

Министерство образования и науки Российской Федерации



федеральное государственное автономное образовательное учреждение
высшего образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Институт Кибернетики
Направление подготовки 09.04.03 Прикладная информатика _____
Кафедра Программной инженерии _____

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Определение степени патологии при воспалении легких с помощью снимков компьютерной томографии

УДК 616.24-001:620.179.152.1

Студент

Группа	ФИО	Подпись	Дата
8KM51	Лигай Светлана Дореновна		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ПИ	Аксенов С.В.	к.т.н.		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент каф. МЕН	Баннова К.А.	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ЭБЖ	Пустовойтова М.И.	к.х.н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Зав. кафедрой	ФИО	Ученая степень, звание	Подпись	Дата
ПИ	Иванов М.А.	к.т.н.		

Томск – 2017 г.

Запланированные результаты изучения по ООП

Код результатов	Результат обучения (выпускник должен быть готов)
	<i>Профессиональные компетенции</i>
P1	Применять базовые и специальные естественно-научные и математические знания в области информатики, экономики, маркетинга и менеджмента, достаточные для комплексной инженерной деятельности.
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных и экономических задач.
P3	Ставить и решать задачи комплексного анализа, связанные с созданием новых информационных технологий и информационных систем в экономике, с использованием базовых и специальных знаний, современных аналитических методов и моделей.
P4	Разрабатывать новые и модернизировать уже существующие информационные технологии и системы (в экономике) в соответствии с техническим заданием
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области прикладной информатики. Проводить исследования, связанные с оценкой информационной безопасности проектов.
P6	Внедрять, эксплуатировать и обслуживать современные информационные технологии и системы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.
	<i>Универсальные компетенции</i>
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.
P8	Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности.
P9	Эффективно работать индивидуально и в качестве члена группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации.
P10	Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности.
P11	Демонстрировать способность к самостоятельной к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии.

Форма задания на выполнение выпускной квалификационной работы

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Институт кибернетики _____
Направление подготовки (специальность) прикладная информатика _____
Кафедра Программной инженерии _____

УТВЕРЖДАЮ:

Зав. кафедрой

(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8KM51	Лигай Светлана Дореновна

Тема работы:

Определение степени патологии при воспалении легких с помощью снимков компьютерной томографии

Утверждена приказом директора (дата, номер)

Срок сдачи студентами выполненной работы:

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе

(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).

1. Объект исследования – Алгоритмическое программное обеспечение для определения патологии, вызванных воспалением легких.
2. Характеристики ПО -
3. Средства разработки: Visual Studio 2015, язык реализации C#, RapidMiner Studio
4. Исследовать эффективность применения текстового анализа к снимкам компьютерной томографии больных пневмонией

<p>Перечень подлежащих исследованию, проектированию и разработке вопросов</p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> 1. Задача исследования – разработать алгоритмическое ПО для определения патологий, вызванных воспалением легких 2. Аналитический обзор исследований по диагностике пневмонии <ol style="list-style-type: none"> 2.1. Пневмония и средства диагностики пневмонии 2.2. Обзор методов классификации изображений 2.3. Обзор методов текстурного анализа изображений 3. Проектирование алгоритма классификации патологий легкого при пневмонии <ol style="list-style-type: none"> 3.1. Алгоритм сегментации 3.2. Алгоритм GLCM 4. Разработка ПО на основании разработанных алгоритмов. <ol style="list-style-type: none"> 4.1. Выбор средств разработки 4.2. Реализация сегментации снимков КТ и ее эффективность в зависимости от режима отображения 4.3. Реализация алгоритма GLCM 5. Тестирование алгоритма и оценка качества модели <ol style="list-style-type: none"> 5.1. Построение классификатора для оценки качества модели
<p>Перечень графического материала</p> <p><i>(с точным указанием обязательных чертежей)</i></p>	<ol style="list-style-type: none"> 1. Схема алгоритма сегментации снимков КТ 2. Схема алгоритма GLCM
<p>Консультанты по разделам выпускной квалификационной работы</p>	
<p>Раздел</p>	<p>Консультант</p>
<p>1</p>	<p>Аксенов Сергей Владимирович</p>
<p>2</p>	<p></p>
<p>3</p>	<p></p>
<p>4</p>	<p></p>
<p>5</p>	<p>Сидоренко Марина Ивановна</p>
<p>6</p>	<p>Баннова Кристина Алексеевна</p>
<p>7</p>	<p>Сидоренко Татьяна Валерьевна</p>

Названия разделов, которые должны быть написаны на русском и иностранном языках:

1. Аналитический обзор исследований по диагностике пневмонии
2. Проектирование алгоритма классификации патологий легкого при пневмонии
3. Разработка ПО на основании разработанных алгоритмов.
4. Тестирование алгоритма и оценка качества модели

**Дата выдачи задания на выполнение выпускной
квалификационной работы по линейному графику****Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф.ПИ	Аксенов С.В.	к.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8КМ51	Лигай Светлана Дореновна		

Реферат

Выпускная квалификационная работа содержит 95 страниц, 43 рисунка, 29 таблиц и 3 приложения.

Ключевые слова: Пневмония, классификация данных, сегментация, текстура, классификатор.

Объект исследования: патологии пневмонии на снимках компьютерной томографии.

Цель исследования: Повысить качество диагностирования пневмонии с помощью обработки снимков КТ.

Во введении рассмотрена актуальность темы, научная новизна исследования, определен объект и предмет исследования. В первой главе рассмотрены исследования диагностики пневмонии, методы и подходы классификации и сегментации изображений, в том числе текстурный анализ признаков. Вторая глава посвящена разработке алгоритма сегментации для снимков КТ и алгоритма текстурного анализа признаков. Третья глава описывает результаты разработки программного обеспечения. Четвертая глава посвящена тестированию и оценке точности модели. Также в работе приведены разделы социальной ответственности и финансового менеджмента.

Оглавление

Введение.....	8
Раздел 1. Аналитический обзор исследований по диагностике пневмонии.....	10
1.1. Пневмония и средства диагностики пневмонии	10
1.2. Обзор методов классификации изображений.....	13
1.3. Обзор методов текстурного анализа изображений	24
1.4. Краткие выводы по разделу.....	27
Раздел 2. Проектирование алгоритма классификации патологий легкого при пневмонии.....	28
2.1. Алгоритм сегментации.....	28
2.2. Алгоритм GLCM.....	30
2.3. Краткие выводы по разделу.....	32
Раздел 3. Разработка ПО на основании разработанных алгоритмов.....	33
3.1. Выбор средств разработки.....	33
3.2. Реализация сегментации снимков КТ и ее эффективность в зависимости от режимов отображения.....	36
3.3. Реализация GLCM	40
3.4. Краткие выводы по разделу.....	42
Раздел 4. Тестирование алгоритма и оценка качества модели	44
4.2. Результат оценки качества модели	46
4.3. Краткие выводы по разделу.....	46
Раздел 5. Социальная ответственность.....	48
Раздел 6. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	56
Заключение.....	75
Список публикаций	77
Список литературы.....	78
Приложения	83

Введение

Пневмония является одной из самых распространенных легочных заболеваний в мире и занимает весомую позицию среди глобальных проблем человечества.

Ежегодно публикуется пугающая статистика о распространении пневмонии: острая пневмония встречается у 10-14 человек из 1000, в возрастной группе старше 50 лет – у 17 человек из 1000. Хотя при правильном и своевременном лечении прогноз преимущественно положительный, но процент смертности в среднем составляет 5%-7% у взрослых, 17,5% у детей и 13%-15% у больных пожилого возраста.

Подобный довольно высокий процент летальности обусловлен затяжным течением болезни и возникновением осложнений. Большую роль в возникновении такой ситуации играют ошибки диагностирования. В нашей стране в 20% случаев, то есть у каждого пятого больного пневмонию путают с другими заболеваниями.

Сегментация является основным этапом анализа снимков компьютерной томографии при легочных заболеваниях. Сегментация позволяет выделить пораженные участки легкого, что значительно облегчает диагностику.

Самыми распространенными подходами сегментации снимков компьютерной томографии являются алгоритмы разделения, слияния и наращивания областей. Однако данные подходы имеют некоторые недостатки в плане точности определения.

Целью диссертационной работы является повышение качества диагностики пневмонии я помощью методов текстурного анализа снимков компьютерной томографии и деревьев решений.

Для достижения поставленной цели требовалось решить следующие задачи:

1. Провести аналитический обзор исследований по диагностике пневмонии, в том числе существующих алгоритмов классификации и

сегментации изображений, а также рассмотреть методы текстурного анализа.

2. Разработать алгоритм классификации патологии при пневмонии, а также алгоритм сегментации снимков КТ.
3. Разработать программное обеспечение на основании разработанного алгоритма классификации патологий при пневмонии
4. Провести тестирование и оценку качества модели.

Научная новизна данной работы состоит в применении методов работы с текстурными признаками и деревьев решений для классификации патологий легкого при пневмонии.

Личный вклад автора состоит в разработке алгоритмов классификации и программной реализации данного алгоритма, составления классификатора и проведении тестирования модели.

Раздел 1. Аналитический обзор исследований по диагностике пневмонии

1.1. Пневмония и средства диагностики пневмонии

Пневмония – это воспалительный процесс, происходящий в легочных тканях, а также самое распространенное клиническое проявление пневмококковой инфекции у взрослых и основная причина смерти детей до 5 лет [1].

В течении последних нескольких лет происходило переосмысление пневмонии, в следствии получения все больше противоречивых данных, в том числе выявление новых возбудителей заболевания. Даже классификация пневмонии претерпела изменения. Согласно рабочей классификации пневмония делится на следующие виды:

- внебольничная пневмония (приобретена вне лечебных учреждений);
- нозокомиальная пневмония (приобретенная в лечебных учреждениях);
- аспирационная пневмония;
- пневмония у лиц с тяжелыми дефектами иммунитета.

Во всём мире пневмонией заболевают около 450 миллионов человек в год, около 7 миллионов случаев заканчиваются летальным исходом. Согласно статистике ВОЗ, пневмония является одной из 10 самых распространенных причин смерти – в 2012 году от нее умерло 3,1 млн человек.

На рисунке 1 представлена официальная статистика по внебольничным пневмониям на 100 тысяч человек населения РФ с 2013 года.

Внебольничная пневмония (ВП) наиболее распространена среди городского населения. Ежегодно общее число болеющих ВП в России превышает 1 500 000 человек. Это число превышает официальные статистические данные, поскольку включает в себя все случаи заражения пневмонией, а не только тех, учитывающих больных, у которых пневмония стала основной причиной обращения за медицинской помощью или причиной летального исхода.



Рис 1. Статистика внебольничной пневмонии на 100 тыс. населения

Процент летальности от внебольничной пневмонии составляет 5–7%, у госпитализированных пациентов – от 8 до 14%. Наибольшему риску подвержены больные старшей возрастной группы с сопутствующими заболеваниями или при тяжелом течении болезни летальность составляет 15–30%. Пневмония составляет 10% всех летальных исходов в крупных городах России.

Симптомами пневмонии являются слабость, снижение аппетита, мышечные и головные боли, сильный кашель, одышка, а также гной и жидкость, заполняющая альвеолы, что затрудняет вентиляцию и газообмен в легких, вызывает жар, озноб, затруднение дыхания и другие симптомы. Для пневмонии характерно воспаление тканей легкого и, следовательно, увеличение его в размере.

У взрослых течение заболевания постепенное, поэтому возникают сложности для ранней диагностики. Однако при длительном сроке заболевания возможны осложнения в виде:

- воспаление плевры – плеврит;
- абсцесс легкого – процесс образование полости с гнойным содержимым по причине расплавления локализованного участка воспалительного процесса;

- гангрена легкого – распад легочной ткани;
- бронхообструктивный синдром – больной испытывает одышку, недостаток кислорода;
- острая дыхательная недостаточность – легкие не в состоянии обеспечивать приток в организм необходимое количество кислорода.

Симптоматическое диагностирование острых инфекционных заболеваний не является эффективным, таким образом, к методам диагностики пневмонии относят следующие мероприятия:

- Лабораторное исследование крови;
- Анализ мокроты;
- Рентгенологическое исследование грудной клетки;
- Компьютерная томография.

Однако, учитывая постепенное нарастание симптомов заболевания рентгенологическое исследование не может выявить всю полноту картины течения заболевания. Ошибки в диагностике пневмоний достигают 30%, диагноз пневмонии в первые три дня ставится в 35% случаев.

Таким образом, очевидно, что процент ранней диагностики недостаточно высок, поэтому развитие более совершенных методов диагностирования пневмонии, в том числе анализ снимков компьютерной томографии, становится невероятно актуальным.

Среди методов диагностирования пневмонии наиболее точным является компьютерная томография. На данный момент для наибольшей эффективности определения поражения легких при небольших затемнениях используют топографическую компьютерную томографию. Данный метод показал высокую точность диагностики – 10,9%. [6]

Снимки компьютерной томографии позволяют наиболее точно и своевременно поставить диагноз пациентам, однако, зачастую качество снимков снижено за счет посторонних шумов. В качестве подобных шумов могут выступать слои одежды при сканировании.

Для облегчения диагностирования пациентов и анализа снимков возможно применение различных программных комплексов. В основу таких программных комплексов ложатся алгоритмы классификации и сегментации изображений. Задача, которую должна решать автоматическая сегментация снимков компьютерной томографии больных пневмонией, представляет собой необходимость удаления всех посторонних объектов на изображении, определения пораженных и здоровых областей легкого.

1.2. Обзор методов классификации изображений

Классификация данных является огромным и стремительно развивающимся разделом информационных технологий, ее алгоритмы призваны анализировать данные и распределять их в некоторую упорядоченную систему. Это наиболее простое и общее описание классификации.

Итак, классификация — это форма анализа данных, при которой выделяются модели, описывающие важные классы данных. Такие модели, называемые классификаторами, представляют собой метки для категории классов.

Классификация — двухступенчатый процесс, состоящий из стадии обучения и непосредственной классификации. Первый шаг, то есть процесс обучения или тренировки системы включает в себя построение общей концепции или классификатора на основании уже определенных данных и классов.

Алгоритмы классификации производят построение классификаторов на основании тренировочных наборов или кортежей данных, которым присущи собственные классификационные метки. Кортеж X представлен в виде n -мерного вектора $X=(x_1, x_2, \dots, x_n)$, описывающего n измерений для кортежа из n атрибутов из базы данных A_1, A_2, \dots, A_n , соответственно. Каждый кортеж X предполагается, как принадлежащий к предопределенному классу, то есть определенный другим атрибутом класса, называемым атрибутом классификационной метки.

Точность классификатора на заданном тестовом наборе данных является процентом от тестового кортежа, который корректно распределен этим классификатором.

На рисунке 3 показан основной принцип классификации данных. Для обучения системы используются некоторые наборы тренировочных кортежей, на основании которых строится и уточняется модель. После все данные проходят непосредственный процесс классификации. Для достижения необходимой точности модель обучают неоднократно.

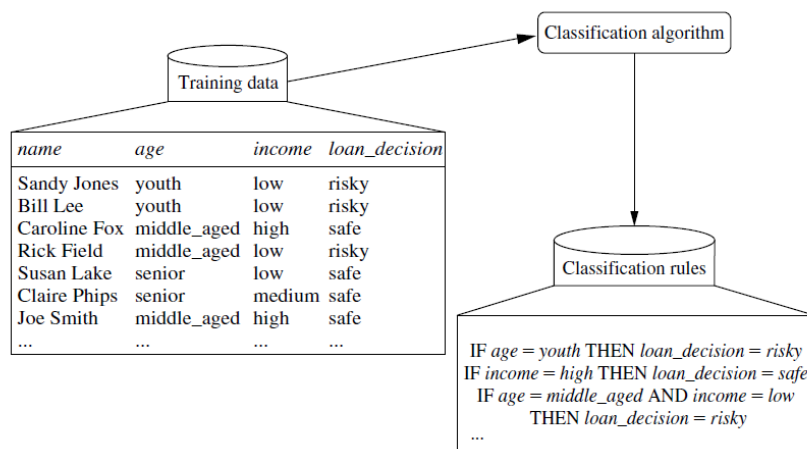


Рис 2. Процесс обучения

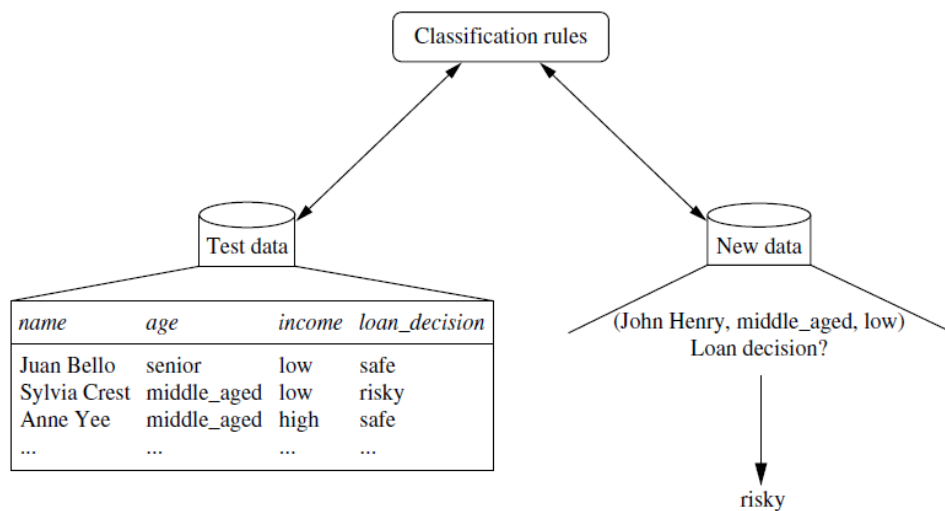


Рис 3 Процесс классификации

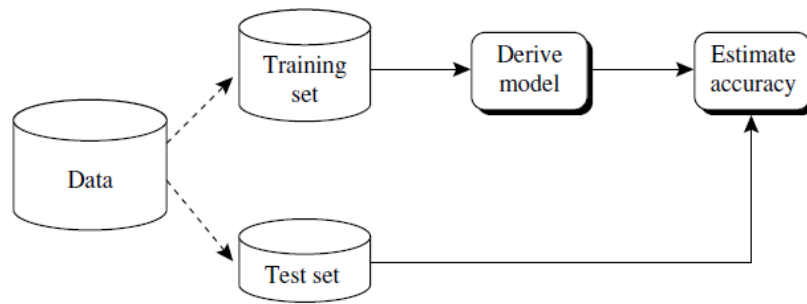


Рис 4. Основной принцип классификации

Большая часть алгоритмов классификации данных построена на данной двухступенчатой схеме, однако они также в большинстве своем резистентны, то есть типично рассчитаны на работу с небольшим количеством данных.

Сегментация – это процесс разбиения изображения на сегменты, которые представляют собой множества пикселей, объединенных по тем или иным признакам [3]. На сегодняшний день сегментация изображений широко применяется в медицине для обнаружения опухолей и других патологий, определения объемов тканей, изучения анатомической структуры [4].

Для выявления наиболее эффективных методов сегментации изображений был произведен сравнительный анализ. Среди методов, основанных на операторах выделения краев, методов водораздела, метода выращивания регионов, метод нормальных разрезов и методов оценки изменения формы спектра. Также предложена модификация метода нормальных разрезов. [7]

Среди трех поколений методов сегментации изображений были выделены методы, отличающиеся точностью, автономностью, сложностью, возможностью конкретизации данных алгоритмов для определенных задач и многими другими параметрами. [8]

Сегментация изображений является развивающейся областью. Множество уже существующих методов и алгоритмов сегментации получают развитие. Одним из таких алгоритмов является алгоритм, основанный на основе анализа связанных компонентов и новых принципов представления полутоновых изображений. [9]

Также в результате исследований сегментации снимков земной поверхности был разработан алгоритм автоматической сегментации, который основан на светочувствительности элементов различных спектральных каналов. Для реализации алгоритма используются коллометрический алгоритм и Байесовский классификатор. [11]

Благодаря применению алгоритмов сегментации и классификации спутниковых радарных изображений позволяет выделять однородные участки изображений без удаления шумов снимков. [12]

В целях диагностики скелетных метастазов была разработана система автоматизированной компьютерной диагностики, которая обладает функцией экспертного анализа снимков. Система включает в себя сегментацию изображений, создание обучающей выборки изображений после медицинской обработки. [13]

Сегментация спутниковых снимков высокого разрешения позволяет разработать эффективные методы формирования текстурных признаков для полутоновых изображений. Подобные исследования позволяют ознакомиться с непараметрическими алгоритмами кластеризации и методами сегментации любых изображений высокого разрешения.

Для определения текстурных признаков следует выделять несколько подходов: статистический, структурный и спектральный. [14]

Исследование текстурных признаков снимков с большим разрешением требует рассмотрения статистических методов, которые оценивают статистические характеристики. Данные характеристики определяются на основании текстурных матрицы, которая является матрицей совместимости пар соседних пикселей на уровне одного тона. [15]

На основании проведенных исследований и комплексного анализа возможно определить наиболее эффективные методы и алгоритмы, которые могут лечь в основу алгоритма сегментации снимков компьютерной томографии для больных пневмонией. В основе всех высокоуровневых методов сегментации лежат базовые алгоритмы. Все алгоритмы сегментации могут быть разделены по

принципу своего действия, то есть по применяемым действиям для формирования кластеров. Можно выделить группы: попиксельные, граничные, на основании выделения однородных областей и гибридные. [16]

Попиксельные методы не учитывают отношения соседних пикселей или их положение на изображении, а разделяют кластеры на основе только пространства признаков. Разделение при данном подходе может осуществляться с помощью фиксированных порогов, например, дерева решений. Однако эффективным данный подход может только при наличии априорных знаний. При наличии таких знаний данная задача может представляться задачей классификации с обучением или кластеризацией без обучения. [17]

Граничные методы представляют собой методы поиска граничных пикселей с помощью градиентных фильтров, таких как Робертса, Собела, Лапласа и других. После происходит замыкание границ, которое и определяет выделяемые зоны. Основным алгоритмом, применяемым в качестве метода заполнения контуров, является алгоритм водораздела. Сильная зависимость от шума является существенным недостатком данного подхода. [18]

Методы, основанные на анализе областей, включают в себя методы разделения изображений на однородные области, учитывая спектральные и текстурные признаки, при помощи анализа пространственной близости пикселей. Методы наращивания областей, разделения и слияния областей также входят в данный раздел. [19]

Задача кластеризации представляет собой разделение множество объектов на сравнительно небольшое количество подмножеств, которые называются кластерами. Кластеры должны отличаться друг от друга существенно в пространстве признаков, а объекты внутри кластеров должны быть схожими. [20] Для каждой конкретной задачи необходимо подбирать меры схожести для объектов и количества кластеров, так как определенного ответа на данный вопрос не существует. [21]

На настоящий момент известно несколько сотен методов кластеризации данных. Данная тематика рассматривается во множестве статей и монографий

(например, 20-27). Обзор методов кластеризации позволяет сделать выводы о сложности данной проблемы и невозможностью создания универсального алгоритма кластеризации.

Данное утверждение подтверждается данными представленными на рисунке ниже. На нем изображена модель данных, которая содержит кластеры различной формы и плотности, а также «шум» (рис.5а). На данный момент не существует алгоритма способного правильно распознать и выделить все кластеры. Часть б рисунка изображает желаемый результат кластеризации изображения.

Качественная сегментация должна сочетать в себе не только спектральные, но и текстурные признаки. Таким образом, исследования, позволяющие сочетать спектральные и текстурных признаков являются активно развивающимся направлением.

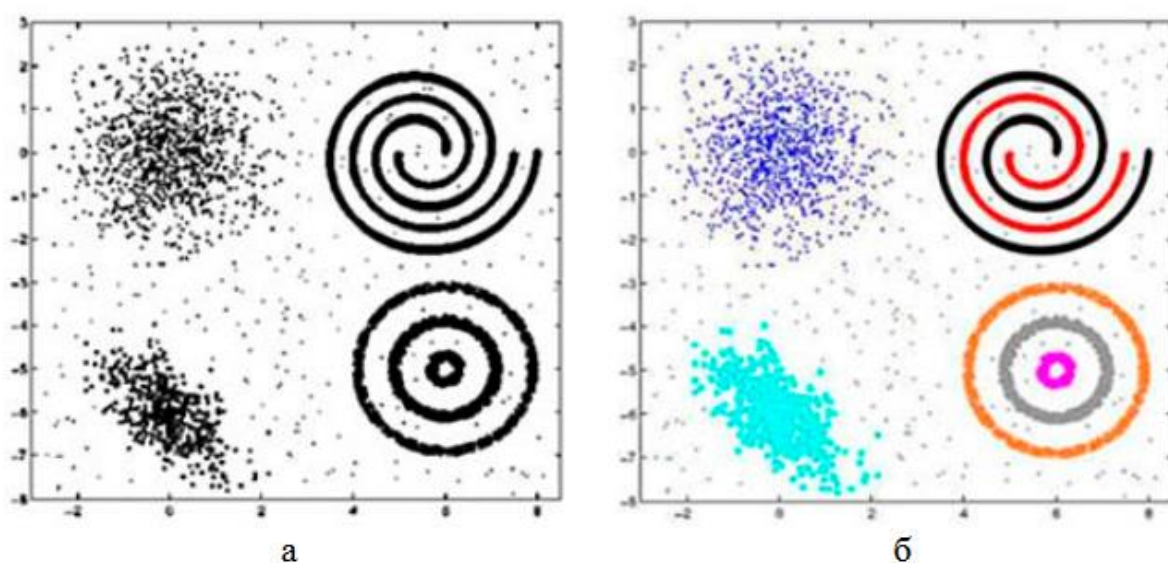


Рис 5. а) Кластеры различной формы, размера и плотности; б) желаемый результат кластеризации

Извлечение спектральных и текстурных признаков различается в зависимости от подходов: методы на основе последовательного извлечения спектральных и текстурных характеристик, методы, основанные на независимом извлечении характеристик и вероятностно-статистические методы. [29]

Первый подход представляет собой последовательное извлечение текстурных и спектральных характеристик. Однако существенным недостатком подобного подхода является то, что критерии извлеченные первыми имеют больший вес при сегментации. Ярким примером последовательного извлечения является алгоритм JSEG.

Алгоритм JSEG производится в три этапа. На первом этапе проводится адаптивное квантование, затем формируется J-изображение, которое содержит локальную однородность, которая зависит от равномерности распределения цветовых меток в окрестности каждого пикселя. Это необходимо для определения границ кластеров, в качестве идентификатора. Заключительный этап представляет собой использование метода наращивания кластеров, при этом соседние сегменты близкие по цвету могут объединяться.

Недостатком алгоритма JSEG является возможная пересегментация за счет отсутствия учета неоднородностей текстур. На данный момент существуют модификации алгоритма JSEG, которые производят дополнительную кластеризацию по текстурным признакам.

Второй подход основан на независимом извлечении спектральных и текстурных характеристик, при чем с учетом пространственных связей между пикселями. К ним относятся алгоритмы слияния и разделения [30-32], алгоритмы наращивания областей [14], алгоритм на основе активных контуров [34].

Алгоритмы разделения и слияния производятся также в три этапа. Первоначально производится разделение изображения на сегменты, в основном квадратные, затем соседние сегменты объединяются, если расстояние между ними не превышает заданного значения. Расстояние представляет собой сумму расстояний по текстурным и спектральным признакам. На заключительном этапе происходит попиксельная коррекция границ сегментов. Преимуществом данного подхода является быстрое действие, поэтому он лег в основу многих алгоритмов.

Более точное определение границ сегментов возможно при использовании алгоритмов наращивания областей. Первоначально они

подбирают «семена» - исходные точки, затем итеративно присоединяют к ним соседние пиксели. Основным недостатком данного подхода является сильная зависимость результата от правильного выбора «семян» [9].

Алгоритмы на основе активных контуров представляют собой итеративное изменение контуров сегмента, при этом полагается, что пиксели внутри первоначального контура имеют высокую степени однородности. Недостатком также является высокая зависимость от выбора исходных параметров. Преимуществом является высокая точность определения границ сегмента [34].

Третий подход основан на вероятно-статистическом подходе. Данный подход рассматривает все изображение целиком как единое спектрально-текстурное пространство. Примерами подобного подхода могут быть алгоритмы К-средних, нейронные сети, нечеткой кластеризации. Недостатками таких подходов является необходимость изначального задания количества кластеров, большие вычислительные затраты [35].

Рассмотрим несколько конкретных алгоритмов классификации, использующие приведенные выше подходы:

- искусственные нейронные сети

Искусственные нейронные сети (ИНС) – сети, в качестве вершин которых выступают искусственные нейроны, которые основаны на работе биологических нейронов.

ИНС позволяет интерпретировать вектор входных параметров с учетом их весов в выходной вектор. Подобный механизм рассматривается как грубое приближение к работе человеческого мозга.

ИНС способна обучаться на «эталонных» или тестовых векторах, выделять характерные признаки и использовать их в дальнейшем для оценки степени схожести входных векторов с эталонными.

Ниже представлена обобщенная схема искусственного нейрона на основании которых строится ИНС (рис 4).

Здесь $X = [x_1, x_2, \dots, x_N]^T$ - входной вектор;

$x_0 = 1$ - константный "псевдосигнал", называемый сигналом поляризации или, просто, поляризатором.

x_j — действительные числа, возможно, размерные, но чаще нормированные.

$W_i = [w_{i1}, w_{i2}, \dots, w_{iN}]^T$ — вектор весов входного вектора. Веса входных сигналов используются для вычисления взвешенной суммы u_i по формуле $u_i = \sum_{j=0}^N w_{ij}x_j$.

Взвешенная сумма входных сигналов u_i служит аргументом функции активации ИН $f(u_i)$, определяющей значение выходного сигнала y_i .

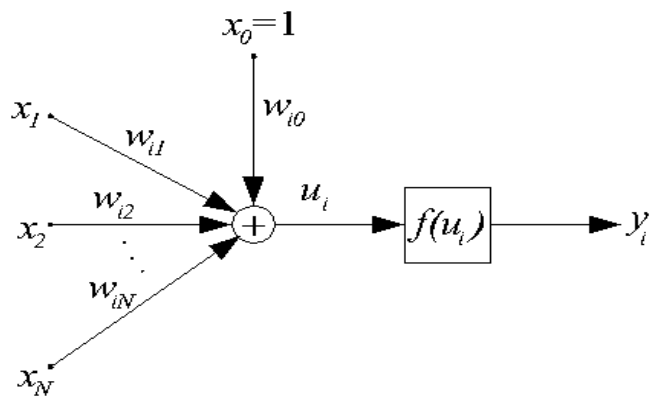


Рис 6. Обобщенная структурная схема искусственного нейрона

- деревья решений, символьные правила

Деревья решений – это способ представления правил в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел, дающий решение.

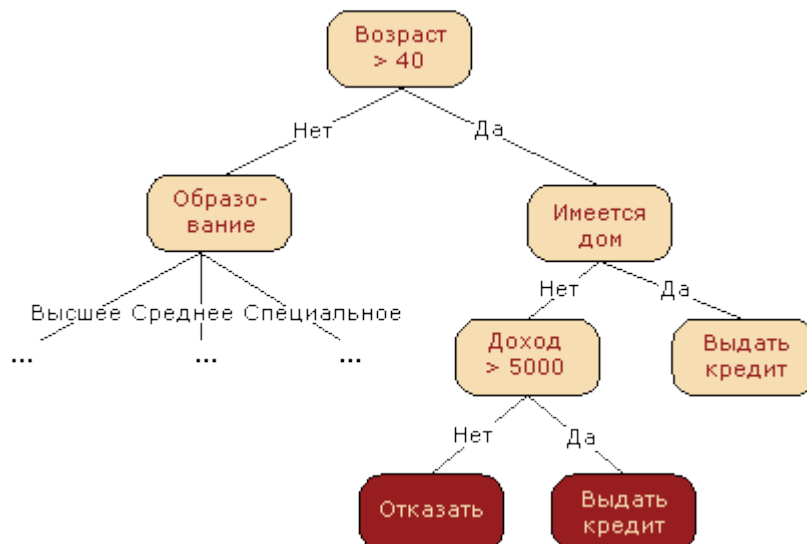


Рис 7. Простейший пример дерева решений

Под правилом понимается логическая конструкция, представленная в виде "если ... то ...".

- методы ближайшего соседа и k-ближайшего соседа

Метод ближайшего соседа является, пожалуй, самым простым алгоритмом классификации. Классифицируемый объект x относится к тому классу y_i , которому принадлежит ближайший объект обучающей выборки x_i .

Метод k ближайших соседей. Для повышения надёжности классификации объект относится к тому классу, которому принадлежит большинство из его соседей — k ближайших к нему объектов обучающей выборки x_i . В задачах с двумя классами число соседей берут нечётным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным классам.

Метод взвешенных ближайших соседей. В задачах с числом классов 3 и более нечётность уже не помогает, и ситуации неоднозначности всё равно могут возникать. Тогда i-му соседу приписывается вес w_i , как правило, убывающий с ростом ранга соседа i . Объект относится к тому классу, который набирает больший суммарный вес среди k ближайших соседей.

- линейная регрессия

Линейная регрессия — метод восстановления зависимости между двумя переменными. Ниже приведен пример программы, которая строит линейную модель зависимости по заданной выборке и показывает результат на графике.

Для заданного множества из m пар (x_i, y_i) , $i=1,2,\dots, m$, значений свободной и зависимой переменной требуется построить зависимость. Назначена линейная модель $y_i = f(w, x_i) + \varepsilon_i$ с аддитивной случайной величиной ε . Переменные x, y принимают значения на числовой прямой \mathbb{R} . При таких предположениях параметры w регрессионной модели вычисляются с помощью метода наименьших квадратов.

- разнообразные методы визуализации данных

К визуальным методам представления данных относятся построение графиков, диаграмм, гистограмм и т.д.

Различные методы классификации характеризуются определенными свойствами, которые могут быть определяющими при выборе метода анализа данных. Методы можно сравнивать между собой, оценивая характеристики их свойств.

Основные свойства и характеристики методов классификации: точность, масштабируемость, трудоемкость, быстрота.

Масштабируемость – свойство вычислительной системы, которое обеспечивает предсказуемый рост системных характеристик, например, быстроты реакции, общей производительности и пр., при добавлении к ней вычислительных ресурсов.

В таблице 1 приведена сравнительная характеристика некоторых распространенных методов. Оценка каждой из характеристик проведена следующими категориями, в порядке возрастания: чрезвычайно низкая, очень низкая, низкая/нейтральная, нейтральная/низкая, нейтральная, нейтральная/высокая, высокая, очень высокая.

Приведенные методы классификации данных могут быть использованы для машинного обучения на основании набора компьютерной томографии пациентов больных пневмонией для определения пораженных участков

Таблица 1. Сравнительный анализ методов классификации

Алгоритм	Точность	Масштабируемость	Трудоемкость	Быстрота
Линейная регрессия	нейтральная	высокая	нейтральная	высокая
Нейронные сети	высокая	низкая	нейтральная	очень низкая

Методы визуализации	высокая	очень низкая	очень высокая	Чрезвычайно низкая
Деревья решений	низкая	высокая	высокая	высокая
k-ближайшего соседа	низкая	очень низкая	нейтральная	высокая

На основании данного анализа можно сделать выводы об использовании тех или иных методов в реализации конкретной задачи. Методы визуального представления не подходят для решения сложных задач по классификации кластеров изображений.

1.3. Обзор методов текстурного анализа изображений

Текстура является важной характеристикой, которая присутствует в любом изображении. Однако само понятие и концепция текстуры не определена достаточно точно. В большей степени это связано с восприятием изображений человеческой визуальной системой.

Существует две основные концепции определения текстуры. Первая заключается в представлении как повторение базовых примитивов, которые имеют различную ориентацию в пространстве. Данное определение ориентировано на структурированной природе текстуры [37]. Вторая концепция представляет собой определение текстуры, как однородный аспект, который не имеет ярко выраженных краев [38].

Сочетание этих двух определений может стать лучшим определением термина «текстура». Ее можно определить, как двухуровневую структуру, которая является пространством базовых примитивов, имеющих случайный аспект [39].

Текстуры можно подразделить по признакам базовых примитивов и их взаимодействию между ними на мелкозернистые, крупнозернистые, гладкие, гранулированные и холмистые. Например, при рассмотрении маленького

участка изображения текстура может быть мелкозернистой, но с возрастанием участка изображения базовые элементы охватывают все больше элементов раstra, то текстура увеличивает зернистость.

Текстуры могут быть слабыми и сильными в зависимости от пространственного взаимодействия между базовыми примитивами, то есть частоты их повторения на некотором участке изображения.

Для анализа и измерения текстур изображения существуют ряд методов. Все эти методы можно разделить на несколько групп:

- Статистические методы

Статистические методы основаны на выявлении статистических текстурных признаков областей изображения. В основе статистического подхода лежит вычисление матриц смежности, которые учитывают уровень яркости элементов и их взаимное расположение. Для вычисления матрицы смежности используются такие текстурные признаки как угловой момент, инерция, энтропия, корреляция и т.д.

Недостатками данного подхода являются высокая детальность сегментации, то есть большое количество небольших сегментов, которые могут быть объединены в несколько больших, также высокая вычислительная сложность.

- Структурные методы

Структурный подход основан на том, что текстура состоит из регулярно повторяющейся совокупности базовых примитивов, которые расположены в соответствии с некоторыми правилами размещения. Таким образом, после определения примитивов и правил их расположения возможно применение двух основных подходов к анализу текстур. Первый подход основан на вычислении статистических признаков и использовании их в качестве текстурных признаков, второй подход определяет принципы расположения примитивов.

- Фрактальные методы

Фрактальные методы отличаются тем, что текстуры в них называются фракталами, то есть составная структура, части которой подобны целому. Таким

образом, фрактальный подход использует самоподобие фрактала. Фрактальные размерности искусственных объектов существенно отличаются от естественной, что является неоспоримым преимуществом при необходимости поиска искусственных изменений.

Недостатком данных методов является невозможность построения общей фрактальной модели всего изображения, а также высокая трудоемкость.

- **Спектральные методы**

Спектральные методы основаны на спектральных мерах текстур, то есть на спектре Фурье, который подходит для описания направлений в периодических текстурах. Пространственные частоты определяют размер зерна текстуры, то есть высокие пространственные частоты характерны для мелкозернистых текстур, а низкие – для крупнозернистых.

- **Комбинированные или смешанные методы**

Комбинированные методы основаны на использовании сочетания признаков различных групп.

Любая группа методов обладает как преимуществами, так и недостатками. Так как целью данной исследовательской работы является исследование автоматического определения патологии легкого при пневмонии, то для дальнейшего рассмотрения не подходят фрактальные и структурные методы, так как направлены на анализ конкретных периодических структур. И хотя спектральные методы были тестированы на анализ текстур легкого, однако данные исследования основывались на отделении снимков больных пациентов от здоровых и основывался на текстурном отличии здоровых легких от больных. Таким образом, можно сделать вывод о целесообразности рассмотрения статистических методов.

Группа статистических методов анализа текстур насчитывает множество методов, однако для сравнения рассмотрим два метода, которые хорошо зарекомендовали себя при решении сложных и комплексных задач: метод основанный на получении статистических данных с помощью гистограммы

абсолютных разностей (Grey-Level Difference) и метод матрицы смежности уровней серого тона (Grey-Level Co-occurrence Matrix).

В [36] были проведены исследования о эффективности данных методов на примере сегментации спутниковых снимков. Для тестирования были выбраны набор снимков из альбома Бродатца и фрагменты снимков спутника QuickBird. В ходе исследований было установлено, что информативность признаков, получаемых с помощью метода GLCM зависит от уровня квантования серого тона. В целом точность классификации для метода GLD составила 95,43%, а для GLCM – 98,29%.

1.4. Краткие выводы по разделу

Пневмония является распространенным и опасным заболеванием, что подтверждают проценты заболеваемости и летальности. Среди причин высокой летальности не последнее место занимает несвоевременная диагностика и ошибки диагностирования. Самым эффективным средством диагностики является компьютерная томография, и для облегчения распознавания патологий на снимках возможно использование классификации данных, в том числе сегментация и текстурный анализ изображения.

На основании рассмотрения работ [4-35], в которых приведены примеры и разработки алгоритмов сегментации изображений можно сделать вывод о целесообразности применения алгоритмов разделения и слияния для сегментации снимков компьютерной томографии больных пневмонией. В силу небольшого количества снимков вычислительные затраты не превысят пороговое значение, а также быстроедействие алгоритмов является неоспоримым преимуществом.

В процессе тестирования и оценки эффективности алгоритмов текстурного анализа метод GLCM показал наибольшую эффективность и точность при распознавании текстур.

Раздел 2. Проектирование алгоритма классификации патологий легкого при пневмонии

2.1. Алгоритм сегментации

Сегментация изображений позволяет разделить исходное изображение на несколько однородных областей, все элементы внутри которых сходны по некоторым параметрам. Результатом сегментации является совокупность сведений об исходном изображении, называемая картой изображения. Сегментация позволяет уменьшить объем информации на изображении, а также облегчить его последующий анализ.

Для сегментации снимков компьютерной томографии больных пневмонией был использован алгоритм, основанный на методах разделения и слияния, так как данный подход показал свою эффективность в ходе аналитических исследований.

Каждый пиксель на снимке компьютерной томографии содержит номер цвета в палитре – от 0 до 255, где 0 – это белый, а 255 – черный. Так как на снимке здоровые области легкого имеют черный цвет, а пораженные – белый, для облегчения анализа изображения необходимо проанализировать оттенки серого. Для человеческого глаза оттенки серого трудноразличимы на снимке, поэтому врач не всегда может точно определить является ли область пораженной или нет, особенно это касается очень маленьких областей.

Каждый пиксель анализируется на близость цвета к белому или черному и получает соответствующий номер из палитры. Например, если пиксель имеет цвет под номером 10, то ему присваивается номер 0, который соответствует белому цвету. Граница разделения между 0 и 255 задается в программном коде и стандартно установлена на 60, то есть если номера пикселей находятся в диапазоне от 60 до 255, то им присваиваются номер 0, то есть белый цвет. После подобной сегментации определение пораженных и здоровых областей легких значительно облегчается.

Также данный алгоритм позволяет объединять маленькие пораженные области, невидимые невооруженным взглядом, в общие сегменты. Это

происходит за счет рассмотрения небольших квадратных участков изображения, размером 9×9 пикселей. Каждый пиксель анализируется и ему присваивается номер, соответствующий его диапазону. После вычисляется среднее значение, и весь квадрат получает номер, соответствующий номеру среднего диапазона.

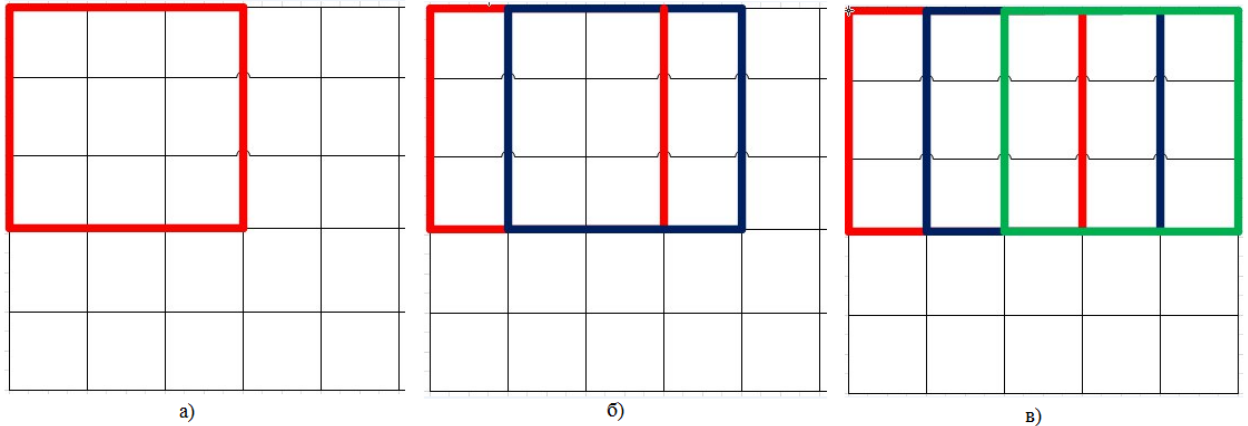


Рис 8. Пример сегментации изображения

Для примера рассмотрим участок изображения 5×5 пикселей и квадрат 3×3 . На рисунке 13а) показан участок изображения, на котором рассматривается квадрат размером 3×3 пикселя. Алгоритм определяет номера цветов каждого пикселя внутри квадрата, после чего выбирает среднее и сравнивает с границей заданного диапазона и присваивает ему номер 0 или 255. Сдвиг квадрата осуществляется на 1 пиксель, как показано на рисунке 13б), и после прохождения первого ряда (рисунок 13в), сдвиг производится на 1 пиксель по вертикали и процесс повторяется (рисунок 14).

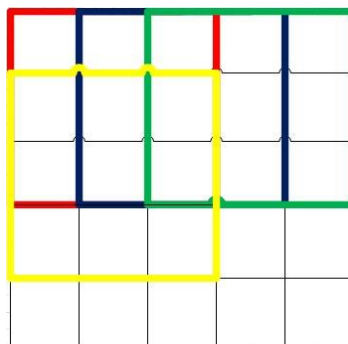


Рис 9. Прохождение второго ряда

После полученный номер цвета задается в верхний левый пиксель нового изображения (рис 15).

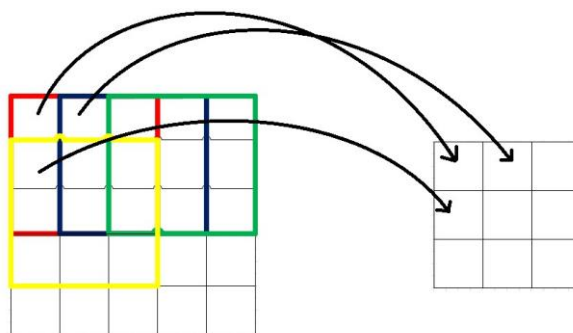


Рис 10. Запись номера цвета в пиксели нового изображения

Данный алгоритм позволяет объединять маленькие пораженные области в несколько больших, это позволяет определить степень поражения легочной ткани визуально и облегчает последующий программный анализ снимков.

2.2. Алгоритм GLCM

GLCM (Grey – Level Co-occurrence Matrix) – матрица совпадений градаций серого была предложена в 1973 году Робертом Хараликом. Она представляет собой таблицу, которая показывает частоту появления различных комбинаций оттенков серого на изображении.

Алгоритм вычисления текстурных признаков по методу GLCM представляет собой построение функции распределения вероятности совместных событий, каждое из которых является характеристикой уровня яркости для пары пикселей, отстоящих друг от друга на некотором расстоянии и угле.

Первоначально рассчитываются статистические данные, которые принимают отклонение без учета соседствующих пикселей. Затем рассчитываются отношения между парами таких соседних пикселей в исходном изображении.

Вокруг каждого пикселя изображения выделяются соседние пиксели, расположенные под углами 0, 45, 90, 135 градусов. Затем строится функция распределения вероятности по формуле 1.

$$P_{d,\theta}(i, j) = \frac{1}{M} \sum_x^{n_x} \sum_y^{n_y} \delta(I(x, y) = i, I(x + \Delta x, y + \Delta y) = j), \quad (1)$$

, где $\Delta x, \Delta y$ – относительные смещения в соответствии с расстоянием d и углом θ , а M – это количество соседних пар пикселей.

На рисунке 18 представлена схема относительных смещений Δx , Δy .

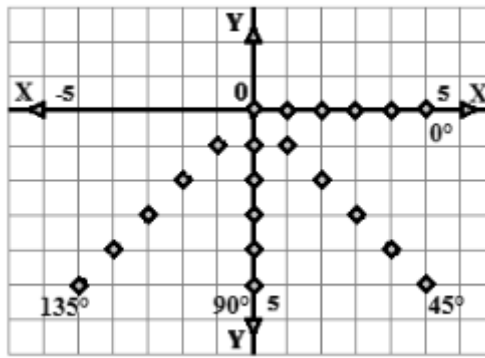


Рис 11. Относительные смещения между пикселями

Рассмотрим пример построения матрицы градаций серого на примере изображения:

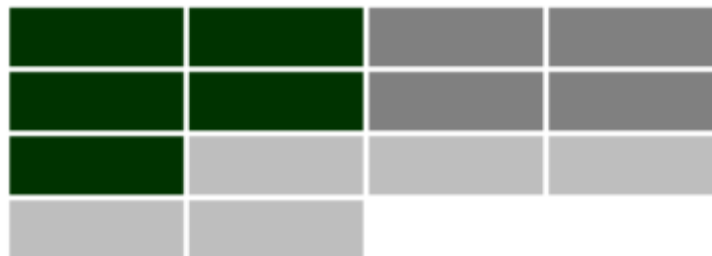


Рис 12. Исходное изображение

Тогда матрица градаций серого будет иметь вид:

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Примем расстояние равным одному пикселю. При рассмотрении большого изображения целесообразно принять расстояние большим, что никак не скажется на методе расчёта. Результатом расчёта станет матрица вида:

0	0,0	0,1	0,2	0,3
1	1,0	1,1	1,2	1,3
2	2,0	2,1	2,2	2,3
3	3,0	3,1	3,2	3,3

Каждая ячейка заполнена количеством комбинаций 0,0 т.е. количеством попаданий в эту область пикселя с уровнем яркости 0 и соседствующего с ним, имеющего яркость 0. Матрицы градаций уровня серого симметричны, таким

образом в результирующую таблицу входят главная диагональ и все ячейки что выше нее.

2.3. Краткие выводы по разделу.

В данном разделе был предложен алгоритм сегментации изображений, который позволит выделить пораженные области легкого.

Также предложен алгоритм текстурного анализа изображения.

Раздел 3. Разработка ПО на основании разработанных алгоритмов

3.1. Выбор средств разработки

Эффективность разработки во многом зависит не только от разработчиков, но и от правильного выбора средств и технологий для разработки. Дипломная работа подразумевает разработку алгоритмического программного обеспечения для определения степени патологии легкого при пневмонии на основании снимков компьютерной томографии. Разнообразие платформ и технологий разработки, представленных на рынке, заставляет задуматься над выбором.

Выбор языка разработки обуславливается различными факторами, такими как уровень языка, степень его проработки и многое другое, а также учитывая цели разработки.

Для реализации разработки был выбран язык C#. На данный момент язык программирования C# один из самых мощных языков, по сравнению с другими - относительно новый объектно-ориентированный язык, разработанный компанией Microsoft. Первый релиз данного языка вышел вместе с релизом Microsoft Visual Studio .Net. Текущей версией C# является версия 7.0, вышедшая 7 марта 2017 года вместе с Visual Studio 2017. Однако при написании данной дипломной работы была использована более ранняя версия языка.

C# относится к языкам с C-подобным синтаксисом, и частично основан на языках C++ и Java. Он поддерживает полиморфизм, перегрузку операторов, наследование, статическую типизацию, а также атрибуты, делегаты, свойства, события и т.д.

Объектно-ориентированный подход позволят решить проблемы по разработке крупных, но в тоже время гибких и масштабируемых приложений.

В качестве сравнения рассмотрим 2 среды разработки, которые поддерживают язык разработки C# - Microsoft Visual Studio и SharpDevelop.

Компания Microsoft является разработчиком пакета продуктов Microsoft Visual Studio, которые включают в себя интегрированную среду разработки. Microsoft Visual Studio позволяют разрабатывать консольные приложения,

приложения с графическим интерфейсом, веб-сайты, веб-приложения, веб-службы для всех платформ, поддерживаемых продуктами компании Microsoft, в том числе Windows, Windows Mobile, Windows CE, .NET Framework, Silverlight и другие.

Visual Studio включает в себя редактор исходного кода и возможность простейшего рефакторинга кода. Встроенный отладчик может работать как отладчик уровня исходного кода, так и как отладчик машинного уровня. Остальные встраиваемые инструменты включают в себя редактор форм для упрощения создания графического интерфейса приложения, веб-редактор, дизайнер классов и дизайнер схемы базы данных. Visual Studio позволяет создавать и подключать сторонние дополнения (плагины) для расширения функциональности практически на каждом уровне, включая добавление поддержки систем контроля версий исходного кода, добавление новых наборов инструментов или инструментов для прочих аспектов процесса разработки программного обеспечения.

Интерфейс среды Visual Studio 2015 построен по расширяемому и управляемому принципу. Набором пунктов GUI (глобальный уникальный идентификатор объекта) и способами их визуализации можно управлять. Интерфейс пользователя содержит большое число полезных инструментов и возможность обращения к внешним инструментам, набор которых также расширяем. В целом пользовательский интерфейс Visual Studio 2015 можно охарактеризовать как современный, гибкий, расширяемый и удобный. По своей сути и назначению, среда Visual Studio открыта для расширения новыми языками и видами проектов.

SharpDevelop – свободная среда разработки для C#, Visual Basic, IronPython, IronRuby, F#, C++. Обычно используется как альтернатива Visual Studio .NET.

Интерфейс SharpDevelop весьма похож на интерфейс Visual Studio. SharpDevelop может легко подключать дополнительные фреймворки, например, для работы с базами данных (Entity Framework). В числе достоинств данного

продукта – то, что SharpDevelop позволяет открывать проекты, созданные в Visual Studio. Среда имеет механизм загрузки пакетов библиотек и компонентов – NuGet, который позволяет легко интегрировать нужные библиотеки в проект.

SharpDevelop имеет возможность конвертировать приложение C# в приложение на Ruby, Python или VB.NET. Также содержит отладчик кода, визуальный редактор форм, подсветку синтаксиса, интегрированную поддержку NUnit, MbUnit и NCover и т.д.

SharpDevelop довольно мощный продукт, он обладает богатым функционалом, быстро работает, занимает мало места.

Microsoft Visual Studio 2015 – это среда разработки, которая позволяет разрабатывать сложные, комплексные, гибкие и масштабируемые проекты благодаря мощному инструментарию и интегрированной поддержке разработки через тестирование и новые инструменты отладки ПО для многоядерных процессоров. Подобный подход позволяет качественно и быстро находить и устранять дефекты, обеспечивая создание высококачественного решения. Visual Studio можно использовать для реализации идей и решений для широкого спектра платформ, включая Windows, Windows Server, веб-среду, облачную среду, Office и SharePoint и многое другое.

В ходе сравнения средой нашей разработки был выбран продукт Microsoft Visual Studio 2015, так как всесторонний набор инструментов и служб этой среды разработки гораздо больше, чем инструментов SharpDevelop.

RapidMiner – программная платформа, разработанная одноименной компанией в 2011 году и призванная избавить компании от необходимости разрабатывать собственную аналитическую систему. RapidMiner написан на языке программирования Java. Она обеспечивает интегрированную среду для машинного обучения и поддерживает все его этапы, в том числе подготовку данных, визуализацию результатов, проверку и оптимизацию модели. Платформа снабжена широким набором операторов, которые решают задачи получения и обработки информации из разнообразных источников (файлы, базы данных и т.п.).

По данным Bloor Research, RapidMiner обеспечивает 99% расширенного аналитического решения с помощью шаблонных структур, которые уменьшают количество ошибок, почти устраняя необходимость писать код. RapidMiner предоставляет процедуры интеллектуального анализа данных и машинного обучения, включая: предварительную обработку данных, визуализацию, интеллектуальную аналитику, статистическое моделирование, оценку и развертывание. Также предоставляет графический интерфейс для проектирования и выполнения аналитических рабочих процессов.

Функциональность RapidMiner может быть дополнена дополнительными плагинами, которые доступны через RapidMiner Marketplace. RapidMiner Marketplace предоставляет платформу для разработчиков, чтобы создавать алгоритмы анализа данных и публиковать их для сообщества.

Версия 7.0 RapidMiner включает в себя обновленный пользовательский интерфейс и расширенные возможности по подготовке данных.

Данный программный продукт является Open Source решением и предоставляется бесплатно с небольшими ограничениями, такими как 1 логический процессор и работа с 10000 строк данных. Однако существует возможность получить бесплатную образовательную лицензию, которая снимает упомянутые ограничения.

3.2. Реализация сегментации снимков КТ и ее эффективность в зависимости от режимов отображения

Снимки компьютерной томографии разнятся не только в зависимости от типа заболевания, но и от способа представления этого заболевания. Различные режимы отображения снимков КТ позволяют рассматривать снимки пациента более комплексно, таким образом, можно сделать вывод о необходимости различать результаты анализа в зависимости от этих режимов. Определение оптимальных режимов для индивидуальных особенностей заболевания при сегментации становится острой необходимостью.

Все снимки компьютерной томографии пациентов хранятся в специализированном формате DICOM (Digital Imaging and Communications in

Medicine). В данном формате медицинских изображений в каждый файл помимо собственно картинки «вшита» другая информация — имя пациента, его возраст, номер снимка, обозначения режимов сканирования и т.п.

Использование специализированного формата позволяет работать с файлами в связи с сопутствующими данными пациента, что для медицины является важной частью диагностирования и лечения.

Для открытия и просмотра подобных снимков необходимо специальное программное обеспечение. Также в формате DICOM невозможно проводить компьютерный анализ снимков. Для открытия и последующего сохранения снимков компьютерной томографии используется DICOMViewer.

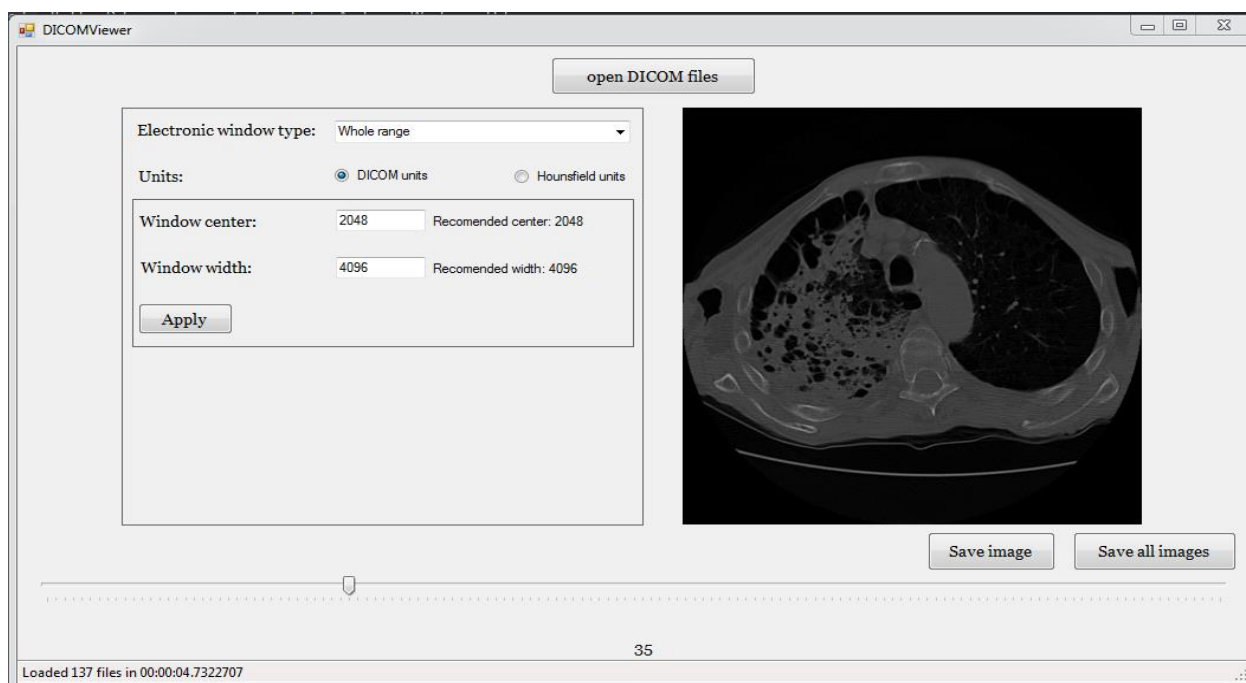


Рис 13. Пример отображения снимков КТ

Данное программное обеспечение позволяет открывать файлы снимков и отображать различные режимы проецирования легких, так называемые окнами.

Режимы или «окна» отображения формируются за счет различий в рентгеновской плотности тканей в организме. Существует шкала Хаунсфилда показывающая отличие рентгеновской плотности тканей от нулевой, за которую принята дистиллированная вода. Ниже приведена таблица средних показателей.

Таблица 2. Средние денситометрические показатели

Вещество	HU
----------	----

Воздух	-1000
Жир	-120
Вода	0
Мягкие ткани	+40
Кости	+400 и выше

Таким образом, существуют «окна»: легочные, плевральные, мягкотканые, костные и общие. Ниже приведены примеры отображения одного среза в различных режимах отображения (рис 2,3).

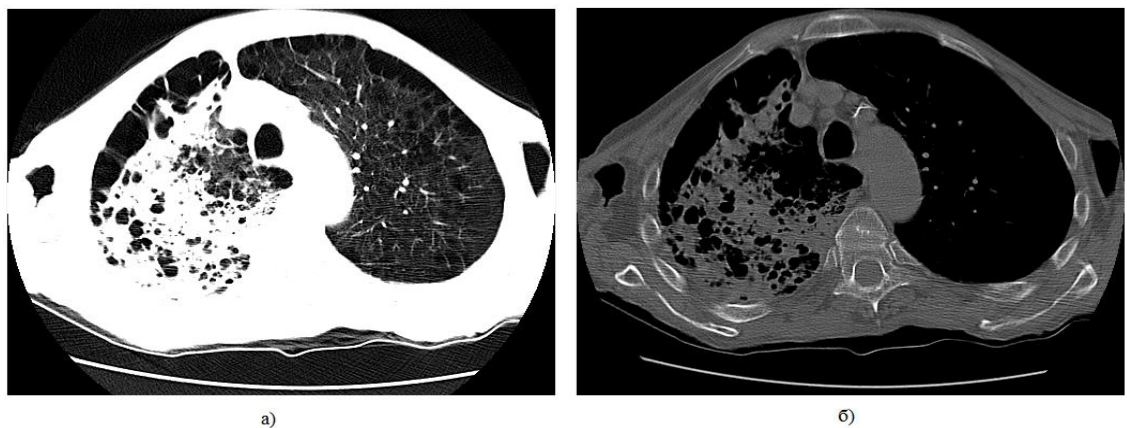


Рис 14. Пример отображения снимка в легочном режиме (а) и костном (б).

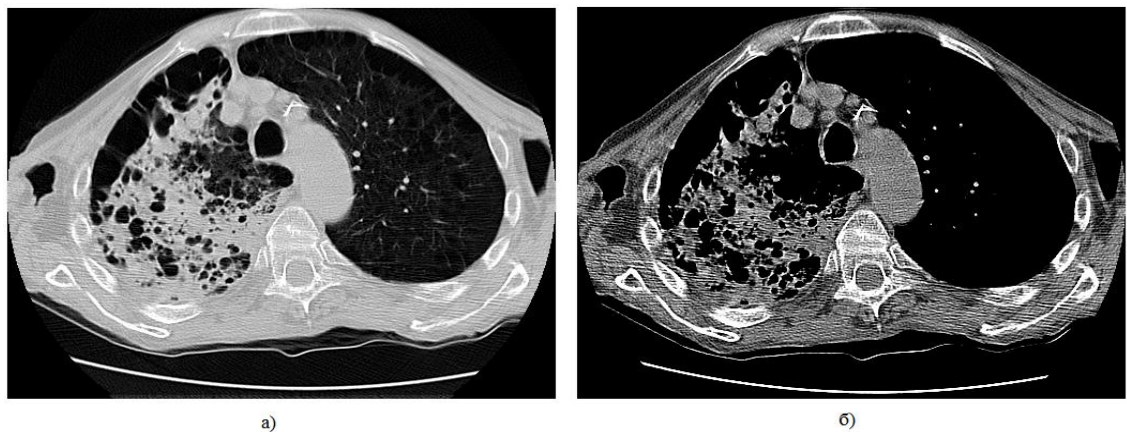


Рис 15. Пример отображения снимка в плевральном (а) и мягкотканом (б) режиме.

В зависимости от режима отображения различные области легкого воспаленные сегменты выглядят иначе и приобретают другой цвет, таким образом, можно выделить те режимы, при которых поврежденные ткани легкого видны наиболее четко.

Однако невозможно определить четкость изображения невооруженным взглядом, поэтому для облегчения дальнейшей работы со снимками компьютерной томографии производится сегментация изображения, благодаря которой пораженные области приобретают более ярко выраженные цвет и форму, а посторонние шумы, такие как ложе, одежда и окружающий воздух затемняются.

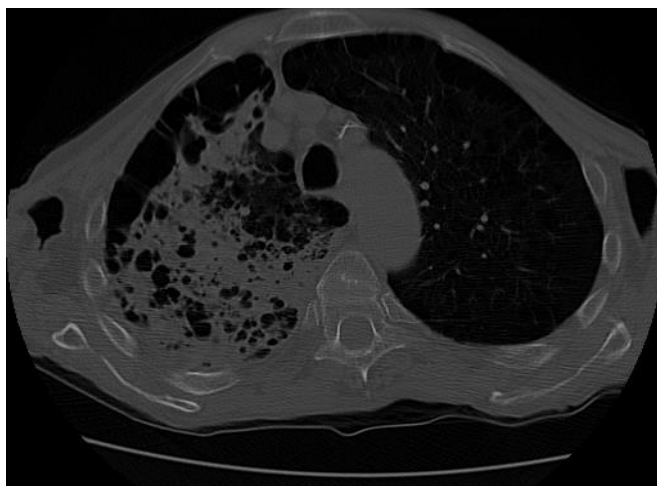


Рис 16. Срез легкого в общем режиме

На рисунке 3 явно видны ложе томографа и слои одежды. При определении пораженных областей эти шумы будут создавать ложные сегменты. Также на рисунке видны темные области, т.е. воздух. Таким образом, темные области внутри легкого представляют собой здоровую ткань, содержащую воздух. Для облегчения поиска пораженных областей легкого необходимо исключить также и воздух окружающий тело.

На рисунках ниже представлены результаты обработки и сегментации снимков КТ различных режимов отображения с помощью программного обеспечения. Для обработки изображений ниже были использованы изображения одного среза томографии пациента рассмотренные выше при сохранении снимков в различных режимах отображения (рис 2,3).

На рисунках 7,8 явно видно, что из исходного снимка удалено большее количество помех, таких как ложе и слои одежды, а также пораженные и здоровые области легкого приобрели более четкое цветовое разделение.

Однако даже визуальный осмотр явно показывает, что в костном и мягкотканом режимах пораженные области видны намного хуже, таким образом, можно сделать вывод, что легочный и плевральный режим наиболее подходящие режимы для последующего анализа снимков компьютерной томографии.

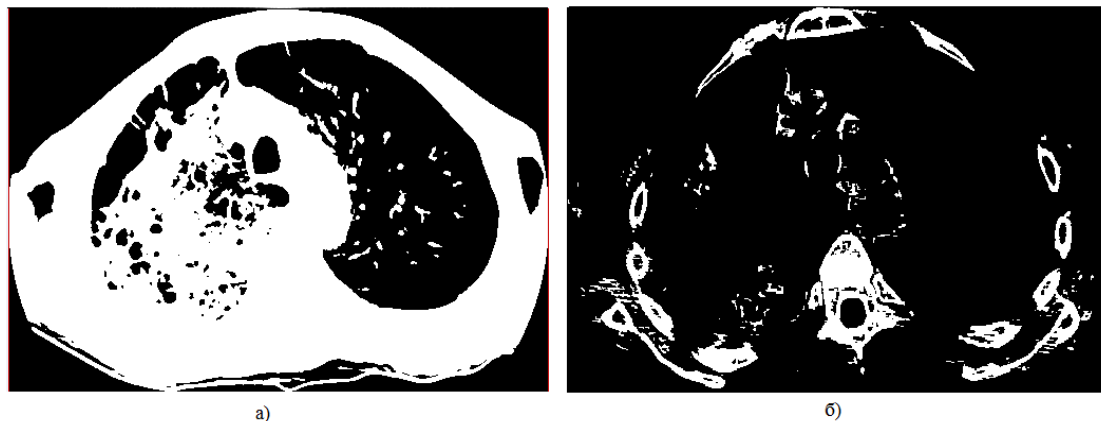


Рис 17. Результат сегментации снимков в легочном (а) и костном (б) режиме

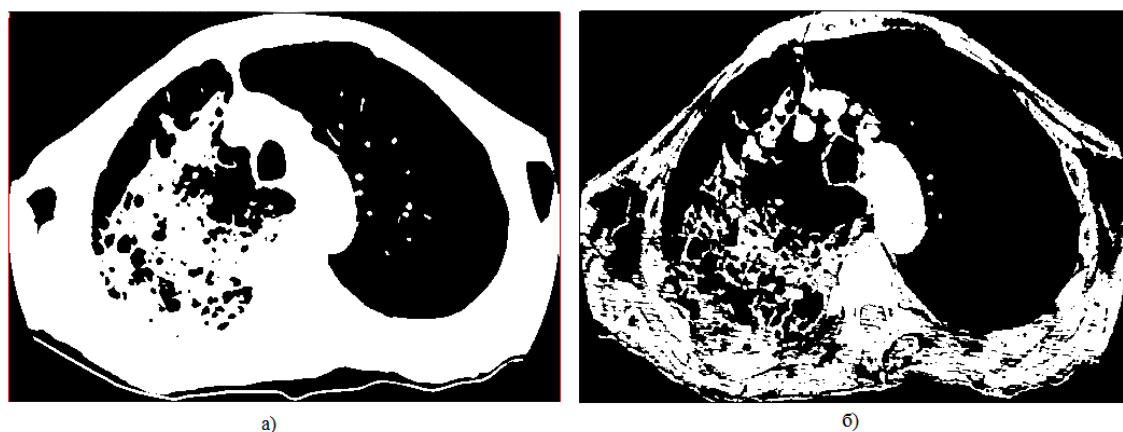


Рис 18. Результат сегментации снимков в плевральном (а) и мягкотканом (б) режиме

3.3. Реализация GLCM

В рамках диссертационной работы было разработано программное обеспечение, предназначенное для определения патологии на снимках КТ у больных пневмонией. Оно было реализовано на языке C# и включает в себя рассмотренный алгоритм выделения текстурных признаков GLCM, также позволяет задавать исходные параметры анализа и предоставляет результативную сводную таблицу градаций серого.

Графический интерфейс пользователя представлен на рисунке 18.

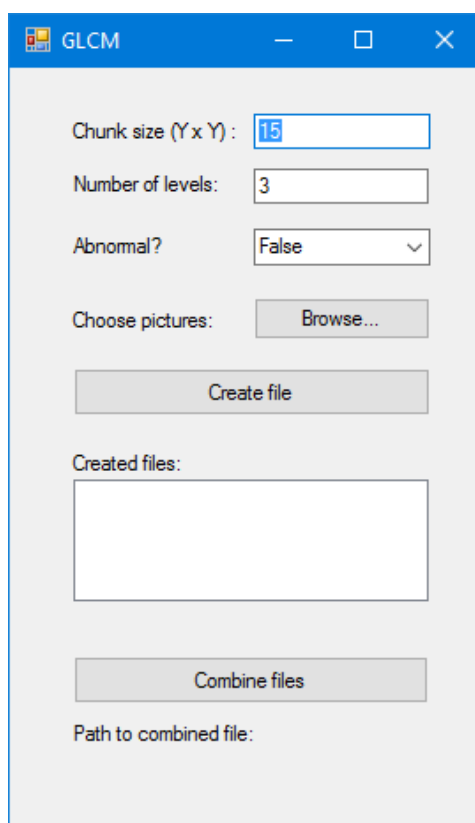


Рис 19. Графический интерфейс

В поле «Chunk size» вводится размер квадратного окна, на который разделяется исходное изображение. Поле «Numbers of levels» предназначено для ввода количества уровней серого.

В приложении можно выделить 2 функциональных блока:

1. Блок работы с файлами предназначен для загрузки и сохранения данных. Возможна загрузка изображений в формате BMP. Результат анализа представляется в виде сводной таблицы уровней градации серого и сохраняется на локальном хранилище.
2. Блок анализа текстурных признаков предназначен для обработки изображений посредством алгоритма GLCM.

После загрузки изображений ПО проводит анализ на основании алгоритма GLCM и формирует сводную таблицу уровней градации серого, которую записывает в файл формата Excel.

Пример фрагмента сводной таблицы уровней градации серого представлен на рисунке 19.

0,0,0,6,29,0,0,0,0,3,27,0,0,0,0,10,29,0,0,0,0,10
0,0,0,5,22,0,0,0,0,2,21,0,0,0,0,8,24,0,0,0,0,7,20
0,0,0,3,16,0,0,0,0,2,13,0,0,0,0,6,18,0,0,0,0,5,15
0,0,0,2,10,0,0,0,0,2,7,0,0,0,0,4,12,0,0,0,0,4,8,0
0,0,0,2,5,0,0,0,0,2,3,0,0,0,0,4,5,0,0,0,0,4,3,0,1
0,0,0,1,2,0,0,0,0,1,1,0,0,0,0,4,1,0,0,0,0,3,0,0,1
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,1
0,1
0,1

Рис 20. Фрагмент результатов применения алгоритма GLCM

В качестве исходных данных для алгоритма GLCM были использованы фрагменты снимков легких больных пневмонией, а также фрагменты здоровых легких. Фрагменты были отобраны у трех пациентов.

Все фрагменты брались в двух экземплярах в легочном и плевральном режиме, чтобы установить точность определения патологии легкого в обоих режимах. Также для сравнения точности классификации в зависимости от режима отображения фрагменты выделялись на одном участке легкого.

Для определения наиболее эффективных входных параметров, была проведена классификация для нескольких размеров окон.

Результаты работы ПО позволяют определять поврежденные участки и выделять патологию при пневмонии.

3.4. Краткие выводы по разделу

Важную роль при разработке программного обеспечения играют средства разработки. Для разработки алгоритмического ПО для определения патологии легкого наилучшим выбором являются Microsoft Visual Studio 2015 в качестве среды разработки и язык разработки – С#. Также для тестирования и оценки точности модели классификации была использована платформа RapidMiner Studio версии 7.3.

На основании проведенной сегментации изображений в легочном, костном, жировом и плевральном режимах отображения был сделан вывод, что сегментация в легочном и плевральном режиме отображения показала более высокую точность.

Было создано программное обеспечение, предназначенное для анализа снимков компьютерной томографии для определения патологии легкого у больных пневмонией.

Раздел 4. Тестирование алгоритма и оценка качества модели

Для тестирования и оценки качества модели был построен классификатор с помощью платформы RapidMiner, который позволил проанализировать результаты работы разработанного программного обеспечения и визуализировать результаты анализа. При оценке качества модели использовались деревья решений.

4.1. Построение классификатора для оценки качества модели.

Результатом классификации стали процентные показатели точности работы модели по определению пораженных и здоровых участков.

На рисунке 20 представлен классификатор для оценки качества модели.

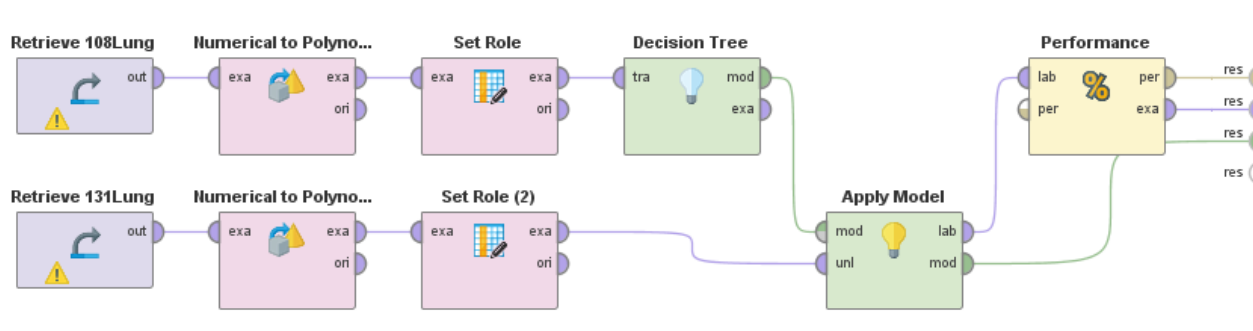


Рис 21. Классификатор для определения точности модели

Результатом работы классификатора стали:

1. Дерево решений, которое отображает выбор поврежденных участков легкого (рис 22).
2. Таблица, отображающая предсказания матрицы относительно того, является ли участок пораженным и степень уверенности в предсказании (рис 23).
3. Таблица, отражающая общую точность модели и точность предсказания пораженных и здоровых участков (рис 21).

accuracy: 60.76%

	true 1	true 0	class precision
pred. 1	35264	23206	60.31%
pred. 0	350	1215	77.64%
class recall	99.02%	4.98%	

Рис 22. Пример таблицы общей точность модели

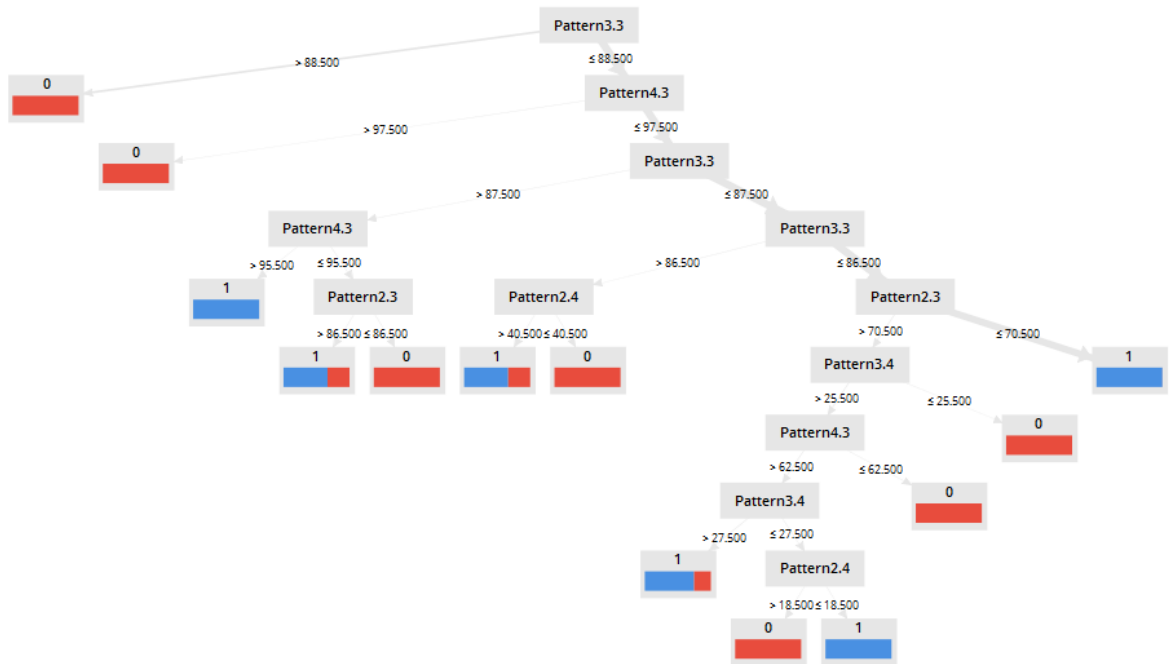


Рис 23. Пример дерева решений

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4
1	1	1	0.994	0.006	0	0	0	35	2
2	1	1	0.994	0.006	0	0	0	31	2
3	1	1	0.994	0.006	0	0	0	27	2
4	1	1	0.994	0.006	0	0	0	22	1
5	1	1	0.994	0.006	0	0	0	19	0
6	1	1	0.994	0.006	0	0	0	16	0
7	1	1	0.994	0.006	0	0	0	13	0
8	1	1	0.994	0.006	0	0	0	11	0
9	1	1	0.994	0.006	0	0	0	10	0
10	1	1	0.994	0.006	0	0	0	8	0
11	1	1	0.994	0.006	0	0	0	6	0
12	1	1	0.994	0.006	0	0	0	4	0
13	1	1	0.994	0.006	0	0	0	3	0
14	1	1	0.994	0.006	0	0	0	3	0
15	1	1	0.994	0.006	0	0	0	2	0
16	1	1	0.994	0.006	0	0	0	2	0

Рис 24. Пример таблицы предсказания модели

Обучение модели проходило на трети всего набора данных. Для обучения были выбраны сводные таблицы пациента под номером 108 в легочном и плевральном режиме.

Проверку модель проходила на двух других пациентах под номерами 131, 254, фрагменты легких которых также были отобраны в легочном и плевральном режиме.

4.2. Результат оценки качества модели

В результате анализа данных двух пациентов была составлена таблица качества модели (таб.3).

Таблица 3. Оценка качества модели

	131	254
Легочный	60,76%	87,27%
Плевральный	80,72%	59,99%

Результаты классификации данных пациентов представлены в приложении В.

Классификация для различных размеров окон показала, что с возрастанием размера окна точность модели возрастает, однако при переходе порога 9×9 пикселей эффективность работы модели уменьшается. Таким образом, можно сделать вывод, что размер 9×9 пикселей является оптимальным для метода GLCM при определении патологий легкого при пневмонии (таб 4).

Таблица 4. Точность модели в зависимости от размера окна

	3×3	5×5	9×9	15×15	18×18
Легочный	60,21%	74,36%	90,39%	83,37%	71,28%
Плевральный	54,39%	63,41%	78,16%	74%	68,52%

Также исследования показали, что при классификации наборов данных в легочном режиме модель имеет более высокую точность, чем в плевральном режиме.

4.3. Краткие выводы по разделу

Тестирование модели было проведено на основании выборок фрагментов легких двух пациентов в легочном и плевральном режимах отображения, а также для пяти размеров рассматриваемых окон в обоих режимах отображения.

По результатам тестирования был сделан вывод о высокой точности работы классификатора в легочном режиме и с размером окна 9×9 пикселей.

В целом классификатор показал высокую точность определения пораженных и здоровых областей легкого.

Раздел 5. Социальная ответственность

Введение

Социальная ответственность подразумевает под собой обеспечение безопасности охраны труда и окружающей среды, которое включает в себя охрану здоровья работников, обеспечение безопасных условий работы, защита от профессиональных заболеваний и производственного травматизма.

Для регламентирования приведенных выше аспектов существует международный стандарт ICCSR26000:2011 «Социальная ответственность организации», который направлен на принятие решений, позволяющих избежать несчастные случаи на производстве и позволяет снизить негативное воздействие на окружающую среду.

Разработка раздела, посвященного социальной ответственности организации, является обязательной частью процесса научно-исследовательской деятельности и написания дипломной работы. Для полного и комплексного изучения необходимо рассмотреть несколько обязательных составляющих, таких как, мероприятия для обеспечения безопасности, законодательное регулирование проектных решений, безопасность в чрезвычайных ситуациях, а также техногенная и региональная безопасность.

В данной научно-исследовательской работе произведена разработка алгоритмической модели для определения степени патологии легкого при пневмонии и предполагает непосредственную работу с персональным компьютером. Поэтому определение оптимальных условий труда является неотъемлемым критерием безопасности.

5.1. Техногенная безопасность

5.1.1. Недостаточная освещенность рабочей зоны

От степени освещенности помещения зависит не только здоровье глаз, но и общая работоспособность человека. Недостаточность освещения приводит к ослаблению внимания, снижению четкости зрения, повышению утомляемости.

Требования по освещенности в различных помещениях должны отличаться, в зависимости от их назначения. Нормы освещенности помещений

учитываются в обязательном порядке в административных и жилых зданиях, образовательных и дошкольных учреждений, предприятия бытового обслуживания, общественного питания и т.д. Требования к освещению определяются санитарными правилами и нормами (СанПиН). Норма освещенности определяется в Люксах (Лк) – это 1 люмен на кв.м. Ниже приведены нормы освещенности для основных групп помещений.

Таблица 1. Нормы освещенности для основных групп помещений

Вид помещения	Норма освещенности согласно СНиП, Лк
Офис общего назначения	200-300
Офис большой площади со свободной планировкой	400
Офис, в котором осуществляются чертежные работы	500
Зал для конференций	200
Эскалаторы, лестницы	50-100
Холл, коридор	50-75
Архив	75
Кладовая	50

Освещения разделяется на естественное – освещение дневным светом, искусственное – применяется при отсутствии дневного освещения и совмещенное – применяется при недостаточности естественного освещения.

Источниками искусственного освещения служат как правило люминесцентные лампы, которые попарно равномерно располагают над рабочими поверхностями. Требования к освещенности в помещениях с компьютерами: при выполнении зрительных работ высокой точности общая освещенность должна составлять 300лк, а совмещенная - 750лк; аналогичные требования при выполнении работ средней точности - 200 и 300лк соответственно.

5.1.2. Отклонение показателей микроклимата

В соответствии с санитарными правилами и нормами микроклимат помещения измеряется с помощью нескольких показателей, в том числе температура воздуха, температура поверхностей, относительная влажность воздуха, интенсивность теплового облучения. Данные показатели могут варьироваться в зависимости от назначения помещения и времени года.

Компьютеры в процессе работы могут повышать температуру помещения и снижать относительную влажность. В санитарных нормах - СанПиН 2.2.4.548 – 96 установлены величины параметров микроклимата, создающие комфортные условия.

Таблица 1. Параметры микроклимата для помещений с компьютерами

Период	Параметр микроклимата	Величина
Холодный	Температура воздуха в помещении	22...24°C
	Относительная влажность	40...60%
	Скорость движения воздуха	до 0,1м/с
Теплый	Температура воздуха в помещении	23...25°C
	Относительная влажность	40...60%
	Скорость движения воздуха	0,1...0,2м/с

Таблица 2. Нормы подачи свежего воздуха в помещения с компьютерами

Характеристика помещения	Объемный расход подаваемого в помещение
Объем до 20 м ³ на человека	Не менее 30
20...40 м ³ на человека	Не менее 20
Более 40 м ³ на человека	Естественная вентиляция

Объем рабочих помещений не должен быть меньше 19,5 м³/человека с учетом максимального числа одновременно работающих в смену.

5.1.3. Повышенный уровень шума на рабочем месте

Повышенный уровень шума не только вызывает дискомфорт работников, но и может привести к нарушению слуха, сердечно-сосудистым заболеваниям, а также вызывать головные боли, понижение аппетита, снижения памяти и т.д.

Стены, облицованные звукопоглощающими материалами, позволяют снизить уровень шума в помещении с компьютерами.

Таблица 3. Предельные уровни звука, дБ, на рабочих местах.

Категория напряженности труда	Категория тяжести труда			
	I. Легкая	II. Средняя	III. Тяжелая	IV. Очень тяжелая
I. Мало напряженный	8 0 дБ	80 дБ	75 дБ	75 дБ
II. Умеренно напряженный	7 0 дБ	70 дБ	65 дБ	65 дБ
III. Напряженный	6 0 дБ	60 дБ	-	-
IV. Очень напряженный	5 0 дБ	50 дБ	-	-

5.1.4. Повышенный уровень электромагнитных излучений

Электромагнитное излучения от персонального компьютера представляет собой проблему по нескольким причинам:

- Персональный компьютер имеет 2 источника излучения (системный блок и монитор).
- Маленькое расстояние между пользователем и ПК
- Длительное время воздействия.

Для персонального компьютера допускается рентгеновское излучение, не превышающее 10мкбэр/ч, и ультрафиолетовое и инфракрасное излучение – в пределах 10-100мВт/м².

Для снижения воздействия этих видов излучения рекомендуется применять мониторы с пониженным уровнем излучения (MPR-II, TCO-92, TCO-99), устанавливать защитные экраны, а также соблюдать регламентированные режимы труда и отдыха.

Таблица 4. Допустимые значения параметров неионизирующих электромагнитных излучений (в соответствии с СанПиН 2.2.2.542-96)

Наименование параметра	Допустимые
Напряженность электрической составляющей электромагнитного поля на расстоянии 50см от поверхности	10В/м
Напряженность магнитной составляющей электромагнитного поля на расстоянии 50см от поверхности видеомонитора	0,3А/м
Напряженность электростатического поля не должна превышать:	
для взрослых пользователей	20кВ/м
для детей дошкольных учреждений и учащихся средних специальных и высших учебных заведений	15кВ/м

5.2. Электробезопасность

На корпусах мониторов, системного блока, мыши и клавиатуры собираются токи статического электричества, которые приводят к возникновению разрядов при прикосновении. Возникновение заряда с большим потенциалом может породит электрическое поле с повышенной напряжённостью, которое может быть опасно для человека.

Данное поле может вывести из строя компьютер, но и при длительном нахождении нанести вред нервной и сердечно-сосудистых систем человека.

Использование нейтрализаторов статического электричества, которые создают положительные и отрицательные ионы вблизи диэлектрического наэлектризованного объекта, позволяет избежать выше перечисленных проблем.

5.3. Региональная безопасность

При производстве компьютерной техники используются различные материалы, которые потенциально могут нанести вред окружающей среды. Поэтому утилизация данной техники представляет собой сложный технологический процесс и были приняты жесткие законы, ограничивающие

бытовую утилизацию технических отходов. Утилизация люминесцентных ламп должна производиться коммунальными службами, которые занимаются вывозом специальных отходов, так как эти лампы содержат от 3 до 5 мг ртути каждая. В производстве энергосберегающих ламп также используется небольшое количество ртути, поэтому их утилизация должна осуществляться также специальными службами. Транспортировка ламп осуществляется организациями, которые специализируются на утилизации опасных отходов.

5.4. Организационные мероприятия обеспечения безопасности

Одно рабочее место ПЭВМ должно занимать площадь не менее 6 м². Расстояние между рабочими столами и мониторами должны учитываться при размещении рабочих мест, а также удаленность от силовых кабелей и вводов, высоковольтных трансформаторов, технологического оборудования, создающего помехи в работе ПЭВМ.

Показатели микроклимата помещений должны соответствовать санитарным правилам и нормам микроклимата производственных помещений.

Использование диффузно-отражающих материалов с коэффициентами отражения для потолка 0,7-0,8; для стен 0,5-0,6; для пола 0,3-0,5 необходимо для внутренней отделки помещений с компьютерами.

В помещении должны находиться углекислотный огнетушитель, аптечка первой медицинской помощи.

Согласно СанПиНу 2.2.2.542-96 при 8-ми часовой рабочей смене на ВДТ и ПЭВМ перерывы в работе должны составлять от 10 до 20 минут каждые два часа работы.

5.5. Особенности законодательного регулирования проектных решений

Продолжительность рабочего дня должна составлять 40 часов в неделю согласно ст. 91 ТК РФ. Сокращение продолжительности рабочего дня – не более 35 часов в неделю возможно для инвалидов I или II группы, в соответствии с ст. 92 ТК РФ. Гибкий режим работы возможен при согласовании с руководством.

Беременным женщинам в связи с медицинским заключением снижают норму выработки по ст. 254 ТК РФ с сохранением средней заработной платы.

5.6. Безопасность в чрезвычайных ситуациях

Пожарная безопасность — состояние защищённости личности, имущества, общества и государства от пожаров. Пожарная безопасность обеспечивается системой пожарной защиты и системой предотвращения пожара.

Опасный фактор пожара (ОФП):

- пламя и искры;
- тепловой поток;
- повышенная температура окружающей среды;
- повышенная концентрация токсичных продуктов горения и термического разложения;
- пониженная концентрация [кислорода](#);
- снижение видимости в дыму.

Действия при пожаре в здании:

- При наличии телефона, «112» или «01» сообщить о пожаре и своем местоположении;
- Не входить в места с высокой концентрацией дыма и видимостью менее, чем 10 метров.
- Покинуть помещение, используя запасные и основные пути эвакуации;
- Попутно отключить электроэнергию;
- Передвигаться к выходу на четвереньках, при этом закрывая рот и нос подручными средствами защиты;
- Плотно закрыть дверь при выходе;

Если дым и пламя в соседних помещениях не позволяет выйти наружу:

- Стараться не поддаваться панике;
- Проверить возможности спуститься по пожарной лестнице или выйти на крышу;

- При отсутствии возможности эвакуироваться для защиты от дыма и тепла необходимо как можно надёжней загерметизировать своё помещение:
- Плотно закрыть двери, окна и форточки, заткнуть щели изнутри, используя при этом любую, желательно мокрую, ткань;
- При наличии воды, постоянно смачивать двери и пол.

Пожарная безопасность в ТПУ обеспечивается в соответствии с требованиями ФЗ пожарной безопасности № 69-ФЗ от 21.12.1994 г., правилами противопожарного режима в РФ.

Рабочее место располагается в кибернетическом центре ТПУ 212 аудитория. На рисунке 1 представлен план эвакуации второго этажа кибернетического центра.



Рисунок 25. План эвакуации.

Раздел 6. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Введение

В настоящий момент успех научного исследования определяется не только преобладанием параметров открытия над предыдущими разработками, но и тем будет ли продукт коммерчески успешен или востребован на рынке. Также важными вопросами для исследователя являются такие вопросы, как – какова будет цена продукта, каков необходимый бюджет научного проекта и т.д.

Проектирование и создание разработок, способных конкурировать на рынке, а также отвечающих требованиям в области ресурсоэффективности и ресурсосбережения является целью раздела «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение».

В рамках данной научной работы исследуется разработка алгоритмической модели определения степени патологии легкого при пневмонии на основании снимков компьютерной томографии.

6.1. Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения

6.1.1. Потенциальные потребители результатов исследования

Основными потребителями услуг приложения для определения степени патологии легкого при пневмонии являются врачи и больницы, в том числе диспансеры, специализирующиеся на легочных заболеваниях. Сведения о сегментах рынка продаж услуг приложений для обработки снимков компьютерной томографии представлены в таблице 1.

Согласно карте сегментирования рынка, можно сделать вывод о полностью свободном сегменте. На данный момент времени не существует отечественных разработок, позволяющие определять степень патологии легкого при пневмонии. Целевым сегментом будет распространение приложения на отечественном рынке.

Таблица 1. Карта сегментирования рынка

		Виды заболеваний	
		Пневмония	Другие легочные заболевания
Виды компаний	Зарубежные компании		
	Отечественные компании		



Tuberculosis.by

6.1.2. Анализ конкурентных технических решений

Для анализа конкурирующих разработок была построена оценочная карта для сравнения конкурентных разработок. Была рассмотрена конкурентоспособность разработки относительно уже существующей компании - Tuberculosis(к), предоставляющая услуги по определению степени патологии при пневмонии. Анализ конкурентных технических решений определяется по формуле:

$$K = \sum B_i * B_i$$

где K – конкурентоспособность научной разработки или конкурента; B_i – вес показателя (в долях единицы); B_i – балл i -го показателя.

Коэффициент конкурентоспособности предприятия:

$$k_{kc} = K_{\phi} / K_{к} = 4,4/4,1 = 1,07.$$

$k_{kc} > 1$, следовательно предприятие конкурентоспособно.

Таблица 2. Оценочная карта для сравнения конкурентных разработок

Критерии оценки	Вес критерия	Баллы		Конкурентоспособность	
		Б _ф	Б _к	К _ф	К _к
1	2	3	4	5	6
1.Точность определения степени патологии	0,2	4	4	0,8	0,8
2. Скорость обработки снимков	0,1	4	3	0,4	0,3
3.Интуитивная понятность интерфейса	0,1	5	4	0,5	0,4
4. Устойчивость к сбоям при загрузке изображений	0,1	4	5	0,4	0,5
5. Поддержка нескольких режимов отображения	0,1	5	5	0,5	0,5
6.Потребность в ресурсах памяти	0,1	3	4	0,3	0,4
7. Устойчивость к отказам серверов	0,1	5	4	0,5	0,4
8. Цена	0,2	5	4	1	0,8
Итого	1			4,4	4,1

6.1.3. Технология QuaD

В данном разделе для оценки конкурентоспособности исследования была построена оценочная карта по технологии QuaD.

Таблица 3. Оценочная карта для сравнения конкурентных разработок

Критерии оценки	Вес критерия	Баллы	Максимальный балл	Относительное значение	Средневзвешенное значение
-----------------	--------------	-------	-------------------	------------------------	---------------------------

1	2	3	4	5	6
Показатели оценки качества разработки					
1.Точность определения степени патологии	0,2	70	100	0,7	0,14
2.Скорость обработки снимков	0,1	75	100	0,75	0,075
3.Интуитивная понятность интерфейса	0,1	90	100	0,9	0,09
4. Устойчивость к сбоям при загрузке изображений	0,1	60	100	0,6	0,06
5.Поддержка нескольких режимов отображения	0,1	75	100	0,75	0,075
6.Потребность в ресурсах памяти	0,1	60	100	0,6	0,06
7. Устойчивость к отказам серверов	0,1	80	100	0,8	0,08
8. Цена	0,2	80	100	0,8	0,16
Итого	1				0,74

$$P_{cp} = \sum V_i * B_i$$

где P_{cp} – средневзвешенное значение показателя качества и перспективности научной разработки; V_i – вес показателя (в долях единицы); B_i – средневзвешенное значение i -го показателя. Значение P_{cp} составило 74, что показывает перспективность разработки выше среднего.

6.1.4. SWOT-анализ

Посредством SWOT-анализа выявим внешние и внутренние факторы среды проекта, и выявим степень соответствия сильных и слабых сторон проекта внешним условиям окружающей среды.

Таблица 4. Матрица SWOT

	Сильные стороны	Слабые стороны
	С1. Низкая стоимость услуги С2. Высокая точность определения патологии С3. Квалифицированный персонал С4. Безопасность хранения данных С5.Отсутствие аналогов в РФ	Сл1.Узкопрофильность продукта Сл2.Недостаточная аппаратно-техническая база Сл3.Отсутствие опыта Сл4.Возможные сбои программного продукта Сл5.Отсутствие долгосрочных соглашений с клиентами

Возможности	В1. Привлечение крупных клиентов В2. Расширение аппаратно-технической базы В3. Привлечение зарубежных клиентов В4. Получение дополнительного финансирования В5. Выход на международные рынки	В1С1С2С3С4С5 – Привлечение большего количества клиентов, увеличение доходов В3С1С2С4 - Индивидуальный подход к клиенту В4С2С3С5- Участие в конкурсах финансирования проектов В5С1С2С3С4-Расширение бизнеса для международных рынков.	В2Сл2Сл4 - Приобретение нового оборудования В4Сл2 –Участие в конкурсах финансирования проектов.
	Угрозы	У1С5 – Снижение стоимости услуги У2С4 – Изменение политики компании У3С1С3 – Повышение стоимости услуги У5С4 – Улучшение оборудования	У1Сл1 – Доработка программного продукта У3Сл5 - Проведение рекламной политики У4Сл1Сл2Сл3Сл4Сл5 - Привлечение клиентов посредством скидок У5Сл2Сл4 – Расширение аппаратно-технической базы

Таблица 5. Интерактивная матрица проекта

Сильные стороны						
Возможности		С1	С2	С3	С4	С5
	В1	+	+	+	+	+
	В2	-	-	0	-	-
	В3	+	+	0	+	-
	В4	-	+	+	-	+
	В5	+	+	+	+	-

Направления реализации проекта: В1С1С2С3С4С5, В3С1С2С4, В4С2С3С5, В5С1С2С3С4.

Таблица 6. Интерактивная матрица проекта

Слабые стороны						
Возможности		Сл1	Сл2	Сл3	Сл4	Сл5
	В1	-	-	-	-	-
	В2	0	+	0	+	-
	В3	-	-	-	-	-
	В4	-	+	-	-	-
	В5	-	-	-	-	-

Направление реализации проекта: В2Сл2Сл4, В4Сл2.

Таблица 7. Интерактивная матрица проекта

Сильные стороны

Угрозы		C1	C2	C3	C4	C5
	У1	-	-	-	-	+
	У2	-	-	-	+	0
	У3	+	0	+	-	0
	У4	-	0	-	-	-
У5	0	-	-	+	0	

Направления реализации проекта: У1С5, У2С4, У3С1С3, У5С4.

Таблица 8. Интерактивная матрица проекта

Слабые стороны						
Угрозы		Сл1	Сл2	Сл3	Сл4	Сл5
	У1	+	-	-	-	-
	У2	-	-	-	-	-
	У3	-	-	-	-	+
	У4	+	+	+	+	+
У5	-	+	-	+	-	

Направления реализации проекта: У1Сл1, У3Сл5, У4Сл1Сл2Сл3Сл4Сл5,

7.2. Планирование научно-исследовательских работ

7.2.1 Структура работ в рамках научного исследования

Научное исследование было выполнено студентом-дипломником под руководством научного руководителя. Порядок и очередность выполнения работ представлены в таблице 9.

Таблица 9. Перечень этапов, работ и распределение обязанностей

Основные этапы	№ раб.	Содержание работ	Исполнитель
Разработка технического задания	1	Выбор темы	Студент, Руководитель
	2	Составление и утверждение технического задания	Студент, Руководитель
Анализ предметной области	3	Календарное планирование работ	Студент
	4	Подбор и изучение материалов по теме	Студент
	5	Анализ литературных источников	Студент, Руководитель
Основная	6	Проектирование структуры системы	Студент, Руководитель
	7	Проектирование и разработка базы данных	Студент
	8	Разработка интерфейса пользователя	Студент
	9	Разработка системы загрузки снимков КТ	Студент
	10	Разработка системы сегментации изображений	Студент
	11	Разработка системы определения патологии	Студент

	12	Тестирование и доработка системы	Студент, Руководитель
Заключительная	13	Написание пояснительной записки	Студент, Руководитель
	14	Подготовка и оформление дипломной презентации	Студент

7.2.2 Определение трудоемкости выполнения работ

Для расчета ожидаемого значения продолжительности работ $t_{ож}$ применяется две оценки: t_{min} и t_{max} (метод двух оценок).

$$t_{ож} = \frac{3 \cdot t_{min} + 2 \cdot t_{max}}{5},$$

где t_{min} – минимальная трудоемкость работ, чел/дн.; t_{max} – максимальная трудоемкость работ, чел/дн.

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях T_p , учитывающая параллельность выполнения работ несколькими исполнителями.

$$T_{pi} = \frac{t_{ожi}}{Ч_i},$$

где T_{pi} – продолжительность одной работы, раб. дн.; $t_{ожi}$ – ожидаемая трудоемкость выполнения одной работы, чел.-дн. $Ч_i$ – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни.

$$T_{ki} = T_{pi} * k_{кал}$$

где T_{ki} – продолжительность выполнения i -й работы в календарных днях;
 T_{pi} – продолжительность выполнения i -й работы в рабочих днях; $k_{кал}$ – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}}$$

где $T_{\text{кал}}$ – количество календарных дней в году; $T_{\text{вых}}$ – количество выходных дней в году; $T_{\text{пр}}$ – количество праздничных дней в году.

Для выполнения перечисленных в таблице 9 работ требуются специалисты: студент, научный руководитель. Результаты расчетов представлены в таблице 10.

Таблица 10. Временные показатели проведения научного исследования

№ Работы	Трудоёмкость работ									Исполнители			Длительность работ в рабочих днях p_i			Длительность работ в календарных днях k_i		
	t_{\min} , чел-дни			t_{\max} , чел-дни			$t_{ож\ i}$, чел-дни											
	И.1	И.2	И.3	И.1	И.2	И.3	И.1	И.2	И.3	И.1	И.2	И.3	И.1	И.2	И.3	И.1	И.2	И.3
Выбор темы, С, Р	2	2	2	4	4	4	2,8	2,8	2,8	2	2	2	1,4	1,4	1,4	2	2	2
Составление и утверждение технического задания, С, Р	1	2	3	2	4	5	1,4	2,8	3,2	2	2	2	0,7	1,4	1,6	1	2	2
Календарное планирование работ, С	1	1	1	2	2	2	1,4	1,4	1,4	1	1	1	1,4	1,4	1,4	2	2	2
Подбор и изучение материалов по теме, С	3	5	6	5	7	8	3,8	5,8	6,8	1	1	1	3,8	5,8	6,8	6	9	10
Анализ литературных источников, С, Р	7	5	8	10	8	11	8,2	6,2	9,2	2	2	2	4,1	3,1	4,6	6	5	7
Проектирование структуры системы, С, Р	3	2	5	5	5	9	3,8	3,2	6,6	2	2	2	1,9	1,6	3,3	3	2	5
Проектирование и разработка базы данных системы, С	7	9	10	10	12	14	8,2	10,2	10,8	1	1	1	8,2	10,2	10,8	12	15	16
Разработка интерфейса пользователя, С	5	6	8	10	12	14	7	8,4	10,4	1	1	1	7	8,4	10,4	10	12	15
Разработка системы загрузки снимков КТ, С	3	5	6	5	7	8	5	7,8	6,8	1	1	1	5	7,8	6,8	7	11	10
Разработка системы сегментации изображений, С	3	5	6	5	7	8	5	5,8	6,8	1	1	1	5	5,8	6,8	7	9	10
Разработка системы определения патологии, С	10	15	18	15	20	22	12	17	19,6	1	1	1	12	17	19,6	18	25	29
Тестирование и доработка системы, С, Р	2	5	7	5	9	10	3,2	8,4	8,2	2	2	2	1,6	4,2	4,1	2	6	6
Написание пояснительной записки, С, Р	4	7	10	7	10	14	5,2	8,2	11,5	2	2	2	2,6	4,1	5,75	4	6	9
Подготовка и оформление презентации дипломного проекта, С	2	4	3	4	5	6	2,8	4,4	4,2	1	1	1	2,8	4,4	4,2	4	7	6
Итого	53	73	93	89	112	135	69,8	92,4	108,3	-	-	-	57,5	76,6	87,55	84	113	129

$$K_{\text{кал}} = 365 / (365 - 118) = 1,48$$

По данным расчетам, программа будет разработана:

- в первом исполнении 84 дней
- во втором исполнении 113 дней
- в третьем исполнении 129 дней

Следовательно, можно сделать вывод, что в первом исполнении работы будет выполнена быстрее. На основе таблицы 10 построен календарный план-график, представленный на рисунке 1 и 2. График строился для максимального по длительности исполнения работ в рамках научно-исследовательского проекта с разбивкой по месяцам за период времени дипломирования.

6.2.3. Разработка графика проведения научного исследования

Для построения диаграммы Ганта определим минимальное по длительности исполнение. Это исполнение номер 3, и время его исполнения составит 129 дней. Календарный план-график для студента представлен на рисунке 1, а для руководителя – на рисунке 2.

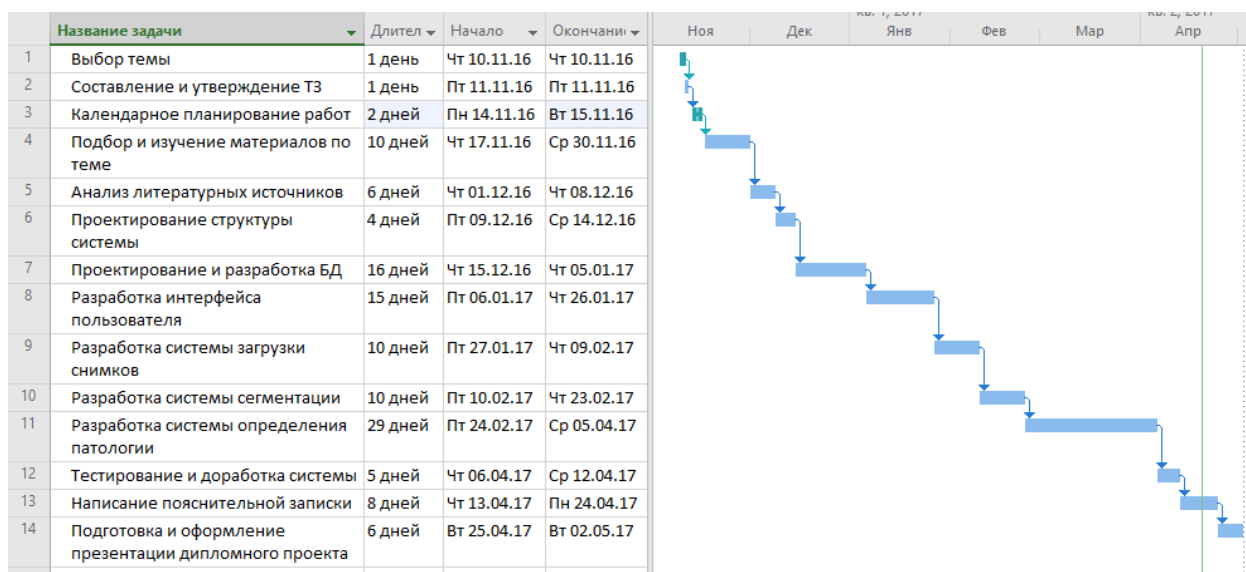


Рис 26. Календарный план-график (Студент)

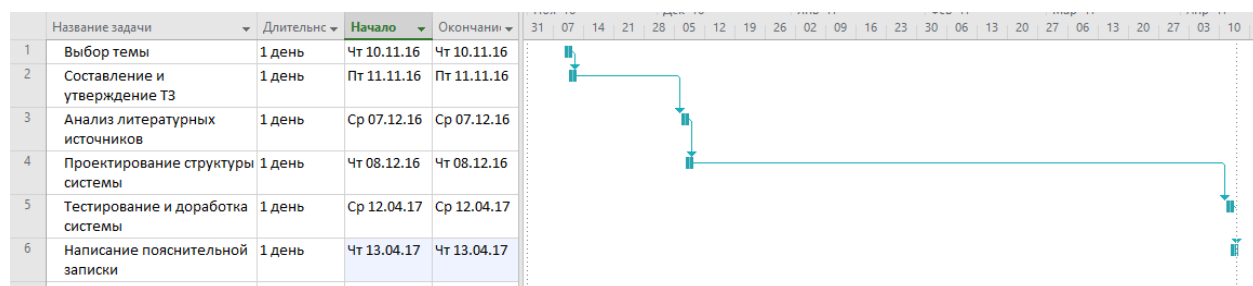


Рис 27. Календарный план-график (Руководитель)

6.2.4. Бюджет научно-технического исследования

При планировании бюджета НТИ должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. В процессе формирования бюджета НТИ используется следующая группировка затрат по статьям:

- материальные затраты НТИ;
- основная заработная плата исполнителей темы;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- накладные расходы.

6.2.4.1 Расчет материальных затрат НТИ

Расчет материальных затрат осуществляется по формуле 6:

$$Z_M = (1 + k_T) * \sum_{i=1}^m C_i * N_{расxi}$$

где m – количество видов материальных ресурсов, потребляемых при выполнении научного исследования; $N_{расxi}$ – количество материальных ресурсов i -го вида, планируемых к использованию при выполнении научного исследования (шт., кг, м, м² и т.д.); C_i – цена приобретения единицы i -го вида потребляемых материальных ресурсов (руб./шт., руб./кг, руб./м, руб./м² и т.д.); k_T – коэффициент, учитывающий транспортно-заготовительные расходы.

Для разработки данного продукта необходимы следующие материальные ресурсы:

- Системный блок
- Принтер
- Монитор

Расчет материальных затрат представлен в таблице 11.

Таблица 11. Материальные затраты

Наименование	Единица измерения	Количество			Цена за ед., руб.			Затраты на материалы, (З _м), руб.		
		Исп.1	Исп.2	Исп.3	Исп.1	Исп.2	Исп.3	Исп.1	Исп.2	Исп.3
Системный блок	Шт.	1	1	1	30499	30999	31999	35073	35648	36798
Принтер	Шт.	1	1	1	4599	7699	9499	5288	8853	10923
Монитор	Шт.	2	2	2	4199	5999	7499	9657	13797	17247
Сервер Depo Storm 3400 A2	Шт.	1	1	2	27924 6	27924 6	285995	279246	279246	657788
Итого								329264	337544	628486

6.2.4.2. Основная заработная плата исполнителей темы

Основная заработная плата рассчитывается по формуле:

$$C_{\text{осн/зп}} = \sum_{i=1}^n t_i * C_{\text{зп}_i}$$

где n – количество видов работ; t_i – затраты труда на выполнение i -го вида работ, в днях; $C_{\text{зп}_i}$ – среднедневная заработная плата работника, выполняющего i -ый вид работы, руб/день.

Среднедневная заработная плата рассчитывается по формуле:

$$C_{\text{зп}_i} = \frac{D * K * M_p}{F_0}$$

где D – месячный должностной оклад работника; K – коэффициент, учитывающий коэффициент по премиям и районный коэффициент ($K=1,3$); M_p – количество месяцев работы без отпуска в течение года; F_0 – действительный годовой фонд рабочего времени работника, в днях. При отпуске 28 дня $M_p=11,08$.

Результаты расчета действительного годового фонда проведены в таблице 12.

Таблица 12. Годовой фонд рабочего времени

Показатели рабочего времени	Руководитель	Студент
Календарное число дней	365	365
Количество нерабочих дней - выходные дни - праздничные дни	119	119
Потери рабочего времени - отпуск - невыходы по болезни	28	28
Действительный годовой фонд рабочего времени	218	218

Расчет затрат на основную заработную плату приведен в таблице 16. При этом затраты на оплату труда студента-дипломника определяются как оклад студента $D = 6976,22$ рублей, а оклад руководителя проекта составляет 23264,86. Коэффициент K , учитывающий коэффициент по премиям и районный коэффициент равен 1,3, а M_p равно 11,08.

Таблица 13. Затраты на основную заработную плату

Исполнители	Среднедневная заработная плата $C_{зн}$ (руб.)			Трудоемкость (t_i), чел-дни			Затраты на основную зарплату (руб.)		
	Исп.1	Исп.2	Исп.3	Исп.1	Исп.2	Исп.3	Исп.1	Исп.2	Исп.3
Руководитель	1530,2			7	9	8	10711,4	13771,8	12241,6
Студент	458,8			43	44	45	19728,4	20187,2	20646
Итого							30439,8	33959	32887,6

6.2.4.3. Дополнительная заработная плата

Дополнительная заработная плата включает заработную плату за не отработанное рабочее время, но гарантированную действующим законодательством.

Расчет дополнительной заработной платы ведется по формуле:

$$Z_{\text{доп}} = k_{\text{доп}} * Z_{\text{осн}}$$

где $k_{\text{доп}}$ – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,11 – 0,15).

Примем $k_{\text{доп}}$ равным 0,12. Результаты по расчетам дополнительной заработной платы сведены в таблицу 14.

Таблица 14. Затраты на дополнительную заработную плату

Исполнители	Основная зарплата(руб.)			Коэффициент дополнительной заработной платы ($k_{\text{доп}}$)	Дополнительная зарплата(руб.)		
	Исп.1	Исп.2	Исп.3		Исп.1	Исп.2	Исп.3
Руководитель	10711,4	13771,8	12241,6	0,12	1285,4	1652,6	1469
Студент	19728,4	20187,2	20646	0,12	2367,4	2422,5	2477,5
Итого					3652,8	4075,1	3946,5

6.2.4.4. Отчисления во внебюджетные фонды (страховые отчисления)

Величина отчислений во внебюджетные фонды определяется исходя из формулы 11:

$$Z_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}})$$

где $k_{\text{внеб}}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

На 2014 г. в соответствии с Федеральным законом от 24.07.2009 №212-ФЗ установлен размер страховых взносов равный 30,2%. На основании пункта 1 ст.58 закона №212-ФЗ для учреждений, осуществляющих образовательную и научную деятельность в 2014 году водится пониженная ставка – 27,1%.

Отчисления во внебюджетные фонды представлены в таблице 18.

Таблица 15. Отчисления во внебюджетные фонды

Исполнитель	Основная заработная плата, руб.			Дополнительная заработная плата, руб.		
	Исп. 1	Исп. 2	Исп. 3	Исп. 1	Исп. 2	Исп. 3
Руководитель проекта	10711,4	13771,8	12241,6	1285,4	1652,6	1469
Студент	19728,4	20187,2	20646	2367,4	2422,5	2477,5
Коэффициент отчислений во внебюджетные фонды	27,1%					
Итого						
Исполнение 1	43331,69					
Исполнение 2	48341,34					
Исполнение 3	46816,14					

6.2.4.5. Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии, почтовые и телеграфные расходы, размножение материалов и т.д. Их величина определяется по формуле:

$$Z_{\text{накл}} = (\text{сумма статей } 1 \div 5) * k_{\text{нр}}$$

где $k_{\text{нр}}$ – коэффициент, учитывающий накладные расходы.

Величина коэффициента накладных расходов равна 50%.

Исполнение 1: $(329264/5)*0,5 = 32926,4$

Исполнение 2: $(337544/5)*0,5 = 33754,4$

Исполнение 3: $(628486/5)*0,5 = 62848,6$

6.2.4.6. Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно-исследовательской работы (темы) является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции.

Определение бюджета затрат на научно-исследовательский проект по каждому варианту исполнения приведен в таблице 16.

Таблица 16. Расчет бюджета затрат НИИ

Наименование статьи	Сумма, руб.		
	Исп.1	Исп.2	Исп.3
1. Материальные затраты НИИ	329264	337544	628486
2. Затраты по основной заработной плате исполнителей темы	30439,8	33959	32887,6
3. Затраты по дополнительной заработной плате исполнителей темы	3652,8	4075,1	3946,5
4. Отчисления во внебюджетные фонды	9239,09	10307,24	9982,04
5. Накладные расходы	32926,4	33754,4	62848,6
6. Бюджет затрат НИИ	405522,09	419639,74	738150,74

Вывод: Основываясь на данных, полученных в пунктах 2.4.1 – 2.4.6, был рассчитан бюджет затрат научно-исследовательской работы. Наиболее низким по себестоимости оказался проект первого исполнения, затраты на его полную реализацию составляют 405522,09 рублей.

6.3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Интегральный финансовый показатель разработки определяется по формуле:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}}$$

где $I_{\text{финр}}^{\text{исп.}i}$ – интегральный финансовый показатель разработки; Φ_{pi} – стоимость i -го варианта исполнения; Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта (в т.ч. аналоги).

Расчет:

$$I_{\text{финр}}^{\text{исп1}} = \frac{405522,09}{738150,74} = 0,55$$

$$I_{\text{финр}}^{\text{исп2}} = \frac{419639,74}{738150,74} = 0,56$$

$$I_{\text{финр}}^{\text{исп3}} = \frac{738150,74}{738150,74} = 1$$

Интегральный финансовый показатель разработки позволяет определить численное увеличение или уменьшение бюджета затрат разработки в разгах. На основании расчета интегрального финансового показателя разработки было определено, что при первом и втором исполнении бюджет разработки уменьшится в 0,55 и 0,56 раз соответственно. В третьем исполнении интегральный показатель равен 1, это показывает, что данный вариант не подразумевает увеличение или уменьшение бюджета затрат.

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить по формуле:

$$I_{pi} = \sum a_i * b_i$$

где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i^a, b_i^p – балльная оценка i -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Расчет интегрального показателя ресурсоэффективности приведен в таблице 17.

Таблица 17. Сравнительная оценка характеристик вариантов исполнения проекта

Объект исследования Критерии	Весовой коэффициент параметра	Исп.1	Исп.2	Исп.3
1. Точность	0,2	5	5	3

2. Удобство в эксплуатации	0,1	5	4	3
3. Помехоустойчивость	0,1	5	5	3
4. Энергосбережение	0,15	4	4	4
5. Надежность	0,25	4	4	4
6. Материалоемкость	0,2	4	5	3
Итого:	1			

$$I_{p-исп1} = 5*0,2 + 5*0,1 + 5*0,1 + 4*0,15 + 4*0,25 + 4*0,2 = 4,4$$

$$I_{p-исп2} = 5*0,2 + 4*0,1 + 5*0,1 + 4*0,15 + 4*0,25 + 5*0,2 = 4,5$$

$$I_{p-исп3} = 3*0,2 + 3*0,1 + 3*0,1 + 4*0,15 + 4*0,25 + 3*0,2 = 3,4$$

На основе сравнительной оценки характеристик вариантов исполнения можно сделать вывод, что второе исполнение наиболее ресурсоэффективно.

Интегральный показатель эффективности вариантов исполнения разработки ($I_{испi}$)

$$I_{исп 1} = \frac{I_{p-исп1}}{I_{финр.1}} = \frac{4,4}{0,55} = 8$$

$$I_{исп 2} = \frac{I_{p-исп2}}{I_{финр.2}} = \frac{4,5}{0,56} = 8,03$$

$$I_{исп 3} = \frac{I_{p-исп3}}{I_{финр.3}} = \frac{3,4}{1} = 3,4$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволяет определить сравнительную эффективность проекта и выбрать наиболее целесообразный вариант из предложенных. Сравнительная эффективность проекта ($\mathcal{E}_{ср}$) рассчитывается по формуле:

$$\mathcal{E}_{ср} = \frac{I_{исп1}}{I_{исп2}}$$

Сравнительная эффективность разработки, представлена в таблице 18.

Таблица 18. Сравнительная эффективность разработки

№ п/п	Показатели	Исп.1	Исп.2	Исп.3
1	Интегральный финансовый показатель разработки	0,55	0,56	1
2	Интегральный показатель ресурсоэффективности разработки	4,4	4,5	3,4
3	Интегральный показатель эффективности	8	8,03	3,4
4	Сравнительная эффективность вариантов исполнения	0,996	2,36	0,425

По результатам расчета сравнительной эффективности разработки можно сделать вывод о эффективности разработки каждого варианта исполнения. Данный анализ показывает, что разработка второго исполнения наиболее эффективная.

Вывод: Для определения целесообразности проведения исследования был проведен анализ уже существующих компаний, занимающихся продажей услуг по определению степени патологии легкого. В результате было выявлено, что отечественных разработок и аналогов в данном направлении не существует, что предоставляет возможность занять абсолютно пустую нишу на отечественном рынке.

Для определения конкурентоспособности было проведено два исследования по оценочной карте и по технологии QuaD. В результате исследования по оценочной карте было выявлено, что коэффициент конкурентоспособности равен 1,07, что показывает высокую конкурентоспособность. Результаты QuaD анализа подтвердили данный вывод, показав средневзвешенное значение показателя качества и перспективности научной разработки равным 74, что соответствует значению выше среднего.

SWOT-анализ показал степень соответствия сильных и слабых сторон проекта внешним условиям окружающей среды. На его основании было направлено развитие проекта.

Для оптимизации работы над проектом был составлен календарный график, включающий в себя количество дней, необходимых для разработки продукта, в трех вариантах исполнения, которые составили 84, 113 и 129 дней соответственно. Также было проведено разделение всей разработки на отдельные этапы и распределение данных этапов и работ между студентом-дипломником и руководителем проекта. В дальнейшем на основании данного календарного графика были определены затраты на заработную плату участникам проекта.

Оклад студента-дипломника был установлен в размере 6976,22 рублей, а руководителя проекта – в размере 23264,86 рублей. Для определения фонда заработной платы были рассчитаны затраты на основную и дополнительную заработную плату, а также отчисления во внебюджетные фонды.

Для определения материальных затрат была составлена смета, включающая в себя материальные ресурсы необходимые для осуществления проекта. Затраты на данные материалы составили 329264, 337544 и 628486 рублей в первом, втором и третьем исполнении соответственно. Также для данной статьи расходов были рассчитаны накладные расходы с коэффициентом в 50%.

На основании всех статей затрат в трех вариантах исполнения был сформирован бюджет затрат НИИ равный 405522,09 в первом исполнении, 419639,74 во втором исполнении и 738150,74 в третьем.

Для определения ресурсной, финансовой и экономической эффективности были проведены расчеты интегральных показателей разработки, ресурсоэффективности и эффективности вариантов исполнения разработки. На основании этих показателей был определен сравнительный интегральный показатель эффективности, который составил 0,996 в первом варианте, 2,36 во втором варианте и 0,425 в третьем варианте исполнения.

В результате проведенного исследования можно сказать, что с позиции технической и финансовой ресурсоэффективности наиболее целесообразным является второе исполнение.

Заключение

В диссертационной работе была приведена статистика заболеваемости и летальности от пневмонии, на основании которой был сделан вывод о важности своевременной и эффективной диагностике данного заболевания, а также о наибольшей эффективности компьютерной томографии в качестве средства диагностирования патологий легкого и применения методов классификации к улучшения этого процесса.

Был проведен анализ методов классификации и основных подходов к сегментации изображений. Результатом данного анализа стало выделения подхода к сегментации, основанного на алгоритмах слияния и разделения как наиболее подходящего для сегментации снимков компьютерной томографии. Также был проведен обзор методов текстурного анализа изображения, на основании которого был выделен алгоритм GLCM в качестве алгоритма для работы со снимками компьютерной томографии.

В качестве языка разработки был выбран C#, в качестве среды разработки – Microsoft Visual Studio 2015, а также RapidMner v7.3. В качестве платформы для тестирования и анализа качества модели.

Для определения наиболее подходящего режима отображения снимков КТ была проведена сегментация снимков в легочном, плевральном, костном и жировом режимах отображения. Легочный и плевральный подход показали большую эффективность при сегментации.

На основании алгоритма GLCM было разработано программное обеспечение для определения патологии легкого у больных пневмонией. Результаты работы ПО были протестированы с помощью классификатора, разработанного в приложении RapidMner Studio с помощью деревьев решений. Тестирование было проведено для двух пациентов и для пяти размеров рассматриваемого окна.

По результатам тестирования был сделан вывод о высокой точности определения патологии в легочном режиме, а также определен размер окна показавший наиболее высокую точность - 9×9 пикселей.

Список публикаций

1. Эффективность сегментации КТ-сканов легкого в различных режимах отображения при пневмонии. - III Международная научно-практическая конференция «Advances in Science and Technology»
2. Основные принципы классификации данных. - III Всероссийская научно-практическая конференция «Информационные технологии в прикладных исследованиях»

Список литературы

1. Н.А. Таубалдинова, Л.И. Вансович, А.М. Джексенбиева, А.А. Бушанская Распространенность заболеваемости и применение физиотерапии в лечении пневмонии // Вестник КазНМУ - 2013. - №3 – С. 3.
2. Всемирная организация здравоохранения – Пневмония [Электронный ресурс] - <http://www.who.int/mediacentre/factsheets/fs331/ru/>
3. Пневмония [Электронный ресурс] - <https://www.smed.ru/guides/43578#>
4. Фисенко В.Т., Фисенко Т.Ю. Компьютерная обработка и распознавание изображений: Учеб.пособие.– СПб: СПбГУ ИТМО, 2008.– 192 с.
5. Грузман И.С., Киричук В.С. Цифровая обработка изображений в информационных системах: Учеб.пособие. – Новосибирск: Изд-во НГТУ, 2002. – 352 с.
6. Коновалов, В.К., Дуков, Л.Г., Колмогоров, В.Г., Лукьяненко, Н.Я., Бобкова, А.В. Топографическая КТ-диагностика затемнений небольших размеров в легких – Вестник рентгенологии и радиологии, 2010. – С.15.
7. Косюра О.В., Филипов М.В. Модифицированный метод нормальных разрезов для сегментации цветных изображений - Новые информационные технологии в автоматизированных системах, №19, 2016. – С.26.
8. Марусина, М.Я., Волгарева, А.П. Подходы к разработке методов автоматической сегментации томографических изображений – Фундаментальные и прикладные проблемы техники и технологии, №15, 2015. – С.136-141.
9. Недзьведь, А.М., Белоцерковский, А.М., Исмаил-Заде, Р.С., Казакевич, А.П. 2D-сегментация изображений компьютерной томографии

на основе комплексного анализа окрестности - *Таврический вестник информатики и математики.*, № 2, 2008. – С.121-129.

10. Пеганова, Е.В., Батухтин, Д.М., Митракова, Н.Н. Автоматизированная система сегментации узкоспектральных изображений для оптимизации эндоскопической диагностики при патологии пищевода. - *Экспериментальная и клиническая гастроэнтерология.*, №2, 2014. – С.32-37.

11. Ветров, А.А., Кузнецов, А.Е. Автоматическая сегментация облачных объектов на снимках земной поверхности высокого пространственного разрешения – *Исследование Земли из космоса.*, №2, 2014. – С.27.

12. Евтюшкин, А.В. Брыксин, В.М., Данилова, А.Е., Филатов, А.В. Сегментация многомерных радарных изображений земных покровов. – *Вестник Балтийского федерального университета им. И.Канта.*, №10, 2014 – С.125-132.

13. Барабаш, П.И., Гостюшкин, В.В., Косых, Н.Э., Мазур, А.И., Савин, С.З. Проблема оценки качества изображений в радионуклидной диагностике. – *Вестник Тихоокеанского государственного университета.*, №3, 2013 – С.17-22.

14. Гонсалес, Р. Цифровая обработка изображений/Р. Гонсалес, Р. Вудс // М.:Техно-сфера, 2005.–С. 1072.

15. Рылов, С.А. Методы и алгоритмы сегментации мультиспектральных спутниковых изображений высокого пространственного разрешения кандидата технических наук: 05.13.18 – Новосибирск, 2016. – С. 11.

16. Dey, V. A review on image segmentation techniques with remote sensing perspective /V. Dey, Y. Zhang, M. Zhong // Proc. ISPRS TC VII Symposium, IAPRS.–2010.–Vol.38.–Part 7A.–P.31-42.

17. Sarmah, S. A grid-density based technique for finding clusters in satellite image/ S.Sarmah, D.K.Bhattacharyya // Pattern Recognition Letters. – 2012. –Vol.33. – No.5. –P.589-604.

18. Chauhan, A.S. Image segmentation methods: A survey approach/ A.S.Chauhan, S.Silakari, M.Dixit // Proc. Fourth International Conference on Communication Systems and Network Technologies. – IEEE, 2014.–P.929-933.
19. Барталев, С.А. Анализ возможностей применения методов сегментации спутниковых изображений для выявления изменений в лесах / С.А. Барталев, Т.С. Ховратович // Современные проблемы дистанционного зондирования Земли из космоса. – 2011. –Т.8. – No1. – С. 44-62.
20. Jain, A.K. Data clustering: 50 years beyond K-means // Pattern recognition letters. – 2010. – Vol.31. – No.8. – P.651-666.
21. Xu, R. Survey of clustering algorithms / R. Xu, D. Wunsch // IEEE Transactions, Neural Networks. – 2005. – Vol.16. – No.3. – P.645-678.
22. Jain, A.K. Data clustering: a review / A.K. Jain, M.N. Murty, P.J. Flynn // ACM computing surveys (CSUR). – 1999. – Vol.31. –No.3. –P.264-323.
23. Filippone, M. A survey of kernel and spectral methods for clustering / M. Filippone, F. Camastra, F. Masulli, S. Rovetta // Pattern recognition. – 2008. – Vol.41. –No.1.– P.176-190.
24. Fahad, A. A survey of clustering algorithms for big data: Taxonomy and empirical analysis / A. Fahad, N. Alshatri, Z. Tarietal. // IEEE Transactions, Emerging Topics in Computing –2014. – Vol.2. – No.3. – P.267-279.
25. Ahmed, N. Recent review on image clustering // Image Processing, IET. – 2015. – Vol.9 – No.11. – P.1020 - 1032.
26. Nagpal, A. Review based on data clustering algorithms / A. Nagpal, A. Jatain, D. Gaur // IEEE Conference, Information & Communication Technologies (ICT). – IEEE, 2013. – P.298-303.
27. Arumugadevi,S. Comparison of clustering methods for segmenting color images / S. Arumugadevi, V.
28. Seenivasagam // Indian Journal of Science and Technology. – 2015. – Vol.8. – No.7. – P.670-677.

29. Kashef, R. Cooperative clustering / R. Kashef, M.S. Kamel // Pattern Recognition. – 2010. – Vol.43. – No.6. – P.2315-2329.
30. Filkov, V. Integrating microarray data by consensus clustering / V. Filkov, S. Skiena // International Journal on Artificial Intelligence Tools. – 2004. – Vol.13. – No.04. – P.863-880.
31. Ilea, D.E. Color image segmentation using a self-initializing EM algorithm. / D.E. Ilea, P.F. Whelan // Proc. 6th International Conference Visualization, Imaging, Image Processing (IASTED). – 2006.–P. 417-424.
32. Nammalwar, P. Integration of feature distributions for colour texture segmentation / P. Nammalwar, O. Ghita, P.F. Whelan // Proc. 17th International Conference on Pattern Recognition (ICPR). – IEEE, 2004. – Vol.1. – P. 716-719.
33. Sima, H. Texture Region Merging with Histogram Feature for Color Image Segmentation / H. Sima, P. Guo // Proc. 9th International Conference on Computational Intelligence and Security (CIS). – IEEE, 2013. – P.224-228.
34. Yang, Y. Multilayer graph cuts based unsupervised color–texture image segmentation using multivariate mixed student's t-distribution and regional credibility merging / Y.Yang, S.Han, T.Wang, W.Tao, X.C.Tai // Pattern Recognition. – 2013. – Vol. 46. – No.4. – P.1101-1124.
35. Ilea, D.E. Image segmentation based on the integration of colour–texture descriptors – A review /D.E. Ilea, P.F. Whelan // Pattern Recognition. – 2011. – Vol.44. – No.10.–P.2479-2501.
36. Пестунов И.А., Рылов С.А. Сегментация спутниковых изображений высокого разрешения по спектральным и текстурным признакам. – Интерэкспо Гео-Сибирь – 2012. - № 4 – С.24.
37. F. D’Astous and M.E. Jernigan Texture Discrimination Based on Detailed Measures of the Power Spectrum 83-86.
38. Patric C. Chen and T. Pavlidis. Segmentation by Texture Using Correlation. 551-553.

39. Andre Gagalowicz and Song De Ma. Sequential Synthesis of Natural Textures Computer Vision, Graphics and Image Processing 30, 1985, 289-315.
40. John-Peter Lewis Texture Synthesis for Digital Painting. Computer Graphics. Volume 18, Number 3 July 1984, 245-252.
41. Павлидис Т. Алгоритмы машинной графики и обработки изображения. Перевод с английского: - М., "Радио и связь" -1986 г., - 400 с.
42. Сидорова В.С. Текстуальный анализ аэрокосмических изображений на ЭВМ. 30-36.
43. Прэйтт У. К., Фожра О. Д., Гагалович А. Применение моделей стохастических текстур для обработки изображений. ТИИЭР 5, 1981, 54-64.
44. Andre Gagalowicz, Sond De Ma, Catherine Tournier-Lasserve 1-D Model for Color Texture. - Computer Vision for Robots -1985 - 45-49.
45. David D. Garber and Alexander A. Sawchuk, Texture Simulation Using a Best-Fit Model. Pattern Recognition and Image Processing, August 1981, 603-608.
46. Noboru Funakubu Region Segmentation of Biomedical Tissue Image Using Color Texture Feature Pattern Recognition and Image Processing, Juli30 –August 2, vol. 1, 1984, 30-32.
47. Hidefumi Kobatake and Jun Moroo Partition of Texture Image Using Two-Dimensional Linear Prediction Model.

Приложения

Приложение А Приложение А

Раздел 7 Обзор литературы

Студент:

Группа	ФИО	Подпись	Дата
8км51	Лигай Светлана Дореновна		

Консультант кафедры ПИ :

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ПИ	Чердынцев Е.С.	К.Т.Н.		

Консультант – лингвист кафедры ИЯИК :

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Зав.каф.	Сидоренко Т.В.	К.П.Н.		

Приложение Б

Обзор литературных источников на иностранном языке

Pneumonia is one of the most common diseases on earth, which naturally gives rise to interest in its study. The high mortality rate persists for decades, despite the ever-increasing level of medical treatment methods. To a greater degree, this state of affairs is a consequence of the shortcomings in the level of methods for diagnosing this disease.

Among the methods for diagnosing pneumonia, the most accurate is computerized tomography. At the moment, topographic computed tomography is used for the most effective detection of pulmonary lesions with small blackouts. This method showed a high accuracy of diagnosis - 10.9%. [6]

However, for the recognition of small damaged areas of the lung, methods of automatic image segmentation are used. These algorithms make it possible to isolate the affected areas of the lung. There are many studies of image segmentation techniques.

Classification of data is a huge and rapidly developing part of information technology, its algorithms are designed to analyze data and distribute them in some orderly system. This is the simplest and most general description of the classification.

So, classification is a form of data analysis, in which models that describe important data classes are distinguished. Such models, called classifiers, are labels for the category of classes.

Classification is a two-step process, consisting of a stage of training and direct classification. The first step, that is, the process of training or training the system involves building a general concept or classifier based on already defined data and classes.

Classification algorithms build classifiers based on training sets or tuples of data that have their own classification marks. The tuple X is represented as an n -dimensional vector $X = (x_1, x_2, \dots, x_n)$ describing n dimensions for a tuple of n attributes from the database A_1, A_2, \dots, A_n , respectively. Each tuple X is assumed to

belong to a predefined class, that is, defined by another class attribute, called a classification label attribute.

The accuracy of the classifier on a given test set of data is the percentage of the test tuple that is correctly distributed by this classifier. To train the system, some sets of training tuples are used, on the basis of which the model is built and refined. After all the data goes through a direct classification process. To achieve the necessary accuracy, the model is taught repeatedly.

Different methods of classification are characterized by certain properties that can be decisive when choosing a method for analyzing data. Methods can be compared among themselves, evaluating the characteristics of their properties.

The main properties and characteristics of classification methods: accuracy, scalability, labor intensity, speed.

Scalability is a property of a computer system that provides predictable growth of system characteristics, for example, speed of response, overall performance, etc., when adding computing resources to it.

In [36], the information content of textural attribute systems is analyzed using four methods for selecting textural features.

When analyzing images, their important characteristic is the texture that is present in all images, from images obtained with the help of aircraft and satellite devices to microscopic images in biomedical research. However, despite this, the concept of texture is not well defined and is understood even today. From the literature, there are two types of texture definitions:

- Firstly, this is the interpretation of the texture as a repetition of basic primitives having different orientation in space. That is, this definition insists on the structured nature of the texture. Examples in this case are textures of fabric, brick wall, etc. Supporters of such a definition focused themselves on spectral analysis and representation of the texture [37].

- Secondly, the texture is considered as an anarchic and homogeneous aspect, which does not have pronounced edges. For supporters of this method, there are no conspicuous patterns or dominant frequency in the texture (for example, turf, crust,

ground and so on, viewed from a long distance), i.e. They focused themselves on the probabilistic method of solving the texture problem [38].

In fact, a better definition of the term "texture" is achieved by the synthesis of both methods described above. Texture should be considered a two-level structure. That is, it is a spatial organization (the highest level) of basic primitives (or non-derivative elements as they call Kharalik [37]), which themselves have a random aspect (the lowest level) [39].

A number of methods are used to measure and describe image textures. Among them are the following groups:

- methods based on measurement of spatial frequency (fine textures are dominated by high, and in coarse textures - low spatial frequencies);
- methods based on the calculation of the number of drops per unit area of the image (on coarse textures this value is small, with a decrease in the graininess of the texture it increases);
- methods using the adjacency matrix of brightness values (with increasing distance between the points being evaluated, in coarse-grained textures, the change in the brightness distribution is much slower than in fine-grained textures);
- methods that describe textures with lengths of series (lines with constant brightness of points, on coarse textures these series are longer than on fine-grained textures);
- Autoregressive methods for describing the texture use the coefficients of linear estimates of the brightness of the point element of the image from the given values of the elements of some of its neighborhoods (these coefficients are almost identical for coarse-grained and essentially different for Fine-grained textures);
- methods based on the histogram of the spatial brightness difference;

- methods that look for regularity in the form of structural elements;
- methods based on the analysis of the microstructure of the texture field.

The advantage of spatial frequency as a means of describing the texture is the familiarity of this concept. However, there is a significant difficulty associated with quantizing the brightness of the image. The quantization procedures are not invariant even with respect to monotonic brightness transformations.

Currently, textural information is widely used in image processing algorithms and computer graphics. The class of problems solved using the texture is large enough. Most often, the use of texture can be found in the problems of segmentation and modeling of three-dimensional images (the term "three-dimensional images" means not actually three-dimensional images, but their two-dimensional projections).

Segmentation allows you to select areas of the image that seem to the observer the same; This operation ensures the image is divided into homogeneous areas. The homogeneity of the region can be determined in terms of brightness levels, but it is possible on the basis of the texture information of the region under consideration. That is, we can say that the regions are perceived by the observer as homogeneous if they have a similar texture.

A comparative analysis was performed to identify the most effective methods of image segmentation. Comparative analysis was carried out among methods based on edge separation operators, watershed methods, the method of growing regions, the method of normal sections, methods of growing regions and methods for estimating the change in the shape of the spectrum. Based on this analysis, conclusions were drawn about the effectiveness of the normal section method. A modification of the method of normal sections is also proposed. [7]

Among the three generations of image segmentation methods, methods that differ in accuracy, autonomy, complexity, the ability to specify these algorithms for specific tasks and many other parameters have been singled out. [8]

Segmentation of images is a developing area. A number of already existing methods and algorithms for segmentation are developing. One such algorithm is an algorithm based on an analysis of related components and new principles for representing halftone images. [9]

Segmentation of images can be applied in a variety of areas. For each specific task, it is necessary to develop a specific algorithm. In studies for 238 patients with pathologies of the upper divisions of the digestive tract, image segmentation algorithms were applied. [10]

Also, as a result of studies of segmentation of earth surface images, an automatic segmentation algorithm was developed, which is based on the photosensitivity of elements of different spectral channels. To implement the algorithm, a colometric algorithm is used and the Bayesian classifier. [11]

Thanks to the application of algorithms for segmentation and classification of satellite radar images, it is possible to isolate homogeneous parts of images without removing the noise of images. [12]

In order to diagnose skeletal metastases, an automated computer diagnostic system was developed that has the function of expert analysis of images. The system includes the segmentation of images, the creation of a training sample of images after medical treatment. [13]

Segmentation of satellite images of high resolution allows us to develop effective methods for the formation of texture attributes for halftone images. Such studies allow you to get acquainted with nonparametric algorithms of clustering and methods of segmentation of any high-resolution images.

To determine texture attributes, several approaches should be distinguished: statistical, structural and spectral. [14]

The study of texture characteristics of high-resolution images requires consideration of statistical methods that evaluate statistical characteristics. These characteristics are determined based on texture matrices, which is the matrices of compatibility of pairs of neighboring pixels at a single tone level. [15]

Based on the studies and complex analysis it is possible to determine the most effective methods and algorithms that can form the basis of the algorithm for segmentation of CT images for patients with pneumonia. At the heart of all high-level methods of segmentation are the basic algorithms. All the segmentation algorithms can be divided according to the principle of their operation, that is, according to the applied actions for the formation of clusters. You can distinguish groups: per-pixel, boundary, based on the selection of homogeneous regions and hybrid. [16]

Per-pixel methods do not take into account the relationship of neighboring pixels or their position on the image, but separate clusters based only on the feature space. Separation in this approach can be carried out using fixed thresholds, for example, a decision tree. However, this approach is effective only if there is a priori knowledge. With such knowledge, this task may seem to be a classification task with learning or clustering without training. [17]

Boundary methods are methods for finding boundary pixels using gradient filters, such as Roberts, Sobel, Laplace and others. After there is a closure Which determines the allocated zones. The main algorithm used as a method of filling contours is the watershed algorithm. A strong dependence on noise is an important drawback of this approach. [18]

Methods based on the analysis of areas include methods for dividing images into homogeneous areas, taking into account spectral and textural features, by analyzing the spatial proximity of pixels. Methods for scaling up areas, dividing and merging areas are also included in this section. [19]

The task of clustering is the separation of many objects into a relatively small number of subsets, which are called clusters. Clusters must differ significantly from each other in feature space, and objects within clusters must be similar. [20] For each specific task, it is necessary to select measures of similarity for objects and the number of clusters, since there is no definite answer to this question. [21]

At the moment, there are several hundred methods of clustering data. This topic is considered in a variety of articles and monographs (for example, 20-27). An

overview of clustering methods allows one to draw conclusions about the complexity of this problem and the inability to create a universal clustering algorithm.

Qualitative segmentation should combine not only spectral, but also textural features. Thus, studies that allow the combination of spectral and textural characteristics are an actively developing direction.

The extraction of spectral and textural features differs depending on the approaches: methods based on sequential extraction of spectral and texture characteristics, methods based on independent extraction of characteristics and probability-statistical methods. [29]

The first approach is the sequential extraction of texture and spectral characteristics. However, a significant drawback of this approach is that the criteria extracted first have a greater weight in segmentation. A vivid example of sequential extraction is the JSEG algorithm.

The JSEG algorithm is performed in three stages. At the first stage, adaptive quantization is carried out, then a J-image is created that contains local homogeneity, which depends on the uniformity of the distribution of color marks in the neighborhood of each pixel. This is necessary to determine the boundaries of clusters, as an identifier. The final stage is the use of the method of building up clusters, while neighboring segments that are close in color can be combined.

The disadvantage of the JSEG algorithm is the possible re-intersection due to the lack of consideration for the heterogeneity of textures. At the moment, there are modifications to the JSEG algorithm, which produce additional clustering by texture features.

The second approach is based on independent extraction of spectral and texture characteristics, with due regard for spatial relationships between the pixels. These include the algorithms of merging and separation [30-32], area-building algorithms [14], an algorithm based on active contours [34].

Algorithms for separation and merging are also produced in three stages. Initially, the image is divided into segments, mostly square ones, then adjacent

segments are combined if the distance between them does not exceed the specified value. The distance is the sum of the distances along the textural and spectral features. At the final stage, the pixel boundaries are corrected pixel-by-pixel. The advantage of this approach is the speed, so it is the basis of many algorithms.

A more accurate definition of the boundaries of segments is possible with the use of area-building algorithms. Initially, they select the "seeds" - the starting points, then iteratively attach adjacent pixels to them. The main disadvantage of this approach is the strong dependence of the result on the correct choice of "seeds" [9].

Active-loop algorithms represent an iterative change in the contours of a segment, while it is assumed that the pixels within the original contour have a high degree of uniformity. The disadvantage is also a high dependence on the choice of the initial parameters. The advantage is the high accuracy of segment boundaries [34].

The third approach is based on a probabilistic-statistical approach. This approach considers the whole image as a single spectral-texture space. Examples of such an approach can be algorithms of K-means, neural networks, fuzzy clustering. The disadvantages of such approaches are the need to initially set the number of clusters, large computational costs [35].

The most common approach is the separation and merging algorithms. Based on the review of [4-35], in which examples and algorithms of image segmentation are given, it can be concluded that the separation and merging algorithms are useful for segmentation of computer tomography images of patients with pneumonia. Due to the small number of images, the computational costs do not exceed the threshold value, and the speed of the algorithms is an indisputable advantage.

Приложение В

Результаты оценки качества модели

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4
1	1	1	0.994	0.006	0	0	0	35	2
2	1	1	0.994	0.006	0	0	0	31	2
3	1	1	0.994	0.006	0	0	0	27	2
4	1	1	0.994	0.006	0	0	0	22	1
5	1	1	0.994	0.006	0	0	0	19	0
6	1	1	0.994	0.006	0	0	0	16	0
7	1	1	0.994	0.006	0	0	0	13	0
8	1	1	0.994	0.006	0	0	0	11	0
9	1	1	0.994	0.006	0	0	0	10	0
10	1	1	0.994	0.006	0	0	0	8	0
11	1	1	0.994	0.006	0	0	0	6	0
12	1	1	0.994	0.006	0	0	0	4	0
13	1	1	0.994	0.006	0	0	0	3	0
14	1	1	0.994	0.006	0	0	0	3	0
15	1	1	0.994	0.006	0	0	0	2	0
16	1	1	0.994	0.006	0	0	0	2	0

Рис 1. Таблица предсказаний для пациента 131 в легочном режиме

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4
1	1	0	0.005	0.995	0	0	0	10	0
2	1	0	0.005	0.995	0	0	0	9	0
3	1	0	0.005	0.995	0	0	0	7	0
4	1	0	0.005	0.995	0	0	0	6	0
5	1	0	0	1	0	0	0	4	0
6	1	1	0.934	0.066	0	0	0	3	0
7	1	1	0.934	0.066	0	0	0	2	0
8	1	1	0.934	0.066	0	0	0	1	0
9	1	1	0.934	0.066	0	0	0	0	0
10	1	1	0.934	0.066	0	0	0	0	0
11	1	1	0.934	0.066	0	0	0	0	0
12	1	1	0.934	0.066	0	0	0	0	0
13	1	1	0.934	0.066	0	0	0	0	0
14	1	1	0.934	0.066	0	0	0	0	1
15	1	1	0.934	0.066	0	0	0	0	2
16	1	1	0.977	0.023	0	0	0	0	3

Рис 2. Таблица предсказаний для пациента 131 в плевральном режиме

accuracy: 60.76%

	true 1	true 0	class precision
pred. 1	35264	23206	60.31%
pred. 0	350	1215	77.64%
class recall	99.02%	4.98%	

Рис 3. Таблица оценки точности модели для пациента 131 в легочном режиме

accuracy: 80.72%

	true 1	true 0	class precision
pred. 1	29792	12637	70.22%
pred. 0	1670	30112	94.75%
class recall	94.69%	70.44%	

Рис 4. Таблица оценки точности модели для пациента 131 в плевральном режиме

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4
1	1	0	0.017	0.983	0	0	0	79	92
2	1	0	0.017	0.983	0	0	0	75	97
3	1	0	0.017	0.983	0	0	0	72	103
4	1	0	0.017	0.983	0	0	0	68	109
5	1	0	0.017	0.983	0	0	0	66	112
6	1	0	0.017	0.983	0	0	0	62	115
7	1	0	0.017	0.983	0	0	0	57	117
8	1	0	0.017	0.983	0	0	0	51	118
9	1	0	0.017	0.983	0	0	0	53	112
10	1	0	0.017	0.983	0	0	0	46	110
11	1	0	0.017	0.983	0	0	0	46	106
12	1	1	0.836	0.164	0	0	0	42	105
13	1	1	0.836	0.164	0	0	0	42	102
14	1	1	0.836	0.164	0	0	0	38	102
15	1	1	0.836	0.164	0	0	0	40	96
16	1	1	0.836	0.164	0	0	0	40	88

Рис 5. Таблица предсказаний для пациента 254 в легочном режиме

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4
1	1	0	0.112	0.888	0	0	0	51	42
2	1	0	0.112	0.888	0	0	0	52	42
3	1	0	0.112	0.888	0	0	0	54	42
4	1	0	0.112	0.888	0	0	0	56	44
5	1	0	0.112	0.888	0	0	0	60	46
6	1	0	0.112	0.888	0	0	0	59	52
7	1	0	0.112	0.888	0	0	0	54	59
8	1	0	0.112	0.888	0	0	0	51	64
9	1	0	0.112	0.888	0	0	0	48	69
10	1	0	0.112	0.888	0	0	0	53	67
11	1	0	0.112	0.888	0	0	0	54	67
12	1	0	0.112	0.888	0	0	0	53	69
13	1	0	0.112	0.888	0	0	0	54	72
14	1	0	0.112	0.888	0	0	0	51	78
15	1	0	0.112	0.888	0	0	0	50	84
16	1	0	0.112	0.888	0	0	0	53	88

Рис 6. Таблица предсказаний для пациента 254 в плевральном режиме

accuracy: 87.26%

	true 1	true 0	class precision
pred. 1	41517	2540	94.23%
pred. 0	5068	10607	67.67%
class recall	89.12%	80.68%	

Рис 7. Таблица оценки точности модели для пациента 254 в легочном режиме

accuracy: 59.99%

	true 1	true 0	class precision
pred. 1	21749	1035	95.46%
pred. 0	23283	14720	38.73%
class recall	48.30%	93.43%	

Рис 8. Таблица оценки точности модели для пациента 254 в плевральном режиме

ExampleSet (186376 examples, 4 special attributes, 24 regular attributes)

Filter (186,376 / 186,376 examples):

all

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4	Pattern1.5	Pattern2.0
1	1	1	0.869	0.131	0	0	0	4	16	0	0
2	1	1	1	0	0	0	0	3	12	0	0
3	1	1	0.987	0.013	0	0	0	1	8	0	0
4	1	1	0.987	0.013	0	0	0	0	4	0	0
5	1	1	0.987	0.013	0	0	0	0	2	0	0
6	1	1	0.987	0.013	0	0	0	0	1	0	0
7	1	1	0.987	0.013	0	0	0	0	0	0	0
8	1	1	0.987	0.013	0	0	0	0	0	0	0
9	1	1	0.987	0.013	0	0	0	0	0	0	0
10	1	1	0.987	0.013	0	0	0	0	0	0	0
11	1	1	0.987	0.013	0	0	0	0	0	0	0
12	1	1	0.987	0.013	0	0	0	0	0	0	0
13	1	1	0.987	0.013	0	0	0	0	0	0	0
14	1	1	0.987	0.013	0	0	0	0	0	0	0
15	1	1	0.987	0.013	0	0	0	0	0	0	0
16	1	1	0.987	0.013	0	0	0	0	0	0	0

Рис 9. Таблица предсказаний в легочном режиме для окна размером 9×9

accuracy: 90.39%

	true 1	true 0	class precision
pred. 1	113312	1205	98.95%
pred. 0	16706	55153	76.75%
class recall	87.15%	97.86%	

Рис 10. Таблица оценки точности модели в легочном режиме для окна размером 9×9

ExampleSet (283069 examples, 4 special attributes, 24 regular attributes)

Filter (283,069 / 283,069 examples):

all

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4	Pattern1.5	Pattern2.0	P
1	1	1	0.966	0.034	0	0	0	2	11	0	0	0
2	1	1	0.951	0.049	0	0	0	2	12	0	0	0
3	1	1	0.951	0.049	0	0	0	2	13	0	0	0
4	1	1	0.951	0.049	0	0	0	3	15	0	0	0
5	1	1	0.951	0.049	0	0	0	4	19	0	0	0
6	1	1	0.951	0.049	0	0	0	4	23	0	0	0
7	1	1	0.951	0.049	0	0	0	4	23	0	0	0
8	1	1	0.951	0.049	0	0	0	4	19	0	0	0
9	1	1	0.951	0.049	0	0	0	2	17	0	0	0
10	1	1	0.951	0.049	0	0	0	2	15	0	0	0
11	1	1	0.951	0.049	0	0	0	2	13	0	0	0
12	1	1	0.966	0.034	0	0	0	1	11	0	0	0
13	1	1	0.966	0.034	0	0	0	0	7	0	0	0
14	1	1	0.966	0.034	0	0	0	0	3	0	0	0
15	1	1	0.966	0.034	0	0	0	0	2	0	0	0
16	1	1	0.966	0.034	0	0	0	0	4	0	0	0

Рис 11. Таблица предсказаний в плевральном режиме для окна размером 9×9

accuracy: 78.16%

	true 1	true 0	class precision
pred. 1	88362	26708	76.79%
pred. 0	35106	132893	79.10%
class recall	71.57%	83.27%	

Рис 12. Таблица оценки точности модели в плевральном режиме для окна размером 9×9

ExampleSet (200772 examples, 4 special attributes, 24 regular attributes)

Filter (200,772 / 200,772 examples):

all

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4	Pattern1.5	Pattern2.0
1	1	1	0.994	0.006	0	0	0	6	29	0	0
2	1	1	0.994	0.006	0	0	0	5	22	0	0
3	1	1	0.994	0.006	0	0	0	3	16	0	0
4	1	1	0.994	0.006	0	0	0	2	10	0	0
5	1	1	0.994	0.006	0	0	0	2	5	0	0
6	1	1	0.994	0.006	0	0	0	1	2	0	0
7	1	1	0.994	0.006	0	0	0	0	0	0	0
8	1	1	0.994	0.006	0	0	0	0	0	0	0
9	1	1	0.994	0.006	0	0	0	0	0	0	0
10	1	1	0.994	0.006	0	0	0	0	0	0	0
11	1	1	0.994	0.006	0	0	0	0	0	0	0
12	1	1	0.994	0.006	0	0	0	0	0	0	0
13	1	1	0.994	0.006	0	0	0	0	0	0	0
14	1	1	0.994	0.006	0	0	0	0	0	0	0
15	1	1	0.994	0.006	0	0	0	0	0	0	0
16	1	1	0.994	0.006	0	0	0	0	0	0	0

Рис 13. Таблица предсказаний в легочном режиме для окна размером 15×15

accuracy: 83.37%

	true 1	true 0	class precision
pred. 1	107665	26494	80.25%
pred. 0	6903	59710	89.64%
class recall	93.97%	69.27%	

Рис 14. Таблица оценки точности модели в легочном режиме для окна размером 15×15

ExampleSet (11528 examples, 4 special attributes, 24 regular attributes) Filter (11,528 / 11,528 examples): all

Row No.	Abnormal	prediction(A...	confidence(1)	confidence(0)	Pattern1.0	Pattern1.1	Pattern1.2	Pattern1.3	Pattern1.4	Pattern1.5	Pattern2.0
1	1	1	0.977	0.023	0	0	0	6	29	0	0
2	1	1	0.977	0.023	0	0	0	5	22	0	0
3	1	1	0.977	0.023	0	0	0	3	16	0	0
4	1	1	0.977	0.023	0	0	0	2	10	0	0
5	1	1	0.977	0.023	0	0	0	2	5	0	0
6	1	0	0	1	0	0	0	1	2	0	0
7	1	1	0.934	0.066	0	0	0	0	0	0	0
8	1	1	0.934	0.066	0	0	0	0	0	0	0
9	1	1	0.934	0.066	0	0	0	0	0	0	0
10	1	1	0.934	0.066	0	0	0	0	0	0	0
11	1	1	0.934	0.066	0	0	0	0	0	0	0
12	1	1	0.934	0.066	0	0	0	0	0	0	0
13	1	1	0.934	0.066	0	0	0	0	0	0	0
14	1	1	0.934	0.066	0	0	0	0	0	0	0
15	1	1	0.934	0.066	0	0	0	0	0	0	0
16	1	1	0.934	0.066	0	0	0	0	0	0	0

Рис 13. Таблица предсказаний в плевральном режиме для окна размером 15×15

accuracy: 68.41%

	true 1	true 0	class precision
pred. 1	5158	0	100.00%
pred. 0	3642	2728	42.83%
class recall	58.61%	100.00%	

Рис 14. Таблица оценки точности модели в плевральном режиме для окна размером 15×15