

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Кафедра «Прикладная математика и информатика»
(наименование)

09.04.03 Прикладная информатика
(код и наименование направления подготовки)

Управление корпоративными информационными процессами
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему «Повышение эффективности деятельности учреждений здравоохранения на основе интеллектуального анализа данных»

Обучающийся

В.А. Дьяченко
(Инициалы Фамилия)

(личная подпись)

Научный
руководитель

к.т.н., Н. В. Хрипунов

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Содержание

Введение	3
1 Современное состояние проблемы анализа данных в информационных системах здравоохранения.....	5
1.1 Известные подходы к применению искусственного интеллекта в здравоохранении	9
1.2 Двойное прочтение рентгенологических исследований	11
1.3 Риски и выгоды применения искусственного интеллекта в скрининге: перспективы и ограничения	17
1.4 Применение методов нейронных сетей в проведении двойного прочтения рентгенографии ЗНЖС	19
2 Компьютерная диагностика в маммографии на основе искусственного интеллекта	22
2.1 Стандартные алгоритмы машинного обучения.....	22
2.2 Алгоритмы искусственного интеллекта на основе глубокого обучения	25
2.3 Применение сверточной нейронной сети для анализа изображений молочной железы	29
3 Методология исследования и используемые данные для обработки.....	35
3.1 Подготовка данных для обучения нейронной сети.....	39
3.2 Архитектура нейронной сети. Слои, функции активации и параметры	40
3.3 Описание процесса обучения нейронной сети. Метрики используемые для оценки производительности системы.	43
4 Апробация данных и перспектива развития	48
4.1 Оценка основных показателей и тестируемые данные	55
4.2 Исследуемая группа и процесс прочтения.....	57
4.3 Экономическая эффективность	64
Заключение	66
Список используемой литературы	68

Введение

Во всем мире цифровые технологии являются приоритетом развития сферы здравоохранения. Цифровизация обеспечивает доступность и качество услуг без увеличения затрат на здравоохранение. Данный процесс в России идет медленно, о чем свидетельствует низкий уровень использования средств цифровизации в медицинских учреждениях на региональном уровне, а также отставание в использовании и внедрении новых технологий и сервисов, улучшающих качество медицинских услуг.

Качественный анализ медицинской информации, вопросы сбора информации, ее передача и хранение, вопросы репрезентативности, оперативности и достоверности информации на всех уровнях могут быть качественно решены посредством использования современных информационных технологий.

Многие европейские страны полагаются на двойное считывание для достижения высоких показателей выявления рака при сохранении низких показателей отзыва. Однако высокие требования к ресурсам, предъявляемые двойным считыванием, приводят к устойчивому давлению в условиях нехватки рабочей силы. Система искусственного интеллекта, выполняющая функции надежного и прочного независимого считывания при скрининге рака молочной железы, может помочь удовлетворить как клинические, так и социально-экономические потребности и сделать высококачественный скрининг более доступным.

Таким образом, актуальность данной работы обусловлена необходимостью повышения эффективности процесса постановки диагноза при прочтении рентгенологических снимков при скрининге.

Объектом исследования является деятельность учреждений здравоохранения при проведении скрининга.

Предметом исследования являются методы и технологии интеллектуального анализа данных.

Целью исследования является повышение эффективности прочтения рентгеновских снимков при проведении скрининга, путем применения инструментов интеллектуального анализа данных.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- провести анализ литературных источников, содержащих информацию о проводимых исследованиях в данной предметной области,
- исследовать практику применения систем искусственного интеллекта в рентгенологических исследованиях,
- разработать модель двойного прочтения с применением технологий искусственного интеллекта,
- провести апробацию предлагаемых решений.

Гипотеза исследования – применение методов интеллектуального анализа данных для двойного прочтения рентгенологических снимков позволит повысить достоверность ранней диагностики раковых заболеваний при проведении скринингов.

Научная новизна заключается в разработке и исследовании эффективности нового подхода к диагностике рака молочной железы, сочетающего экспертное медицинское мнение с возможностями искусственного интеллекта.

Практическая значимость состоит в возможности улучшения качества и скорости диагностики, что может способствовать раннему обнаружению рака и снижению доли ложноположительных и ложноотрицательных результатов

Методы исследования данной работы включают в себя: анализ и обобщение специальной литературы, публикаций в периодических изданиях, методы и технологии интеллектуального анализа данных.

Работа состоит из введения, 5 разделов, заключения и списка литературы. Объем работы 72 страницы, содержит 22 рисунка, 11 таблиц, 1 формулу. Список литературы состоит из 42 источников.

1 Современное состояние проблемы анализа данных в информационных системах здравоохранения

«В настоящее время все более актуальной становится задача обработки и анализа больших объемов информации для решения лечебно-диагностических задач и задач управления. В связи с постоянным совершенствованием процесса диагностики и лечения, а также повышением требований к оперативности и доступности информации объем информационных потоков в организации увеличивается, а ее структура все более усложняется.

Эффективное решение задач, связанных с вводом, обработкой и архивированием медицинской информации, практическим управлением информационными потоками, краткосрочным и долгосрочным планированием, статистическим и финансовым анализом требует использования медицинских информационных систем (МИС), позволяющих обрабатывать всю нужную информацию, следуя привычному движению лечения пациента: поступление - диагностика - лечение - реабилитация – наблюдение» [12].

Всестороннее изучение аспектов работы отрасли здравоохранения, а также процесса работы медицинских учреждений в регионах, выявило острую проблему информационной интеграции, в основном связанной с недостаточной автоматизацией процессов электронного документооборота на различных уровнях. Постепенно идет процесс выхода из ручного труда, выстраиваются процессы выдачи и оформления больничных листов, электронных рецептов, автоматизации работы регистратуры, налаживаются связи между информационными системами [9].

Начало было положено с выходом Указа Президента Российской Федерации от 7 мая 2018 г. «О национальных целях и стратегических задачах развития РФ на период до 2024 года» [13], в задачи которого входила разработка эффективных путей трансформации приоритетных отраслей, в том

числе и здравоохранения, путем внедрения цифровых технологий и платформенных решений.

«В рамках государственной программы «Цифровая экономика»] реализуется развитие цифрового здравоохранения и общенационального проекта «Цифровая медицина», в частности, федеральной программы по созданию уникального цифрового контура на базе Единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ). Процесс компьютеризации здравоохранения осуществляется давно, и во многих регионах страны последовательно решаются поставленные задачи. Масштабы этого исследования очень разнообразны и были освещены в нескольких научных статьях.» [8]

Работы С.В. Панасенко [7], Т.О. Бедрина [1] отражают специфику цифровой революции в здравоохранении. Позднее эта идея нашла отражение и в других работах. Все эти исследования позволяют глубже понять подходы к использованию информационных систем и особенности их применения в важных областях здравоохранения.

Стоит рассмотреть, как сейчас устроено отраслевое управление и какую роль играет медицинская статистика. Во-первых, руководитель в сфере здравоохранения, будь то главный врач медицинской организации, специалисты министерства здравоохранения или другие лица принимающие решения в этой сфере, вынуждены работать в ситуации, когда буквально каждый день им нужно принимать сотни различных тактических, а иногда и стратегических решений. При этом все это осуществляется в ситуации, когда доступ к информации ограничен, она меняется очень часто и руководители работают сильными перегрузками [6].

Методология управления на основе данных направлена на то, чтобы повысить эффективность и снизить издержки работы управляемой нами медицинской организации, сети медицинских клиник, региональной или федеральной системы здравоохранения за счет использования цифровых

платформ и более качественного принятия решений, именно в условиях быстро изменяющейся ситуации.

Членом Экспертного совета Минздрава РФ по вопросам использования информационно-коммуникационных технологий Гусевым А.В. [5] были выделены следующие ключевые принципы управления на основе данных:

- измеряемый процесс не должен управляться;
- принятие решений на основе данных подразумевает автоматический сбор этих данных из учетных систем, из первичных данных и медицинских информационных систем;
- отсутствие ручного ввода данных для формирования отчета, если видно, что данные в отчет вносятся вручную, то это не управление на основе данных.

Структура информационных систем в сфере здравоохранения в Российской Федерации, представленная на рисунке 1, состоит из нескольких уровней.

Федеральный уровень: единая государственная информационная система в сфере здравоохранения (ГИС СЗ), в которой есть множество крупных блоков, паспорт медицинской организации, реестр медицинских работников и так далее.

Региональный уровень, который также называется государственной информационной системой сферы здравоохранения: он консолидирует информацию о том, как работают медицинские организации и как используются ресурсы внутри субъекта Российской Федерации.

Учрежденческий уровень: медицинские информационные системы фармацевтических организаций.

Эти 3 уровня позволяют иметь срез информации как на уровне отдельной клиники, субъекта, так и в целом федерации.

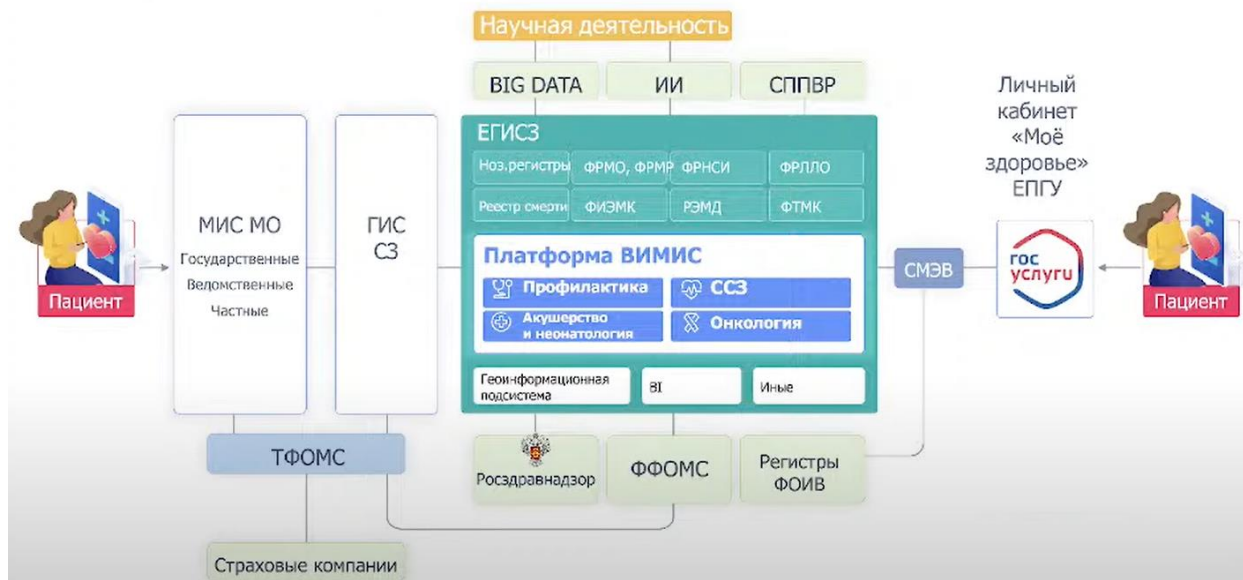


Рисунок 1 – Структура информационных систем в сфере здравоохранения в Российской Федерации

Принятие решение на основе данных идет не только обычным наблюдением, но и с использованием инжиниринга. Инжиниринг подразумевает обязательное применение технологии бережливой поликлиники или бережливого производства. Подробно концепт бережливой поликлиники был рассмотрен Э. Вергазовой [4]. Он заключается в выявлении узких мест, где осуществляются потери времени, ресурсов, качества, других ключевых метрик, перестройке этих процессов, измерении заново процессных результативных показателей и в основном основан на человеческом факторе и качествах управляющего аппарата больницы.

В стратегии цифровой трансформации в рамках программы «Цифровая экономика» ключевым местом является управление системы здравоохранения на основе данных, внедрения систем поддержки принятия управленческих решений, аналитических систем, ситуационных центров и всех инструментов, о которых уже упоминалось ранее [11].

1.1 Известные подходы к применению искусственного интеллекта в здравоохранении

В целях обеспечения медицинских организаций информационной поддержкой решения задач управления и медицинской диагностики предлагается проводить анализ эффективности на основе методов математической статистики и теории принятия решений в областях анализа данных и оценки использования ресурсов, для повышения эффективности деятельности медицинских организаций, а также качества оказания медицинской помощи.

«Оценка эффективности деятельности медицинских организаций в современных социально-экономических условиях обусловлена необходимостью оптимизации процесса управления и оказания медицинских услуг пациентам, а также ростом объема медицинской информации» [12]. Другими словами, эффективность – это зависимость используемых ресурсов и конечного результата.

Аналитика больших данных – это растущая область, способная принести большой вклад в развитие здравоохранения. Несмотря на это, многие аспекты больших данных все еще представляют проблемы и вызывают большую озабоченность в их использовании, например, такие как управление объемом, разнообразием, скоростью, достоверностью, точностью и т.д.

Однако вышеупомянутые проблемы не помешали исследованию больших данных в качестве основного источника данных в здравоохранении. Это обусловлено необходимостью контроля и снижения растущих расходов, а также поиска вариантов для улучшения результатов лечения пациентов. В то время как существует ряд широко разрекламированных примеров использования больших данных в здравоохранении за рубежом, таких как Google Flu и HealthMap, на сегодняшний день нет общей классификации их использования [2].

В последние годы концепция больших данных была введена в систему здравоохранения как решение целого ряда проблем в информационных системах, поскольку они становятся все более сложными и дорогостоящими. Эффективная интеграция таких данных с помощью интеллектуального анализа данных и медицинской информации может привести к снижению затрат и улучшению обслуживания пациентов благодаря принятию обоснованных решений.

Большие данные являются основой для электронных медицинских карт EHR (Electronic Health Records), используемых преимущественно в США. Они способны объединять операционные, финансовые и клинические аналитические системы, таким образом осуществляя поддержку медицины, основанной на доказательных данных [2]. Такой подход предполагает систематический анализ предыдущих клинических данных с целью предоставления информации лицам, принимающим решения.

В статье [20] ученые из Университета Эдит Коуэн провели анализ основных областей использования больших данных в медицине. Собранная литература была отсортирована от высококачественного полезного материала до низкокачественной исследовательской литературы. Высококачественный материал относится к статьям, найденным из рецензируемых журнальных статей, которые содержали соответствующую информацию. Учитывая разнообразие контекста, а также для проведения предварительного анализа, были выделены 5 наиболее используемых областей больших данных в медицине.

Администрирование и снабжение: управление затратами на оказание медицинской помощи. Была отмечена закономерность в отношении подкатегорий, они прямо или косвенно влияли на управление оказанием медицинской помощи.

Поддержка принятия клинических решений: различные компоненты, позволяющие врачам получать дополнительную информацию для помощи в процессе принятия решений.

Клиническая информация: эта категория представляет собой информацию и наборы данных, для последующего анализа, например, информационные системы, хранящие клиническую информацию и так далее.

Поведение/потребитель: в категории "поведение/потребитель" признано использование больших данных для целей демографического анализа, а именно поведение и факторы образа жизни как отдельных людей, так и населения, которые могут быть использованы аналитиками больших данных для составления карты потенциальных вспышек заболеваний среди населения.

Вспомогательная информация: эта категория относится к использованию тех данных, которые нельзя однозначно отнести к одной из перечисленных групп. Либо они превосходят одну категорию и подходят под несколько, либо соответствуют своей собственной категории использования.

1.2 Двойное прочтение рентгенологических исследований

Злокачественные новообразования молочной железы (ЗНМЖ) являются одной из наиболее актуальных проблем современного здравоохранения. Согласно данным ВОЗ на 2022 год ЗНМЖ было диагностировано у 2,3 миллиона человек по всему миру, а количество смертей от данной патологии приравнивается к 685 тысяч женщин. Для России рак молочной железы является наиболее частым онкологическим заболеванием среди женщин, его доля от всех злокачественных новообразований составляет 22%. За последние десять лет заболеваемость выросла на 21,3%, только в 2022 году в России было выявлено более 74 тысяч новых случаев РМЖ [41]. Наиболее часто диагностируемыми типами раков у женщин являются протоковая и дольковая карцинома. ЗНМЖ – это прогрессирующее заболевание, которое становится системным без адекватного лечения на ранней стадии.

На данный момент насчитывается множество методов получения изображений молочных желез, совокупность которых образует современный диагностический комплекс, среди них рентгеновская маммография (МГ),

термография груди (ТГ), магнитно-резонансная томография (МРТ), позитронно-эмиссионная томография (ПЭТ), компьютерная томография (КТ), трехмерное ультразвуковое исследование (УЗИ) и гистопатология (ГП) [25].

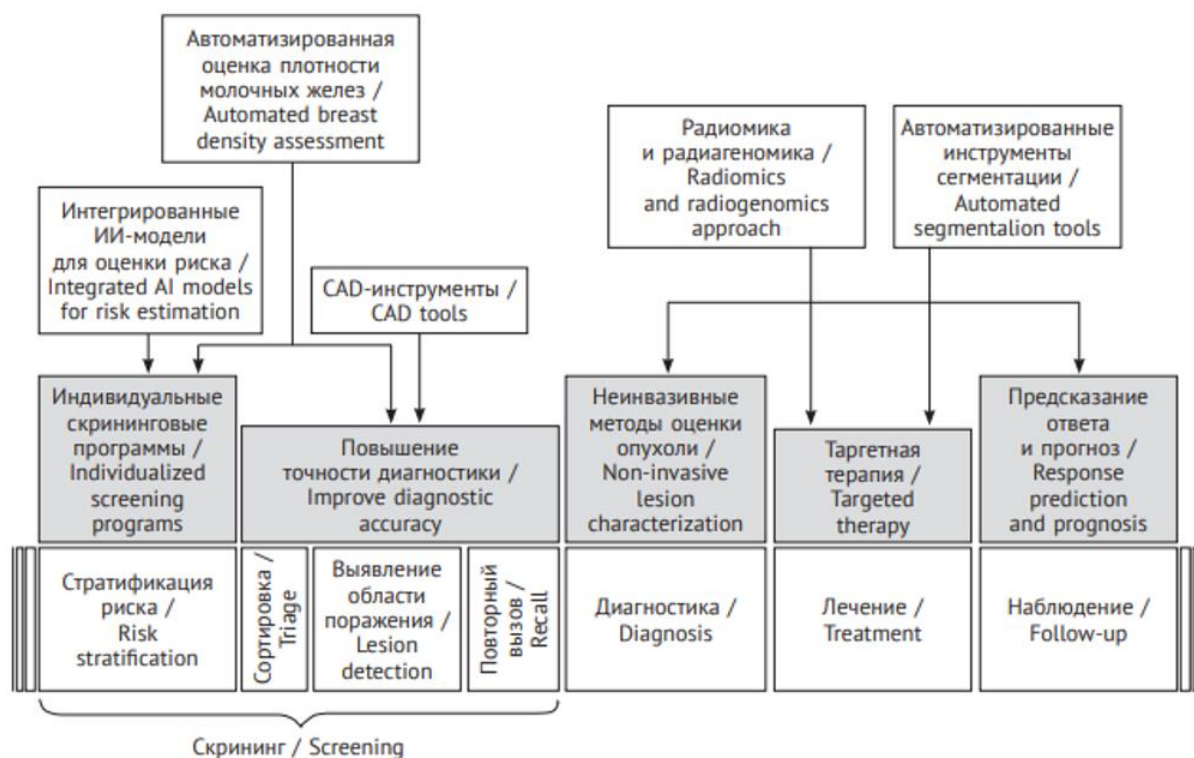


Рисунок 2 – Области применения ИИ для обнаружения рака молочной железы

Искусственный интеллект имеет несколько применений в визуализации рака молочной железы [16, 18-19, 34, 38, 42] (рисунок 2): – стратификация риска для создания индивидуальных программ скрининга; – использование ИИ в качестве «второго чтения»; – анализ текстуры опухоли (определение ее подтипа, оценка перитуморальной области); – радиомический и радиогеномный анализ для планирования лекарственной терапии и динамического наблюдения; – прогнозирование ответа на лечение, риска наличия метастазов в аксиллярные лимфоузлы, риска рецидива и общей выживаемости.

Однако ведущую роль прочно заняла признанная в 1960 году маммография, которая на сегодняшний день является основным методом диагностики лечения заболеваний молочных желез и проводится как по клиническим показателям, так и с целью скрининга. Скрининг охватывает весь процесс, начиная с приглашения на диагностику определенной группы лиц, соответствующей определенным критериям риска и заканчивая непосредственным лечением. Целью любой программы скрининга является снижение смертности, а его основной задачей выявление бессимптомного рака или предшествующих ему состояний у потенциально здоровой группы лиц. В отличие от ранней диагностики ЗНМЖ, во время скрининга предполагается, что у целевой группы искомое заболевание обнаружено не будет.

Скрининг – это сложная система, которая включает в себя несколько этапов (рисунок 3). Сначала формируется группа риска, затем целевому контингенту населения рассылаются приглашения на скрининг. После этого проводится скрининговый тест, результаты которого контролируются, а пациенты с положительными результатами направляются на дополнительное обследование.

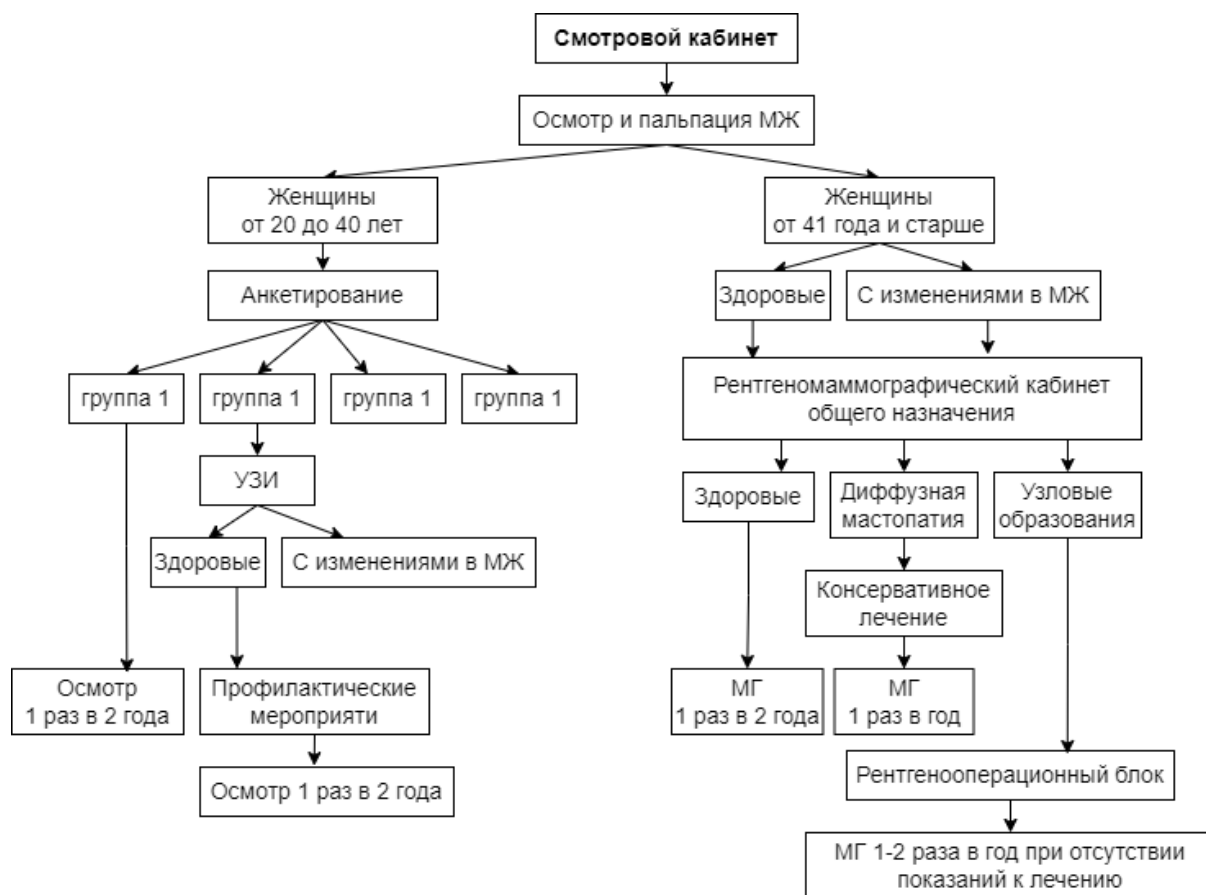


Рисунок 3 – Процесс проведения скрининга

Для подтверждения диагноза в программу скрининга также входит проведение патоморфологического исследования новообразования, что позволяет определить стадию развития опухолевого процесса и обеспечить эффективное лечение. Изначально скрининг проводился на основе экранной пленки с использованием аналоговых систем, но на сегодняшний день весь процесс полностью перешел на цифровые платформы FFDM (fullfield digital mammography). Данный процесс позволил сохранять всю информацию о пациентах в электронных медицинских картах.

Ручной осмотр изображений молочной железы является широко используемым методом, однако некоторые неизбежные факты, могут привести к не точному обнаружению и затянуть процесс диагностики, например:

- недоступность экспертов в отдаленных районах (слаборазвитые страны);
- недоступность экспертов с достаточными знаниями для точного анализа мультиклассовых изображений (изображений с возможными множественными характеристиками заболевания);
- ежедневный осмотр большого количества медицинских изображений может быть утомительной и обременительной практикой;
- тонкая природа опухоли молочной железы и сложная структура тканей молочной железы затрудняют ручной анализ.

Чтение маммограмм это сложный и трудоемкий процесс, который в большинстве своем зависит от опыта врачей. Данный фактор значительно повышает процент получения ложноположительных результатов, что в свою очередь влечет за собой дополнительные расходы в рамках системы здравоохранения. С целью снижения числа ложноположительных результатов в зарубежную практику введено слепое чтение двумя независимыми врачами радиологами, однако это еще больше увеличило рабочую нагрузку.

При проведении скрининга в большинстве европейских стран двойное прочтение рентгенограмм молочной железы является стандартной практикой, выполняемой для сведения к минимуму ложноположительных и ложноотрицательных результатов (рисунок 4). Обзор проводится двумя радиологами либо по отдельности, либо совместно. Если их мнение не совпадает, маммографию смотрит третий специалист или группа специалистов, этот процесс получил название «стратегия арбитража». Частота отзыва (направление на последующее обследование) является важным маркером ложноположительных отзывов и должна быть сведена к минимуму настолько, насколько это возможно. Мета-анализ показал, что увеличение частоты выявления рака при двойном чтении по сравнению с одиночным дало более высокие показатели отзыва, чем одиночное чтение: 3,6% против 3,0% и 6,2% против 4,8% [39]. Не существует последовательной закономерности,

указывающей на то, что какой-либо один подход к отзыву приводит к большему выявлению рака. В целом, ЗНМЖ встречается реже при двойном чтении (0,6-3,0 на 1000 экранов), чем при одиночном (0,9-6,1 на 1000 экранов) [31]. Маммографический вид обычно имеет микрокальцификаты, поэтому второй читатель систематически выявляет более мелкие маммографические признаки и микрокальцификаты, которые могли быть либо пропущены при поиске первым читателем, либо отклонены первым читателем как низкий риск.



Рисунок 4 – Процесс двойного прочтения с арбитражем

Важным фактором, оказывающим влияние на процент понижения выявления рака в программах скрининга, является осведомленность второго читателя решениями первого, что привело к появлению метода «слепого чтения». Ослепление приводит к более независимому принятию решений, поэтому между читателями возникает больше разногласий, что приводит к снижению количества как ложноотрицательных, так и ложноположительных результатов. Метод «слепого чтения» повысил эффективность двойного прочтения (83,1% против 75,5%), а также привел к снижению интервальной частоты рака (1,5% против 3,1%) по сравнению с первоначальным вариантом прочтения [40].

1.3 Риски и выгоды применения искусственного интеллекта в скрининге: перспективы и ограничения

Применение искусственного интеллекта (ИИ) для интерпретации маммографических исследований представляет собой перспективное направление, демонстрируя эффективность как инструмента поддержки принятия решений для радиологов, так и в роли самостоятельного читателя. ИИ в качестве самостоятельного читателя предложен для выявления нормальных случаев, которые можно исключить из анализа радиологом, для замены второго читателя и для выявления случаев высокого риска, требующих более детальной оценки. Ряд исследований по использованию ИИ в качестве инструмента поддержки принятия решений при проведении скрининга с использованием данных о раке показали сокращение времени анализа с сохранением или увеличением точности. Если ИИ сможет облегчить бремя интерпретации, это откроет новые перспективы для внедрения данной технологии в программы скрининга для широких населенных групп. Общая цель данного исследования - ретроспективная оценка эффективности обнаружения рака с использованием имеющейся в продаже системы искусственного интеллекта на материале одноразового широкоугольного скрининга в сравнении с двойным чтением маммографии и двукратным скринингом в качестве эталона. Конкретно, нашей задачей является рассмотрение возможности оптимизации процесса скрининга с использованием искусственного интеллекта с целью уменьшения рабочей нагрузки на радиолога. Это включает в себя выявление нормальных случаев, которые могут быть исключены из дальнейшего анализа, а также возможность замены второго читателя или обоих читателей.

Вариант "AI Gatekeeper", при котором система искусственного интеллекта применяется для исключения нормальных обследований из ручного считывания, в то время как остальные обследования читаются дважды. Проводились оценки при различных пороговых значениях для оценки

ИИ. Особое внимание уделялось уровню порога, при котором 50% обследований были бы исключены, чтобы поддержать постоянную нагрузку на скрининг по сравнению со скринингом маммографии (при условии, что время чтения би-томографии в два раза больше, чем маммографии).

Вариант "Одно считывание + ИИ", где система искусственного интеллекта используется вместо второго считывания, что также приводит к уменьшению общего количества считываний в два раза. Для облегчения сравнения, искусственный интеллект был использован как отдельный считыватель с высоким порогом специфичности, выбранным таким образом, чтобы достичь того же количества консенсусных обсуждений, что и при двойном считывании би-томографии.

Вариант "Только AI", при котором скрининговые исследования анализируются исключительно системой искусственного интеллекта. Система ИИ анализирует все обследования, и они классифицируются как подлежащие консенсусному обсуждению, если оценка ИИ превышает установленное пороговое значение, при этом количество обсуждений остается таким же, как и при двойном считывании би-томографии.

На рисунке 5 представлена графическая схема всех возможных вариаций прочтения рентгеновских снимков, с применением ИИ и без него.

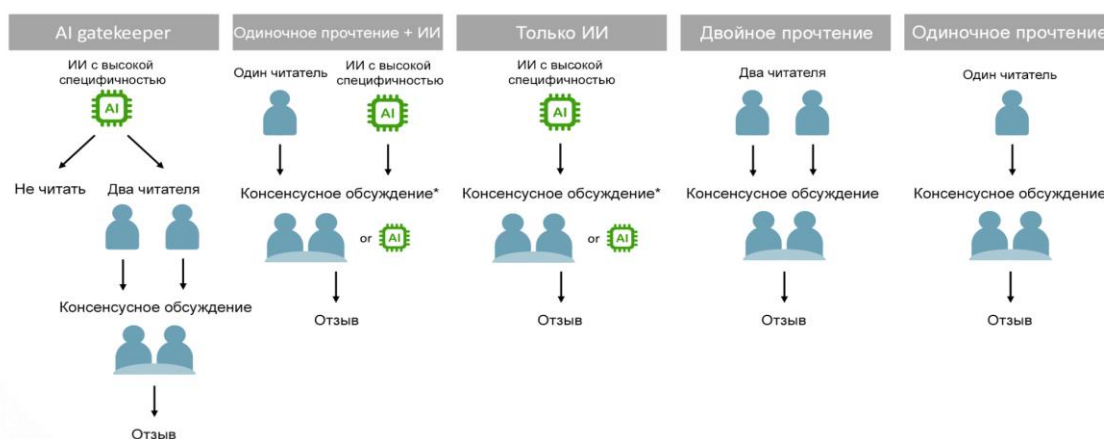


Рисунок 5 – Варианты прочтения маммографии

Результаты фактических консенсусных обсуждений, возникших из двойного считывания, использовались в случаях, когда они были доступны. Однако, в ситуациях, где консенсусное обсуждение было вызвано только искусственным интеллектом (ИИ), фактических консенсусных обсуждений не проводилось. В таких случаях пришлось прибегнуть к использованию заменительных консенсусных обсуждений в рамках рабочих сценариев "одно считывание + ИИ" и "только ИИ".

Для каждой модели проводился анализ количества обнаруженных раков, ложноположительных результатов, пропущенных случаев рака. Базовая истина определялась с использованием всех обнаруженных раков в результате скрининга (маммография + ДБТ) и случаев интервальных раков [37]. Учитывая, что некоторые интервальные раки могут быть выявлены в процессе скрининга, включение их в базовую истину позволяет системе искусственного интеллекта обнаруживать раки, которые могли быть упущены радиологами. Количество и характеристики пропущенных раков были также подвергнуты анализу. Полученные результаты также были сопоставлены с текущим стандартом скрининга, который предусматривает двукратное считывание маммографии.

1.4 Применение методов нейронных сетей в проведении двойного прочтения рентгенографии ЗНЖС

Одним из наиболее ожидаемых изменений в области скрининга рака молочной железы является возможность использования глубокой нейронной сети в качестве одного из посредников в двойном прочтении маммограмм. Существует множество потенциальных ролей для ИИ, однако на сегодняшний день известны лишь случаи замены второго читателя с сохранением правила арбитражности.

В статье [35] эффективность двойного прочтения с использованием ИИ оценивалась с помощью симуляции с использованием исторических мнений

читателей. В таблице 1 видно представлены статистические тесты, которые показали, что двойное прочтение было как минимум не хуже, а в некоторых случаях и вовсе превосходило показатели частоты отзыва. Если рассматривать понятия чувствительности и специфичности в рамках машинного обучения, то они будут подразумевать вероятность получения истинно положительных и истинно отрицательных результатов соответственно. В рамках маммографических исследований чувствительность означает способность модели на базе искусственного интеллекта выявлять патологии, если они имеются, в свою очередь специфичность будет отвечать за истинно отрицательные результаты.

Таблица 1 – Эффективность двойного прочтения с использованием ИИ

Показатель эффективности	Исторические данные двойного прочтения	Данные двойного прочтения с ИИ	Результаты сравнения
Частота отзыва	3.8% (3.8, 3.9)	3.8% (3.7, 3.9)	Не уступают 0.99 (0.98, 1.01)
Чувствительность	86.1% (84.5, 87.6)	83.9% (82.3, 85.6)	Не уступают 0.98 (0.97, 0.98)
Специфичность	97.1% (96.9, 97.2)	97.1% (97.0, 97.3)	Превосходят 1.00 (1.00, 1.00)

Следует отметить, что некоторые виды микрокальцификатов, которые были бы истинно положительными для ИИ, на самом деле распознаются как ложноположительные, что приводит к более низким значениям чувствительности и специфичности.

Влияние внедрения ИИ в данном случае будет зависеть от нескольких факторов. Во-первых, требуется применение метода «слепого чтения». Если читатели не будут «ослеплены», они могут слишком сильно положиться на систему ИИ и согласовать свои решения с ее решением. Во-вторых, если существует арбитраж несопадающих решений в случае одного читателя и системы ИИ, точность арбитража может быть снижена, если арбитражный(ые) читатель(и) слишком сильно или слишком слабо уверен(ы) в точности ИИ. В-

третьих, если отзыв происходит при любом из предложенных читателями вариантов, то потребуется очень высокая специфичность системы ИИ, чтобы предотвратить неуправляемо высокий уровень отзыва. В-четвертых, изменчивость между считывателями. В системах, где считыватели с очень высоким коэффициентом отзыва систематически объединялись со считывателями с низким коэффициентом отзыва для контроля общего коэффициента отзыва, добавление третьего считывателя в качестве системы ИИ может привести к увеличению коэффициента отзыва.

Таким образом процесс двойного прочтения будет выглядеть следующим образом (рисунок 6):

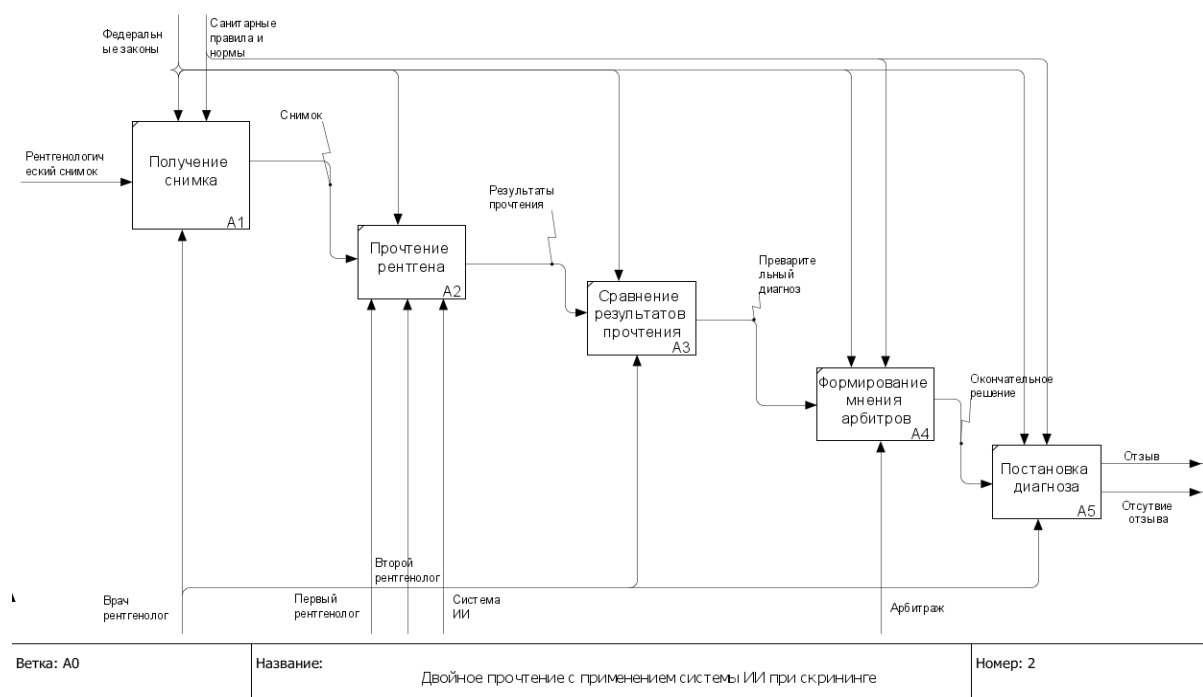


Рисунок 6 – Процесс двойного прочтения с системой ИИ и арбитражем

Предполагается, что внедрение системы ИИ в качестве второго независимого читателя позволит сохранить все преимущества использования нейронных сетей в прочтении рентгенологических снимков, а также поможет сократить риск появления ложноположительных и ложноотрицательных ОТЗЫВОВ.

2 Компьютерная диагностика в маммографии на основе искусственного интеллекта

В анализе и диагностике медицинских изображений искусственный интеллект (ИИ) успешно способствует распознаванию сложных моделей изображений для обеспечения лучшей количественной оценки автоматизированным и надежным способом. Существует две категории алгоритмов ИИ, а именно алгоритмы, использующие созданные вручную характеристики, и алгоритмы, обрабатывающие первичные данные. Первая категория алгоритмов ИИ относится к стандартному или обычному ИИ, в то время как вторая категория алгоритмов ИИ относится к современным подходам глубокого машинного обучения.

2.1 Стандартные алгоритмы машинного обучения

Машинное обучение (МО) – это термин, впервые введенный Артуром Самуэлем в 1959 году, который описал его как подобласть ИИ. Это метод, который распознает закономерности на основе заданных входных данных, изучает эти закономерности без явного программирования и решает задачи на их основе. Для медицинской организации входными данными. В случае медицинской визуализации в качестве входных данных будут использоваться вручную извлеченные особенности изображений, такие как форма, площадь, текстура и так далее. Далее извлеченные признаки могут быть обработаны алгоритмами отбора признаков для выбора наиболее релевантных и, наконец, задача алгоритмов МО заключается в объединении выбранных входных признаков в единое значение, такое как сигнатура опухоли, которая может представлять вероятность состояния заболевания (рисунок 7).

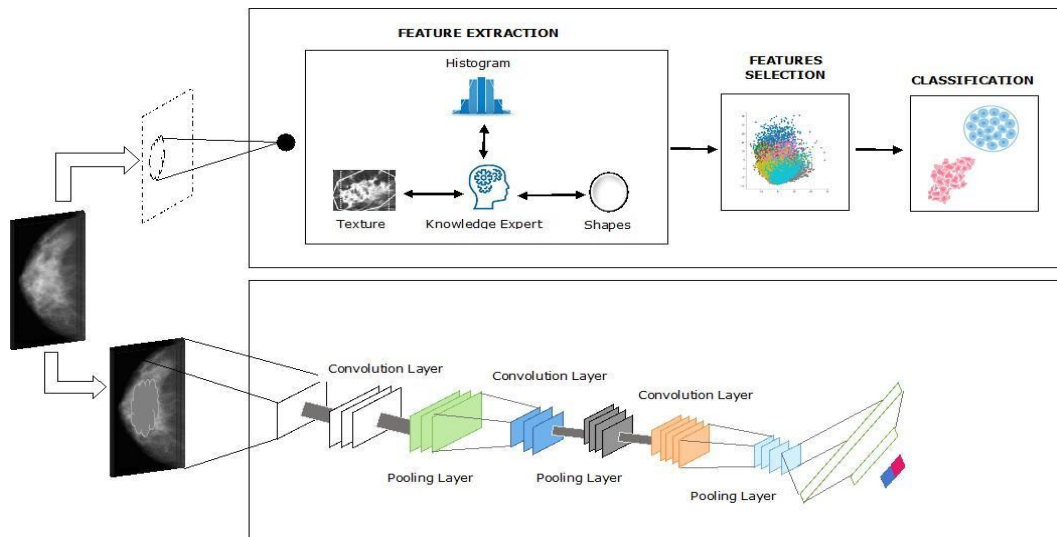


Рисунок 7 – Изучение маммографии методами машинного обучения и глубокого обучения

Алгоритмы МО делятся на два основных типа – алгоритмы контролируемого обучения и алгоритмы неконтролируемого обучения (рисунок 8). Эти два типа далее делятся на классификацию, регрессию (контролируемое обучение) и кластеризацию (неконтролируемое обучение) на основе получаемых результатов.

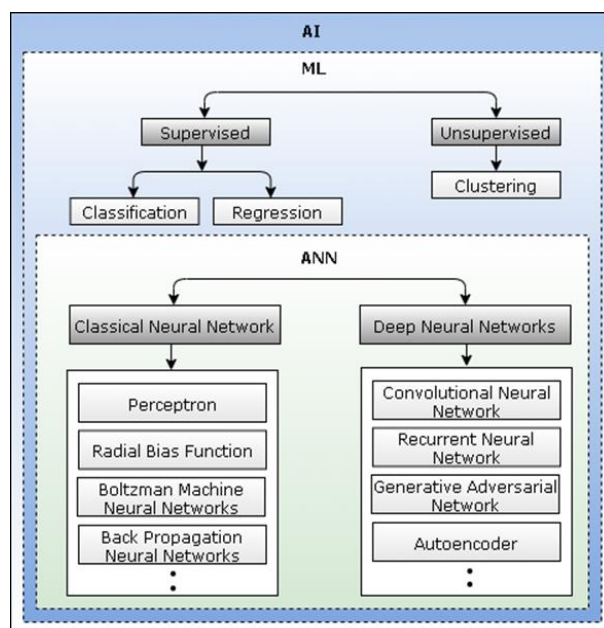


Рисунок 8 – Категории алгоритмов искусственного интеллекта

При контролируемом обучении алгоритм обучается на помеченных данных, другими словами, на основе заданного распределения вероятностей классов (прогнозируемого заболевания) или предыдущих знаний. При неконтролируемом обучении алгоритм изучает шаблоны из данных, связывая их с обнаруженными кластерами точек данных. Для обнаружения и анализа рака молочной железы использовались различные алгоритмы МО, основанные на ручной обработке признаков, в частности, дерево решений (DT), метод случайного леса (RF), метод опорных векторов (SVM), наивный байесовский классификатор (NB), метод k-ближайших соседей (KNN), линейный дискриминантный анализ (LDA) и логистическая регрессия (LR) [23].

Хотя созданные вручную признаки считаются дискриминационными по своей природе, и алгоритмы МО, основанные на таких признаках, в некоторых случаях достигли замечательных результатов в области медицинской визуализации, извлечение таких признаков всегда является сложной задачей, поскольку для этого требуются знания экспертной области, а для того, чтобы сделать их подходящими для алгоритмов МО, необходима всесторонняя переработка [33].

Однако зависимость от экспертных определений повышает шансы упустить некоторые важные факторы, из-за чего точность диагностического процесса может быть низкой. Кроме того, созданные вручную признаки фокусируются только на одном аспекте, например, на текстуре или на краях изображений. Так, например, алгоритм, извлекающий признаки по форме, не сможет извлечь признаки по текстуре. Таким образом, системы, обученные на одном наборе элементов данных, не смогут работать с другим набором данных и проводить дискриминационный анализ. Следовательно, всегда существует необходимость в классификаторе для принятия классификационных решений и обработки полученного пространства признаков. Однако выбор подходящего классификатора всегда является сложной задачей.

2.2 Алгоритмы искусственного интеллекта на основе глубокого обучения

Для устранения ограничений, присущих алгоритмам МО, последние исследования направлены на подходы, основанные на глубоком обучении (DL). Алгоритмы глубокого обучения работают с необработанными данными и не требуют ручного извлечения признаков для получения результата. Они способны адаптироваться ко всем видам признаков входных данных, поэтому обеспечивают более надежные результаты по классификации и сегментации, чем традиционные методы машинного обучения. Алгоритмы глубокого обучения также успешно применяются для оценки и анализа медицинских изображений, таких как КТ и МРТ. Они возникли на основе искусственной нейронной сети (ANN), которая является подмножеством алгоритмов машинного обучения. Искусственная нейронная сеть имитирует работу биологических нейронов, другими словами, простую работу мозга. На рисунке 9 изображена основная структура искусственного нейрона, которая показывает, что взвешенные входы подаются на нейрон для получения выхода или предсказания. Нейрон суммирует полученные входы и применяет нелинейную функцию активации для получения выходного ответа (обычно между 0 и 1).

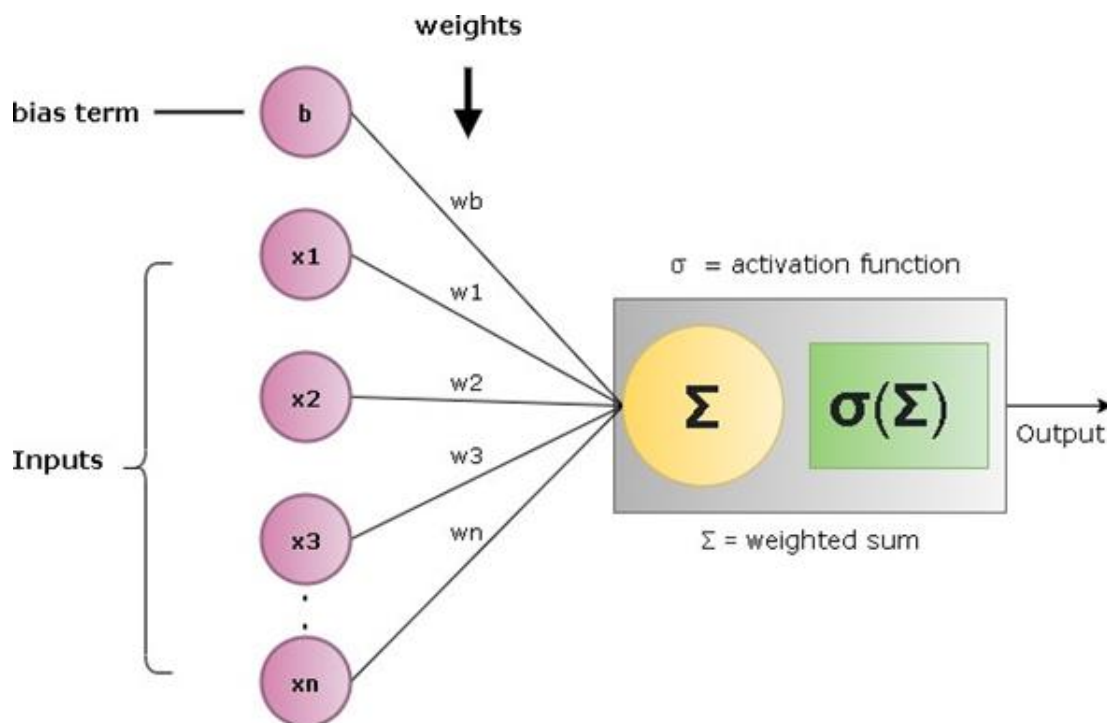


Рисунок 9 – Базовая структура нейрона

«В ANN узлы (нейроны) соединены друг с другом в виде одного или нескольких иерархических слоев и могут посылать и принимать сигналы, как показано на рисунке 10. Реакция (отклонение или принятие) этих посылаемых и принимаемых сигналов зависит от выхода нелинейной функции активации. С входом каждого нейрона или узла сети связан вес, который может повлиять на данный вход и полезен для передачи данных на выходной слой. Входной слой передает данные в виде вектора признаков со взвешенным значением скрытому слою. Скрытый слой, состоящий из активационных единиц, переносит вектор признаков из первого слоя со взвешенным значением и выполняет вычисления в качестве выхода. Выходной слой состоит из активационных единиц, каждая из которых соответствует метке/классам, присутствующим в наборе данных, переносит взвешенный выход скрытого слоя и предсказывает соответствующий класс. ANN использует функцию обратного распространения на этапе обучения для уменьшения функции

ошибки. Ошибка уменьшается путем обновления значений весов в каждом слое.»[7]

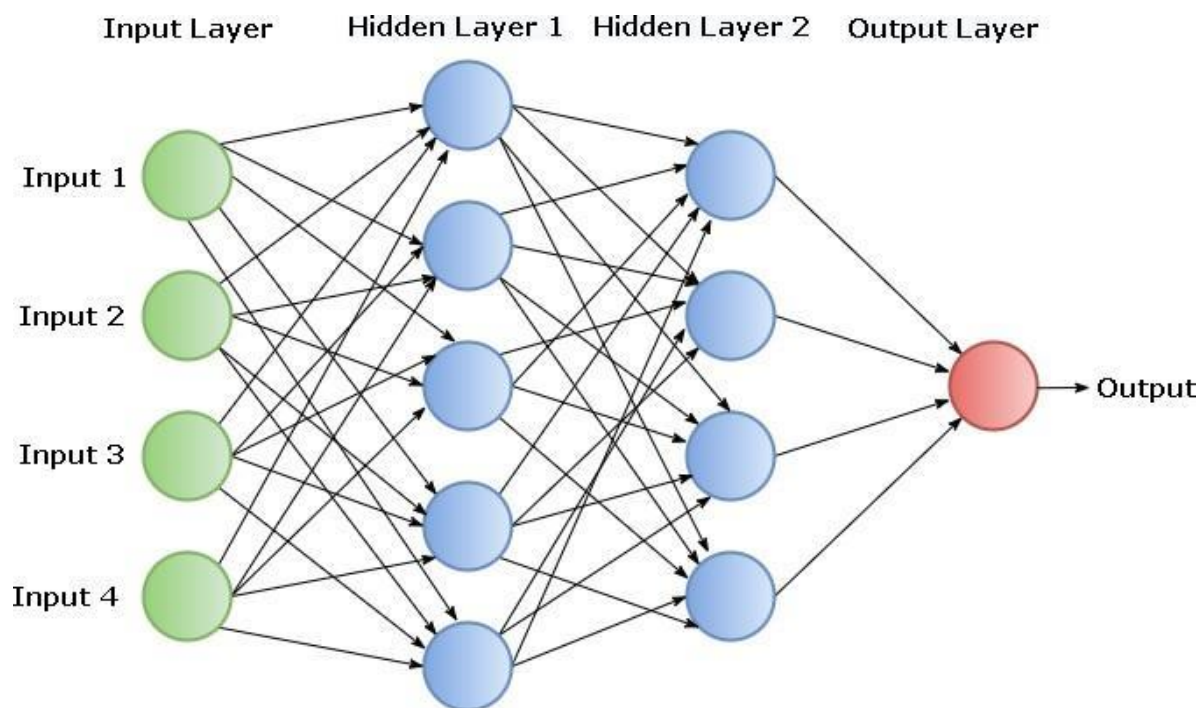


Рисунок 10 – Схема работы искусственной нейронной сети

«Для классификации рака молочной железы в основном используются два типа ANN, а именно: неглубокая нейронная сеть (SNN) и глубокая нейронная сеть (DNN). Среди этих двух нейронных сетей, ГНС является наиболее популярным типом ANN, который широко используется для анализа и диагностики медицинских изображений. DNN обладает способностью автоматически извлекать соответствующие характеристики из исходных данных без каких-либо экспертных знаний и вмешательства человека. Благодаря вышеупомянутым характеристикам, подходы глубокого обучения обеспечивают существенное улучшение процессов диагностики, анализа и принятия клинических решений с использованием данных медицинской визуализации.»[7]

После получения передовых результатов в области анализа изображений, в литературе было использовано множество вариантов DNN для

классификации рака молочной железы, таких как многослойная сеть (ML-NN), сеть глубокого доверия (DBN), метод главных компонент (PCA.Net), автоэнкодер (SDAE), а также различные варианты сверточной нейронной сети (CNN).

В таблице 2 приведены сильные и слабые стороны моделей DNN при использовании их для анализа медицинских изображений.

Таблица 2 – Сравнительная характеристика моделей глубокой нейронной сети

Модель	Описание	Преимущества	Ограничения
SNN	Однослойная прямая сеть со скрытым слоем	Из-за небольшого размера сети требуется меньше вычислительных ресурсов и времени на обучение. Возможность получения приемлемой производительности в случае небольшого набора данных. Хорошая результативность на данных низкой размерности.	Низкая производительность при работе с данными высокой размерности и при решении многоклассовых задач. Высокая вычислительная сложность при множестве гиперпараметров. Требуются большие наборы данных для обучения.
SDAE	Эта модель извлекает дискриминантные скрытые паттерны из данных, используя внутренний метод реконструкции	Обладает способностью подавления шума. Легкая регуляризация и оптимизация параметров обучения. Способность автоматического устранения шумов помогает выделять значимые признаки.	Низкая производительность при работе с данными малой размерности. Плохая корреляция между измерениями. Свойство автоматического шумоподавления иногда не подходит для данных низкой размерности.
DBN	Генеративная модель, состоящая из нескольких слоев, обучаются по методу послойного обучения признаков. Все скрытые слои обучаются по одному слою за раз.	Послойное обучение помогает улучшить обобщение признаков. Легко оптимизировать гиперпараметр каждого слоя. Слои сети могут быть обучены в неконтролируемой манере, что очень требовательно в классификации ЗНЖС.	Не может отследить потери.

Продолжение таблицы 2

Модель	Описание	Преимущества	Ограничения
CNN (De novo)	Различные слои (входной, сверточный, объединяющий, полностью связанные слои) иерархически организованы настраиваемым образом. Модель обучается с нуля под наблюдением.	Индивидуальная модель может быть разработана в зависимости от типа и количества изображений. Предпочтительна, когда исходных изображений недостаточно для обучения.	В случае наборов данных из разных областей, трудно оптимизировать обучение модели. Сложно решать многоклассовые задачи, если количество доступных изображений невелико
CNN (Pre-trained)	Модель обучается с помощью трансферного обучения с использованием предварительно обученных доступных сетей.	Модель может быть обучена быстрее, чем de novo, в случае меньшего количества доступных ресурсов. Показывает приемлемую производительность в случае небольших целевых данных.	Получение ненадежных результатов в случае очень маленьких наборов данных. Сложно обрабатывать и оптимизировать в случае новых добавленных слоев

Среди вышеупомянутых архитектур, CNN является наиболее мощной, эффективной и широко используемой для анализа медицинских изображений, в особенности изображений молочной железы.

2.3 Применение сверточной нейронной сети для анализа изображений молочной железы

Как уже говорилось ранее, сверточная нейронная сеть (CNN) – это подтип глубоких нейронных сетей с прямой передачей данных, которые показали надежные результаты при анализе изображений. В отличие от обычной DNN, архитектура CNN может принимать на вход целые данные. Для ее обучения обычно используются два метода: трансферное обучение (TL) и метод de novo. В методе de novo архитектура CNN обучается с нуля, после чего различные обученные модели комбинируются для получения оптимальной, в

то время как в методе трансферного обучения для анализа используются уже предварительно обученные модели. Трансферное обучение – это удобный метод, однако он в основном работает с ограниченными или небольшими наборами данных.

Архитектура CNN состоит из иерархических слоев с обучаемыми фильтрами (операция свертки) и операциями объединения (уменьшение размера представления изображения), которые захватывают дискриминационные признаки из необработанных входных данных для повышения эффективности классификации. Далее, слой классификации вычисляет вероятность или оценку изученных классов на основе извлеченных признаков. Иллюстрация глубокой CNN (DCNN, т.е. CNN с большим количеством сверточных и объединяющих слоев или), примененной для классификации изображений молочной железы (для обнаружения рака молочной железы), представлена на рисунке 11.

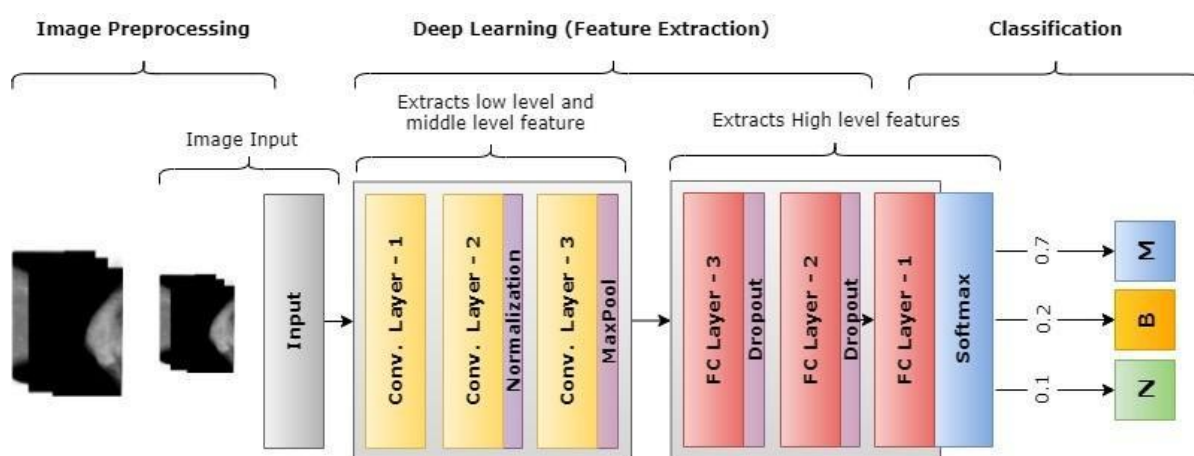


Рисунок 11 – Архитектура глубокой CNN, применяемая для обнаружения и классификации рака молочной железы

Для анализа изображений молочной железы (или любых других медицинских изображений) входной слой в архитектуре CNN принимает входное изображение. Сверточные слои содержат фильтры, известные как ядра. Эти слои используют операции свертки для захвата высокоуровневых

характеристик, таких как края, цвета, формы и пятна. Количество слоев и сверточных фильтров иерархически расположены таким образом, что уровень извлечения от низкоуровневых признаков к высокоуровневым возрастает по мере увеличения глубины слоев.

В таблице 3 представлен обзор избранных статей [34], опубликованных за последние 4-5 лет, в которых использовались методы глубокого обучения (в частности, на основе CNN) для обнаружения рака молочной железы. Общей чертой всех этих статей является то, что почти все (за исключением нескольких) представили результаты с использованием набора данных маммограмм для доказательства надежности предложенной схемы.

Таблица 3 – Сравнительная характеристика методов глубокого обучения

Исследовательская группа	Классификация	Источник данных	Кол-во изображений	Тип ГНС и его архитектура	Размер изображений	Точность
Ragab и др. [32]	Доброкачественные и злокачественные	DDSM	5257	DCNN: (Alexnet[5 Conv+2 FC] + SVM Classifier)	227 x 227	87.20
Ionescu и др. [21]	Раковые, нераковые	PROCAS	73128	DCNN:4-conv2D layer with 1-fc layer	640 x 512	61.00
Singh и др. [36]	Форма: Неправильная, дольчатая, овальная, круглая	InbreastD DSM	410 inbreast, 1168 DDSM	GAN+CNN: 3 Conv Layer+2 FC+ Softmax Classifier	64 x 64	83.00
Junior и др. [22]	Обнаружение архитектурных искажений	Личное FFDM	280	VGG-16:13-Con2D,4 pooling, 3 Dense	224 x 224	89.00

Продолжение таблицы 3

Исследовательская группа	Классификация	Источник данных	Кол-во изображений	Тип ГНС и его архитектура	Размер изображений	Точность
Lee и др. [24]	Нормальные, доброкачественные, злокачественные	MIAS	322	DCNN+RPN:4 - con2d (Dense+Relu+batchNorm+Dropout)+Softmax	256 x 256	97.00
Altaf и др. [14]	Нормальные, аномальные доброкачественные, аномальные злокачественные	DDMS, InbreastB CDR	900 DDM S 300 Inbreast 450 BCD R	Transfer Learning Based PCNN (Pulse-Coupled Neural Network)+DCNN	224 x 224	98.72 (DDMS), 97.50 (InBreast),
Khamparia и др. [23]	Нормальные, аномальные доброкачественные, аномальные злокачественные	DDMS patch	10713	MVGG16 + ImageNet	224 x 224	94.30
Lee и др. [24]	Нормальные, доброкачественные, злокачественные	MIAS	322	DCNN+RPN:4 - con2d (Dense+Relu+batchNorm+Dropout)+Softmax	256 x 256	97.00

Если проанализировать таблицу 3, можно сделать вывод, что различные исследователи использовали глубокую сверточную нейронную сеть (DCNN) для выполнения различных задач при анализе рака молочной железы, поскольку архитектуры DCNN достигают самых современных результатов. Некоторые исследования были посвящены классификации структуры и геометрии тканей молочной железы. Другие исследователи классифицировали аномальную структуру ткани как злокачественную и доброкачественную. Из таблицы также видно, что производительность DNN в основном зависит от

глубины архитектуры, наличия больших наборов данных, вычислительных ресурсов и больших размеров входных данных для обучения.

Современные зарубежные и российские сервисы ИИ, направленные на раннее выявление РМЖ, представлены в таблице 4. Первой компанией, получившей знак Европейского сертификата качества за применение методов глубокого обучения в радиологии, стала Kheiron Medical Technologies и их система Mia (Mammography intelligent assessment). В ее основе лежат запутанные нейронные сети, а обучение проходило на более чем 3 млн изображений, что позволяет ей с большой долей вероятности выявлять рак на ранних стадиях. Система работает следующим образом: исходное изображение анализируется с предположением, что либо никаких действий больше не требуется, либо пациенту потребуются дальнейшее обследование. Окончательный диагноз не ставится без контроля врача.

Таблица 4 – Сравнительная характеристика известных систем ИИ, направленных на выявление РМЖ

Параметр	Система ИИ				
	Mia	AlexNet	Lunit	CureMetrix	Botkin.AI
Приоритизация данных	+	Нет данных	Нет данных	+	+
Автоматическое формирование протокола	+	Нет данных	Нет данных	Нет данных	-
Автоматическое позиционирование молочной железы	+	Нет данных	Нет данных	Нет данных	-
Двойное прочтение	+	+	+	+	+
Результаты	Чувствительность 76,1% Специфичность 96,3%	Точность 93,2%	Точность 95,0%	AUC 0,95	Нет данных
Выявление РМЖ при повышенной плотности ткани	Нет данных	Нет данных	+	+	-
Присвоение категории BI-RADS	Нет данных	Нет данных	+	Нет данных	+

«Другая методология была представлена Н. Li et al. [27], где предварительно обученная модель AlexNet и функция извлечения были применены к набору изображений цифровых ММГ из двух групп высокого риска, а также здоровых контрольных групп. Авторы показали, что без какой-либо дальнейшей точной настройки ММГ-изображений признаки первого полностью связанного слоя модели могут эффективно отличать здоровую контрольную группу от обеих групп высокого риска. Сервис Lunit дает возможность дистанционно загрузить исследования и проанализировать работу модели. Данные о выявленных патологиях отображаются в виде тепловых карт. До 96% AUC Lunit Insight ММГ точно обнаруживает на ММГ поражения, подозрительные на РМЖ [28]. Сервис CureMetrix имеет преимущество в оценке анализа рентгенологических изображений молочных желез с высокой плотностью тканей и наблюдении выявленных изменений в динамике» [29].

3 Методология исследования и используемые данные для обработки

Сбор данных является важным этапом в подготовке и обучении нейронной сети системы двойного прочтения рентгеновских снимков. Этот процесс направлен на создание обширного, разнообразного и репрезентативного набора данных, который будет использоваться для обучения модели. Рассмотрим подробнее, зачем необходим сбор данных и что это представляет собой в данном контексте:

Для обучения и тестирования системы, основанной на искусственном интеллекте, используются разнообразные рентгеновские снимки молочных желез, собранные из различных медицинских и онкологических источников (рисунок 12). Набор должен включать изображения высокого разрешения, охватывающие различные формы и стадии рака молочной железы, такие как опухоли, микрокальцинаты, кисты и другие характерные признаки [17].

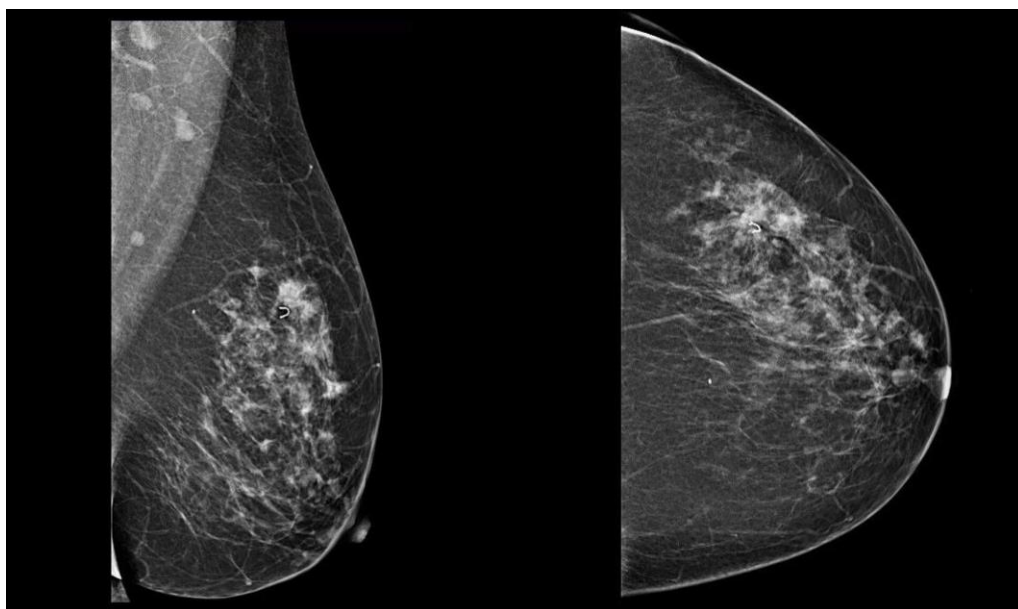


Рисунок 12 – Пример изображения для обработки

Помимо рентгеновских снимков, база должна включать в себя анонимизированные медицинские записи пациентов, включая данные об

истории болезни, результаты лабораторных тестов и предыдущие диагнозы. Эти данные являются важными для обогащения обучающего набора и повышения точности классификации системы в области выявления и анализа рака молочной железы.

В рамках исследования будут применены методы аугментации изображений для предотвращения переобучения модели и расширения разнообразия обучающего набора данных (рисунок 13). Такие техники, как вращение, отражение и изменение контраста, использовались с целью создания более устойчивого и разнообразного обучающего набора, способного повысить обобщающую способность модели [15].

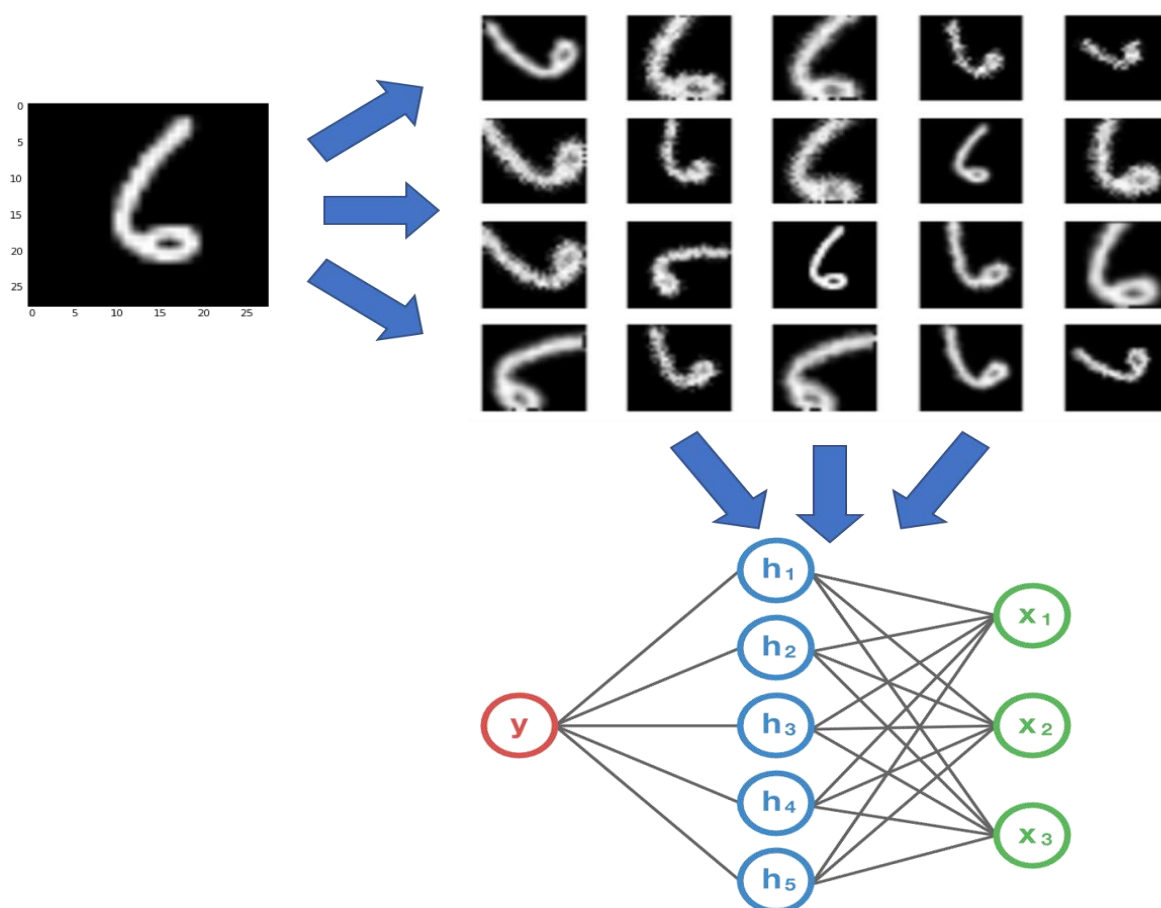


Рисунок 13 – Аугментация изображений

Также необходимо осуществить нормализацию изображений с целью снижения влияния изменений в условиях освещения и характере съемки на результаты классификации. Этот этап направлен на повышение стабильности работы модели при различных условиях эксплуатации.

Выбор архитектуры нейронной сети: в ходе исследования рассмотрены сверточные нейронные сети (CNN), включая предварительно обученные модели, такие как ResNet и Inception. В таблице 5 будет проведен сравнительный анализ с собственно разработанными архитектурами для выявления оптимальной структуры сети, соответствующей поставленной задаче.

Таблица 5 – Сравнительный анализ моделей

	Преимущества	Недостатки
Предварительно обученные модели (ResNet и Inception)	Предварительно обученные модели, такие как ResNet и Inception, обладают высокой точностью на стандартных наборах данных, что может ускорить обучение и повысить производительность. Они успешно выявляют сложные признаки в изображениях, что важно для анализа рентгеновских снимков.	Возможно, что предварительно обученные модели не полностью соответствуют специфике задачи по двойному прочтению рентгеновских снимков, требуя дополнительной настройки.
Собственно разработанные архитектуры	Позволяют более гибкую адаптацию к специфике задачи и типу данных, используемых в исследовании. Могут быть оптимизированы для конкретного вида изображений и требований к обработке данных, что важно в медицинских приложениях.	Требуют более продолжительного обучения и большего объема данных, чем предварительно обученные модели, чтобы достичь высокой производительности.

Сравнительный анализ указывает на то, что использование предварительно обученных моделей, таких как ResNet и Inception, может

обеспечить высокую начальную точность, особенно при ограниченных ресурсах. Однако собственно разработанные архитектуры предоставляют большую гибкость и специализацию для конкретной задачи, что может привести к лучшей адаптации к медицинским изображениям и требованиям двойного прочтения рентгеновских снимков.

Обучение и тестирование: Данные будут разделены на обучающий, валидационный тестовые наборы. Будет проведено несколько итераций обучения, с использованием стандартных методов оптимизации, таких как стохастический градиентный спуск. Процесс обучения должен быть проконтролирован с использованием метрик точности, чувствительности и специфичности.

В рамках исследования активно используется метод кросс-валидации, играющий ключевую роль в оценке устойчивости и обобщающей способности разрабатываемой модели двойного прочтения рентгеновских снимков. Кросс-валидация становится фундаментальным инструментом, направленным на повышение достоверности результатов и обеспечение устойчивости модели в различных сценариях использования.

В рамках данной работы, цель которой заключаются в создании более точного и эффективного метода медицинской диагностики, кросс-валидация является неотъемлемой частью процесса тестирования и оценки предлагаемой модели. Ее применение обеспечивает проверку модели на различных подмножествах данных, что важно для устойчивого и эффективного функционирования в реальных клинических условиях.

Для тщательной оценки эффективности разработанной модели двойного прочтения рентгеновских снимков, будет проведена работа со стандартными метриками, такими как точность, полнота и F1-мера, чтобы не только измерить ее производительность, но и получить более глубокое понимание ее способностей к точной диагностике выявления рака молочной железы.

Этот методический подход охватывает весь цикл работы с данными — от их сбора до окончательного анализа. Применение стандартных метрик

предоставляет нам обширный инструментарий для измерения эффективности модели, в то время как комплексный и системный подход дает возможность глубокого понимания того, как наша модель взаимодействует с различными аспектами диагностического процесса.

3.1 Подготовка данных для обучения нейронной сети

Первоочередной задачей будет являться выборка методологии и алгоритмов, специально адаптированных под поставленную задачу, а именно избавление от выбросов и артефактов на рентгеновских изображениях, таких как снимки с низким разрешением, искаженные фотографии или снимки с неверной меткой класса, аномалии в распределении цветовых каналов и экстремальные значения яркости.

Необходимо провести валидацию медицинских данных для проверки их корректности и соответствия стандартам. Удаление дубликатов и записей с недостаточной информацией, оцифровку по необходимости.

Данные необходимо разделить на обучающие, валидационные и тестовые наборы: 70% данных были отнесены к обучающему набору, что обеспечит модели широкий спектр входных данных для обучения на разнообразных примерах. Дополнительные 15% данных выделены для валидационного набора, который используется в процессе настройки гиперпараметров, обеспечивая улучшенную устойчивость и оптимизацию модели. Оставшиеся 15% данных составили тестовый набор, на котором проводится финальная оценка производительности модели. Этот метод разделения данных имеет критическое значение для достоверной оценки модели, поскольку он гарантирует, что она тестируется на реальных сценариях и данных, которые не входили в обучающий процесс. Этот этап позволяет избежать переобучения и обеспечивает модели возможность обобщения на новые, ранее не виденные данные, что важно для успешного применения в практике медицинской диагностики.

Также к каждому изображению по возможности необходимо прикрепить метаданные из медицинских записей, такие как симптомы, характеристики заболеваний и лабораторные результаты для создания комплексного восприятия моделью.

Балансировка классов в обучающих данных играет немаловажную роль. Это способствует более равномерному обучению модели, что важно для точности ее диагностических предсказаний в разнообразных клинических сценариях.

Этот детальный процесс подготовки данных обеспечивает надежную основу для обучения нейронной сети, учитывая как рентгеновские изображения, так и медицинские записи, и обеспечивая их интеграцию для более глубокого анализа и повышения эффективности модели.

3.2 Архитектура нейронной сети. Слои, функции активации и параметры

Выбор оптимальной архитектуры нейронной сети оказывает существенное воздействие на способность модели обрабатывать и анализировать мультимодальные данные, такие как изображения и текст. Разрабатывая архитектуру, мы должны включать в нее совокупность сверточных и рекуррентных слоев, обеспечивая эффективное взаимодействие с данными различных модальностей. Сверточные слои специализируются на обработке изображений, выделяя визуальные особенности, в то время как рекуррентные слои заботятся о текстовой информации, улавливая ее последовательные и контекстуальные аспекты. Интеграция сверточных и рекуррентных слоев позволяет модели создавать более глубокие и информативные представления данных, снимая ограничения, связанные с обработкой отдельных модальностей. Этот подход существенно улучшает способность к адаптации к разнообразным входным данным, следовательно, повышает ее точность.

Для извлечения признаков из рентгеновских изображений необходимо использовать несколько сверточных слоев с целью захвата различных уровней абстракции в изображении. После сверток используется пулинг для уменьшения размерности данных и повышения устойчивости к мелким изменениям (рисунок 14).

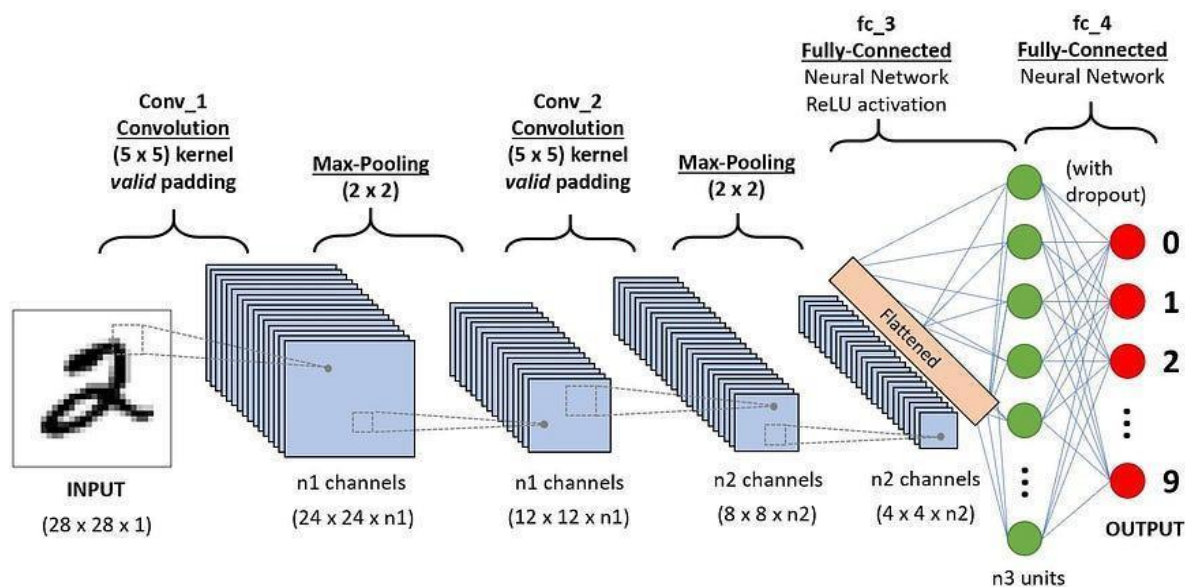
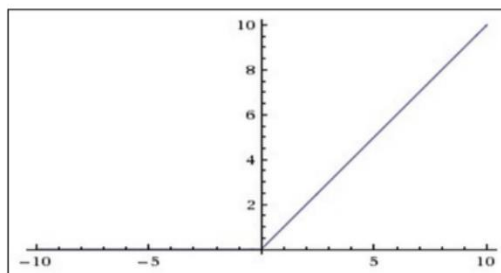


Рисунок 14 – Сверточная нейронная сеть

Основываясь на необходимости извлечения сложных и нелинейных зависимостей в данных, сверточные слои архитектуры нейронной сети применяют функцию активации ReLU (Rectified Linear Unit), представленную на рисунке 15. Эта функция не только способствует эффективному обнаружению и выделению важных признаков в изображениях рентгеновских снимков, но также обеспечивает более быструю сходимость в процессе обучения. Для предотвращения возможного переобучения модели к обучающим данным применяются методы регуляризации, а именно слои исключения (Dropout), которые случайным образом "выключают" определенные нейроны во время обучения, тем самым предотвращая слишком

сильную специализацию нейронов на конкретные признаки и повышая обобщающую способность модели.

ReLU (rectified linear unit)



$$f(s) = \max(0, s)$$

$$f'(s) = \begin{cases} 1, & s > 0 \\ \text{rand}(0.01, 0.05), & s \leq 0 \end{cases}$$

Рисунок 15 – ReLU, функция активации в нейронных сетях

Дополнительно, для улучшения способности модели фокусироваться на ключевых моментах текстовой информации, нужно внедрить механизм внимания. Этот механизм позволяет модели динамически выделять и уделять больше внимания различным частям входных данных, повышая точность и интерпретируемость результатов.

Последний слой нейронной сети снабжен функцией softmax. Она применяется для преобразования выходных значений модели в вероятностные распределения по различным классам. Softmax обеспечивает нормализацию и превращает выходы сети в значения, интерпретируемые как вероятности принадлежности объекта к каждому классу. Такой подход позволяет получать четкие и интерпретируемые результаты в задачах многоклассовой классификации.

Для оптимизации параметров сети используется стохастический градиентный спуск (SGD) с адаптивным шагом обучения (learning rate). Стохастический градиентный спуск позволяет эффективно минимизировать

функцию потерь, обновляя параметры сети на основе случайных подвыборок данных. Адаптивный шаг обучения позволяет контролировать скорость обучения, что важно для стабильности и быстроты сходимости процесса обучения. Выбор функции потерь имеет ключевое значение для успешной обучаемости модели. В данном случае используется подходящая функция потерь. Это может быть, например, бинарная кросс-энтропия, учитывая, что задача связана с бинарной классификацией или детекцией аномалий.

Разработка оптимальной архитектуры для двойного прочтения рентгеновских снимков включала тщательную настройку различных гиперпараметров, чтобы обеспечить высокую производительность модели. Эксперименты должны проводиться на валидационном наборе данных, где осуществлялся подбор оптимальных значений для параметров, таких как количество сверточных и рекуррентных слоев, размер фильтров, количество нейронов в плотных слоях, а также коэффициенты регуляризации и dropout. Количество сверточных слоев определяет способность модели выделять важные признаки на изображениях, в то время как рекуррентные слои обеспечивают анализ текстовых данных, учитывая контекст медицинских записей. Размер фильтров и количество нейронов в плотных слоях влияют на сложность и емкость модели, что требует тщательной настройки для достижения баланса между производительностью и избеганием переобучения.

Итоговая архитектура нейронной сети объединяет в себе преимущества сверточных и рекуррентных архитектур для достижения точных и надежных результатов в задаче двойного прочтения рентгеновских снимков.

3.3 Описание процесса обучения нейронной сети. Метрики используемые для оценки производительности системы.

С целью еще более эффективной обработки изображений и текста в архитектуре должны быть созданы отдельные ветви, специализированные на каждом типе данных. Это предоставляет модели дополнительные

возможности для улучшения диагностических результатов, так как она более тщательно анализирует каждый тип в изоляции и в контексте.

Для улучшения интерпретируемости модели и акцентирования внимания на ключевых моментах медицинских записей должен быть внедрен механизм внимания (Attention) в рекуррентные слои текстовых данных. Внедрение механизма внимания призвано сделать модель более внимательной к важным элементам, что в свою очередь улучшит ее способность делать взвешенные и информативные выводы.

Таблица 5 – Цели и ожидаемые результаты от метрик

Аугментация текстовых данных	
Цель	Цель данного этапа заключается в усилении обучения модели на разнообразных вариантах медицинской терминологии путем аугментации текстовых данных. Это включает в себя применение случайных перестановок и замен слов в текстах медицинских записей.
Ожидаемые выгоды	Аугментация текстовых данных способствует улучшению обобщающей способности модели, обучая ее более гибко реагировать на новые термины и различные вариации описаний симптомов. Этот процесс также содействует созданию модели, способной более эффективно адаптироваться к разнообразным стилям медицинских записей.
Тюнинг параметров регуляризации	
Цель	Основной целью этого этапа является внимательная настройка параметров регуляризации, таких как коэффициенты dropout и L2-регуляризации. Это необходимо для предотвращения переобучения модели, особенно при наличии ограниченных объемов данных.
Ожидаемые выгоды	Тщательная настройка параметров регуляризации позволяет модели повысить свою устойчивость к перекосам и аномалиям.

Продолжение таблицы 5

Инициализация	
Цель	Цель этапа инициализации заключается в использовании предварительно обученных моделей для инициализации весов нейронной сети. Это направлено на ускорение процесса сходимости и повышение эффективности обучения, особенно при ограниченном объеме данных.
Ожидаемые выгоды	Инициализация весов с использованием предварительно обученных моделей способствует улучшению обучения на ограниченном объеме данных. Модель может быстрее достигнуть оптимальных результатов, так как веса уже настроены на определенные паттерны, что особенно важно при ограниченных ресурсах данных.

Для более глубокого и разностороннего обучения нужно внедрить синтетические данные. Основные метрики производительности представлены в таблице 6.

Таблица 6 – Метрики производительности

Метрики	Определение	Применение
Точность (Accuracy):	Точность — это метрика, измеряющая общую правильность классификации модели. Это соотношение правильно классифицированных объектов ко всем объектам в наборе данных.	Точность особенно полезна для оценки общей эффективности системы, когда классы в данных сбалансированы. Эта метрика позволяет оценить, насколько точно модель предсказывает все классы.
Полнота (Recall) и Точность (Precision):	Полнота измеряет, как много реальных положительных случаев было правильно обнаружено моделью, в то время как точность измеряет, насколько много из предсказанных положительных случаев действительно являются положительными.	Полнота и точность предоставляют дополнительный уровень детализации, оценивая способность модели находить все случаи конкретного заболевания и определять точность этих обнаружений.

Продолжение таблицы 6

Метрики	Определение	Применение
F1-мера (F1-score):	F1-мера представляет собой гармоническое среднее полноты и точности, обеспечивая баланс между этими двумя метриками.	Эта метрика дает комплексную оценку, учитывая и полноту, и точность.
AUC-ROC (Площадь под ROC-кривой):	AUC-ROC измеряет площадь под кривой ROC и предоставляет оценку способности модели разделять классы, особенно в условиях дисбаланса.	Эта метрика дополняет оценку, предоставляемую точностью и полнотой, и обеспечивает дополнительную информацию о способности системы различать различные классы заболеваний.
Матрица ошибок:	Матрица ошибок представляет собой детальный анализ ложноположительных и ложноотрицательных случаев для каждого класса.	Анализ матрицы ошибок позволяет выявить особенности ошибок модели, что может быть полезно для их дальнейшего улучшения и оптимизации.
Визуализация результатов:	Визуализация результатов включает создание графиков и визуализаций, таких как кривые обучения, матрицы ошибок, ROC-кривые.	Эти визуализации обеспечивают наглядную оценку производительности системы, что полезно для понимания ее работы и выявления возможных улучшений.

Процедура разделения данных:

- модель обучается на тестовом наборе данных с целью выявления закономерностей и особенностей данных;
- гиперпараметры модели, такие как количество слоев, нейронов и другие, настраиваются на валидационном наборе для оптимизации производительности и предотвращения переобучения;
- после настройки модель оценивается на пробном наборе данных для объективной оценки ее производительности и генерализации на новых данных.

Следующий шаг отказаться от извлечения признаков и перенаправить усилия на создание новых моделей, способных автоматически извлекать признаки из изображений, а не оптимизировать и настраивать гиперпараметры и стратегии обучения существующих алгоритмов. Наиболее эффективным способом достижения этой цели в настоящее время являются сверточные нейронные сети (CNN), которые воспринимают исходные изображения маммограмм в двумерном пространстве, а не извлеченные из них признаки, где теряется вся пространственная информация. Генерация синтетических изображений и текстовых данных позволит модели обучаться на более разнообразных вариантах раковых состояний и улучшит ее обобщающую способность. Этот подход был согласован с основной целью исследования – создать модель, способную эффективно работать с разными формами данных и обеспечивать точное двойное прочтение рентгеновских снимков.

4 Апробация данных и перспектива развития

В медицинских учреждениях используется типовая конфигурация аппаратно-программных средств (рисунок 16). Она включает в себя:

- сервер региональной медицинской информационной системы (РМИС), находящийся непосредственно в медицинской организации (МО);
- сервер региональной радиологической информационной системы (РРИС).;
- клиентские компьютеры (АРМ медицинских работников);
- клиентские компьютеры (АРМ рентгенолога);

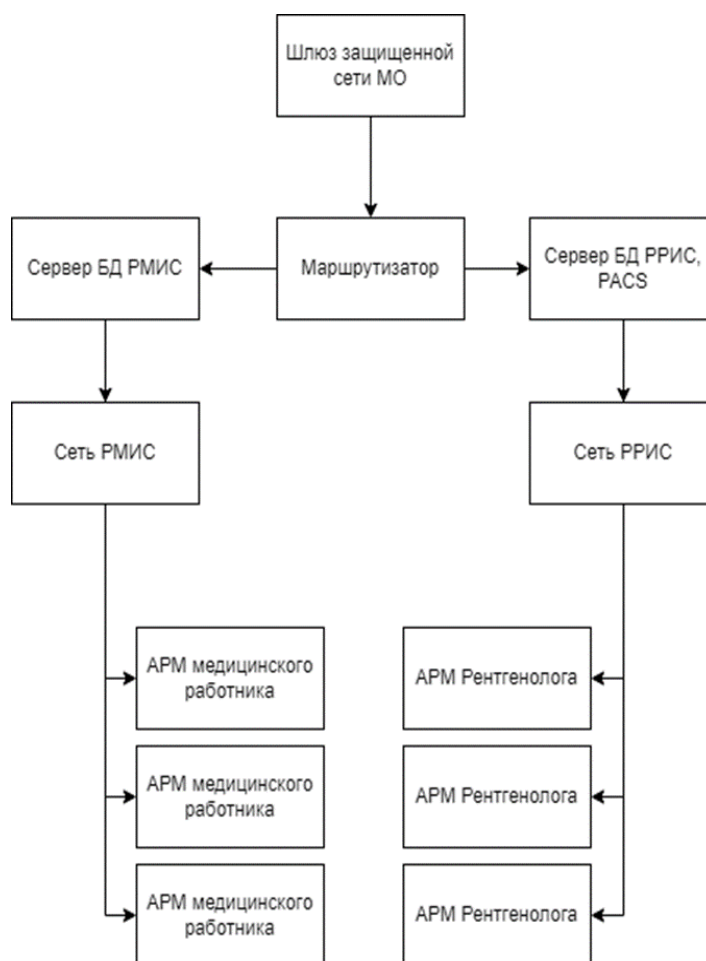


Рисунок 16 – Типовая схема инфраструктуры медицинской организации

Для лучшего понимания взаимодействия между системой и врачом, а также для ясного представления этапов окончательного вывода о заболевании, мы использовали диаграмму BPMN (Business Process Model and Notation). Эта диаграмма визуализирует ключевые шаги, обеспечивая наглядное представление процесса взаимодействия.

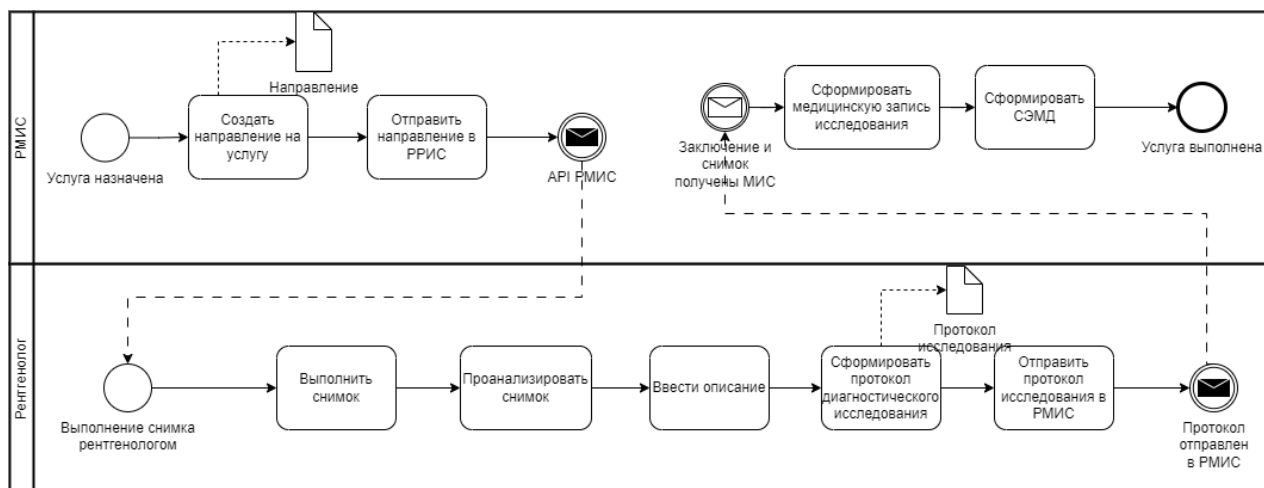


Рисунок 17 – Модель прочтения врачом рентгенологом «Как есть»

Визуализация процесса модели «Как есть» (рисунок 17):

- иллюстрация этапов взаимодействия врача и пациента в процессе анализа рентгеновских снимков для более глубокого понимания врачебной консультации;
- наглядная постановка и указание ролей врача и пациента, а также четкое определение их обязанностей на различных этапах взаимодействия;
- представление последнего этапа, на котором формируется окончательный вывод о заболевании.

Как правило выполнение снимка и само прочтение маммографии выполняется одним и тем же врачом, а хранение данных и последующая привязка результата с медицинскими данными происходит на серверах РМИС.

Структура диаграммы BPMN:

- обозначение начала процесса анализа рентгеновских снимков между врачом и нейронной сетью;
- показ шагов, на которых нейронная сеть предоставляет врачу результаты анализа снимков;
- представление этапа анализа медицинских записей пациента для получения полной клинической картины;
- выделение этапа, на котором объединяются результаты анализа изображений и медицинских записей для создания комплексного вывода (мультиmodalный анализ);
- иллюстрация шагов консультации врача на основе полученных результатов, включая формулирование выводов и рекомендаций;
- представление процесса, на котором формируется окончательный вывод о заболевании, учитывая все аспекты мультиmodalного анализа;
- обозначение завершения процесса анализа рентгеновских снимков и формирования окончательного диагностического вывода.

Также чтобы подробнее понять динамику и важность процесса двойного прочтения рентгеновских снимков с использованием нейронной сети, необходимо рассмотреть расширенную диаграмму вариантов использования. Расширенная диаграмма вариантов использования служит фундаментом для более глубокого понимания взаимозависимости участников процесса (рисунок 18).

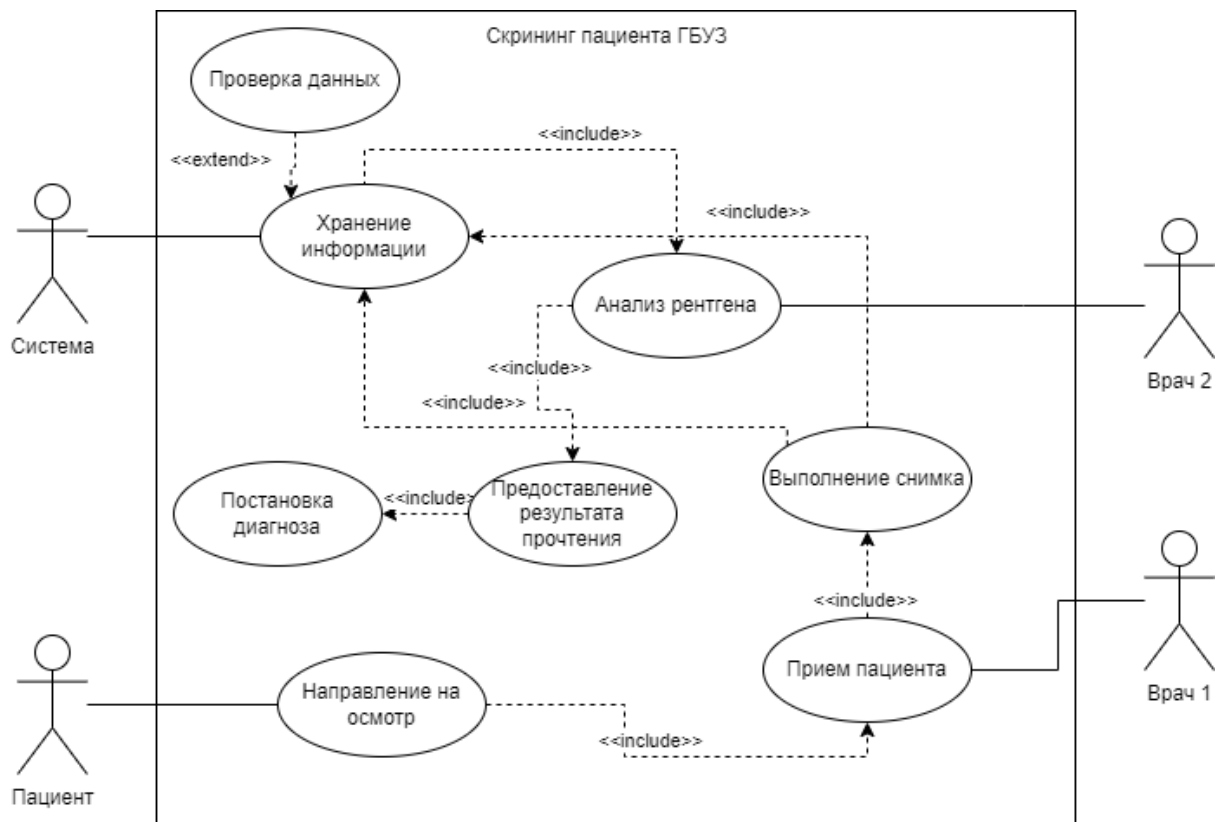


Рисунок 18 – Диаграмма процесса проведения скрининга

Представленная диаграмма выполняет следующие важные функции:

- позволяет визуализировать многоэтапный, сложный процесс совместной работы врача и нейронной сети, начиная с передачи данных и заканчивая формированием окончательного диагноза;
- подчеркивает и выделяет важные моменты сотрудничества, где взаимодействие данных и обмен информацией с врачом играют критическую, важную роль в достижении точности диагноза;
- основа для понимания ролей и задач: определяет роли и ответственности как врача, так и искусственного интеллекта на каждом этапе процесса, создавая и показывая понятную картину основ взаимодействия;
- является фундаментом для успешного внедрения метода двойного прочтения, облегчая взаимопонимание и согласование действий участников.

Согласно части 1 статьи 10 Федерального закона № 152-ФЗ «О персональных данных», информация о состоянии здоровья относится к особой категории персональных данных. Именно поэтому, вся сетевая инфраструктура медицинских организаций должна быть обеспечена техническими мерами по обеспечению информационной безопасности.

Двойное прочтение с использованием искусственного интеллекта в качестве второго независимого читателя реализуется следующим образом (рисунок 19):

- врач через медицинскую информационную систему (МИС) формирует для пациента направление на услугу в электронном виде;
- направление передаётся в РРИС (региональная рентгеноинформационная система) с помощью API (Application Programming Interface);
- пациент направляется на снимок. Для создания проекта в РРИС рентген-лаборанты сопоставляют (связывают) направление и цифровой снимок;
- средствами РРИС снимок отправляется на независимое описание двум врачам-рентгенологам;
- при готовности двух описаний РРИС формирует протокол исследования и передаёт его в РМИС (региональную медицинскую информационную систему), где он доступен для ознакомления направившему врачу;
- в МИС формируется структурированный электронный медицинский документ, который направляется в региональное хранилище РС ЕГИСЗ БО и затем, в порядке очереди, — на регистрацию в хранилище Минздрава;

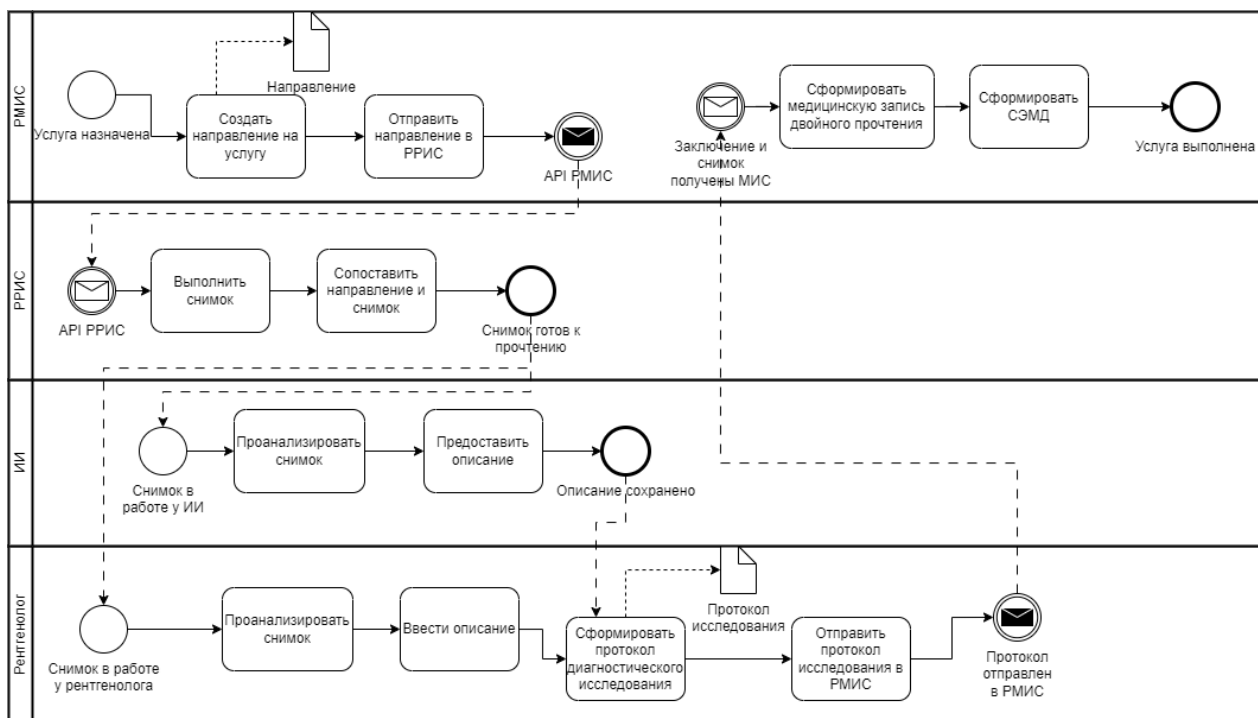


Рисунок 19 – Разработанная модель двойного прочтения с ИИ

Модуль сбора данных отвечает за сбор, предварительную обработку и передачу данных на следующие этапы системы.

База данных результатов хранит результаты двойного прочтения, включая диагнозы, вероятности и другую метаинформацию. Располагается в специализированном хранилище данных, обеспечивающее сохранность и доступность результатов.

Интерфейс пользователя позволяет врачам взаимодействовать с системой, загружать данные, просматривать результаты и вносить коррективы. Располагается на медицинских компьютерах, планшетах или рабочих станциях, подключенных к системе.

Диаграмма развертывания, которая поможет визуализировать систему и модули, подключаемые к ней, а также интерфейсы и рабочие станции пользователей, представлена на рисунке 20.

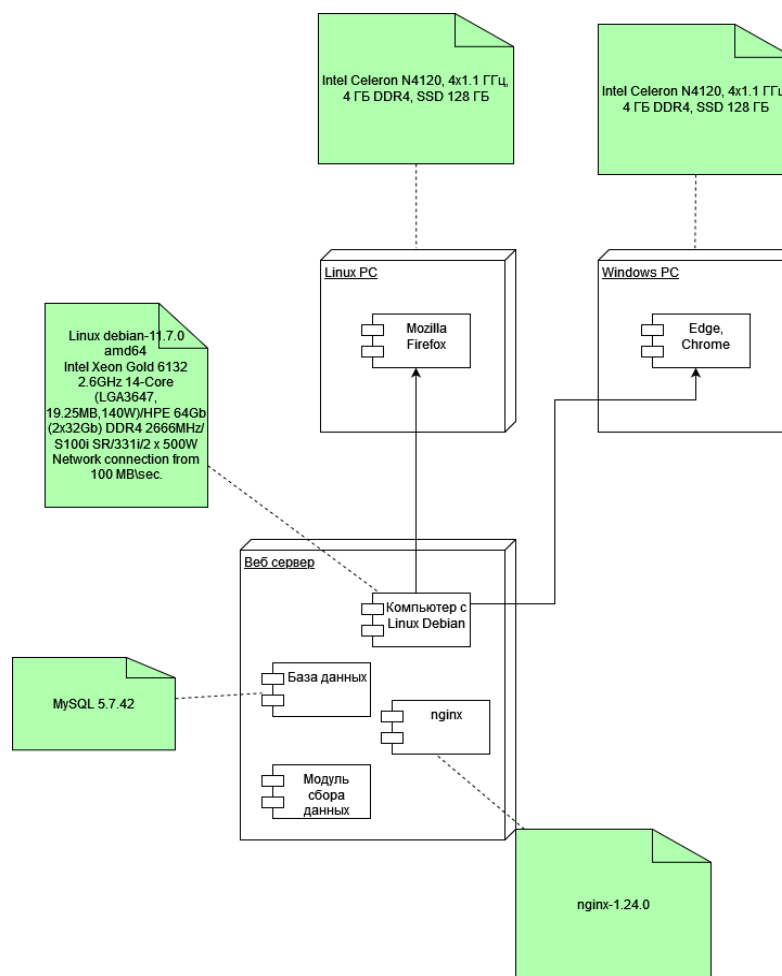


Рисунок 20 - Диаграмма развертывания

Система управления данными обеспечивает централизованное управление данными, их хранение, резервное копирование и обеспечивает безопасность данных. Располагается на специализированном сервере или облачном хранилище.

Сервер приложений обрабатывает запросы от интерфейса пользователя, взаимодействует с моделью нейронной сети и базой данных результатов. Сервер, обеспечивающий выполнение бизнес-логики и обработку запросов.

4.1 Оценка основных показателей и тестируемые данные

Исследование проводилось на готовом наборе данных mini-MIAS, состоящем из 2,750 маммограмм и включающем в себя историческое описание патологий. Программное обеспечение принимало на вход цифровые изображения молочных желез в формате DICOM, проводило их анализ и предоставляло результаты в виде текстовых описаний (DICOM SR), изображений с обозначением патологических областей (DICOM SC) и вероятности наличия патологических изменений в исследуемой области. Результаты в последствии должны быть интегрированы в ЕРИС ЕМИАС. В рамках исследования анализировались показатели вероятности наличия патологии (в частности, вероятность злокачественного процесса), а также оценивалась корректность присвоения рейтинга BI-RADS (стандартизированная шкала оценки результатов маммографии) и точность определения локализации изменений.

В скрининге наиболее информативными показателями являются частота отзыва и выявления рака, однако для оценки системы ИИ следует опираться на характеристики, приведенные в таблице 7.

Таблица 7 – Основные характеристики оценки РМЖ при использовании системы ИИ

Показатель	Сильные стороны	Ограничения
Частота отзыва	Стандартная метрика скрининга. Измеряется для всей группы, включая неподтвержденные случаи. Равносилен специфичности.	Актуально только для несбалансированных групп.

Продолжение таблицы 7

Показатель	Сильные стороны	Ограничения
Частота выявления рака (ЧВР)	Стандартная метрика скрининга. Равносилен чувствительности. Интуитивно понятный.	Зависит от программы скрининга.
Чувствительность	Стандартная метрика диагностики.	Не является стандартной метрикой для скрининга. Снижается с увеличением количества данных по ИР. Зависит от истинно положительных результатов.
Специфичность	Стандартная метрика диагностики.	Зависит от истинно отрицательных результатов. Не является стандартной метрикой для скрининга.
Положительная прогностическая оценка (ППО)	Стандартная метрика диагностического исследования, которая иногда применяется во время скрининга.	В зависимости от используемого определения (по всем отзывам или по определенным), может означать верхнюю или нижнюю границу.
Частота арбитража	Измеряется по всей группе, включая неподтвержденные случаи, а не только подтвержденные положительные и подтвержденные отрицательные.	Результаты зависят от характера арбитражного процесса, который может отличаться.

Результаты работы системы ИИ основаны на одном прогнозе для каждого случая. Система использовала заранее определенные пороги принятия решения "отзыв" или "отсутствие отзыва", которые были установлены на основе исторических данных, полученных от пациентов, не включенных в исследование. Таким образом происходил баланс метрик чувствительности и специфичности, так как при расчете показателя

чувствительности учитываются только выявленные скринингом раковые заболевания.

Чувствительность, частота выявления рака (ЧВР) и положительное прогностическое значение (ППЗ) определялись в совокупности как "положительные результаты, выявленные скринингом" и "интервальные раки (ИР)". К положительным результатам выявления злокачественной опухоли относились те случаи скрининга, которые после исторического двойного прочтения подтверждались рядом исследований в течение 180 дней после скринингового обследования.

4.2 Исследуемая группа и процесс прочтения

В таблице 8 представлены характеристики исследуемой группы. Из 2,750 общих случаев было обнаружено 272 (1,0%) положительных результатов, из которых 210 (0,84%) были выявлены скринингом. Наблюдаемая доля ИР составила 0,14%.

Таблица 8 – Характеристика исследуемой группы

Характеристики		Кол-во случаев	Доля исследуемого населения
Всего		275,900	100,0%
Возраст	<40	483	0,2%
	40-49	37,696	13,7%
	50-59	114,524	41,5%
	60-69	98,289	35,6%

Продолжение таблицы 8

Характеристики		Кол-во случаев	Доля исследуемого населения
Всего		275,900	100,0%
Возраст	70-79	23,359	8,5%
	80-89	15	0,6%
	>90	2,683	<0,1%
Позитивные случаи	Всего	2,310	0,97%
	Скрининг	373	0,84%
	Интервальные	289	0,14%

На этапе определения пороговых значений выборка данных была сокращена до 100 снимков, из которых 50 определены как наличие патологии и 50 как ее отсутствие. При выставлении порогового значения в 0,93, что соответствует 100,0% специфичности, отмечалось отсутствие ложноположительных результатов. Таким образом, ИИ корректно определил 29 снимков (55,0%) из категории «Присутствие патологии». Аналогичная ситуация наблюдается и при выставлении порогового значения в 0,25, что соответствует 100% чувствительности и группе «Отсутствие патологии» (28 снимком, 56,0%), с полным отсутствием ложноотрицательных результатов.

Если руководствоваться методом индекса Юдена [38], при пороговом значении 0,82 значение чувствительности составило 80%, количество истинно отрицательно результатов – 45, истинно положительных 40, что соответствует 92% и 81% верно определенных результатов всех исследований, определенных ранее групп соответственно. При такой настройке системы получаем чувствительность 81,0% (95% ДИ 67,8–92,1), специфичность 92,0% (95% ДИ 82,76–97,3), точность 85,0% (95% ДИ 77,8–92,1).

Для визуального представление данных были построены соответствующие графики посредством ROC-анализа (метод построения и анализа характеристической кривой), с помощью web-инструмента (<https://roc-analysis.mosmed.ai/>) [10, 30]. Результатом являлась вероятность наличия рака, где 0 – отсутствие, 1 – наличие патологии (рисунок 21).

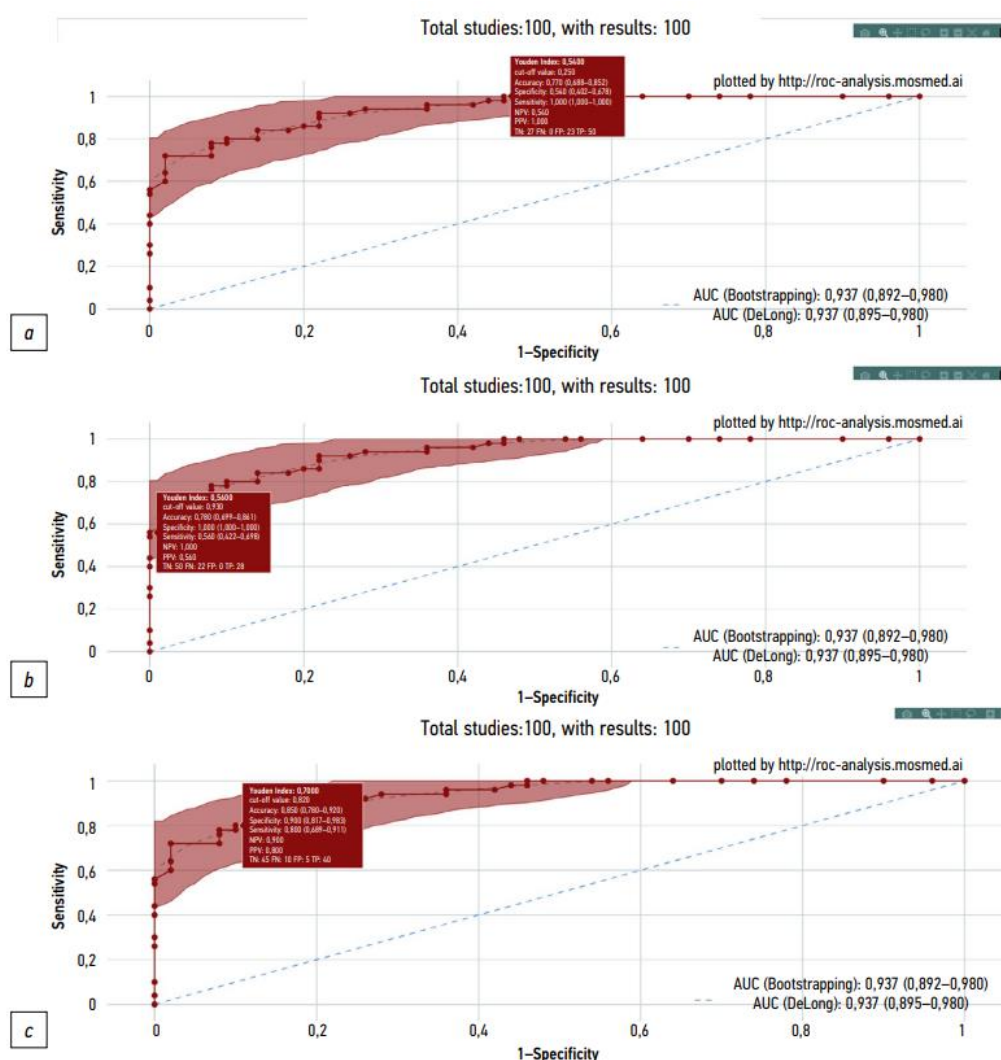


Рисунок 21 – Визуализация пороговых значений

Применение программного обеспечения, основанного на технологии искусственного интеллекта (ИИ), для услуги «Описание и интерпретация данных маммографического исследования с использованием ИИ» позволяет настроить пороговые значения таким образом, чтобы показатели

чувствительности, специфичности и точности соответствовали или даже превосходили результаты, достигаемые при двойном чтении маммограмм врачами-рентгенологами. Эта настройка различных пороговых уровней обеспечивает более точное выявление патологических изменений, что подтверждается многочисленными исследованиями и практическими данными.

Распределение маммографий по группам в зависимости от выбранного порогового значения представлено в таблице 9.

Таблица 9 – Группировка данных по пороговым значениям

Количество снимков	Результат	100% чувствительность	100% специфичность	Сбалансированные значения
50	Истинно положительный	50	29 55%	40 81%
	Ложноположительный	22	0	6
50	Истинно отрицательный	28 56%	50	45 92%
	Ложноотрицательный	0	21	9

Для всех показателей были представлены двусторонние 95% доверительные интервалы (ДИ), с количеством итераций в 1000. Для расчета относительных расхождений при проверке на показатели “превосходит” и “не превосходит” использовались пропорциональные значения, данные отражены в таблице 10. Коэффициент “не превосходит” был установлен на уровне 10%. Для показателей, где более высокий результат указывает на лучшую эффективность, вывод о непревзойденности делался если соотношение было $> 0,9$. Поскольку в таких тестах использовалась только одна сторона 95 % ДИ, односторонняя альфа составляла 2,5%. Ранее для оценки маммографического скрининга с помощью систем САД использовался 10-процентный интервал,

однако 97,5-процентный доверительный интервал является более строгим показателем, чем обычно используемые 90-95% [26].

Таблица 10 – Результаты двойного прочтения с ИИ и без него

Показатель эффективности	Историческое двойное прочтение	Двойное прочтение с ИИ	Результат
Частота отзыва	3.6% (3.5, 3.8)	3.6% (3.4, 3.7)	Превосходит 0.98 (0.96, 0.9981)
ЧВР	9.3 на 1000 (8.6, 10.1)	9.1 на 1000 (8.4, 9.9)	Не превосходит 0.97 (0.96, 0.99)
Чувствительность	85,6% (82.7, 88.1)	83.4% (80.4, 86.1)	Не превосходит 0.97 (0.96, 0.99)
Специфичность	97.4% (97.0, 97.7)	97.5% (97.1, 97.8)	Не превосходит 1.00 (0.9994, 1.00)
ППО	25.7% (23.9, 27.6)	25.6% (23.7, 27.5)	Не превосходит 1.00 (0.98, 1.01)

Только 24,3% (27/111) ИР, обнаруженных системой ИИ, были также обнаружены при моделировании двойного прочтения с помощью системы ИИ, в то время как большинство исторических читателей не определяли эти случаи как «отзыв». Только 20,9 % (37/177) раков, исторически выявленных в следующем раунде скрининга и системой ИИ, были также выявлены при двойном прочтении с помощью системы ИИ.

В таблице 11 представлены случаи, когда ИИ и исторический первый читатель согласны или не согласны с мнением исторического арбитра или не согласны с мнением исторического второго читателя. В среднем исторический второй читатель более чувствителен и менее специфичен по исторически арбитражным делам по сравнению с исторически не арбитражными делами. Когда ИИ и первый читатель приходят к согласию, в среднем исторический второй читатель оказывается значительно менее чувствительным и менее

специфичным по исторически арбитражным делам по сравнению с исторически неарбитражными делами.

Таблица 11 – Сравнительная характеристика исторически первого и второго читателя, арбитра и системы ИИ

	ИИ и первый читатель согласны					ИИ и первый читатель не согласны				
	Случаи	Исторически й арбитр		Второй читатель		Случаи	Исторический арбитр		Второй читатель	
		Кол-во (%)	Чувст. (%)	Спец. (%)	Чувст. (%)		Спец. (%)	Кол-во (%)	Чувст. (%)	Спец. (%)
Исторически й арбитраж проводился	446 (1.8%)	94.3%	56.3%	48.9%	24.4%	4943 (14.9%)	96.2%	66.3%	85.7%	78.9%
Исторически й арбитраж не проводился	2273 (98.2%)			87.8%	99.0%	2248 (85.1%)			66.7%	80.2%

Чувствительность системы ИИ варьировалась от 76,9 до 85,7 %, а специфичность - от 89,2 до 96,1 %. В целом система ИИ отметила 2 037 из 2 310 (88,2%) выявленных скринингом раков, 111 из 373 (29,8%) исторических случаев ИР и 177 из 631 (28,1%) случаев, когда рак был исторически выявлен в следующем раунде скрининга. Для сравнения, исторический первый читатель отметил 2 086 из 2 310 (90,3%) выявленных скринингом раков, 26 из 373 (7,0%) исторических ИР и 41 из 631 (6,5%) случая, когда рак был исторически выявлен в следующем раунде скрининга.

На основании этой информации предлагается провести классификацию прогнозируемых значений в одну из трех групп, отражающих разнообразные результаты классификации и уровень уверенности программного обеспечения, а именно:

- «зелёная зона» определяет значения в диапазоне от 0 до 0,25, указывая на отсутствие патологии с полной уверенностью в 100%;

- «красная зона» описывает прогнозируемые результаты в пределе от 0,93 до 1,0, гарантируя 100% вероятность наличия патологии;
- «желтая зона» описывает предсказанные значения, которые находятся в интервале от 0,25 до 0,93, охватывая как "отсутствие целевой патологии", так и "присутствие целевой патологии", но с несовершенной точностью в классификации.

Рекомендуется внедрить в описание маммографии прогнозируемое значение и связанную с ним зону, с целью предоставления информации врачу-рентгенологу при возникновении арбитража, что позволит оценить степень надежности полученных результатов (рисунок 22). Предложенная стратегия поможет врачам поддерживать максимальную бдительность при анализе маммографий, отнесенных к категории «желтая зона». Это особенно важно, учитывая, что даже при высокой точности текущие системы искусственного интеллекта не обладают способностью абсолютно точно классифицировать каждое маммографическое изображение.

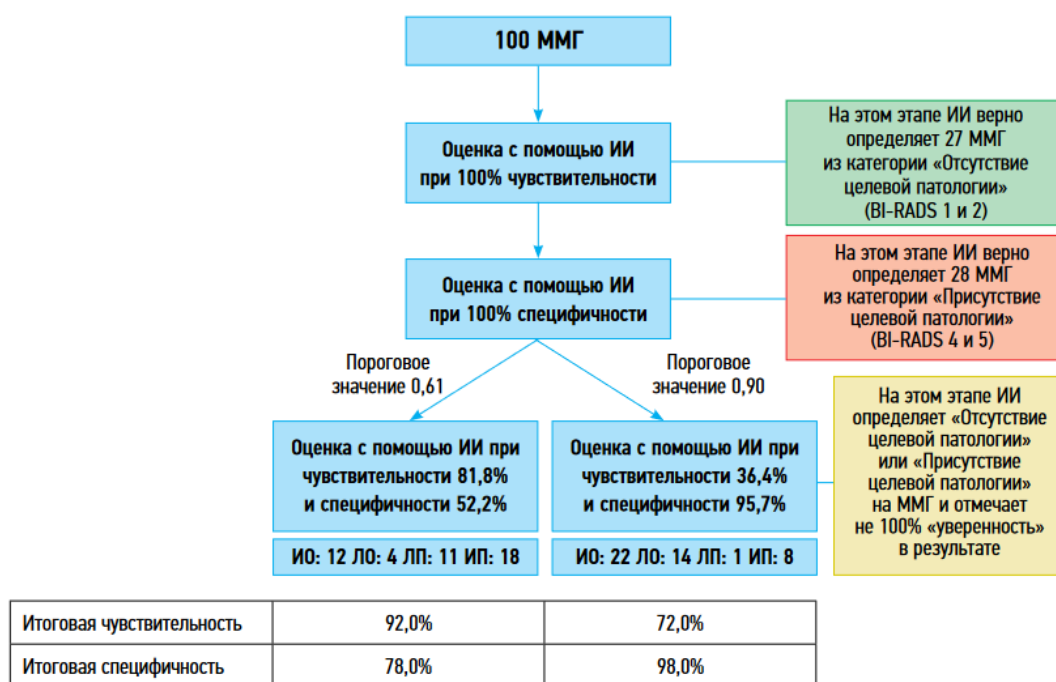


Рисунок 22 – Схематическое представление принципа первичной аналитики маммографических данных

Предполагается, что использование системы ИИ в качестве второго независимого читателя позволит сократить число читателей рентгенологов, но в тоже время значительно увеличит долю случаев, требующих арбитража с 3,2% до 12,4%, что позволило бы сократить количество прочтений, требующих участия человека примерно на 250000. Учитывая, что время, затрачиваемое на арбитраж, может быть в четыре раза больше по сравнению с первым и вторым прочтением, общая доля снижения рабочей нагрузки составила бы от 31 до 43%.

4.3 Экономическая эффективность

В ходе исследования была проанализирована средняя стоимость медицинской услуги "А06.20.004 Маммография" в соответствии с региональными программами обязательного медицинского страхования на 2024 год. В рамках диспансеризации в 18 регионах России на данную услугу установлена отдельная стоимость, кроме того, в 4 регионах из 18 отдельно оценивается и услуга "А06.30.002 Описание и анализ рентгенографических изображений (повторное описание маммографии)".

Диапазон цен на описание маммографии варьируется от 115,95 до 1047,98 рублей за одно прочтение. На данный момент известно, что услуга по анализу маммографических данных с использованием искусственного интеллекта используется исключительно в Москве, а ее стоимость установлена на уровне 235,00 рублей.

Для данного исследования были рассмотрены два метода определения объема средств, требуемых для финансирования программ скрининговой маммографии на территории России. В среднем рентгенологическое обследование молочных желез оценивается в 176,00 рублей, следовательно, двукратное проведение маммографии рентгенологами составит 352,00 рублей, а оплата за двойной анализ каждого снимка 236,00 рублей [3]. В связи с этим, при ожидаемом числе маммографий в год в России в размере 8,1 миллиона,

расходы на повторное описание снимков рентгенологами составят 2,8 миллиарда рублей, а использование искусственного интеллекта в паре с рентгенологами снизит затраты до 1,9 миллиарда рублей при том же объеме исследований. Возможное сокращение расходов при применении программного обеспечения, основанного на технологии искусственного интеллекта, оценивается в 1,0 миллиарда рублей ежегодно.

Второй метод состоял в анализе стоимости медицинских услуг в различных регионах Российской Федерации на 2024 год. Учитывались как экономическая выгода при анализе маммографии искусственным интеллектом и рентгенологом, так и двойное прочтение врачами-рентгенологами.

$$\frac{176,00 \times 2 - 236,00}{176,00 \times 2} \times 100\% = 32,9\% \quad (1)$$

Если как было сказано ранее, цена на проведение медицинских процедур по описанию маммографии врачом-рентгенологом варьируется от 115,95 до 1047,98 рублей, то ценовой диапазон двойного описания маммографии уже оценивается от 227,93 до 2067,87. Учитывая, что стоимость анализа в регионах будет на 32,9% меньше, чем двойное прочтение врачами рентгенологами оценка стоимости описания маммографии искусственным интеллектом и врачом-рентгенологом попадает в диапазон от 152,53 до 1387,93 рублей.

В заключении, при среднегодовом объеме маммографий в 8,1 миллиона в России, расходы на парный анализ врачами-рентгенологами могут колебаться от 1,7 до 16,8 миллиарда рублей, в то время как использование искусственного интеллекта и рентгенолога позволит сократить эти издержки до 1,3-11,5 миллиарда рублей. Эффективное применение программного обеспечения на основе технологий искусственного интеллекта, может привести к потенциальной экономии на обще региональном уровне от 0,5 до 5,4 миллиардов рублей каждый год.

Заключение

В ходе выполнения данной работы была дана краткая характеристика применения анализа данных в управлении учреждениями здравоохранения, их область применения, особенности и основной функционал. После проведения аналитической работы с литературными источниками были определены основные направления применения интеллектуального анализа в сфере управления здравоохранения, что подтверждает правильность выбора методологии.

Был рассмотрен процесс двойного прочтения рентгенологических снимков при выявлении злокачественных новообразований молочной железы, а также основные методы и алгоритмы искусственного интеллекта, используемые при диагностике раковых заболеваний. Дана сравнительная характеристика методов глубокого обучения при индивидуальном прочтении рентгенологических снимков.

Мы применили системный подход, охватывающий весь цикл обработки данных, начиная от сбора и подготовки до выбора архитектуры нейронной сети. Для обучения модели будут использоваться разнообразные мультимодальные данные, объединяющие рентгеновские изображения и текстовые данные медицинских записей. Этот комплексный подход обеспечил достоверность и репрезентативность результатов.

Важным этапом была подробная подготовка данных, включая интеграцию изображений и текстовых данных, а также балансировку классов для предотвращения перекаса в обучении. Этот детальный процесс обеспечил надежную основу для обучения нейронной сети, учитывая разнообразие информации.

Выбранная архитектура включала сверточные и рекуррентные слои для эффективной работы с мультимодальными данными. Применение функций активации ReLU, LSTM и механизма внимания позволило учесть особенности как изображений, так и текстовых данных.

Была проанализирована типовая конфигурация аппаратно-технических средств действующей медицинской организации, а также исследован текущий процесс прочтения рентгенологических снимков, путем построения соответствующих диаграмм и моделей.

На основе предыдущих исследований, а также текущего состояния процесса прочтения маммографических снимков при проведении скрининга, была предложена модель двойного прочтения с применением технологий искусственного интеллекта, проведен анализ эффективности использования искусственного интеллекта на тестовом наборе данных mini-MIAS в различных сценариях прочтения по целевым группам, а также дана оценка перспективы внедрения и использования системы в дальнейшем, как с функциональной точки зрения, так и с экономической.

Список используемой литературы

1. Бедрина Т. О проекте «Создания единого цифрового контура на основе ЕГИСЗ». URL: <https://docplayer.ru/147170461-O-proekte-sozdaniya-edinogo-cifrovogo-kontura-na-osnove-egisz.html>. (дата обращения: 10 ноября 2022 г.)
2. Богдан И. В., Гурылина М. В., Чистякова Д. П. Цифровые информационные системы московского здравоохранения: взгляд населения //Московская медицина. – 2020. – №. 4. – С. 18-20.
3. Васильев Ю. А. и др. Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента //Издательские решения. – 2023.
4. Вергазова Э. Бережливая поликлиника: неожиданные проблемы и как их решить //Заместитель главного врача. – 2018. – №. 3. – С. 14-25.
5. Гусев А. В. Перспективы дальнейшего развития службы медицинской статистики путем перехода к управлению на основе данных //Врач и информационные технологии. – 2018. – №. 2. – С. 6-22.
6. Заболотная Н. В., Гатилова И. Н., Заболотный А. Т. Цифровизация здравоохранения: достижения и перспективы развития //Экономика. Информатика. – 2020. – Т. 47. – №. 2. – С. 380-389.
7. Панасенко С. В., Мкртчян В. С. Перспективы использования нейротехнологий в различных отраслях цифровой экономики //Российское предпринимательство. – 2018. – Т. 19. – №. 11. – С. 3269-3278.
8. Паспорт национальной программы «Цифровая экономика Российской Федерации» (утв. президиумом Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам 24 декабря 2018 г. N 16). URL: <https://base.garant.ru/72190282/> (дата обращения: 10 ноября 2022 г.).
9. Постановление «О единой государственной информационной системе в сфере здравоохранения» (внес. изм. 02.02.2019 г.) от 05.05.2018 г. №

555. URL: <http://docs.cntd.ru/document/557308809> (дата обращения 15 ноября 2022 г.).

10. Патент РФ на изобретение № 2022617324/05.04.2022. Бюл. № 4. Морозов С.П., Андрейченко А.Е., Четвериков С.Ф., и др. Веб-инструмент для выполнения ROC-анализа результатов диагностических тестов. Режим доступа: <https://www.elibrary.ru/item.asp?id=48373757>. (дата обращения: 10.03.2024)

11. Солодкий В. А. и др. Современные системы поддержки принятия врачебных решений на базе искусственного интеллекта для анализа цифровых маммографических изображений //Вестник рентгенологии и радиологии. – 2023. – Т. 104. – №. 2. – С. 151-162.

12. Титов В. А., Цыганов С. Н. Математические методы и инструментальные средства повышения эффективности деятельности медицинских организаций //Открытое образование. – 2016. – Т. 20. – №. 6. – С. 70-76.

13. Указ Президента РФ от 07.05.2018 г. № 204 «О национальных целях и стратегических задачах развития РФ на период до 2024 года». URL: <https://www.garant.ru/hotlaw/federal/1195467/> (дата обращения 10 ноября 2022 г.).

14. Altaf M. M. A hybrid deep learning model for breast cancer diagnosis based on transfer learning and pulse-coupled neural networks //Mathematical Biosciences and Engineering. – 2021. – Т. 18. – №. 5. – С. 5029-5046.

15. Beeravolu A. R. et al. Preprocessing of breast cancer images to create datasets for deep-CNN //IEEE Access. – 2021. – Т. 9. – С. 33438-33463.

16. Cè M. et al. Artificial intelligence in breast cancer imaging: Risk stratification, lesion detection and classification, treatment planning and prognosis—A narrative review //Exploration of Targeted Anti-tumor Therapy. – 2022. – Т. 3. – №. 6. – С. 795.

17. Clark K. et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository //Journal of digital imaging. – 2013. – T. 26. – C. 1045-1057.
18. Couture H. D. et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype //NPJ breast cancer. – 2018. – T. 4. – №. 1. – C. 30.
19. Hayashi M., Yamamoto Y., Iwase H. Clinical imaging for the prediction of neoadjuvant chemotherapy response in breast cancer //Chinese Clinical Oncology. – 2020. – T. 9. – №. 3. – C. 31-31.
20. Hermon R., Williams P. A. H. Big data in healthcare: What is it used for? – 2014.
21. Ionescu G. V. et al. Prediction of reader estimates of mammographic density using convolutional neural networks //Journal of Medical Imaging. – 2019. – T. 6. – №. 3. – C. 031405-031405.
22. Junior O. P. et al. A novel fusion-based texture descriptor to improve the detection of architectural distortion in digital mammography //Journal of Digital Imaging. – 2021. – T. 34. – №. 1. – C. 36-52.
23. Khamparia A. et al. Diagnosis of breast cancer based on modern mammography using hybrid transfer learning //Multidimensional systems and signal processing. – 2021. – T. 32. – C. 747-765.
24. Lee R. S. et al. A curated mammography data set for use in computer-aided detection and diagnosis research //Scientific data. – 2017. – T. 4. – №. 1. – C. 1-9.
25. Lee R. S. et al. Comparison of segmentation-free and segmentation-dependent computer-aided diagnosis of breast masses on a public mammography dataset //Journal of biomedical informatics. – 2021. – T. 113. – C. 103656. 2016.
26. Lehman C. D. et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection //JAMA internal medicine. – 2015. – T. 175. – №. 11. – C. 1828-1837.

27. Li H. et al. Deep learning in breast cancer risk assessment: evaluation of convolutional neural networks on a clinical dataset of full-field digital mammograms //Journal of medical imaging. – 2017. – Т. 4. – №. 4. – С. 041304-041304.
28. Lunit Insight MMG. AI solution for mammography. URL: <https://www.lunit.io/en/products/mmg/> (дата обращения 20.05.2024).
29. Metrix C. Empowering radiologists for better outcomes in women's health. URL: <https://curemetrix.com/> (дата обращения 20.05.2024).
30. Morozov S. P. et al. Clinical acceptance of software based on artificial intelligence technologies (radiology) //arXiv preprint arXiv:1908.00381. – 2019.
31. Posso M. et al. Cost-effectiveness of double reading versus single reading of mammograms in a breast cancer screening programme //PloS one. – 2016. – Т. 11. – №. 7. – С. e0159806.
32. Ragab D. A. et al. Breast cancer detection using deep convolutional neural networks and support vector machines //PeerJ. – 2019. – Т. 7. – С. e6201.
33. Ryazantsev S. V., Vyazemskaya E. E., Budkovaya M. A. The measures of informativeness of diagnostic medical technologies in otorhinolaryngology: calculation and interpretation //Rossiiskaya otorinolaringologiya. – 2020. – Т. 19. – №. 1. – С. 46-55.
34. Shah S. M. et al. Artificial intelligence for breast cancer analysis: Trends & directions //Computers in Biology and Medicine. – 2022. – Т. 142. – С. 105221.
35. Sharma N. et al. Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms //BMC cancer. – 2023. – Т. 23. – №. 1. – С. 460.
36. Singh V. K. et al. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network //Expert Systems with Applications. – 2020. – Т. 139. – С. 112855.
37. Skaane P. Breast cancer screening with digital breast tomosynthesis //Digital Breast Tomosynthesis: A Practical Approach. – 2016. – С. 11-28.

38. Skarping I., Larsson M., Förnvik D. Analysis of mammograms using artificial intelligence to predict response to neoadjuvant chemotherapy in breast cancer patients: proof of concept //European Radiology. – 2022. – С. 1-11.
39. Taylor P., Potts H. W. W. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate //European Journal of Cancer. – 2008. – Т. 44. – №. 6. – С. 798-807.
40. Taylor-Phillips S., Stinton C. Double reading in breast cancer screening: considerations for policy-making //The British journal of radiology. – 2020. – Т. 93. – №. 1106. – С. 20190610.
41. World Health Organization. Estimated number of cases in 2022, worldwide, both sexes, all URL: [https:// gco.iarc.fr/today/data/factsheets/cancers /](https://gco.iarc.fr/today/data/factsheets/cancers/) (дата обращения 10 мая 2023 г.).
42. Yala A. et al. A deep learning model to triage screening mammograms: a simulation study //Radiology. – 2019. – Т. 293. – №. 1. – С. 38-46.