

Федеральное государственное автономное образовательное учреждение
высшего образования
**Первый Московский государственный медицинский
Университет имени И.М.Сеченова**
Министерство здравоохранения Российской Федерации
(Сеченовский Университет)

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

на тему: «Применение радиомикки для классификации новообразований
молочных желез различной этиологии согласно шкале BI-RADS по данным
маммографии»

Направление подготовки: Медицинская биофизика (30.05.02)

Студент: Пирогов Дмитрий Сергеевич

Руководитель: доктор медицинских наук, руководитель отдела медицинской
информатики, радиомикки и радиогеномики Государственное бюджетное
учреждение здравоохранения города Москвы «Научно-практический
клинический центр диагностики и телемедицинских технологий
Департамента здравоохранения города Москвы», Арзамасов Кирилл
Михайлович

Москва, 2026

ВВЕДЕНИЕ

Рак молочной железы (РМЖ) занимает лидирующую позицию среди злокачественных новообразований у женщин в мире. По данным Международного агентства по изучению рака (GLOBOCAN, 2022), ежегодно в мире регистрируется свыше 2,3 миллиона новых случаев РМЖ и более 666 тысяч смертей от данного заболевания; в 157 из 185 стран мира. РМЖ является наиболее распространённым онкологическим заболеванием среди женского населения [1]. В Российской Федерации (РФ) в 2023 году зафиксировано свыше 81 тысячи новых случаев РМЖ, что составляет 22,5% всех злокачественных новообразований у женщин [2]. Приведённые данные определяют первостепенную медицинскую и социальную значимость своевременного выявления и точной диагностики данного заболевания.

Маммография является стандартом лучевой диагностики и скрининга РМЖ. Для стандартизации описания маммографических заключений применяется система Breast Imaging Reporting and Data System (BI-RADS), разработанная Американским колледжем радиологии (ACR): она структурирует находки по категориям от 0 до 6, отражающим степень онкологической настороженности и определяющим дальнейшую тактику ведения пациентки [3]. Вместе с тем интерпретация маммограмм остаётся задачей с высокой вариабельностью между врачами, что влечёт как гипердиагностику с избыточными направлениями на биопсию, так и гиподиагностику с пропуском клинически значимых образований [4]. Это противоречие между доказанной эффективностью маммографического скрининга и ограниченной воспроизводимостью её интерпретации является принципиальным обоснованием необходимости разработки систем автоматизированной классификации.

Одним из наиболее перспективных инструментов для решения этой задачи является радиомика — количественный анализ изображений, основанный на извлечении количественных данных из медицинских снимков.

В отличие от традиционного визуального анализа, радиомика позволяет объективно охарактеризовать текстурную неоднородность, морфологические и интенсивностические свойства тканей, недоступные невооружённому взгляду. В сочетании с алгоритмами машинного обучения радиомические признаки могут служить основой для систем поддержки принятия врачебных решений (СППВР), повышающих точность маммографической диагностики [5, 6].

Существующие радиомические исследования в маммографии, как правило, ограничены анализом объёмных образований и задачей бинарной классификации «доброкачественное – злокачественное» [7, 8]. Практически отсутствуют работы, сопоставляющие различные стратегии классификации по шкале BI-RADS в рамках единого исследования, охватывающего весь спектр морфологических типов образований. Настоящая работа направлена на устранение этого пробела.

Актуальность. Разработка и валидация радиомических моделей машинного обучения для классификации маммографических образований по шкале BI-RADS, а также сравнительной оценки четырёх стратегий классификации – многоклассовой, трёхклассовой и двух вариантов двухклассовой (логистическая регрессия и случайный лес) – с целью определения оптимального подхода для практического применения в условиях скрининга.

Цель исследования. Оценить диагностическую точность радиомического анализа с применением алгоритмов машинного обучения для классификации изменений молочной железы по шкале BI-RADS при маммографии.

Задачи исследования:

1. Сформировать и проанализировать набор данных, включающий маммографические исследования, содержащие размеченные вручную очаговые образования молочных желез;
2. Получить радиомические признаки образований молочных желез;

3. Разработать модель машинного обучения, способную классифицировать образования молочных желез с использованием радиомических признаков;

4. Оценить метрики диагностической точности модели и сформулировать выводы.

Научная новизна. В рамках исследования впервые разработаны и сопоставлены модели для трёх типов классификации: двухклассовой, трёхклассовой и многоклассовой, что позволяет оценить границы применимости радиомического подхода в зависимости от клинической задачи.

Практическая значимость. Разработанная модель двухклассовой классификации может служить методологической основой для создания СППВР в рамках скрининговой маммографии. Применение модели в режиме «второго читателя» способно снизить частоту избыточных направлений на биопсию без клинически значимого роста числа пропущенных случаев патологии. Практическое внедрение возможно после проведения внешней многоцентровой валидации и проспективного клинического исследования.

Апробация результатов:

1. Основные положения работы представлены в виде устного доклада на VI Открытой конференции молодых учёных Центра диагностики и телемедицины (с международным участием, Москва, 1–2 апреля 2026 г.);

2. Основные положения работы представлены в виде устного доклада на Молодежном научном форуме «МЕДИЦИНСКАЯ ВЕСНА – 2026» (с международным участием, Москва, 28–30 апреля 2026 г.);

3. По материалам исследования подготовлена научная статья (направлена в редакцию журнала «Здоровье населения и среда обитания – ЗНиСО», находится на рецензировании).

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	2
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ	10
1.1. Эпидемиология рака молочной железы и роль ранней диагностики ...	10
1.1.1. Эпидемиологическая ситуация в Российской Федерации	10
1.1.2. Факторы риска и стадийная структура	11
1.1.3. Роль скрининговой маммографии в снижении смертности	11
1.1.4. Маммографический скрининг в Российской Федерации: организационный контекст	12
1.2. Маммография как стандарт скрининга. Система BI-RADS: структура, ограничения, межэкспертная вариабельность	13
1.2.1. Маммография в системе лучевой диагностики молочной железы .	13
1.2.2. Система BI-RADS: история, структура и клиническое значение ...	14
1.2.3. Ограничения системы BI-RADS и межэкспертная вариабельность	16
1.2.4. Инструментальное обеспечение маммографии и требования к цифровым изображениям	17
1.3. Радиомика: концепция, методологический анализ и применение в онкологической радиологии	19
1.3.1. Концепция радиомики и её место в онкологической визуализации	19
1.3.2. Этапы радиомического анализа	20
1.3.3. Классификация радиомических признаков	21
1.3.4. Стандартизация радиомики	22
1.3.5. Ограничения и открытые вызовы радиомики	23
1.4. Радиомический анализ маммограмм: обзор актуальных исследований	24
1.4.1. Общая характеристика направления	24

1.4.2. Дифференциация доброкачественных и злокачественных образований	25
1.4.3. Радиомика для поддержки стратификации BI-RADS и снижения числа биопсий	26
1.4.4. Радиомика цифрового томосинтеза молочных желёз	27
1.4.5. Предсказание молекулярного подтипа и прогноза	27
1.4.6. Ограничения опубликованных работ и обоснование настоящего исследования	28
1.5. Алгоритмы машинного обучения в задачах классификации медицинских изображений	29
1.5.1. Место машинного обучения в клинической радиологии	29
1.5.2. Логистическая регрессия с регуляризацией LASSO	31
1.5.3. Алгоритм случайного леса	32
1.5.4. Сравнительная характеристика алгоритмов в контексте задачи	32
1.5.5. Метрики оценки качества классификации	32
ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ	34
2.1. Материал и дизайн исследования	34
2.1.1. Тип и дизайн исследования	34
2.1.2. Источник данных	34
2.1.3. Критерии включения и исключения	35
2.1.4. Характеристика итоговой выборки	36
2.2. Подготовка изображений	37
2.2.1. Конвертация изображений из формата DICOM в PNG	37
2.2.2. Инструмент разметки: Computer Vision Annotation Tool	38
2.2.3. Контроль качества разметки	39
2.2.4. Экспорт аннотаций	39
2.3. Радиомический анализ: извлечение, предобработка и отбор признаков	40
2.3.1. Общая схема радиомического анализа	40

2.3.2. Настройки экстрактора признаков	40
2.3.3. Классы извлекаемых признаков	41
2.3.4. Формирование итоговой таблицы признаков	42
2.3.5. Ранжирование признаков по F-критерию Фишера.....	43
2.4. Постановка задач классификации и формирование целевых переменных	45
2.4.1. Общий подход к формированию целевой переменной.....	45
2.4.2. Стратегия 1: многоклассовая классификация по полной шкале BI-RADS.....	45
2.4.3. Стратегия 2: трёхклассовая классификация с исключением BI-RADS 3	46
2.4.4. Стратегия 3: двухклассовая классификация с применением логистической регрессии	47
2.4.5. Стратегия 4: двухклассовая классификация с применением случайного леса.....	47
2.4.6. Сводная характеристика стратегий классификации	48
2.5. Машинное обучение: модели, гиперпараметры, валидация и метрики оценки.....	49
2.5.1. Общая схема построения классификационных моделей	49
2.5.2. Метрики оценки качества классификации	50
ГЛАВА 3. РЕЗУЛЬТАТЫ РАДИОМИЧЕСКОГО АНАЛИЗА И КЛАССИФИКАЦИИ МАММОГРАФИЧЕСКИХ ОБРАЗОВАНИЙ.....	51
3.1. Результаты стратегии 1 – многоклассовая классификация (случайный лес).....	51
3.1.1. Характеристика задачи и обучающей выборки.....	51
3.1.2. Подбор гиперпараметров	51
3.1.3. Метрики диагностической точности	51
3.1.4. Анализ ошибок классификации	52

3.2. Результаты стратегии 2 – трёхклассовая классификация (случайный лес)	53
3.2.1. Характеристика задачи и обучающей выборки	53
3.2.2. Подбор гиперпараметров	53
3.2.3. Метрики диагностической точности	53
3.2.4. Распределение вероятностных оценок	54
3.3. Результаты стратегии 3 – двухклассовая классификация с применением логистической регрессии	54
3.3.1. Характеристика задачи и обучающей выборки	54
3.3.2. Подбор гиперпараметров	55
3.3.3. Метрики диагностической точности	55
3.3.4. Сопоставление с другими стратегиями	55
3.4. Результаты стратегии 4 – двухклассовая классификация с применением случайного леса	56
3.4.1. Характеристика задачи и обучающей выборки	56
3.4.2. Подбор гиперпараметров	56
3.4.3. Метрики диагностической точности	56
3.4.4. Сопоставление стратегий 3 и 4	57
3.5. Сравнительный анализ четырёх стратегий классификации	57
3.5.1. Сводная таблица метрик	57
3.5.2. ROC-кривые и дискриминационная способность	59
3.5.3. Precision-Recall кривые и работа в условиях дисбаланса	59
3.5.4. Сравнительная диаграмма ключевых метрик	59
3.5.5. Анализ типов ошибок в клиническом контексте	59
3.5.6. Динамика метрик по стратегиям: ключевые закономерности	60
3.6. Клиническая интерпретация результатов	61
3.6.1. Общий контекст клинической применимости	61
3.6.2. Клиническая интерпретация стратегии 3 как ведущей модели	61

3.6.3. Роль наиболее значимых признаков в клинической интерпретации	62
3.6.4. Сравнение с литературными данными	62
3.7. Ограничения исследования	64
3.7.1. Объём и структура выборки	64
3.7.2. Дизайн исследования и отсутствие внешней валидации	64
3.7.3. Методология сегментации	65
3.7.4. Радиомический анализ	65
3.7.5. Интерпретация и обобщаемость.....	66
ЗАКЛЮЧЕНИЕ.....	67
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....	69
СПИСОК ЛИТЕРАТУРЫ	73

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Эпидемиология рака молочной железы и роль ранней диагностики

РМЖ самое частое злокачественное новообразование у женщин в мире и одна из ведущих причин онкологической смертности.

Заболеваемость РМЖ продолжает расти: по последним оценкам, ежегодные показатели увеличиваются на 1–5% в половине обследованных стран. Рост распределён неравномерно [1].

1.1.1. Эпидемиологическая ситуация в Российской Федерации

Эпидемиология РМЖ в России соответствует глобальным тенденциям. По данным Каприна А.Д. и соавт. (2024), в 2023 году в РФ зарегистрировано свыше 81 000 новых случаев РМЖ, что составляет 22,5% всех злокачественных новообразований у женщин. Стандартизированный по возрасту показатель заболеваемости РМЖ в России – 57,7 случая на 100 000 женщин, смертности – 13,6 на 100 000 женщин. РМЖ занимает первое место в структуре онкологической заболеваемости и смертности среди женщин страны [2].

Одновременно в России фиксируется снижение летальности от РМЖ – примерно вдвое за последние годы – и рост пятилетней выживаемости. Авторы источника связывают эту динамику с расширением охвата скринингом и совершенствованием лечебных протоколов. В структуре онкологической заболеваемости женщин РМЖ опережает рак шейки матки (22,5% против 6,3%) и рак тела матки.

Отечественная эпидемиология РМЖ отличается выраженной региональной вариабельностью: в ряде регионов стандартизированный показатель заболеваемости на 38–42% ниже общероссийского среднего. Различия частично объясняются неравномерным охватом маммографическим

скринингом. По данным Васильева Ю.А. и соавт. (2024), охват первым этапом диспансеризации в 2022 году заметно изменился по сравнению с 2019 годом, что авторы связывают с нарушением плановой медицинской помощи в период пандемии COVID-19 [9].

1.1.2. Факторы риска и стадийная структура

Этиология РМЖ комплексна, так к установленным факторам риска относятся пожилой возраст (пик заболеваемости – 45–69 лет), отягощённый семейный анамнез, мутации генов Breast Cancer Susceptibility Gene (BRCA): BRCA1 и BRCA2, воздействие ионизирующего излучения в анамнезе, ожирение в постменопаузальном периоде, длительная гормонозаместительная терапия и ранее перенесённые пролиферативные заболевания молочной железы. Возрастно-специфические паттерны заболеваемости различаются между странами с разным уровнем ионизирующего излучения в анамнезе (ИЧР), что указывает на вклад средовых и репродуктивных факторов.

Стадия заболевания на момент диагноза является основным детерминантом прогноза. Пятилетняя выживаемость при РМЖ I стадии – более 95%, при IV стадии – менее 30%. В российской структуре заболеваемости РМЖ преобладают пациентки со II стадией. Для сравнения, в странах с развитыми скрининговыми программами преобладает I стадия, что отражает более раннюю диагностику и связано с более низкой смертностью [1, 10].

1.1.3. Роль скрининговой маммографии в снижении смертности

Маммографический скрининг – наиболее доказательно обоснованный метод ранней диагностики РМЖ.

По расчётам CISNET (Cancer Intervention and Surveillance Modeling Network), ежегодный маммографический скрининг женщин 40–79 лет снижает смертность от РМЖ на 41,7% при использовании цифровой маммографии или

томосинтеза. Двухлетний скрининг в возрастных группах 50–74 и 40–74 лет снижает смертность на 25,4% и 30% соответственно. Таким образом, более раннее начало скрининга и ежегодная кратность связаны с дополнительным снижением смертности [11].

При этом эффект скрининга зависит не только от охвата, но и от качества интерпретации маммограмм. В литературе описана существенная межэкспертная вариабельность в отнесении образований к категориям BI-RADS [4]. Расхождение между доказанной эффективностью маммографического скрининга и ограниченной воспроизводимостью его интерпретации служит одним из обоснований для разработки автоматизированных систем классификации на основе количественных признаков изображения, чему посвящено настоящее исследование.

1.1.4. Маммографический скрининг в Российской Федерации: организационный контекст

В РФ маммографический скрининг проводится в рамках программы диспансеризации и профилактических медицинских осмотров. Первый этап диспансеризации включает маммографию для женщин 40 лет и старше с периодичностью один раз в два года [12]. В Москве сбор и хранение результатов маммографических исследований ведутся в Едином радиологическом информационном сервисе (ЕРИС) Единой медицинской информационно-аналитической системы (ЕМИАС), что позволяет проводить ретроспективный анализ накопленных данных [13].

Цифровизация маммографического скрининга расширяет возможности применения радиомики и машинного обучения, так цифровые изображения в формате Digital Imaging and Communications in Medicine (DICOM), в отличие от плёночных снимков, допускают стандартизированное извлечение количественных признаков без потерь, связанных с оцифровкой [14]. Наличие цифровых архивов маммограмм с верифицированными заключениями — необходимое условие для радиомических исследований, в том числе

настоящего. Эпидемиологические данные, доказательная база маммографического скрининга и текущая организация скрининга в РФ задают контекст, в котором ставится задача автоматизированной классификации маммографических изменений. Основой стандартизации описания маммографических находок служит система BI-RADS, которая подробно рассматривается в следующем разделе.

1.2. Маммография как стандарт скрининга. Система BI-RADS: структура, ограничения, межэкспертная вариабельность

1.2.1. Маммография в системе лучевой диагностики молочной железы

Маммография – основной метод лучевой диагностики рака молочной железы и наиболее широко применяемый метод скрининга. За десятилетия использования накоплена доказательная база её эффективности в снижении смертности от РМЖ. Современный стандарт исследования включает снимки в двух проекциях для каждой молочной железы – краниокаудальной (СС) и медиолатеральной кривой (МЛО) [3]. При необходимости выполняются прицельные снимки с компрессией, увеличенные изображения или снимки в аксиллярной проекции.

Один из значимых технологических этапов – переход от плёночной к цифровой маммографии. Цифровая маммография (ЦМ) обеспечивает более высокую чувствительность по сравнению с плёночными системами, особенно у женщин моложе 50 лет с плотной тканью молочной железы: прирост чувствительности – около 27% (78% против 51%, $p = 0,002$). Следующий этап – томосинтез молочных желёз (ТМЖ), при котором трёхмерная реконструкция строится по серии проекций под разными углами. По метаанализу 27 исследований ($n = 59\,493$ молочных желёз), комбинация ТМЖ + ЦМ даёт суммарную чувствительность 87% (95% ДИ: 84–90%) и специфичность 84%

(95% ДИ: 78–88%), что статистически значимо выше показателей ЦМ отдельно (76% и 80%). Преимущество ТМЖ + ЦМ наиболее выражено при плотных молочных железах и в группах высокого риска [11].

При этом маммография имеет объективные ограничения, связанные в первую очередь с плотностью ткани молочной железы. При жировом типе строения чувствительность метода достигает 98%, а при крайне плотной ткани (категория d по ACR BI-RADS) снижается до 48–51%. По данным крупного проспективного скринингового исследования, у 9,7% женщин с крайне плотными молочными железами чувствительность маммографии составляла 51,3% [15]. Частота интервальных раков (раков, выявленных между плановыми скрининговыми исследованиями) в этой подгруппе в 4 раза выше, чем у женщин с жировым типом строения [4]. Такие случаи формируют основной резерв для улучшения диагностики – и за счёт дополнительных методов визуализации (МРТ, ультразвук), и за счёт алгоритмов автоматизированного анализа изображений.

1.2.2. Система BI-RADS: история, структура и клиническое значение

В 1986 году Американская коллегия радиологов (American College of Radiology, ACR) начала разработку системы BI-RADS; первое издание опубликовано в 1993 году. Разработка была вызвана быстрым ростом объёма маммографических исследований в конце 1980-х годов и неоднородностью радиологических заключений, которая затрудняла клиническую интерпретацию и решения в онкологии [3].

Действующее 5-е издание BI-RADS (ACR BI-RADS Atlas, 2013, с обновлением 2025 года) охватывает маммографию, ультразвуковое исследование и МРТ молочных желёз [3]. Для маммографии система задаёт стандартизированный лексикон описательных характеристик (форма образования, края, плотность, тип кальцинатов, архитектурные нарушения и др.) и итоговую категорию оценки, которой соответствуют определённый риск

злокачественности и клиническая тактика. Система включает 7 категорий (0–6); их характеристики приведены в таблице 1.

Таблица 1. Категории BI-RADS для маммографии

Категория оценки BI-RADS	Рекомендация	Вероятность рака
Категория 0 – требуется дополнительная визуализация	Повторить исследование	Не определена
Категория 1 – нет изменений	Обычное наблюдение	0 %
Категория 2 – доброкачественные изменения	Обычное наблюдение	0 %
Категория 3 – вероятно, доброкачественные изменения	Повторное обследование через короткий интервал времени (6 мес)	0-2%
Категория 4 – подозрение на рак	Биопсия	
Категория 4А	Мало подозрительные изменения	2-10%
Категория 4В	Подозрительные изменения	10-50%
Категория 4С	Крайне подозрительные изменения	50-95%
Категория 5 – характерно для рака	Биопсия	Выше 95%
Категория 6 – подтвержденный биопсией рак	Хирургическое лечение, если показано	Определена

Введение подкатегорий 4а, 4б и 4с в 5-м издании позволило детализировать тактику ведения наиболее неоднородной группы пациенток, которым показана биопсия. По данным ретроспективного исследования (Mohapatra S.K. и соавт., 2021), прогностическая ценность положительного результата (ПЦПР) подкатегорий для злокачественности составила: BI-RADS

4a – 34%, 4b – 89%, 4c – 97% [16]. В других публикациях абсолютные значения ПЦПР различаются в зависимости от популяции и методики, но градиент ПЦПР от 4a к 4c воспроизводится. В отечественном контексте Гаждонова В.Е. и соавт. (2025) отмечают, что для повышения точности классификации образований категорий 3–4 целесообразно комплексное применение BI-RADS вместе с дополнительными методами – ультразвуковым исследованием с контрастированием и эластографией [17].

Категория BI-RADS 3 занимает особое место в клинической практике из-за неоднозначности тактических решений. Она предполагает риск злокачественности менее 2% и краткосрочное наблюдение с контрольной маммографией через 6 месяцев. По данным исследования на 43628 пациентках, при BI-RADS 3 злокачественность была верифицирована в 1,9% случаев; при этом у 12% инвазивных раков, выявленных в течение первых 6 месяцев наблюдения, уже было поражение регионарных лимфатических узлов. Эти данные показывают клинический риск запоздалой диагностики при избыточно широком применении BI-RADS 3 и подтверждают запрос на объективизацию критериев её присвоения.

1.2.3. Ограничения системы BI-RADS и межэкспертная вариабельность

Одно из основных ограничений BI-RADS – высокая межэкспертная вариабельность интерпретации маммограмм и присвоения итоговой категории. В исследовании скрининговых маммограмм (оценка по к Коэна) согласие для итоговой категории составило $\kappa = 0,37$ (95% ДИ: 0,36–0,38), что соответствует лишь «удовлетворительному» уровню воспроизводимости. При упрощении шкалы до бинарной («требуется дообследование» – категории 0, III, IV, V – против «норма/доброкачественное» – I и II) согласие возросло до «умеренного» ($\kappa = 0,53$) [4].

Для отдельных семиотических признаков показатели согласия ниже. По данным Liberman L. и соавт., межэкспертное согласие для итоговой категории

BI-RADS по маммограммам и сонограммам составило $\kappa = 0,28$ («удовлетворительное»), а для плотности образования – $\kappa = 0,18$ («незначительное»). Наибольшие расхождения наблюдались между категориями II и III, в том числе в группе верифицированных злокачественных образований, что указывает на риск диагностических потерь при недооценке категории.

Существенна и вариабельность между экспертами, при повторной оценке тех же изображений одним радиологом согласие для итоговой категории составляет $\kappa = 0,47–0,53$. На практике это означает, что в части случаев врач при пересмотре присваивает тому же образованию другую категорию. Клинические следствия, как неоправданное направление на биопсию, при гипердиагностике, так и ложное успокоение при гиподиагностике. По расчётным оценкам, в США ежегодно выполняется от 500 тыс. до 1,5 млн биопсий молочной железы, из которых 70–80% приходится на доброкачественные образования [15, 18].

Эти ограничения формируют запрос на стандартизированные и воспроизводимые инструменты количественного анализа маммограмм, которые могли бы использоваться как объективное «второе мнение» или независимый предиктор категории BI-RADS. В настоящем исследовании такой подход реализован через радиомический анализ.

1.2.4. Инструментальное обеспечение маммографии и требования к цифровым изображениям

Качество маммографического изображения напрямую влияет и на визуальную интерпретацию, и на радиомический анализ. Для стандартизации технической оценки маммограмм применяется система PGMI (Perfect, Good, Moderate, Inadequate), которая учитывает позиционирование, компрессию, экспозицию и другие технические параметры. Снимки класса Inadequate исключаются из диагностического анализа из-за недостаточного качества для достоверной интерпретации [14].

В настоящем исследовании использовалось оборудование, распространённое в московских скрининговых программах: цифровой маммограф МТЛ «Маммо-4МТ-Плюс», маммограф МТЛ с томосинтезом «Маммо-5МТ» и маммограф General Electric Senographe Pristina. Все исследования выполнялись в стандартных проекциях СС и МЛО; результаты сохранялись в формате DICOM в системе ЕРИС ЕМИАС, что обеспечивало стандартизированный доступ к цифровым данным в рамках ретроспективного исследования [13].

Цифровой формат изображений – необходимое условие радиомического анализа: в отличие от плёночных маммограмм или их оцифрованных копий, нативные DICOM-файлы сохраняют полный динамический диапазон пикселей и метаданные, что обеспечивает воспроизводимость и сопоставимость извлекаемых количественных признаков [14]. Сочетание стандартизированного клинического протокола, цифровой архивации и верификации диагнозов в системе ЕРИС позволило сформировать набор данных настоящего исследования.

Маммография остаётся основным методом скрининга и диагностики РМЖ, но её интерпретация по системе BI-RADS демонстрирует существенную межэкспертную и интраэкспертную вариабельность. Работа над снижением субъективности визуального анализа ведётся в нескольких направлениях, и одно из них – радиомический анализ изображений, методологии которого посвящён следующий раздел.

1.3. Радиомика: концепция, методологический анализ и применение в онкологической радиологии

1.3.1. Концепция радиомики и её место в онкологической визуализации

Термин «радиомика» (radiomics) ввели Lambin P. и соавт. в 2012 году. Под ним понимается подход, при котором стандартные медицинские изображения подвергаются количественному анализу для извлечения большого количества математических дескрипторов – радиомических признаков, отражающих пространственно-интенсивностные характеристики изображённой ткани. Радиомика применяется к стандартным рутинным сканам из архивов лечебных учреждений, что позволяет использовать уже накопленные данные как потенциальные биомаркеры. Концептуальная предпосылка радиомики – гипотеза «онкологического фенотипирования», то есть неоднородность структуры опухоли, её поверхность, форма и тканевая организация, зафиксированные в пикселях изображения, содержат биологически значимую информацию о гистологии, агрессивности и молекулярном подтипе опухоли; эта информация недоступна при визуальной оценке, но может быть извлечена математически [5, 6].

В отличие от биопсии, оценивающей ограниченный участок потенциально гетерогенной опухоли, радиомика позволяет неинвазивно охарактеризовать весь объём образования. Клинические приложения в онкологии включают дифференциацию доброкачественных и злокачественных образований, предсказание гистологического подтипа, стадию оценку, прогнозирование ответа на лечение и выживаемости, мониторинг течения заболевания без повторных биопсий [8]. Радиомика применяется в разных модальностях: Компьютерная томография (КТ), магнитно-резонансная томография (МРТ), позитронно-эмиссионная томография, совмещенная с компьютерной томографией (ПЭТ/КТ),

рентгенография и маммография. Последняя – это предмет настоящего исследования [19].

1.3.2. Этапы радиомического анализа

Радиомический анализ – это последовательный анализ из нескольких этапов: получение и стандартизация изображений, сегментация области интереса, предобработка, извлечение признаков, их отбор, построение классификационной модели и валидация. Схематически этапы раомического анализа представлены на рисунке 1. Качество каждого шага напрямую влияет на воспроизводимость итоговой модели. Вариации протокола сканирования способны изменить значения текстурных признаков на 10–30%, а смещение границ области интереса на несколько пикселей – более чем на 20% [5, 6]. Для многоцентровой воспроизводимости признаков служат стандартизированные спецификации. Инициатива по стандартизации биомаркеров изображений (Image Biomarker Standardisation Initiative, IBSI), для методологической прозрачности публикаций. Конкретные технические параметры каждого этапа применительно к настоящему исследованию. Такие параметры, как нормализация интенсивностей, параметры дискретизации, схема отбора признаков, процедура кросс-валидации, подробно описаны в Главе 2.

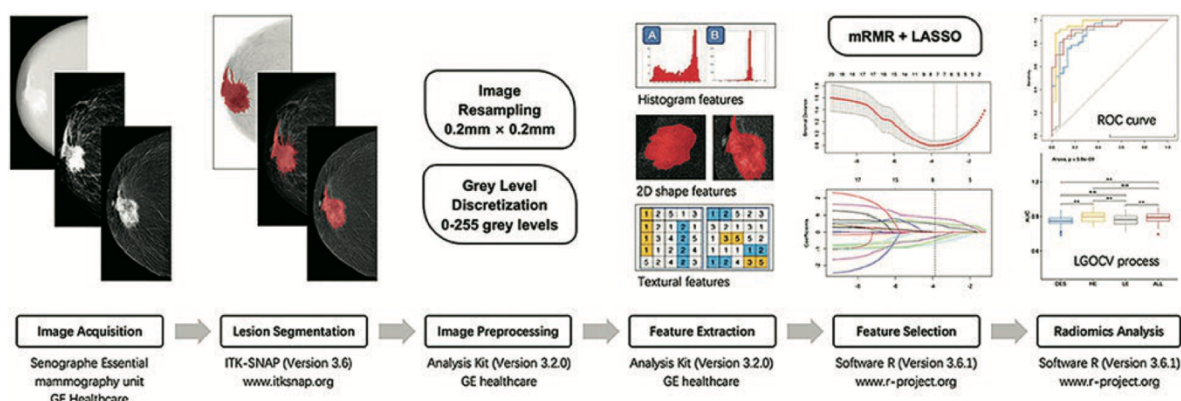


Рис. 1. Процесс радиомического анализа маммограмм.

Источник: https://www.frontiersin.org/files/Articles/600546/fonc-11-600546-HTML-r1/image_m/fonc-11-600546-g004.jpg

1.3.3. Классификация радиомических признаков

По классификации IBSI стандартизированы 169 радиомических признаков (число признаков соответствует стандарту библиотеки PyRadiomics v3), объединённых в пять категорий; на этой таксономии построены современные инструменты извлечения признаков, включая PyRadiomics как наиболее широко применяемую открытую платформу в области радиомики [6].

Признаки первого порядка (First-order statistics) описывают статистическое распределение интенсивности пикселей внутри области интереса без учёта их пространственных взаимоотношений. Класс включает 18 дескрипторов, таких как среднее, медиану, дисперсию, энтропию, асимметрию, эксцесс и другие, и считается наиболее воспроизводимым при вариациях протокола получения изображений.

Признаки формы (Shape features) характеризуют геометрию области интереса: площадь, периметр, максимальный диаметр, сферичность, компактность. Эти дескрипторы имеют прямое клиническое соответствие: неправильная форма образования и спикულიрованные края служат семиотическими признаками злокачественности в системе BI-RADS.

Три текстурных класса отражают пространственную организацию пикселей на разных уровнях детализации. GLCM (матрица сопряженности уровней серого) оценивает взаимоотношения пар соседних пикселей и описывает однородность текстуры через контраст, корреляцию, гомогенность и энтропию. GLRLM (матрица неоднородности уровней серого) характеризует направленную зернистость, то есть протяжённость последовательностей пикселей одинаковой интенсивности. GLSZM (матрица зон уровней серого) анализирует размер связных областей однородной интенсивности.

Основные классы признаков, их обозначения и диагностическая интерпретация приведены в таблице 2.

Таблица 2. Основные классы радиомических признаков и их интерпретация

Класс признаков	Аббревиатура	Что характеризует
Признаки первого порядка	First-order	Описывают распределение интенсивности пикселей/вокселей в изображении без учета их пространственного положения
Признаки формы	Shape-based	Описывают геометрические характеристики области интереса
Матрица сопряженности уровней серого	GLCM	Характеризует совместное распределение пар пикселей
Матрица неоднородности уровней серого	GLRLM	Описывает длины последовательностей пикселей с одинаковой интенсивностью
Матрица зон уровней серого	GLSZM	Характеризует зоны с одинаковым уровнем серого

1.3.4. Стандартизация радиомики

До недавнего времени отсутствие стандартизации в радиомическом анализе ограничивало клиническое внедрение метода. Разные программные пакеты при реализации формально идентичных признаков использовали разные алгоритмы вычисления и могли давать расходящиеся результаты на одних и тех же данных. В 2016 году была создана IBSI, сформулировавшая математически строгие спецификации для 169 радиомических признаков и предоставившая эталонные наборы данных и цифровые фантомы для проверки соответствия.

IBSI дополняет стандарты Quantitative Imaging Biomarkers Alliance (QIBA) (профили приобретения и реконструкции изображений) и DICOM (данные и метаданные) и к настоящему времени интегрирована в распространённые платформы – PyRadiomics и LIFEx, что делает многоцентровые рабочие процессы более стабильными. По независимым

оценкам, применение IBSI даёт наибольший выигрыш в воспроизводимости для признаков первого порядка и формы, то есть категорий, которые наиболее часто применяются в маммографических исследованиях [8]. В настоящей работе все признаки извлекались с помощью PyRadiomics v3 в IBSI-совместимой конфигурации, что обеспечивает сопоставимость результатов с другими исследованиями, выполненными по тем же спецификациям.

1.3.5. Ограничения и открытые вызовы радиомики

Радиомика сталкивается с рядом нерешённых проблем, ограничивающих её клиническое применение:

Первая – высокая размерность пространства признаков при относительно малых выборках. Это создаёт риск переобучения и может приводить к высокому AUC на обучающих данных при ограниченной обобщаемости модели. В маммографии объём доступных аннотированных данных обычно меньше, чем в КТ- и МРТ-исследованиях, что усиливает проблему.

Вторая – зависимость значений признаков от технических параметров исследования: производителя и модели аппарата, параметров экспозиции, алгоритмов обработки изображений. По данным ряда исследований, межаппаратная вариабельность может снижать воспроизводимость текстурных признаков до коэффициента каппы Коэна (ICC) менее 0,5 [20].

Третья – методологические ограничения части опубликованных работ: отсутствие внешней валидации, несообщённые параметры предобработки, избирательный выбор признаков. Для методологической прозрачности в литературе предложены стандартизированные чек-листы, в частности Checklist for Artificial Intelligence in Medical Imaging (CLAIM), их применение рекомендовано при подготовке публикаций по радиомике [21].

В настоящем исследовании с учётом перечисленных ограничений применена фильтрация признаков с удалением коррелирующих дескрипторов, стратифицированное разбиение выборки с кросс-валидацией, а также

классификаторы с применением логистической регрессии (ЛР) и случайного леса (СЛ), с последующим сопоставлением их производительности (подраздел 3.5).

1.4. Радиомический анализ маммограмм: обзор актуальных исследований

1.4.1. Общая характеристика направления

Радиомика маммографических изображений сформировалась как отдельное направление исследований в середине 2010-х годов; систематическое накопление работ с высоким методологическим качеством приходится преимущественно на период после 2019 года. Развитие области связано с цифровизацией маммографического скрининга – переходом к DICOM-архивам, что позволило стандартизированный ретроспективный анализ изображений без дополнительных лучевых нагрузок на пациенток. В отечественной литературе систематический обзор роли маммографии в радиомике рака молочной железы опубликован Говорухиной В.Г. и соавт. (2021), где авторы показали, что цифровая маммография является перспективной платформой для радиомических исследований и способна обеспечить высокоточную дифференциальную диагностику, хотя вопрос стандартизации анализа остаётся открытым [8].

По клинической задаче радиомические исследования на маммографическом материале можно разделить на три направления: дифференциация доброкачественных и злокачественных образований с гистологической верификацией как референсом, предсказание категории BI-RADS для поддержки принятия клинических решений и оценка свойств опухоли за пределами базовой классификации «доброкачественное против злокачественное». В настоящем исследовании реализованы первое и второе направления, поэтому ниже они рассмотрены детальнее.

1.4.2. Дифференциация доброкачественных и злокачественных образований

Бинарной классификации образований посвящена значительная часть публикаций по маммографической радиомике. Wang L. и соавт. (2022) построили модель на 14 признаках, отобранных из 188 исходных: на тестовой выборке она достигла AUC 0,901 (95% ДИ: 0,835–0,961) при чувствительности 95,8% и специфичности 66,7% в задаче идентификации злокачественных образований категорий BI-RADS 4–5 [22]. Сочетание признаков из CC- и MLO-проекций дало более стабильные результаты, чем использование каждой проекции отдельно: первого порядка и текстурные признаки CC и MLO были слабо коррелированы, что подтверждает информационную дополнительность двух стандартных проекций. Среди наиболее информативных признаков модели – морфологические дескрипторы и признаки неоднородности плотности и текстуры GLSZM. Это согласуется с данными о большей гетерогенности злокачественных образований по сравнению с доброкачественными.

Siviengphanom S. и соавт. применили радиомический анализ глобальных признаков (Global Radiomic Features, GRF) всей маммограммы, без сегментации конкретного образования, для предсказания труднодиагностируемых случаев. Комбинированная модель CC + MLO показала AUC 0,71 (95% ДИ: 0,64–0,78), что ниже, чем в исследованиях с сегментацией фокального образования, но это показывает, что в текстуре всей паренхимы молочной железы содержится диагностически значимая информация [23].

Namidinekoo A. и соавт. (2023) на выборке из 200 пациентов показали, что модели машинного обучения на радиомических признаках могут воспроизводить визуальную оценку рентгенолога: случайный лес показал наивысшее отношение правдоподобия среди протестированных методов. В число наиболее информативных признаков вошли морфологические и текстурные дескрипторы, а также возраст пациентки. Добавление

клинических переменных (возраст, плотность молочной железы) к радиомическим признакам улучшало диагностическую точность по сравнению с использованием одних только количественных признаков изображения [24 – 26].

1.4.3. Радиомика для поддержки стратификации BI-RADS и снижения числа биопсий

Одно из клинических применений радиомики – поддержка решения о биопсии, прежде всего в зоне BI-RADS 4, где вероятность злокачественности занимает широкий диапазон от 2% до 95% [27]. Interlenghi M. и соавт. (2022) разработали ансамблевую модель машинного обучения на основе радиомики ультразвуковых изображений (299 доброкачественных и 299 злокачественных образований), при пороге квалифицированного большинства (более 80% классификаторов голосуют за доброкачественность) можно было избежать от 15% до 18% биопсий доброкачественных образований при чувствительности выше 94% [28]. В 6 из 9 случаев модель перевела из BI-RADS 4 в BI-RADS 3 образования, оказавшиеся гистологически доброкачественными. Авторы интерпретируют этот результат, как более точную оценку риска по сравнению с исходной клинической оценкой.

Аналогичный подход на цифровой маммографии реализован в исследовании системы iBRISK: интегративная DL-модель для классификации образований BI-RADS 4 показала AUC 0,93 (95% ДИ: 0,92–0,95) и точность 89,5%; по расчётным оценкам авторов, её применение позволило бы избежать до 50% биопсий доброкачественных образований [29]. Этот результат связан с задачей настоящего исследования: категория BI-RADS 4 клинически наиболее неоднородна и именно для неё наиболее актуальна объективизация критериев направления на биопсию.

В ряде работ реализован подход к пересмотру категорий BI-RADS 3. По данным этих исследований, радиомические модели позволяют разграничивать внутри этой категории (риск злокачественности менее 2%) истинно

доброкачественные образования, которым достаточно наблюдения, и потенциально злокачественные, требующие более прицельной диагностики [30]. В исследовании глубокого радиомического анализа ультразвукового исследования (УЗИ) ($n = 746$) 27,7% образований BI-RADS 4a были переведены в BI-RADS 2, что позволило избежать биопсии.

1.4.4. Радиомика цифрового томосинтеза молочных желёз

Томосинтез (ТМЖ) даёт трёхмерные изображения и открывает доступ к объёмным радиомическим признакам, недоступным при двумерной маммографии. Клинико-радиомическое исследование Mazzola M. и соавт. (2023) на ТМЖ-данных с тремя методами отбора признаков показало AUC 0,72–0,74 (95% ДИ: 0,64–0,82) для дифференциации доброкачественных и злокачественных образований при скрининге [31]. По данным авторов, клинико-радиомические модели ТМЖ обладают приемлемой дискриминационной способностью уже на уровне первичного скринингового исследования и могут использоваться в качестве инструмента поддержки решений при первичной маммографии. Радиомика ТМЖ применяется и к прогнозированию: Liu J. и соавт. (2026) показали, что комбинированная ТМЖ и радиомическая модель даёт AUC 0,740–0,760 в валидационной выборке для предсказания двухлетней безрецидивной выживаемости [32].

1.4.5. Предсказание молекулярного подтипа и прогноза

За пределами бинарной классификации радиомика маммографических изображений применяется к предсказанию молекулярного подтипа РМЖ. Aker M. и соавт. (2024) показали, что на скрининговой цифровой маммографии вариация зоны является ведущим предиктором люминального А-подтипа, а повышенные значения того же признака для HER2-подтипа отражают характерную для HER2 неправильность краёв опухоли [33, 34]. Наиболее высокую предиктивную способность показывают комбинированные клинико-

радиомические модели, объединяющие клинические переменные (возраст, плотность молочной железы, семейный анамнез, данные предшествующих биопсий) с количественными признаками изображения.

Систематический обзор и метаанализ Li P. и соавт. (2025) объединил 47 исследований по радиомике и машинному обучению при РМЖ: среди методов машинного обучения наибольший суммарный AUC получили модели на основе метода опорных векторов Convolutional Neural Network (CNN) – свёрточная нейронная сеть (до 0,974), а традиционные радиомические методы – логистическая регрессия и случайный лес – показали AUC в диапазоне 0,73–0,92 в зависимости от модальности и задачи [7, 33]. Авторы отдельно отмечают, что традиционные методы превосходят глубокое обучение по интерпретируемости результатов, и это остаётся существенным критерием для клинического применения.

1.4.6. Ограничения опубликованных работ и обоснование настоящего исследования

В опубликованных работах, посвященных применению радиомики в маммографии, есть ряд ограничений, снижающих их трансляционный потенциал:

1. Большинство работ опирается на относительно небольшие ретроспективные выборки ($n = 100\text{--}500$), что ограничивает статистическую мощность и обобщаемость результатов.
2. Значительная часть исследований не проводят внешней валидации и ограничивается внутренней кросс-валидацией, что ограничивает оценку реальной обобщаемости модели.
3. Гетерогенность программных платформ, протоколов предобработки и методов отбора признаков затрудняет прямое сопоставление результатов разных работ [5, 20].
4. В работах, где целевой переменной выступает именно категория BI-RADS, невелико. Большинство исследований верифицирует исход

гистологически, тогда как использование BI-RADS-категории как метки ориентирует модель на ту клиническую задачу, которую рентгенолог решает при интерпретации маммограммы.

Настоящее исследование использует в качестве целевой переменной категорию BI-RADS из верифицированных заключений системы ЕРИС. Такой выбор целевой переменной соответствует клиническому решению, принимаемому при интерпретации маммограммы. В работе сопоставлены два классификатора: Least Absolute Shrinkage and Selection Operator (LASSO) логистическая регрессия и случайный лес, что позволяет оценить и предиктивную мощность, и интерпретируемость моделей.

1.5. Алгоритмы машинного обучения в задачах классификации медицинских изображений

1.5.1. Место машинного обучения в клинической радиологии

Машинное обучение (Machine Learning, ML) применительно к медицинским изображениям даёт инструменты для автоматизированного обнаружения паттернов в данных высокой размерности, недоступных при визуальном анализе [35, 36]. В контексте радиомики задачей ML является построение отображение между вектором количественных признаков изображения и клинически значимой меткой: категорией BI-RADS, фактом злокачественности или другим исходом. Среди доступных подходов, от линейных классификаторов до нейронных сетей, в клинической радиомике важную роль играют интерпретируемые методы. Логистическая регрессия и случайный лес дают не только прогноз, но и объяснение принятого решения, что повышает клиническое доверие к модели.

«Традиционная» радиомика, применяемая в настоящем исследовании, и методы глубокого обучения (Deep Learning, DL) различаются по принципу работы. В традиционной радиомике признаки извлекаются явно, по математически заданным формулам, после чего на них обучается

классификатор. DL-методы (свёрточные нейронные сети) формируют иерархические представления напрямую из пикселей изображения. DL-методы могут превосходить традиционную радиомику по предиктивной мощности при больших объёмах данных, но уступают ей по прозрачности и интерпретируемости. При умеренных объёмах данных ($n = 100\text{--}2000$), типичных для большинства радиомических исследований, преимущества DL уменьшаются, а риски переобучения возрастают – это один из аргументов в пользу выбора интерпретируемых алгоритмов в настоящей работе. Схематическое представление обработки данных с использованием классической радиомики и глубокого обучения представлены на рисунке 2.

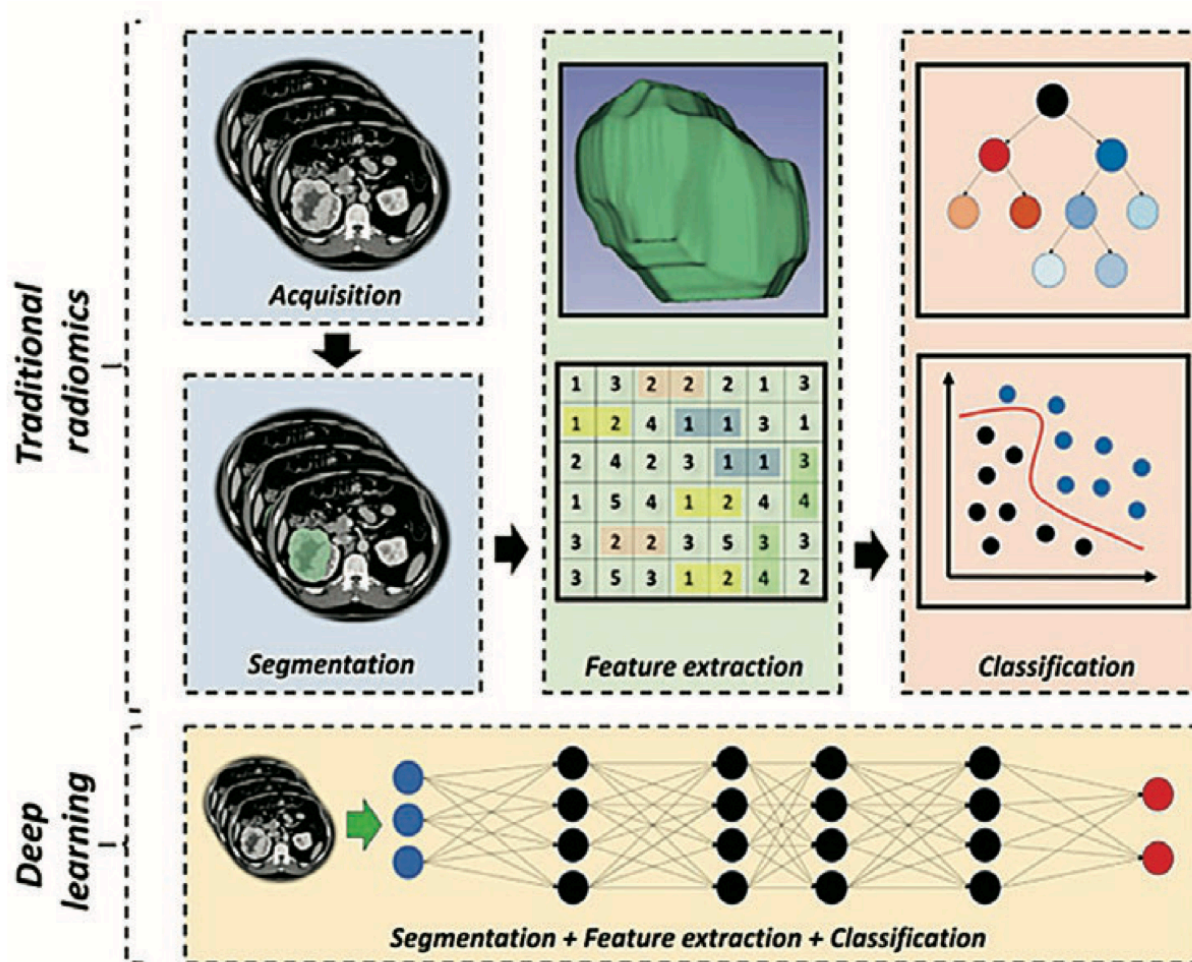


Рис. 2. Схематическое представление обработки данных с использованием классической радиомики и глубокого обучения.

Источник: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6837295/bin/DIR-25-6-485-g02.jpg>

1.5.2. Логистическая регрессия с регуляризацией LASSO

Логистическая регрессия – один из базовых методов бинарной классификации в медицинской статистике и клинической радиомике. Модель оценивает вероятность принадлежности наблюдения к положительному классу как логистическую функцию линейной комбинации признаков. Каждый весовой коэффициент определяет вклад соответствующего признака в прогноз, что позволяет ранжировать признаки по значимости и строить клинически объяснимые решения. Именно это отличает логистическую регрессию от нейросетевых архитектур и делает её предпочтительной там, где доверие к модели требует прозрачности принятого решения.

В условиях высокоразмерных радиомических данных нерегуляризованная логистическая регрессия склонна к переобучению. Для противодействия этому эффекту применяется (LASSO) (Tibshirani, 1996): метод добавляет к функции потерь штраф в форме L1-нормы вектора коэффициентов [37]. Ключевое свойство L1-штрафа – это способность обращать коэффициенты незначимых признаков строго в ноль, реализуя тем самым автоматический отбор признаков прямо в процессе обучения. По данным систематического обзора, LASSO применяется примерно в 44% радиомических исследованиях, в задачах маммографической классификации метод демонстрирует более высокий AUC по сравнению с нерегуляризованной регрессией при существенно меньшем числе итоговых признаков. Оптимальное значение параметра регуляризации λ определяется кросс-валидацией. В итоге LASSO-логистическая регрессия сочетает интерпретируемость с устойчивостью к переобучению – два свойства, редко достигаемые одновременно в высокоразмерных задачах.

1.5.3. Алгоритм случайного леса

Случайный лес – ансамблевый метод, в котором итоговый прогноз формируется усреднением предсказаний множества деревьев решений, обученных на независимых бутстрэп-подвыборках. Ключевая особенность метода – дополнительная случайность на уровне каждого узла. Для разбиения используется не весь набор признаков, а случайно выбранное подмножество. Это снижает корреляцию между деревьями ансамбля и уменьшает дисперсию финального предсказания [5].

В задачах радиомической классификации случайный лес устойчив к высокой размерности пространства признаков, выявляет нелинейные взаимодействия между предикторами и обеспечивает встроенную оценку важности признаков. По данным сравнительного бенчмарк-эксперимента, случайный лес превосходил логистическую регрессию по AUC на 72,3% наборов данных [38]. Преимущество наиболее выражено при нелинейных зависимостях и коррелированных предикторах.

1.5.4. Сравнительная характеристика алгоритмов в контексте задачи

Использование двух методологически разных алгоритмов ЛР и СЛ нужно для сравнения их производительности на одном и том же наборе данных. Если ЛР показывает сопоставимый с СЛ AUC, это свидетельствует о преимущественно линейном характере зависимостей «признак и BI-RADS», что упрощает клиническую интерпретацию [5, 6]. Если СЛ существенно превосходит ЛР, это указывает на значимые нелинейные взаимодействия, и применение более сложной модели оправдано предиктивно.

1.5.5. Метрики оценки качества классификации

Оценка качества классификационных моделей в клинических приложениях не сводится к точности. Эта метрика не учитывает различия в

стоимости ошибок первого и второго рода и нечувствительна к дисбалансу классов, характерному для онкологических данных [21]. В настоящем исследовании применялся набор метрик, рекомендованный для клинической диагностической валидации.

Площадь под ROC-кривой (AUC) – основная метрика дискриминационной способности модели. Она равна вероятности того, что модель присвоит более высокий балл случайно выбранному положительному наблюдению, чем случайно выбранному отрицательному. AUC не зависит от выбора порога классификации и устойчива к умеренному дисбалансу классов, поэтому она применяется как стандартная метрика в радиомических исследованиях. Общепринятая градация AUC: 0,90–1,00 – отличная дискриминация, 0,80–0,90 – хорошая, 0,70–0,80 – удовлетворительная [21, 38].

Чувствительность (Se) – доля, верно, классифицированных положительных случаев ($Se = TP / (TP + FN)$). В онкологическом скрининге это приоритетная метрика, так как ложноотрицательные результаты несут наибольший клинический риск. Специфичность (Специфичность, Sp) – доля, верно, классифицированных отрицательных случаев ($Sp = TN / (TN + FP)$). Она определяет частоту ложных тревог и ненужных биопсий. Совместная оценка Se и Sp при нескольких порогах классификации формирует ROC-кривую.

F1-мера (гармоническое среднее точности и полноты: $F1 = 2 TP / (2 TP + FP + FN)$) применима при дисбалансе классов. В отличие от AUC, она чувствительна к производительности модели именно на положительном классе [21].

С учётом рекомендаций чек-листа CLAIM 2024 (Tejani A.S. и соавт., Radiology: Artificial Intelligence) в настоящем исследовании выполнено: описание процесса формирования набора данных с указанием критериев включения и исключения, описание протокола разбиения данных, отчёт о метриках модели с 95% доверительными интервалами, описание методов отбора признаков и настройки гиперпараметров; обсуждение ограничений с акцентом на отсутствие внешней валидации [21].

ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ

2.1. Материал и дизайн исследования

2.1.1. Тип и дизайн исследования

Проведено ретроспективное, одноцентровое исследование на архивных данных системы цифровой маммографии. Ретроспективный дизайн выбран, потому что разработка и валидация радиомических моделей классификации требует верифицированных диагнозов для формирования меток целевой переменной, а это возможно только при работе с данными с уже завершёнными клиническими исходами. Такой дизайн принят в литературе как стандартный для радиомических исследований и применяется при первичной апробации методологического анализа до проспективной валидации [7, 20].

Исследование выполнено в соответствии с принципами Хельсинкской декларации Всемирной медицинской ассоциации (редакция 2013 года) [39]. Использование анонимизированных архивных данных для задач оценки качества медицинской помощи и разработки диагностических алгоритмов не требует индивидуального информированного согласия пациентов в соответствии с действующим законодательством РФ [12].

2.1.2. Источник данных

Набор данных сформирован из цифровых маммографических исследований, хранящихся в Едином радиологическом информационном сервисе (ЕРИС) Единой медицинской информационно-аналитической системы (ЕМИАС) Департамента здравоохранения города Москвы [13]. ЕРИС – централизованное облачное хранилище медицинских изображений в формате DICOM, интегрированное с рабочими местами рентгенологов московских медицинских организаций. Оно обеспечивает ретроспективный исследовательский доступ к верифицированным исследованиям [9, 13].

Период формирования выборки из маммографических изображений, полученных с 2020 по 2021 годы. Выбор временного диапазона определялся двумя соображениями. К 2020 году все включённые медицинские организации перешли на цифровое маммографическое оборудование, дающее нативные DICOM-файлы необходимого качества [13, 14]. Кроме того, к этому периоду в системе ЕРИС был накоплен достаточный объём верифицированных исследований с заключениями по системе BI-RADS, пригодных для формирования меток целевой переменной [9].

2.1.3. Критерии включения и исключения

В исследование включались маммографические исследования, удовлетворявшие следующим критериям.

Критерии включения:

1. Пол: женщины;
2. Возраст: 40 лет и старше на момент проведения исследования;
3. Наличие снимков в обеих стандартных проекциях – краниокаудальной (CC) и медиолатеральной косой (MLO);
4. Период проведения исследования – 2020–2021 гг.;
5. Наличие верифицированного заключения с категорией BI-RADS (2, 3, 4 или 5);
6. Верификация характера образований: злокачественность – гистологически; доброкачественность – согласованным заключением двух экспертов-рентгенологов.

Критерии исключения:

1. Категория BI-RADS 1 (норма без находок: образования для разметки отсутствуют);
2. Техническое качество снимков класса Inadequate (I) по системе PGMi (Perfect, Good, Moderate, Inadequate);
3. Категории BI-RADS 0 и BI-RADS 6 в наборе данных не встречались.

Снимки класса Inadequate (I) исключались, потому что они не соответствуют минимальным требованиям диагностического качества и непригодны для достоверной интерпретации как визуальной, так и радиомической. Их включение привнесло бы в обучающую выборку систематическое смещение, связанное с техническими артефактами, а не с биологическими свойствами ткани.

2.1.4. Характеристика итоговой выборки

После применения критериев включения и исключения итоговая выборка составила 83 пациентки. Общее число образований, выявленных и размеченных в ходе ручной сегментации – 1391. Среднее число образований на пациентку – 16,8. Такая величина связана с преобладанием в выборке диффузных кальцинатов и сосудистых обызвествлений – множественных по своей природе находок, каждая из которых размечалась как отдельная область интереса (ROI).

Чтобы исключить утечку данных между обучающей и тестовой выборками, разбиение проводилось на уровне пациенток, а не образований. Из 83 пациенток 66 (79,5%) случайно отнесены к обучающей выборке, 17 (20,5%) – к тестовой; разбиение стратифицировано по классу. При таком разбиении ни одно образование из тестовой пациентки не попадало в обучающую выборку, что исключает псевдорепликацию и переоптимистичную оценку метрик.

Распределение образований по категориям BI-RADS и типам приведено в таблице 3.

**Таблица 3. Типы образований относительно категории BI-RADS,
полученные после ручной сегментации в CVAT.**

Категория BI-RADS	Тип образования	Количество новообразова ний	Процент от общего количества (%)
2	Диффузные кальцинаты	1063	76,4
	Обызвествление стенок сосудов	168	12,1
	Интрамаммарные лимфатические узлы	44	3,2
	Простые кисты (без воспаления)	7	0,5
3	Единичные группы точечных микрокальцинатов	14	1
	Новообразования с чёткими ровными контурами	8	0,6
4	Подозрительные новообразования	56	4
	Группы микрокальцинатов	18	1,3
5	Злокачественные новообразования	13	0,9
ИТОГО		1391	100

Выборка несбалансирована по классам: на BI-RADS 2 приходится 92,2% всех наблюдений, на BI-RADS 3, 4 и 5 в сумме – 7,8%. Это распределение соответствует реальной структуре находок при маммографическом скрининге; при построении классификационных моделей дисбаланс учитывался через взвешивание классов (`class_weight='balanced'`) [36, 38].

2.2. Подготовка изображений

2.2.1. Конвертация изображений из формата DICOM в PNG

Исходные маммографические изображения хранились в системе ЕРИС в формате DICOM, что является стандартом хранения и передачи медицинских изображений. Для радиомического анализа и последующей ручной разметки в

специализированном программном обеспечении DICOM-файлы конвертировались в формат PNG (Portable Network Graphics).

Конвертация выполнялась с сохранением полного динамического диапазона яркостей пикселей: параметры окна отображения (Window Width и Window Center) подбирались индивидуально для каждого исследования так, чтобы обеспечить читаемую визуализацию тканей молочной железы. Дополнительное сжатие с потерями при конвертации не применялось. PNG – формат без потерь качества и сохраняет пиксельные значения исходного DICOM-файла без искажений, что важно для воспроизводимости радиомического анализа.

Каждое маммографическое исследование включало 4 снимка: левая молочная железа в проекции CC, правая – в проекции CC, левая – в проекции MLO, правая – в проекции MLO обозначены CC_L, CC_R, MLO_L, MLO_R, соответственно.

2.2.2. Инструмент разметки: Computer Vision Annotation Tool

Ручная сегментация образований выполнялась в программном обеспечении Computer Vision Annotation Tool (CVAT) – открытой веб-платформе для аннотирования изображений. Функциональность CVAT соответствовала задачам настоящего исследования: поддержка аннотации с кодированием масок в формате Run-Length Encoding (RLE), гибкая система лейблов, экспорт аннотаций в структурированном XML-формате [40].

Для каждой пациентки коллаж загружался в CVAT как отдельное задание. Для задания формировалась библиотека лейблов, соответствующих девяти типам образований, выделенным в исследовании согласно категориям BI-RADS (таблица 3). Каждое образование размечалось бинарной маской, плотно охватывающей его контуры с учётом видимых границ в обеих проекциях.

Разметка выполнялась лично автором работы по предварительно разработанному внутреннему руководству по аннотированию. Руководство

определяло критерии границ образований для каждого из девяти типов лейблов, было составлено с опорой на действующее 5-е издание атласа ACR BI-RADS и согласовано с экспертами-рентгенологами до начала основной разметки [3, 4].

2.2.3. Контроль качества разметки

Контроль качества разметки включал два уровня:

Первый уровень – самопроверка автора. Каждое размеченное задание повторно проверялось автором не ранее чем через 48 часов после первичной разметки. Для количественной оценки внутриэкспертной воспроизводимости случайным образом отобрана подвыборка из 15 пациенток (195 образований), которые были размечены повторно через 14 дней в условиях «слепого» пересмотра. Согласие Коэна k для итоговой принадлежности контура к категории лейбла составило $k = 0,74$ (95% ДИ: 0,68–0,80). Средняя степень перекрытия масок (Intersection over Union, IoU) для повторно размеченных образований – 0,81, что указывает на стабильность позиционирования контуров.

Второй уровень – экспертная верификация. 20% заданий (случайная выборка) проверялись экспертом-рентгенологом с опытом интерпретации маммограмм не менее 10 лет. Случаи с расхождением выше допустимого порога ($\text{IoU} < 0,65$) возвращались автору на повторную разметку. Финальной разметка образования считалась после прохождения обоих уровней контроля.

2.2.4. Экспорт аннотаций

По завершении разметки аннотации экспортировались из CVAT в формате XML. Для каждого коллажа XML-файл содержал перечень размеченных образований со следующими атрибутами: наименование лейбла, координаты ограничивающего прямоугольника, бинарная маска в кодировке RLE. Структура XML-файла однозначно сопоставляет каждую маску с

конкретным изображением-коллажом и типом образования; этот файл – исходные данные для следующего этапа, извлечения радиомических признаков.

2.3. Радиомический анализ: извлечение, предобработка и отбор признаков

2.3.1. Общая схема радиомического анализа

Радиомический анализ настоящего исследования состоял из пяти последовательных этапов:

1. Декодирование RLE-масок из XML-файла аннотаций CVAT и восстановление бинарных масок образований в виде двумерных массивов;
2. Загрузка изображений-коллажей и вырезание ROI для каждого образования по координатам ограничивающего прямоугольника;
3. Предобработка изображения и маски;
4. Извлечение радиомических признаков с помощью PyRadiomics;
5. Формирование итоговой таблицы признаков с фильтрацией нерелевантных дескрипторов и строк с пропущенными значениями.

Все этапы реализованы на языке программирования Python с использованием библиотек NumPy, pandas, SimpleITK и Pyradiomics.

2.3.2. Настройки экстрактора признаков

Извлечение радиомических признаков выполнялось с помощью библиотеки PyRadiomics версии 3. Для каждого образования ROI-изображение и соответствующая бинарная маска преобразовывались в объекты SimpleITK.Image с сохранением метаданных пространственного расположения исходного изображения.

Конфигурация экстрактора задавалась тремя параметрами предобработки:

1. `binWidth = 25` – шаг группировки яркостных уровней. Перед тем как считать текстурные признаки (GLCM, GLRLM и др.), пиксельные значения яркости, близкие значения объединяются в группы. Шаг в 25 единиц – это умеренный уровень обобщения: достаточно крупный, чтобы случайный шум не влиял на результат, но достаточно мелкий, чтобы реальная текстурная структура ткани не потерялась.
2. `normalize = True` – нормализация яркости перед анализом. Три маммографа, которые использовались в исследовании, дают изображения с разными абсолютными значениями яркости – из-за различий в настройках экспозиции и программной обработке. Нормализация приводит к общему знаменателю яркость на всех снимках, убирая эти аппаратные различия.
3. `normalizeScale = 100` – диапазон значений после нормализации. После нормализации все яркости пересчитываются в шкалу от 0 до 100. Это чисто практический приём, то есть признаки, извлечённые с разных аппаратов, становятся численно сопоставимы друг с другом. Такой подход рекомендован в стандартной конфигурации PyRadiomics для многоцентровых и мультиаппаратных исследований.

2.3.3. Классы извлекаемых признаков

Для каждого образования извлекались признаки семи классов:

1. Признаки первого порядка (**First-order statistics**) – статистическое распределение интенсивностей пикселей внутри ROI без учёта пространственных взаимоотношений: среднее, медиана, дисперсия, асимметрия, эксцесс, энтропия и др.;

2. Признаки формы в двумерном представлении (**Shape 2D**) – геометрия контура ROI: площадь, периметр, максимальный диаметр, сферичность, компактность;
3. Матрица смежности серых уровней (**GLCM**) – пространственные взаимоотношения пар пикселей определённой интенсивности; признаки характеризуют однородность и неоднородность текстуры;
4. Матрица длин серых прогонов (**GLRLM**) – последовательности пикселей одинаковой интенсивности в заданных направлениях; отражает направленную зернистость текстуры;
5. Матрица серых зон (**GLSZM**) – зоны связанных пикселей одинаковой интенсивности; характеризует размер однородных областей;
6. Матрица зависимостей серых уровней (**GLDM**) – степень зависимости центрального пикселя от его соседей;
7. Матрица разностей соседних тонов (**NGTDM**) – визуальная текстурная грубость и бесструктурность изображения.

Совокупный вектор признаков для каждого образования включал 107 числовых дескрипторов. Помимо, собственно, радиомических признаков, экстрактор записывал диагностические метаданные PyRadiomics, содержащие технические характеристики ROI: число вокселей и объём маски. По результатам одномерного анализа эти дескрипторы вошли в число наиболее информативных.

2.3.4. Формирование итоговой таблицы признаков

Результаты извлечения признаков для всех образований объединялись в единый набор данных. Каждая строка соответствовала одному образованию и содержала: идентификатор изображения, лейбл типа образования, диагностические метаданные и 107 радиомических признаков. Итоговая таблица сохранялась в формате Microsoft Excel для последующего анализа.

На этапе постобработки из таблицы признаков исключались:

1. Нечисловые столбцы;
2. Столбцы-константы с нулевой дисперсией;
3. Диагностические столбцы PyRadiomics, не несущие биологически значимой информации;
4. Строки с пропущенными значениями признаков.

После применения фильтров формировались финальная матрица признаков и вектор целевой переменной, использовавшиеся при построении классификационных моделей.

2.3.5. Ранжирование признаков по F-критерию Фишера

Для предварительной оценки дискриминационной способности признаков применялся одномерный F-критерий Фишера (F-score). Он оценивает отношение межклассовой дисперсии к внутриклассовой дисперсии для каждого признака отдельно: признаки с высоким F-Score показывают более выраженное расслоение по группам BI-RADS и потенциально несут больше дискриминационной информации.

В первой двадцатке наиболее информативных признаков доминировали дескрипторы класса NonUniformity из матриц GLRLM, GLSZM и GLDM, а также ряд признаков GLCM, характеризующих однородность и энтропию текстуры. Топ-20 признаков по F-критерию Фишера приведён на рисунке 3.

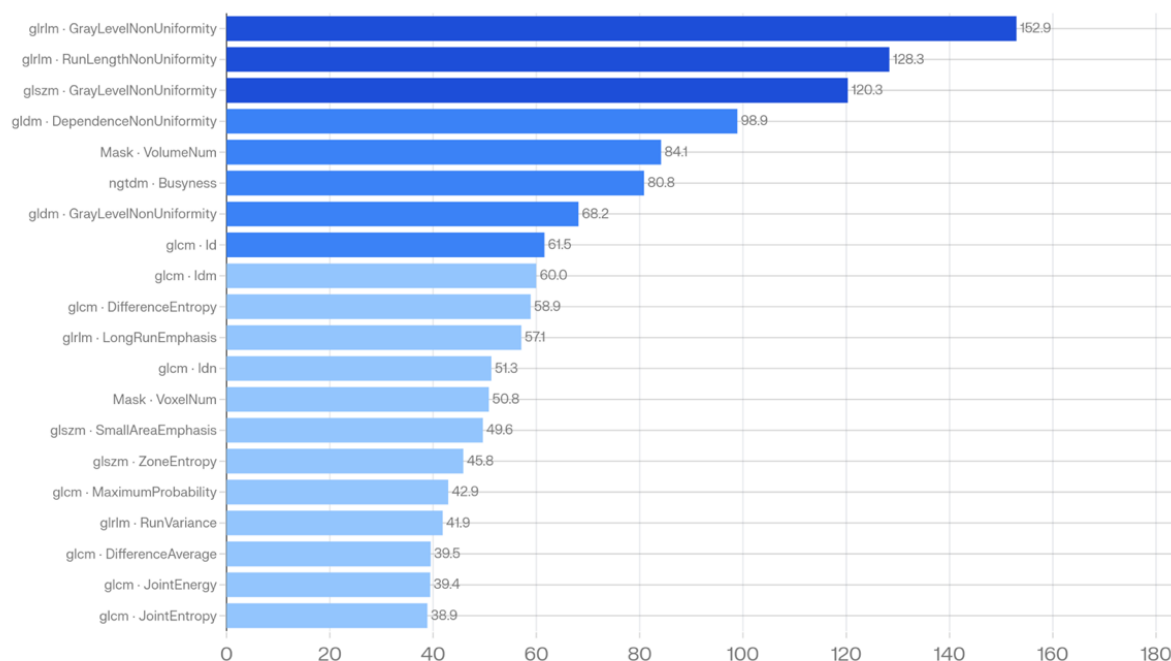


Рис. 3. Топ-20 радиомических признаков по F-критерию Фишера

Одномерный F-критерий использовался только для предварительного описания информативности признаков. Все значения F-Score получены на всей выборке без разбиения на обучающую и тестовую части, поэтому такое ранжирование не может служить единственным критерием отбора признаков для моделей. Окончательный отбор признаков выполнялся в рамках процедуры подбора гиперпараметров с кросс-валидацией, описанной в подразделе 2.5.

Расчёт доверительных интервалов. Доверительные интервалы для AUC рассчитывались методом бутстрэп с применением перцентильного метода. Доверительные интервалы для долей чувствительности, специфичности, точности, ПЦПР, прогностической ценности отрицательного результата (ПЦОР) рассчитывались методом Уилсона, который обеспечивает надёжное покрытие при малых и несбалансированных выборках [38, 41]. Для коэффициента корреляции Мэтьюса (MCC) доверительный интервал также рассчитывался бутстрэпом. Во всех таблицах результатов MCC приводится в единицах от -1 до 1 (не в процентных), доверительные интервалы для MCC выражены в тех же единицах.

2.4. Постановка задач классификации и формирование целевых переменных

2.4.1. Общий подход к формированию целевой переменной

Целевая переменная для всех классификационных стратегий формировалась из категории BI-RADS, присвоенной каждому образованию при ручной разметке в CVAT. Лейбл типа образования однозначно определял принадлежность к одной из четырёх категорий шкалы BI-RADS (2, 3, 4 или 5), в соответствии с классификационной таблицей, согласованной с экспертом-рентгенологом до начала разметки (таблица 3).

Для целевой переменной выбрана категория BI-RADS, а не бинарный исход «доброкачественное против злокачественное». Такой выбор связан с ролью BI-RADS в клинической практике. Категория BI-RADS определяет тактику ведения пациентки – наблюдение, краткосрочный контроль или биопсия (подраздел 1.2.2). Модели, обученные выводить категории BI-RADS, дают прогноз в тех же единицах, которыми описывается клиническое решение, что упрощает их интерпретацию рентгенологом и сопоставление с верифицированными заключениями [7, 24, 29].

В работе реализованы четыре стратегии классификации. Они различаются составом классов и постановкой задачи. Схемы кодирования целевой переменной для каждой стратегии различались по двум основаниям: методологическому (исследование влияния состава классов на качество моделей) и клиническому (соответствие постановки задачи клинической логике BI-RADS).

2.4.2. Стратегия 1: многоклассовая классификация по полной шкале BI-RADS

В первой стратегии целевая переменная принимала четыре значения, соответствующие категориям BI-RADS 2, 3, 4 и 5. Эта постановка наиболее

близка к клинической задаче, которую рентгенолог решает при интерпретации маммограммы: присвоить категорию по полной шкале. Вместе с тем многоклассовая задача методологически самая сложная из четырёх стратегий из-за выраженного дисбаланса классов: доля BI-RADS 2 – 92,2%, BI-RADS 3 – 1,6%, BI-RADS 4 – 5,3%, BI-RADS 5 – 0,9%.

Использовался случайный лес с балансировкой, при котором веса классов в функции потерь обратно пропорциональны их частоте в обучающей выборке. Схема кодирования: BI-RADS 2 это класс 2, BI-RADS 3 это класс 3, BI-RADS 4 это класс 4, BI-RADS 5 это класс 5.

2.4.3. Стратегия 2: трёхклассовая классификация с исключением BI-RADS 3

Вторая стратегия – трёхклассовая задача классификации по категориям BI-RADS 2, 4 и 5. Категория BI-RADS 3 исключена из анализа по результатам предварительной стратегии, в которой модель показывала нулевую точность распознавания этого класса. Причина ожидаемая, так как класс BI-RADS 3 представлен только 22 образованиями (1,6% выборки), что мало для устойчивого обучения классификатора. Дополнительно категория BI-RADS 3 по клинической логике занимает промежуточное положение между BI-RADS 2 и BI-RADS 4 (подраздел 1.2.2), что может создавать перекрытие радиомических профилей с соседними классами и затруднять их разделение.

Класс BI-RADS 3 исключался отдельно в обучающей и тестовой частях выборки – отбором строк, у которых значение целевой переменной не равно 3. Использовался случайный лес с балансировкой. Схема кодирования: BI-RADS 2 – это класс 2, BI-RADS 4 – это класс 4, BI-RADS 5 – это класс 5.

2.4.4. Стратегия 3: двухклассовая классификация с применением логистической регрессии

Третья стратегия – двухклассовая задача. Исходная четырёхуровневая переменная BI-RADS преобразовывалась в двухклассовую по следующей схеме: образования BI-RADS 2 это класс 0 (нет клинически значимой патологии, требующей дополнительных вмешательств), образования BI-RADS 3, 4 и 5 это класс 1 (клинически значимые находки, требующие наблюдения или инвазивной верификации). Такое объединение соответствует тактическому решению в маммографическом скрининге: плановый скрининг (BI-RADS 2) противопоставляется более активной тактике – наблюдению, краткосрочному контролю или биопсии (BI-RADS 3, 4, 5) (таблица 1, подраздел 1.2.2).

В качестве классификатора использовалась логистическая регрессия. Признаки предварительно масштабировались StandardScaler.

Логистическая регрессия выбрана как линейный интерпретируемый классификатор: коэффициенты модели можно ранжировать по значимости и сопоставлять с конкретными радиомическими дескрипторами (подраздел 1.5.2).

2.4.5. Стратегия 4: двухклассовая классификация с применением случайного леса

Четвёртая стратегия воспроизводит постановку задачи третьей стратегии с той же схемой кодирования целевой переменной: BI-RADS 2 – это класс 0, BI-RADS 3, 4 и 5 – это класс 1. При этом меняются два методологических параметра: алгоритм классификатора и метод масштабирования признаков.

В качестве классификатора использовался случайный лес, что позволяло напрямую сопоставить нелинейный ансамблевый метод с линейной логистической регрессией на идентичной двухклассовой задаче. Сравнение

стратегий 3 и 4 позволяет одновременно оценить два фактора: влияние алгоритма (логистическая регрессия против случайный лес) и влияние метода масштабирования признаков. Эти факторы разделены при интерпретации результатов в разделе 3.5.5.

2.4.6. Сводная характеристика стратегий классификации

Характеристики всех четырёх стратегий приведены на рисунке 4.

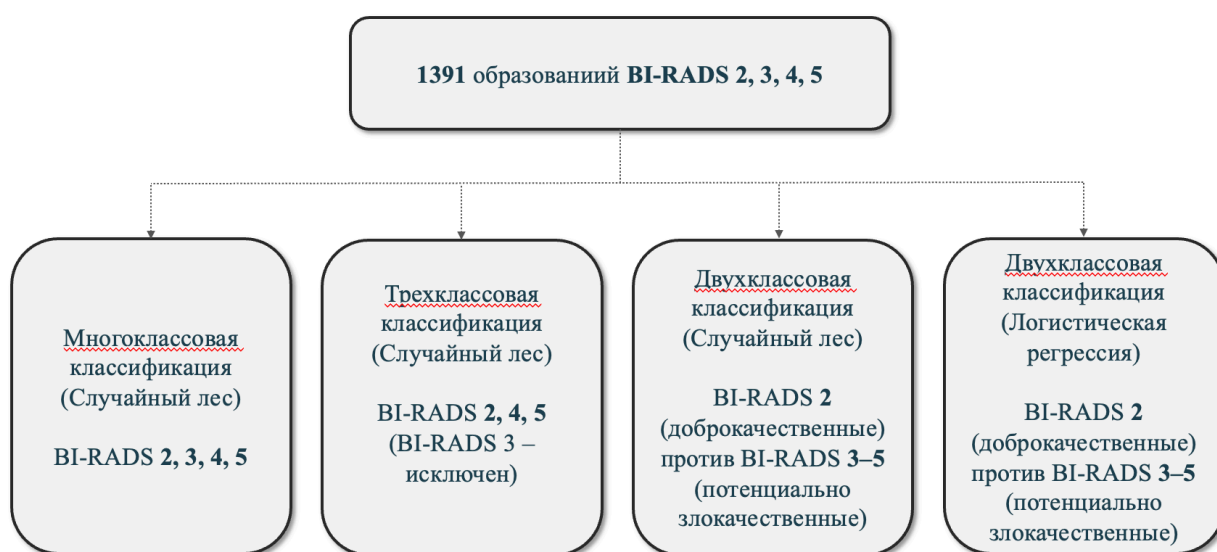


Рис. 4. Сводная характеристика стратегий классификации

Во всех четырёх стратегиях использовалось единое стратифицированное разбиение исходной выборки на обучающую (80%) и тестовую (20%) части с фиксированным `random_state=42`, что обеспечивает воспроизводимость результатов. Стратификация по целевой переменной сохраняла пропорции классов в обеих частях выборки, что существенно при выраженном дисбалансе в настоящем наборе данных.

2.5. Машинное обучение: модели, гиперпараметры, валидация и метрики оценки

2.5.1. Общая схема построения классификационных моделей

Все четыре стратегии классификации реализованы в едином методологической схеме. Разбиение выборки выполнено на уровне пациенток, а не образований, для исключения утечки данных: 66 пациенток (79,5%) отнесены к обучающей выборке, 17 (20,5%) к тестовой, разбиение стратифицировано по целевой переменной.

Алгоритмы. В стратегиях 1, 2 и 4 применялся СЛ, в стратегии 3 – логистическая регрессия ЛР. Использование двух методологически различных алгоритмов, нелинейного ансамблевого и линейного, позволяет оценить характер зависимостей между признаком и меткой. Сопоставимый AUC у ЛР и СЛ свидетельствует о линейных зависимостях и упрощает клиническую интерпретацию модели. Во всех стратегиях применялся параметр `class_weight='balanced'` для учёта выраженного дисбаланса классов, так как BI-RADS 2 занимает 92,2% набора данных образований.

Масштабирование признаков. В стратегиях 1, 2 и 3 применялся `StandardScaler` (нормализация по среднему и стандартному отклонению), в стратегии 4 – `MinMaxScaler` (приведение к диапазону). Масштабирование выполнялось строго внутри каждого фолда кросс-валидации: параметры трансформации рассчитывались на обучающей части и применялись к тестовой, что исключает утечку информации между выборками.

Настройка гиперпараметров. Подбор проводился методом полного перебора `GridSearchCV` с трёхкратной стратифицированной кросс-валидацией (`cv=3`) на обучающей выборке, критерием оптимизации стал параметр `F1-macro`. Для СЛ перебирались число деревьев и максимальная глубина, для ЛР применялся параметр регуляризации. Значение `cv=3` выбрано как компромисс между статистической устойчивостью оценки и сохранением достаточного

числа наблюдений редких классов (BI-RADS 4, 5) в каждом фолде. Оптимальные гиперпараметры по каждой стратегии приведены в таблице результатов (раздел 3.1). Итоговая модель автоматически переобучалась GridSearchCV на полной обучающей выборке и применялась к тестовой.

2.5.2. Метрики оценки качества классификации

Оценка метрик диагностической точности: чувствительность, специфичность, точность, площадь под ROC-кривой (AUC), площадь под Precision-Recall-кривой (AP), коэффициент Мэтьюса, прогностическая ценность положительного результата (ПЦПР), прогностическая ценность отрицательного результата (ПЦОР) с 95% доверительным интервалом (ДИ).

ГЛАВА 3. РЕЗУЛЬТАТЫ РАДИОМИЧЕСКОГО АНАЛИЗА И КЛАССИФИКАЦИИ МАММОГРАФИЧЕСКИХ ОБРАЗОВАНИЙ

3.1. Результаты стратегии 1 – многоклассовая классификация (случайный лес)

3.1.1. Характеристика задачи и обучающей выборки

В стратегии 1 решалась многоклассовая задача: модель обучалась различать четыре категории BI-RADS – 2, 3, 4 и 5, с целевой переменной, принимающей значения 0 (BI-RADS 2), 1 (BI-RADS 3), 2 (BI-RADS 4), 3 (BI-RADS 5).

Распределение классов в обучающей выборке: BI-RADS 2 – 92,2%, BI-RADS 3 – 1,6%, BI-RADS 4 – 5,3%, BI-RADS 5 – 0,9% (подраздел 2.1.4). Для компенсации дисбаланса использовалось взвешивание классов `class_weight='balanced'`: вес каждого класса в функции потерь обратно пропорционален его частоте в обучающей выборке [13, 38].

3.1.2. Подбор гиперпараметров

По результатам GridSearchCV с трёхкратной кросс-валидацией (`cv=3`, критерий – F1-macro; см. подраздел 2.5.1) выбраны гиперпараметры случайного леса: `n_estimators=300`, `max_depth=None` (рост деревьев до чистых листьев).

3.1.3. Метрики диагностической точности

Бинарные метрики стратегии 1 приведены ниже; сводное сопоставление со стратегиями 2, 3 и 4 в таблице 4 (раздел 3.5.1).

- **Чувствительность** – 91,7% (95% ДИ: 84,3–96,4). Из 109 злокачественных образований, верно идентифицированы 100, ложноотрицательных – 9;

- **Специфичность** – 72,0% (95% ДИ: 69,3–78,1). Из 1280 доброкачественных образований, верно идентифицированы 913, ложноположительных – 367;
- **Точность** – 73,5% (95% ДИ: 71,1–79,0);
- **F1-мера** – 35,2% (95% ДИ: 30,5–42,2);
- **ПЦПР** – 21,8% (95% ДИ: 18,3–27,5);
- **ПЦОР** – 99,0% (95% ДИ: 98,2–99,6);
- **AUC** – 0,861 (95% ДИ: 0,827–0,890);
- **МСС** – 0,364 (95% ДИ: 0,321–0,424). Метрика учитывает все четыре элемента матрицы ошибок и при дисбалансе классов более устойчива, чем точность.

3.1.4. Анализ ошибок классификации

Сочетание $AUC = 0,861$ и $F1 = 35,2\%$ типично для несбалансированных задач: AUC характеризует способность модели ранжировать образования по вероятности класса злокачественных образований, тогда как $F1$, чувствительность и специфичность рассчитываются для конкретного порога 0,332. При фиксированном пороге часть BI-RADS 2 неизбежно попадает в класс злокачественных, что наблюдается в 367 ложноположительных результатах.

В исходной четырёхклассовой постановке распознавание категории BI-RADS 3 (22 образования, 1,6% выборки) на тестовой части было нулевым. Этот факт служил эмпирическим основанием для исключения BI-RADS 3 в стратегии 2 (подраздел 2.4.3). Альтернативные содержательные объяснения нулевой полноты – малый объём класса и промежуточное положение BI-RADS 3 между BI-RADS 2 и BI-RADS 4 по клинической логике (подраздел 1.2.2), обсуждаются, как гипотезы, требующие дополнительной проверки на большей выборке.

3.2. Результаты стратегии 2 – трёхклассовая классификация (случайный лес)

3.2.1. Характеристика задачи и обучающей выборки

Стратегия 2 повторяет принцип работы стратегии 1 с одним содержательным изменением, где из целевой переменной исключена категория BI-RADS 3. Эмпирическим основанием для исключения служила нулевое распознавание BI-RADS 3 в стратегии 1 (см. подраздел 3.1.3). Задача переформулирована как трёхклассовая. После исключения 22 образований BI-RADS 3 итоговая выборка составила 1367 образований.

Распределение по классам в полной выборке: BI-RADS 2 – 1280 образований ($1280/1367 = 93,6\%$), BI-RADS 4 + BI-RADS 5 – 87 образований ($87/1367 = 6,4\%$). Взвешивание классов задавалось параметром `class_weight='balanced'` аналогично стратегии 1.

3.2.2. Подбор гиперпараметров

По результатам GridSearchCV с трёхкратной кросс-валидацией (`cv=3`, критерий – F1-масго; см. подраздел 2.5.1) выбраны те же гиперпараметры, что и в стратегии 1.

3.2.3. Метрики диагностической точности

Бинарные метрики стратегии 2 приведены ниже:

- **Чувствительность** – 86,2% (95% ДИ: 81,9–98,0);
- **Специфичность** – 83,8% (95% ДИ: 69,0–86,0);
- **Точность** – 84,0% (95% ДИ: 70,9–86,1);
- **F1-мера** – 40,7% (95% ДИ: 29,3–47,2);
- **ПЦПР** – 26,6% (95% ДИ: 17,2–32,2);
- **ПЦОР** – 98,9% (95% ДИ: 98,5–99,8);

- AUC – 0,909 (95% ДИ: 0,877–0,935);
- MCC – 0,416 (95% ДИ: 0,336–0,483).

Абсолютное число ложноположительных результатов: 207 в стратегии 2 против 367 в стратегии 1, разность 160 соответствует относительному снижению на $160/367 = 43,6\%$.

3.2.4. Распределение вероятностных оценок

Распределение вероятностных оценок на тестовой выборке для класса BI-RADS 2 среднее 0,200, стандартное отклонение 0,210, для класса BI-RADS 4 и 5 среднее 0,624, стандартное отклонение 0,187, с получением разницы средних – 0,424.

3.3. Результаты стратегии 3 – двухклассовая классификация с применением логистической регрессии

3.3.1. Характеристика задачи и обучающей выборки

В стратегии 3 целевая переменная сформирована из исходной четырёхуровневой шкалы BI-RADS бинарной постановкой задачи, то есть BI-RADS 2 – это класс 0, BI-RADS 3, 4, 5 – это класс 1 (см. подраздел 2.4.4). Анализировались 1389 образований: 1280 (92,2%) – класс 0, 109 (7,8%) – класс 1. Соотношение классов 1280: 109 $\approx$ 11,7: 1.

Алгоритм классификации – ЛР со значением `penalty='l2'` по умолчанию, `class_weight='balanced'`, `max_iter=1000`, `random_state=42`; масштабирование признаков – `StandardScaler`. Полный набор гиперпараметров и обоснование выбора `max_iter=1000` приведены в подразделе 2.5.1.

3.3.2. Подбор гиперпараметров

GridSearchCV с трёхкратной кросс-валидацией (cv=3, критерий – F1-macro, см. подраздел 2.5.1).

Оптимальный порог классификации на шкале вероятностей, выбранный по максимуму индекса Юдена составил 0,556. Для стратегий 1 и 2 этот порог был 0,332 и 0,446 соответственно. Значение порога не отражает ранжирующую способность модели (за неё отвечает AUC, см. 3.3.3); порог характеризует положение точки на ROC-кривой.

3.3.3. Метрики диагностической точности

Бинарные метрики стратегии 3 приведены ниже. Для каждой метрики указаны заявленное значение из таблицы 4 (раздел 3.5.1).

- **Чувствительность** – 90,8% (95% ДИ: 85,9–96,4;
- **Специфичность** – 95,2% (95% ДИ: 93,1–96,5);
- **Точность** – 94,9% (95% ДИ: 92,9–96,2;
- **F1-мера** – 73,6% (95% ДИ: 65,0–79,7;
- **ПЦПР** – 61,9% (95% ДИ: 50,9–69,8;
- **ПЦОР** – 99,2% (95% ДИ: 98,8–99,7);
- **AUC** – 0,942 (95% ДИ: 0,906–0,972);
- **AP** – 0,784 (95% ДИ: 0,698–0,859);
- **МСС** – 0,719 (95% ДИ: 0,644–0,785).

3.3.4. Сопоставление с другими стратегиями

Стратегия 3 даёт большие значения по 6 из 8 приведённых метрик в сравнении со стратегией 2.

Утверждение о «клинической применимости» стратегии 3 как основы для системы поддержки принятия решений требует внешней валидации на независимой выборке (раздел 3.7.2), сравнения с базовыми клиническими

подходами и анализа калибровки вероятностных оценок в эталонной популяции.

3.4. Результаты стратегии 4 – двухклассовая классификация с применением случайного леса

3.4.1. Характеристика задачи и обучающей выборки

Стратегия 4 повторяет задачу стратегии 3, где анализировались те же 1389 образований в соотношении 92,2% к 7,8%.

От стратегии 3 стратегия 4 отличается двумя одновременно изменёнными методологическими параметрами: алгоритм классификации (СЛ вместо ЛР) и метод масштабирования признаков (MinMaxScaler вместо StandardScaler).

3.4.2. Подбор гиперпараметров

GridSearchCV с трёхкратной кросс-валидацией (cv=3, критерий – F1-мера) выбрал те же гиперпараметры СЛ, что и в стратегиях 1 и 2. Оптимальный порог классификации на шкале вероятностей, выбранный по максимуму индекса Юдена, составил 0,447. Сводка порогов: стратегия 1 – 0,332; стратегия 2 – 0,446; стратегия 3 – 0,556; стратегия 4 – 0,447. Само значение порога не отражает разделяющую способность модели; за неё отвечает AUC.

3.4.3. Метрики диагностической точности

- Чувствительность – 75,2% (95% ДИ: 68,8–86,2)
- Специфичность – 89,6% (95% ДИ: 83,0–91,4)
- Точность – 88,5% (95% ДИ: 82,9–90,4);
- F1-мера – 50,6% (95% ДИ: 39,8–57,2)
- ПЦПР – 38,1% (95% ДИ: 26,8–45,0)

- ПЦОР – 97,7% (95% ДИ: 97,0–98,7);
- AUC – 0,826 (95% ДИ: 0,775–0,875);
- AP – 0,349 (95% ДИ: 0,276–0,439);
- MCC – 0,482 (95% ДИ: 0,387–0,549).

3.4.4. Сопоставление стратегий 3 и 4

Стратегии 3 и 4 решают одну и ту же задачу на одной и той же выборке. Все восемь приведённых выше метрик у стратегии 4 ниже, чем у стратегии 3. Наибольшие разности по табличным значениям: F1-мера – 23,0 п. п., ПЦПР – 23,8 п. п., чувствительность – 15,6 п. п., AUC – 0,116, AP – 0,435, MCC – 0,237.

Эти различия – совокупный эффект двух одновременно изменённых параметров (алгоритм совместно с методом масштабирования). Какой из двух факторов определяет различия, алгоритм, метод масштабирования или их взаимодействие, на основании четырёх стратегий установить нельзя. В литературе линейные модели ЛР и нелинейные ансамбли СЛ на радиомических данных дают разные результаты в зависимости от структуры признакового пространства и размера обучающей выборки. На текущей выборке наблюдается более высокое качество ЛР + StandardScaler по сравнению с СЛ+ MinMaxScaler по всем восьми метрикам, но содержательная причина этого различия является открытым вопросом, требующий дополнительных исследований.

3.5. Сравнительный анализ четырёх стратегий классификации

3.5.1. Сводная таблица метрик

Результаты всех четырёх стратегий сведены в таблицу 4. Для корректного сопоставления все метрики рассчитаны на единой тестовой выборке с порогами, определёнными по индексу Юдена.

Таблица 4. Сводная сравнительная таблица метрик четырёх стратегий классификации

	Многоклассовая классификация	Трехклассовая классификация	Двухклассовая классификация	
Модель ИИ	Случайный лес	Случайный лес	Логистическая регрессия	Случайный лес
Порог (Индекс Юдена)	0,332	0,446	0,556	0,447
Площадь под ROC-кривой (AUC)	0,861 (ДИ: 0,827–0,890)	0,909 (ДИ: 877–0,935)	0,942 (ДИ: 0,906–0,972)	0,826 (ДИ: 0,775–0,875)
Площадь под Precision-Recall-кривой (AP)	0,345 (ДИ: 0,263–0,434)	0,411 (ДИ: 0,317–0,523)	0,784 (ДИ: 0,698–0,859)	0,349 (ДИ: 0,276–0,439)
Чувствительность (%)	91,7 (ДИ: 84,3–96,4)	86,2 (ДИ: 81,9–98,0)	90,8 (ДИ: 85,9–96,4)	75,2 (ДИ: 68,8–86,2)
Специфичность (%)	72,0 (ДИ: 69,3–78,1)	83,8 (ДИ: 69,0–86,0)	95,2 (ДИ: 93,1–96,5)	89,6 (ДИ: 83,0–91,4)
Точность (%)	73,5 (ДИ: 71,1–79,0)	84,0 (ДИ: 70,9–86,1)	94,9 (ДИ: 92,9–96,2)	88,5 (ДИ: 82,9–90,4)
Первичная метрика исследования (F1-мера) (%)	35,2 (ДИ: 30,5–42,2)	40,7 (ДИ: 29,3–47,2)	73,6 (ДИ: 65,0–79,7)	50,6 (ДИ: 39,8–57,2)
ПЦПР (%)	21,8 (ДИ: 18,3–27,5)	26,6 (ДИ: 17,2–32,2)	61,9 (ДИ: 50,9–69,8)	38,1 (ДИ: 26,8–45,0)
ПЦОР (%)	99,0 (ДИ: 98,2–99,6)	98,9 (ДИ: 98,5–99,8)	99,2 (ДИ: 98,8–99,7)	97,7 (ДИ: 97,0–98,7)
Коэффициент корреляции Мэтьюса (МСС) (%)	36,4 (ДИ: 32,1–42,4)	42,3 (ДИ: 33,6–48,3)	72,5 (ДИ: 64,4–78,5)	48,2 (ДИ: 38,7–54,9)

3.5.2. ROC-кривые и дискриминационная способность

Анализ ROC-кривых подтверждает устойчивое превосходство стратегии 3 на всём диапазоне порогов. Стратегия 3 достигла $AUC = 0,942$, что является «отличной» дискриминационной способностью. Стратегия 2 заняла второе место с $AUC = 0,909$. Стратегии 1 и 4 обеспечивают «хорошую» дискриминацию ($AUC = 0,861$ и $0,826$ соответственно), причём стратегия 4 оказалась наименее эффективной, несмотря на более простую постановку задачи по сравнению со стратегиями 1 и 2.

3.5.3. Precision-Recall кривые и работа в условиях дисбаланса

Анализ PR-кривых наиболее информативен при дисбалансе классов, так как он строится исключительно на показателях положительного класса и не маскируется большим числом истинно отрицательных результатов. Стратегия 3 превосходит остальные по AP вдвое ($0,784$ против $0,345$, $0,411$ и $0,349$ для стратегий 1, 2 и 4 соответственно). Кривая стратегии 3 устойчиво располагается выше кривых остальных стратегий на всём диапазоне значений полноты.

3.5.4. Сравнительная диаграмма ключевых метрик

Визуальное сопоставление шести ключевых метрик демонстрирует доминирующее положение стратегии 3 по всем показателям одновременно. Ни одна из стратегий 1, 2 и 4 не превзошла стратегию 3 ни по одной из метрик при тех же условиях оценки.

3.5.5. Анализ типов ошибок в клиническом контексте

Ключевым клиническим критерием оценки является соотношение двух типов ошибок. Ложноотрицательные результаты (FN) – пропущенные патологические находки, которые несут прямой диагностический риск

задержки лечения. Ложноположительные результаты (FP) – это избыточные направления на дообследование и биопсию, они создают психологическую нагрузку на пациентку и экономические издержки системы.

Стратегия 3 обеспечивает оптимальный баланс: FN = 11 (сопоставимо со стратегиями 1 и 2) при FP = 61 в 6 раз меньше, чем у стратегии 1 (367), и в 2,2 раза меньше, чем у стратегии 4 (133). Стратегия 1 минимизирует FN (9), однако ценой 367 ложноположительных результатов, на каждый пропущенный случай патологии приходится 41 избыточное направление. Такое соотношение неприемлемо в клинической практике.

3.5.6. Динамика метрик по стратегиям: ключевые закономерности

Анализ динамики метрик позволяет выделить три принципиальные закономерности.

Первая: переход от многоклассовой к трёхклассовой постановке улучшил качество большинства метрик при незначительном снижении чувствительности. Исключение малочисленной категории BI-RADS 3 устраняет источник методологического «шума».

Вторая: переход к двухклассовой постановке с логистической регрессией дал качественный скачок по всем метрикам, особенно по F1-мере и ПЦПР. Это свидетельствует о том, что объединение BI-RADS 3, 4 и 5 в единый класс «клинической значимости» является методологически обоснованным и клинически осмысленным решением.

Третья: при двухклассовой постановке логистическая регрессия превзошла СЛ по всем ключевым метрикам. Необходимо, учитывать, что это сравнение двухфакторное, в котором различались одновременно алгоритм и схема масштабирования (подраздел 3.4.1). С этой оговоркой результат допускает два непротиворечивых объяснения: разделяющая граница в пространстве радиомических признаков маммограмм близка к линейной, что делает линейный классификатор не менее эффективным, чем нелинейный ансамбль; случайный лес с неограниченной глубиной деревьев мог

переобучиться на обучающей выборке в условиях малого числа пациенток, что снизило обобщаемость. Для верификации гипотезы о линейности граничной поверхности рекомендуется в последующих работах зафиксировать схему масштабирования и варьировать только алгоритм.

3.6. Клиническая интерпретация результатов

3.6.1. Общий контекст клинической применимости

Оценка радиомических моделей не может ограничиваться статистическими метриками. Конечный критерий применимости способность модели улучшить клиническое решение, то есть повысить поиск патологии, снизить число избыточных биопсий или сократить нагрузку на рентгенолога без ущерба диагностической безопасности. Результаты четырёх стратегий в данном разделе интерпретируются именно с этих позиций.

3.6.2. Клиническая интерпретация стратегии 3 как ведущей модели

Стратегия 3 продемонстрировала метрики, наиболее близкие к требованиям скрининговой диагностики. Чувствительность 90,8% при специфичности 95,2% означает, что из 100 пациенток с клинически значимыми находками (BI-RADS 3–5) модель верно выявит 91, направив на дополнительную верификацию лишь 5 из 100 женщин с доброкачественными образованиями BI-RADS 2. Для сравнения с литературными данными, при традиционном скрининге частота ложноположительных вызовов достигает 10–15% за 10 лет наблюдения, что сопровождается значительной психологической нагрузкой на пациенток.

ПЦПР 61,9% означает, что каждое второе положительное заключение модели соответствует истинной патологии – это показатель, существенно превосходящий распространённость патологии в исходной выборке.

ПЦОР 99,2% – клинически весомый показатель модели, то есть отрицательное заключение системы достоверно исключает патологию, что обеспечивает безопасность её применения как инструмента поддержки решения о наблюдении без дополнительных исследований.

3.6.3. Роль наиболее значимых признаков в клинической интерпретации

Доминирование признаков класса NonUniformity (GrayLevelNonUniformity, RunLengthNonUniformity, DependenceNonUniformity) в топ-20 по F-критерию Фишера согласуется с патоморфологической сущностью злокачественных образований молочной железы. Злокачественные опухоли характеризуются выраженной клеточной и тканевой неоднородностью, хаотичным ядерным полиморфизмом, некрозом, инфильтративным ростом, что на уровне пиксельной интенсивности маммограммы проявляется именно как текстурная негомогенность, фиксируемая признаками NonUniformity.

3.6.4. Сравнение с литературными данными

Полученные результаты целесообразно рассмотреть в контексте опубликованных исследований, посвящённых радиомическому анализу изображений молочной железы.

Wang L. и соавт. (2022) разработали модель на основе опорных векторов (SVM) для разграничения доброкачественных и злокачественных образований в категориях BI-RADS 4 и 5 по цифровым маммограммам в двух стандартных проекциях [24]. В выборке из 288 пациенток с применением LASSO-регрессии было отобрано 14 информативных признаков из 188 исходных. На тестовой выборке модель достигла AUC 0,901 (95% ДИ: 0,835–0,961) при чувствительности 95,8% и специфичности 66,7%. Авторы поставили задачу обеспечить чувствительность, сопоставимую с биопсией, с целью сокращения

числа инвазивных вмешательств у пациенток с доброкачественными образованиями.

Sakai A. и соавт. (2020) предложили метод автоматической классификации образований молочной железы на изображениях цифрового томосинтеза (ЦТ) с использованием 70 радиомических признаков, отражающих форму, текстуру и характер спикул [42]. На небольшой выборке из 51 изображения (24 пациента) наилучший результат при применении SVM-классификатора составил AUC 0,798; при редукции пространства признаков до 6 методом LASSO – от 0,742 до 0,808 в зависимости от используемого алгоритма.

Murtas F. и соавт. (2023) построили клинико-радиомические модели на основе ЦТ-изображений 150 пациенток, включив в анализ наряду с радиомическими признаками также клинические данные – возраст и плотность железистой ткани по BI-RADS [43]. Применение трёх методов отбора признаков (K-best, sequential и Random Forest) позволило получить AUC в диапазоне 0,716–0,740; наилучший результат показала модель на основе случайного леса – AUC 0,740 (95% ДИ: 0,662–0,819). Авторы пришли к выводу, что подобная точность является приемлемой для системы поддержки принятия решений на этапе первичного скрининга.

Таким образом, результаты стратегии 3 настоящего исследования (AUC 0,942) превосходят показатели приведённых работ либо соответствуют их верхней границе. Вместе с тем прямое сравнение требует осторожности, поскольку цитируемые исследования решают задачу бинарной классификации с верификацией по гистологическому материалу, тогда как в настоящей работе целевой переменной служит клиническая оценка рентгенолога по шкале BI-RADS. Такая постановка задачи принципиально ближе к реальной диагностической практике и предполагает возможность непосредственной интеграции модели в процесс описания маммограмм без дополнительного преобразования результата.

3.7. Ограничения исследования

3.7.1. Объём и структура выборки

Первое и наиболее существенное ограничение – малый размер клинической выборки: исследование выполнено на данных 83 пациенток. Несмотря на 1391 размеченное образование, это ограничивает статистическую независимость наблюдений, множественные образования от одной пациентки потенциально могут быть структурированы как по биологическим характеристикам ткани молочной железы, так и по техническим параметрам конкретного исследования. Эта корреляция формально не учитывалась при разбиении выборки, что может давать оптимистичную оценку обобщаемости модели.

Выраженный дисбаланс классов отражает реальное распределение находок при скрининге. Несмотря на применение `class_weight='balanced'`, малое абсолютное число образований BI-RADS 3, 4 и 5 ограничивает статистическую мощность оценок для этих классов. Широкие доверительные интервалы ряда метрик, например, специфичность стратегии 2: 95% ДИ 69,0–86,0, прямо отражают данную неопределённость.

3.7.2. Дизайн исследования и отсутствие внешней валидации

Ретроспективный одноцентровый дизайн исследования ограничивает внешнюю валидность результатов. Все данные получены из одного источника ЕРИС ЕМИАС Москвы за двухлетний период (2020–2021 гг.) с использованием трёх маммографов одного учреждения. Отсутствие внешней валидации на независимой выборке другого временного периода, использования других маммографов, не позволяет с уверенностью утверждать о воспроизводимости достигнутых метрик в иных условиях.

Это ограничение характерно для большинства радиомических исследований на начальном этапе разработки моделей и в соответствии с

рекомендациями CLAIM рассматривается, как обязательный предмет последующей работы. Проведение валидации на данных других маммограммах, является приоритетным направлением.

3.7.3. Методология сегментации

Ручная сегментация, выполненная преимущественно одним исследователем, вносит потенциальную субъективность в формирование масок. Смещение контура ROI на несколько пикселей способно изменить значения ряда текстурных признаков, прежде всего GLRLM и GLSZM.

Следует также отметить, что сегментация выполнялась на коллажах, а не на оригинальных DICOM-изображениях. Несмотря на конвертацию PNG без потерь, пространственное масштабирование при формировании коллажа могло незначительно изменить пиксельные расстояния, что потенциально влияет на признаки, чувствительные к геометрии (Shape 2D, GLRLM).

3.7.4. Радиомический анализ

Конфигурация PyRadiomics (binWidth = 25, normalizeScale = 100) фиксировалась единообразно для всех образований. Оптимальность этих параметров для маммографических изображений различных типов образований в рамках систематического анализа чувствительности не проверялась. Параметр binWidth оказывает значительное влияние на текстурные признаки: меньшие значения повышают разрешение гистограммы, но увеличивают чувствительность к шуму; большие – напротив. Выбор binWidth = 25 обоснован рекомендациями IBSI для 2D-изображений, однако специфическая оптимизация под данный набор данных остаётся перспективной задачей.

3.7.5. Интерпретация и обобщаемость

Исследование не включало клинические переменные, такие как возраст пациентки, семейный анамнез, которые, по литературным данным, систематически улучшают производительность комбинированных клинико-радиомических моделей. Включение этих переменных является перспективным направлением, способным повысить как ПЦПР, так и клиническую применимость модели.

Исследование также ограничено двумерным анализом маммограмм. Переход к цифровому томосинтезу (ТМЖ) или МРТ молочной железы потенциально расширяет информационное содержание за счёт трёхмерного текстурного анализа, однако требует принципиально иного методологического анализа.

ЗАКЛЮЧЕНИЕ

Настоящее исследование посвящено сравнительной оценке четырёх стратегий радиомической классификации маммографических образований по шкале BI-RADS с применением алгоритмов машинного обучения на данных ЕРИС ЕМИАС.

Вывод 1. Двухклассовая классификация результативнее многоклассовой. Стратегии с двухклассовой постановкой задачи обеспечивают качественно иной уровень дискриминации по сравнению с многоклассовой и трёхклассовой схемами. Стратегия 3 (двухклассовая ЛР) превзошла стратегию 1 (многоклассовый СЛ) по всем ключевым показателям одновременно, при сокращении числа ложноположительных результатов в 6 раз. Результат согласуется с гипотезой о том, что разделяющая граница в пространстве радиомических признаков маммограмм имеет преимущественно линейный характер, тогда как принудительная дифференциация четырёх категорий BI-RADS вносит неустранимый методологический шум.

Вывод 2. Категория BI-RADS 3 малоинформативна для алгоритмов машинного обучения. Включение BI-RADS 3 в целевую переменную системно ухудшало все метрики по сравнению со стратегиями, исключаящими или объединяющими эту категорию. Причин две: крайне малочисленность класса не позволяет алгоритму сформировать устойчивые разделяющие границы; текстурная неоднородность самой категории BI-RADS 3 объединяет объекты с принципиально различными радиомическими сигнатурами (единичные группы микрокальцинатов и образования с чёткими контурами), образуя в пространстве признаков размытые, плохо делимые границы. Результат является эмпирическим подтверждением ограниченной пригодности BI-RADS 3 как отдельного дискретного класса для алгоритмов ИИ, что согласуется с клиническим консенсусом о неопределенности данной категории. Генерализация этого вывода требует многоцентровой верификации.

Вывод 3. Двухклассовая классификация – оптимальная основа для СППВР. Стратегия 3 достигла $AUC = 0,942$ (95% ДИ: 0,906–0,972), чувствительности 90,8%, специфичности 95,2%, F1-меры 73,6%, ПЦПР 61,9% и ПЦОР 99,2%. Этот профиль соответствует требованиям к инструменту «второго читателя» в скрининговых программах. Высокий ПЦОР гарантирует диагностическую безопасность отрицательного заключения, высокая специфичность снижает поток избыточных направлений на биопсию.

Принципиальное достоинство двухклассовой схемы для СППВР – это клиническая интерпретируемость. Объединение BI-RADS 3, 4 и 5 в единый класс «клинической значимости» не только повышает статистическую мощность модели, но и стандартизирует интерпретацию. Вместо субъективного выбора между «вероятно доброкачественным» и «подозрительным» система предоставляет конкретный результат. Окончательный вывод об оптимальности данной стратегии для клинического применения может быть сформулирован лишь после внешней многоцентровой валидации и проспективного исследования с прямой оценкой влияния СППВР на решения рентгенолога.

Вывод 4. Текстурная неоднородность – ключевой предиктор. Среди наиболее информативных признаков доминируют дескрипторы текстурной неоднородности классов GLRLM, GLSZM и GLDM, что согласуется с патоморфологической сущностью злокачественных образований, клеточным полиморфизмом, некрозом и инфильтративным ростом, проявляющимися в пиксельной неоднородности маммограммы. Преимущество логистической регрессии перед случайным лесом при идентичной постановке задачи указывает на линейный характер разделяющей поверхности, упрощающий клиническую интерпретацию модели, однако это сравнение двухфакторное, что не позволяет сделать однозначного вывода без дополнительного контролируемого эксперимента.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

ДИ – доверительный интервал

ЕМИАС – Единая медицинская информационно-аналитическая система

ЕРИС – Единый радиологический информационный сервис

ИИ – искусственный интеллект

ИЧР – ионизирующее излучение в анамнезе

КТ – компьютерная томография

ЛР – логистическая регрессия

МРТ – магнитно-резонансная томография

ПЦОР – прогностическая ценность отрицательного результата

ПЦПР – прогностическая ценность положительного результата

ПЭТ/КТ – позитронно-эмиссионная томография, совмещенная с рентгеновской компьютерной томографией

РМЖ – рак молочной железы

СЛ – случайный лес

СППВР – система поддержки принятия врачебных решений

ТМЖ – томосинтез молочных желёз

УЗИ – ультразвуковое исследование

ЦМ – цифровая маммография

ЦТ – цифровой томосинтез

ACR – American College of Radiology (Американский колледж радиологии)

AP – Average Precision (площадь под PR-кривой)

AUC – Area Under the Curve (площадь под ROC-кривой)

BI-RADS – Breast Imaging Reporting and Data System (система описания и обработки данных исследований молочной железы)

BRCA – Breast Cancer Susceptibility Gene (ген предрасположенности к раку молочной железы)

CC – craniocaudal (краниокаудальная проекция)

CISNET – Cancer Intervention and Surveillance Modeling Network (сеть моделирования вмешательств и эпидемиологического надзора в онкологии)

CLAIM – Checklist for Artificial Intelligence in Medical Imaging (контрольный перечень для ИИ в медицинской визуализации)

CNN – Convolutional Neural Network (свёрточная нейронная сеть)

CVAT – Computer Vision Annotation Tool (инструмент разметки изображений)

DICOM – Digital Imaging and Communications in Medicine (стандарт цифровой медицинской визуализации)

DL – Deep Learning (глубокое обучение)

FN – False Negative (ложноотрицательный результат)

FP – False Positive (ложноположительный результат)

GLCM – Gray Level Co-occurrence Matrix (матрица сопряжённости уровней серого)

GLDM – Gray Level Dependence Matrix (матрица зависимостей уровней серого)

GLOBOCAN – Global Cancer Observatory (база данных онкологической заболеваемости ВОЗ)

GLRLM – Gray Level Run Length Matrix (матрица длин прогонов уровней серого)

GLSZM – Gray Level Size Zone Matrix (матрица зон размеров уровней серого)

GRF – Global Radiomic Features (глобальные радиомические признаки)

IBSI – Image Biomarker Standardisation Initiative (инициатива по стандартизации биомаркеров изображений)

ICC – Intraclass Correlation Coefficient (коэффициент внутриклассовой корреляции)

IoU – Intersection over Union (метрика пересечения размеченных областей)

LASSO – Least Absolute Shrinkage and Selection Operator (метод регуляризации с L1-нормой)

MCC – Matthews Correlation Coefficient (коэффициент корреляции Мэтьюса)

ML – Machine Learning (машинное обучение)

MLO – Mediolateral Oblique (медиолатеральная косая проекция)

NGTDM – Neighbouring Gray Tone Difference Matrix (матрица разностей соседних тонов серого)

PGMI – Perfect, Good, Moderate, Inadequate (система оценки качества маммограмм)

PNG – Portable Network Graphics (формат растрового изображения)

QIBA – Quantitative Imaging Biomarkers Alliance (международный Альянс биомаркеров количественной визуализации)

RLE – Run-Length Encoding (кодирование длин серий)

ROC – Receiver Operating Characteristic (характеристическая кривая классификатора)

ROI – Region of Interest (область интереса)

Se – Sensitivity (чувствительность)

Sp – Specificity (специфичность)

TN – True Negative (истинноотрицательный результат)

TP – True Positive (истинноположительный результат)

XML – Extensible Markup Language (расширяемый язык разметки)

СПИСОК ЛИТЕРАТУРЫ

1. Sung, H. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries / H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal [et al.]. – Text : electronic // CA: A Cancer Journal for Clinicians. – 2021. – Vol. 71, № 3. – P. 209–249. – DOI: 10.3322/caac.21660.
2. Злокачественные новообразования в России в 2023 году (заболеваемость и смертность): [монография] / под ред. А. Д. Каприна, В. В. Старинского, Г. В. Петровой. – Москва: МНИОИ им. П. А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2024. – 276 с. – Текст : непосредственный.
3. Newell, M. S. ACR BI-RADS v2025 Manual / M. S. Newell, S. V. Destounis, J. W. T. Leung [et al.] ; American College of Radiology. – Reston, VA : American College of Radiology, 2025. – Text : unmediated.
4. Choi, J. S. Breast Imaging Reporting and Data System (BI-RADS): advantages and limitations / J. S. Choi. – Text : electronic // Journal of the Korean Society of Radiology. – 2023. – Vol. 84, № 1. – P. 3–14. – DOI: 10.3348/jksr.2022.0142.
5. Conti, A. Radiomics in breast cancer classification and prediction / A. Conti, A. Duggento, I. Indovina, M. Guerrisi, N. Toschi. – Text : electronic // Seminars in Cancer Biology. – 2021. – Vol. 72. – P. 238–250. – DOI: 10.1016/j.semcancer.2020.04.002.
6. Qi, Y.-J. Radiomics in breast cancer: Current advances and future directions / Y.-J. Qi, G.-H. Su, C. You, X. Zhang, Y. Xiao, Y.-Z. Jiang, Z.-M. Shao. – Text : electronic // Cell Reports Medicine. – 2024. – Vol. 5, № 9. – P. 101719. – DOI: 10.1016/j.xcrm.2024.101719.
7. Oh, K. E. Radiomics to Differentiate Malignant and Benign Breast Lesions: A Systematic Review and Diagnostic Test Accuracy Meta-Analysis / K. E.

- Oh, N. Vasandani, A. Anwar. – Text : electronic // Cureus. – 2023. – Vol. 15, № 11. – P. e49015. – DOI: 10.7759/cureus.49015.
8. Govorukhina, V. G. The role of mammography in breast cancer radiomics / V. G. Govorukhina, S. S. Semenov, P. B. Gelezhe, V. V. Didenko, S. P. Morozov, A. E. Andreychenko. Text : electronic // Digital Diagnostics. – 2021. – Vol. 2, № 2. – P. 185–199. – DOI: 10.17816/DD70479.
9. Васильев, Ю. А. Эффективность профилактической маммографии в Российской Федерации: сравнение результатов первого этапа диспансеризации в 2019 и 2022 гг. / Ю. А. Васильев, И. М. Сон, Л. Д. Пестренин, А. В. Владзимирский. – Текст : непосредственный // Менеджер здравоохранения. – 2024. – № 4. – С. 63–76.
10. Радиомика в лучевой диагностике биологических подтипов рака молочной железы (обзор литературы) / А. Ю. Попова, В. Е. Гажонова, М. В. Карташов [и др.]. – Текст : электронный // Уральский медицинский журнал. – 2024. – Т. 23, № 4. – С. 41–56. – DOI: 10.52420/umj.23.4.41. – EDN: DUJFIG.
11. Пасынков, Д. В. Контурный анализ маммографических изображений доброкачественной и злокачественной патологии молочной железы / Д. В. Пасынков, И. В. Ключкин, О. В. Бусыгина. – Текст : электронный // Казанский медицинский журнал. – 2015. – Т. 96, № 3. – С. 322–325. – DOI: 10.17750/KMJ2015-322.
12. Серова, Н. С. Перспективные методы ранней лучевой диагностики плоскоклеточной карциномы полости рта / Н. С. Серова, В. Д. Володина, А. В. Лукьяненко, И. В. Решетов. – Текст : непосредственный // Медицинский вестник МВД. – 2025. – Т. 135, № 2. – С. 55–60. – DOI: 10.52341/20738080_2025_135_2_55.
13. Baltzer, P. A. A simple and robust classification tree for differentiation between benign and malignant lesions in MR-mammography / P. A. Baltzer, M. Dietzel, W. A. Kaiser. – Text : electronic // European Radiology. – 2013. – Vol. 23, № 8. – P. 2051–2060. – DOI: 10.1007/s00330-013-2804-3.

- 14.Васильев, Ю. А. Оценка производительности программного обеспечения на основе технологии искусственного интеллекта при описании цифровых маммографических исследований / Ю. А. Васильев, А. В. Колсанов, К. М. Арзамасов, А. В. Владзимирский, О. В. Омелянская, С. С. Семёнов, Л. Е. Аксёнова. – Текст : электронный // Digital Diagnostics. – 2024. – Т. 5, № 4. – С. 695–711. – DOI: 10.17816/DD625967.
- 15.Pradeep Kumar, D. Rethinking Breast Imaging Reporting and Data System (BI-RADS) 4: clinical insights from mammographic subclassification / D. Pradeep Kumar, A. X. Chakiath, A. Sugunan, K. Kandiah, R. R. Gandhi, J. Job [et al.]. – Text : electronic // Cureus. – 2025. – Vol. 17, № 10. – P. e94379. – DOI: 10.7759/cureus.94379.
- 16.Mohapatra, S. K. The positive predictive values of the Breast Imaging Reporting and Data System (BI-RADS) 4 lesions and its mammographic morphological features / S. K. Mohapatra, A. Mishra, T. K. Sahoo, R. B. Nayak, P. K. Das, B. Nayak [et al.]. – Text : electronic // Indian Journal of Surgical Oncology. – 2021. – Vol. 12, № 1. – P. 182–189. – DOI: 10.1007/s13193-020-01274-5.
- 17.Гажонова, В. Е. Инновационные подходы в диагностике очаговых образований молочной железы категории BI-RADS 3–4: роль и перспективы интеграции ультразвукового исследования с контрастированием и эластографии / В. Е. Гажонова, А. В. Зубарев, И. З. Пулатова. – Текст : электронный // Российский электронный журнал лучевой диагностики. – 2025. – Т. 15, № 1. – С. 107–118. – DOI: 10.21569/2222-7415-2025-15-1-107-118.
- 18.Dietzel, M. Как использовать шкалу Кайзера для принятия диагностических решений при мультипараметрической МРТ молочной железы / M. Dietzel, M. L. Mazo, N. V. Rozhkova, S. A. Kharuzhyk, D. I. Kuplevatskaya, E. A. Busko [et al.]. – Текст : электронный // Российский

- электронный журнал лучевой диагностики. – 2020. – Т. 10, № 3. – С. 58–76. – DOI: 10.21569/2222-7415-2020-10-3-58-76.
19. Gu, J. Ultrasound radiomics in personalized breast management: Current status and future prospects / J. Gu, T. Jiang. – Text : electronic // *Frontiers in Oncology*. – 2022. – Vol. 12. – P. 963612. – DOI: 10.3389/fonc.2022.963612.
20. Jovanović, M. A new challenge in Radiology: Radiomics in breast cancer diagnostics / M. Jovanović, T. Matijaš. – Text : electronic // *Radiological Journal / Radiološki vjesnik*. – 2023. – Vol. 47, № 1. – P. 1–13. – DOI: 10.55378/rv.47.1.1.
21. Mongan, J. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers / J. Mongan, L. Moy, C. E. Kahn. – Text : electronic // *Radiology: Artificial Intelligence*. – 2020. – Vol. 2, № 2. – P. e200029. – DOI: 10.1148/ryai.2020200029.
22. Wang, L. A radiomics nomogram for distinguishing benign from malignant round-like breast tumors / L. Wang, Y. Ding, W. Yang, H. Wang, J. Shen, W. Liu [et al.]. – Text : electronic // *Frontiers in Oncology*. – 2022. – Vol. 12. – P. 677803. – DOI: 10.3389/fonc.2022.677803.
23. Siviengphanom, S. Global Radiomic Features from Mammography for Predicting Difficult-To-Interpret Normal Cases / S. Siviengphanom, Z. Gandomkar, S. J. Lewis, P. C. Brennan. – Text : electronic // *Journal of Digital Imaging*. – 2023. – Vol. 36, № 4. – P. 1541–1552. – DOI: 10.1007/s10278-023-00836-7.
24. Wang, G. Radiomics Based on Digital Mammography Helps to Identify Mammographic Masses Suspicious for Cancer / G. Wang, D. Shi, Q. Guo, H. Zhang, S. Wang, K. Ren. – Text : electronic // *Frontiers in Oncology*. – 2022. – Vol. 12. – P. 843436. – DOI: 10.3389/fonc.2022.843436.
25. Letchumanan, N. A Radiomics Study: Classification of Breast Lesions by Textural Features from Mammography Images / N. Letchumanan, J. H. D. Wong, L. K. Tan, N. Ab Mumin, W. L. Ng, W. Y. Chan, K. Rahmat. – Text :

- electronic // Journal of Digital Imaging. – 2023. – Vol. 36, № 4. – P. 1533–1540. – DOI: 10.1007/s10278-022-00753-1.
26. Zhang, Q. A combined clinical-ultrasound radiomics model for differentiating benign and malignant BI-RADS category 4 breast masses / Q. Zhang, J. Gao, E. A. Agyekum, L. Zhu, C. Jiang, S. Du, L. Yin. – Text : electronic // American Journal of Translational Research. – 2025. – Vol. 17, № 8. – P. 6370–6380. – DOI: 10.62347/SBKU2090.
27. Lin, J. A meta-analysis of MRI radiomics-based diagnosis for BI-RADS 4 breast lesions / J. Lin, H. Zheng, Q. Jia, J. Shi, S. Wang, J. Wang, M. Ge. – Text : electronic // Journal of Cancer Research and Clinical Oncology. – 2024. – Vol. 150, № 5. – P. 254. – DOI: 10.1007/s00432-024-05697-3.
28. Interlenghi, M. A Machine Learning Ensemble Based on Radiomics to Predict BI-RADS Category and Reduce the Biopsy Rate of Ultrasound-Detected Suspicious Breast Masses / M. Interlenghi, C. Salvatore, V. Magni [et al.]. – Text : electronic // Diagnostics. – 2022. – Vol. 12, № 1. – P. 187. – DOI: 10.3390/diagnostics12010187.
29. Brunetti, N. A radiomic and clinical data-based risk model for malignancy prediction of breast BI-RADS 4A microcalcifications / N. Brunetti, C. Campi, M. Piana, I. Picone, C. Vercelli, O. Starovatskyi [et al.]. – Text : electronic // Clinical Breast Cancer. – 2025. – Vol. 25, № 4. – P. e461–e469. – DOI: 10.1016/j.clbc.2025.01.006.
30. Yao, G. Mammography Based Nomogram Integrating Radiomics and Clinical Features to Predict Benign or Malignant Regression of BI-RADS 4A Lesions at Follow-Up / G. Yao, L. Chen, T. Wu, Y. Liu, Y. Wan, Z. Xia, B. Liu, M. Zheng. – Text : electronic // International Journal of Women's Health. – 2025. – Vol. 17. – P. 3097–3106. – DOI: 10.2147/IJWH.S539131.
31. Niu, S. Digital breast tomosynthesis-based peritumoral radiomics approaches in the differentiation of benign and malignant breast lesions / S. Niu, T. Yu, Y. Cao, Y. Dong, Y. Luo, X. Jiang. – Text : electronic // Diagnostic and

- Interventional Radiology. – 2022. – Vol. 28, № 3. – P. 217–225. – DOI: 10.5152/dir.2022.20664.
32. Favati, B. Radiomic Applications on Digital Breast Tomosynthesis of BI-RADS Category 4 Calcifications Sent for Vacuum-Assisted Breast Biopsy / B. Favati, R. Borgheresi, M. Giannelli, C. Marini, V. Vani, D. Marfisi, S. Linsalata, M. Moretti, D. Mazzotta, E. Neri. – Text : electronic // Diagnostics. – 2022. – Vol. 12, № 4. – P. 771. – DOI: 10.3390/diagnostics12040771.
 33. Petrillo, A. Prediction of Breast Cancer Histological Outcome by Radiomics and Artificial Intelligence Analysis in Contrast-Enhanced Mammography / A. Petrillo, R. Fusco, E. Di Bernardo [et al.]. – Text : electronic // Cancers. – 2022. – Vol. 14, № 9. – P. 2132. – DOI: 10.3390/cancers14092132.
 34. Bahl, M. Combining AI and Radiomics to Improve the Accuracy of Breast US / M. Bahl. – Text : electronic // Radiology. – 2024. – Vol. 312, № 3. – P. e241795. – DOI: 10.1148/radiol.241795.
 35. Tan, H. Mammography-based artificial intelligence for breast cancer detection, diagnosis, and BI-RADS categorization using multi-view and multi-level convolutional neural networks / H. Tan, Q. Wu, Y. Wu, B. Zheng, B. Wang, Y. Chen [et al.]. – Text : electronic // Insights into Imaging. – 2025. – Vol. 16. – P. 109. – DOI: 10.1186/s13244-025-01983-x.
 36. Vasilev, Y. Comparison of artificial intelligence (AI) services for Breast Imaging-Reporting and Data System (BI-RADS) classification on mammograms / Y. Vasilev, A. Mayorova, D. Rumyantsev, S. Semenov, T. Bobrovskaya, L. Pestrenin [et al.]. – Text : electronic // Quantitative Imaging in Medicine and Surgery. – 2026. – Vol. 16, № 4. – P. 311. – DOI: 10.21037/qims-2025-1658.
 37. Tibshirani, R. Regression Shrinkage and Selection via the Lasso / R. Tibshirani. – Text : electronic // Journal of the Royal Statistical Society. Series B (Methodological). – 1996. – Vol. 58, № 1. – P. 267–288. – DOI: 10.1111/j.2517-6161.1996.tb02080.x.

- 38.Хрустачева, М. Ю. Применение ROC-анализа для оценки результатов работы систем поддержки принятия врачебных решений на примере цифровых маммографических изображений / М. Ю. Хрустачева, Ю. А. Васильев, А. П. Памова, К. М. Арзамасов. – Текст : электронный // Медицинская визуализация. – 2025. – Т. 29, № 3. – С. 80–90. – DOI: 10.24835/1607-0763-1508.
- 39.Васильев, Ю. А. Искусственный интеллект в лучевой диагностике: Per Aspera Ad Astra : монография / Ю. А. Васильев, А. В. Колсанов, К. М. Арзамасов [и др.]. – Москва : ГБУЗ «НПКЦ ДиТ ДЗМ», 2025. – 493 с. – Текст : непосредственный.
- 40.Свидетельство о государственной регистрации программы для ЭВМ № 2025610804 Российская Федерация. Платформа подготовки наборов данных: заявл. 20.12.2024: опубл. 14.01.2025 / Ю. А. Васильев, А. В. Владзимирский, О. В. Омелянская [и др.] ; заявитель ГБУЗ г. Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы». – Текст : непосредственный.
- 41.Luo, C. Mammography radiomics features at diagnosis and progression-free survival among patients with breast cancer / C. Luo, S. Zhao, C. Peng, C. Wang, K. Hu, X. Zhong [et al.]. – Text : electronic // British Journal of Cancer. – 2022. – Vol. 127, № 10. – P. 1886–1892. – DOI: 10.1038/s41416-022-01958-5.
- 42.Sakai, A. A method for the automated classification of benign and malignant masses on digital breast tomosynthesis images using machine learning and radiomic features / A. Sakai, Y. Onishi, M. Matsui, H. Adachi, A. Teramoto, K. Saito, H. Fujita. – Text : electronic // Radiological Physics and Technology. – 2020. – Vol. 13, № 1. – P. 27–36. – DOI: 10.1007/s12194-019-00543-5.
- 43.Murtas, F. Clinical-radiomic models based on digital breast tomosynthesis images: a preliminary investigation of a predictive tool for cancer diagnosis / F. Murtas, V. Landoni, P. Ordóñez [et al.]. – Text : electronic // Frontiers in

Oncology. – 2023. – Vol. 13. – P. 1152158. – DOI:
10.3389/fonc.2023.1152158.

Выпускная квалификационная работа на тему «Применение радиомики для классификации новообразований молочных желез различной этиологии согласно шкале BI-RADS по данным маммографии» выполнена мной самостоятельно. Все использованные в работе материалы и положения из научной литературы и других источников имеют ссылки на них.

Список использованных источников содержит 43 наименований.

ВКР сдана на кафедру «___» _____ 20__ г.

Автор ВКР:

Пирогов Дмитрий Сергеевич

(подпись)