Министерство науки и высшего образования Российской Федерации

**Муромский институт (филиал)**

федерального государственного бюджетного образовательного учреждения высшего образования

**«Владимирский государственный университет**

**имени Александра Григорьевича и Николая Григорьевича Столетовых»**

**(МИ ВлГУ)**

Факультет ИТР

Кафедра информационных систем

КУРСОВОЙ ПРОЕКТ

по. Геоинформационные технологии

тема: Анализ расположения абстрактных пространственных фигур

Руководитель

к. т. н., доцент, каф. ИС

(уч. степень, звание)

Еремеев С.В.

(оценка) (фамилия, инициалы)

(подпись) (дата)

Члены комиссии Студент ИСм-121

(группа)

Толмачев А.С.

(подпись) (Ф.И.О.) (фамилия, инициалы)

(подпись) (Ф.И.О.) (подпись) (дата)

Муром 2022

Лист задания.

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 4](#_Toc68975011)

[1 Анализ технического задания 6](#_Toc68975012)

[1.1 Обзор предметной области 6](#_Toc68975013)

[1.2 Исходные данные к проекту 7](#_Toc68975014)

[1.3 Требования к разрабатываемому проекту 9](#_Toc68975015)

[2 Разработка алгоритма 10](#_Toc68975016)

[3 Исследование работы алгоритма 14](#_Toc68975017)

[ЗАКЛЮЧЕНИЕ 30](#_Toc68975018)

[СПИСОК ЛИТЕРАТУРЫ 31](#_Toc68975019)

# ВВЕДЕНИЕ

Цель данной работы - разработка алгоритма сегментации изображений на основе персистентной гомологии для решения задач поиска фиброаденом. Алгоритм направлен на нахождение фиброаденом на маммограммах.

Целью работы является исследования алгоритмов сегментирования изображений на основе персистентной гомологии для решения задач поиска фиброаденом на маммографических снимках, а также использование его для разработки программы по поиску фиброаденом на маммограммах.

Для достижения этой цели были поставлены следующие задачи:

a) проведение анализа предметной области, предъявление требований к разрабатываемому алгоритму;

b) предобработка набора исходных изображений, подобранного для обнаружения дефектов;

d) поиск дефектов и анализ полученных результатов.

Таким образом, структура данного курсового проекта состоит из трех глав и выглядит следующим образом:

a) в первой главе производится анализ технического задания, который включает в себя:

- обзор предметной области;

- определение исходных данных к проекту;

- установление требований к проекту;

b) во второй главе описывается процесс разработки алгоритма сегментации на основе персистентной гомологии;

c) в третьей главе описывается процесс тестирования работы разработанного в предыдущей главе алгоритма и производится анализ полученных результатов.

1. **Анализ технического задания**

**1.1 Обзор предметной области**

Персистентная гомология подразумевает наличие дыр – групп пикселей, объединённых по цветовым характеристикам. Для объединения пикселей в группы необходимо представить изображение в виде списка объектов  где n – результат произведения размеров изображения.

Каждый элемент списка  где , содержит информацию о положении пикселя, его яркости и то, к какой дыре он принадлежит. При этом c = (r, g, b) – значение яркостей цветовых каналов пикселя, x – координата по оси X, y – координата по оси Y, h – дыра, которой принадлежит пиксель. Далее, необходимо установить список связей  где m – общее количество связей. Каждый элемент списка  где  - информация о том, к каким точкам относится связь, i – величина расстояния между ними. Расстояние между точками измеряется по формуле (1) и напрямую зависит от величины яркости цветовых каналов выбранных точек. Согласно формуле (1) для определения расстояния между точками необходимо получить разность значения яркости каждого цветового канала пикселя, а затем посчитать среднеарифметическое значение разности всех каналов. После чего необходимо прибавить к значению разности каждого канала его отклонения от среднеарифметического значения разности:

Функция avg обозначает вычисление среднеарифметического значения. Такой подход позволит выделить зоны, цветовые каналы которых меняются с разной интенсивностью. При построении связей следует учитывать, что соединяются лишь соседние точки в плоскости OX и OY, соседние точки по диагонали не связываются между собой. Таким образом, можно утверждать, что каждая точка, в зависимости от ее расположения, будет иметь до четырех связей.

Установив все связи между точками, необходимо отсортировать массив связей в порядке возрастания значения их длины l. Проделав подобные операции, мы получим некоторую схему связей.

Дальнейший процесс подразумевает проход по списку связей и объединение пикселей в некоторые группы, называемые дырами h. Таким образом, обработав первую связь в списке, мы объединим и  в дыру h, которой будет присвоено имя в соответствии с порядковым номером ее возникновения.

Новое цветовое значение вычисляется по формуле с = ()/2.

По мере обработки списка связей будут образовываться новые дыры. В случаях обработки связи между точкой и дырой точка поглощается дырой и приобретает цвет дыры, цвет дыры при этом не изменяется.

# 1.2 Исходные данные к проекту

В качестве исходных данных к проекту были определены следующие:

1. 12 маммографических изображений, которые будут являться базой фиброаденом.

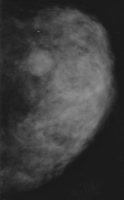
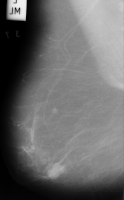
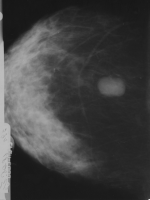


Рис.1 – маммограммы с фиброаденомами

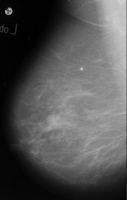
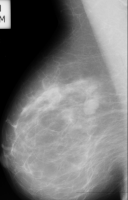
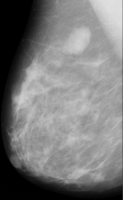


Рис.2 – маммограммы с фиброаденомами

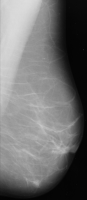
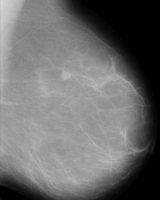
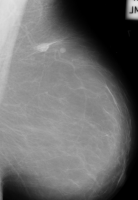


Рис.3 – маммограммы с фиброаденомами

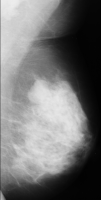
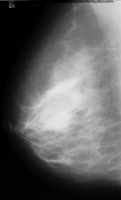
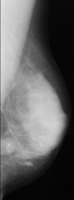


Рис.4 – маммограммы с фиброаденомами

1. 10 маммографических изображений, которые будут являться тестовыми изображениями

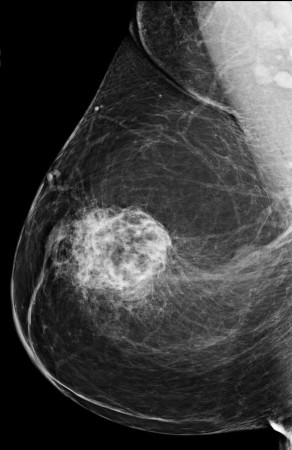
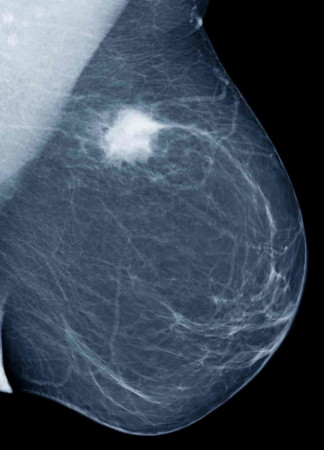
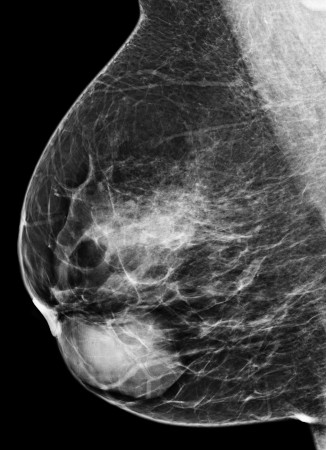


Рис.5 – маммограммы с фиброаденомами

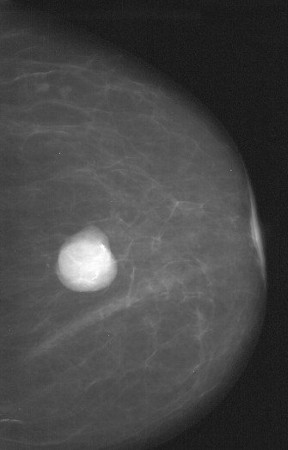
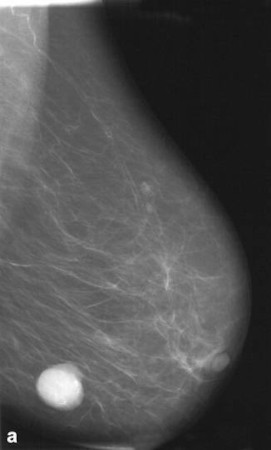


Рис.6 – маммограммы с фиброаденомами

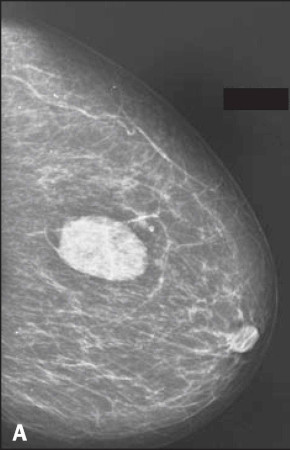
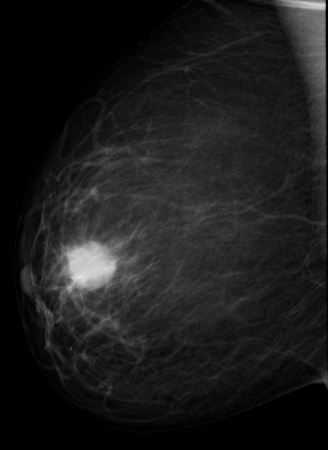


Рис.7 – маммограммы с фиброаденомами

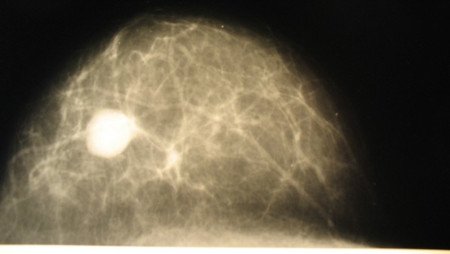


Рис.8 – маммограммы с фиброаденомами

1. Среда разработки Microsoft Visual Studio.
   1. **Требования к разрабатываемому проекту**

Исходя из задач, поставленных перед началом работы, к разрабатываемому проекту необходимо выставить следующие требования:

1. Поиск дефектов на входном изображении.
2. **Разработка алгоритма**

Для реализации предлагаемого алгоритма было разработано приложение на языке C#. Применение разработанного алгоритма позволит сегментировать изображения с дефектами, сохраняя при этом всю историю преобразований. Таким образом, можно утверждать, что результатом работы является сегментация, содержащая в себе всевозможные вариации при изменении порогового значения яркости. В процессе разработки программного обеспечения предусмотрена возможность визуализации с выбором текущей итерации, которая эквивалентна значению порога яркости (Рисунок 4).

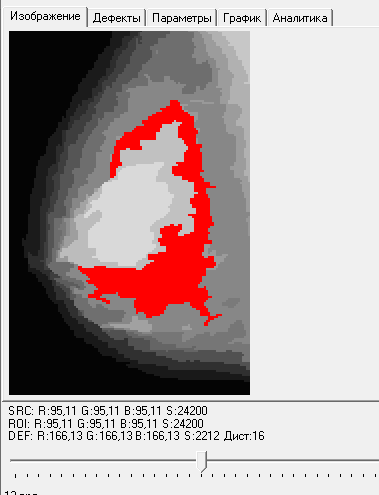


Рисунок 9 – Окно разбиения изображения на сегменты и выделение областей с нужными дефектами

Технически сегментирование реализовано при помощи установки пикселю принадлежности к той или иной дыре для различных интервалов яркости. Проведенные эксперименты показали, что соотношения размеров дыр и их цветовых значений могут выступать в качестве признака обнаружения и классификации дефекта. Выбранные параметры будут значительно отличаться от типа дефекта и анализируемого материала.

На Рисунке 5 приведен пример маммограммы, где можно сразу обнаружить фиброаденому:

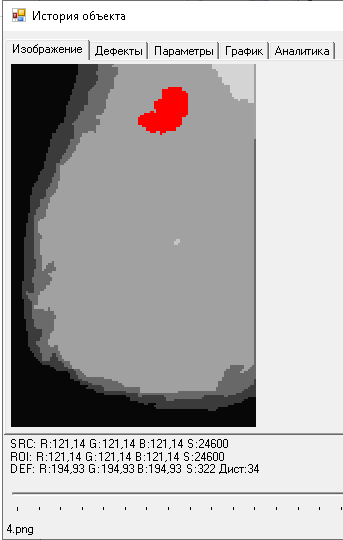


Рисунок 10 – Пример разбиения изображения с фиброаденомой на маммограмме

Различие заметно из-за большой разницы яркости, а также по количеству отображаемых объектов, что говорит об их объемности.

Для обнаружения дефектов был выбран порог равный 12, который подходит для большинства изображений с выявлеными дефектами. Данный порог отсекает ненужные сегменты изображения, оставляя только реализуемые для алгоритма детали.

Для нахождения фиброаденом и определения других обектов была разработана база данных, которая способна хранить информацию о занесенных туда объектах (Рисунок 6).

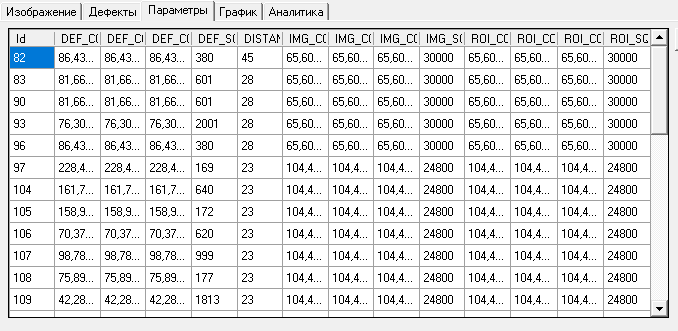


Рисунок 11 – База данных записаных сегментов изображения

Также в программе были определены классы для классификации этих сегментов (Рисунок 12):

1)Фиброаденома;

2)Молочная железа;

3)Фон;

4)Блик и прочее;

5)Мышца;

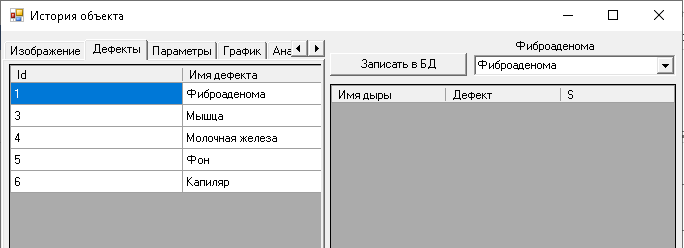


Рисунок 12 – Создание классов для классификации сегментов

Для отображения результатов разбиваемых сегментов в интерфейс была добавлена таблица с колонками: номером объекта, ее класс и площадь, равная количеству пикселей на изображении (Рисунок 13).

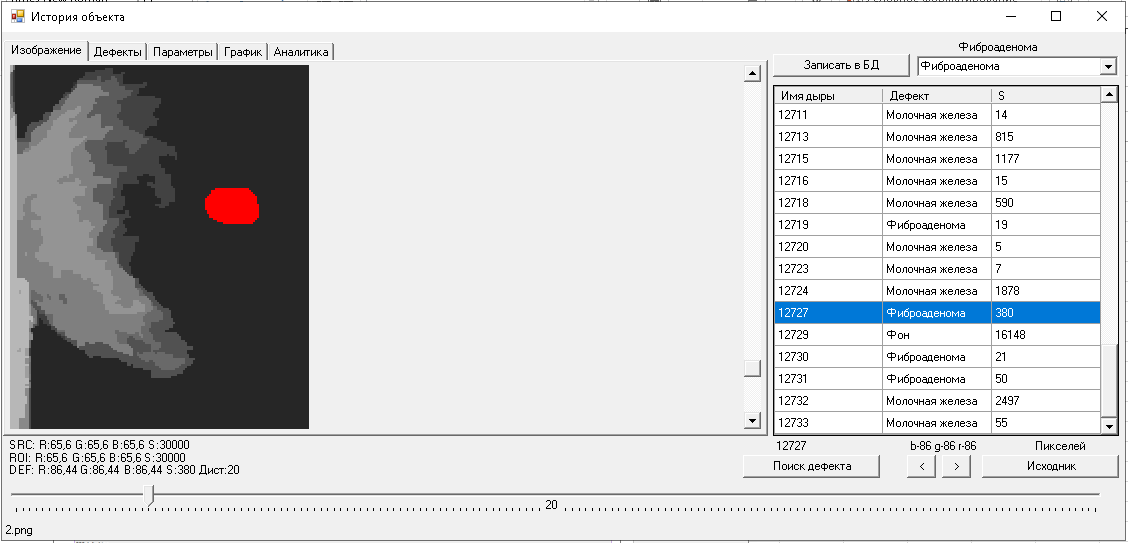


Рисунок 13 – Отображение сегментов снимка молочной железы в таблице

1. **Исследование работы алгоритма**

Тестрирование работы реализованных систем будет проводится следующим образом: на вход программы поступит 12 изображений, далее разделим изображение на сегменты используя порог яркости. Из полученных сегментов выборочно записываем нужные нам дефекты в базу данных для дальнейшего анализа алгоритма.

Разновидности объектов классов, являющимися сегментами изображения, отображены на рисунке 10:

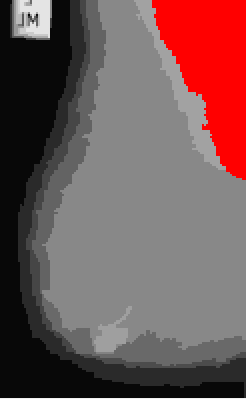


Рисунок 14 – Классификация сегментов изображения  
а)Фиброаденома; б)Молочная железа; в)Мышца;

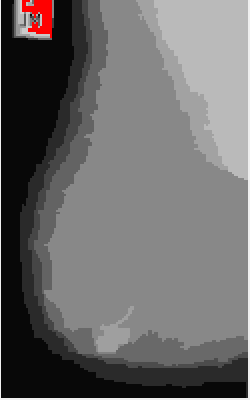
 

Рисунок 15 – Классификация сегментов изображения

г)Фон; д) Блик и прочее

Изначально вручную производился поиск и добавление нужных нам объектов для обучения системы. Для этого применяем нажатие ЛКМ на нужный нам объект на изображении, после чего сегмент выделится в красный цвет. Чтобы добавить объект в базу данных, требуется нажать на кнопку «Записать в БД», выбрав при этом соответствующий объекту класс в соседнем выпадающем списке. Пример добавления представлен на рисунке 12.

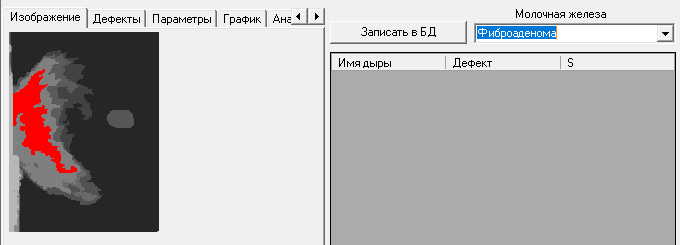


Рисунок 16 – Добавление объекта в базу данных

После множественного добавления информации об объектах в базу данных система стала способна к собственному определению принадлежности сегмента к тому или иному классу. На Рисунке 13 отображен результат работы программы, который показывает, что в правом верхнем углу отмеченный нами сегмент принадлежит классу Фиброаденома. Также в процессе проверки работы системы с добавленными данными были выявлены ошибки определения классификации (Рисунок 14). Данная ошибка связана с тем, что сегменты схожи по своим числовым характеристикам (параметрам), что усложняет точность определения класса разработанной системой. Информацию о характеристике объекта можно увидеть под табличной частью интерфейса, при выбранном объекте на изображении. После обнаружения ошибки классификации, повторяя приведенный выше пример, добавляем данные о данном объекте в базу данных, для повышения точности определения.

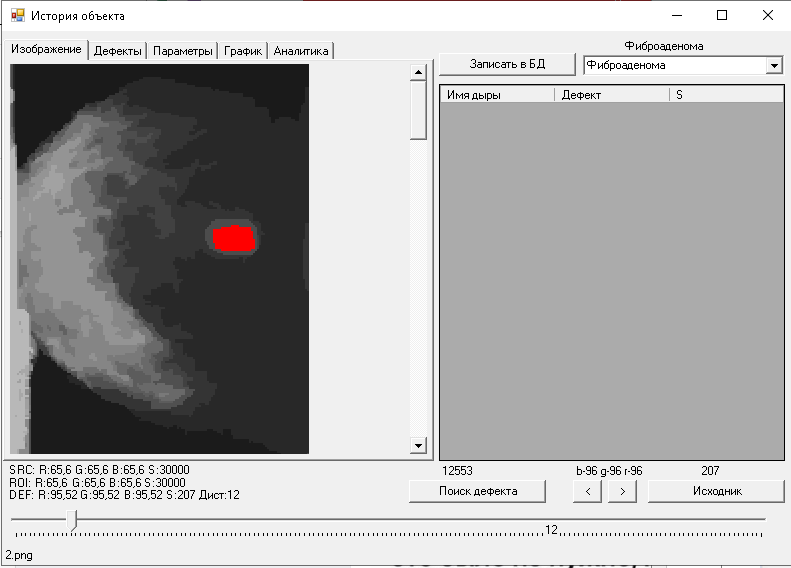


Рисунок 17 – Результат выявления фиброаденомы

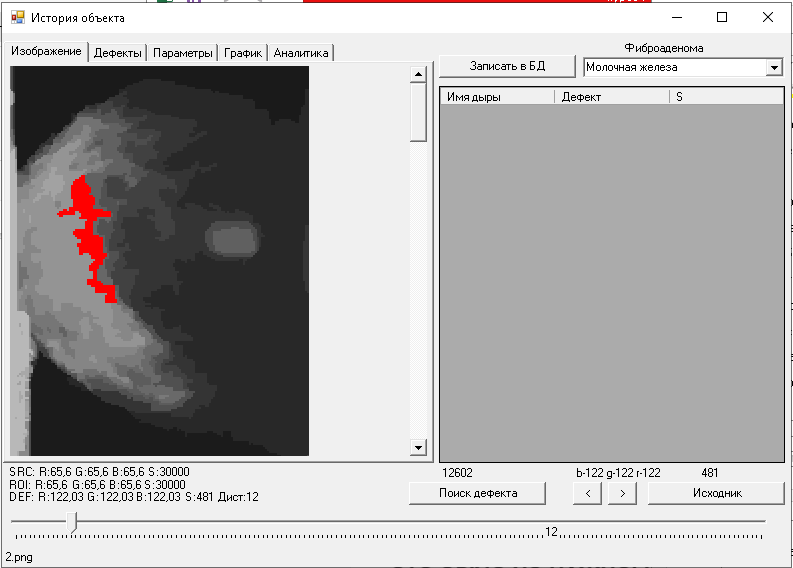


Рисунок 18 – Исправление ошибки классификации системой

Для проверки точности системы вручную проводился анализ правильности принадлежности объекту нужному классу. Данный анализ производился на изображениях с умеренным количеством сегментов, поэтому существует погрешность в вычислении точности классификации. Все исследованные изображения показаны ниже:

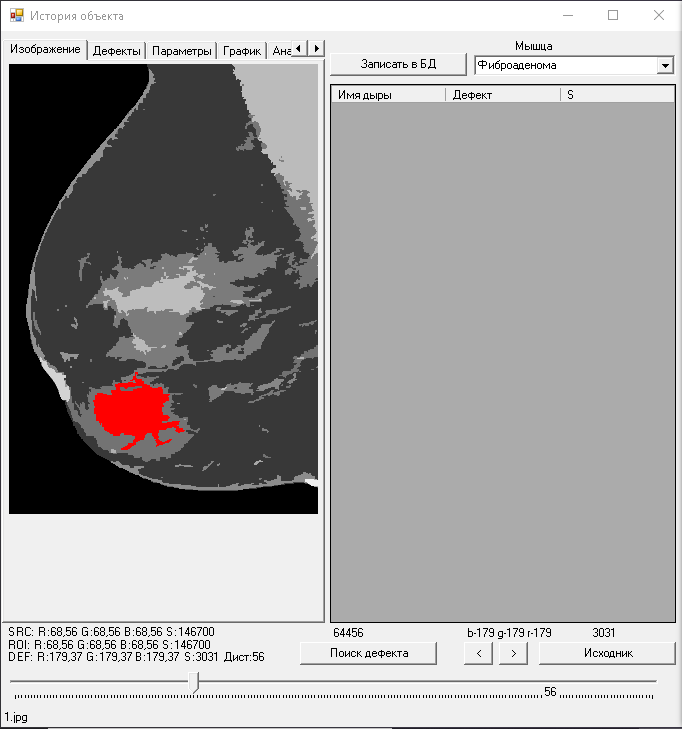


Рисунок 19 – 1.png – Неудача

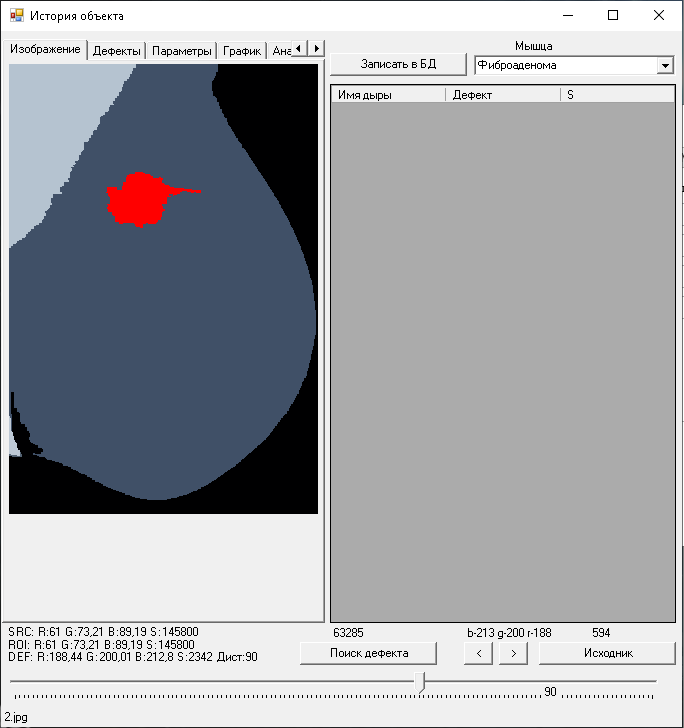


Рисунок 20 – 2.png – Неудача

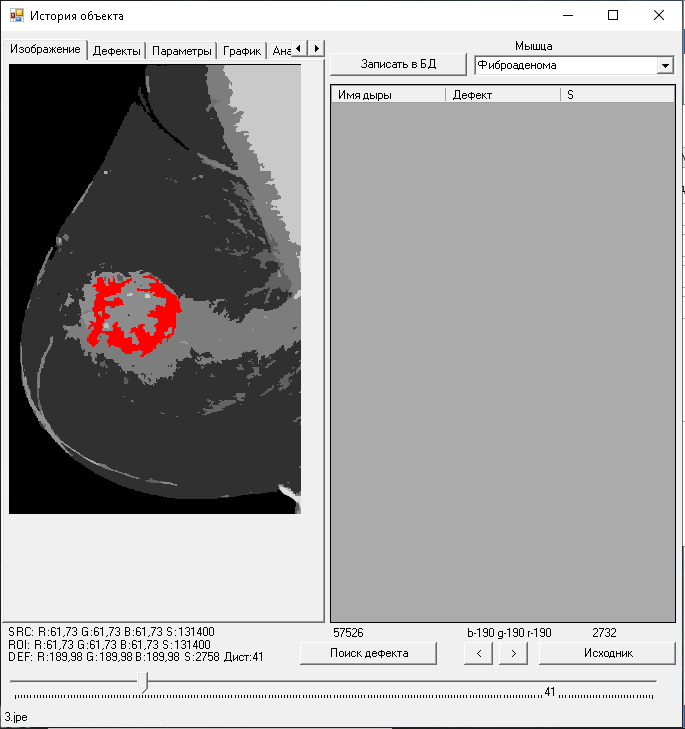


Рисунок 21 – 3.png – Неудача

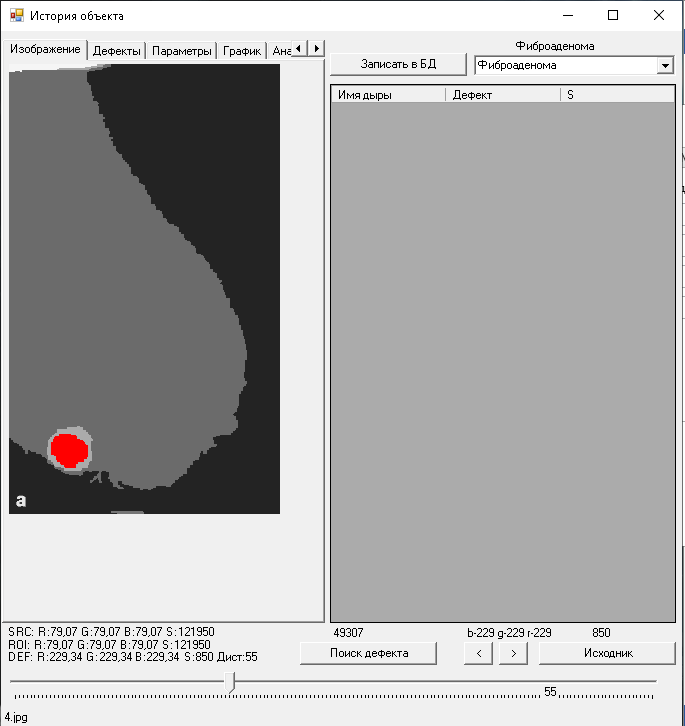


Рисунок 22 – 4.png – Успех

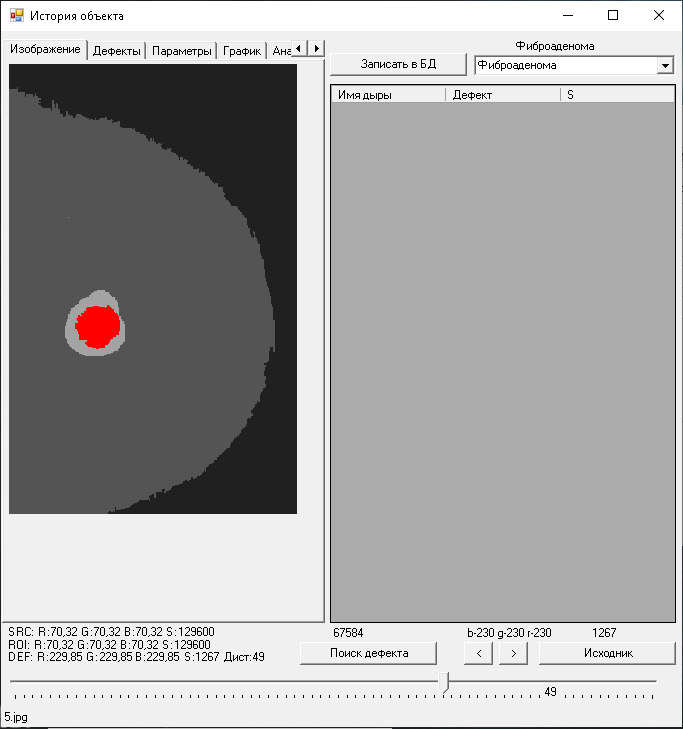


Рисунок 23 – 5.png – Успех

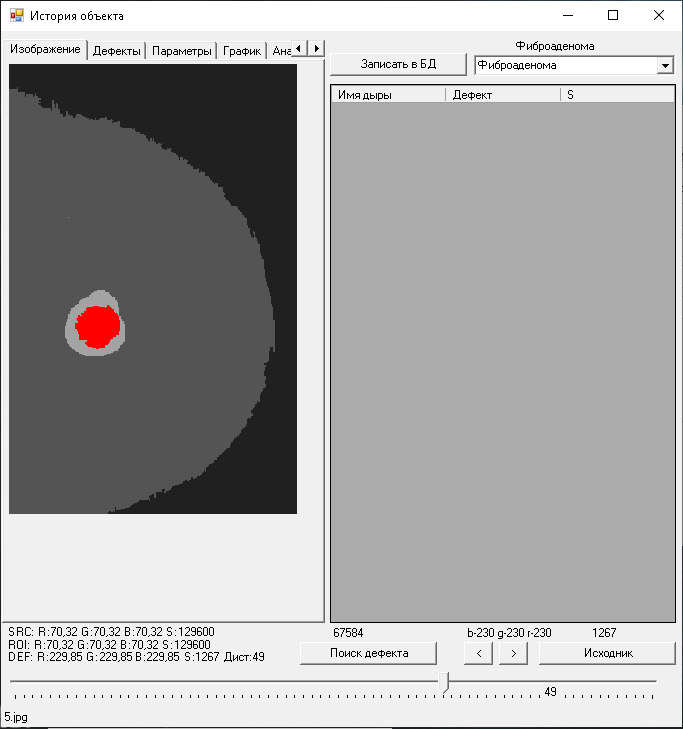


Рисунок 24 – 6.png – успех.

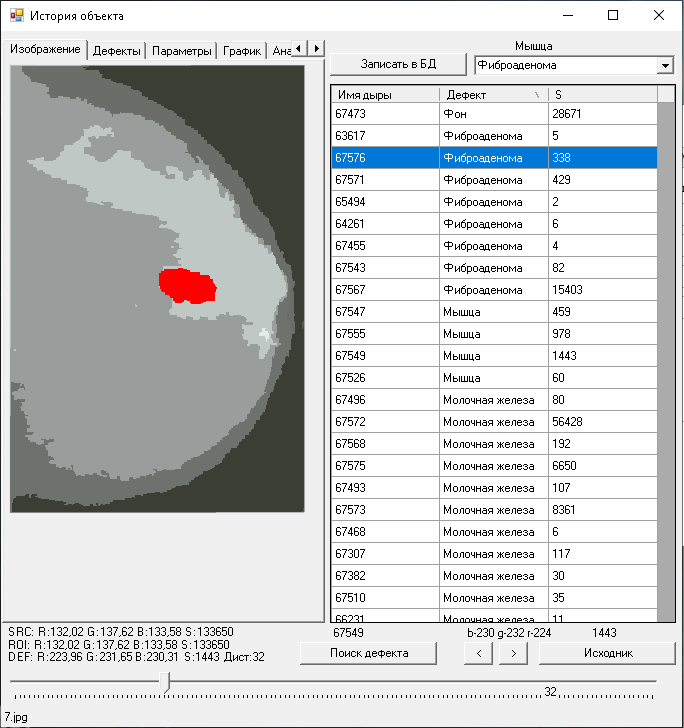


Рисунок 25 – 7.png – Неудача

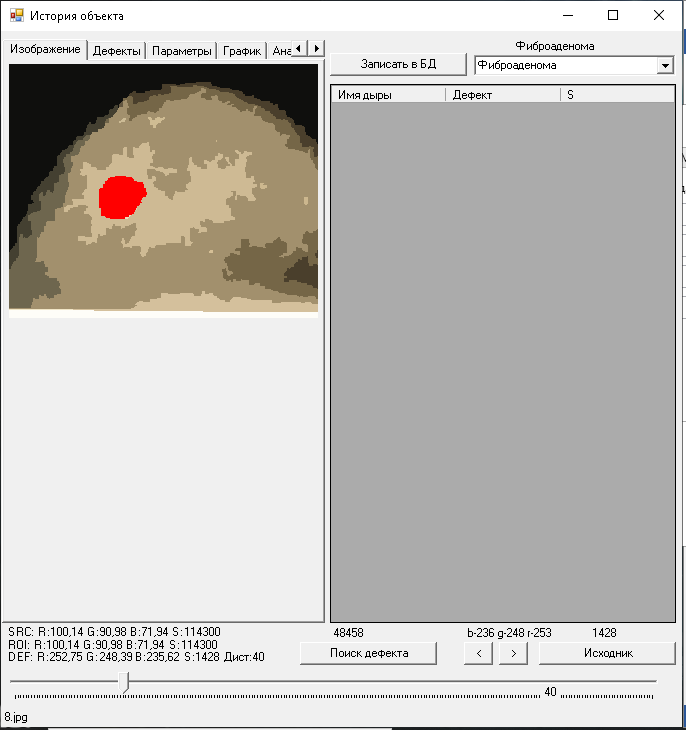


Рисунок 26 – 8.png – успех.

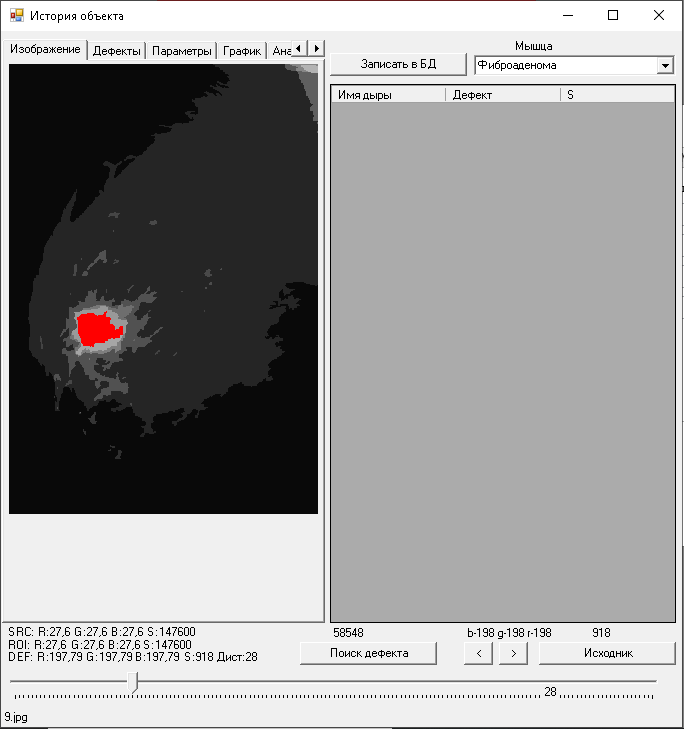


Рисунок 27 – 9.png – Неудача.

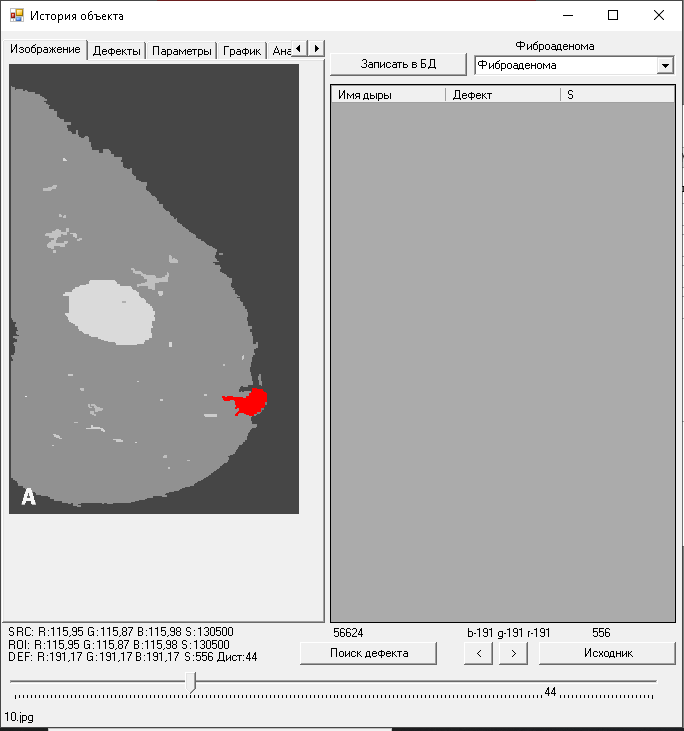


Рисунок 28 – 10.png – Успех

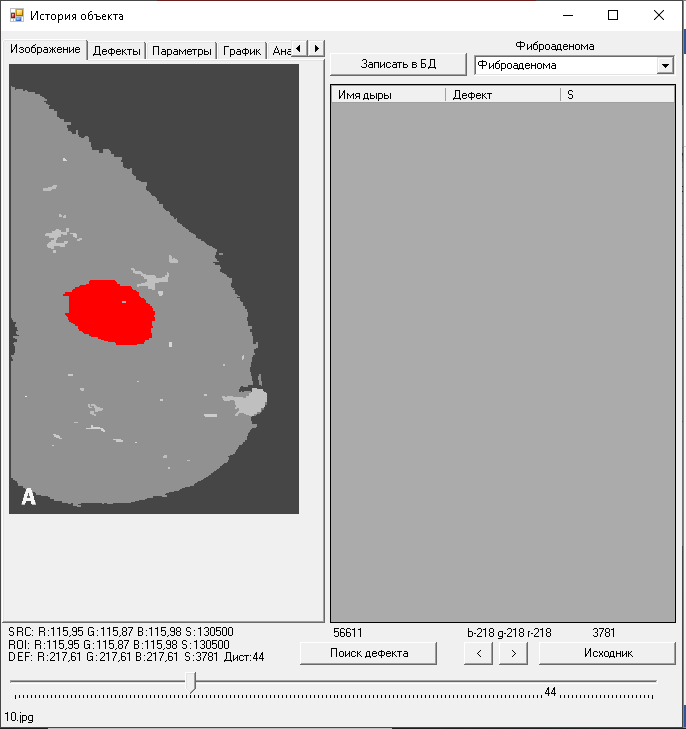


Рисунок 29 – 10.png – Успех

Исходя из проведенных расчетов можно сделать вывод о том, что точность классификации системы равна приблизительно 54%. Данная цифра говорит о сложной структуре строения молочных желез, которое содержит в себе множество различных сегментов, требуемые тчательного анализа специалиста.

# ЗАКЛЮЧЕНИЕ

В данной пояснительной записке приведен процесс реализации алгоритма сегментации маммографических снимков, основанного на методах персистентной гомологии. Данная методология позволяет анализировать изображение, основываясь на информации всего объединения пикселей для всех пороговых значений яркости.

В рамках работы над проектом был проведен анализ предметной области, определены данные к проекту, а также предъявлены требования к разрабатываемому алгоритму. Далее была произведена разработка алгоритма, который классифицирует сегменты изображения.

Исходя из проведенных тестов можно сказать, что алгоритм не является надежным и не может использоваться как основное средство сегментации маммограм, однако может быть использован в качестве помощника для той же сегментации. Этот вывод сделан на основании результатов тестирования, которые показывают точность классификации сегментов изображения равную 54%. Данный процент указывает на необходимость в доработки разработанного алгоритма для выявления более точного результата.

Итогом работы является сбор статистики и анализ полученных данных в процессе тестирования программы для выявления точности работы алгоритма.

# СПИСОК ЛИТЕРАТУРЫ

1. Персистентные гомологии. //URL:http://intsysjournal.org/pdfs/23-2/ kushnar.pdf (дата обращения: 10.03.21).
2. Фиброаденома. //URL:https://radiopaedia.org/articles/fibroadenoma-breast (дата обращения: 8.03.21).
3. Фиброаденома. //URL: https://www.sciencedirect.com/topics/medicine-and-dentistry/fibroadenoma (дата обращения: 9.03.21).