. Алгоритмы анализа медицинских изображений. Методы испытаний. Общие требования : национальный стандарт Российской Федерации : дата введения 2022-09-01. – Москва : Стандартинформ, 2022. – 10 с.

14. ГОСТ Р 71672-2024. Системы прогнозной аналитики на основе искусственного интеллекта в клинической медицине. Основные положения : национальный стандарт Российской Федерации. – Москва : Стандартинформ, 2024.
15. ГОСТ Р 71673-2024. Системы искусственного интеллекта в лучевой диагностике. Алгоритмы анализа медицинских изображений. Методы испытаний на определение точности измерений : национальный стандарт Российской Федерации. – Москва : Стандартинформ, 2024. – 13 с.
16. ГОСТ Р 71674-2024. Системы искусственного интеллекта в клинической медицине. Набор данных в формате DICOM для тестирования алгоритмов. Методы обезличивания набора данных и контроля набора данных на отсутствие персональных данных : национальный стандарт Российской Федерации. – Москва : Стандартинформ, 2024. – 9 с.
17. ГОСТ Р 71738-2024. Системы искусственного интеллекта в лучевой диагностике. Алгоритмы анализа медицинских изображений. Методы испытаний на способность и устойчивость работы с разнородными данными : национальный стандарт Российской Федерации. – Москва : Стандартинформ, 2024. – 8 с.
18. Границы профессиональной ответственности врача общей практики (семейного врача): результаты всероссийского исследования / О. М. Драпкина, В. В. Фомин, Т. Е. Морозова [и др.] // Кардиоваскулярная терапия и профилактика. – 2026. – Т. 25. – № 2S. – С. 6–17. – DOI: 10.15829/1728-8800-2026-4851.
19. Дерматоскопическая характеристика различных сосудистых признаков в опухолях кожи / И. А. Ламоткин, Д. И. Корабельников, Е. Г. Перевалова, А. И. Ламоткин // Медицинский вестник ГВКГ им. Н.Н. Бурденко. – 2023. – № 1 (11). – С. 12–21. – DOI: 10.53652/2782-1730-2023-4-1-12-21.
20. Диагностика диспластических невусов по клиническим, дерматоскопическим и гистологическим признакам / И. А. Ламоткин, Г. Н. Тарасенко, Д. И. Корабельников [и др.] // Медицинский вестник МВД. – 2023. – Т. 124. – № 3 (124). – С. 42–47. – DOI: 10.52341/20738080_2023_124_3_42.
21. Диагностика и маршрутизация пациентов с подозрением на злокачественные новообразования кожи в условиях первичного звена здравоохранения: пробелы и

перспективы / Т. А. Гайдина, А. С. Дворников, В. Н. Ларина [и др.] // Архивъ внутренней медицины. – 2024. – Т. 14. – № 6. – С. 419–434. – DOI: 10.20514/2226-6704-2024-14-6-419-434.

22. Дифференциальная диагностика доброкачественных и злокачественных меланоцитарных опухолей кожи с применением программы искусственного интеллекта «Derma Onko Melanoma Check» / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин, О. В. Гладько // Медицинский вестник МВД. – 2026. – № 4. – С. 38–45. – DOI: 10.52341/20738080_2026_143_4_38.

23. Искусственный интеллект в здравоохранении и медицине: история ключевых событий, его значимость для врачей, уровень развития в разных странах / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин [и др.] // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2024. – Т. 17. – № 2. – С. 243–250. – DOI: 10.17749/2070-4909/farmakoeconomika.2024.254.

24. Исследование точности медицинского изделия «ПроРодинки» для диагностики базальноклеточной карциномы / Ю. А. Васильев, В. Н. Галкин, Р. А. Раводин [и др.] // Врач и информационные технологии. – 2026. – № 1. – С. 52–63. – DOI: 10.25881/18110193_2026_1_52.

25. Кадровые проблемы дерматовенерологии / И. М. Сон, М. А. Иванова, Т. А. Соколовская, Ю. И. Матушевская // Современные проблемы здравоохранения и медицинской статистики. – 2016. – № 4. – С. 17–26.

26. Клинико-дерматоскопический анализ динамических изменений невусов у пациентов с высоким риском меланомы кожи / И. И. Вахитова, А. В. Миченко, К. С. Титов [и др.] // Клиническая дерматология и венерология. – 2021. – Т. 20. – № 1. – С. 72–79. – DOI: 10.17116/klinderma202120011.

27. Кодекс этики применения искусственного интеллекта в сфере охраны здоровья. Версия 2.1 : [утвержден Межведомственной рабочей группой при Минздраве России, протокол от 14 февраля 2025 г. № 90/18-0/117]. – Текст : электронный // КонсультантПлюс : сайт. – URL: https://www.consultant.ru/document/cons_doc_LAW_501066/ (дата обращения: 05.06.2026). – Текст : электронный.

28. Корабельников, Д. И. Алгоритм маршрутизации пациентов при дифференциальной диагностике новообразований кожи при сочетанном применении программ искусственного интеллекта Derma Onko Check и Melanoma Check / Д. И. Корабельников, А. И. Ламоткин // Клинический разбор в общей медицине. – 2025. – Т. 6. – № 12. – С. 71–79. – DOI: 10.47407/kr2025.6.11.00715.
29. Корабельников, Д. И. Анализ причин ошибочной классификации изображений немеланоцитарных опухолей кожи программами на основе технологий искусственного интеллекта / Д. И. Корабельников, А. И. Ламоткин // Сибирский журнал клинической и экспериментальной медицины. – 2026. – Т. 41. – № 1. – С. 221–231. – DOI: 10.29001/2073-8552-2026-2881.
30. Корабельников, Д. И. Дифференциальная диагностика новообразований кожи и дерматозов с применением программы Derma Check: результаты клинической апробации / Д. И. Корабельников, А. И. Ламоткин // Вестник последипломного медицинского образования. – 2026. – № 3. – С. 28–32. – DOI: 10.22363/2221-741X-2026-3-xx-xx.
31. Корабельников, Д. И. Искусственный интеллект в дерматовенерологии: сравнительный анализ применяемых программ компьютерного зрения на основе моделей машинного обучения / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2025. – Т. 18. – № 4. – С. 571–581. – DOI: 10.17749/2070-4909/farmakoeconomika.2025.340.
32. Корабельников, Д. И. Искусственный интеллект в здравоохранении: история развития, термины и понятия, классификации / Д. И. Корабельников, А. И. Ламоткин, И. А. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2026.372> (дата обращения: 05.07.2026). – Текст : электронный.
33. Корабельников, Д. И. Искусственный интеллект в здравоохранении: мировой опыт внедрения, правовое регулирование, проблемы и этические аспекты / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная

фармакоэкономика и фармакоэпидемиология. – 2025. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2025.328> (дата обращения: 05.06.2026). – Текст : электронный.

34. Корабельников, Д. И. Искусственный интеллект в онкологии: мировой опыт использования и перспективы развития / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2025. – Т. 18. – № 3. – С. 437–447. – DOI: 10.17749/2070-4909/farmakoeconomika.2025.302.

35. Корабельников, Д. И. Классификация применяемых в клинической медицине и здравоохранении профессиональных специализированных программ для ЭВМ, включая основанные на технологиях искусственного интеллекта / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2026.407> (дата обращения: 24.08.2026). – Текст : электронный.

36. Корабельников, Д. И. Методологические подходы к оценке точности и эффективности диагностики программами искусственного интеллекта на основе логики клинического мышления в дерматоонкологии / Д. И. Корабельников, А. И. Ламоткин, И. А. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2026.385> (дата обращения: 05.07.2026). – Текст : электронный.

37. Корабельников, Д. И. Модуль непрерывного профессионального развития врача «Derma Check Educational Module» в повышении компетенций в дерматоонкологии / Д. И. Корабельников, А. И. Ламоткин, И. А. Ламоткин // Вестник медицинского института непрерывного образования. – 2026. – Т. 6. – № 3. – С. 29–35. – DOI: 10.36107/2782-1714_2026-6-3-29-35.

38. Корабельников, Д. И. Отношение врачей-терапевтов к применению технологий искусственного интеллекта в клинической медицине (по результатам

опроса) / Д. И. Корабельников, А. И. Ламоткин // Клинический разбор в общей медицине. – 2026. – Т. 7. – № 5. – С. 138–144. – DOI: 10.47407/kr2026.7.5.00855.

39. Корабельников, Д. И. Проблемы диагностики злокачественных новообразований кожи с применением программ искусственного интеллекта и пути их решения на примере программы Derma Onko Melanoma Check / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2026.403> (дата обращения: 15.08.2026). – Текст : электронный.

40. Корабельников, Д. И. Проблемы дифференциальной диагностики доброкачественных и злокачественных новообразований кожи / Д. И. Корабельников, И. А. Ламоткин, А. И. Ламоткин // Клинический разбор в общей медицине. – 2026. – Т. 7. – № 9. – С. 91–98. – DOI: 10.47407/kr2026.7.9.00p4574.

41. Корабельников, Д. И. Программы компьютерного зрения в медицине: методология формирования наборов данных, классификация и особенности клинического применения / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2026.374> (дата обращения: 05.07.2026). – Текст : электронный.

42. Корабельников, Д. И. Результаты опроса врачей-специалистов, ординаторов и студентов медицинских вузов о применении технологий ИИ в медицине / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2026.402> (дата обращения: 21.08.2026). – Текст : электронный.

43. Корабельников, Д. И. Способ оценки медико-экономической эффективности диагностики с применением алгоритмов маршрутизации пациентов на основании заключений моделей искусственного интеллекта / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и

фармакоэпидемиология. – 2026. – Т. 19. – № 2. – С. 314–333. – DOI: 10.17749/2070-4909/farmakoeconomika.2026.358.

44. Корабельников, Д. И. Способ оценки эффективности диагностики с использованием алгоритмов на основании заключений моделей искусственного интеллекта / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – Т. 19. – № 1. – С. 79–91. – DOI: 10.17749/2070-4909/farmakoeconomika.2026.355.

45. Корабельников, Д. И. Способ повышения эффективности программ компьютерного зрения на основе моделей искусственного интеллекта с помощью модуля коррекции параметров фотоизображений / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – Т. 19. – № 1. – С. 158–167. – DOI: 10.17749/2070-4909/farmakoeconomika.2026.364.

46. Корабельников, Д. И. Способ расчета финансовых затрат на диагностику с применением моделей искусственного интеллекта случая заболевания и оценки ее медико-экономической эффективности на уровне региона / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – Т. 19. – № 2. – С. 369–380. – DOI: 10.17749/2070-4909/farmakoeconomika.2026.362.

47. Корабельников, Д. И. Сравнительный анализ медицинских изделий с искусственным интеллектом в государственных реестрах США, Китая, России / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2026.404> (дата обращения: 28.08.2026). – Текст : электронный.

48. Корабельников, Д. И. Эволюция семейства программ компьютерного зрения «Derma Check» / Д. И. Корабельников, А. И. Ламоткин // Менеджер здравоохранения. – 2026. – № 11. – С. 78–87. – DOI: 10.21045/1811-0185-2026-11-78-87.

49. Корабельников, Д. И. Эффективность применения искусственного интеллекта в клинической медицине / Д. И. Корабельников, А. И. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2025. – Т. 18. – № 1. – С. 114–124. – DOI: 10.17749/2070-4909/farmakoeconomika.2025.287.
50. Купеева, И. А. Интеллектуальные обучающие системы как элемент системы повышения качества медицинской помощи / И. А. Купеева, Р. А. Раводин, А. А. Ефремов // Менеджер здравоохранения. – 2015. – № 1. – С. 51–56.
51. Купеева, И. Информационно-диагностические и обучающие системы в дерматовенерологии / И. Купеева, Р. Раводин // Врач. – 2015. – № 8. – С. 85–88.
52. Курылев, А. А. Реальная клиническая практика при оценке технологий здравоохранения: состояние на 2022 год / А. А. Курылев, А. А. Журавков, А. С. Колбин // Реальная клиническая практика: данные и доказательства. – 2022. – Т. 2. – № 4. – С. 1–9. – DOI: 10.37489/2782-3784-mvrwd-21.
53. Ламоткин, А. И. Алгоритм маршрутизации пациентов при дифференциальной диагностике новообразований кожи с использованием программы искусственного интеллекта Melanoma Check / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин // Медицинский вестник ГВКГ им. Н.Н. Бурденко. – 2025. – № 4. – С. 6–13. – DOI: 10.53652/2782-1730-2025-6-4-06-13.
54. Ламоткин, А. И. Анализ причин ошибочной классификации меланоцитарных опухолей кожи программами компьютерного зрения на основе технологий искусственного интеллекта / А. И. Ламоткин, Д. И. Корабельников // Вестник медицинского института непрерывного образования. – 2026. – Т. 6. – № 3. – С. 102–110. – DOI: 10.36107/2782-1714_2026-6-3-102-110.
55. Ламоткин, А. И. Веб-приложение «Derma Check»: архитектура мультимодульной системы компьютерного зрения для диагностики новообразований и другой патологии кожи с модулями бинарной и многоклассовой классификации, анамнеза, объяснимого искусственного интеллекта, динамического наблюдения и кабинетом врача / А. И. Ламоткин, Д. И.

Корабельников // Менеджер здравоохранения. – 2026. – № 10. – С. 90–101. – DOI: 10.21045/1811-0185-2026-10-68-79.

56. Ламоткин, А. И. Диагностика меланомы кожи с помощью искусственного интеллекта с применением программы компьютерного зрения / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин // Клиническая медицина. – 2025. – Т. 103. – № 7. – С. 540–545. – DOI: 10.30629/0023-2149-2025-103-7-540-545.

57. Ламоткин, А. И. Искусственный интеллект в дерматовенерологии: обзор применяемых мобильных приложений на основе моделей искусственного интеллекта / А. И. Ламоткин, Д. И. Корабельников // Клинический разбор в общей медицине. – 2026. – Т. 7. – № 1. – С. 95–104. – DOI: 10.47407/kr2026.7.1.00761.

58. Ламоткин, А. И. Искусственный интеллект: основные термины и понятия, применение в здравоохранении и клинической медицине / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2024. – Т. 17. – № 3. – С. 409–415. – DOI: 10.17749/2070-4909/farmakoeconomika.2024.267.

59. Ламоткин, А. И. Методология анализа причин ошибочной классификации фотоизображений программами компьютерного зрения на основе искусственного интеллекта в медицинской визуализации / А. И. Ламоткин, Д. И. Корабельников // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2026. – URL: <https://doi.org/10.17749/2070-4909/farmakoeconomika.2026.382> (дата обращения: 05.07.2026). – Текст : электронный.

60. Ламоткин, А. И. Сверточные нейронные сети и визуальные трансформеры в диагностике опухолей кожи: сравнительный анализ эффективности моделей искусственного интеллекта в программах компьютерного зрения / А. И. Ламоткин, Д. И. Корабельников // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2025. – Т. 18. – № 3. – С. 365–375. – DOI: 10.17749/2070-4909/farmakoeconomika.2025.327.

61. Ламоткин, А. И. Сравнение диагностической эффективности моделей компьютерного зрения программы Derma Check при использовании дерматоскопических и клинических фотоизображений / А. И. Ламоткин, Д. И.

Корабельников // Вестник последипломного медицинского образования. – 2026. – № 3. – С. 35–44. – DOI: 10.22363/2221-74IX-2026-3-xx-xx.

62. Ламоткин, А. И. Сравнение точности классификации опухолей кожи с помощью программ искусственного интеллекта при использовании изображений, полученных с применением цифровых зеркальных фотокамер и камер смартфонов / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин // Медицинский вестник ГВКГ им. Н.Н. Бурденко. – 2025. – № 3 (21). – С. 36–41. – DOI: 10.53652/2782-1730-2025-6-3-36-41.

63. Ламоткин, И. А. Дерматология. Иллюстрированная энциклопедия. Дерматовенерология. Микология. Онкодерматология. Дерматоскопия / И. А. Ламоткин, А. И. Ламоткин. – Москва : Лаборатория знаний, 2025. – 752 с.

64. Ламоткин, И. А. Диагностика меланом по клиническим, дерматоскопическим и гистологическим признакам / И. А. Ламоткин, Д. И. Корабельников, А. И. Ламоткин // Медицинский вестник МВД. – 2024. – Т. 131. – № 4 (131). – С. 66–71. – DOI: 10.52341/20738080_2024_131_4_66.

65. Ламоткин, И. А. Онкодерматология: атлас : учебное пособие / И. А. Ламоткин. – 2-е изд., электрон. – Москва : Лаборатория знаний, 2021. – 878 с.

66. Ламоткин, И. А. Эффективность дифференциальной диагностики доброкачественных и злокачественных новообразований кожи с помощью программы «Derma Onko Check» (версия 2.0) / И. А. Ламоткин, Д. И. Корабельников, А. И. Ламоткин // Вестник медицинского института непрерывного образования. – 2026. – Т. 6. – № 3. – С. 46–54. – DOI: 10.36107/2782-1714_2026-6-3-46-54.

67. Лукичев, П. М. Экономика искусственного интеллекта: возможности и проблемы использования в здравоохранении / П. М. Лукичев, О. П. Чекмарев // Вопросы инновационной экономики. – 2022. – Т. 12. – № 2. – С. 1111–1130. – DOI: 10.18334/vines.12.2.114782.

68. Методические рекомендации по организации и проведению клинико-экономического анализа медицинских изделий / Л. С. Краснова, Е. С. Арькова, Е.

- А. Лучинин, М. А. Холонья-Волоскова. – Москва : ГБУЗ «НИИ организации здравоохранения и медицинского менеджмента ДЗМ», 2022. – 60 с.
69. Методы цифровой обработки медицинских оптических изображений для улучшения качества диагностики заболеваний: обзор / Т. М. Акаева, А. В. Каменский, Н. Бородина [и др.] // Компьютерная оптика. – 2025. – Т. 49. – № 4. – С. 593–623. – DOI: 10.18287/2412-6179-CO-1554.
70. Неретин, Е. Ю. Первичная ранняя диагностика меланомы кожи после индивидуального обучения врачей / Е. Ю. Неретин, К. С. Титов, Г. М. Запиров // Клиническая дерматология и венерология. – 2023. – Т. 22. – № 1. – С. 99–105. – DOI: 10.17116/klinderma20232201199.
71. О лечении предраковых поражений кожи с помощью наружных местных препаратов / И. А. Ламоткин, Е. В. Селезнева, А. Н. Хлебникова [и др.] // Военно-медицинский журнал. – 2021. – Т. 342. – № 4. – С. 76–77.
72. Об утверждении профессионального стандарта «Врач общей практики (семейный врач)» : Проект Приказа Министерства труда и социальной защиты РФ (подготовлен Минтрудом России 27.11.2018) // Справочно-правовая система «Гарант». – URL: <https://www.garant.ru/products/ipo/prime/doc/56675186/> (дата обращения: 18.06.2026). – Текст : электронный.
73. Онуфриенко, Ю. И. Изучение основных причин смертности населения РФ / Ю. И. Онуфриенко // Стратегия устойчивого развития регионов России. – 2016. – № 30. – С. 131–133.
74. Особенности преподавания искусственного интеллекта в рамках программ высшего медицинского образования / Т. Г. Авачева, О. А. Милованова, А. А. Кривушин, С. А. Прохина // Российский журнал персонализированной медицины. – 2025. – Т. 5. – № 2. – С. 154–165. – DOI: 10.18705/2782-3806-2025-5-2-154-165.
75. Патент 2853036 С1 Российская Федерация, МПК G06N 3/02. Способ дифференциальной диагностики доброкачественных и злокачественных новообразований кожи с применением модели искусственного интеллекта / Д. И. Корабельников, А. И. Ламоткин ; патентообладатель АНО ДПО «Московский

медико-социальный институт имени Ф.П. Гааза». – № 2853036 ; заявл. 04.11.2025 ; опубли. 17.12.2025.

76. Патент 2853090 С1 Российская Федерация, МПК G06N 3/02. Способ дифференциальной диагностики меланомы кожи и доброкачественных меланоцитарных и немеланоцитарных новообразований кожи с применением модели искусственного интеллекта / А. И. Ламоткин, Д. И. Корабельников ; патентообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза». – № 2853090 ; заявл. 06.11.2025 ; опубли. 18.12.2025.

77. Петрунин, Ю. Ю. Доверие к технологиям генеративного искусственного интеллекта как зеркало доверия к институтам / Ю. Ю. Петрунин, Н. З. Нуралиева // Государственное управление. Электронный вестник. – 2025. – № 113. – С. 22–30. – DOI: 10.55959/MSU2070-1381-113-2025-22-30.

78. Предварительная дифференциальная диагностика доброкачественных и злокачественных опухолей из эпидермальной ткани кожи с применением программы искусственного интеллекта «Derma Onko Check» / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин [и др.] // Современные проблемы здравоохранения и медицинской статистики. – 2025. – № 2. – С. 223–242. – DOI: 10.24412/2312-2935-2025-2-223-242.

79. Преимущества и барьеры внедрения искусственного интеллекта в медицину: результаты опроса 354 российских специалистов / Ю. С. Драпкина, Н. П. Макарова, А. П. Калинин, В. Ю. Смольникова // Медицинский оппонент. – 2025. – Т. 3. – № 31. – С. 10–16. – DOI: 10.64831/2619-0001-2025-3-10-16.

80. Раводин, Р. А. Оценка эффективности системы дистанционного обучения в первичной подготовке врачей-дерматовенерологов / Р. А. Раводин // Курский научно-практический вестник «Человек и его здоровье». – 2016. – № 4. – С. 49–53. – DOI: 10.21626/vestnik/2016-4/08.

81. Результаты эпидемиологического мониторинга заболеваемости, смертности, распространенности злокачественных новообразований на основе действующего государственного ракового регистра / Ж. В. Хайлова, А. Д. Каприн, С. А. Иванов

[и др.] // Менеджер здравоохранения. – 2024. – № 11. – С. 39–51. – DOI: 10.21045/1811-0185-2024-11-39-51.

82. Российская Федерация. Законы. О персональных данных : Федеральный закон № 152-ФЗ : [принят Государственной Думой 8 июля 2006 года : одобрен Советом Федерации 14 июля 2006 года]. – Текст : электронный // КонсультантПлюс : сайт. – URL: https://www.consultant.ru/document/cons_doc_LAW_61801/ (дата обращения: 05.06.2026). – Текст : электронный.

83. Российская Федерация. Законы. Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации : Федеральный закон № 258-ФЗ : [принят Государственной Думой 22 июля 2020 года : одобрен Советом Федерации 24 июля 2020 года]. – Текст : электронный // КонсультантПлюс : сайт. – URL: https://www.consultant.ru/document/cons_doc_LAW_358738/ (дата обращения: 05.06.2026). – Текст : электронный.

84. Российская Федерация. Правительство. Об утверждении Правил государственной регистрации медицинских изделий : Постановление Правительства РФ от 30 ноября 2024 г. № 1684. – Текст : электронный // КонсультантПлюс : сайт. – URL: https://www.consultant.ru/document/cons_doc_LAW_491966/ (дата обращения: 05.06.2026). – Текст : электронный.

85. Свидетельство о государственной регистрации базы данных № 2025624714 Российская Федерация. База данных новообразований кожи – база данных: № 2025624278: заявл. 16.10.2025: опубл. 27.10.2025 / Корабельников Д. И., Ламоткин А. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 11.

86. Свидетельство о государственной регистрации базы данных № 2025624747 Российская Федерация. База данных оригинальных и аугментированных фотоизображений новообразований кожи с характеристиками новообразований и пациентов – база данных: № 2025624361: заявл. 19.10.2025: опубл. 27.10.2025 / Ламоткин А. И., Корабельников Д. И. ; правообладатель АНО ДПО «Московский

медико-социальный институт имени Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 11.

87. Свидетельство о государственной регистрации программы для ЭВМ № 2026683871 Российская Федерация. Модуль профессионального развития знаний и компетенций врача в дерматологии и дерматоонкологии со справочниками и автоматизированным формированием интерактивных тестов и клинических задач «Derma Check Educational Module» – программа для ЭВМ: № 2026683616: заявл. 04.08.2026: опубл. 04.08.2026 / Ламоткин А. И., Корабельников Д. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федерация службы по интеллектуальной собственности.

88. Свидетельство о государственной регистрации программы для ЭВМ № 2025661515 Российская Федерация. Программа для автоматизированной визуальной идентификации злокачественных и доброкачественных опухолей кожи «Derma Onko Melanoma Check» – программа для ЭВМ: № 2025660546: заявл. 06.05.2025: опубл. 22.05.2025 / Ламоткин А. И., Ламоткин И. А., Корабельников Д. И. // Федерация службы по интеллектуальной собственности. – Бюллетень № 6.

89. Свидетельство о государственной регистрации программы для ЭВМ № 2026662061 Российская Федерация. Мульти модельная программа для автоматизированной визуальной идентификации и дифференциальной диагностики доброкачественных и злокачественных новообразований кожи «Диагностика злокачественных новообразований кожи» («Derma Onko Melanoma Check New») – программа для ЭВМ: № 2026661893: заявл. 25.04.2026: опубл. 25.04.2026 / Ламоткин А. И., Корабельников Д. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федерация службы по интеллектуальной собственности. – Бюллетень № 5.

90. Свидетельство о государственной регистрации программы для ЭВМ № 2026683792 Российская Федерация. Веб-приложение мульти модельной системы поддержки принятия врачебных решений для автоматизированной визуальной идентификации и дифференциальной диагностики патологии кожи с модулями анализа изображений, анамнеза, объяснимого искусственного интеллекта,

кабинетом врача и функцией динамического наблюдения образований «Derma Check Web» – программа для ЭВМ: № 2026683503: заявл. 04.08.2026: опубл. 04.08.2026 / Ламоткин А. И., Корабельников Д. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федерация службы по интеллектуальной собственности. – Бюллетень № 8.

91. Свидетельство о государственной регистрации программы для ЭВМ № 2026615027 Российская Федерация. Программа для автоматизированного расчета стоимости диагностики случая и оценки медико-экономической эффективности диагностики с применением моделей искусственного интеллекта (Health care efficiency calculator for AI-diagnostics algorithms) – программа для ЭВМ: № 2026614147: заявл. 20.02.2026: опубл. 20.02.2026 / Корабельников Д. И., Ламоткин А. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 2.

92. Свидетельство о государственной регистрации программы для ЭВМ № 2025661513 Российская Федерация. Программа для автоматизированной визуальной идентификации злокачественных и доброкачественных опухолей кожи «Derma Onko Check» ver. 2.0 – программа для ЭВМ: № 2025660544: заявл. 06.05.2025: опубл. 22.05.2025 / Ламоткин А. И., Ламоткин И. А., Корабельников Д. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 6.

93. Свидетельство о государственной регистрации программы для ЭВМ № 2026616231 Российская Федерация. Программа для автоматизированной визуальной идентификации и дифференциальной диагностики доброкачественных и злокачественных опухолей кожи с модулем нормализации параметров фотоизображений и алгоритмами маршрутизации пациентов "Derma Onko Melanoma Check Plus" – программа для ЭВМ: № 2026615441: заявл. 04.03.2026: опубл. 04.03.2026 / Корабельников Д. И., Ламоткин А. И. ; правообладатель АНО

ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 3.

94. Свидетельство о государственной регистрации программы для ЭВМ № 2025661514 Российская Федерация. Программа для автоматизированной визуальной идентификации меланомы кожи «Melanoma check» ver. 2.0 – программа для ЭВМ: № 2025660545: заявл. 06.05.2025: опубл. 22.05.2025 / Ламоткин А. И., Ламоткин И. А., Корабельников Д. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 6.

95. Свидетельство о государственной регистрации программы для ЭВМ № 2026613742 Российская Федерация. Программа для автоматизированной оценки эффективности моделей искусственного интеллекта и алгоритмов маршрутизации на основании заключений моделей искусственного интеллекта в диагностике заболеваний "CalcRDAI&RNDAI" – программа для ЭВМ: № 2026612771: заявл. 10.02.2026: опубл. 10.02.2026 / Корабельников Д. И., Ламоткин А. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 2.

96. Свидетельство о государственной регистрации программы для ЭВМ № 2024668566 Российская Федерация. Программа для визуальной идентификации злокачественных и доброкачественных опухолей кожи "Derma Onko Check" – программа для ЭВМ: № 2024668247: заявл. 08.08.2024: опубл. 08.08.2024 / Ламоткин А. И., Ламоткин И. А., Корабельников Д. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 11.

97. Свидетельство о государственной регистрации программы для ЭВМ № 2024668565 Российская Федерация. Программа для визуальной идентификации меланомы кожи "Melanoma check" – программа для ЭВМ: № 2024668246: заявл. 08.08.2024: опубл. 08.08.2024 / Ламоткин А. И., Ламоткин И. А., Корабельников Д. И. ; правообладатель АНО ДПО «Московский медико-социальный институт имени

Ф.П. Гааза» // Федеральная служба по интеллектуальной собственности. – Бюллетень № 8.

98. Современные методы неинвазивной диагностики и профилактики меланомы / А. А. Лаптей, К. С. Кильмаева, М. Ю. Герасименко [и др.] // Клиническая медицина. – 2024. – Т. 102. – № 11-12. – С. 812–823. – DOI: 10.30629/0023-2149-2024-102-11-12-812-823.

99. Состояние онкологической помощи населению России в 2023 году / Под ред. А. Д. Каприна, В. В. Старинского, А. О. Шахзадовой. – Москва : МНИОИ им. П. А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2024. – 262 с. : ил. – ISBN 978-5-85502-297-1.

100. Состояние онкологической помощи населению России в 2024 году / Под ред. А. Д. Каприна, В. В. Старинского, А. О. Шахзадовой. – Москва : МНИОИ им. П. А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2025. – 275 с. : ил. – ISBN 978-5-85502-309-1.

101. Состояние онкологической помощи населению России в 2025 году / Под ред. А. Д. Каприна [и др.]. – Москва : МНИОИ им. П. А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2026. – 275 с. : ил. – ISBN 978-5-85502-314-5.

102. Токарева, О. М. Концептуальная модель оценки экономической эффективности ИИ-решений в здравоохранении / О. М. Токарева // Экономика и управление. – 2025. – Т. 31. – № 6. – С. 804–815. – DOI: 10.35854/1998-1627-2025-6-804-815.

103. Точность предварительной диагностики доброкачественных и злокачественных меланоцитарных опухолей кожи с применением программы искусственного интеллекта для смартфона Derma Onko Check / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин [и др.] // Медицинский вестник ГВКГ им. Н.Н. Бурденко. – 2025. – № 2(20). – С. 39–48. – DOI: 10.53652/2782-1730-2025-6-2-39-48.

104. Точность предварительной диагностики злокачественных меланоцитарных опухолей кожи с помощью программы искусственного интеллекта / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин [и др.] // Медицинский вестник

ГВКГ им. Н.Н. Бурденко. – 2025. – № 1(19). – С. 42–51. – DOI: 10.53652/2782-1730-2025-6-1-42-51.

105. Улумбекова, Г. Э. Внедрение технологий искусственного интеллекта в медицинское образование: обзор подходов / Г. Э. Улумбекова, И. Ю. Худова, И. М. Еникеев // ОРГЗДРАВ: новости, мнения, обучение. Вестник ВШОУЗ. – 2026. – Т. 12. – № 1 (43). – С. 4–31. – DOI: 10.33029/2411-8621-2026-12-1-4-31.

106. Частота встречаемости различных стадий меланомы кожи / И. А. Ламоткин, Е. В. Мухина, О. Г. Капустина [и др.] // Актуальные вопросы дерматовенерологии : сборник научных трудов по материалам Всероссийской научно-практической конференции с международным участием, посвященной 80-летию кафедры дерматовенерологии КГМУ и 100-летию со дня рождения профессора В.А. Леонова (Курск, 11 мая 2018 года) / под ред. Л. В. Силиной, Т. П. Исаенко. – Курск : Курский государственный медицинский университет, 2018. – С. 91–95.

107. Эффективность обучения диагностике новообразований кожи с применением программного модуля на основе технологий искусственного интеллекта Derma Check Educational Module / Д.И. Корабельников, А.И. Ламоткин, В.В. Гладько [и др.] // Вестник медицинского института непрерывного образования. – 2026. – Т. 6, № 4. – С. 72–77. – DOI: 10.36107/2782-1714_2026-6-4-72-77.

108. Эффективность предварительной диагностики доброкачественных опухолей из сосудистой и фиброзной тканей кожи с применением программы искусственного интеллекта «Derma Onko Check» / А. И. Ламоткин, Д. И. Корабельников, И. А. Ламоткин, В. В. Гладько // Медицинский вестник МВД. – 2025. – Т. 137. – № 4. – С. 45–52. – DOI: 10.52341/20738080_2025_137_4_45.

109. Эффективность предварительной дифференциальной диагностики доброкачественных и злокачественных новообразований кожи с помощью программы искусственного интеллекта Derma Onko Check / А. И. Ламоткин, Д. И. Корабельников, О. Ю. Олисова, И. А. Ламоткин // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2025. – Т. 18. – № 2. – С. 261–270. – DOI: 10.17749/2070-4909/farmakoeconomika.2025.294.

110. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis / X. Liu, L. Faes, A. U. Kale [et al.] // *The Lancet Digital Health*. – 2019. – Vol. 1. – № 6. – P. e271–e297. – DOI: 10.1016/S2589-7500(19)30123-2.
111. A comparison of dermatologists' and primary care physicians' accuracy in diagnosing melanoma: a systematic review / S. C. Chen, D. M. Bravata, E. Weil, I. Olkin // *Archives of Dermatology*. – 2001. – Vol. 137. – № 12. – P. 1627–1634. – DOI: 10.1001/archderm.137.12.1627.
112. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task / T. J. Brinker, A. Hekler, A. H. Enk [et al.] // *European Journal of Cancer*. – 2019. – Vol. 111. – P. 148–154. – DOI: 10.1016/j.ejca.2019.02.005.
113. A methodology for developing dermatological datasets: lessons from retrospective data collection for AI-based applications / A. Pedro, P. Romero, S. Vidaurre [et al.] // *BMC Medical Research Methodology*. – 2025. – Vol. 25. – № 1. – P. 251. – DOI: 10.1186/s12874-025-02706-y.
114. A patient-centric dataset of images and metadata for identifying melanomas using clinical context / V. Rotemberg, N. Kurtansky, B. Betz-Stablein [et al.] // *Scientific Data*. – 2021. – Vol. 8. – № 1. – P. 34. – DOI: 10.1038/s41597-021-00815-z.
115. A practical guide to the implementation of AI in orthopaedic research, Part 6: how to evaluate the performance of AI research? / F. C. Oettl, A. Pareek, P. W. Winkler [et al.] // *Journal of Experimental Orthopaedics*. – 2024. – Vol. 11. – № 3. – P. e12039. – DOI: 10.1002/jeo2.12039.
116. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology / J. Scheetz, P. Rothschild, M. McGuinness [et al.] // *Scientific Reports*. – 2021. – Vol. 11. – № 1. – P. 5193. – DOI: 10.1038/s41598-021-84698-5.
117. A survey on deep learning in medical image analysis / G. Litjens, T. Kooi, B. E. Bejnordi [et al.] // *Medical Image Analysis*. – 2017. – Vol. 42. – P. 60–88. – DOI: 10.1016/j.media.2017.07.005.

118. A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis / M. P. Salinas, J. Sepúlveda, L. Hidalgo [et al.] // NPJ Digital Medicine. – 2024. – Vol. 7. – № 1. – P. 125. – DOI: 10.1038/s41746-024-01103-x.
119. A systematic review of the barriers to the implementation of artificial intelligence in healthcare / M. I. Ahmed, B. Spooner, J. Isherwood [et al.] // Cureus. – 2023. – Vol. 15. – № 10. – P. e46454. – DOI: 10.7759/cureus.46454.
120. Abohashish, S. M. M. Enhanced melanoma and non-melanoma skin cancer classification using a hybrid LSTM-CNN model / S. M. M. Abohashish, H. H. Amin, E. I. Elsedimy // Scientific Reports. – 2025. – Vol. 15. – № 1. – P. 24994. – DOI: 10.1038/s41598-025-08954-8.
121. Accuracy of Computer-Aided Diagnosis of Melanoma: A Meta-analysis / V. Dick, C. Sinz, M. Mittlböck, H. Kittler, P. Tschandl // JAMA Dermatology. – 2019. – Vol. 155. – № 11. – P. 1291–1299. – DOI: 10.1001/jamadermatol.2019.1375.
122. Actinic keratoses: Natural history and risk of malignant transformation in the Veterans Affairs Topical Tretinoin Chemoprevention Trial / V. D. Criscione, M. A. Weinstock, M. F. Naylor [et al.] // Cancer. – 2009. – Vol. 115. – № 11. – P. 2523–2530. – DOI: 10.1002/cncr.24284.
123. Adamson, A. S. Machine Learning and Health Care Disparities in Dermatology / A. S. Adamson, A. Smith // JAMA Dermatology. – 2018. – Vol. 154. – № 11. – P. 1247–1248. – DOI: 10.1001/jamadermatol.2018.2348.
124. Adamson, A. S. Melanoma Overdiagnosis Is Not the Key Problem-Reply / A. S. Adamson, E. A. Suarez, H. G. Welch // JAMA Dermatology. – 2022. – Vol. 158. – № 9. – P. 1083–1084. – DOI: 10.1001/jamadermatol.2022.2926.
125. Addressing overdiagnosis and overtreatment in cancer: a prescription for change / L. J. Esserman, I. M. Thompson, B. Reid [et al.] // The Lancet Oncology. – 2014. – Vol. 15. – № 6. – P. e234–e242. – DOI: 10.1016/S1470-2045(13)70598-9.
126. Advances in the noninvasive diagnosis of melanoma-40 years beyond the ABCDs / J. Burshtein, A. Witkowski, D. Zakria [et al.] // CA: A Cancer Journal for Clinicians. – 2026. – Vol. 76. – № 1. – P. e70065. – DOI: 10.3322/caac.70065.

127. Adversarial attacks on medical machine learning / S. G. Finlayson, J. D. Bowers, J. Ito [et al.] // *Science*. – 2019. – Vol. 363. – № 6433. – P. 1287–1289. – DOI: 10.1126/science.aaw4399.
128. AI for IMPACTS framework for evaluating the long-term real-world impacts of AI-powered clinician tools: systematic review and narrative synthesis / C. Jacob, N. Brasier, E. Laurenzi [et al.] // *Journal of Medical Internet Research*. – 2025. – Vol. 27. – P. e67485. – DOI: 10.2196/67485.
129. AI knowledge, attitudes, perceptions, and willingness to use AI among oncologists in China: a nationwide cross-sectional study / X. Xiong, M. Xiao, H. Liu [et al.] // *BMC Medical Education*. – 2026. – Vol. 26. – № 1. – P. 573. – DOI: 10.1186/s12909-026-08951-z.
130. AI-driven healthcare: Fairness in AI healthcare: A survey / S. V. Chinta, Z. Wang, A. Palikhe [et al.] // *PLOS Digital Health*. – 2025. – Vol. 4. – № 5. – P. e0000864. – DOI: 10.1371/journal.pdig.0000864.
131. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies / K. Freeman, J. Dinnes, N. Chuchu [et al.] // *BMJ*. – 2020. – Vol. 368. – P. m127. – DOI: 10.1136/bmj.m127.
132. An artificial intelligence based app for skin cancer detection evaluated in a population based setting / A. M. Smak Gregoor, T. E. Sangers, L. J. Bakker [et al.] // *npj Digital Medicine*. – 2023. – Vol. 6. – № 1. – P. 90. – DOI: 10.1038/s41746-023-00831-w.
133. An image is worth 16x16 words: transformers for image recognition at scale / A. Dosovitskiy, L. Beyer, A. Kolesnikov [et al.] // *Proceedings of the International Conference on Learning Representations (ICLR)*. – 2021. – Vol. 2021. – P. 2010.11929.
134. An, F. Medical Image Classification Algorithm Based on Visual Attention Mechanism-MCNN / F. An, X. Li, X. Ma // *Oxidative Medicine and Cellular Longevity*. – 2021. – Vol. 2021. – P. 6280690. – DOI: 10.1155/2021/6280690.
135. Analysis of the ISIC image datasets: Usage, benchmarks and recommendations / B. Cassidy, C. Kendrick, A. Brodzicki [et al.] // *Medical Image Analysis*. – 2022. – Vol. 75. – P. 102305. – DOI: 10.1016/j.media.2021.102305.

136. Applications of artificial intelligence and the challenges in health technology assessment: a scoping review and framework with a focus on economic dimensions / M. Ramezani, A. Bakhtiari, R. Daroudi [et al.] // *Health Economics Review*. – 2025. – Vol. 15. – № 1. – P. 46. – DOI: 10.1186/s13561-025-00645-4.
137. Applications of Artificial Intelligence in Cancer Diagnosis and Treatment / Y. Xie, Z. Wang, Z. Zeng [et al.] // *Cancer Medicine*. – 2026. – Vol. 15. – № 8. – P. e72143. – DOI: 10.1002/cam4.72143.
138. Applications of generative adversarial networks in the diagnosis, prognosis, and treatment of ophthalmic diseases / R. Doorly, J. Ong, E. Waisberg [et al.] // *Graefes Archive for Clinical and Experimental Ophthalmology*. – 2025. – Vol. 263. – № 8. – P. 2117–2134. – DOI: 10.1007/s00417-025-06830-9.
139. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis / U. J. Muehlematter, P. Daniore, K. N. Vokinger [et al.] // *The Lancet Digital Health*. – 2021. – Vol. 3. – № 3. – P. e195–e203. – DOI: 10.1016/S2589-7500(20)30292-2.
140. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians / O. Asan, A. E. Bayrak, A. Choudhury // *Journal of Medical Internet Research*. – 2020. – Vol. 22. – № 6. – P. e15154. – DOI: 10.2196/15154.
141. Artificial Intelligence and Its Effect on Dermatologists' Accuracy in Dermoscopic Melanoma Image Classification: Web-Based Survey Study / R. C. Maron, J. S. Utikal, A. Hekler [et al.] // *Journal of Medical Internet Research*. – 2020. – Vol. 22. – № 9. – P. e18091. – DOI: 10.2196/18091.
142. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review / O. T. Jones, R. N. Matin, M. van der Schaar [et al.] // *The Lancet Digital Health*. – 2022. – Vol. 4. – № 6. – P. e466–e476. – DOI: 10.1016/S2589-7500(22)00023-1.
143. Artificial Intelligence Applications in Dermatology: Where Do We Stand? / A. Gomolin, E. Netchiporouk, R. Gniadecki, I. V. Litvinov // *Frontiers in Medicine*. – 2020. – Vol. 7. – P. 100. – DOI: 10.3389/fmed.2020.00100.

144. Artificial intelligence in dermatology: challenges and perspectives / K. Liopyris, S. Gregoriou, J. Dias, A. J. Stratigos // *Dermatology and Therapy*. – 2022. – Vol. 12. – № 12. – P. 2637–2651. – DOI: 10.1007/s13555-022-00833-8.
145. Artificial intelligence in dermatology: current applications and future innovations // *Qualitative Research Journal for Social Studies*. – 2025. – Vol. 2. – № 3. – P. 1335–1344. – DOI: 10.63878/qrjs444.
146. Artificial Intelligence in Medical Education: Transformative Potential, Current Applications, and Future Implications / J. S. Izquierdo-Condoy, M. Arias-Intriago, L. Montero Corrales, E. Ortiz-Prado // *JMIR Medical Education*. – 2026. – Vol. 12. – P. e77127. – DOI: 10.2196/77127.
147. Artificial intelligence in mobile health for skin cancer diagnostics at home (AIM HIGH): a pilot feasibility study / A. M. Smak Gregoor, T. E. Sangers, J. A. Eekhof [et al.] // *EClinicalMedicine*. – 2023. – Vol. 60. – P. 102019. – DOI: 10.1016/j.eclinm.2023.102019.
148. Artificial intelligence in oncology: current landscape, challenges, and future directions / W. Lotter, M. J. Hassett, N. Schultz [et al.] // *Cancer Discovery*. – 2024. – Vol. 14. – № 5. – P. 711–726. – DOI: 10.1158/2159-8290.CD-23-1199.
149. Artificial Intelligence in Skin Cancer Diagnostics: The Patients' Perspective / T. B. Jutzi, E. I. Krieghoff-Henning, T. Holland-Letz [et al.] // *Frontiers in Medicine*. – 2020. – Vol. 7. – P. 233. – DOI: 10.3389/fmed.2020.00233.
150. Artificial intelligence in the detection of skin cancer / E. J. Beltrami, A. C. Brown, P. J. M. Salmon [et al.] // *Journal of the American Academy of Dermatology*. – 2022. – Vol. 87. – № 6. – P. 1336–1342. – DOI: 10.1016/j.jaad.2022.08.028.
151. Artificial Intelligence in Undergraduate Medical Education: A Scoping Review / J. Lee, A. S. Wu, D. Li, K. M. Kulasegaram // *Academic Medicine*. – 2021. – Vol. 96. – № 11S. – P. S62–S70. – DOI: 10.1097/ACM.0000000000004291.
152. Artificial intelligence in undergraduate medical education: an updated scoping review / J. Simoni, J. Urtubia-Fernandez, E. Mengual [et al.] // *BMC Medical Education*. – 2025. – Vol. 25. – № 1. – P. 1609. – DOI: 10.1186/s12909-025-08188-2.

153. Artificial Intelligence Revolutionizing the Field of Medical Education / S. Narayanan, R. Ramakrishnan, E. Durairaj, A. Das // *Cureus*. – 2023. – Vol. 15. – № 11. – P. e49604. – DOI: 10.7759/cureus.49604.
154. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies / M. Nagendran, Y. Chen, C. A. Lovejoy [et al.] // *BMJ*. – 2020. – Vol. 368. – P. m689. – DOI: 10.1136/bmj.m689.
155. Assessing the performance of prediction models: a framework for traditional and novel measures / E. W. Steyerberg, A. J. Vickers, N. R. Cook [et al.] // *Epidemiology*. – 2010. – Vol. 21. – № 1. – P. 128–138. – DOI: 10.1097/EDE.0b013e3181c30fb2.
156. Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition / J. K. Winkler, C. Fink, F. Toberer [et al.] // *JAMA Dermatology*. – 2019. – Vol. 155. – № 10. – P. 1135–1141. – DOI: 10.1001/jamadermatol.2019.1735.
157. Attitudes Toward Artificial Intelligence Within Dermatopathology: An International Online Survey / S. Polesie, P. H. McKee, J. M. Gardner [et al.] // *Frontiers in Medicine*. – 2020. – Vol. 7. – P. 591952. – DOI: 10.3389/fmed.2020.591952.
158. Attitudes towards artificial intelligence within dermatology: an international online survey / S. Polesie, M. Gillstedt, H. Kittler [et al.] // *British Journal of Dermatology*. – 2020. – Vol. 183. – № 1. – P. 159–161. – DOI: 10.1111/bjd.18875.
159. Attitudes towards the use of artificial intelligence in dermatology: a survey of Australian dermatologists / B. Partridge, M. Janda, N. Gillespie [et al.] // *Australasian Journal of Dermatology*. – 2025. – Vol. 66. – № 5. – P. e279–e286. – DOI: 10.1111/ajd.14524.
160. Augmented Intelligence Dermatology: Deep Neural Networks Empower Medical Professionals in Diagnosing Skin Cancer and Predicting Treatment Options for 134 Skin Disorders / S. S. Han, I. Park, S. Eun Chang [et al.] // *The Journal of Investigative Dermatology*. – 2020. – Vol. 140. – № 9. – P. 1753–1761. – DOI: 10.1016/j.jid.2020.01.019.

161. Automated Classification of Skin Lesions: From Pixels to Practice / A. Narla, B. Kuprel, K. Sarin [et al.] // *The Journal of Investigative Dermatology*. – 2018. – Vol. 138. – № 10. – P. 2108–2110. – DOI: 10.1016/j.jid.2018.06.175.
162. Automated magnetic resonance image segmentation of the anterior cruciate ligament / S. W. Flannery, A. M. Kiapour, D. J. Edgar [et al.] // *Journal of Orthopaedic Research*. – 2021. – Vol. 39. – № 4. – P. 831–840. – DOI: 10.1002/jor.24926.
163. Barata, C. A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer / C. Barata, M. E. Celebi, J. S. Marques // *IEEE Journal of Biomedical and Health Informatics*. – 2019. – Vol. 23. – № 3. – P. 1096–1109. – DOI: 10.1109/JBHI.2018.2845939.
164. Basal cell carcinoma: Epidemiology; pathophysiology; clinical and histological subtypes; and disease associations / M. C. Cameron, E. Lee, B. P. Hibler [et al.] // *Journal of the American Academy of Dermatology*. – 2019. – Vol. 80. – № 2. – P. 303–317. – DOI: 10.1016/j.jaad.2018.03.060.
165. BCN20000: dermoscopic lesions in the wild / M. Combalia, N. C. F. Codella, V. Rotemberg [et al.] // *arXiv*. – 2019. – Vol. 2019. – P. 1908.02288.
166. Behara, K. AI in dermatology: a comprehensive review into skin cancer detection / K. Behara, E. Bhero, J. T. Agee // *PeerJ Computer Science*. – 2024. – Vol. 10. – P. e2530. – DOI: 10.7717/peerj-cs.2530.
167. Beyond Post hoc Explanations: A Comprehensive Framework for Accountable AI in Medical Imaging Through Transparency, Interpretability, and Explainability / Y. Singh, Q. A. Hathaway, V. Keishing, [et al.] // *Bioengineering*. – 2025. – Vol. 12. – № 8. – P. 879. – DOI: 10.3390/bioengineering12080879.
168. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden / L. Petersson, I. Larsson, J. M. Nygren [et al.] // *BMC Health Services Research*. – 2022. – Vol. 22. – № 1. – P. 850. – DOI: 10.1186/s12913-022-08215-8.
169. Char, D. S. Implementing Machine Learning in Health Care – Addressing Ethical Challenges / D. S. Char, N. H. Shah, D. Magnus // *The New England Journal of Medicine*. – 2018. – Vol. 378. – № 11. – P. 981–983. – DOI: 10.1056/NEJMp1714229.

170. Characteristics of publicly available skin cancer image datasets: a systematic review / D. Wen, S. M. Khan, A. Ji Xu [et al.] // *The Lancet Digital Health*. – 2022. – Vol. 4. – № 1. – P. e64–e74. – DOI: 10.1016/S2589-7500(21)00252-1.
171. Chen, Y. A. Analysis of dermoscopy teaching modalities in United States dermatology residency programs / Y. A. Chen, J. Rill, E. V. Seiverling // *Dermatology Practical & Conceptual*. – 2017. – Vol. 7. – № 3. – P. 38–43. – DOI: 10.5826/dpc.070308.
172. Cheng, C. H. Artificial intelligence in cancer: applications, challenges, and future perspectives / C. H. Cheng, S. S. Shi // *Molecular Cancer*. – 2025. – Vol. 24. – № 1. – P. 274. – DOI: 10.1186/s12943-025-02450-3.
173. Combating COVID-19 using object detection techniques for next-generation autonomous systems / H. Shenai, J. Gala, K. Kekre [et al.] // *Cyber-Physical Systems*. – 2022. – Vol. 2022. – P. 55–73. – DOI: 10.1016/B978-0-12-824557-6.00007-8.
174. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study / P. Tschandl, N. Codella, B. N. Akay [et al.] // *The Lancet Oncology*. – 2019. – Vol. 20. – № 7. – P. 938–947. – DOI: 10.1016/S1470-2045(19)30333-X.
175. Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults / L. Ferrante di Ruffano, Y. Takwoingi, J. Dinnes [et al.] // *The Cochrane Database of Systematic Reviews*. – 2018. – Vol. 12. – № 12. – P. CD013186. – DOI: 10.1002/14651858.CD013186.
176. Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks / A. Esteva, B. Kuprel, R. A. Novoa [et al.] // *Nature*. – 2017. – Vol. 546. – № 7660. – P. 686. – DOI: 10.1038/nature22985
177. Cost-effectiveness of artificial intelligence aided vessel occlusion detection in acute stroke: an early health technology assessment / K. G. van Leeuwen, F. J. A. Meijer, S. Schalekamp [et al.] // *Insights into Imaging*. – 2021. – Vol. 12. – № 1. – P. 133. – DOI: 10.1186/s13244-021-01077-4.
178. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study / M. Areia, Y. Mori, L. Correale [et al.] // *The Lancet Digital Health*. – 2022. – Vol. 4. – № 6. – P. e436–e444. – DOI: 10.1016/S2589-7500(22)00042-5.

179. Cost-Effectiveness of Screening for Primary Aldosteronism and Subtype Diagnosis in the Resistant Hypertensive Patients / C. C. Lubitz, K. P. Economopoulos, S. Sy [et al.] // *Circulation: Cardiovascular Quality and Outcomes*. – 2015. – Vol. 8. – № 6. – P. 621–630. – DOI: 10.1161/CIRCOUTCOMES.115.002002.
180. Current approach of skin lesions suspected of malignancy in general practice in the Netherlands: a quantitative overview / K. Ahmadi, E. Prickaerts, J. G. E. Smeets [et al.] // *Journal of the European Academy of Dermatology and Venereology*. – 2018. – Vol. 32. – № 2. – P. 236–241. – DOI: 10.1111/jdv.14484.
181. De-constructing bias on skin lesion datasets / A. Bissoto, M. Fornaciali, E. Valle, S. Avila // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. – 2019. – Vol. 2019. – P. 2766–2774. – DOI: 10.1109/CVPRW.2019.00335.
182. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images / A. Hekler, J. S. Utikal, A. H. Enk [et al.] // *European Journal of Cancer*. – 2019. – Vol. 118. – P. 91–96. – DOI: 10.1016/j.ejca.2019.06.012.
183. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task / T. J. Brinker, A. Hekler, A. H. Enk [et al.] // *European Journal of Cancer*. – 2019. – Vol. 113. – P. 47–54. – DOI: 10.1016/j.ejca.2019.04.001.
184. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2016. – Vol. 2016. – P. 770–778. – DOI: 10.1109/CVPR.2016.90.
185. Definitions, methods, and applications in interpretable machine learning / W. J. Murdoch, C. Singh, K. Kumbier [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2019. – Vol. 116. – № 44. – P. 22071–22080. – DOI: 10.1073/pnas.1900654116.
186. Densely connected convolutional networks / G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2017. – Vol. 2017. – P. 4700–4708. – DOI: 10.1109/CVPR.2017.243.

187. Dermatologist-like explainable AI enhances trust and confidence in diagnosing melanoma / T. Chanda, K. Hauser, S. Hobelsberger [et al.] // *Nature Communications*. – 2024. – Vol. 15. – № 1. – P. 524. – DOI: 10.1038/s41467-023-43095-4.
188. Dermatoscopy of facial actinic keratosis, intraepidermal carcinoma, and invasive squamous cell carcinoma: a progression model / I. Zalaudek, J. Giacomel, K. Schmid [et al.] // *Journal of the American Academy of Dermatology*. – 2012. – Vol. 66. – № 4. – P. 589–597. – DOI: 10.1016/j.jaad.2011.02.011.
189. Dermoscopic features of basal cell carcinoma and its subtypes: A systematic review / O. Reiter, I. Mimouni, S. Dusza [et al.] // *Journal of the American Academy of Dermatology*. – 2021. – Vol. 85. – № 3. – P. 653–664. – DOI: 10.1016/j.jaad.2019.11.008.
190. Dermoscopy Education in Undergraduate Medical Education: Scoping Review of Pedagogical Approaches and Gaps / S. Moussa, A. S. Aldien, S. A. E. Aboushawareb [et al.] // *Journal of Medical Education and Curricular Development*. – 2025. – Vol. 12. – P. 23821205251385824. – DOI: 10.1177/23821205251385824.
191. Dermoscopy for the diagnosis of melanoma: primary care diagnostic technology update / A. Plüddemann, C. Heneghan, M. Thompson [et al.] // *British Journal of General Practice*. – 2011. – Vol. 61. – № 587. – P. 416–417. – DOI: 10.3399/bjgp11X578142.
192. Dermoscopy in basal cell carcinoma: an updated review / M. Álvarez-Salafranca, M. Ara, P. Zaballos // *Actas Dermo-Sifiliográficas*. – 2021. – Vol. 112. – № 4. – P. 330–338. – DOI: 10.1016/j.ad.2020.11.011.
193. Dermoscopy in China: current status and future prospective / X. Shen, R. X. Yu, C. B. Shen [et al.] // *Chinese Medical Journal*. – 2019. – Vol. 132. – № 17. – P. 2096–2104. – DOI: 10.1097/CM9.0000000000000396.
194. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet / G. Argenziano, H. P. Soyer, S. Chimenti [et al.] // *Journal of the American Academy of Dermatology*. – 2003. – Vol. 48. – № 5. – P. 679–693. – DOI: 10.1067/mjd.2003.281.

195. Diagnosis and treatment of basal cell carcinoma: European consensus-based interdisciplinary guidelines / K. Peris, M. C. Fargnoli, C. Garbe [et al.] // *European Journal of Cancer*. – 2019. – Vol. 118. – P. 10–34. – DOI: 10.1016/j.ejca.2019.06.003.
196. Diagnostic accuracy of artificial intelligence compared to family physicians and dermatologists for skin conditions: a systematic review and meta-analysis / N. Nadour, T. Duguet, S. Zahedi [et al.] // *BMC Primary Care*. – 2025. – Vol. 26. – № 1. – P. 384. – DOI: 10.1186/s12875-025-03073-9.
197. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis / R. Aggarwal, V. Sounderajah, G. Martin [et al.] // *npj Digital Medicine*. – 2021. – Vol. 4. – № 1. – P. 65. – DOI: 10.1038/s41746-021-00438-z.
198. Diagnostic accuracy of dermoscopy for 934 basal cell carcinomas: a single-center retrospective study / A. Yuki, S. Takatsuka, R. Abe, T. Takenouchi // *The Journal of Dermatology*. – 2023. – Vol. 50. – № 1. – P. 64–71. – DOI: 10.1111/1346-8138.16607.
199. Disparities in dermatology AI performance on a diverse, curated clinical image set / R. Daneshjou, K. Vodrahalli, R. A. Novoa [et al.] // *Science Advances*. – 2022. – Vol. 8. – № 32. – P. eabq6147. – DOI: 10.1126/sciadv.abq6147.
200. Early detection of metastatic risk in primary cutaneous melanoma using weakly supervised learning / F. Dahlén, I. Shujski, F. Yacob [et al.] // *Scientific Reports*. – 2026. – Vol. 16. – № 1. – P. 11234. – DOI: 10.1038/s41598-026-45588-w.
201. Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria / N. R. Abbasi, H. M. Shaw, D. S. Rigel [et al.] // *JAMA*. – 2004. – Vol. 292. – № 22. – P. 2771–2776. – DOI: 10.1001/jama.292.22.2771.
202. Economic evaluation of AI-assisted technologies in healthcare: A systematic review / W. T. Wu, Y. W. Chao, T. K. Lin [et al.] // *Journal of Food and Drug Analysis*. – 2025. – Vol. 33. – № 4. – P. 487–500. – DOI: 10.38212/2224-6614.3570.
203. Economic Evaluations and Equity in the Use of Artificial Intelligence in Imaging Exams for Medical Diagnosis in People With Skin, Neurological, and Pulmonary Diseases: Protocol for a Systematic Review / G. O. Santana, R. M. Couto, R. M. Loureiro [et al.] // *JMIR Research Protocols*. – 2023. – Vol. 12. – P. e48544. – DOI: 10.2196/48544.

204. Economics of Artificial Intelligence in Healthcare: Diagnosis vs. Treatment / N. N. Khanna, M. A. Maindarkar, V. Viswanathan [et al.] // Healthcare. – 2022. – Vol. 10. – № 12. – P. 2493. – DOI: 10.3390/healthcare10122493.
205. El Arab, R. A. Systematic review of cost effectiveness and budget impact of artificial intelligence in healthcare / R. A. El Arab, O. A. Al Moosa // NPJ Digital Medicine. – 2025. – Vol. 8. – P. 548. – DOI: 10.1038/s41746-025-01722-y.
206. Ennab, M. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions / M. Ennab, H. Mcheick // Frontiers in Robotics and AI. – 2024. – Vol. 11. – P. 1444763. – DOI: 10.3389/frobt.2024.1444763.
207. Epidemiological trends in skin cancer / Z. Apalla, A. Lallas, E. Sotiriou [et al.] // Dermatology Practical & Conceptual. – 2017. – Vol. 7. – № 2. – P. 1–6. – DOI: 10.5826/dpc.0702a01.
208. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermoscopy and a new 7-point checklist based on pattern analysis / G. Argenziano, G. Fabbrocini, P. Carli [et al.] // Archives of Dermatology. – 1998. – Vol. 134. – № 12. – P. 1563–1570. – DOI: 10.1001/archderm.134.12.1563.
209. Erdat, E. C. Turkish medical oncologists' perspectives on integrating artificial intelligence: knowledge, attitudes, and ethical considerations / E. C. Erdat, Ş. Çay Şenler // BMC Medical Ethics. – 2025. – Vol. 26. – № 1. – P. 82. – DOI: 10.1186/s12910-025-01249-7.
210. Errichetti, E. Dermoscopy in General Dermatology: A Practical Overview / E. Errichetti, G. Stinco // Dermatology and Therapy. – 2016. – Vol. 6. – № 4. – P. 471–507. – DOI: 10.1007/s13555-016-0141-6.
211. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review / C. Mennella, U. Maniscalco, G. De Pietro, M. Esposito // Heliyon. – 2024. – Vol. 10. – № 4. – P. e26297. – DOI: 10.1016/j.heliyon.2024.e26297.
212. Ethical challenges in the algorithmic era: a systematic rapid review of risk insights and governance pathways for nursing predictive analytics and early warning systems / Y.

- Cao, L. Deng, X. Liu [et al.] // BMC Medical Ethics. – 2025. – Vol. 26. – № 1. – P. 151. – DOI: 10.1186/s12910-025-01308-z.
213. Ethical limitations of algorithmic fairness solutions in health care machine learning / M. D. McCradden, S. Joshi, M. Mazwi, J. A. Anderson // The Lancet Digital Health. – 2020. – Vol. 2. – № 5. – P. e221–e223. – DOI: 10.1016/S2589-7500(20)30065-0.
214. Ethics and governance of artificial intelligence for health : WHO guidance / World Health Organization. – Geneva : World Health Organization, 2021. – 148 p. – ISBN 978-92-4-002920-0.
215. European consensus-based interdisciplinary guideline for melanoma. Part 1: Diagnostics: Update 2022 / C. Garbe, T. Amaral, K. Peris [et al.] // European Journal of Cancer. – 2022. – Vol. 170. – P. 236–255. – DOI: 10.1016/j.ejca.2022.03.008.
216. Evaluating artificial intelligence in clinical settings – let us not reinvent the wheel / K. Cresswell, N. de Keizer, F. Magrabi [et al.] // Journal of Medical Internet Research. – 2024. – Vol. 26. – P. e46407. – DOI: 10.2196/46407.
217. Evaluation framework to guide implementation of AI systems into healthcare settings / S. Reddy, W. Rogers, V. P. Makinen [et al.] // BMJ Health & Care Informatics. – 2021. – Vol. 28. – № 1. – P. e100444. – DOI: 10.1136/bmjhci-2021-100444.
218. Evaluation of the Clinical, Technical, and Financial Aspects of Cost-Effectiveness Analysis of Artificial Intelligence in Medicine: Scoping Review and Framework of Analysis / J. Gomez Rossi, B. Feldberg, J. Krois, F. Schwendicke // JMIR Medical Informatics. – 2022. – Vol. 10. – № 8. – P. e33703. – DOI: 10.2196/33703.
219. Evans, H. Understanding the errors made by artificial intelligence algorithms in histopathology in terms of patient impact / H. Evans, D. Snead // NPJ Digital Medicine. – 2024. – Vol. 7. – № 1. – P. 89. – DOI: 10.1038/s41746-024-01093-w.
220. Expectations for Artificial Intelligence (AI) in Psychiatry / S. Monteith, T. Glenn, J. Geddes [et al.] // Current Psychiatry Reports. – 2022. – Vol. 24. – № 11. – P. 709–721. – DOI: 10.1007/s11920-022-01378-5.
221. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks / P. Tschandl, C. Rosendahl, B. N. Akay [et al.] // JAMA Dermatology. – 2019. – Vol. 155. – № 1. – P. 58–65. – DOI: 10.1001/jamadermatol.2018.4378.

222. Explainability agreement between dermatologists and five visual explanations techniques in deep neural networks for melanoma AI classification / M. Giavina-Bianchi, W. G. Vitor, V. Fornasiero de Paiva [et al.] // *Frontiers in Medicine*. – 2023. – Vol. 10. – P. 1241484. – DOI: 10.3389/fmed.2023.1241484.
223. Explainable artificial intelligence (XAI) in medical imaging: a systematic review of techniques, applications, and challenges / F. Ahmed, N. S. Naz, S. Khan [et al.] // *BMC Medical Imaging*. – 2026. – Vol. 26. – № 1. – P. 37. – DOI: 10.1186/s12880-025-02118-w.
224. Explainable artificial intelligence in skin cancer recognition: a systematic review / K. Hauser, A. Kurz, S. Haggenmüller [et al.] // *European Journal of Cancer*. – 2022. – Vol. 167. – P. 54–69. – DOI: 10.1016/j.ejca.2022.02.025.
225. Exploring the potential of artificial intelligence in improving skin lesion diagnosis in primary care / A. Escalé-Besa, O. Yélamos, J. Vidal-Alaball [et al.] // *Scientific Reports*. – 2023. – Vol. 13. – № 1. – P. 4293. – DOI: 10.1038/s41598-023-31340-1.
226. Exposure to Trace Elements and Risk of Skin Cancer: A Systematic Review of Epidemiologic Studies / N. H. Matthews, K. Fitch, W. Q. Li [et al.] // *Cancer Epidemiology, Biomarkers & Prevention*. – 2019. – Vol. 28. – № 1. – P. 3–21. – DOI: 10.1158/1055-9965.EPI-18-0286.
227. Final version of 2009 AJCC melanoma staging and classification / C. M. Balch, J. E. Gershenwald, S. J. Soong [et al.] // *Journal of Clinical Oncology*. – 2009. – Vol. 27. – № 36. – P. 6199–6206. – DOI: 10.1200/JCO.2009.23.4799.
228. Foundation Models Meet Medical Image Interpretation / L. Jiao, J. Hao, R. Li [et al.] // *Research*. – 2026. – Vol. 9. – P. 1024. – DOI: 10.34133/research.1024.
229. From data to diagnosis: skin cancer image datasets for artificial intelligence / D. Wen, A. Soltan, E. Trucco, R. N. Matin // *Clinical and Experimental Dermatology*. – 2024. – Vol. 49. – № 7. – P. 675–685. – DOI: 10.1093/ced/llae112.
230. Fu, S. Dermatological guide for primary care physicians: full body skin checks, skin cancer detection, and patient education on self-skin checks and sun protection / S. Fu, S. Kim, C. Wasko // *Proceedings (Baylor University. Medical Center)*. – 2024. – Vol. 37. – № 4. – P. 647–654. – DOI: 10.1080/08998280.2024.2351751.

231. Generation and Evaluation of Synthetic Melanoma Images Using Generative Adversarial Networks / P. Y. Lin, Y. Shen, N. Mathew [et al.] // *Bioengineering*. – 2026. – Vol. 13. – № 2. – P. 245. – DOI: 10.3390/bioengineering13020245.
232. Gerke, S. Ethical and legal challenges of artificial intelligence-driven healthcare / S. Gerke, T. Minssen, G. Cohen // *Artificial Intelligence in Healthcare*. – 2020. – Vol. 260. – P. 295–336. – DOI: 10.1016/B978-0-12-818438-7.00012-5.
233. Ghassemi, M. The false hope of current approaches to explainable artificial intelligence in health care / M. Ghassemi, L. Oakden-Rayner, A. L. Beam // *The Lancet Digital Health*. – 2021. – Vol. 3. – № 11. – P. e745–e750. – DOI: 10.1016/S2589-7500(21)00208-9.
234. Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040 / M. Arnold, D. Singh, M. Laversanne [et al.] // *JAMA Dermatology*. – 2022. – Vol. 158. – № 5. – P. 495–503. – DOI: 10.1001/jamadermatol.2022.0160.
235. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries / H. Sung, J. Ferlay, R. L. Siegel [et al.] // *CA: A Cancer Journal for Clinicians*. – 2021. – Vol. 71. – № 3. – P. 209–249. – DOI: 10.3322/caac.21660.
236. Gordon, L. G. Health system costs of skin cancer and cost-effectiveness of skin cancer prevention and screening: a systematic review / L. G. Gordon, D. Rowell // *European Journal of Cancer Prevention*. – 2015. – Vol. 24. – № 2. – P. 141–149. – DOI: 10.1097/CEJ.0000000000000056.
237. Grad-CAM: visual explanations from deep networks via gradient-based localization / R. R. Selvaraju, M. Cogswell, A. Das, [et al.] // *International Journal of Computer Vision*. – 2020. – Vol. 128. – № 2. – P. 336–359. – DOI: 10.1007/s11263-019-01228-7.
238. Gu, C. Deep Transfer Learning Using Real-World Image Features for Medical Image Classification, with a Case Study on Pneumonia X-ray Images / C. Gu, M. Lee // *Bioengineering*. – 2024. – Vol. 11. – № 4. – P. 406. – DOI: 10.3390/bioengineering11040406.

239. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension / S. Cruz Rivera, X. Liu, A. W. Chan [et al.] // *Nature Medicine*. – 2020. – Vol. 26. – № 9. – P. 1351–1363. – DOI: 10.1038/s41591-020-1037-7.
240. Guidelines of care for the management of basal cell carcinoma / J. Y. S. Kim, J. H. Kozlow, B. Mittal [et al.] // *Journal of the American Academy of Dermatology*. – 2018. – Vol. 78. – № 3. – P. 540–559. – DOI: 10.1016/j.jaad.2017.10.006.
241. Haddadin, S. The Use of Artificial Intelligence to Detect Malignant Skin Lesions / S. Haddadin, L. Ganti // *Mayo Clinic Proceedings: Digital Health*. – 2024. – Vol. 2. – № 2. – P. 241–245. – DOI: 10.1016/j.mcpdig.2024.04.003.
242. Harrison, K. The Accuracy of Skin Cancer Detection Rates with the Implementation of Dermoscopy Among Dermatology Clinicians: A Scoping Review / K. Harrison // *The Journal of Clinical and Aesthetic Dermatology*. – 2024. – Vol. 17. – № 9-10 Suppl 1. – P. S18–S27.
243. Haykal, D. Keep an Eye on AI: How Artificial Intelligence Is Redefining Patient Education in Dermatology / D. Haykal // *Journal of Cosmetic Dermatology*. – 2025. – Vol. 24. – № 5. – P. e70225. – DOI: 10.1111/jocd.70225.
244. Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging / L. Oakden-Rayner, J. Dunnmon, G. Carneiro, C. Ré // *Proceedings of the ACM Conference on Health, Inference, and Learning*. – 2020. – Vol. 2020. – P. 151–159. – DOI: 10.1145/3368555.3384468.
245. Hinton, G. E. A fast learning algorithm for deep belief nets / G. E. Hinton, S. Osindero, Y. W. Teh // *Neural Computation*. – 2006. – Vol. 18. – № 7. – P. 1527–1554. – DOI: 10.1162/neco.2006.18.7.1527.
246. Human-computer collaboration for skin cancer recognition / P. Tschandl, C. Rinner, Z. Apalla [et al.] // *Nature Medicine*. – 2020. – Vol. 26. – № 8. – P. 1229–1234. – DOI: 10.1038/s41591-020-0942-0.
247. ImageNet large scale visual recognition challenge / O. Russakovsky, J. Deng, H. Su [et al.] // *International Journal of Computer Vision*. – 2015. – Vol. 115. – № 3. – P. 211–252. – DOI: 10.1007/s11263-015-0816-y.

248. ImageNet: a large-scale hierarchical image database / J. Deng, W. Dong, R. Socher [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2009. – Vol. 2009. – P. 248–255. – DOI: 10.1109/CVPR.2009.5206848.
249. Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images / V. Thambawita, I. Strümke, S. A. Hicks [et al.] // Diagnostics. – 2021. – Vol. 11. – № 12. – P. 2183. – DOI: 10.3390/diagnostics11122183.
250. Improving dataset transparency in dermatologic artificial intelligence using a dataset nutrition label / Y. Li, M. Taylor, K. S. Chmielinski [et al.] // NPJ Digital Medicine. – 2025. – Vol. 8. – № 1. – P. 641. – DOI: 10.1038/s41746-025-02125-9.
251. Imran, N. Artificial intelligence in medical education: Are we ready for it? / N. Imran, M. Jawaid // Pakistan Journal of Medical Sciences. – 2020. – Vol. 36. – № 5. – P. 857–859. – DOI: 10.12669/pjms.36.5.3042.
252. Incidence and mortality of nonmelanoma skin cancer in Europe: current trends and challenges / M. Sendín-Martín, R. C. Bueno-Molina, J. C. Hernández-Rodríguez [et al.] // Clinical & Translational Oncology. – 2026. – Vol. 28. – № 1. – P. 302–319. – DOI: 10.1007/s12094-025-03985-z.
253. Insights 2024: Attitudes toward AI. – Elsevier, 2024. – 43 p. – URL: <https://assets.ctfassets.net> (дата обращения: 12.05.2026). – Текст : электронный.
254. Instruments and new technologies for the in vivo diagnosis of melanoma / A. A. Marghoob, L. D. Swindle, C. Z. Moricz [et al.] // Journal of the American Academy of Dermatology. – 2003. – Vol. 49. – № 5. – P. 777–799. – DOI: 10.1016/s0190-9622(03)02470-8.
255. Introducing Artificial Intelligence Training in Medical Education / K. Paranjape, M. Schinkel, R. Nannan Panday [et al.] // JMIR Medical Education. – 2019. – Vol. 5. – № 2. – P. e16048. – DOI: 10.2196/16048.
256. Jobin, A. The global landscape of AI ethics guidelines / A. Jobin, M. Ienca, E. Vayena // Nature Machine Intelligence. – 2019. – Vol. 1. – № 9. – P. 389–399. – DOI: 10.1038/s42256-019-0088-2.

257. Kaushal, A. Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms / A. Kaushal, R. Altman, C. Langlotz // JAMA. – 2020. – Vol. 324. – № 12. – P. 1212–1213. – DOI: 10.1001/jama.2020.12067.
258. Key challenges for delivering clinical impact with artificial intelligence / C. J. Kelly, A. Karthikesalingam, M. Suleyman [et al.] // BMC Medicine. – 2019. – Vol. 17. – № 1. – P. 195. – DOI: 10.1186/s12916-019-1426-2.
259. Khalifa, N. E. A comprehensive survey of recent trends in deep learning for digital images augmentation / N. E. Khalifa, M. Loey, S. Mirjalili // Artificial Intelligence Review. – 2022. – Vol. 55. – № 3. – P. 2351–2377. – DOI: 10.1007/s10462-021-10066-4.
260. Khatun, N. Technology innovation to reduce health inequality in skin diagnosis and to improve patient outcomes for people of color: a thematic literature review and future research agenda / N. Khatun, G. Spinelli, F. Colecchia // Frontiers in Artificial Intelligence. – 2024. – Vol. 7. – P. 1394386. – DOI: 10.3389/frai.2024.1394386.
261. Kittler, H. Evolution of the Clinical, Dermoscopic and Pathologic Diagnosis of Melanoma / H. Kittler // Dermatology Practical & Conceptual. – 2021. – Vol. 11. – № Suppl 1. – P. e2021163S. – DOI: 10.5826/dpc.11S1a163S.
262. Krizhevsky, A. ImageNet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G. E. Hinton // Communications of the ACM. – 2017. – Vol. 60. – № 6. – P. 84–90. – DOI: 10.1145/3065386.
263. Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review / R. Daneshjou, M. P. Smith, M. D. Sun [et al.] // JAMA Dermatology. – 2021. – Vol. 157. – № 11. – P. 1362–1369. – DOI: 10.1001/jamadermatol.2021.3129.
264. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions / A. Abd-Alrazaq, R. AlSaad, D. Alhuwail [et al.] // JMIR Medical Education. – 2023. – Vol. 9. – P. e48291. – DOI: 10.2196/48291.
265. Large language models in medicine / A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, [et al.] // Nature Medicine. – 2023. – Vol. 29. – № 8. – P. 1930–1940. – DOI: 10.1038/s41591-023-02448-8.

266. LeCun, Y. Deep learning / Y. LeCun, Y. Bengio, G. Hinton // *Nature*. – 2015. – Vol. 521. – № 7553. – P. 436–444. – DOI: 10.1038/nature14539.
267. Leiter, U. Epidemiology of Skin Cancer: Update 2019 / U. Leiter, U. Keim, C. Garbe // *Advances in Experimental Medicine and Biology*. – 2020. – Vol. 1268. – P. 123–139. – DOI: 10.1007/978-3-030-46227-7_6.
268. Lin, H. Artificial intelligence (AI) -integrated educational applications and college students' creativity and academic emotions: students and teachers' perceptions and attitudes / H. Lin, Q. Chen // *BMC Psychology*. – 2024. – Vol. 12. – № 1. – P. 487. – DOI: 10.1186/s40359-024-01979-0.
269. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness / S. Vollmer, B. A. Mateen, G. Bohner [et al.] // *BMJ*. – 2020. – Vol. 368. – P. 16927. – DOI: 10.1136/bmj.16927.
270. Maity, S. Large language models in healthcare and medical applications: a review / S. Maity, M. J. Saikia // *Bioengineering*. – 2025. – Vol. 12. – № 6. – P. 631. – DOI: 10.3390/bioengineering12060631.
271. Maleki Varnosfaderani, S. The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century / S. Maleki Varnosfaderani, M. Forouzanfar // *Bioengineering*. – 2024. – Vol. 11. – № 4. – P. 337. – DOI: 10.3390/bioengineering11040337.
272. Malvehy, J. Follow-up of melanocytic skin lesions with digital total-body photography and digital dermoscopy: a two-step method / J. Malvehy, S. Puig // *Clinics in Dermatology*. – 2002. – Vol. 20. – № 3. – P. 297–304. – DOI: 10.1016/s0738-081x(02)00220-1.
273. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists / H. A. Haenssle, C. Fink, R. Schneiderbauer [et al.] // *Annals of Oncology*. – 2018. – Vol. 29. – № 8. – P. 1836–1842. – DOI: 10.1093/annonc/mdy166.

274. Marks, R. Malignant transformation of solar keratoses to squamous cell carcinoma / R. Marks, G. Rennie, T. S. Selwood // *The Lancet*. – 1988. – Vol. 1. – № 8589. – P. 795–797. – DOI: 10.1016/s0140-6736(88)91658-3.
275. Medical students' attitudes toward AI in education: perception, effectiveness, and its credibility / A. Sami, F. Tanveer, K. Sajwani [et al.] // *BMC Medical Education*. – 2025. – Vol. 25. – P. 82. – DOI: 10.1186/s12909-025-06704-y.
276. MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images / I. Giotis, N. Molders, S. Land [et al.] // *Expert Systems with Applications*. – 2015. – Vol. 42. – № 19. – P. 6578–6585. – DOI: 10.1016/j.eswa.2015.04.034.
277. Melanoma / D. Schadendorf, A. C. J. van Akkooi, C. Berking [et al.] // *The Lancet*. – 2018. – Vol. 392. – № 10151. – P. 971–984. – DOI: 10.1016/S0140-6736(18)31559-9.
278. Melanoma awareness and prevalence of dermoscopic examination among internet users: a cross-sectional survey / G. Kaminska-Winciorek, J. Wydmanski, M. Gajda, A. Tukiendorf // *Postępy Dermatologii i Alergologii*. – 2016. – Vol. 33. – № 6. – P. 421–428. – DOI: 10.5114/pdia.2016.63297.
279. Melanoma skin cancer statistics derived from 7442 Japanese patients: Japanese melanoma study / Y. Fujisawa, S. Yoshikawa, T. Takenouchi [et al.] // *International Journal of Clinical Oncology*. – 2025. – Vol. 30. – № 5. – P. 844–855. – DOI: 10.1007/s10147-025-02747-9.
280. Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual / J. E. Gershenwald, R. A. Scolyer, K. R. Hess [et al.] // *CA: A Cancer Journal for Clinicians*. – 2017. – Vol. 67. – № 6. – P. 472–492. – DOI: 10.3322/caac.21409.
281. Meta-analysis of digital dermoscopy follow-up of melanocytic skin lesions: a study on behalf of the International Dermoscopy Society / G. Salerni, T. Terán, S. Puig [et al.] // *Journal of the European Academy of Dermatology and Venereology*. – 2013. – Vol. 27. – № 7. – P. 805–814. – DOI: 10.1111/jdv.12032.
282. Meta-analysis of number needed to treat for diagnosis of melanoma by clinical setting / A. J. Petty, B. Ackerson, R. Garza [et al.] // *Journal of the American Academy*

of Dermatology. – 2020. – Vol. 82. – № 5. – P. 1158–1165. – DOI: 10.1016/j.jaad.2019.12.063.

283. Metrics reloaded: recommendations for image analysis validation / L. Maier-Hein, A. Reinke, P. Godau [et al.] // Nature Methods. – 2024. – Vol. 21. – № 2. – P. 195–212. – DOI: 10.1038/s41592-023-02151-z.

284. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist / B. Norgeot, G. Quer, B. K. Beaulieu-Jones [et al.] // Nature Medicine. – 2020. – Vol. 26. – № 9. – P. 1320–1324. – DOI: 10.1038/s41591-020-1041-y.

285. Morelli, S. Recent Advances in AI-Driven Mobile Health Enhancing Healthcare-Narrative Insights into Latest Progress / S. Morelli, D. Giansanti // Bioengineering. – 2025. – Vol. 13. – № 1. – P. 54. – DOI: 10.3390/bioengineering13010054.

286. Must-have Qualities of Clinical Research on Artificial Intelligence and Machine Learning / B. Koçak, R. Cuocolo, D. P. dos Santos [et al.] // Balkan Medical Journal. – 2023. – Vol. 40. – № 1. – P. 3–12. – DOI: 10.4274/balkanmedj.galenos.2022.2022-11-51.

287. Naik, P. P. Cutaneous Malignant Melanoma: A Review of Early Diagnosis and Management / P. P. Naik // World Journal of Oncology. – 2021. – Vol. 12. – № 1. – P. 7–19. – DOI: 10.14740/wjon1349.

288. Nguyen, V. D. Skin lesion classification on imbalanced data using deep learning with soft attention / V. D. Nguyen, N. D. Bui, H. K. Do // Sensors. – 2022. – Vol. 22. – № 19. – P. 7530. – DOI: 10.3390/s22197530.

289. Noninvasive Imaging Tools in the Diagnosis and Treatment of Skin Cancers / R. M. Bakos, T. P. Blumetti, R. Roldán-Marín, G. Salerni // American Journal of Clinical Dermatology. – 2018. – Vol. 19. – № 1. – P. 3–14. – DOI: 10.1007/s40257-018-0367-4.

290. Opinion research among Russian Physicians on the application of technologies using artificial intelligence in the field of medicine and health care / I. A. Orlova, Z. A. Akopyan, A. G. Plisyuk [et al.] // BMC Health Services Research. – 2023. – Vol. 23. – № 1. – P. 749. – DOI: 10.1186/s12913-023-09493-6.

291. Opportunities for improving the efficiency of keratinocyte carcinoma care in primary and specialist care: Results from population-based Dutch cohort studies / M.

- Wakkee, S. van Egmond, M. Louwman [et al.] // *European Journal of Cancer*. – 2019. – Vol. 117. – P. 32–40. – DOI: 10.1016/j.ejca.2019.05.010.
292. Outcomes of referral to dermatology for suspicious lesions: implications for teledermatology / K. V. Viola, W. L. Tolpinrud, C. P. Gross [et al.] // *Archives of Dermatology*. – 2011. – Vol. 147. – № 5. – P. 556–560. – DOI: 10.1001/archdermatol.2011.108.
293. Overview of skin cancer types and prevalence rates across continents / A. H. Roky, M. M. Islam, A. M. F. Ahasan [et al.] // *Cancer Pathogenesis and Therapy*. – 2024. – Vol. 3. – № 2. – P. 89–100. – DOI: 10.1016/j.cpt.2024.08.002.
294. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones / A. G. C. Pacheco, G. R. Lima, A. S. Salomão [et al.] // *Data in Brief*. – 2020. – Vol. 32. – P. 106221. – DOI: 10.1016/j.dib.2020.106221.
295. Park, S. H. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction / S. H. Park, K. Han // *Radiology*. – 2018. – Vol. 286. – № 3. – P. 800–809. – DOI: 10.1148/radiol.2017171920.
296. Patel, P. The need for improved dermoscopy training in residency: a survey of US dermatology residents and program directors / P. Patel, S. Khanna, B. McLellan, K. Krishnamurthy // *Dermatology Practical & Conceptual*. – 2017. – Vol. 7. – № 2. – P. 17–22. – DOI: 10.5826/dpc.0702a03.
297. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study / C. A. Nelson, L. M. Pérez-Chada, A. Creadore [et al.] // *JAMA Dermatology*. – 2020. – Vol. 156. – № 5. – P. 501–512. – DOI: 10.1001/jamadermatol.2019.5014.
298. Perspectives on artificial intelligence in dermatology: an international cross-sectional study / E. Karampinis, C. M. Zoumpourli, A. Lallas [et al.] // *Medicina*. – 2026. – Vol. 62. – № 4. – P. 759. – DOI: 10.3390/medicina62040759.
299. PH² – a dermoscopic image database for research and benchmarking / T. Mendonca, P. M. Ferreira, J. S. Marques [et al.] // *Annual International Conference of the*

- IEEE Engineering in Medicine and Biology Society. – 2013. – Vol. 2019. – P. 5437–5440. – DOI: 10.1109/EMBC.2013.6610779.
300. Qamar, T. Understanding the black-box: towards interpretable and reliable deep learning models / T. Qamar, N. Z. Bawany // *PeerJ Computer Science*. – 2023. – Vol. 9. – P. e1629. – DOI: 10.7717/peerj-cs.1629.
301. Que, S. K. T. Cutaneous squamous cell carcinoma: Incidence, risk factors, diagnosis, and staging / S. K. T. Que, F. O. Zwald, C. D. Schmults // *Journal of the American Academy of Dermatology*. – 2018. – Vol. 78. – № 2. – P. 237–247. – DOI: 10.1016/j.jaad.2017.08.059.
302. Real-World Evaluation of AI-Driven Diabetic Retinopathy Screening in Public Health Settings: Validation and Implementation Study / M. Duggal, A. Chauhan, V. Gupta [et al.] // *JMIR Medical Informatics*. – 2025. – Vol. 13. – P. e67529. – DOI: 10.2196/67529.
303. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine / G. D. Sanders, P. J. Neumann, A. Basu [et al.] // *JAMA*. – 2016. – Vol. 316. – № 10. – P. 1093–1103. – DOI: 10.1001/jama.2016.12195.
304. Reinehr, C. P. H. Actinic keratoses: review of clinical, dermoscopic, and therapeutic aspects / C. P. H. Reinehr, R. M. Bakos // *Anais Brasileiros de Dermatologia*. – 2019. – Vol. 94. – № 6. – P. 637–657. – DOI: 10.1016/j.abd.2019.10.004.
305. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI / B. Vasey, M. Nagendran, B. Campbell [et al.] // *Nature Medicine*. – 2022. – Vol. 28. – № 5. – P. 924–933. – DOI: 10.1038/s41591-022-01772-9.
306. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines / H. Ibrahim, X. Liu, S. C. Rivera [et al.] // *Trials*. – 2021. – Vol. 22. – № 1. – P. 11. – DOI: 10.1186/s13063-020-04951-6.
307. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images /

- M. A. Marchetti, N. C. F. Codella, S. W. Dusza [et al.] // *Journal of the American Academy of Dermatology*. – 2018. – Vol. 78. – № 2. – P. 270–277.e1. – DOI: 10.1016/j.jaad.2017.08.016.
308. Rethinking progressive low-light image enhancement: A frequency-aware tripartite multi-scale network / Y. Li, K. Zhang, X. Zhou [et al.] // *Neural Networks*. – 2026. – Vol. 196. – P. 108351. – DOI: 10.1016/j.neunet.2025.108351.
309. Rigel, D. S. The Evolution of Melanoma Diagnosis: 25 Years Beyond the ABCDs / D. S. Rigel, J. Russak, R. Friedman // *CA: A Cancer Journal for Clinicians*. – 2010. – Vol. 60. – № 5. – P. 301–316. – DOI: 10.3322/caac.20074.
310. Ring, C. Dermatoscopy / C. Ring, N. Cox, J. B. Lee // *Clinics in Dermatology*. – 2021. – Vol. 39. – № 4. – P. 635–642. – DOI: 10.1016/j.clindermatol.2021.03.009.
311. Robust Visual Identification of Under-resourced Dermatological Diagnoses with Classifier-Steered Background Masking / M. Dominguez, J. R. Wolf, P. Prasad [et al.] // *AMIA. Annual Symposium Proceedings. AMIA Symposium*. – 2024. – Vol. 2024. – P. 368–377.
312. Rosendahl, C. Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions / C. Rosendahl // *Journal of the American Academy of Dermatology*. – 2011. – Vol. 64. – № 6. – P. 1068–1073. – DOI: 10.1016/j.jaad.2010.03.039.
313. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead / C. Rudin // *Nature Machine Intelligence*. – 2019. – Vol. 1. – № 5. – P. 206–215. – DOI: 10.1038/s42256-019-0048-x.
314. Saito, T. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets / T. Saito, M. Rehmsmeier // *PLOS ONE*. – 2015. – Vol. 10. – № 3. – P. e0118432. – DOI: 10.1371/journal.pone.0118432.
315. Saroha, S. Artificial Intelligence in Medical Education: Promise, Pitfalls, and Practical Pathways / S. Saroha // *Advances in Medical Education and Practice*. – 2025. – Vol. 16. – P. 1039–1046. – DOI: 10.2147/AMEP.S523255.

316. Sat, M. The impact of AI integration in project preparation in education course on pre-service teachers' innovativeness, AI anxiety, attitudes, and acceptance / M. Sat // BMC Psychology. – 2025. – Vol. 13. – № 1. – P. 1297. – DOI: 10.1186/s40359-025-03647-3.
317. Score-CAM: score-weighted visual explanations for convolutional neural networks / H. Wang, Z. Wang, M. Du, [et al.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. – 2020. – Vol. 2020. – P. 24–25. – DOI: 10.1109/CVPRW50498.2020.00020.
318. Screening for reducing morbidity and mortality in malignant melanoma / M. Johansson, J. Brodersen, P. C. Gøtzsche, K. J. Jørgensen // The Cochrane Database of Systematic Reviews. – 2019. – Vol. 6. – № 6. – P. CD012352. – DOI: 10.1002/14651858.CD012352.pub2.
319. Screening for Skin Cancer in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force / K. J. Wernli, N. B. Henrikson, C. C. Morrison [et al.] // JAMA. – 2016. – Vol. 316. – № 4. – P. 436–447. – DOI: 10.1001/jama.2016.5415.
320. Shorten, C. A survey on image data augmentation for deep learning / C. Shorten, T. M. Khoshgoftaar // Journal of Big Data. – 2019. – Vol. 6. – № 1. – P. 60. – DOI: 10.1186/s40537-019-0197-0.
321. Skin and Syntax: Large Language Models in Dermatopathology / A. Shah, S. Wahood, D. Guerhazi [et al.] // Dermatopathology. – 2024. – Vol. 11. – № 1. – P. 101–111. – DOI: 10.3390/dermatopathology11010009.
322. Skin cancer and dermoscopy training for primary care physicians: a pilot study / V. De Bedout, N. M. Williams, A. M. Muñoz [et al.] // Dermatology Practical & Conceptual. – 2021. – Vol. 11. – № 1. – P. e2021145. – DOI: 10.5826/dpc.1101a145.
323. Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review / T. J. Brinker, A. Hekler, J. S. Utikal [et al.] // Journal of Medical Internet Research. – 2018. – Vol. 20. – № 10. – P. e11936. – DOI: 10.2196/11936.
324. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts / S. Haggenmüller, R. C. Maron, A. Hekler [et al.] //

- European Journal of Cancer. – 2021. – Vol. 156. – P. 202–216. – DOI: 10.1016/j.ejca.2021.06.049.
325. Skin cancer classification with deep learning: a systematic review / Y. Wu, B. Chen, A. Zeng [et al.] // *Frontiers in Oncology*. – 2022. – Vol. 12. – P. 893972. – DOI: 10.3389/fonc.2022.893972.
326. Skin cancer diagnosis by lesion, physician, and examination type: a systematic review and meta-analysis / J. Y. Chen, K. Fernandez, R. P. Fadadu [et al.] // *JAMA Dermatology*. – 2025. – Vol. 161. – № 2. – P. 135–146. – DOI: 10.1001/jamadermatol.2024.4382.
327. Skin Cancer Education Interventions for Primary Care Providers: A Scoping Review / A. E. Brown, M. Najmi, T. Duke [et al.] // *Journal of General Internal Medicine*. – 2022. – Vol. 37. – № 9. – P. 2267–2279. – DOI: 10.1007/s11606-022-07501-9.
328. Skin Lesion Segmentation in Dermoscopic Images with Noisy Data / N. Lama, J. Hagerty, A. Nambisan [et al.] // *Journal of Digital Imaging*. – 2023. – Vol. 36. – № 4. – P. 1712–1722. – DOI: 10.1007/s10278-023-00819-8.
329. Smartphone applications for melanoma detection by community, patient and generalist clinician users: a review / A. P. Kassianos, J. D. Emery, P. Murchie, F. M. Walter // *The British Journal of Dermatology*. – 2015. – Vol. 172. – № 6. – P. 1507–1518. – DOI: 10.1111/bjd.13665.
330. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition / K. He, X. Zhang, S. Ren, J. Sun // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2015. – Vol. 37. – № 9. – P. 1904–1916. – DOI: 10.1109/TPAMI.2015.2389824.
331. Squamous Cell Carcinoma – Similarities and Differences among Anatomical Sites / W. Yan, I. I. Wistuba, M. R. Emmert-Buck, H. S. Erickson // *American Journal of Cancer Research*. – 2011. – Vol. 1. – № 3. – P. 275–300.
332. Standardization of terminology in dermoscopy/dermatoscopy: Results of the third consensus conference of the International Society of Dermoscopy / H. Kittler, A. A. Marghoob, G. Argenziano [et al.] // *Journal of the American Academy of Dermatology*. – 2016. – Vol. 74. – № 6. – P. 1093–1106. – DOI: 10.1016/j.jaad.2015.12.038.

333. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies / P. M. Bossuyt, J. B. Reitsma, D. E. Bruns [et al.] // *BMJ*. – 2015. – Vol. 351. – P. h5527. – DOI: 10.1136/bmj.h5527.
334. Suitability of the Current Health Technology Assessment of Innovative Artificial Intelligence-Based Medical Devices: Scoping Literature Review / L. Farah, I. Borget, N. Martelli, A. Vallee // *Journal of Medical Internet Research*. – 2024. – Vol. 26. – P. e51514. – DOI: 10.2196/51514.
335. Teledermatology for diagnosing skin cancer in adults / N. Chuchu, J. Dinnes, Y. Takwoingi [et al.] // *The Cochrane Database of Systematic Reviews*. – 2018. – Vol. 12. – № 12. – P. CD013193. – DOI: 10.1002/14651858.CD013193.
336. The ABCD rule of dermatoscopy. High prospective value in the diagnosis of doubtful melanocytic skin lesions / F. Nachbar, W. Stolz, T. Merkle [et al.] // *Journal of the American Academy of Dermatology*. – 1994. – Vol. 30. – № 4. – P. 551–559. – DOI: 10.1016/s0190-9622(94)70061-3.
337. The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards / S. Holland, A. Hosny, S. Newman [et al.] // *arXiv*. – 2018. – Vol. 2018. – P. 1805.03677. – DOI: 10.48550/arXiv.1805.03677.
338. The dermatoscopic universe of basal cell carcinoma / A. Lallas, Z. Apalla, G. Argenziano [et al.] // *Dermatology Practical & Conceptual*. – 2014. – Vol. 4. – № 3. – P. 11–24. – DOI: 10.5826/dpc.0403a02.
339. The diagnostic accuracy of dermoscopy for basal cell carcinoma: a systematic review and meta-analysis / O. Reiter, I. Mimouni, M. Gdalevich [et al.] // *Journal of the American Academy of Dermatology*. – 2019. – Vol. 80. – № 5. – P. 1380–1388. – DOI: 10.1016/j.jaad.2018.12.026.
340. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review / J. Wolff, J. Pauling, A. Keck, J. Baumbach // *Journal of Medical Internet Research*. – 2020. – Vol. 22. – № 2. – P. e16866. – DOI: 10.2196/16866.
341. The ethics of AI in health care: A mapping review / J. Morley, C. C. V. Machado, C. Burr [et al.] // *Social Science & Medicine*. – 2020. – Vol. 260. – P. 113172. – DOI: 10.1016/j.socscimed.2020.113172.

342. The FAIR Guiding Principles for scientific data management and stewardship / M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg [et al.] // *Scientific Data*. – 2016. – Vol. 3. – P. 160018. – DOI: 10.1038/sdata.2016.18.
343. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review / B. D. Simon, K. Ozyoruk, D. G. Gelikman [et al.] // *Diagnostic and Interventional Radiology*. – 2024. – Vol. 30. – № 6. – P. 385–395. – DOI: 10.4274/dir.2024.242631.
344. The Impact of Artificial Intelligence on Healthcare: A Comprehensive Review of Advancements in Diagnostics, Treatment, and Operational Efficiency / M. Faiyazuddin, S. J. Q. Rahman, G. Anand [et al.] // *Health Science Reports*. – 2025. – Vol. 8. – № 1. – P. e70312. – DOI: 10.1002/hsr2.70312.
345. The Importance of Incorporating Human Factors in the Design and Implementation of Artificial Intelligence for Skin Cancer Diagnosis in the Real World / C. M. Felmingham, N. R. Adler, Z. Ge [et al.] // *American Journal of Clinical Dermatology*. – 2021. – Vol. 22. – № 2. – P. 233–242. – DOI: 10.1007/s40257-020-00574-4.
346. The role of dermoscopy and digital dermoscopy follow-up in the clinical diagnosis of melanoma: clinical and dermoscopic features of 99 consecutive primary melanomas / G. Salerni, T. Terán, C. Alonso, R. Fernández-Bussy // *Dermatology Practical & Conceptual*. – 2014. – Vol. 4. – № 4. – P. 39–46. – DOI: 10.5826/dpc.0404a07.
347. Thind, B. S. Artificial intelligence in oncology: promise, peril, and the future of patient–physician interaction / B. S. Thind, C. K. Tsao // *Frontiers in Digital Health*. – 2025. – Vol. 7. – P. 1633577. – DOI: 10.3389/fdgth.2025.1633577.
348. Three-point checklist of dermoscopy. A new screening method for early detection of melanoma / H. P. Soyer, G. Argenziano, I. Zalaudek [et al.] // *Dermatology*. – 2004. – Vol. 208. – № 1. – P. 27–31. – DOI: 10.1159/000075042.
349. Tjiu, J. W. Equity and generalizability of artificial intelligence for skin-lesion diagnosis using clinical, dermoscopic, and smartphone images: a systematic review and meta-analysis / J. W. Tjiu, C. F. Lu // *Medicina*. – 2025. – Vol. 61. – № 12. – P. 2186. – DOI: 10.3390/medicina61122186.

350. Transfer learning for medical image classification: a literature review / H. E. Kim, A. Cosa-Linan, N. Santhanam [et al.] // BMC Medical Imaging. – 2022. – Vol. 22. – № 1. – P. 69. – DOI: 10.1186/s12880-022-00793-7.
351. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement / G. S. Collins, J. B. Reitsma, D. G. Altman, K. G. Moons // BMC Medicine. – 2015. – Vol. 13. – P. 1. – DOI: 10.1186/s12916-014-0241-z
352. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods / G. S. Collins, K. G. M. Moons, P. Dhiman [et al.] // BMJ. – 2024. – Vol. 385. – P. e078378. – DOI: 10.1136/bmj-2023-078378.
353. Tsang, M. W. Even patients with changing moles face long dermatology appointment wait-times: a study of simulated patient calls to dermatologists / M. W. Tsang, J. S. Resneck Jr // Journal of the American Academy of Dermatology. – 2006. – Vol. 55. – № 1. – P. 54–58. – DOI: 10.1016/j.jaad.2006.04.001.
354. Tschandl, P. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions / P. Tschandl, C. Rosendahl, H. Kittler // Scientific Data. – 2018. – Vol. 5. – P. 180161. – DOI: 10.1038/sdata.2018.161.
355. Turcian, D. Artificial intelligence in primary care: an overview / D. Turcian, V. Stoicu-Tivadar // Studies in Health Technology and Informatics. – 2022. – Vol. 289. – P. 208–211. – DOI: 10.3233/SHTI210896.
356. Ugly Duckling Sign as a Major Factor of Efficiency in Melanoma Detection / C. Gaudy-Marqueste, Y. Wazaefi, Y. Bruneu [et al.] // JAMA Dermatology. – 2017. – Vol. 153. – № 4. – P. 279–284. – DOI: 10.1001/jamadermatol.2016.5500.
357. Uncertainty Quantification for Clinical Outcome Predictions with (Large) Language Models / Z. Chen, P. Li, X. Dong, P. Hong // Findings of the Association for Computational Linguistics: NAACL. – 2025. – Vol. 2025. – P. 7512–7523. – DOI: 10.18653/v1/2025.findings-naacl.419.

358. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations / L. Seyyed-Kalantari, H. Zhang, M. B. A. McDermott [et al.] // *Nature Medicine*. – 2021. – Vol. 27. – № 12. – P. 2176–2182. – DOI: 10.1038/s41591-021-01595-0.
359. Use of a Large Language Model as a Dermatology Case Narrator: Exploring the Dynamics of a Chatbot as an Educational Tool in Dermatology / E. Karampinis, D. A. Bozi Tzetzzi, G. Pappa [et al.] // *JMIR Dermatology*. – 2025. – Vol. 8. – P. e72058. – DOI: 10.2196/72058.
360. Use of Deep Neural Networks in the Detection and Automated Classification of Lesions Using Clinical Images in Ophthalmology, Dermatology, and Oral Medicine-A Systematic Review / R. F. T. Gomes, L. F. Schuch, M. D. Martins [et al.] // *Journal of Digital Imaging*. – 2023. – Vol. 36. – № 3. – P. 1060–1070. – DOI: 10.1007/s10278-023-00775-3.
361. Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge / M. Combalia, N. Codella, V. Rotemberg [et al.] // *The Lancet Digital Health*. – 2022. – Vol. 4. – № 5. – P. e330–e339. – DOI: 10.1016/S2589-7500(22)00021-8.
362. Value of dermoscopy in a population-based screening sample by dermatologists / I. Hoorens, K. Vossaert, S. Lanssens [et al.] // *Dermatology Practical & Conceptual*. – 2019. – Vol. 9. – № 3. – P. 200–206. – DOI: 10.5826/dpc.0903a05.
363. Varoquaux, G. Machine learning for medical imaging: methodological failures and recommendations for the future / G. Varoquaux, V. Cheplygina // *npj Digital Medicine*. – 2022. – Vol. 5. – № 1. – P. 48. – DOI: 10.1038/s41746-022-00592-y.
364. Views on mobile health apps for skin cancer screening in the general population: an in-depth qualitative exploration of perceived barriers and facilitators / T. E. Sangers, M. Wakkee, E. C. Kramer-Noels [et al.] // *The British Journal of Dermatology*. – 2021. – Vol. 185. – № 5. – P. 961–969. – DOI: 10.1111/bjd.20441.
365. Virtual patient simulation to enhance medical students' clinical communication and decision-making skills: a pilot study / A. Dávidovics, K. Dávidovics, P. Hillebrand [et

al.] // BMC Medical Education. – 2025. – Vol. 26. – № 1. – P. 171. – DOI: 10.1186/s12909-025-08507-7.

366. Virtual Patient Simulations in Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration / A. A. Kononowicz, L. A. Woodham, S. Edelbring [et al.] // Journal of Medical Internet Research. – 2019. – Vol. 21. – № 7. – P. e14676. – DOI: 10.2196/14676.

367. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration / B. M. Kyaw, N. Saxena, P. Posadzki [et al.] // Journal of Medical Internet Research. – 2019. – Vol. 21. – № 1. – P. e12959. – DOI: 10.2196/12959.

368. Vital signs: melanoma incidence and mortality trends and projections – United States, 1982-2030 / G. P. Guy Jr, C. C. Thomas, T. Thompson [et al.] // MMWR. Morbidity and Mortality Weekly Report. – 2015. – Vol. 64. – № 21. – P. 591–596.

369. Wang, M. Recent global patterns in skin cancer incidence, mortality, and prevalence / M. Wang, X. Gao, L. Zhang // Chinese Medical Journal. – 2025. – Vol. 138 – № 2. – P. 185–192. – DOI: 10.1097/CM9.00000000000003416.

370. Wartman, S. A. Medical Education Must Move From the Information Age to the Age of Artificial Intelligence / S. A. Wartman, C. D. Combs // Academic Medicine. – 2018. – Vol. 93. – № 8. – P. 1107–1109. – DOI: 10.1097/ACM.0000000000002044.

371. Whiteman, D. C. The Growing Burden of Invasive Melanoma: Projections of Incidence Rates and Numbers of New Cases in Six Susceptible Populations through 2031 / D. C. Whiteman, A. C. Green, C. M. Olsen // The Journal of Investigative Dermatology. – 2016. – Vol. 136. – № 6. – P. 1161–1171. – DOI: 10.1016/j.jid.2016.01.035.

372. WHO classification of skin tumours / D. E. Elder, D. Massi, R. A. Scolyer, R. Willemze (Eds.). – World Health Organization, 2018. – 500 p.

373. WHO classification of skin tumours. 4th edition / D.E. Elder, D. Massi, R.A. Scolyer, R. Willemze (Eds.). Lyon: World Health Organization, 2018. – 500 pp.

374. Wu, X. Dermoscopy: not just for dermatologists / X. Wu, M. A. Marchetti, A. A. Marghoob // Melanoma Management. – 2015. – Vol. 2. – № 1. – P. 63–73. – DOI: 10.2217/mmt.14.32.

ПРИЛОЖЕНИЕ А



Рисунок А.1 – Диплом