



На правах рукописи

Егошин Иван Александрович

**Исследование методов и разработка алгоритмов обработки
суммационных маммографических изображений**

Специальность 1.2.2. –
«Математическое моделирование,
численные методы и комплексы программ»

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Йошкар-Ола – 2024

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Марийский государственный университет».

Научный руководитель:

Колчев Алексей Анатольевич
кандидат физико-математических наук, доцент

Научный консультант:

Пасынков Дмитрий Валерьевич
кандидат медицинских наук, доцент

Официальные оппоненты:

Сизиков Валерий Сергеевич
доктор технических наук, профессор,
профессор ФГАОУ ВО «Санкт-Петербургский
национальный исследовательский университет
информационных технологий, механики и
оптики»

Шлеймович Михаил Петрович
кандидат технических наук, доцент,
заведующий кафедрой автоматизированных
систем обработки информации и управления
ФГБОУ ВО «Казанский национальный
исследовательский технический университет
им. А.Н. Туполева-КАИ»

Ведущая организация:

Федеральное государственное бюджетное
образовательное учреждение высшего
образования «Московский государственный
университет имени М. В. Ломоносова»

Защита состоится 202 г. в часов минут на заседании диссертационного совета 24.1.224.01 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН) по адресу 117312, г. Москва, проспект 60-летия Октября, 9 (конференц-зал, 1-й этаж).

С диссертацией можно ознакомиться в библиотеке ФИЦ ИУ РАН по адресу: г. Москва, ул. Вавилова, д. 40 и на официальном сайте ФИЦ ИУ РАН: <http://www.frcsc.ru>.

Отзывы на автореферат в двух экземплярах, заверенные печатью учреждения, выслать по адресу: 119333, г. Москва, ул. Вавилова, д. 44, кор. 2, ученому секретарю диссертационного совета 24.1.224.01.

Автореферат разослан « » 202 года
Телефон для справок:

Ученый секретарь
диссертационного совета 24.1.224.01,
кандидат физико-математических наук, доцент

И.В. Смирнов

Общая характеристика работы

Актуальность темы. Проблемой маммографических изображений является то, что различные структуры, располагающиеся на разной глубине, могут проецироваться на одни и те же точки на получающемся двумерном изображении (суммация), что сильно затрудняет их последующий анализ. Все это приводит к утрате текстуры, характерной для патологических очагов, и, соответственно, к сложностям при визуальном выявлении мелких и атипичных проявлений рака молочной железы (РМЖ).

Современные подходы к автоматической обработке маммографических изображений можно разделить на четыре группы используемых методов.

К первой группе относятся методы на основе адаптивного порога, которые в разное время разрабатывали A. Qayyum, N. Singh и др. Их преимущество заключается в том, что порог определяется для градации яркости локальной области изображения, который способен выделить отличающиеся от фона интересующие объекты. Однако данный контекст характерен только для типичных случаев изменений молочной железы (МЖ) на однородном фоне.

Ко второй группе относятся методы на основе анализа областей изображения, предложенные, в том числе, Ю.А. Подгорновой, С.С. Садыковым и др., которые делают сегментацию устойчивой к изменениям интенсивности, так как динамически определяются критерии сходства пикселей на основе локальных характеристик изображения. Однако они зависят от правильного выбора первоначальной точки приближения.

Третья группа методов основана на анализе краев или контурного поиска на изображении, когда производится поиск градиентного перепада яркости на изображении, предложенная S.S. Yasiran, N. Al-Najdawi и др. Это раскрывает очертания визуальных данных и аналогично человеческому восприятию изображения. Главным недостатком таких методов является чувствительность к шуму и невозможность правильно анализировать плавные переходы.

Четвертая группа методов обработки маммографических изображений основана на применении методов машинного обучения, включая сети глубокого обучения, которые заключаются в извлечении классифицирующих признаков из

изображения и последующего отнесения анализируемых областей к заданным классам, разрабатываемые Li H., Jafari Z. и др. Однако они сталкиваются с некоторыми серьезными проблемами, которые тормозят их дальнейшее развитие. Например, результат прогнозирования может быть недостаточно устойчив из-за уровня обобщения модели. Кроме того, сама модель часто недостаточно обоснована, так как отсутствует интерпретируемость обученной модели, а обучающих данных не всегда достаточно, особенно применительно к сложным, атипичным и многовариантным вариантам изменений.

Проведенный анализ состояния проблемы обработки маммографических изображений предопределили выбор темы диссертационной работы, постановку цели и задач исследования.

Целью диссертационного исследования является разработка и улучшение алгоритмов обработки и анализа суммационных маммографических изображений для повышения точности и достоверности неинвазивной диагностики рака молочной железы.

Для достижения поставленной цели решаются следующие **задачи**:

1. Разработать алгоритм обнаружения подозрительных изменений молочной железы на маммограммах.
2. Провести анализ эффективности предложенного алгоритма обнаружения изменений в сравнении с моделью глубокого обучения.
3. Разработать методику сегментации подозрительных изменений на маммографических изображениях.
4. Развить методики классификации подозрительных изменений на маммограммах.
5. Разработать метод обнаружения скоплений кальцинатов на маммографических изображениях.
6. Разработать методику классификации доброкачественных и злокачественных кальцинатов на маммографических изображениях.

Научная новизна работы состоит в следующем.

1. Предложен новый метод обнаружения патологических изменений на маммограммах на основе алгоритма вложенных контуров и рассматриваемой

модели образований, позволяющий повысить чувствительность маммографии в выявлении рака молочной железы.

2. Разработана оригинальная методика сегментации подозрительных изменений на маммографических изображениях, основанная на предложенной модели, где модуль градиента растет в направлении от вершины образования к его границе, и не искажающий границы исследуемого объекта, а также впоследствии не влияющий на извлечение текстурных и/или геометрических признаков изменений молочной железы.

3. Разработан новый метод обнаружения кальцинатов на маммографических изображениях, обеспечивающий повышение чувствительности маммографии в выявлении рака молочной железы.

4. Разработана новая методика классификации доброкачественных и злокачественных скоплений кальцинатов на маммографических изображениях, отдельно учитывающая скопления кальцинатов в сосудах в виде модели линейных объектов и обеспечивающая снижение частоты ложноположительных откликов алгоритма вложенных контуров.

Научная и практическая значимость работы. Использование разработанных методик и алгоритмов улучшает качество анализа маммографических изображений. На основе данных методов может быть реализована система компьютерной диагностики образований на ранней стадии, способствующая поддержке принятия решений.

Реализованный метод на основе алгоритма вложенных контуров потенциально способен обнаруживать типичные образования на более ранней стадии. Это имеет высокую важность, так как стадия опухоли на момент постановки диагноза представляет собой один из важнейших факторов.

В целом, внедрение разработанных алгоритмов в клиническую практику может способствовать повышению выживаемости пациенток за счет улучшения выявляемости различных вариантов РМЖ в более ранней стадии.

Соответствие диссертации паспорту научной специальности. В соответствии с формулой специальности 1.2.2 «Математическое моделирование, численные методы и комплексы программ» (технические науки) в работе

выполнены разработка, исследование и реализация алгоритмов и методов обнаружения, сегментации и классификации патологических изменений МЖ на маммографических изображениях на основе предложенной модели функции яркости образования. Работа соответствует следующим пунктам паспорта специальности: п. 2. «Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий», п. 3. «Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента», п. 8. «Комплексные исследования научных и технических проблем с применением современной технологии математического моделирования и вычислительного эксперимента» и п. 9. «Постановка и проведение численных экспериментов, статистический анализ их результатов, в том числе с применением современных компьютерных технологий (технические науки)».

Методы исследования. Для решения поставленных задач и получения основных научно-практических результатов использованы методы диагностики РМЖ (маммография), методы машинного обучения, методы математического анализа, математического и информационного моделирования, системный подход к анализу и синтезу алгоритмов автоматической обработки медицинских изображений. Теоретические результаты получены с использованием аналитических методов диагностики РМЖ, теории вероятностей и математической статистики, теории статистических решений и численных методов расчета на ЭВМ с использованием программного пакета Matlab версии 2020b. Экспериментальные исследования проведены с использованием натурных данных скрининга МЖ и специально разработанного автором программного обеспечения (ПО).

Основные положения, выносимые на защиту:

1. Предложен алгоритм вложенных контуров, который обеспечивает эффективное обнаружение подозрительных изменений МЖ на маммограммах при широком диапазоне соотношения яркостей образования и фона.

2. Разработана методика сегментации подозрительных образований на маммографических изображениях без искажения границ исследуемого объекта, позволяющая повысить точность классификации на основе методов машинного обучения.

3. Предложен метод обнаружения кальцинатов на маммографических изображениях, обеспечивающий эффективное обнаружение признаков раннего рака молочной железы.

4. Разработана методика классификации доброкачественных и злокачественных кальцинатов на маммографических изображениях, способствующая поддержке принятия клинических решений.

Достоверность обусловлена соответствием результатов, полученных путем аналитического и численного моделирования, результатам экспериментальных исследований, выполненных в рамках данной работы и другими исследователями. Эффективность разработанных алгоритмов подтверждена натурными испытаниями и повторяемостью результатов на больших объемах экспериментальных данных, полученных в период 2016-2024 гг.

Апробация работы. Основные результаты диссертационного исследования были доложены и обсуждены на следующих научно-технических и научно-медицинских конференциях: Европейских конференциях «Компьютер-ассистированная лучевая диагностика и хирургия» (CARS) (Гейдельберг, 2016), (Берлин, 2018), (Ренн, 2019), (Мюнхен, 2020), (Мюнхен, 2021); Международной конференции по микроволновым технологиям, антеннам и электронным системам (COMCAS) (Тель-Авив, 2017); IX и X международных конгрессах «Невский радиологический форум – 2017 и 2018» (Санкт-Петербург, 2017 и 2018); конференции Европейского общества специалистов по ультразвуковой диагностике (EUROSON) (Познань, 2018).

Исследования автора в 2022-2023 гг. и 2024 г. получили финансовую поддержку Российского научного фонда (проекты 22-71-10070 и 24-21-00031, соответственно).

Личный вклад. Все основные результаты диссертационной работы получены и обоснованы автором самостоятельно. Постановка задач и обсуждение

результатов проводились совместно с научным руководителем и научным консультантом. Личный вклад соискателя состоит в разработке методов [1-7], алгоритмов [1; 15-17; 20; 22], написании программ [23-25] и обработке экспериментальных данных, анализе литературных источников. Весь материал, представленный в работах [1-7; 10-12; 14-17; 20; 22], проанализирован и статистически обработан автором самостоятельно. В [1-2; 5; 11; 14; 15-17; 20] автором проведены аналитические построения и математическое моделирование, получены основные результаты и сформулированы выводы. Учитывая, что обработку значительного объёма экспериментальных данных невозможно провести единолично, роль диссертанта, в работах [4; 8-10; 13; 18-19; 21] заключалась в систематизации данных, участии в проведении эксперимента. По этой же причине многие публикации диссертанта имеют соавторство.

Публикации. Основные результаты докторской диссертации опубликованы в 22 научных публикациях, из них 1 работа в журнале, рекомендованном ВАК, 6 публикаций в научных изданиях, индексируемых Web of Science и Scopus по группе научных специальностей докторской диссертации, включая 4 статьи в изданиях Q1 и Q2 квадриля, 8 – в трудах конференций, всего 11 публикаций в изданиях категории К1 и К2 и приравненных к ним, получено 3 свидетельства о государственной регистрации программы для ЭВМ.

Объем и структура работы. Докторская диссертация состоит из введения, трех глав, заключения, выводов, практических рекомендаций и списка литературы. Диссертация изложена на 155 страницах машинописного текста, содержит 23 таблицы и 53 рисунка. Список литературы включает 175 наименования.

Содержание работы

Во введении обоснованы актуальность работы, цель и задачи исследований, объект, предмет и методы исследований, научная новизна, научная и практическая значимость, представлены защищаемые положения и личный вклад соискателя.

В первой главе анализируются особенности маммографических изображений. В *параграфах 1.1-1.2* дается обзор особенностей маммограмм, их видов и характеристик, а также характеристик различных изменений на них, и

приводятся данные о частоте их выявления. Методы обнаружения и сегментации изменений МЖ приведены в *параграфах 1.3-1.5*. На основе проведенного анализа состояния проблемы обработки медицинских изображений уточнены задачи диссертационного исследования.

Вторая глава посвящена разработке новых методик и алгоритмов обработки маммографических изображений, при этом исследуется эффективность существующих алгоритмов, предлагаются их модификации и некоторые качественно новые методы.

В *параграфе 2.1* представлены наборы данных маммографии (маммограммы) на которых тестировались предложенные методики и характеристики изменений, включенных в анализ.

Математическая модель яркости пикселей образования МЖ на маммограмме в виде модели поверхности второго порядка рассматривается в *параграфе 2.2*. Экспериментальные исследования маммограмм с присутствующим РМЖ показывают, что патологическое изменение можно рассматривать как яркое « пятно », значение яркости пикселей, в центральной области которого, обычно выше чем у окружающих его тканей МЖ и, как правило, патологическое изменение имеет округлую форму (см. рис. 1 а-г).

В качестве обобщенной математической модели образования МЖ на маммограмме может быть предложена модель поверхности второго порядка, например, эллиптический параболоид, предлагаемую ранее Han Z. и др., 2012 г. (см. рис. 1 д):

$$\begin{cases} z = C - (ax^2 + by^2) \\ z \geq 0 \end{cases}$$

Хотя экспериментальные поверхности для функций яркости пикселей изображения, отличаются от таких поверхностей, в диссертации считались справедливыми следующие положения:

1. Линии уровня в окрестности вершины образования являются подобными фигурами (см. рис. 1 е).

2. В сечении поверхности плоскостью $Ax + By = 0$ получается кривая, абсолютные значения производных в точках которой возрастают в направлении от вершины к основанию образования (см. рис. 1 ж).

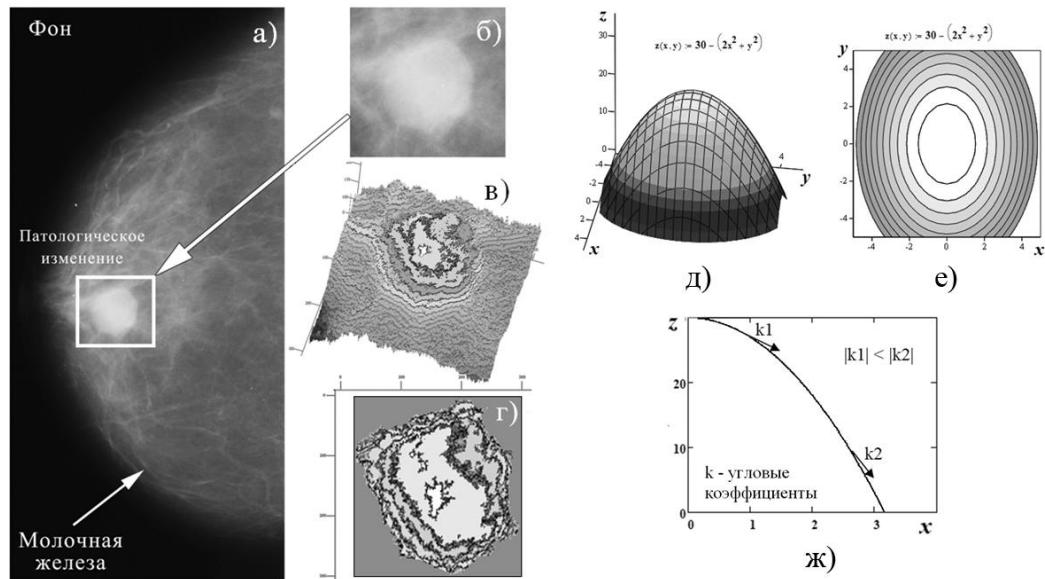


Рис. 1. а) Маммограмма, б) образование МЖ на маммограмме, в) трехмерная поверхность яркости пикселей изображения построенная для образования, г) линии уровня полученные для поверхности яркости пикселей образования, д) трехмерная поверхность второго порядка – эллиптический параболоид; е) линии уровня поверхности; ж) кривая, полученная в сечении поверхности плоскостью.

С учетом указанных выше положений разработан метод обнаружения изменений МЖ на маммограммах с помощью *алгоритма вложенных контуров (ABK)*, который изложен в **параграфе 2.2.1**.

Этапы алгоритма:

Этап 1. Строятся линии уровня для всего яркостного диапазона маммограммы с шагом k . Далее с помощью алгоритма обхода бинаризованных пикселей на каждом шаге градации n сохраняются контуры линий уровня $U_{n,i}$, где n – текущий шаг по градации яркости, $i = 1, \dots, M$, M – количество линий уровня, найденных на n -м шаге градации.

Этап 2. Из набора линий уровня выделяются серии вложенных друг в друга:

$$\begin{cases} \text{вложенный}, & \frac{S(U_{n,i} \cap U_{n+1,i})}{S(U_{n,i})} = 1 \\ \text{не вложенный}, & \frac{S(U_{n,i} \cap U_{n+1,i})}{S(U_{n,i})} < 1 \end{cases}$$

где: $S(U_{n,i})$ – площадь (в пикселях) занимаемая линией уровня U полученными на n -м шаге градации, $S(U_{n,i} \cap U_{n+1,i})$ – площадь (в пикселях) области пересечения линий уровня U полученных на n -м и $n+1$ -м шаге градации яркости.

Этап 3. Рассчитывается набор из характеристик линий уровня:

1) периметр P , равный количеству пикселей линии уровня N ;

2) компактность $C = \frac{P^2}{S}$, где P – периметр, S – площадь линии уровня;

3) координаты центра масс линии уровня: $x_c = \frac{1}{N} \sum_{j=1}^N x_j$, $y_c = \frac{1}{N} \sum_{j=1}^N y_j$;

4) стандартное отклонение расстояния от центра масс линии уровня (x_c, y_c) до

каждой ее точки: $\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^N (r_j - \bar{r})^2}$, где $\bar{r} = \frac{\sum_{j=1}^N r_j}{N}$, $r_j = \sqrt{(x_j - x_c)^2 + (y_j - y_c)^2}$;

5) эксцентриситет линии уровня: $E = \frac{2\sqrt{(0,5A_{\max})^2 - (0,5A_{\min})^2}}{A_{\max}}$;

6) нормированные центральные моменты: $\eta_{p,q} = \frac{\mu_{p,q}}{m_{0,0}^{(p+q)/2+1}}$.

Для оценки степени периодичности линии уровня проведен спектральный анализ величин r_j : $S_k = |F(r_j)|^2$, где F – преобразование Фурье, k – номер гармоники спектра ($k = 1, \dots, N/2$). В качестве параметра, характеризующего сложность линии уровня, используется энтропия нормированных гармоник:

$H = - \sum_{k=1}^{N/2} S_k \ln \bar{S}_k$, где $\bar{S}_k = \frac{S_k}{\sum_{k=1}^{N/2} S_k}$ – доля мощности.

Этап 4. Определяются меры подобия двух вложенных линий уровня A и B , выделенных на разных градациях яркости. Используется четыре меры:

1) сумма абсолютных разностей $K_1(A, B) = \sum_{i=1}^M |w_i^A - w_i^B|$ и ее относительное

значение $K_2(A, B) = \sum_{i=1}^M \left| \frac{w_i^A - w_i^B}{w_i^A} \right|$ (для нормированных центральных моментов);

2) отношение $K_3(A, B) = \frac{w^A}{w^B}$ (для периметра линий уровня);

3) евклидово расстояние между центрами масс линий уровня

$$K_4(A, B) = \sqrt{(x_c^A - x_c^B)^2 + (y_c^A - y_c^B)^2};$$

4) абсолютные разности $K_5(A, B) = |w^A - w^B|$ (по всем остальным характеристикам линии уровня).

Критические точки для мер подобия были определены на основе обучающего набора мер подобия, полученных для маммограмм без патологических изменений и с изменениями МЖ. Чем больше найдено подобных серий вложенных контуров линий уровня, тем больше вероятность того, что эта область является подозрительной.

В *параграфе 2.2.2* предложен метод обнаружения подозрительных изменений на маммограммах с помощью одноэтапного детектора глубокого обучения YOLOv4. Архитектура YOLO основана на построении полностью сверточной нейронной сети (рис. 2), которая содержит 53 сверточных слоя, за каждым из которых следует функция активации Mish: $f(x) = x \cdot \text{th}(\ln(1+e^x))$.

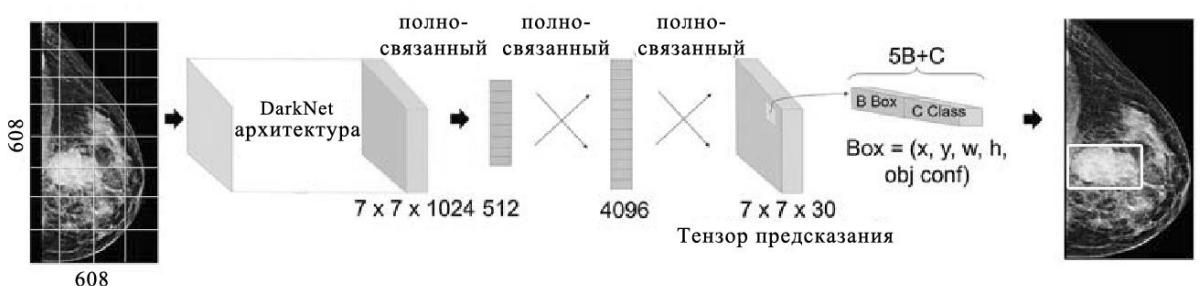


Рис. 2. Архитектура одноэтапного детектора YOLOv4.

Для обучения модели YOLOv4 использовались 535 маммографических изображений, где все патологические области были сегментированы сертифицированным рентгенологом.

Для улучшения работы YOLOv4 предлагается использовать предварительную обработку, которая заключалась в синтезе нового 3-канального изображения из изображения, полученного путем нормализации по усечению и двумя изображениями, полученными после обработки алгоритмом выравнивания аддитивной гистограммы с ограниченной контрастностью (CLAHE).

При обучении более чем в 4000 итераций было получено значение $mAP_{0,5} = 96,23\%$ (mean Average Precision (mAP) при пороге $IoU = 50\%$).

В параграфе 2.2.3 развита методика *сегментации подозрительных изменений на маммографических изображениях*.

Этапы методики сегментации:

Этап 1. С помощью АВК, описанного в параграфе 2.2.1, определяется область интереса (ОИ) на маммограмме (см. рис. 3а) и рассчитываются координаты ее центра масс (x_c, y_c) (точка A на рис. 3б):

$$x_c = \frac{1}{I_0} \sum_{i=1}^N \sum_{j=1}^M x_i I(x_i, y_j), \quad y_c = \frac{1}{I_0} \sum_{i=1}^N \sum_{j=1}^M y_j I(x_i, y_j), \quad I_0 = \sum_{i=0}^N \sum_{j=0}^M I(x_i, y_j),$$

где: $I(x_i, y_j)$ – яркость пикселя ОИ, N и M – ширина и высота, соответственно.

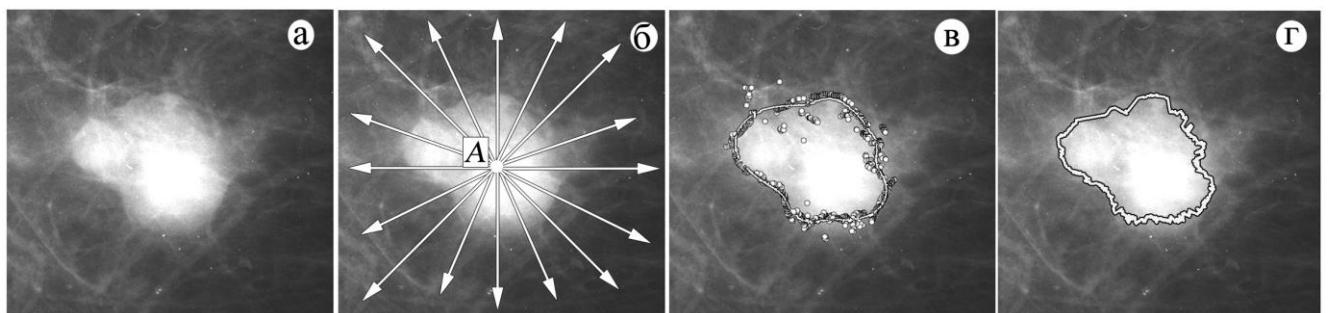


Рис. 3. а) Область интереса маммографического изображение подозрительного изменения МЖ; б) лучи, проведенные из центра изменения; в) примерная граница изменения для сегментирования (белые точки) и линии регрессии (сплошная линия); г) результат сегментирования контура (белая линия).

Этап 2. От точки центра A , по всей окружности изменения проводятся лучи (см. рис. 3б) и выполняется переход к полярной системе координат (r, ϕ) . По пикселям, находящимся на лучах, рассчитывается перепад их яркости одномерным скользящим окном:

$$\Delta I_j = \sum_{t=1}^S I_{j+t} - \sum_{t=1}^S I_{j-t}, \quad j = S, \dots, N-S,$$

где: N – количество пикселей на l -м луче, $I_j = I(r_j, \phi_j)$ – яркость пикселя на l -м луче, S – размер скользящего окна. Наибольший перепад яркости $\max(\Delta I_j)$ соответствует примерной границе (r_l, ϕ_l) необходимой области для сегментирования (см. рис. 3в белые точки).

Этап 3. Фильтрация точек границы – если евклидово расстояние $\Delta r_l = |\bar{r}_l^{avr} - r_l|$ между точками границы области (r_l, ϕ_l) (рис. 3в белые точки) и точками их линии регрессии $(\bar{r}_l^{avr}, \phi_l)$ (рис. 3в белая линия), лежащими на одном l -м луче, больше порога Нibbleка $B = m + k \cdot s$, то такая точка заменялась интерполированной, m и s – среднее и среднеквадратическое отклонение Δr_l , соответственно, коэффициент $k = 0,2$.

На рис. 3 г) белой линией представлен результат сегментирования.

В *параграфе 2.2.4* путем анализа проблем классификации изменений с использованием методов машинного обучения развита методика *классификации очаговых изменений на маммограммах*.

Пусть для множества объектов X задано множество допустимых ответов $Y=\{1, \dots, M\}$ и существует целевая функция $y^*: X \rightarrow Y$, значения которой $y_i = y^*(x_i)$ известны только на конечном подмножестве признаков $\{x_1, \dots, x_l\} \subset X$. Тогда задача классификации заключается в том, чтобы по выборке признаковых описаний F подмножества X^l восстановить зависимость y^* , то есть построить решающую функцию $a: F \rightarrow Y$, которая приближала бы целевую функцию $y^*(x)$, на всём множестве X .

В диссертационной работе в качестве таких объектов x использовались ОИ маммограмм, содержащих изменения МЖ, а для решения задачи классификации изменений исследовалось два похода к извлечению признаков f объектов x .

На рис. 4 представлена блок-схема первой методики классификации изменений молочной железы.

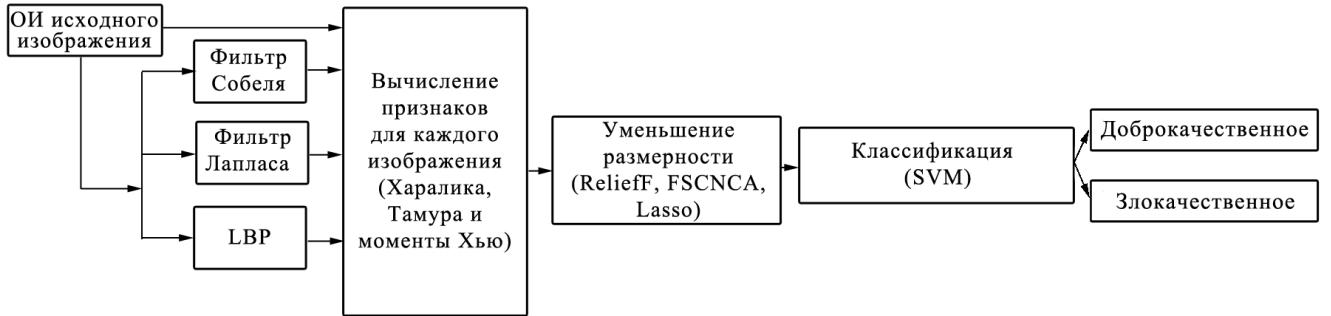


Рис. 4. Блок-схема первой методики классификации изменений.

На первом этапе рассчитывались признаки области изменения МЖ как для исходных значений яркости пикселей изображения, так и для их производных: фильтр Собеля, фильтр Лапласа, дескрипторы локальных бинарных паттернов.

В качестве классифицирующих признаков f использовались признаки Тамуры (Tamura), Харалика (Haralick), моменты Хью (Hu). Для уменьшения размерности полученного набора признаков предлагается использовать методы ReliefF, FSCNCA и LASSO. Данные методы рассчитывают веса для каждого признака, где вес незначимых признаков близок к нулю. Для построения классифицирующей решающей функции a использовался метод опорных векторов (SVM), основанный на построении оптимальной разделяющей гиперплоскости:

$$a(f) = \text{sign}\left(\sum_{j=1}^n w_j f_j - w_0\right) = \text{sign}(\langle w, f \rangle - w_0),$$

где уравнение $\langle w, f \rangle = w_0$, описывает гиперплоскость, разделяющую классы.

Отличительной особенностью данной методики классификации изменений МЖ является то, что классифицирующие признаки извлекались не по одному изображению, а использовались также их производные, которые были получены дополнительной фильтрацией исходного изображения.

В качестве второй методики для извлечения признаков f использовалась глубокая сеть Inception-ResNet, которая была модифицирована путем удаления последних слоев: слоя SoftMax и заключительного классифицирующего слоя. На выходе модифицированной сети получили вектор из 1536 признаков. Этот вектор признаков, так же как и в первом случае, подавался на три блока отбора значимых признаков ReliefF, FSCNCA и LASSO. Различающиеся признаки, отобранные этими тремя методами, объединялись в один вектор признаков, который учитывает только 36 значимых признака. Для построения классифицирующей решающей функции a использовался метод опорных векторов (SVM).

Отличительной особенностью второй методики является использование сети глубокого обучения Inception-ResNet только для извлечения признаков из изображения, а классификация производится методом опорных векторов. Кроме того для отбора значимых признаков используется сразу несколько методов.

На рис. 5 представлена блок-схема второй методики классификации изменений молочной железы.

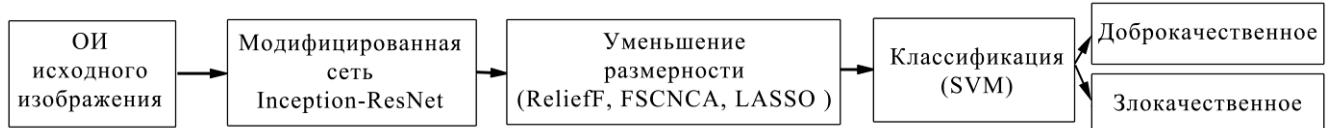


Рис. 5. Блок-схема второй методики классификации изменений.

В параграфе 2.2.5 на основе проведения вычислительных экспериментов и статистического анализа их результатов предложен метод обнаружения кальцинатов на маммограммах.

Этапы метода:

Этап 1. Применение нелинейного фильтра для значений яркости пикселей изображения, который уменьшает яркость пикселей с малой яркостью и увеличивает яркость пикселей с изначально большим значением яркости:

$$f^*(x, y) = \frac{255}{1 + e^{-0,103 \cdot (f(x, y) - 195)}},$$

где: $f(x,y)$ – яркость пикселя представленного своими координатами (x,y) .

Этап 2. Применение метода бинаризации Отцу, суть которого заключается в том, чтобы выставить порог между классами таким образом, чтобы каждый из них был как можно более «плотным».

Этап 3. Фильтрация полученных объектов по перепаду яркости. Производится расчет отношения средней яркости пикселей, расположенных внутри объекта \bar{f}_{in} и снаружи \bar{f}_{out} от него, и проверяется условие:

$$\begin{cases} \text{кальцинат} ; & \bar{f}_{in} / \bar{f}_{out} \geq 1,2 \\ \text{ложный} ; & \bar{f}_{in} / \bar{f}_{out} < 1,2 \end{cases}.$$

Этап 4. Применение алгоритма пространственной кластеризации с присутствием шума (DBSCAN), для группировки тесно расположенных точек, и алгоритма построения выпуклой оболочки Склански, для обвода кластера.

В параграфе 2.2.6 обсуждается метод классификации доброкачественных и подозрительных кальцинатов на маммограммах. Для классификации доброкачественных и подозрительных кальцинатов предлагается использовать их количественные и геометрические признаки: 1) количество объектов в обнаруженному кластере N ; 2) доля объектов малой площади (площадь <50 пикселей): $D_{<50} = N_{<50} / N$; 3) доля объектов большой площади в кластере (площадь >1500 пикселей): $D_{>1500} = N_{>1500} / N$; 4) эксцентризитет кластера (степень его отклонения от окружности): $E = \sqrt{1 - b^2/a^2}$, где a и b – соответственно, большая и малая полуоси кластера; 5) отношение площади, занимаемой объектами кластера, к площади всего кластера кальцинатов: $U = S^* / S$, где S^* – количество пикселей относящихся к объектам кластера, S – количество пикселей всего кластера.

Для сосудистых доброкачественных кальцинатов использовались дополнительные признаки, полученные с помощью преобразования Хафа и индекса корреляции с аппроксимирующей полиномиальной кривой третьей степени:

$$R_{cor} = \sqrt{1 - \frac{\sum (y_i - y^*)^2}{\sum (y_i - \bar{y})^2}}, \quad \bar{y} = \frac{\sum_{i=1}^N y_i}{N},$$

где $y^* = ax^3 + bx^2 + cx + d$ – кубический полином, рассчитанный по координатам (x,y) бинаризованных пикселей кальцинатов, N – количество пикселей.

В качестве классификатора предлагается использовать линейный метод опорных векторов (SVM).

Третья глава посвящена изложению результатов тестирования предложенных методик, численных экспериментов и их обсуждению. Для оценки эффективности работы разработанных алгоритмов и методов была выполнена их программная реализация на языках C++ и Python 3 в виде отдельных модулей.

В качестве тестируемых данных использовались реальные маммограммы. Представлены таблицы с результатами работы каждой из предложенных методик с расчетом оценки эффективности для каждого из типов изменений.

В параграфе 3.1 представлены результаты эффективности работы алгоритма вложенных контуров. Оценка проводилась отдельно как по пленочным, так и по цифровым маммограммам. Общая чувствительность АВК в исследуемой выборке составила 90,73% (323 случая из 356).

Приведено сравнение предложенного АВК с другими современными методиками на маммограммах из баз данных InBreast и CBIS-DDSM (см. Табл. 1).

Табл. 1. Сравнение АВК с современными методиками

База данных CBIS-DDSM		База данных INbreast	
Методы и источники	Точность, (%)	Методы и источники	Точность, (%)
Shams S. и др., 2018, сеть Deep GeneRAtive Multi-task.	89 %	Dhungel N. и др., 2017, каскад методов глубокого обучения.	90 %
Tsochatzidis L. и др., 2019, контролируемый подход к поиску объектов на изображении на основе его контента.	81 %	Carneiro G. и др., 2017, сверточная нейронная сеть ConvNet.	90 %
Falconi L. и др., 2020, сеть глубокого обучения VGG16.	84,4 %	Shi P. и др., 2019, сверточная нейронная сеть.	83,6 %
Zhang H. и др., 2020, классификация с кросс-модальным семантическим анализом.	87,05 %	Zhang H. и др., 2020, классификация с кросс-модальным семантическим анализом.	87,93 %
Himanish S.D. и др., 2022,	89,2 %	El Houby E.M. и др., 2021,	94,52 %

классификатор Xception с тонкой настройкой.		сверточная нейронная сеть.	
Li H., 2022, сеть DualCoreNet.	93 %	Himanish S.D. и др., 2022, классификатор Xception.	95,1 %
Предложенный алгоритм, 2022	97 %	Предложенный алгоритм, 2022	95,6 %
Jafari Z. и др., 2023, CNN с выбором признаков.	96%	Jafari Z. и др., 2023, CNN с выбором признаков.	94,5 %

В параграфе 3.2 представлена оценка эффективности одноэтапного детектора YOLOv4. Оценка предлагаемого одноэтапного детектора проведена в сравнении с АВК. Показано что, в анализируемом наборе общая чувствительность YOLOv4 составила 60%, для АВК аналогичное значение составило 93%.

В параграфе 3.3 представлены результаты сегментации подозрительных изменений на маммографических изображениях и приведено сравнение предложенной методики сегментации с другими современными методиками (см. Табл. 2).

Табл. 2. Сравнение предложенной методики с другими методами.

Методы и источники	Точность, (%)
Lbachir и др., 2021, HRAK алгоритм.	93,15 %
Pezeshki H., 2022, пороговая сегментация.	91,32 %
Li H. и др., 2022, DUALCORENET метод.	93,69 %
Ma Y. и др., 2023, вариационный автоэнкодер на основе перекрестного просмотра.	93,70 %
Предложенная методика, 2023	94 %

В параграфе 3.4 проанализированы результаты работы метода классификации изменений МЖ на маммограммах для двух методик извлечения признаков из изображения. Для первой методики извлечения признаков были правильно классифицированы только 66,6% случаев. Для второй методики, с использованием модифицированной сети глубокого обучения Inception-ResNet, были правильно классифицированы 80,9% случаев.

В параграфах 3.5-3.6 приведены результаты оценки обнаружения и классификации доброкачественных и подозрительных кальцинатов на маммограммах. Общая точность обнаружения кальцинатов всех типов составила

— 405 из 435 (93,1%) и общая оценка классификации доброкачественных и злокачественных изменений составила — 375 из 390 (96,15%).

В заключении формулируются основные результаты работы:

1. Разработан и программно реализован метод обнаружения подозрительных изменений МЖ на маммограммах на основе АВК и рассматриваемой модели образований, который демонстрирует высокую чувствительность (90,73%) не только в случаях типичных и хорошо видимых изменений, но и в случаях атипичных и невидимых изменений, которые не рассматривались в предыдущих работах.

2. Представленные результаты оценки эффективности АВК по сравнению с одноэтапным детектором YOLOv4 показывают, что использование метода АВК для выявления очагов на маммограммах является более значимым. Также рентгенологи, которым были продемонстрированы все маркировки предложенного АВК, изменили свое решение о случае на основании результатов работы алгоритма в 6 из 100 случаев. Результаты YOLOv4 не повлияли на решения рентгенологов.

3. Разработана и программно реализована методика сегментации изменений на маммограммах на основе предложенной модели, где модуль градиента растет в направлении от вершины изменения к его границе. Предлагаемая методика обеспечивает достаточную точность (средняя метрика «Пересечение над объединением» (IoU) по всем типам изменений составила 76%, Precision — 94%) сегментации изменений МЖ, что необходимо для последующей классификации предполагаемых как типичных, так и нетипичных очагов.

4. Полученные результаты классификации изменений МЖ на маммограммах с использованием методов машинного обучения и алгоритмов отбора значимых признаков изображения показали недостаточную точность для клинической практики. Вследствие этого использовать маммограммы для дифференцирования выявленных изменений не представляется перспективным.

5. Разработан метод обнаружения кальцинатов на маммограммах, который обеспечивает высокую точность обнаружения, составляющую 93,1%. Метод программно реализован.

6. Разработана и программно реализована методика классификации доброкачественных и подозрительных кальцинатов на маммографических изображениях с учетом скоплений кальцинатов в сосудах в виде модели линейных объектов, точность классификации которой составляет 96,15%.

Публикации автора по теме диссертации

В изданиях, рекомендованных ВАК по группе научных специальностей диссертации:

1. *Egoшин И.А.* Оценка обобщающей способности алгоритма вложенных контуров при анализе маммограмм // Труды ИСА РАН. – 2024. – Том 74. – № 3. – С. 67-77. (**K-2**)

В изданиях, индексируемых Scopus / WoS по группе научных специальностей диссертации:

2. *Egoshin, I.A., Pasynkov, D.V., Kolchev, A.A., Kliouchkin, I.V., Pasynkova, O.O.* Segmentation of Breast Focal Lesions on the Ultrasound Image // Biomedical Engineering. 2020. — vol. 54. — no. 2. — pp. 99-103. (**K-1, Scopus**)
3. *Kolchev, A., Pasynkov, D., Egoshin, I., Kliouchkin, I., Pasynkova, O., Tumakov, D.* YOLOv4-Based CNN Model versus Nested Contours Algorithm in the Suspicious Lesion Detection on the Mammography Image: A Direct Comparison in the Real Clinical Settings // Journal of Imaging. — 2022. — 8(4): 88. (**K-1, Scopus Q2, WoS**)
4. *Kolchev, A.A., Pasynkov, D.V., Egoshin, I.A., Kliouchkin, I.V., Pasynkova, O.O.* Cystic (including atypical) and solid breast lesion classification using the different features of quantitative ultrasound parametric images // International Journal of Computer Assisted Radiology and Surgery. — 2022. — vol. 17. — Iss. 2. — pp. 219-228. (**K-1, Scopus Q1, WoS**)

5. *Pasynkov, D.V., Kolchev, A.A., Egoshin, I.A., Klioushkin, I.V., Pasynkova, O.O.* A Method for Automatic Calculation of the Diameter of a Pulsating Blood Vessel on Ultrasound Images in Video Stream Mode // Biomedical Engineering. — 2023. — vol. 57. — no. 1. — pp. 23-27. (**K-1, Scopus**)
6. *Пасынков Д.В., Колчев А.А., Егошин И.А., Клюшкин И.В., Пасынкова О.О.* Метод сегментации солидного очагового образования молочной железы и его периферического компонента на ультразвуковых изображениях // Компьютерная оптика. — 2023. — Т. 47. — № 3. — С. 407-414. (**K-1, Scopus Q1, WoS**)
7. *Колчев А.А., Пасынков Д.В., Егошин И.А., Клюшкин И.В., Пасынкова О.О.* Classification of benign and malignant solid breast lesions on the ultrasound images based on the textural features: the importance of the perifocal lesion area // Компьютерная оптика. — 2024. — 48(1): С. 157-165. (**K-1, Scopus Q1, WoS**)

В изданиях, индексируемых Scopus по смежным научным специальностям:

8. *Пасынков Д.В., Егошин И.А., Колчев А.А., Клюшкин И.В., Пасынкова О.О.* Эффективность системы компьютерного анализа маммограмм в диагностике вариантов рака молочной железы, трудно выявляемых при скрининговой маммографии // Российский электронный журнал лучевой диагностики. — 2019. — 9(2): С. 107-118. (**K-1, Scopus Q3, RSCI**)
9. *Пасынков Д.В, Тухбатуллин М.Г., Егошин И.А., Колчев А.А., Клюшкин И.В.* Ультразвуковой скрининг рака молочной железы при маммографически плотной паренхиме: роль системы компьютерного анализа маммограмм (одноцентровое проспективное рандомизированное клиническое исследование) // Российский электронный журнал лучевой диагностики. — 2021 — 11(1): С. 103-113. (**K-1, Scopus Q3, RSCI**)

В изданиях, рекомендованных ВАК по смежным научным специальностям:

10. *Пасынков Д.В., Егошин И.А., Колчев А.А., Клюшкин И.В., Бусыгина О.В.* Сравнительный анализ диагностической ценности систем компьютерного анализа маммограмм I и II поколений // Медицинская визуализация. — 2017. — №1. — С. 90-102. (**K-1**)

11. *Kolchev, A., Egoshin, I.* Use of Hazard Function for Signal Detection on Ionograms // IEEE Geoscience and Remote Sensing Letters. — 2018. — Vol. 15, Iss. 6. — pp. 803-807. (**K-1, Scopus Q1**)
12. *Колчев А.А., Егошин И.А., Недопекин А.Е., Шумаев В.В.* Применение модели смеси вероятностных распределений при обнаружении радиофизических сигналов // Труды конференции PPB–26. — 2019. — Т. 2. — С. 177-180. (**Scopus**)
13. *Пасынков Д.В., Егошин И.А., Колчев А.А., Клюшкин И.В., Пасынкова О.О.* Цветовое картирование серой шкалы с математическим анализом ультразвукового изображения в дифференциальной диагностике кистозных и солидных образований молочной железы // Вестник рентгенологии и радиологии. — 2020. — Т. 101. — №3. — С. 136–146. (**K-3**)
14. *Kolchev, A., Egoshin, I.* Some Problems of Mathematical Modeling of Radiophysical Sounding Signals // Springer Proceedings in Mathematics and Statistics. — 2022. — vol. 379. — pp. 107–117. (**Scopus**)

В трудах конференций:

15. *Egoshin, I., Pasynkov, D., Kolchev, A., Kliouchkin, I.* Value of nested contours analysis algorithm in mammographic image processing // Int. J. of Computer Assisted Radiology and Surgery. — 2016. — Vol. 11. — Suppl. 1. — P. S278.
16. *Egoshin, I., Pasynkov, D., Kolchev, A., Kliouchkin, I., Pasynkova, O.* A segmentation approach for mammographic images and its clinical value // IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems, COMCAS 2017. — 2018. — P. 1-6. (**WoS**)
17. *Egoshin, I., Pasynkov, D., Kolchev, A., Kliouchkin, I., Pasynkova, O.* Automatic differentiation cystic and solid breast lesions at ultrasonic images // Int. J. of Computer Assisted Radiology and Surgery. — 2018. — Vol. 131. Suppl. 1. — P. S255.
18. *Пасынков Д.В., Егошин И.А., Колчев А.А., Клюшкин И.В.* Дифференциальная диагностика атипичных кистозных и солидных образований молочной железы: ценность математической обработки ультразвукового изображения // Лучевая диагностика и терапия. — 2018. — №1 (S). — С. 105.

19. *Pasynkov, D., Egoshin, I., Kolchev, A., Klioushkin, I., Pasynkova, O.* Value of Machine Learning Technique in the Differentiation of Cystic (Including Atypical) and Solid Breast Lesions at Ultrasonic Images // EUROSON, 6-9.09.2018, Poznan, Poland. Abstract book. — P. 168.
20. *Egoshin, I., Pasynkov, D., Kolchev, A., Kliouchkin, I., Pasynkova, O.* Automated differential diagnosis of benign and malignant breast lesions on the mammograms // Int. J. of Computer Assisted Radiology and Surgery. — 2019. — Vol. 14. — Suppl. 1. — pp. S179-S180.
21. *Пасынков Д.В., Егошин И.А., Колчев А.А., Клюшкин И.В., Пасынкова О.О.* Раки, трудно выявляемые при маммографии: причины и ценность систем компьютерного анализа (CAD) // Лучевая диагностика и терапия. — 2019. — №1 (S). — С. 87.
22. *Pasynkov, D., Egoshin, I., Kolchev, A., Klioushkin, I., Pasynkova, O., Shafigulina, V.* The study of machine learning methods for the lesion differentiation on mammograms // Int. J. of Computer Assisted Radiology and Surgery. — 2020. — Vol. 15. — Suppl 1. — pp. S209–S210.

Свидетельства о государственной регистрации программ для ЭВМ:

23. *Пасынков Д.В., Егошин И.А., Колчев А.А., Пасынкова О.О.* Электронная экспертная система дифференциальной диагностики доброкачественных и злокачественных солидных образований молочной железы на ультразвуковых изображениях (версия 1.0): Свидетельство о государственной регистрации программы для ЭВМ № 2021662473 от 29.07.2021.
24. *Пасынков Д.В., Егошин И.А., Колчев А.А., Пасынкова О.О.* Электронная экспертная система дифференциальной диагностики кистозных и солидных образований молочной железы на ультразвуковых изображениях (версия 1.0): Свидетельство о государственной регистрации программы для ЭВМ № 2021662649 от 02.08.2021.
25. *Егошин И.А., Колчев А.А., Пасынков Д.В.* Программа для обработки данных скрининга молочной железы: Свидетельство о государственной регистрации программы для ЭВМ № 2023611125 от 17.01.2023.