

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

Школа Инженерная школа информационных технологий и робототехники  
Направление подготовки 09.04.03. Прикладная информатика  
Отделение Информационных технологий

**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**

Тема работы
Система поддержки принятия врачебных решений при выборе тактики лечения детей с эндокринопатиями

УДК 005.53:616.43-053.3-08

Студент

Группа	ФИО	Подпись	Дата
8КМ61	Емельянова Юлия Андреевна		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Марухина О.В.	к. т. н.		

**КОНСУЛЬТАНТЫ:**

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент отделения СГН	Старикова Е.В.	к. ф. н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Авдеева И.И.			

Консультант – лингвист отделения ИЯ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Краснова Т.И.			

**ДОПУСТИТЬ К ЗАЩИТЕ:**

Руководитель ООП 09.04.03 Прикладная информатика	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Марухина О.В.	к. т. н.		

**Запланируемые результаты обучения по программе  
09.04.03 – Прикладная информатика**

Код	Результат обучения
P1	Применяет базовые и специальные знания в области современных информационно-коммуникационных технологий для решения междисциплинарных инженерных задач.
P2	Проводит теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретацию полученных данных в области информатизации и автоматизации прикладных процессов и создания, внедрения, эксплуатации и управления информационными системами в прикладных областях
P3	Внедряет, сопровождает и эксплуатирует современные информационные системы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья и безопасности труда, выполнять требования по защите окружающей среды
P4	Активно владеет иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты инновационной инженерной деятельности.
P5	Владеет и применяет методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе глобальных компьютерных сетей.
P6	Эффективно работает индивидуально, в качестве члена и руководителя группы, состоящей из специалистов различных направлений и квалификаций, демонстрирует ответственность за результаты работы и готовность следовать корпоративной культуре организации
P7	Самостоятельно учится и непрерывно повышает квалификацию в течение всего периода профессиональной деятельности
<b>Профиль «Системы корпоративного управления»</b>	
P8	Применяет глубокие профессиональные знания основ построения информационных технологий и систем, достаточные для решения научных и профессиональных задач производства. Знает современные проблемы и методы прикладной информатики и научно-технического развития информационных технологий.
P9	Ставит и решает задачи комплексного анализа, связанные с информатизацией и автоматизацией прикладных процессов; созданием, внедрением, эксплуатацией и управлением информационными системами в прикладных областях, с использованием базовых и специальных знаний, современных аналитических методов и моделей
P10	Способен организовывать работы по моделированию прикладных ИС и реинжинирингу прикладных и информационных процессов предприятия и организации. Способен управлять проектами по информатизации прикладных задач и созданию ИС предприятий и организаций.

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

Школа Инженерная школа информационных технологий и робототехники  
 Направление подготовки 09.04.03. Прикладная информатика  
 Отделение Информационных технологий

УТВЕРЖДАЮ:

Руководитель ООП 09.04.03. Прикладная информатика

\_\_\_\_\_ Марухина О.В.

(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
8КМб1	Емельяновой Юлии Андреевне

Тема работы:

Система поддержки принятия врачебных решений при выборе тактики лечения детей с эндокринопатиями	
Утверждена приказом директора (дата, номер)	20.04.2018 г., № 2796/С

Срок сдачи студентом выполненной работы:	
--	--

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<p><b>Исходные данные к работе</b>  <i>(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.).</i></p>	<p>Клинико-лабораторные показатели детей с эндокринопатиями, собранные Томским НИИ Курортологии и физиотерапии.</p>
<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b>  <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>	<p>Анализ предметной области                  Поиск методов и инструментария для решения поставленных задач                  Реализация проекта</p>

<b>Перечень графического материала</b> <i>(с точным указанием обязательных чертежей)</i>	Диаграммы размахов Графики информативных и неинформативных признаков Графическое построение дерева решений
---	--

**Консультанты по разделам выпускной квалификационной работы**  
*(с указанием разделов)*

Раздел	Консультант
<b>Разделы 1-3</b>	Марухина Ольга Владимировна
<b>Раздел 4</b>	Старикова Екатерина Васильевна
<b>Раздел 5</b>	Авдеева Ирина Ивановна

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

Введение
Глава 1. Обзор литературы
Глава 2. Поиск методов решения задачи
Глава 3. Реализация проекта

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	
---	--

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Марухина О. В.	К. Т. Н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8КМб1	Емельянова Ю.А.		

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

Школа Инженерная школа информационных технологий и робототехники  
 Специальность 09.04.03. Прикладная информатика  
 Уровень образования Магистратура  
 Отделение Информационных технологий  
 Период выполнения Весенний семестр 2018 учебного года  
 Форма представления работы:

магистерская диссертация
--------------------------

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН  
 выполнения выпускной квалификационной работы**

Срок сдачи студентом выполненной работы:	
--	--

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
01.02.2018	Получение задания на ВКР	
01.03.2018	Получение задания по социальной ответственности	
10.03.2018	Получение задания по финансовому менеджменту	
19.03.2018	Глава 1. Анализ предметной области	
02.04.2018	Глава 2. Поиск методов и инструментария для решения поставленных задач	
23.04.2018	Глава 3. Реализация проекта	
30.04.2018	Глава 4. Финансовый менеджмент	
07.05.2018	Глава 5. Социальная ответственность	
20.05.2018	Проверка оформления. Заключение.	
31.05.2018	Сдача выполненной работы.	

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Марухина О. В.	к. т. н.		

**СОГЛАСОВАНО:**

Руководитель ООП 09.04.03 Прикладная информатика	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Марухина О.В.	к. т. н.		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И  
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
8KM61	Емельяновой Юлии Андреевне

<b>Школа</b>	Инженерная школа информационных технологий и робототехники	<b>Отделение</b>	Информационных технологий
<b>Уровень образования</b>	Магистратура	<b>Направление/специальность</b>	09.04.03 Прикладная информатика

**Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:**

<i>1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Система поддержки принятия врачебных решений на основе анализа клинико-лабораторных показателей детей и подростков с эндокринопатиями.
<i>2. Нормы и нормативы расходования ресурсов</i>	
<i>3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

<i>1. Оценка коммерческого и инновационного потенциала НТИ</i>	Оценка конкурентоспособности, рассмотрение альтернатив проведения НИ, SWOT анализ, QuaD технология.
<i>2. Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок</i>	Планирование этапов работ, определение трудоемкости, построение диаграммы Ганта.
<i>3. Определение ресурсной, финансовой, экономической эффективности</i>	Сравнительный анализ интегральных показателей эффективности, формирование бюджета НТИ.

**Перечень графического материала (с точным указанием обязательных чертежей):**

<ol style="list-style-type: none"> <li>1. Оценочная карта для сравнения конкурентных технических решений</li> <li>2. SWOT анализ, QuaD технология</li> <li>3. Расчет трудозатрат, диаграмма Ганта</li> <li>4. Бюджет НТИ</li> <li>5. Оценка ресурсной, финансовой и экономической эффективности НТИ</li> </ol>
--

<b>Дата выдачи задания для раздела по линейному графику</b>	10.03.2018
---	------------

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент отделения СГН	Старикова Екатерина Васильевна	к. ф. н.		10.03.2018

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8KM61	Емельянова Юлия Андреевна		10.03.2018

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»**

Студенту:

<b>Группа</b>	<b>ФИО</b>
8КМ61	Емельяновой Юлии Андреевне

<b>Школа</b>	Инженерная школа информационных технологий и робототехники	<b>Отделение</b>	Информационных технологий
<b>Уровень образования</b>	Магистратура	<b>Направление/специальность</b>	09.04.03 Прикладная информатика

**Исходные данные к разделу «Социальная ответственность»:**

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Система поддержки принятия врачебных решений при выборе тактики лечения детей с эндокринопатиями будет использоваться на персональном компьютере. Объектом исследования является рабочее место, представляющее собой помещение с персональным компьютером.
--	--

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

<p><b>1. Производственная безопасность</b></p> <p>1.1. Анализ выявленных вредных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности:</p> <ul style="list-style-type: none"> <li>– физико-химическая природа вредности, её связь с разрабатываемой темой;</li> <li>– действие фактора на организм человека;</li> <li>– приведение допустимых норм с необходимой размерностью (со ссылкой на соответствующий нормативно-технический документ);</li> <li>– предлагаемые средства защиты;</li> <li>– (сначала коллективной защиты, затем – индивидуальные защитные средства).</li> </ul> <p>1.2. Анализ выявленных опасных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности:</p> <ul style="list-style-type: none"> <li>– механические опасности (источники, средства защиты);</li> <li>– термические опасности (источники, средства защиты);</li> <li>– электробезопасность (в т.ч. статическое электричество, молниезащита – источники, средства защиты);</li> <li>– пожаровзрывобезопасность (причины, профилактические мероприятия, первичные средства пожаротушения).</li> </ul>	<p>Анализ выявленных вредных факторов включает:</p> <ul style="list-style-type: none"> <li>– недостаточная освещенность рабочего места</li> <li>– микроклимат</li> <li>– повышенный уровень шума</li> <li>– повышенный уровень электромагнитных излучений</li> <li>– зрительное напряжение</li> <li>– монотонность трудового процесса</li> </ul> <p>Анализ выявленных опасных факторов включает:</p> <ul style="list-style-type: none"> <li>– поражение электрическим током</li> <li>– повышенный уровень статического электричества</li> <li>– короткое замыкание</li> </ul>
<p><b>2. Экологическая безопасность:</b></p> <ul style="list-style-type: none"> <li>– защита селитебной зоны</li> <li>– анализ воздействия объекта на атмосферу</li> </ul>	<p>Анализ негативного воздействия на окружающую среду: утилизация люминесцентных ламп, компьютерной</p>

(выбросы); – анализ воздействия объекта на гидросферу (сбросы); – анализ воздействия объекта на литосферу (отходы); – разработать решения по обеспечению экологической безопасности со ссылками на НТД по охране окружающей среды.	техники и оргтехники
<b>3. Безопасность в чрезвычайных ситуациях:</b> – перечень возможных ЧС при разработке и эксплуатации проектируемого решения; – выбор наиболее типичной ЧС; – разработка превентивных мер по предупреждению ЧС; – разработка действий в результате возникшей ЧС и мер по ликвидации её последствий.	Техногенные: Пожар
<b>4. Правовые и организационные вопросы обеспечения безопасности:</b> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	– ГОСТ 12.2.032 – 78. Система стандартов безопасности труда. Рабочее место при выполнении работ сидя. Общие эргономические требования. – СанПиНом 2.2.2/2.4.1340 – 03. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы.

<b>Дата выдачи задания для раздела по линейному графику</b>	01.03.2018
---	------------

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Авдеева Ирина Ивановна			01.03.2018

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8КМ61	Емельянова Юлия Андреевна		01.03.2018



## Реферат

Магистерская диссертация включает в себя: 107 страниц, 30 рисунков, 27 таблиц, 28 формул, 43 источника, 1 приложение.

Объектом исследования является актуальная в наше время проблема избыточной массы тела у детей и подростков.

Цель исследования – разработка алгоритмической базы системы поддержки принятия врачебных решений при выборе тактики лечения детей с эндокринопатиями.

Задачи:

- анализ существующих методов и подходов обработки многомерных данных;
- оценка информативности клинико-лабораторных показателей;
- формирование алгоритма принятия решения для выбора траектории лечения для детей с эндокринопатиями.

В работе рассмотрено понятие многомерных данных, выполнен обзор литературы и анализ предметной области. Рассчитана информативность клинико-лабораторных показателей для 5 групп лечения, на основе, которой сформировали базу знаний. В качестве метода представления базы знаний выбрано построение дерева решений.

Ключевые слова: многомерные данные, система поддержки принятия врачебных решений, информативность, метод Кульбака.

## **Определения и сокращения**

**Система поддержки принятия решений** – компьютерная автоматизированная система, целью которой является помощь людям, принимающим решение в сложных условиях для полного и объективного анализа предметной деятельности.

**Эндокринопатия** – обобщающее понятие для заболеваний, связанных с нарушением деятельности эндокринных желез.

**Ожирение** – это хроническое нарушение обмена веществ, сопровождающееся избыточным отложением жировой ткани.

**Данные** – это совокупность сведений о событиях, явлениях, зафиксированных на носителе в форме, пригодной для постоянного хранения, передачи и обработки.

**Информативные признаки** – полезная информация для конкретной цели, полученная из исходной информации.

**Дерево принятия решений** – средство поддержки принятия решений, используемое в статистике и анализе данных для прогнозных моделей.

СППВР – система поддержки принятия врачебных решений

НИИ – научно-исследовательский институт

## Оглавление

Введение .....	13
Глава 1 Анализ предметной области .....	15
1.1 Обзор литературы .....	15
1.2 Многомерные данные.....	17
1.3 Описание предметной области .....	19
Глава 2 Поиск методов и инструментария для решения поставленных задач .....	22
2.1 Выбор метода решения .....	22
2.1.1 База знаний.....	22
2.1.2 Дерево решений.....	25
2.2 Выбор инструментария .....	26
2.2.1 RStudio.....	27
2.2.2 Rapid Miner.....	28
Глава 3 Реализация проекта.....	29
3.1 Первичная обработка данных .....	29
3.2 Методы сокращения размерности.....	35
3.2.1 Оценка информативности по Кульбаку .....	39
3.2.2 Оценка информативности клиничко-лабораторных показателей .....	40
3.3 Описание операторов в Rapid Miner .....	48
3.4 Построение дерева решений .....	50
3.5 Алгоритм СППВР для выбора тактики лечения.....	52
Глава 4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.....	54
4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения .....	55
4.1.1 Потенциальные потребители результатов исследования .....	55
4.1.2 Анализ конкурентных технических решений .....	56
4.1.3 Технология QuaD.....	57
4.1.4 Диаграмма Исикавы .....	59
4.1.5 SWOT-анализ .....	59
4.1.6 Оценка готовности проекта к коммерциализации .....	61
4.2 Определение возможных альтернатив проведения научных исследований .....	62
4.3 Планирование научно-исследовательских работ .....	63
4.3.1 Структура работ в рамках научного исследования .....	63
4.3.2 Определение трудоемкости выполнения работ.....	64

4.3.3	Разработка графика проведения научного исследования.....	65
4.3.4	Бюджет научно-технического исследования (НТИ).....	67
4.3.4.1	Расчет материальных затрат .....	67
4.3.4.2	Основная заработная плата исполнителей темы.....	68
4.3.4.3	Дополнительная заработная плата исполнителей темы.....	69
4.3.4.4	Отчисления во внебюджетные фонды (страховые отчисления) .....	69
4.3.4.5	Накладные расходы .....	70
4.3.4.6	Формирование бюджета затрат научно-исследовательского проекта.....	70
4.4	Определение ресурсной, финансовой, бюджетной, социальной и экономической эффективности исследования.....	71
Глава 5	Социальная ответственность .....	74
5.1	Производственная безопасность .....	75
5.1.1	Анализ вредных факторов .....	75
5.1.1.1	Недостаточная освещенность рабочего места .....	75
5.1.1.2	Микроклимат .....	78
5.1.1.3	Повышенный уровень шума .....	79
5.1.1.4	Повышенный уровень электромагнитных излучений.....	79
5.1.1.5	Зрительное напряжение и монотонность трудового процесса .....	80
5.1.2	Анализ опасных факторов .....	81
5.2	Экологическая безопасность .....	82
5.3	Безопасность в чрезвычайных ситуациях .....	83
5.4	Правовые и организационные вопросы обеспечения безопасности.....	84
	Заключение.....	86
	Список источников:.....	88
	Приложение А.....	92

## Введение

В настоящее время вся информация хранится в электронном виде в базах данных и занимает большие объемы. За счет того, что объем данных увеличивается в медицине и время на принятия решений сокращается, возникают врачебные ошибки. Решением данных проблем может быть внедрение в область медицины различных информационных технологий.

Объектом исследования является актуальная в наше время проблема избыточной массы тела у детей и подростков. Предмет исследования – многомерный массив данных клинико-лабораторных показателей пациентов, страдающих ожирением разной степени.

Цель исследования – разработка алгоритмической базы системы поддержки принятия врачебных решений при выборе тактики лечения детей с эндокринопатиями.

Задачи выпускной квалификационной работы:

- обзор литературных источников и публикаций для выявления отечественных и зарубежных практик решения задач обработки многомерных данных и получения новых знаний;
- обзор инструментария работы с многомерными данными, обоснование выбора инструментария для решения задачи ВКР;
- первичная обработка и подготовка данных – анализ выбросов и пропущенных значений;
- снижение размерности многомерных данных за счет выявления и отбора информативных признаков;
- формирование набора правил (базы знаний), определяющих выбор траектории лечения, оценка построенной модели;
- формирование алгоритма принятия решения о выборе траектории лечения для детей с эндокринопатиями.

Разработка систем поддержки принятия врачебных решений актуальна по причине частых диагностических ошибок и неправильных назначений врачей. Массовая компьютеризация здравоохранения делает исключительно перспективным создание систем для алгоритмической и информационной поддержки врачебных решений.

Актуальность исследований обусловлена тремя аспектами:

- правительство РФ ставит приоритетным проект РФ «Совершенствование процессов организации медицинской помощи на основе внедрения информационных технологий»;

- важность решения задачи стабилизации доли населения России с ожирением, особенно это важно для детей и молодежи;
- большое количество медицинских данных требует оперативной обработки.

Новизна данного исследования заключается в создании нового алгоритма принятия решения о выборе траектории лечения для детей с эндокринопатиями (на основе анализа клинико-лабораторных показателей).

Построенная на основе этого алгоритма система поддержки принятия решений проходит опытную эксплуатацию на базе НИИ Курортологии и физиотерапии.

По теме исследования было сделано 6 публикаций и 4 доклада на конференциях.

# Глава 1 Анализ предметной области

## 1.1 Обзор литературы

В настоящее время медицинские задачи могут решаться с помощью активно развивающихся информационных технологий, охватывающих практически все стороны деятельности медицинских организаций, включая управление ресурсами, лечебным процессом и оказание медицинской помощи [1]. Одним из наиболее перспективных направлений является разработка систем поддержки принятия врачебных решений (СППВР). Автоматизированные системы оказываются наиболее эффективными в лечении по сравнению с эмпирическими способами, усилия многих специалистов в области информационных технологий направлены на проектирование именно такой системы, которая сможет алгоритмизировать процесс принятия решения, основываясь на знаниях экспертов.

СППВР служат основой для создания принципиально новых подходов к решению традиционных задач в здравоохранении, таких как наблюдение пациента, консультация, принятие решение по дальнейшему лечению [2, 3]. В ходе анализа научных публикаций по данной теме выявлено, что разработки и исследования в этой области ведутся во всем мире и в различных направлениях. В здравоохранении, например, авторами Атьков О., Кудряшов Ю., Прохоров А. [4] была разработана система для использования в области сердечно-сосудистых заболеваний, система помогает врачу выбирать клинические решения при работе с пациентами по восьми сердечно-сосудистым заболеваниям (острый коронарный синдром, артериальная гипертония и другие). Найданов Ч.А. [5] в своей статье описал СППВР для получения рекомендаций по возникновению и снижению риска для появления критических состояний в предметной области «Деформации и дегенеративные заболевания позвоночника». Так же ведутся разработки врачебных систем в дерматовенерологии, фармакотерапии, кардиологии [6-9]. СППВР применима в лечебно-диагностическом процессе больных с сахарным диабетом [10], для диагностики и терапии артериальной гипертензии, для анализа больных с ишемической болезнью сердца [11, 12].

Авторы различных статей анализируют возможности применения в СППВР нечеткой классификации, генетических алгоритмов, методов оценки рисков [3,7]. Многие системы находятся не только на стадии разработки, но уже успешно внедрены. В Российский онкологический научный центр имени Н.Н. Блохина внедрена система для решения задач диагностики острых лейкозов. Данная система позволяет врачу проводить оценку характеристик клеток крови, делать заключения касательно типа клетки крови

[13]. Ещё один пример – информационная система «Перинатальный мониторинг». Она является региональной системой поддержки принятия решений и внедрена в 82 учреждениях здравоохранения Красноярского края [3]. Существуют некоторые экспертные системы в области медицины: MYCIN – диагностирование бактерий и диагностика заболеваний свертываемости крови; DENDRAL – определение молекулярной структуры; CASNET – диагностика глаукомы; DXplain – поддержка клинических решений; Germwatcher – диагностика инфекций; Puff – анализ нарушения дыхания. Среди российских разработок наиболее известны экспертные системы: ДИН (распознавание неотложного состояния) и ДИАГЕН (диагностика наследственных болезней у детей).

Эндокринопатия – обобщающее понятие для заболеваний, связанных с нарушением деятельности эндокринных желез. Симптомами эндокринопатий могут быть: увеличение роста, превышающее норму, увеличение отдельных частей тела, слабость мышц и быстрое утомление от привычных физических нагрузках, увеличение веса, изменения в крови, нарушение работы сердца. Формы эндокринопатии: синдром Шихана (расстройство организма в результате гибели части тела), акромегалия (усиленный рост скелета и внутренних органов), несахарный диабет (болезнь, причиной которой становится нехватка в организме антидиуретического гормона, отвечающего за водный обмен в организме), врожденный гипотиреозом (недостаточность гормонов щитовидной железы). Одной из наиболее распространенных патологий эндокринопатии является ожирение. Актуальной проблемой, как в России, в Соединенных Штатах, так и во всем мире является область медицины, связанная с ожирением. Ожирение – это хроническое нарушение обмена веществ, сопровождающееся избыточным отложением жировой ткани. Первичное ожирение возникает в результате неправильного питания или передается по наследству. Вторичное – вызывается приобретенными заболеваниями, например, нарушение эндокринной системы. Рост заболеваемости данной патологией отмечается не только у взрослого населения, но и в детском возрасте [14, 15]. Последствиями ожирения могут быть сахарный диабет второго типа, сердечно-сосудистые патологии, онкологические заболевания, нарушения костно-мышечного аппарата, артериальной гипертензии, ишемической болезни сердца, заболеваний печени, психологические проблемы у детей и взрослых [16].

Считается, что к началу XXI века избыточную массу тела имело около 30% населения планеты [17]. В журнале The Lancet опубликована статья, в которой Международная группа ученых сообщила статистические данные по ожирению детей за последние 40 лет. Отмечается, что одной из крупнейших угроз здоровью людей на



планете в настоящее время является избыточная масса. Авторы работы использовали данные по массе тела 31,5 миллиона человек из 200 стран мира в возрасте от 5 до 19 лет за период с 1975 по 2016 год. Оказалось, что в 1975 году 11 миллионов детей и подростков имели ожирение. В 2016 же году уже 124 миллиона [18]. Что касается Томской области, в одном исследовании ученые обследовали 524 студентов. Оценка производилась по следующим показателям: индекс массы тела, гигиена питания и физическая активность. В результате исследования оказалось, что 86 % студентов-юношей имеют нормальную массу тела, 4 % – недостаток, избыток веса – 9 %, ожирением страдает 1 % [19].

В качестве основного показателя, используемого для оценивания избыточного веса, широко применяется индекс массы тела (ИМТ). Однако при отнесении пациента к категории с ожирением, следует учитывать и другие показатели, ИМТ следует считать приблизительным критерием, так как у разных людей он может соответствовать разной степени полноты [20]. Из обзора видно, что такие темы, как СППВР в области здравоохранения, лечение детей с избыточной массой тела являются актуальными на сегодняшний день.

## **1.2 Многомерные данные**

Данные – это совокупность сведений о событиях, явлениях, зафиксированных на носителе в форме, пригодной для постоянного хранения, передачи и обработки. В информационных системах они могут быть представлены в различных формах. Описание объектов содержат все доступные наблюдению, измерению параметры, характеристики, признаки, поэтому любой исследуемый объект характеризуется большим числом признаков. В зависимости от количества признаков данные могут быть одномерными, двумерными и многомерными. Многомерные данные содержат информацию о трех или более признаках для каждого объекта. Этими данными можно воспользоваться для получения информации о зависимостях между признаками, о том насколько они взаимосвязаны, для предсказания значения одной переменной на основании значений других [21]. Хранящиеся в электронном виде данные могут иметь большое количество скрытых закономерностей, которые являются важными для принятия стратегических решений. Большое количество данных требует трудоемких работ при их обработке. Следовательно, существует необходимость анализировать подобного рода данные, хранить только те, которые несут в себе необходимую информацию и представлять новые знания в удобной форме для восприятия человеком.

Многомерный статистический анализ – раздел математической статистики, изучающий методы сбора и обработки многомерных статистических данных, их систематизации и обработки с целью выявления характера и структуры взаимосвязей между компонентами исследуемого многомерного признака, получения практических выводов [22]. За последнее время анализ многомерных данных стал активно развивающимся и применяющимся практически во всех областях исследований, он является одной из наиболее востребованных междисциплинарных областей знания.

Любая обработка информации в различных областях, таких как медицина, банковское дело, телекоммуникации, молекулярная генетика посвящена конкретным целям, например, в медицине – это исследование болезней, лечение, анализ, постановка диагноза у больного. При обращении пациента в медицинское учреждение, врачи проводят обследования и назначают множество анализов, в результате чего, врачи работают с многомерными данными. Раньше диагноз ставился, опираясь на опыт и интуицию врача, но с развитием технологий такая задача может быть автоматизирована и решаться с использованием систем поддержки принятия решений. При выборе решения отнесения объекта к классу, могут возникать проблемы, такие как установление обобщенного признака и определение важности признаков, отражающих свойства объектов [23]. Поэтому из множества всех признаков, необходимо выявить те, которые несут полезную информацию. Признаки, присущие пациенту, которые получены и проанализированы, являются информативными признаками, с помощью которых можно распознать заболевание или его отсутствие у пациента.

Система поддержки принятия решений – это компьютерная система, которая путем сбора и анализа информации может влиять на процессы принятия решений в различных областях человеческой деятельности. В области здравоохранения используется термин «системы поддержки принятия врачебных решений» (СППВР). Обработкой информации и выявлением в ней моделей и тенденций, которые помогают принимать решения, занимается интеллектуальный анализ данных, на котором основываются СППВР.

Выделяют несколько преимуществ таких систем. Системы, основываясь на собранных данных, помогают врачу определить ожидаемые диагностические решения. Количество диагностических гипотез, принимаемых к рассмотрению, становится менее зависимым от личного опыта и знаний врача, благодаря чему уменьшается вероятность пропустить и не выявить какое-либо заболевание. База данных в этих системах учитывает современные представления по исследуемой проблеме, это позволяет ей реализовать обучающую функцию, подсказывая врачу не только необходимый объем клинического,

лабораторного и инструментального обследования пациента, но также и алгоритмы принятия решений, обобщающие знания экспертов по данной проблеме.

### 1.3 Описание предметной области

Исследование выполнялось в рамках совместного инициативного проекта Томского НИИ (научно – исследовательский институт) Курортологии и физиотерапии (ТНИИКиФ) и Научно-учебной лаборатории информационных технологий в социальных и медицинских исследованиях (ИШИТР ТПУ), связанного с исследованиями в области обработки данных (лечение избыточной массы тела у детей и подростков). Основные направления ТНИИКиФ: научная деятельность, лечение и оздоровление. В клинике НИИ курортологии и физиотерапии проводят восстановительное лечение, медицинскую профилактику и реабилитацию пациентов неврологического, гинекологического, урологического, педиатрического, общетерапевтического профилей.

На базе детского отделения ТНИИКиФ было проведено клиническое исследование в период с 2006 по 2013 год, в котором приняли участие 464 ребенка, страдающих ожирением в возрасте от 10 до 15 лет. Данные представлены в виде многомерной таблицы, содержащей 212 объекта (пациента), характеризующиеся 55 признаками (значениями клинико-лабораторных показателей до и после лечения). В зависимости от степени ожирения пациенту назначается комплекс лечения (всего 5 комплексов). Признаки разделены на несколько физиологических групп (рис.1).



Рис.1 Физиологические группы показателей

Перечислим основные показатели, на основе которых проводится исследование:

- **Клиника:**

*ОТ* – объем талии

*ОБ* – объем бедер

*Масса* – масса тела

*Избыток, %* – избыток массы тела

*ТМТ* – тощая масса тела

*ИМТ* – индекс массы тела

- **Сердечно-сосудистая система:** САД, мм рт.ст – систолическое артериальное давление и ДАД, мм рт.ст – диастолическое артериальное давление

- **Физическая работоспособность:**

*НОМА* – индекс инсулинорезистентности Хома

*ТФН* – толерантность к физической нагрузке

*Дв.пр.* – двойное произведение (насыщение миокарда кислородом)

*Общая р-ть* – общая работоспособность

- **Липидный обмен:**

*Общие липиды* – общие липиды крови

*ТАГ* – содержание триацилглицеридов в сыворотке крови

*αХС* – содержание альфа-холестерина в сыворотке крови

*ЛПНП* – липопротеиды низкой плотности

*ЛПОНП* – липопротеиды очень низкой плотности

- **Биохимия крови:**

*ЩФ* – щелочная фосфатаза

*Каталаза* – каталаза сыворотки крови

*МДА* – малоновый диальдегид в сыворотке крови

*Церулоплазмин* – содержание сиаловых кислот в сыворотке крови

- **Гормональный статус:**

*Кортизол* – содержание кортизола в сыворотке крови

*Инсулин* – содержание инсулина в сыворотке крови

*ИЛ-6* – содержание интерлейкина 6

*ИЛ-4* – содержание интерлейкина 4

*Лептин* – содержание лептина

*ТТГ (0,23-3,4)* – содержание тиреотропного гормона в сыворотке крови

*Т3 (1,0-2,8)* – содержание трийодтиронина в сыворотке крови

*ФНО (не >2,5 нг/мл)* – содержание фактора некроза опухоли

*АТ ТПО (до 30)* – антитела к тиреопероксидазе в сыворотке крови

*Т4 (53-58)* – содержание тетраiodтиронина

- **Иммунологический статус**

*IgA* – концентрация иммуноглобулина А в сыворотке крови

*IgG* – концентрация иммуноглобулина G

*IgM* – концентрация иммуноглобулина M

*Лизоцим* – активность лизоцима

*ЦИК, Тхел* (Т-хелперы 33-50%). *Т-суп* (Т-супрессоры 16-39 %), *В-лим* (В-лимфоциты 6-23%) – циркулирующие иммунные комплексы в сыворотке крови

- **Состояние калликреин-кининовой системы:**

*КК* – уровень каллекриина

*ПИ* – протромбиновый индекс

*АПФ* – ангиотензинпревращающий фермент

*МГ* – макроглобулин

- **Окислительная способность плазмы крови:**

*NO* – содержание оксида азота в крови

*окЛПНП* – окисление липопротеидов низкой плотности в сыворотке крови

Помимо выше-перечисленных показателей, массив данных содержит информацию о принадлежности пациента к группе лечения (1-5). Для решения поставленной задачи необходимо построить такую систему, которая могла бы для вновь прибывшего пациента определить для него траекторию лечения, то есть одну из 5 групп, основываясь на его первичных клинико-лабораторных показателях.

## **Глава 2 Поиск методов и инструментария для решения поставленных задач**

### **2.1 Выбор метода решения**

#### **2.1.1 База знаний**

Представление знаний, а так же разработка систем, основанных на знаниях, является одним из основных направлений в области изучения искусственного интеллекта. Данное направление связано с разработкой моделей представления знаний, созданием баз знаний, образующих ядро экспертных систем. Это направление ещё называют инженерия знаний.

Знания – это закономерности, принципы и законы предметной области, полученные в результате практической деятельности и профессионального опыта, позволяющие специалистам ставить и решать задачи в этой области. Знания основаны на данных, полученных эмпирическим путем.

База знаний - совокупность фактов и правил вывода, допускающих логический вывод и осмысленную обработку информации [24].

База знаний в отличие от базы данных содержит в себе не только фактическую информацию, но и правила вывода, позволяющие делать автоматические умозаключения об уже имеющихся или вновь вводимых фактах и тем самым производить осмысленную обработку информации. База знаний является основным компонентом интеллектуальных и экспертных систем. Содержимое базы знаний подразделяют на факты и правила. Факты – это простые утверждения о характеристиках объекта, а правила служат для зависимостей между фактами.

Системы, основанные на знаниях, реализуются на базе следующих интеллектуальных алгоритмов.

1. Экспертные системы – системы, способные принимать решения, схожие с решениями экспертов.

При создании базы знаний можно выделить несколько этапов: выбор предметной области, для которой будет разрабатываться база знаний; извлечение знаний; следующий этап – структурирование собранных данных, выделяются основные понятия и взаимосвязи между ними, структура информации на входе и выходе, алгоритм принятия решений; на последнем этапе полученные знания описываются на одном из языков представления знаний.

Так же при создании базы знаний необходимо выбрать модель представления знаний, ведь целью представления знаний является организация нужной информации в форму, удобной для восприятия информации и для принятия решений. Существуют различные модели представления знаний, но большинство из них могут относиться к следующим классам [25]:

- Продукционные модели – в таких моделях знания представляются в виде: «Если (условие), то (действие)». Условие – предложение, по нему происходит поиск в базе знаний, действие – выполняются, если поиск успешно осуществился.

- Семантические сети – позволяют описывать свойства и отношения объектов событий или действий с помощью узлов и дуг в виде ориентированного графа. Объектам соответствуют узлы сети, а отношениям – соединяющие их дуги.

- Фреймы – это структура или абстрактный образ для описания ситуации, состоящая из характеристик и их значений. Характеристики называются слотами, а значения – заполнителями слотов. Слот может содержать не только конкретное значение, но и имя функции, позволяющей вычислить его по заданному алгоритму, а также одно или несколько правил, с помощью которых это значение можно найти. Совокупность фреймов представляет собой иерархическую структуру, на верхнем уровне иерархии находится фрейм, содержащий наиболее общую информацию, истинную для всех остальных фреймов. Фреймы обладают способностью наследовать значения характеристик своих родителей, находящихся на более высоком уровне иерархии[26].

## 2. Нечёткая логика;

Базу знаний можно построить, используя нечеткий логический вывод Мамдани или вывод Сугено. Данные алгоритмы можно реализовать в пакете Fuzzy Logic Toolbox вычислительной среды MATLAB. Так же известны алгоритмы нечеткого вывода Ларсена, Цукамото.

Этапы нечеткого вывода:

- Создание правил в виде ПРАВИЛО\_1: ЕСЛИ "Условие\_1" ТО "Заключение\_1".
- Фаззификация входных переменных - каждому значению отдельной входной переменной ставится в соответствие значение функции принадлежности для переменной  $x$ .
- Нечёткий вывод или нахождение степеней истинности (уровней «отсечки») для предпосылок («Условие\_1», «Условие\_2», ..., «Условие\_n») каждого из правил.

- Активизация подзаключений. Для этой цели можно использовать один из модифицированных методов нечёткой композиции (  $\min$  – активизация,  $\text{prod}$  – активизация,  $\text{average}$  – активизация).

- Аккумуляирование заключений нечетких правил продукций. Дефаззификация в системах нечеткого вывода представляет собой процедуру или процесс нахождения значения для каждой из выходных лингвистических переменных множества. На этапе дефаззификации может быть применён метод центра тяжести или метод биссектрисы площади [27].

### 3. Нейронные сети;

Наибольшее распространение в настоящее время получили архитектуры нечеткой нейронной сети вида ANFIS и TSK для создания базы знаний. ANFIS — это искусственная нейронная сеть, основанная на нечеткой системе вывода Такаги-Сугено. Вывод такой системы соответствует набору нечетких правил «если-то», которые имеют способность к обучению. Простой контроллер Такаги-Сугено с двумя входами и двумя правилами: ЕСЛИ  $P_{11}(x_1)$  И  $P_{12}(x_2)$  ТО  $f_1(x_1, x_2)$ . ЕСЛИ  $P_{21}(x_1)$  И  $P_{22}(x_2)$  ТО  $f_2(x_1, x_2)$ . Задача сетей TSK и Ванга-Менделя состоит в отображении пар данных  $(x, d)$ , чтобы ожидаемое значение, соответствующее входному параметру вектору  $x$ , формировалось выходной функцией сети  $y(x)$ .

### 4. Генетические алгоритмы.

Генетический алгоритм – алгоритм поиска, применяется в решениях задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искоемых параметров с использованием механизмов, аналогичных естественному отбору в природе. Получившаяся случайным образом начальная популяция решений подвергается воздействию генетических операций (например, селекция, скрещивание и мутация). В результате чего производится новая популяция решений с улучшенными значениями целевой функции. Процесс повторяется до нахождения оптимума [28].

Для решения интеллектуальных задач используются специальные представления знаний – PROLOG, LIST, SMALLTALK. Средством создания системы базы знаний могут быть языки программирования – C, Basic, Pascal. Так же популярным является использование оболочек для создания экспертных систем, например, Карра, Exsys Developer и другие, базы знаний которых можно наполнять конкретными знаниями.

5. Другим методом представления базы знаний, использующееся в статистике является **дерево принятия решений**. Основные компоненты дерева: листья и ветки. Атрибуты, от которых зависит целевая функция, записываются на ветках,



значения целевой функции - в листьях. Дерево решений – это способ представления логических правил в иерархической структуре.

Для решения задачи формирования базы знаний был выбран метод построения дерева решений, как один из наиболее популярных и дающих хороший результат.

### 2.1.2 Дерево решений

Общий алгоритм построения дерева решений включает следующие пункты:

- Выбор очередного атрибута  $A$ , помещение данного атрибута в корень. Для выбора атрибута существуют различные алгоритмы, например, ID3, где выбор атрибута происходит на основании прироста информации ( $gain$ ), либо на основании критерия Джини; алгоритм C4.5, где выбор атрибута происходит на основании нормализованного прироста информации ( $gain\ ratio$ ); алгоритм CART предназначен для построения бинарного дерева решений, каждый узел дерева при разбиении имеет только двух потомков. В результате работы этих алгоритмов могут получаться слишком детализированные деревья, которые при их дальнейшем использовании дают множество ошибок. Это связано с явлением переобучения. Для сокращения деревьев используется отсечение ветвей ( $pruning$ ).

- Для всех значений  $i$  выбранного атрибута оставляем только те тестовые примеры, у которых значения атрибута  $A$  равно  $i$ .

- Рекурсивное построение дерева в этом потомке. Рекурсивное разбиение означает неоднократное деление на значения входных атрибутов [29].

В каждом рекурсивном разбиении можно выделить три шага.

1. В начале имеем обучающую выборку  $T$  и целевой атрибут  $G$ , относительно которого будет строиться модель дерева решений. Далее определяем атрибут  $A$  для разбиения. Например, в параметрах оператора «Decision Tree» в Rapid Miner выберем критерий энтропии ( $gain\ ratio$ ). Энтропия – мера неопределенности появления какого-либо значения в множестве. В данной критерии вычисляется энтропия для всех атрибутов, а для разбиения выбираются атрибуты с наименьшей энтропией. Для подсчёта энтропии атрибута берется множество его значений  $S$ , встречающихся в выборке  $T$ .

2. Энтропия для множества  $S$  является суммой с противоположным знаком всех относительных частот появления состояния с номером  $i$ , умноженных на их двоичные логарифмы (формула 1):

$$H(S) = -\sum_{i=1}^n p_i \log_2 p_i, \quad (1)$$

где  $S$  – множество, для которого вычисляется энтропия,

$n$  – количество значений во множестве  $S$ ,

$p_i$  – отношение количества элементов в классе  $i$  к количеству элементов во множестве  $S$ .

Подсчитывается  $H(S)$  для каждого атрибута из выборки  $T$  и выбирается атрибут с наименьшим значением.

3. После того как выбран атрибут  $A$ . Пусть  $S$  - множество значений атрибута. Теперь множество, которое подаётся на входе оператора, разбивается на подмножества по одному для каждого уникального значения атрибута  $A$ . Если  $A$  численный атрибут, то количество поддеревьев может быть достаточно велико, такая проблема решается разбиением на непересекающиеся диапазоны. Например, пусть целевой атрибут  $G$  имеет два значения 0 и 1. Каждое значение из множества можно сравнить со значениями целевого атрибута. Например, если значение ближе к 0, то отмечаем кружком, если ближе к 1, то галочкой, в результате можно построить прямую со всеми значениями. Далее на прямой нужно отметить непересекающиеся диапазоны, при этом диапазон должен включать объекты с одинаковым значением целевого параметра  $G$ .

4. После выбора атрибута и выбора способа разбиения множества разбиваем исходную выборку  $T$ , получаем совокупность поддеревьев  $S_1, \dots, S_r$ , где  $r$  количество уникальных значений у номинального признака и число непересекающихся диапазонов у числового признака. Такой алгоритм повторяется для каждого  $S$  множества отдельно.

## 2.2 Выбор инструментария

В настоящее время на рынке программных продуктов имеется достаточно большой выбор средств статистической обработки данных. Наиболее популярные статистические пакеты:

- Statistica – программный пакет для статистического анализа, разработанный компанией StatSoft, сейчас пакет принадлежит корпорации Dell. Функции: анализ и управление данными, визуализация данных.

- SPSS (Statistical Package for the Social Sciences) с 2009 года называется PASW Statistics (Predictive Analytics SoftWare) – включает в себя следующие возможности: хранение данных, использование переменных разных типов, первичная описательная статистика, таблицы, графики, создание отчетов.

• Stata – пакет статистических прикладных программ аналогичный SAS, SPSS, RATS, предназначенный для решения статистических задач в таких областях как экономика, медицина, биология, социология. Как и в выше перечисленных пакетах Stata включает в себя: базовые статистические методы, линейные модели, непараметрические и многомерные методы, кластерный анализ, графику, операции над данными. В таблице 1 представлено сравнение статистических пакетов.

Таблица 1. Анализ статистических пакетов

	Statistica	SPSS	Rapid Miner	Stata	RStudio
Наглядное представление данных (диаграммы, графики)	+	+	+	+	+
Бесплатная программа	-	-	+	-	+
Возможность импорта, экспорта данных из Microsoft Excel	+	+	+	+	+
Поддержка операционных систем	Windows	Windows	Windows, Mac OS X и Linux	Windows	Windows, Mac OS X и Linux

В ходе анализа для выполнения задач исследования были выбраны следующие статистические пакеты: RStudio, Rapid Miner.

### 2.2.1 RStudio

RStudio – бесплатная среда разработки программного обеспечения с открытым исходным кодом для языка программирования R. Скриптовый язык R используется для статистической обработки данных и работы с графикой. R поддерживает широкий ряд методов и способен к расширению с помощью пакетов. Пакеты – это библиотеки для работы специфических функций или специальных областей применения. RStudio доступна в двух версиях: RStudio Desktop, в которой программа выполняется на локальной машине как обычное приложение; и RStudio Server, в которой предоставляется доступ через браузер.

Достоинства программного обеспечения:

- Свободное распространение пакета и свободный код;
- Большой выбор статистических методов и более 7000 тыс. проверенных пакетов;

- Быстрый переход к функциям;
- Интегрированный отладчик для исправления ошибок и интегрированная документация R;

- Автоматическое завершение кода.

Недостатки:

- Недостаточный объем документации на русском языке;
- Трудность обучения программе из-за большого объема команд;
- Медленная работа программы при использовании функций с циклами, со списками и таблицами данных. Данные функции работают медленнее, чем их аналоги в коммерческих пакетах.

### 2.2.2 Rapid Miner

Rapid Miner – бесплатная комплексная система для решения задач машинного обучения интеллектуального анализа данных. Первая версия была разработана в 2001 году группой Искусственного Интеллекта технологического университета в Дортмунде. Имеет большой набор алгоритмов для обработки и анализа данных, включая работу с большими массивами данных.

Достоинства:

- Хороший графический интерфейс пользователя. Процесс представляет собой набор операторов, соединенных последовательно между собой. Имеется более 400 различных операторов (операторы обучения по прецедентам, работы с признаками, визуализации, загрузки и сохранения данных).

- Наличие пошагово учебника по работе с данными в Rapid Miner.

- Наличие инструмента ETL подготовки данных, включающий в себя извлечение данных из внешних источников и загрузку в хранилище данных.

- Расширение с помощью языка R и интеграция в себя операторов WEKA (набор средств визуализации и алгоритмов для интеллектуального анализа данных).

К недостаткам можно отнести то, что не все задачи могут выполняться на сервере, это характерно, когда модель требует всей базы, чтобы выполнить алгоритм частями.

## Глава 3 Реализация проекта

### 3.1 Первичная обработка данных

Первичная обработка данных направлена на упорядочивание информации об исследуемом объекте. На данной стадии исходные сведения группируются по тем или иным признакам, заносятся по таблицам. Исходные данные необходимо анализировать, выявлять информацию с некорректными данными, предоставлять в удобной для работы форме, хорошо представленные сведения дают исследователю понять общую картину всей совокупности данных. Обработка может быть направлена на решение следующих задач:

- упорядочивание исходных данных;
- обнаружение и устранение ошибок, некорректно введенных данных, выбросов, пробелов в сведениях;
- выявление скрытых закономерностей и связей.

Обработка данных является важной составляющей в анализе данных, ведь, качество обработки влияет на конечный результат исследования. Очень часто в исследуемых данных встречаются выбросы. В статистике выбросами являются значения, которые резко отличаются от других значений в собранном наборе данных. Причинами выбросов могут быть ошибки измерений, ошибки ввода данных, действия конкретных факторов на объект. Диагностируя выбросы, необходимо быть предельно внимательным, так как, если исключить нужные для анализа данные, можно допустить ошибку. Наиболее распространенный способ определения выбросов – диапазон трех стандартных отклонений вокруг среднего значения. Но данный метод теряет смысл, если функция распределения отличается от нормальной или объем выборки недостаточно большой. Диагностируя выбросы, легко допустить ошибку, исключив нужные для анализа данные. Определить выбросы можно, используя гистограммы, точечные и ящичные диаграммы, диаграммы индивидуальных значений, рассеивания, временных рядов.

Для определения выбросов воспользуемся графическим анализом наблюдений, с помощью которого легко распознать значения, лежащие за пределами наблюдений, а именно диаграммой размахов («ящик с усами»). Данная диаграмма может быть построена в различных программных продуктах, наиболее известные из них RStudio, Statistica, Excel. В RStudio для построения диаграммы размахов служит функция **boxplot()**.

В RStudio при построении данной диаграммы используются устойчивые оценки центральной тенденции (медиана) и разброса (интерквартильный размах, ИКР). Нижний

квартиль – это величина, выше которой лежит 25% значений из набора данных. Верхний квартиль – величина, ниже которой лежит 75% значений. Верхний ус - расстояние от верхней границы ящика до наибольшего выборочного значения, находящегося в пределах расстояния  $1.5 \cdot \text{ИКР}$  от этой границы, нижний ус аналогично [30]. Наблюдения, находящиеся за пределами усов считаются выбросами.

Описание диаграммы при использовании данной функции представлено на рис.2.

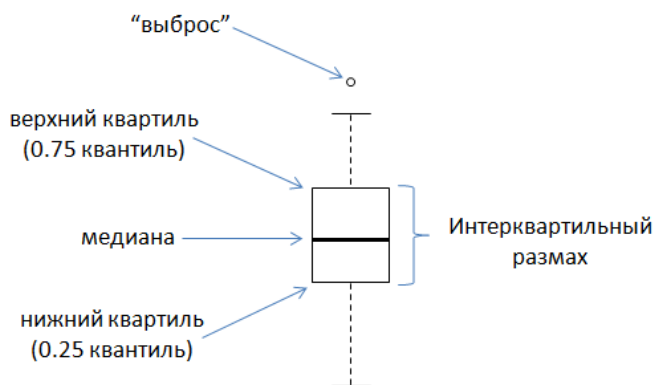


Рис.2 Схема диаграммы размахов

Так как исходные данные хранятся в Excel файле, для считывания данных из файла в RStudio необходимо подключить библиотеку **library(readxl)**. Для построения диаграммы размахов воспользуемся функцией **boxplot(TMT ~ group, data = data)**, пусть показатели (*TMT*, *ИМТ* и другие) будут зависеть от переменной *group*. Например, возьмем показатель *тощая масса тела (TMT)*. Реализация функции представлена на рис.3

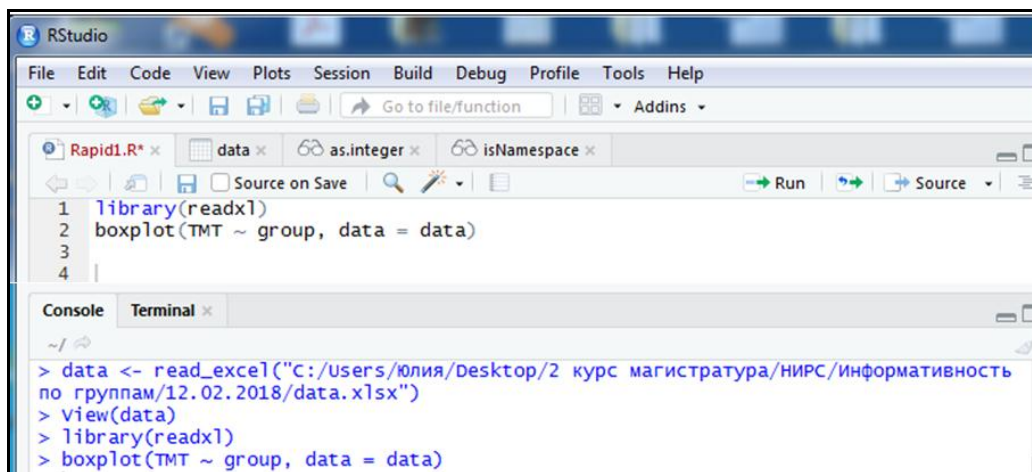


Рис.3 Построение диаграммы выбросов в RStudio

На рис.4 представлена диаграмма размахов признака *TMT*. На диаграмме по вертикали расположены значения показателей, по горизонтали 5 групп лечения.

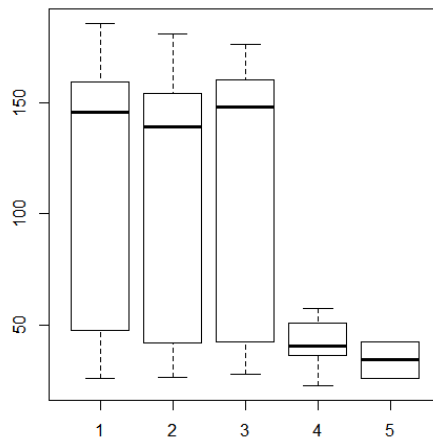


Рис.4 Диаграмма размахов показателя *TMT*

На рисунке видно, что показатель не имеет выбросов в значениях, диаграмма наглядно показывает, что для каждой группы лечения медиана разная. При анализе выбросов показателя *индекса массы тела (ИМТ)* получились выбросы в первой группе 47,3, во второй группе 18,5, в третьей группе 44,3 (рис.5). На рис.6 представлена диаграмма после удаления выбросов.

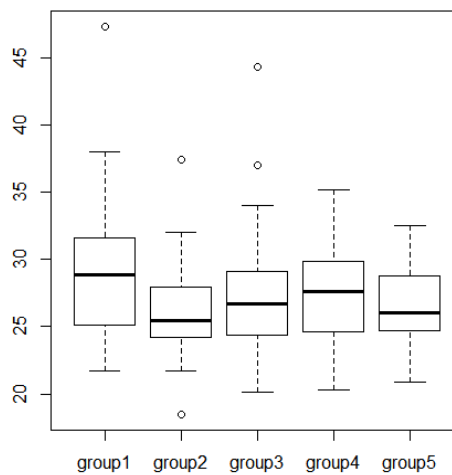


Рис.5 Диаграмма размахов показателя *ИМТ*

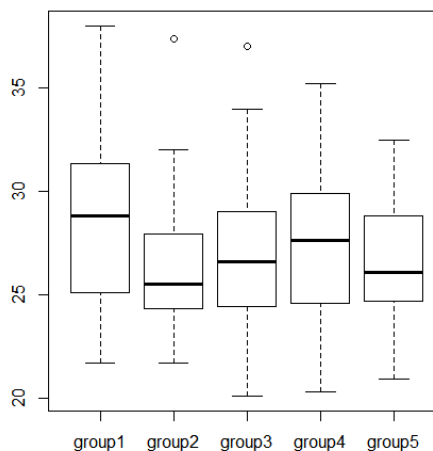


Рис. 6 Диаграмма размахов *ИМТ* после удаления выбросов

Оставшиеся два значения лежат в пределах первой группы, значения являются допустимыми, возможно, пациенты могут относиться к первой группе. При первичной обработке данных, попадались показатели без выбросов, с множеством выбросов, с одним или двумя. Примером диаграммы размахов с одним выбросом является показатель *систолическое артериальное давление (САД)*, получился один выброс в третьей группе со значением 80. На рис. 7, 8 представлены диаграмма размахов САД до и после удаления выброса соответственно.

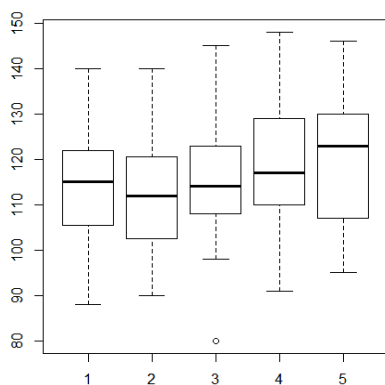


Рис.7 Диаграмма размахов показателя САД

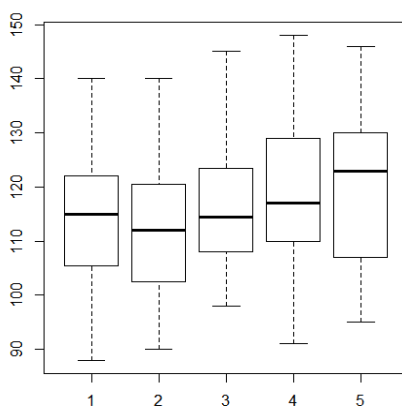


Рис.8 Диаграмма размахов показателя САД без выброса

Далее на рис. 9, 10 представлен анализ выбросов для показателя *циркулирующие иммунные комплексы (ЦИК)*.

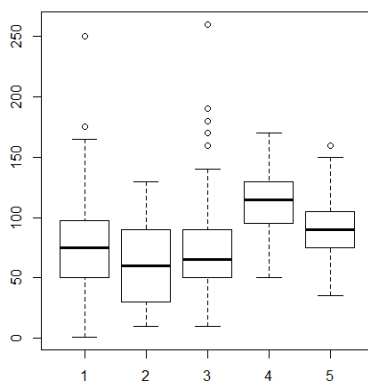


Рис.9 Диаграмма размахов показателя ЦИК



Из диаграммы видно, что в первой группе имеются следующие выбросы: 250, 175, в третьей группе: 170, 190, 260, 180.

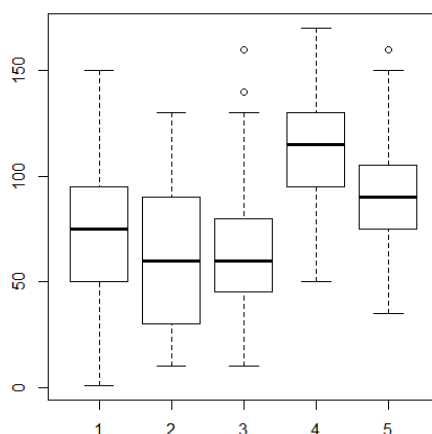


Рис.10 Диаграмма размахов показателя *ЦИК* после удаления выбросов

Анализ выбросов производится совместно с медицинскими экспертами, оставшиеся значения не являются выбросами, они считаются допустимыми значениями. Эти значения не следует удалять, так как удаление может привести к искажению результатов, удаляются только те, которые одобрены медицинскими специалистами. Следующий пример, показывает выброс, получившийся в результате ошибки ввода данных, опечатки специалиста (рис.11).

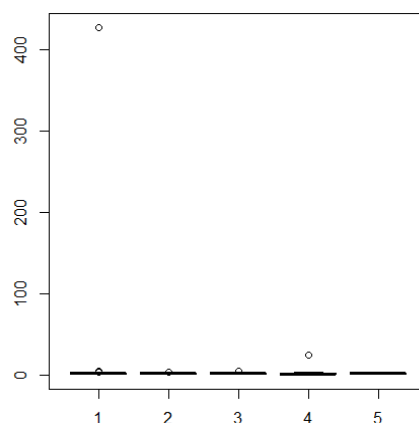


Рис.11 Диаграмма размахов показателя *IgA*

В основном все значения показателя *концентрация иммуноглобулина А в сыворотке крови (IgA)* лежат в пределах от 0 до 5. Значение 427 в первой группе – точно экстремальный выброс, так как при внесении данных специалист должен был ввести значение 4,27. Экстремальный выброс – это те выбросы, которые расположены за пределами первой и третьей квантили более  $3 \cdot \text{ИКР}$ . В четвёртой группе видим выброс равный 24. На рис.12 диаграмма размахов признака *IgA* после удаления выбросов.

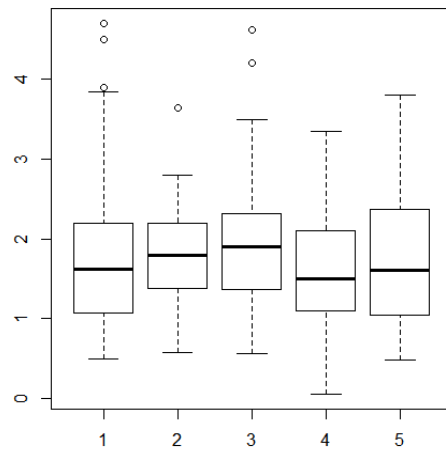


Рис.12 Диаграмма размахов показателя *IgA* после удаления выбросов

Оставшиеся значения являются умеренными выбросами. Это выбросы, которые расположены ниже первой квантили и выше третьей от  $1.5 \cdot \text{ИКР}$ , но не более  $3 \cdot \text{ИКР}$ . В следующих примерах хорошо видны экстремальные и умеренные выбросы. На рис. 13,14 изображены диаграммы выбросов для показателя *лизоцим*. Выброс в первой группе 80, в третьей 4, в четвертой 16.

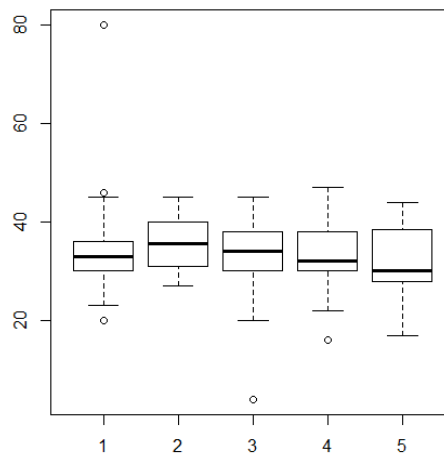


Рис.13 Диаграмма размахов показателя *лизоцим* после удаления выбросов

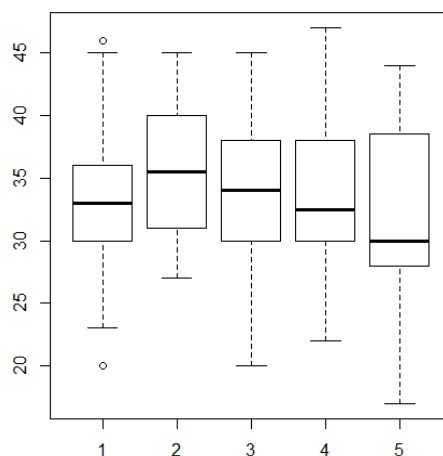


Рис.14 Диаграмма размахов показателя *лизоцим* после удаления выбросов

На рис. 15 изображена диаграмма выбросов для показателя *макроглобулин (МГ)*.

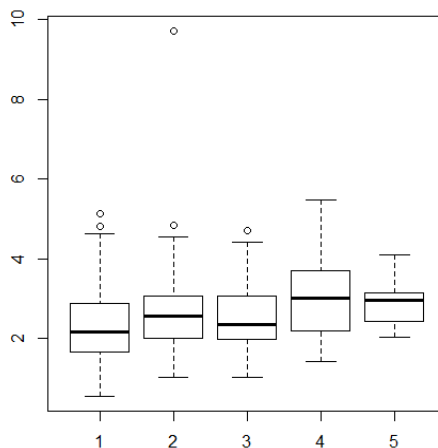


Рис.15 Диаграмма размахов показателя *МГ*

Из рисунка видим один экстремальный выброс равный 9,72. После удаления данного выброса получаем следующую диаграмму (рис.16).

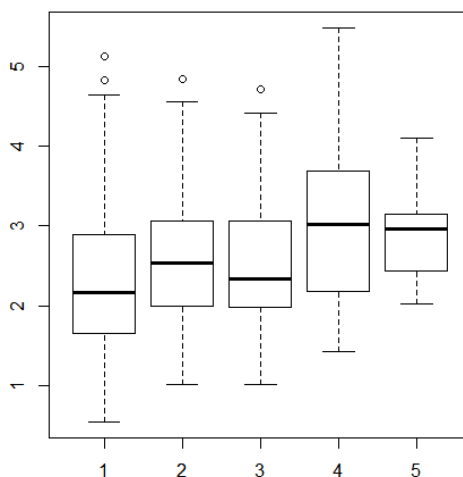


Рис.16 Диаграмма размахов показателя *МГ* без выброса

Оставшиеся умеренные выбросы не удаляются, такие ситуации могут появляться в результате недостаточного набора данных, в результате неверно классифицированных объектов (отнесения пациента к другой группе лечения), присутствия объектов из других выборок (например, снятие определённого показателя у пациента в других условиях).

### 3.2 Методы сокращения размерности

Объекты могут быть представлены сложными многомерными данными, чрезмерный объем информации может привести к снижению эффективности анализа данных, поэтому следует из большого набора данных находить информативные для полезные решения задачи распознавания.

Неинформативные признаки являются источником дополнительного шума и влияют на точность оценки параметров модели. Наборы данных с большим числом признаков могут содержать зависимые переменные. Наличие этих переменных означает дублирование информации, которое может исказить спецификацию модели и влиять на качество оценки ее параметров. Чем выше размерность данных, тем выше объем вычислений при их алгоритмической обработке.

Один из распространенных методов сокращения размерности – метод главных компонент. Суть метода – поиск проекции на гиперплоскость с сохранением большей части дисперсии в данных. Применяется во многих областях, таких как распознавание образов, компьютерное зрение, сжатие данных. В общем случае, процесс выделения главных компонент происходит следующим образом:

- каждой строке исходной матрицы соответствует точка на плоскости;
- начало координат переносится в центр выборки – нулевая главная компонента;
- выбирается направление максимального изменения данных – это первая главная компонента, вдоль которой дисперсия максимальна;
- если шум велик, то выбирается еще одно направление – перпендикулярное к первому, так чтобы описать оставшееся изменение в данных;
- отбрасываются направления с низкой дисперсией.

При таком преобразовании отбрасываются координаты с минимальной дисперсией. Основной недостаток данного метода заключается в том, что переменная с максимальной дисперсией не всегда несет нужную информацию, переменная с минимальной дисперсией разделяет классы. Недостаток: метод не выполняет линейное разделение классов, он только позволяет оптимальным образом восстановить входной вектор на основе неполной информации о нем. Вся дополнительная информация, связанная с вектором (например, принадлежность образа к одному из классов), игнорируется.

Другой метод – метод независимых компонент. Метод обычно рассматривают как расширение метода главных компонент. Целью анализа независимых компонент является не фильтрация данных, а выявление всех сигналов с их возможной кластеризацией при большом количестве на статистически независимые группы, насколько это возможно.

Другие способы уменьшения размерности данных:

1) это многомерное шкалирование – метод анализа и визуализации данных с помощью расположения точек, соответствующих изучаемым объектам, в пространстве меньшей размерности, чем пространство признаков объектов. Точки

размещаются так, чтобы попарные расстояния между ними в новом пространстве как можно меньше отличались от эмпирически измеренных расстояний в пространстве признаков изучаемых объектов;

2) многочисленные нелинейные обобщения: например, поиск наилучшей проекции (метод, состоящий в нахождении такой проекции многомерных данных, для которой достигает максимума некоторая функция качества проекции),

3) самоорганизующаяся карта Кохонена – нейронная сеть с обучением без учителя, выполняющая задачу визуализации и кластеризации. Целью применения данного метода является поиск скрытых закономерностей в данных, основываясь на снижении размерности исходного пространства в пространство меньшей размерности. В результате обучения данной модели получается решетка, состоящая из обученных нейронов. Алгоритм работы сети: первоначальное задание векторов веса для узлов; выбор вектора из множества; нахождение для него лучшего узла на карте, вектор веса которого меньше всего отличается от наблюдения; определение количества соседей и обучение — изменение векторов веса и его соседей с целью их приближения к наблюдению; определение ошибки карты.

Помимо выше пересмотренных методов существуют и другие, выделяют так же поиск информативных признаков.

В медицине одной из важных задач является поиск и отбор признаков достаточно информативных для постановки достоверного диагноза. Специалисты дают определение информативным признакам – это полезная информация для конкретной поставленной цели, полученная из исходной информации. В ходе обзора статей по данной теме были выделены методы, которые используются для оценки информативности учёными, докторами технических наук, студентами. В статье «Оценка информативности диагностических показателей в компьютерных системах медицинской диагностики» авторами предлагается теоретико-информационный подход к формированию системы информативных диагностических признаков на основании количественного выражения их значимости [31].

Авторы И.Б. Богер, Н.В. Шаланов [32] в своей статье рассмотрели методический подход к оценке информативности признаков системы, описывающий многомерный динамический объект. Данный подход включал: оценку степени однородности эмпирических данных по каждому признаку, оценку мультиколлинеарности признаков (линейная зависимость между объектами) на основе матрицы парных коэффициентов корреляции, интегральную оценку потенциала объекта по комплексу показателей.

В статье «Методы анализа информативности разнотипных признаков» были рассмотрены подходы к отбору информативных признаков: на основе дискретных методов поиска в обучающей выборке информативной зоны, на основе методов кластеризации, на основные предположения о нормальности распределения объектов в кластерах, на основе теоретико-информационного понятия энтропии [33].

Выделяют два подхода к оценке информативности – энергетический и информационный. Первый заключается в том, что информативность оценивается по величине показателя. Все показатели упорядочиваются по величине и чем больше величина, тем информативней считается признак. В информационном подходе информацией показателя считается достоверное различие между классами объекта в пространстве признаков. Например, если решается задача отнесения объекта к одному из двух классов, тогда в качестве достоверного различия может быть различие распределений вероятностей показателя, построенных по выборкам из этих классов.

Один из известных методов оценки информативности с использованием различий распределений – метод накопленных частот. При таком методе берутся 2 выборки одного признака, принадлежащие двум различным классам, по обеим выборкам в одних координатных осях строят эмпирические распределения признака  $x$  и подсчитывают накопленные частоты (сумму частот от начального до текущего интервала распределения). Оценкой информативности служит модуль максимальной разности накопленных частот.

В методе Шеннона информативность оценивается как средневзвешенное количество информации, приходящееся на различные градации признака. Формула информативности:

$$I(x_i) = 1 + \sum_{i=1}^G (p_i \cdot \sum_{k=1}^K p_{i,k} \cdot \log_K p_{i,k}), \quad (2)$$

где  $G$  – количество градаций признака;

$K$  – количество классов;

$p_i$  – вероятность  $i$ -ой градации признака.

$$p_i = \frac{\sum_{k=1}^K m_{i,k}}{N}, \quad (3)$$

$m_{i,k}$  – частота появления  $i$ -ой градации в  $K$ -ом классе,

$N$  – общее число наблюдений;

$p_{i,k}$  – вероятность появления  $i$ -ой градации признака в  $K$ -ом классе [34].

$$p_{i,k} = \frac{m_{i,k}}{\sum_{k=1}^K m_{i,k}}, \quad (4)$$

Ещё один известный метод оценки информативности – метод Кульбака. Воспользуемся этим методом в исследовании при решении задачи оценки информативности клинико-лабораторных показателей.

### 3.2.1 Оценка информативности по Кульбаку

Метод Кульбака предлагает в качестве оценки информативности – меру расхождения между двумя классами, которая называется дивергенцией. Информативность вычисляется по формуле:

$$J(x_i / A_1, x_i / A_2) = \sum_j \lg \frac{P(x_{ij} / A_1)}{P(x_{ij} / A_2)} [P(x_{ij} / A_1) - P(x_{ij} / A_2)], \quad (5)$$

$J(x_i / A_1, x_i / A_2)$  – информативность признака,

$P_1$  – вероятность попадания признака в первом классе  $A_1$ ,

$P_2$  – вероятность попадания признака во втором классе  $A_2$ ,

$j$  – номер диапазона признака  $x_j$ .

$J(x_i)$  – величина всегда будет положительной. Это связано со свойством логарифмов. Если числитель логарифмической дроби больше знаменателя, то логарифм отношения вероятностей будет положительной величиной. Если числитель логарифмической дроби меньше знаменателя, то логарифм отношения вероятностей и разность вероятностей будут отрицательными величинами и при перемножении дадут положительную величину. Таким образом, величина  $J(x_i)$ , будучи всегда положительной, отразит абсолютное значение вклада данного диапазона в приближении к любому правильному диагностическому порогу [35]. Метод Кульбака служит для определения информативности признака, который участвует в распознавании двух классов объектов, по данному методу объемы выборки данных могут быть различны.

Этапы расчёта информативности:

1. Были взяты показатели до и после лечения из одной группы.
2. Далее определяли для каждого признака минимальное и максимальное значение.

3. Задавали количество интервалов распределения и подсчитывали количество значений, которые попадают в каждую группу (всего 5 групп).

4. Рассчитывали вероятность попадания признака в группу 1 (до лечения) и в группу 2 (после лечения).

5. По формуле, приведенной выше, рассчитывали информативность.

### 3.2.2 Оценка информативности клинико-лабораторных показателей

Определение информативной ценности связано с необходимостью выделения из множества признаков наиболее информативных, получаемых при обследовании больных, которые характеризуют психофизическое состояние объекта. Исследуем информативность признаков в каждой из пяти групп лечения. Исходные данные были разделены на две выборки: выборка А1 – показатели до лечения, выборка А2 – показатели после лечения. Информативность рассчитывалась с помощью программного продукта «NPP», созданной командой ТПУ, занимающейся данной темой. В программе есть возможность выбрать исходные данные, добавить анализируемые признаки, задать диапазоны, рассчитать вероятность попадания значений признака в каждый диапазон и произвести расчёты информативности. На рис.17 представлен интерфейс программы.

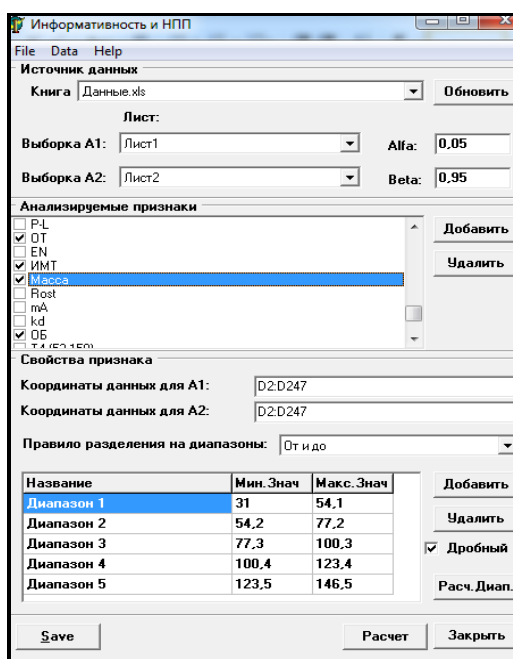


Рис.17 Расчет информативности в программе

Далее приведены результаты расчёта информативности предоставленных физиологических систем организма для каждой группы (таблицы 1-5).



Таблица 1. Информативность группы №1

Физиологическая группа	Признак	Информативность
Клиника	ТМТ	7,91
	ИМТ	0,6
	Масса	0,23
	Избыток	0,16
Сердечно-сосудистая система	ДАД, мм.рт.ст	1,89
	САД, мм.рт.ст	0,8
Физическая работоспособность	ТФН	9,35
	НОМА	0,43
	Дв. Пр.	0,2
Липидный обмен	ЛПОНП	1,7
	ТАГ	1,29
	ЛПНП	0,61
	Холестерин	0,47
	Инд.атерогенности	0,42
	аХС	0,4
	Общие липиды	0,11
Биохимия крови	Церулоплазмин	0,86
	МДА	0,83
	Каталаза	0,63
Иммунологический статус	IgM	0,84
	Лизоцим	0,51
	IgG	0,49
	ЦИК	0,47
	IgA	0,4
	Т-суп	0,4
	В-лим	0,28
	Тхел	0,14
	Т-лим	0,04
Гормональный статус	ИЛ-6	6,32
	Лептин	2,58
	Инсулин	0,69
	ТТГ(0,23-3,4)	0,4
	Кортизол	0,27
	Т3 (1,0-2,8)	0,24
	ФНО (не >2,5 пг/мл)	0,1
	АТ ТПО (до 30)	0,03
	Т4 (53-158)	0
	Состояние калликреин-кининовой системы	ПИ
АПФ		1,05
КК		1,03
МГ		0,44
Окислительная способность плазмы крови	НО	9,56
	окЛПНП	1,72

Таблица 2. Информативность группы №2

<b>Физиологическая группа</b>	<b>Признак</b>	<b>Информативность</b>
Клиника	ИМТ	1,4
	ТМТ	0,55
	Масса	0,18
	Избыток	0,09
Сердечно-сосудистая система	ДАД, мм.рт.ст	0,92
	САД, мм.рт.ст	0,39
Физическая работоспособность	ТФН	0,77
	НОМА	0,51
	Дв. Пр.	0,23
Биохимия крови	Каталаза	9,94
	Церулоплазмин	7,71
	МДА	7,46
	Сиал.к-ты	4,52
Липидный обмен	Холестерин	3,27
	ТАГ	2,42
	Общие липиды	2,21
	ЛПНП	2,2
	ЛПОНП	2,17
	аХС	1,9
	Инд.атерогенности	0,37
Иммунологический статус	ЦИК	11,14
	Тхел	7,9
	IgA	5,76
	Т-суп	4,42
	IgG	3,38
	Лизоцим	2,84
	Т-лим	2,52
	В-лим	1,75
	IgM	1,72
Гормональный статус	T4 (53-158)	4,35
	ФНО (не >2,5 пг/мл)	3,46
	АТ ТПО (до 30)	3,44
	ИЛ-6	2,75
	ТТГ(0,23-3,4)	1,33
	Кортизол	0,6
	T3 (1,0-2,8)	0,53
	Инсулин	0,15
Состояние калликреин-кининовой системы	КК	3,98
	МГ	1,36
	АПФ	0,33
	ПИ	0,33
Окислительная способность плазмы крови	NO	22,8
	окЛПНП	12,25

Таблица 3. Информативность группы №3

<b>Физиологическая группа</b>	<b>Признак</b>	<b>Информативность</b>
Клиника	Избыток	1,76
	ИМТ	0,95
	ОТ	0,22
	Масса	0,11
	ОБ	0
Сердечно-сосудистая система	ДАД, мм.рт.ст	1,18
	САД, мм.рт.ст	0,45
Физическая работоспособность	Дв. Пр.	5,98
	НОМА	2,73
	ТФН	1
Биохимия крови	МДА	11,03
	Сиал.к-ты	2,18
	Церулоплазмин	0,43
	Каталаза	0,25
Липидный обмен	ЛПНП	2,98
	ЛПОНП	2,75
	Инд.атерогенности	2,05
	Холестерин	2
	аХС	1,61
	Общие липиды	1,11
	ТАГ	0,69
Иммунологический статус	IgM	1,82
	ЦИК	1,36
	Лизоцим	1,23
	IgG	1,2
	В-лим	0,78
	Т-лим	0,29
	Тхел	0,21
	IgA	0,07
	Т-суп	0,04
Гормональный статус	ИЛ-6	6,58
	Кортизол	2,69
	АТ ТПО (до 30)	1,73
	Т3 (1,0-2,8)	1,69
	ТТГ(0,23-3,4)	1,04
	Инсулин	0,4
	Т4 (53-158)	0,28
	Лептин	0,16
Состояние калликреин-кининовой системы	КК	4,23
	ПИ	2,68
	АПФ	2,15
	МГ	0,42
Окислительная способность плазмы крови	окЛПНП	49,58
	NO	37,01

Таблица 4. Информативность группы №4

<b>Физиологическая группа</b>	<b>Признак</b>	<b>Информативность</b>
Клиника	Избыток	0,45
	ОБ	0,44
	ОТ	0,3
	ИМТ	0,13
	Масса	0,06
Сердечно-сосудистая система	САД, мм.рт.ст	6,23
	ДАД, мм.рт.ст	0,35
Физическая работоспособность	ТФН	2,44
	НОМА	1,72
	Дв. Пр.	1,59
Липидный обмен	ЛПОНП	4,61
	Общие липиды	1,75
	ТАГ	1,07
	Холестерин	1
	Инд.атерогенности	1
	ЛПНП	0,96
	аХС	0,16
Иммунологический статус	IgG	2,69
	Лизоцим	2,65
	IgA	2,1
	Тхел	1,42
	В-лим	1,34
	Т-лим	0,41
	ЦИК	0,3
	IgM	0,3
	Т-суп	0,11
Гормональный статус	ИЛ-4 (не>20пг/мл)	6,88
	ТТГ(0,23-3,4)	1,86
	Т4св(10-23.2)	1,74
	Т3св(4-8.6)	1,31
	Инсулин	0,94
	ИЛ-1В (не>50 пкг/мл)	0,92
	ФНО (не >2,5 пг/мл)	0,52
	Лептин	0,31
	Кортизол	0,27
Состояние калликреин-кининовой системы	АПФ	5,02
	КК	4,09
	ПИ	3,37
	МГ	2,8

Таблица 5. Информативность группы №5

<b>Физиологическая группа</b>	<b>Признак</b>	<b>Информативность</b>
Клиника	ИМТ	6,73
	Избыток	2,32
	ОБ	0,75
	ОТ	0,3
	Масса	0,08
Сердечно-сосудистая система	САД, мм.рт.ст	7,11
	ДАД, мм.рт.ст	0,4
Физическая работоспособность	НОМА	9,35
	Общая р-ть	8,15
	Дв. Пр.	3,68
	ТФН	3,41
Липидный обмен	ТАГ	2,45
	ЛПНП	0,57
	Холестерин	0,5
	Инд.атерогенности	0,5
	Общие липиды	0,41
	аХС	0,36
	ЛПОНП	0,13
Иммунологический статус	IgA	5,08
	T-лим	3,07
	ЦИК	2,92
	Tхел	0,92
	Лизоцим	0,88
	B-лим	0,8
	IgM	0,5
	IgG	0,21
	T-суп	0,2
Гормональный статус	ТТГ(0,23-3,4)	4,35
	ФНО (не >2,5 пг/мл)	3,79
	ИЛ-6	3,51
	Т3св(4-8.6)	2,3
	Кортизол	0,62
	Инсулин	0,54
	Т4св(10-23.2)	0,43
	ИЛ-4 (не>20пг/мл)	0,36
	ИЛ-1В (не>50 пкг/мл)	0,31
Состояние калликреин-кининовой системы	ПИ	18,43
	МГ	9,57
	КК	5,98
	АПФ	2,8

Исходя из результатов, приведённых выше, можно составить общую таблицу информативности по 5 группам лечения. В таблице 6 приведены наиболее информативные признаки в каждой группе.

Таблица 6. Информативные признаки в каждой группе

Группа	Список информативных признаков
1	NO (9,56), ТФН (9,35), ТМТ (7,91), ИЛ-6 (6,32), лептин (2,58)
2	NO (22,8), окЛПНП (12,25), ЦИК (11,14), Каталаза (9,94), Тхел (7,9), Церулоплазмин (7,71), МДА (7,46)
3	МДА (11,03), ИЛ-6 (6,58), Дв. Пр.(5,98), КК (4,23), ЛПНП (2,98), ЛПОНП (2,75), НОМА (2,73), Кортизол (2,69), ПИ (2,68)
4	ИЛ-4 (6,88), САД, мм.рт.ст (6,23), АПФ (5, 02), ЛПОНП (4,61), КК (4,09), ПИ (3,37), МГ (2,8), IgG (2,69), Лизоцим (2,65), ТФН (2,44), IgA (2,1)
5	ПИ (18,43), МГ (9,57), НОМА (9,35), Общая р-ть (8,15), ИМТ (6,73), САД, мм.рт.ст (7,11), КК (5,98)

По таблице видно, что комплексы лечения по-разному влияют на физиологические системы организма, из-за чего получилось разная информативность признаков. Для каждой физиологической группы были построены графики наиболее информативного и менее информативного признака. Рассмотрим примеры графиков информативных признаков. Возьмем признак *NO* группы лечения 1 из физиологической группы «Окислительная способность плазмы крови», признак *АПФ* группы лечения 4 из физиологической группы «Состояние калликреин-кининовой системы». На рис. 18 представлены вероятности попадания значений данных в диапазоны.

	A	B	C	D
1	NO	Диапазон	Выборка A1	Выборка A2
2	Диапазон 1	4,2..9,1	14%	0%
3	Диапазон 2	9,2..14	41%	5%
4	Диапазон 3	14,1..18,9	32%	50%
5	Диапазон 4	19..23,8	5%	35%
6	Диапазон 5	23,9..29,1	5%	10%
7	АПФ	Диапазон	Выборка A1	Выборка A2
8	Диапазон 1	22,7..27,8	12%	28%
9	Диапазон 2	27,9..33	22%	56%
10	Диапазон 3	33,1..38,2	31%	12%
11	Диапазон 4	38,3..43,4	25%	3%
12	Диапазон 5	43,5..48,7	6%	0%

Рис. 18 Вероятность попадания значений признаков в диапазоны

На рис. 19, 20 представлены информативные признаки *NO* и *АПФ* соответственно.

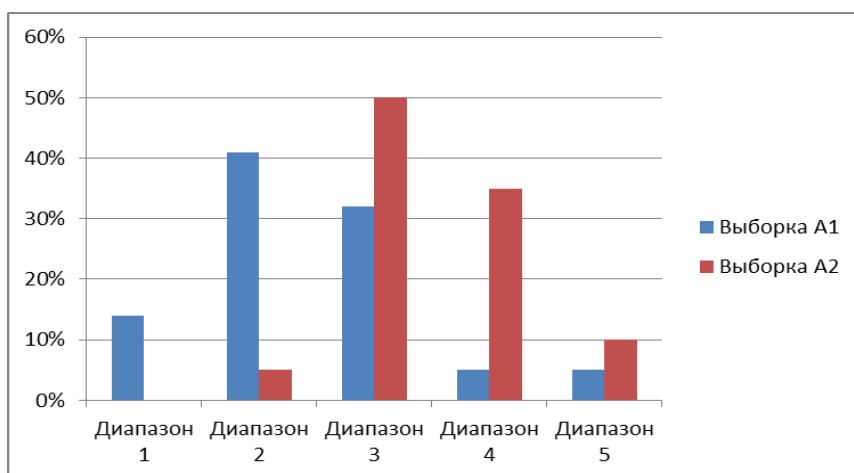


Рис. 19 Информативный признак *NO*

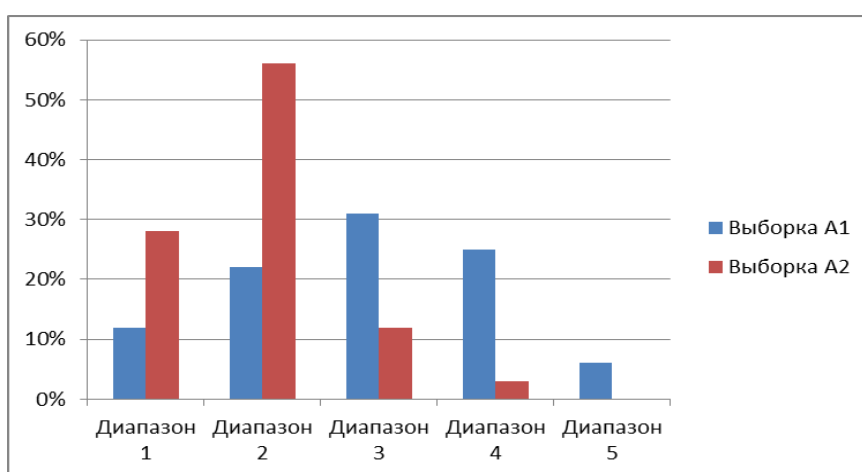


Рис.20 Информативный признак *АПФ*

Далее приведены примеры неинформативных признаков. Например, возьмем признак *масса* из первой группы лечения с информативностью равной 0,23 и признаки *ПИ* из 2-ой группы лечения с информативностью 0,33. На рис. 21 представлены вероятности попадания значений данных в диапазоны, на основе которых построены графики.

	A	B	C	D
1	Масса	Диапазон	Выборка A1	Выборка A2
2	Диапазон 1	33,5..47,3	8%	8%
3	Диапазон 2	47,4..61,1	32%	38%
4	Диапазон 3	61,2..75	34%	34%
5	Диапазон 4	75,1..88,9	14%	12%
6	Диапазон 5	89..103	12%	8%
7	ПИ	Диапазон	Выборка A1	Выборка A2
8	Диапазон 1	5,19..17,5	5%	10%
9	Диапазон 2	17,6..29,8	34%	35%
10	Диапазон 3	29,9..42,1	32%	38%
11	Диапазон 4	42,2..54,4	24%	12%
12	Диапазон 5	54,5..67	2%	5%

Рис.21 Вероятности попадания значений признаков *масса* и *ПИ* в диапазоны

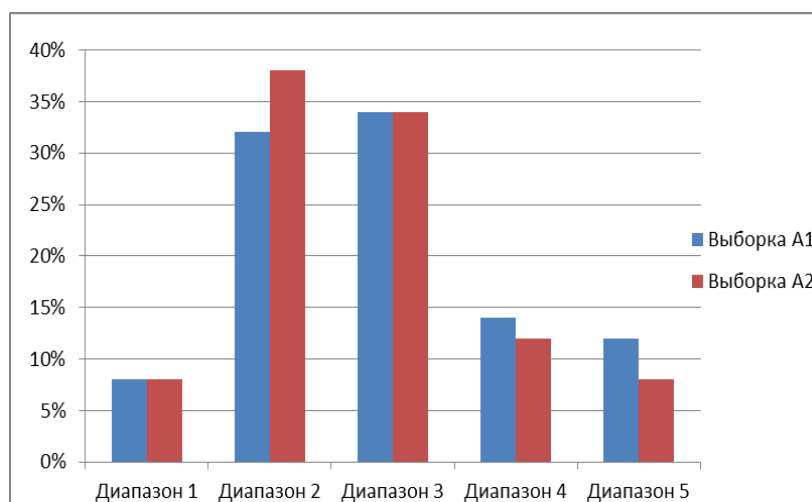


Рис.22 Неинформативный признак *масса*

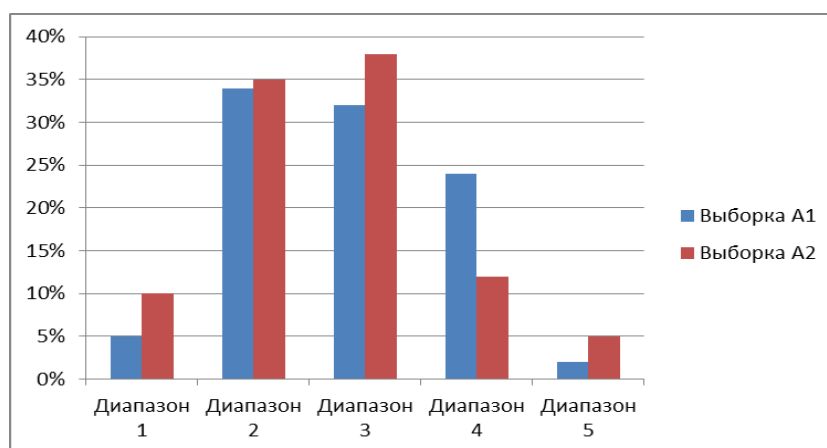


Рис.23 Неинформативный признак *ПИ*

Из графиков видно, что чем больше разница в сдвиге, тем информативнее признак. На рис. 14 хорошо видны сдвиги выборок, выборка A1 смещена в левую стороны, выборка A2 – в правую. На графиках, где разница между выборками не значительная, наглядно можно увидеть, что признаки считаются неинформативными. В результате решения задачи сокращения размерности многомерных данных из 55 клиничко-лабораторных показателей получили 27 информативных, которые несут полезную информацию для определения комплекса лечения.

### 3.3 Описание операторов в Rapid Miner

Схема построения дерева решений в Rapid Miner описывается следующими этапами:

1. Импорт данных, в результате которого получается исходная выборка;
2. На основе полученной обучающей выборки получаем обученную модель (дерево решений);



3. Применение модели для тестовой выборки.

4. Результат критерия производительности задачи классификации (процент правильно классифицированных примеров).

В результате получаем модель, которой можно воспользоваться для новых данных и схему дерева решений, отображающая правила, по которым алгоритм производит свою классификацию. На рис.22 представлены операторы для построения дерева решений.

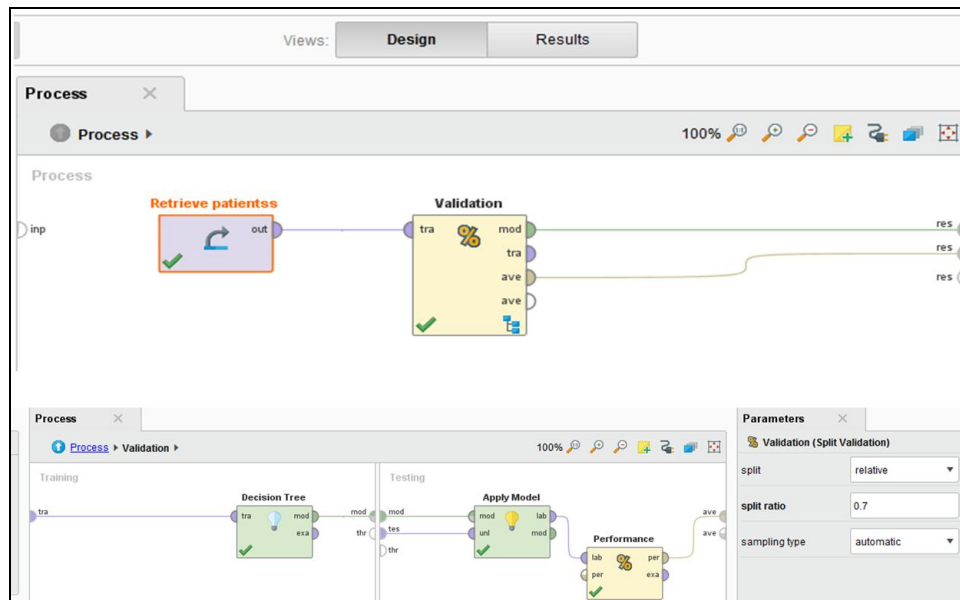


Рис.24 Операторы в Rapid Miner

Оператор «Retrieve» - этот оператор получает доступ к хранящейся в репозитории (хранилище) информации и загружает её в процесс, загрузка клинично-лабораторных показателей.

Оператор «Validation» - имеет два подпроцесса: подпроцесс обучения и подпроцесс тестирования. Подпроцесс обучения используется для обучения или построения модели, затем обученная модель применяется в подпроцессе тестирования. Оператор выполняет проверку для оценки производительности обучающего оператора. В основном оператор используется для оценки того, насколько точно модель, изученная конкретным оператором обучения будет работать на практике. Размер обучающего и тестового подмножеств можно регулировать с помощью различных параметров.

К параметрам относятся split (разбиение) - указывает, как должен быть разделен обучающий набор. Данный параметр имеет следующие значения: относительное (при этом необходимо задать относительный размер обучающего набора, значение должно быть от 1 до 0, где 1 означает, что весь обучающий набор будет использоваться в качестве тренировочного набора) или абсолютное разбиение (где нужно указать точное количество примеров для использования в наборе обучения или теста). Другой параметр sampling

type (тип выборки) позволяет выбрать линейную выборку, когда примеры берутся последовательно, перетасованную, примеры которой выбираются случайным образом, стратифицированную выборку, при которой строятся случайные подмножества, так что каждое подмножество содержит примерно одинаковые пропорции двух значений меток классов и автоматическую. При автоматическом режиме по умолчанию используется стратифицированную выборку. Если обучающая выборка не содержит номинальной метки, вместо этого будет использоваться перетасованная выборка.

Подпроцесс обучение включает в себя оператор «Decision Tree» - генерирует модель дерева решений, которая используется для классификации. В параметрах оператора можно настроить критерий, по которому будут выбраны атрибуты для разбиения, минимальное количество объектов в поддереве, максимальную глубину дерева.

Подпроцесс тестирования включает операторы: «Apply Model», на входе этого оператора подается обученная модель и тестируемые данные, на выходе получаем применение этой модели к тестируемой выборке; и оператор «Performance (Classification)», используемый для статистической оценки эффективности задач классификации. Для этого оператора можно выбрать из списка критерий производительности задачи классификации.

Для исследуемых данных установили относительное распределение с размером обучающего набора 0.7, автоматический режим выборки, для классификации выбран критерий «точность», показывающий процент правильно классифицированных примеров. В параметрах дерева решений установили критерий энтропии для выбора атрибутов для разбиения, минимальное количество объектов в поддереве 2, максимальную глубину дерева 20.

### **3.4 Построение дерева решений**

Исходными данными для построения дерева решений являются информативные признаки, полученные в результате расчета информативности. Процент отнесения объекта к определенной группе равен 96,88% (рис.23).

accuracy: 96.88%						
	true group1	true group2	true group3	true group4	true group5	class precision
pred. group1	21	0	0	0	0	100.00%
pred. group2	2	9	0	0	0	81.82%
pred. group3	0	0	14	0	0	100.00%
pred. group4	0	0	0	13	0	100.00%
pred. group5	0	0	0	0	5	100.00%
class recall	91.30%	100.00%	100.00%	100.00%	100.00%	

Рис.25 Процент правильно классифицированных примеров

В результате получили дерево решений со следующей структурой:

- 10 веток. Ветка – правило, по которому получается поддерева;
- 10 поддеревьев. Поддерево – множество объектов, полученное из некоторого правила.
- 6 листьев. Листья представляют собой некоторое правило решений относительно целевого параметра.
- 5 узлов. Узел - место, в котором происходит разветвление дерева на поддерева, в момент создания узлов решается насколько и как будет расширено дерево.

Описание дерева решений представлено на рис.24.

```

Tree
TMT > 13.200
|  ObRabotos > 150
|  |  NO > 3.500: group3 {group1=0, group2=0, group3=46, group4=0, group5=0}
|  |  NO ≤ 3.500: group4 {group1=0, group2=0, group3=0, group4=42, group5=0}
|  ObRabotos ≤ 150
|  |  Leptin > 58.100
|  |  |  Leptin > 62.550: group1 {group1=5, group2=1, group3=0, group4=0, group5=0}
|  |  |  Leptin ≤ 62.550: group2 {group1=0, group2=28, group3=0, group4=0, group5=0}
|  |  Leptin ≤ 58.100: group1 {group1=71, group2=1, group3=0, group4=0, group5=0}
TMT ≤ 13.200: group5 {group1=0, group2=0, group3=0, group4=0, group5=18}

```

Рис.26 Дерево решений

В ходе построения дерева решений были получены следующие логические правила для определения группы лечения:

1. Если TMT >13.200 и Об.раб-сть >150 и NO >3.500, то третья группа лечения.
2. Если TMT >13.200 и Об.раб-сть >150 и NO <=3.500, то четвертая группа лечения.
3. Если TMT >13.200 и Об.раб-ст<=150 и 58.100<Лептин>62.550, то первая группа лечения.
4. Если TMT >13.200 и Об.раб-ст<=150 и 58.100<Лептин<=62.550, то вторая группа лечения.
5. Если TMT >13.200 и Об.раб-ст<=150 и 58.100<=Лептин, то первая группа лечения.
6. Если TMT <=13.200, то пятая группа лечения.

Графическое построение дерева решений представлено на рис. 25.

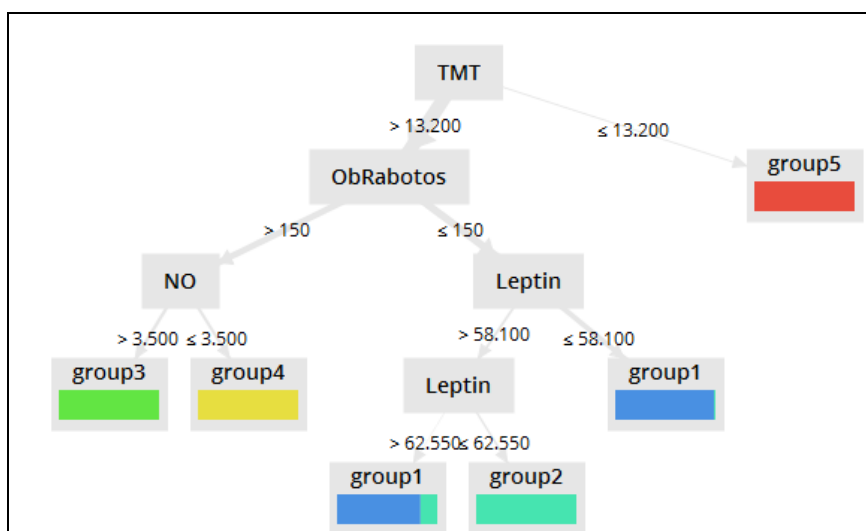


Рис.27 Графическое построение дерева решений

### 3.5 Алгоритм СППВР для выбора тактики лечения

В результате проведения исследования был разработан алгоритм поддержки принятия врачебных решения для выбора тактики лечения для детей и подростков с эндокринопатиями (рис.23).

Алгоритм СППВР состоит из следующих этапов:

1. Массив данных. Томским НИИ курортологии и физиотерапии были предоставлены исходные данные - клинико-лабораторные показатели детей и подростков с эндокринопатиями.

2. Первичная обработка данных. Далее выполнялась очистка данных с помощью анализа выбросов в программном продукте RStudio и анализа пропущенных значений методом средних значений.

3. Сокращение размерности. Для уменьшения размерности данных выбрали расчёт информативности по Кульбаку, из множества признаков отобрали наиболее информативные в каждой группе лечения.

4. В результате очистки данных и сокращения размерности, сформировали обучающую и тестовую выборку для формирования базы знаний. В качестве метода построения базы знаний был выбран - метод построения дерева решений.

5. Дерево решений построено в программном продукте Rapid Miner, на основе которого получили логические правила.

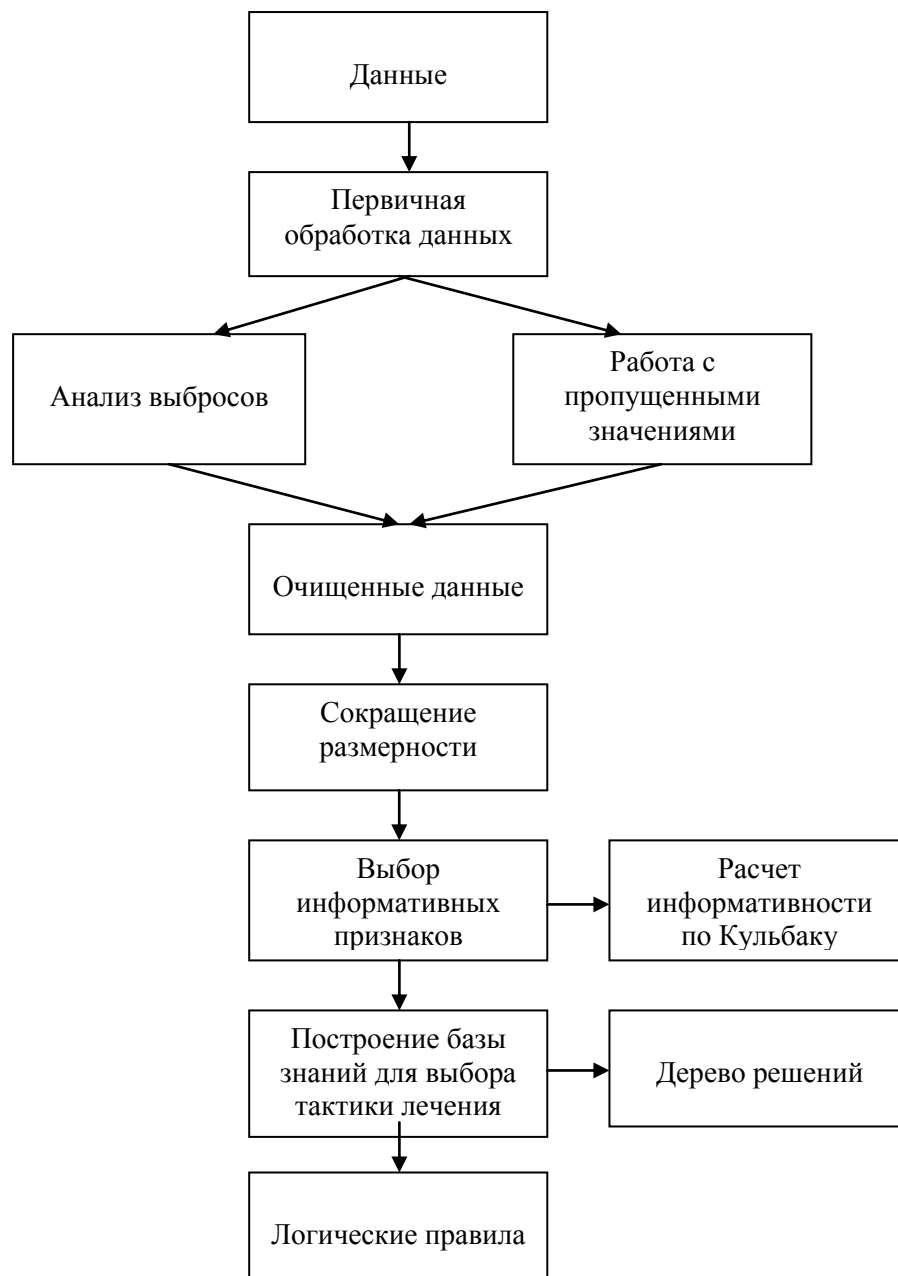


Рис.28 Алгоритм СППВР для выбора траектории лечения

В ходе выполнения разработанного в рамках ВКР алгоритма было произведено сокращение размерности данных, методом оценки информативности по Кульбаку и построение дерева решений. Дерево решений позволило получить логические правила, по которым для вновь прибывшего пациента с эндокринопатиями можно определить группу лечения по его клинико-лабораторным показателям.

## **Глава 4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение**

Данный раздел необходим для определения перспективности научно-исследовательского проекта и проектирования конкурентоспособных разработок.

Задачами раздела являются:

- оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения;
- определение возможных альтернатив проведения научных исследований;
- планирование работ по научно-исследовательскому проекту с использованием линейного графика;
- определение экономической эффективности исследования.

В разделе будет оцениваться разработка системы поддержки принятия врачебных решений при выборе тактики лечения детей с эндокринопатиями.

## 4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения

### 4.1.1 Потенциальные потребители результатов исследования

Для рассмотрения потенциальных потребителей проведем анализ целевого рынка и его сегментирования. Целевым рынком в данном исследовании являются медицинские учреждения. Выделим следующие критерии сегментирования рынка по разработке медицинских информационных систем: тип медицинского учреждения, вид медицинской информационной системы. В таблице 7 представлена карта сегментирования.

Таблица 7. Карта сегментирования рынка услуг по разработке системы поддержки принятия решений

Критерии		Вид медицинской информационной системы		
		Информационные системы территориального, федерального уровней	Комплексная автоматизированная медицинская информационная система	Система поддержки принятия врачебных решений (экспертные системы)
Тип медицинского учреждения	Учреждения стационарного типа (больницы, клиники)	Электронная медицина (г. Ростов-на-Дону)	Авиценна (г. Санкт-Петербург)	
	Медицинские научно-исследовательские институты (НИИ)		Медиалог (г.Москва)	ДИАГЕН , ДИН (г.Москва)
	Амбулаторно-поликлинические учреждения	Барс. Здравоохранение (г. Казань)	Барс. Здравоохранение (г. Казань)	

По таблице видно, какие ниши на рынке услуг по разработке медицинских информационных систем не использованы конкурентами и в каких сегментах конкуренция не очень высокая.

Результаты сегментирования:

- основными сегментами рынка являются – информационные системы территориального, федерального уровней для медицинских научно-исследовательских институтов, системы поддержки принятия врачебных решений для учреждения стационарного типа и для амбулаторно-поликлинических учреждений.

- разрабатываемая система может занять сегмент систем поддержки принятия врачебных решений и применяться не только в амбулаторно-поликлинических и

стационарного типа учреждениях, но также в медицинских научно-исследовательских институтах. Так как можно считать, что в данном сегменте (медицинские НИИ) нет конкурентов в области медицины, связанной с ожирением детей и подростков, системы ДИАГЕН, ДИН ориентированы на другие наследственные болезни у детей.

Разрабатываемая система поддержки принятия врачебных решений может быть создана для НИИ Курортологии и физиотерапии г.Томска, так как исследование проводится на основе предоставленных ими данными.

#### 4.1.2 Анализ конкурентных технических решений

Данный анализ проводится с целью, чтобы оценить эффективность научной разработки по сравнению с конкурентами. Для сравнительного анализа возьмём существующие экспертные системы: «ДИАГЕН» –система диагностики наследственных болезней у детей, «ДИН» –ориентирована на распознавание неотложного состояния у детей. Для данного анализа воспользуемся оценочной картой (таблица 8).

Таблица 8. Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		Б <sub>ф</sub>	Б <sub>к1</sub>	Б <sub>к2</sub>	К <sub>ф</sub>	К <sub>к1</sub>	К <sub>к2</sub>
1	2	3	4	5	6	7	8
<b>Технические критерии оценки ресурсоэффективности</b>							
1. Повышение производительности рудапользователя	0,16	5	5	4	0,8	0,8	0,64
2. Удобство в эксплуатации (соответствует требованиям потребителей)	0,12	5	4	5	0,6	0,48	0,6
3. Надежность	0,09	5	4	4	0,45	0,36	0,36
4. Безопасность	0,06	4	4	3	0,24	0,24	0,18
5. Потребность в ресурсах памяти	0,03	4	4	4	0,12	0,12	0,12
6. Функциональные возможности	0,09	5	4	5	0,45	0,36	0,45
7. Простота в эксплуатации	0,07	4	4	4	0,28	0,28	0,28
8. Качество интеллектуального интерфейса	0,08	4	4	3	0,32	0,32	0,24
9. Возможность подключения в сеть ЭВМ	0,05	5	5	5	0,25	0,25	0,25
<b>Экономические критерии оценки эффективности</b>							
1. Конкурентоспособность продукта	0,05	5	4	4	0,25	0,2	0,2
2. Уровень проникновения на рынок	0,06	3	4	4	0,18	0,24	0,24



3. Цена	0,05	4	4	3	0,2	0,2	0,15
4. Предполагаемый срок эксплуатации	0,02	5	3	4	0,1	0,06	0,08
5. Послепродажное обслуживание	0,04	5	4	3	0,2	0,16	0,12
6. Финансирование научной разработки	0,03	4	3	3	0,12	0,09	0,09
<b>Итого</b>	<b>1</b>	<b>Суммарная оценка</b>			<b>4,56</b>	<b>4,16</b>	<b>4,00</b>

Анализ конкурентных технических решений определяется по формуле 6:

$$K = \sum B_i \cdot B_i, \quad (6)$$

где  $K$  – конкурентоспособность научной разработки или конкурента;

$B_i$  – вес показателя (в долях единицы);

$B_i$  – балл  $i$ -го показателя.

Разрабатываемая система имеет следующие преимущества: повышение производительности труда (за счёт чего будут уменьшаться врачебные ошибки), удобство в эксплуатации.

Коэффициент конкурентоспособности:

$$k_{\text{КС}} = \frac{K_{\text{Ф}}}{K_{\text{К1}}} = (4,56/4,16 + 4,56/4,00)/2 = (1,096 + 1,14)/2 = 1,666.$$

$k_{\text{КС}} > 1$ , следовательно, разрабатываемая система является конкурентоспособной.

### 4.1.3 Технология QuaD

QQualityADvisor – технология для измерения критериев оценивания новой разработки. Эксперты оценивают каждый критерий по 100-ой шкале, где 1 – наиболее слабая позиция, а 100 – наиболее сильная. В таблице 9 представлена оценочная карта для сравнения конкурентных технических решений по технологии QuaD.

Таблица 9. Оценочная карта по технологии QuaD

Критерии оценки	Вес критерия	Баллы	Максимальный балл	Относительное значение (3/4)	Средневзвешенное значение (5x2)
1	2	3	4	5	6
<b>Показатели оценки качества разработки</b>					
1. Повышение производительности труда пользователя	0,16	100	100	1,0	0,16

2. Удобство в эксплуатации (соответствует требованиям потребителей)	0,12	95	100	0,95	0,114
3. Надежность	0,09	80	100	0,8	0,072
4. Безопасность	0,06	75	100	0,75	0,045
5. Потребность в ресурсахпамяти	0,03	50	100	0,5	0,015
6. Функциональные возможности	0,09	70	100	0,7	0,063
7. Простота в эксплуатации	0,07	80	100	0,8	0,056
8. Качество интеллектуального интерфейса	0,08	90	100	0,9	0,072
9. Возможность подключения в сеть ЭВМ	0,05	70	100	0,7	0,035
<b>Показатели оценки коммерческого потенциала разработки</b>					
1. Конкурентоспособность продукта	0,05	80	100	0,8	0,04
2. Уровеньпроникновениян арынок	0,06	75	100	0,75	0,045
3. Цена	0,05	50	100	0,5	0,025
4. Предполагаемыйсрокаэк плуатации	0,02	80	100	0,8	0,016
5. Послепродажноеобслужи вание	0,04	60	100	0,6	0,024
6. Финансированиенаучной разработки	0,03	70	100	0,7	0,021
<b>Итого</b>	<b>1</b>				<b>0,803</b>

Оценка качества и перспективности по технологии QuaD определяется по формуле 7:

$$P_{cp} = \sum B_i \cdot B_i, \quad (7)$$

где  $B_i$  – вес показателя (в долях единицы);

$B_i$  – средневзвешенное значение  $i$ -го показателя.

Значение  $P_{cp}$  говорит о перспективах разработки и качестве проведенного исследования. При расчете значение  $P_{cp}$  получилось 80,3, значение показателя лежит в пределах от 100 до 80, это говорит о том, что перспективность данной разработке является высокой.

#### 4.1.4 Диаграмма Исикавы

Диаграмма Исикавы – графическое представление причинно-следственных связей, диаграмма служит для систематического и полного определения причин возникновения проблемы. Диаграмма представлена на рис.27.

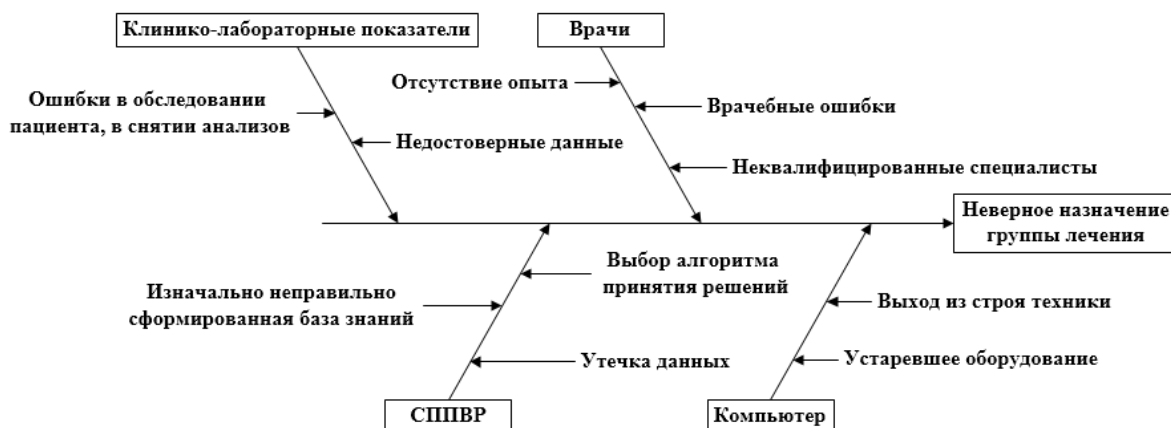


Рис.29 Диаграмма Исикавы

Из рисунка видно, что основной проблемой данной СППВР является неверное назначение тактики лечения детей. Основные факторы, влияющие на проблему: врачи, обследуемые пациенты, клинико-лабораторные показатели различных физиологических групп, оборудование (компьютер) и сама система поддержки принятия врачебных решений. Способы решения проблемы могут быть следующие: повышение квалификации врачей, смена оборудования на более современное, качественная разработка системы, хорошая обучающая и тестовая выборка для методов построения базы знаний.

#### 4.1.5 SWOT-анализ

SWOT–анализ является одним из самых распространенных методов стратегического планирования, при котором определяются сильные, слабые стороны, возможности и угрозы научно-исследовательского проекта (таблица 10).

Таблица 10. SWOT-анализ

	<b>Сильные стороны:</b>	<b>Слабые стороны:</b>
	<p>С1. Уменьшение неверных назначений врачей и сокращение врачебных ошибок</p> <p>С2. Отсутствие аналогов СППВР, направленной на</p>	<p>Сл1. Возможны сбои в работе информационной системы</p> <p>Сл.2 Угроза утечки персональных данных</p>

	<p>выбор траектории лечения детей с избыточной массой тела</p> <p>С3. Количество диагностических предположений становится менее зависимым от личного опыта и знаний врача</p> <p>С4. Увеличение скорости выбора тактики лечения</p>	<p>Сл3. Неправильно сформированная база знаний в системе, вследствие чего неверные назначения группы лечения</p>
<p><b>Возможности:</b></p> <p>В1. Рост спроса, появление новых клиентов</p> <p>В2. Расширение сферы деятельности с одного города на территориальный/ федеральный уровень</p>	<p>В1С1С3С4 – сильные стороны разработки влияют на увеличение спроса</p> <p>В2С2 – за счёт отсутствия аналогов системы, разработка может выйти на новый уровень</p>	<p>В1Сл1Сл3 – за счёт сбоев может произойти уменьшение скорости работы системы</p> <p>В1Сл2 – потеря данных клиентов из-за утечки информации</p> <p>В2Сл1 – нет возможности расширения системы</p>
<p><b>Угрозы:</b></p> <p>У1. Появление конкурентов в области СППВР</p> <p>У2. Отсутствие спроса СППВР</p>	<p>У1С2 – из-за отсутствия аналогов могут появиться конкуренты</p> <p>У2С1 – поиск медицинских учреждений, гибкость системы.</p>	<p>У1Сл1Сл2Сл3 – доработка системы, расширение возможностей, гибкость системы</p> <p>У2Сл2 – улучшение конфиденциальности данных за счёт принятия мер по защите информации</p>

В ходе проведенного анализа внутренней и внешней среды выявили ряд проблем, которые могут возникнуть, поэтому можно предложить ряд рекомендаций:

- грамотный выбор и тестирование инструментария для формирования базы знаний;
- проверка данных на достоверность и первичная их обработка;
- расширение возможностей системы;
- поиск дополнительных медицинских учреждений;

- при разработке системы учесть меры по защите данных.

В совокупности данные рекомендации позволят уменьшить слабые стороны и увеличить результативность информационной системы.

#### 4.1.6 Оценка готовности проекта к коммерциализации

На любом этапе жизненного цикла разработки необходимо определить степень её готовности к коммерциализации (таблица 11).

Таблица 11. Степень готовности научного проекта к коммерциализации

№	Наименование	Степень проработанности проекта	Уровень знаний у разработчика
1.	Определен имеющийся научно-технический задел	5	5
2.	Определены перспективные направления коммерциализации научно-технического задела	2	2
3.	Определены отрасли и технологии (товары, услуги) для предложения на рынке	5	5
4.	Определена товарная форма научно-технического задела для представления на рынок	1	1
5.	Определены авторы	5	5
6.	Проведена оценка стоимости интеллектуальной собственности	3	4
7.	Проведены маркетинговые исследования рынков сбыта	5	5
8.	Разработан бизнес-план коммерциализации научной разработки	2	1
9.	Определены пути продвижения научной разработки на рынок	3	4
10.	Разработана стратегия (форма) реализации научной разработки	5	5
11.	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	1	1
12.	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	1	1

13.	Проработаны вопросы финансирования коммерциализации научной разработки	3	5
14.	Имеется команда для коммерциализации научной разработки	4	4
15.	Проработан механизм реализации научного проекта	5	5
	<b>Итого баллов</b>	45	53

По результатам оценки, можно сделать вывод, что суммарное количество баллов по каждому направлению лежит в пределах от 45 до 59, это говорит о том, что перспективность выше среднего.

Данная разработка является перспективной, возможно, необходимо улучшить направление проведение оценки стоимости интеллектуальной стоимости и точнее определить пути продвижения на рынок.

#### 4.2 Определение возможных альтернатив проведения научных исследований

Проведём морфологический анализ для выявления возможных альтернатив исследования. В таблице 12 представлена морфологическая матрица для СППВР.

Таблица 12. Морфологическая матрица для СППВР

	1	2	3	4
А. Вид системы	Система территориального уровня	Система федерального уровня	Комплексная автоматизированная система для рабочих мест специалистов	Система поддержки принятия врачебных решений (экспертные системы)
Б. Представление базы знаний	Продукционные модели (дерево решений)	Семантические сети	Нейронные сети	Кластерный анализ
В. Статистические пакеты	RStudio	Rapid Miner	Statistica	SPSS

Исходя из результатов морфологического анализа, можно предложить следующие альтернативные варианты научного исследования:

1) A4B1B2 – система поддержки принятия врачебных решений с методом представления базы знаний - построение дерева решений в программном продукте RapidMiner.

2) A4B3B3 – СППВР, построенной на основе нейронной сети в статистическом пакете Statistica.

3) A4B4B1 – СППВР с использованием кластерного анализа, проведённого с помощью скриптового языка R в статистическом пакете RStudio.

### 4.3 Планирование научно-исследовательских работ

#### 4.3.1 Структура работ в рамках научного исследования

Структура работ предполагает следующие этапы:

- описание содержания работ;
- определение участников проекта;
- установление продолжительности работ;
- построение графика проведения исследования.

Рабочая группа научного исследования состоит из двух исполнителей: студента и научного руководителя. Для начала определим основные этапы исследования, опишем содержания работ и распределим исполнителей по видам работы (таблица 13).

Таблица 13. Перечень этапов, работ и распределение исполнителей

Основные этапы	№ работ	Содержание работ	Должность исполнителя
Подготовка технического задания	1	Выбор темы научного исследования	Студент, руководитель
	2	Составление и утверждение технического задания	Студент, руководитель
	3	Календарное планирование работ по теме исследования	Руководитель
Анализ предметной области	4	Обзор литературы	Студент
	5	Описание предметной области	Студент
	6	Создание алгоритма принятия решения	Студент, руководитель
Выбор методов	7	Выбор методов решений	Студент

решения задач	8	Выбор инструментария	Студент
Реализация проекта	9	Обработка медицинских данных	Студент
	10	Анализ данных в статистических пакетах	Студент, руководитель
	11	Оценка полученных результатов (формирование базы знаний)	Студент, руководитель
Оформление отчета по исследованию	12	Составление пояснительной записки	Студент
	13	Подготовка презентации	Студент

### 4.3.2 Определение трудоемкости выполнения работ

Ожидаемое значение трудоемкости  $t_{ожі}$  определим по формуле 8:

$$t_{ожі} = \frac{3 \cdot t_{\min i} + 2 \cdot t_{\max i}}{5}, \quad (8)$$

где  $t_{\min i}$  – минимальная трудоемкость выполнения  $i$ -ой работы, чел.-дн.;

$t_{\max i}$  – максимальная трудоемкость выполнения  $i$ -ой работы, чел.-дн.

Рассчитаем продолжительность каждой работы в рабочих днях  $T_p$ , учитывающая параллельность выполнения работ несколькими исполнителями, используя формулу 9.

$$T_{p_i} = \frac{t_{ожі}}{Ч_i}, \quad (9)$$

где  $T_{p_i}$  – продолжительность одной работы, раб.дн.;

$t_{ожі}$  – ожидаемая трудоемкость выполнения одной работы, чел.-дн.

$Ч_i$  – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Расчет трудозатрат представлен в таблице 14.

Таблица 14. Расчет трудозатрат

№ работы	Трудоёмкость работ
----------	--------------------



	tmin, чел-дни			tmax, чел-дни			t <sub>ожі</sub> , чел-дни		
	Исп.1	Исп.2	Исп.3	Исп.1	Исп.2	Исп.3	Исп.1	Исп.2	Исп.3
1	3	3	3	5	5	5	3,8	3,8	3,8
2	4	4	4	5	5	5	4,4	4,4	4,4
3	2	2	2	3	3	3	2,4	2,4	2,4
4	8	8	8	10	10	10	8,8	8,8	8,8
5	5	5	5	6	6	6	5,4	5,4	5,4
6	10	12	12	14	15	15	11, 6	13,2	13,2
7	4	5	6	5	6	7	4,4	5,4	6,4
8	4	6	6	6	8	8	4,8	6,8	6,8
9	9	9	9	11	11	11	9,8	9,8	9,8
10	20	21	22	22	23	24	20, 8	21,8	22,8
11	12	12	12	14	14	14	12, 8	12,8	12,8
12	10	10	10	11	11	11	10, 4	10,4	10,4
13	5	5	5	7	7	7	5,8	5,8	5,8
Итого	98	104	106	123	128	130	108	113,6	115,6

### 4.3.3 Разработка графика проведения научного исследования

Для отображения графика работ воспользуемся диаграммой Ганта. Для данной диаграммы, необходимо перевести длительность каждого из этапов работ в календарные дни (формула 10).

$$T_{ki} = T_{pi} \cdot k_{\text{кал}}, \quad (10)$$

где  $T_{ki}$  – продолжительность выполнения  $i$ -й работы в календарных днях;

$T_{pi}$  – продолжительность выполнения  $i$ -й работы в рабочих днях;

$k_{\text{кал}}$  – коэффициент календарности.

Коэффициент календарности определим по формуле 11:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}}, \quad (11)$$

где  $T_{\text{кал}}$  – количество календарных дней в году;

$T_{\text{вых}}$  – количество выходных дней в году;

$T_{пр}$  – количество праздничных дней в году.

Коэффициент календарности:  $K_{кал} = 365/(365-118) = 1,48$ . Длительность работ в календарных днях для каждой работы округлим до целого числа. В таблице 15 представлены расчеты временных показателей исследования.

Таблица 15. Временные показатели проведения научного исследования

№ работы	Исполнители			Длительность работ в рабочих днях $T_{pi}$			Длительность работ в календарных днях $T_{ki}$		
	Исп.1	Исп.2	Исп.3	Исп.2	Исп.1	Исп.3	Исп.1	Исп.2	Исп.3
1	2	2	2	1,9	1,9	1,9	3	3	3
2	2	2	2	2,2	2,2	2,2	4	4	4
3	1	1	1	2,4	2,4	2,4	4	4	4
4	1	1	1	8,8	8,8	8,8	14	14	14
5	1	1	1	5,4	5,4	5,4	8	8	8
6	2	2	2	5,8	6,6	6,6	9	10	10
7	1	1	1	4,4	5,4	6,4	7	8	10
8	1	1	1	4,8	6,8	6,8	8	11	11
9	1	1	1	9,8	9,8	9,8	15	15	15
10	2	2	2	10,4	10,9	11,4	16	17	17
11	2	2	2	6,4	6,4	6,4	10	10	10
12	1	1	1	10,4	10,4	10,4	16	16	16
13	1	1	1	5,8	5,8	5,8	9	9	9
Итого	-	-	-	81,3	85,6	87,1	128	134	136

Построим календарный план-график для исполнения с максимальной длительностью работ. Работы выделим цветом в зависимости от исполнителей. Некоторые работы могут выполняться одновременно. На рис.28 представлена диаграмма Ганта, построенная с помощью программного продукта MicrosoftVisio.

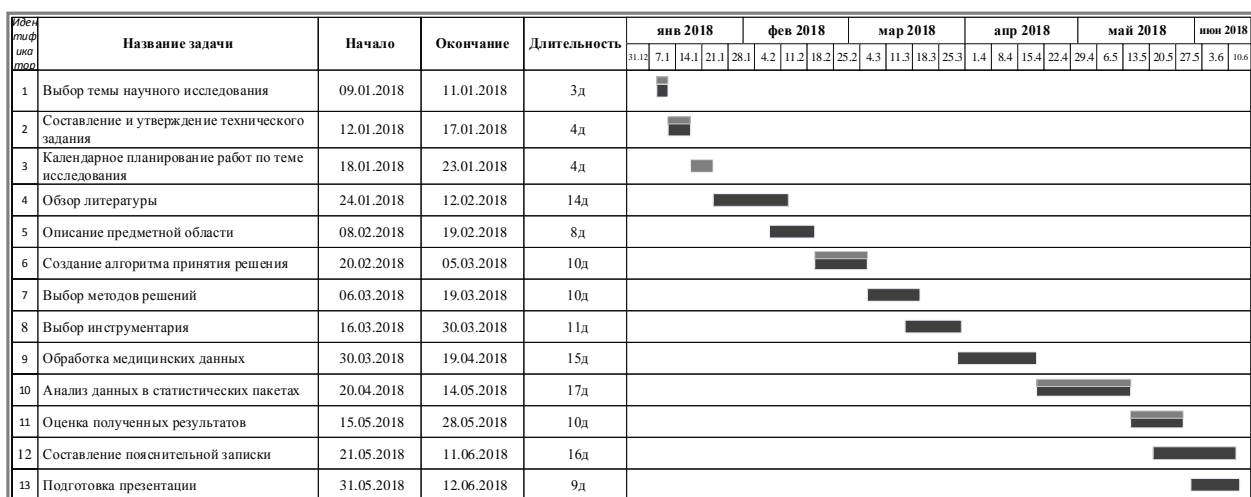


Рис.30 Диаграмма Ганта

#### 4.3.4 Бюджет научно-технического исследования (НТИ)

В бюджет НТИ включаются затраты по следующим статьям:

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

##### 4.3.4.1 Расчет материальных затрат

Для расчета материальных затрат воспользуемся формулой 12:

$$Z_m = (1 + k_T) \cdot \sum_{i=1}^m C_i \cdot N_{расxi}, \quad (12)$$

где  $m$  – количество видов материальных ресурсов;

$N_{расxi}$  – количество материальных ресурсов  $i$ -го вида, планируемых к использованию при выполнении научного исследования (шт., кг, м, м<sup>2</sup> и т.д.);

$C_i$  – цена приобретения единицы  $i$ -го вида потребляемых материальных ресурсов;

$k_T$  – коэффициент, учитывающий транспортно-заготовительные расходы. Пусть транспортные расходы составляют 15% от стоимости.

Занесем в таблицу 16 материальные затраты для данного исследования.

Таблица 16. Материальные затраты

Наименование	Единица измерения	Количество	Цена за ед. руб.			Затраты на материалы, (З <sub>м</sub> ), руб.		
			Исп. 1	Исп. 2	Исп. 3	Исп. 1	Исп. 2	Исп. 3
Системный блок	Шт.	1	10999	11499	11799	12648,85	13223,85	13568,85
Клавиатура	Шт.	1	390	399	410	448,50	458,85	471,50
Мышь	Шт.	1	250	270	299	287,50	310,50	343,85
Монитор	Шт.	1	4550	4599	4699	5232,50	5288,85	5403,85
Принтер	Шт.	1	4599	5299	5599	5288,85	6093,85	6438,85
<b>Итого:</b>						<b>23906,20</b>	<b>25375,90</b>	<b>26226,90</b>

#### 4.3.4.2 Основная заработная плата исполнителей темы

Заработную плату будем рассчитывать для научного руководителя и студента. Статья включает основную и дополнительную заработную плату. Среднедневная заработная плата рассчитывается по формуле 13:

$$Z_{\text{дн}} = \frac{Z_{\text{м}} \cdot M}{F_{\text{д}}}, \quad (13)$$

где  $Z_{\text{м}}$  – месячный должностной оклад работника, руб.;

$M$  – количество месяцев работы без отпуска в течение года:

при отпуске в 24 раб.дня  $M=11,2$  месяца, 5-дневная неделя;

при отпуске в 48 раб.дней  $M=10,4$  месяца, 6-дневная неделя;

$F_{\text{д}}$  – действительный годовой фонд рабочего времени научно-технического персонала, раб.дн. (таблица 17).

Таблица 17. Баланс рабочего времени

Показатели рабочего времени	Научный руководитель	Студент
Календарное число дней	365	365
Количество нерабочих дней	118	118
- выходные дни		
- праздничные дни		
Потери рабочего времени	48	48
- отпуск		
- невыходы по болезни		
Действительный годовой фонд рабочего времени	199	199

Месячный должностной оклад работника (формула 14):

$$Z_{\text{м}} = Z_{\text{тс}} \cdot (1 + k_{\text{пр}} + k_{\text{д}}) \cdot k_{\text{р}}, \quad (14)$$

где  $Z_{тс}$  – заработная плата по тарифной ставке, руб.;

$k_{пр}$  – премиальный коэффициент, равный 0,3 (т.е. 30% от  $Z_{тс}$ );

$k_{д}$  – коэффициент доплат и надбавок составляет примерно 0,2 – 0,5 (в НИИ и на промышленных предприятиях – за расширение сфер обслуживания, за профессиональное мастерство, за вредные условия: 15-20% от  $Z_{тс}$ );

$k_{р}$  – районный коэффициент, равный 1,3 (для Томска).

Расчёт основной заработной платы приведён в таблице 18.

Оклад студента равен 1854, а оклад научного руководителя составляет 23000.

Таблица 18. Расчет основной заработной платы

Исполнители	$k_{р}$	$Z_{м, руб.}$	$Z_{дн, руб.}$	$T_{р, раб. дн.}$			$Z_{осн, руб.}$		
				Исп. 1	Исп. 2	Исп. 3	Исп.1	Исп.2	Исп.3
Научный руководитель	1,3	29900	1562,61	30	31	31	46878,3	48440,91	48440,91
Студент	1,3	2410,2	125,96	82	86	88	10328,72	10832,56	11084,48
Итого:							57207,02	59273,47	59525,39

#### 4.3.4.3 Дополнительная заработная плата исполнителей темы

Дополнительная заработная плата рассчитывается по формуле 15:

$$Z_{доп} = k_{доп} \cdot Z_{осн}, \quad (15)$$

где  $k_{доп}$  – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15). Пусть  $k_{доп}$  равен 0,12. Расчёты представлены в таблице 19.

Таблица 19. Затраты на дополнительную заработную плату

Исполнители	Основная зарплата(руб.)			$k_{доп}$	Дополнительная зарплата(руб.)		
	Исп.1	Исп.2	Исп.3		Исп.1	Исп.2	Исп.3
Научный руководитель	46878,30	48440,91	48440,91	0,12	5625,39	5812,91	5812,91
Студент	10328,72	10832,56	11084,48	0,12	1239,45	1299,91	1330,14
				Итого	6864,84	7112,82	7143,05

#### 4.3.4.4 Отчисления во внебюджетные фонды (страховые отчисления)

Отчисления во внебюджетные фонды найдем с помощью формулы 16:

$$Z_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}), \quad (16)$$

где  $k_{\text{внеб}}$  – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.). Для учреждений осуществляющих образовательную и научную деятельность используется пониженная ставка – 27,1% (таблица 20).

Таблица 20. 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
Научный руководитель	46878,30	48440,91	48440,91	5625,39	5812,91	5812,91
Студент	10328,72	10832,56	11084,48	1239,45	1299,91	1330,14
Коэффициент отчислений во внебюджетные фонды	27,1%					
<b>Итого:</b>						
<b>Исполнение 1</b>	<b>17363,47</b>					
<b>Исполнение 2</b>	<b>17990,68</b>					
<b>Исполнение 3</b>	<b>18067,15</b>					

По таблице виден размер отчислений во внебюджетные фонды по каждому исполнению.

#### 4.3.4.5 Накладные расходы

Величину накладных расходов рассчитаем с помощью формулы 17:

$$Z_{\text{накл}} = (\text{сумма статей } 1 \div 4) \cdot k_{\text{нр}}, \quad (17)$$

где  $k_{\text{нр}}$  – коэффициент, учитывающий накладные расходы. Величина коэффициента накладных расходов равна 16%.

Исполнение 1:  $(105341,53 \cdot 0,16) = 16854,64$ .

Исполнение 2:  $(109752,87 \cdot 0,16) = 17560,46$ .

Исполнение 3:  $(110962,49 \cdot 0,16) = 17753,99$ .

#### 4.3.4.6 Формирование бюджета затрат научно-исследовательского проекта

Определение бюджета затрат на научно-исследовательский проект по каждому варианту исполнения представлено в таблице 21.

Таблица 21. Расчет бюджета затрат НИИ

Наименование статьи	Сумма, руб.			Примечание
	Исп.1	Исп.2	Исп.3	
1. Материальные затраты НИИ	23906,20	25375,90	26226,90	Пункт 1.3.4.1
2. Затраты по основной заработной плате исполнителей темы	57207,02	59273,47	59525,39	Пункт 1.3.4.2
3. Затраты по дополнительной заработной плате исполнителей темы	6864,84	7112,82	7143,05	Пункт 1.3.4.3
4. Отчисления во внебюджетные фонды	17363,47	17990,68	18067,15	Пункт 1.3.4.4
5. Накладные расходы	16854,64	17560,46	17753,99	Пункт 1.3.4.5
<b>Бюджет затрат НИИ</b>	<b>122196,17</b>	<b>127313,33</b>	<b>128716,48</b>	

#### 4.4 Определение ресурсной, финансовой, бюджетной, социальной и экономической эффективности исследования

Для определения интегрального показателя найдем две средневзвешенные величины: финансовую эффективность и ресурсоэффективность.

Формула интегрального финансового показателя (формула 18):

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}}, \quad (18)$$

где  $I_{\text{финр}}^{\text{исп.}i}$  – интегральный финансовый показатель разработки;

$\Phi_{pi}$  – стоимость  $i$ -го варианта исполнения;

$\Phi_{\text{max}}$  – максимальная стоимость исполнения научно-исследовательского проекта (в т.ч. аналоги).

Рассчитаем интегральный финансовый показатель:

$$I_{\text{финр}}^{\text{исп.}1} = \frac{122196,17}{128716,48} = 0,95$$

$$I_{\text{финр}}^{\text{исп.}2} = \frac{127313,33}{128716,48} = 0,98$$

$$I_{\text{финр}}^{\text{исп.}3} = \frac{128716,48}{128716,48} = 1$$

По формуле 19 найдем интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования:

$$I_{pi} = \sum a_i \cdot b_i, \quad (19)$$

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

$a_i$  – весовой коэффициент  $i$ -го варианта исполнения разработки;

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

$n$  – число параметров сравнения.

В таблице 22 представлен расчет интегрального показателя ресурсоэффективности.

Таблица 22. Сравнительная оценка характеристик вариантов исполнения проекта

Критерии \ Объект исследования	Весовой коэффициент параметра	Исп. 1	Исп. 2	Исп. 3
1. Повышение производительности труда пользователя	0,30	5	5	5
2. Удобство в эксплуатации (соответствует требованиям потребителей)	0,25	5	4	3
3. Помехоустойчивость	0,1	4	3	4
4. Энергосбережение	0,1	4	4	3
5. Надежность	0,15	4	4	4
6. Материалоемкость	0,1	4	4	4
<b>Итого:</b>	<b>1</b>	<b>4,55</b>	<b>4,20</b>	<b>3,95</b>

$$I_{p-исп1} = 5*0,30 + 5*0,25 + 4*0,1 + 4*0,1 + 4*0,15 + 4*0,1 = 4,55;$$

$$I_{p-исп2} = 5*0,30 + 4*0,25 + 3*0,1 + 4*0,1 + 4*0,15 + 4*0,1 = 4,20.$$

$$I_{p-исп3} = 5*0,30 + 3*0,25 + 4*0,1 + 3*0,1 + 4*0,15 + 4*0,1 = 4,20.$$

Интегральный показатель эффективности  $I_{исп.i}$  определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя (формуле 20):

$$I_{исп.1} = \frac{I_{p-исп1}}{I_{финр}}, \quad I_{исп.2} = \frac{I_{p-исп2}}{I_{финр}}, \quad I_{исп.3} = \frac{I_{p-исп3}}{I_{финр}} \quad (20)$$

Рассчитаем интегральный показатель для каждого варианта исполнения:

$$I_{исп.1} = \frac{4,55}{0,95} = 4,79$$

$$I_{исп.2} = \frac{4,20}{0,98} = 4,42$$



$$I_{исп.3} = \frac{3,95}{1} = 3,95$$

Сравнительная эффективность проекта  $\mathcal{E}_{cp}$ (формула 21):

$$\mathcal{E}_{cp} = \frac{I_{исп.1}}{I_{исп.2}}, \quad (21)$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволяет определить сравнительную эффективность проекта и выбрать наилучший вариант исполнения (таблица 23).

Таблица 23. Сравнительная эффективность разработки

% п/п	Показатели	Исп.1	Исп.2	Исп.3
1	Интегральный финансовый показатель разработки	0,95	0,98	1
2	Интегральный показатель ресурсоэффективности разработки	4,55	4,20	3,95
3	Интегральный показатель эффективности	4,79	4,42	3,95
4	Сравнительная эффективность вариантов исполнения	1	0,92	0,82

Из таблицы видно, что с позиции финансовой и ресурсной эффективности наиболее эффективным вариантом является первое исполнение.

#### **Выводы по разделу:**

В ходе выполнения задания по разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» был определен целевой рынок разработки, выполнена оценка качества и перспективности по технологии QuaD, которая говорит о том, что данное исследование является перспективным. С помощью SWOT-анализа определены сильные и слабые стороны проекта.

При анализе конкурентных технических решений определена трудоёмкость проекта. В календарных днях она составила 128, что меньше по сравнению с другими разработками. На основе временных показателей проведения научного исследования построен календарный план-график. Бюджет 1-ого исполнения НТИ равен 122196,17, 2-ого исполнения – 127313,33 и 3-его исполнения – 128716,48. По бюджету и по сравнительной эффективности вариантов исполнения видно, что менее затратным и наиболее эффективным получилось НТИ 1-ого исполнения.

## Глава 5 Социальная ответственность

Раздел «Социальная ответственность» охватывает следующие вопросы:

- анализ вредных и опасных факторов при разработке и эксплуатации проектируемого решения;
- экологическая безопасность;
- возможные чрезвычайные ситуации.

Объектом исследования данного раздела является рабочее место, представляющее собой помещение с персональным компьютером.

Работа выполнялась в аудитории 421 Кибернетического центра подразделения ИШИТР. Общая площадь аудитории составляет 35 м<sup>2</sup>. В помещении для выполнения исследования было использовано 1 компьютерное место. При организации рабочего места с компьютером необходимо учитывать санитарные правила и нормы для создания безопасных и комфортных условий труда.

## 5.1 Производственная безопасность

Производственная безопасность включает в себя анализ вредных и опасных факторов. В таблице 24 приведены факторы, характерные для данного исследования.

Таблица 24. Опасные и вредные факторы

Наименование видов работ и параметров производственного процесса	Факторы		Нормативные документы
	Вредные	Опасные	
Работа с компьютером и оргтехникой	Повышенный уровень электромагнитных излучений	Электробезопасность: – поражение электрическим током – повышенный уровень статического электричества – короткое замыкание	СанПин 2.2.4.3359-16, СанПин 2.2.2/2.4.1340-03
	Отсутствие и недостаток естественного и искусственного освещения		СанПин 2.2.1/2.1.1.1278-03
	Микроклимат		СанПин 2.2.2/2.4.1340-03.
	Повышенный уровень шума на рабочем месте		СанПин 2.2.4.548-96.
	Зрительное напряжение		ТОИ Р-45-084-01.
	Монотонность труда		

### 5.1.1 Анализ вредных факторов

#### 5.1.1.1 Недостаточная освещенность рабочего места

Естественное освещение осуществляется через оконные проемы и используется в дневное время, искусственное освещение за счёт освещения лампами – в утреннее и в вечернее время. Недостаточное освещение затрудняет выполнение работы, вызывает утомление, ухудшение зрения. По СанПин 2.2.1/2.1.1.1278-03 для кабинета информатики и вычислительной техники нормируемое значение коэффициента естественной освещенности 3,5%, освещенность при совмещенной системе 200-400 лк [36].

Так как естественного освещения в помещении не хватает, выполним расчёт искусственного освещения. Для помещения с компьютерами выберем систему общего освещения, равномерного. В качестве источников помещения возьмем люминесцентные

лампы дневного света. Наиболее распространенные светильники с люминесцентными лампами – это открытые двухламповые светильники.

Размещение светильников в помещении определяется по формуле 22:

$$h = H - h_c - h_p, \quad (22)$$

где:  $H$  – высота помещения (высота кабинета равна 3,6 м);

$h_c$  – высота светильника, обычно от 0 до 1,5. Пусть  $h_c=0,5$  м.

$h_p$  – высота рабочей поверхности над полом (0,8 м).

$$h = 3,6 - 0,5 - 0,8 = 2,3.$$

Далее определим расстояние между светильниками ( $L$ ) (формула 23):

$$L = \lambda * h \quad (23)$$

где:  $\lambda$  – коэффициент оптимального расстояния между светильниками (для люминесцентных ламп равно 1,1).

$$\text{Тогда } L = 1,1 * 2,3 = 2,53 \text{ м.}$$

Расстояние  $I$  от крайних светильников стены:  $I = L/3 = 2,53/3 = 0,84$  м.

Далее определим количество светильников, для начала по формуле 24 найдем общее количество рядов по ширине помещения:

$$n_{\text{ряд}} = \frac{a - 2/3L}{L} + 1, \quad (24)$$

где:  $b$  – ширина помещения;

$$n_{\text{ш}} = \frac{5 - 2/3 * 2,53}{2,53} + 1 = 2,31 \text{ ряда. Принимаем количество рядов по ширине } n_{\text{ряд}} = 3.$$

Общее количество рядов по длине помещения находим по формуле 25:

$$n_{\text{д}} = \frac{a - 2/3L}{l_{\text{св}} + 0,5}, \quad (25)$$

где:  $a$  – длина помещения;

$l_{\text{св}}$  – длина светильника (1,23 м).

$$n_{\text{д}} = \frac{7 - 2/3 * 2,53}{1,23 + 0,5} = 3,07 \text{ рядов. Принимаем количество рядов по длине } n_{\text{д}} = 4.$$

Определяем общее количество светильников:  $n = 3 * 4 = 12$ .

Рассчитаем индекса помещения:  $i = S / h(a+b) = 35 / 2,3(7+5) = 1,26$ . Состояние потолка 70%, состояние стен 50%, выбраны лампы открытого дневного света,

следовательно по таблице коэффициентов использования светового потока светильников с люминесцентными лампами  $\eta=0,53$ .

Далее определим световой поток ламп по формуле 26:

$$\Phi = \frac{E * S * k * z}{\eta * n * x} \quad (26)$$

где: E – минимальная освещенность;

k – коэффициент запаса равен 1,5, так как помещение с малым выделением пыли;

S – площадь помещения;

n – число светильников в помещении;

$\eta$  – коэффициент использования светового потока;

z – коэффициент неравномерности освещения;

x – число источников света в светильнике.

$$\Phi = \frac{300 \cdot 35 \cdot 1,5 \cdot 1,1}{0,53 \cdot 12 \cdot 2} = \frac{17325}{12,72} = 1362,02 \text{ лм}$$

Выберем лампу дневного света со световым потоком 1650 лм, мощность 30 Вт.

Далее делаем проверку выполнения условия:

$$\Delta = \frac{1650 - 1362,02}{1650} \cdot 100\% = 17,45\%$$

Получаем:  $-10\% \leq 17,45\% \leq +20\%$  выбранная лампа удовлетворяет условиям.

Определим освещенность в помещении (формула 27):

$$E_{\phi} = \frac{N * F * \eta}{S * z * k} \quad (27)$$

где: N – число ламп, шт.;

F – световой поток лампы, лм;

$$E_{\phi} = \frac{24 * 1650 * 0,53}{35 * 1,1 * 1,5} = 364 \text{ лк.}$$

Значение освещенности помещения удовлетворяет требованиям (от 200 до 400 лк).

Определяем мощность осветительной системы помещения по следующей формуле 28 [37]:

$$\sum N = w * n * x \quad (28)$$

где: N - суммарная мощность осветительной системы, Вт;

w - мощность одной выбранной лампы;

n - принятое число светильников, штук;

x - количество ламп в светильнике, штук.

$$\sum N = 30 \cdot 12 \cdot 2 = 720 \text{ Вт}.$$

При проектировании искусственного освещения важно правильное размещение светильников, так чтобы не было прямого света в глаза пользователя.

### 5.1.1.2 Микроклимат

Основными показателями микроклимата в помещении являются: температура воздуха помещения, скорость движения и влажность воздуха. Если все показатели находятся в норме, значит, человек находится в комфортных климатических условиях и у него не возникает никаких дискомфортных ощущений, его механизмы терморегуляции работают с минимальным напряжением и не вызывают отклонений в состоянии здоровья.

В случаях нарушении норм микроклимата у человека может быть тепловое облучение, при недостаточной влажности – нарушение умственной деятельности, снижение зрения. Поэтому следует соблюдать микроклимат, чтобы минимизировать воздействие вредных факторов. Работа за компьютером, в основном работа в сидячем положении, поэтому категория работ рассмотрим I а (работы с интенсивностью энергозатрат до 139 Вт). В таблицах 25,26 представлены оптимальные и допустимые нормы микроклимата соответственно [38].

Таблица 25. Оптимальные нормы микроклимата

Период года	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	22-24	21-25	60-40	0,1
Теплый	23-25	22-26	60-40	0,1

Таблица 26. Допустимые нормы микроклимата

Период года	Температура воздуха, °С		Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с	
	диапазон ниже оптимальных величин	диапазон выше оптимальных величин			ниже оптимальных величин	выше оптимальных величин **

Холодный	20,0-21,9	24,1-25	19,0-26,0	15-75*	0,1	0,1
Теплый	21,0-22,9	25,1-28,0	20,0-29,0	15-75*	0,1	0,1

Для поддержания микроклимата в помещении следует проветривать помещения, проводить проверки и измерения, иметь систему отопления, термометр и гигрометр (показывает процент влажности воздуха). Температура воздуха в помещении от 22 до 24°C, влажность воздуха около 50%. Перед началом рабочего дня и в обеденный перерыв помещение проветривается, параметры микроклимата соответствуют нормам, следовательно, в помещении соблюдаются необходимые для работы условия.

#### **5.1.1.3 Повышенный уровень шума**

Помещение, в котором находится рабочее место, не содержит сильных источников шума и вибрации. В основном шум создается за счет работы компьютеров и другой оргтехники. В компьютере основным источником шума можно выделить вентиляторы охлаждения системных блоков. Влияние шума на организм человека зависит от того, как долго человек находится под его влиянием, от интенсивности звуков и их периодичности.

В первую очередь шум влияет на органы слуха, это может быть заметно не сразу, болезнь может развиваться постепенно. Шум вызывает быстрое утомление, бессонницу, головную боль, влияет на репродуктивную функцию и способствует серьезному расстройству психики.

Благодаря своевременному ремонту и замене комплектующих системных блоков в кабинете уровень шума не превышает 40 дБА. Согласно СанПин 2.2.2/2.4.1340-03 уровень звука в помещении, создаваемый ПЭВМ не должен превышать 50дБА[39].

#### **5.1.1.4 Повышенный уровень электромагнитных излучений**

Все приборы, работающие от сети, в частности компьютер, оказывают влияние на окружающее их электромагнитное поле, большая часть излучения распространяется от задней стенки монитора. От влияния электромагнитных полей на тело человека страдает его нервная система, сердечно-сосудистая система, гормональные нарушения. Первые симптомы: головная боль, общая слабость, замедление пульса. В таблице 27 представлены предельно-допустимые уровни электромагнитных полей на рабочих местах пользователей ПК[40].

Таблица 27. ПДУ электромагнитных полей

Нормируемые параметры		ПДУ
Напряженность электрического поля	5 Гц — 2 кГц	25 В/м
	2 кГц — 400 кГц	2,5 В/м
Напряженность магнитного поля	5 Гц — 2 кГц	250 нТл
	2 кГц — 400 кГц	25 нТл

Некоторые рекомендации, которые могут снизить влияние электромагнитного излучения:

- монитор лучше иметь жидкокристаллический, так как излучение мониторов с электроннолучевой трубкой намного сильнее;
- не оставлять включенным компьютер, если работа на нем закончилась;
- расположение монитора – в углу, чтобы стены поглощали излучение;
- расстояние до монитора должно быть не ближе вытянутой руки;
- влажная уборка помещения и наличие ионизатора послужат дополнительной защитой от электромагнитного излучения компьютера.

#### **5.1.1.5 Зрительное напряжение и монотонность трудового процесса**

На зрение влияет освещение, отражение света, вид монитора, положение тела. При постоянной работе за компьютером может возникнуть зрительное напряжение и усталость, размытие или двоение изображения, сухость глаз, раздражение. Зрительное напряжение возникает из-за того, что зрение постоянно фокусируется на экране, все картинки, находящиеся на экране, находятся в одной плоскости, поэтому человек не переводит взгляд с ближних объектов на дальние, сокращается количество морганий, вследствие чего глаз не увлажняется слезной жидкостью.

Поэтому следует при непрерывной работе за монитором каждые 20-30 минут давать глазам отдохнуть, можно переводить взгляд на удаленные объекты и смотреть на него несколько секунд. Следует чаще моргать, делать гимнастику для глаз, для сохранения зрения можно использовать специальные компьютерные очки, которые созданы для повышения контрастности изображения и устранения бликов, регулярно проходить проверку зрения и выполнять рекомендации врача.

Очень часто человек сталкивается со зрительным напряжением, нарушением осанки, с синдромом запястного канала, когда трудовой процесс происходит монотонно, без перерывов. Поэтому следует устанавливать рациональный режим труда и отдыха. При



введении отдыха в течение трудового дня можно предотвратить и замедлить наступление утомления.

По инструкции Р-45-084-01 длительность работы за компьютером без перерыва может быть не более двух часов. При 8-ми часовом рабочем дне в зависимости от категории работы, минимальное суммарное время регламентированных перерывов составляет 30 минут. Необходимо устраивать каждый час короткие перерывы в работе, во время которых желательно походить по помещению, сделать несколько потягиваний руками, размять шею и ноги, кисти [41].

### **5.1.2 Анализ опасных факторов**

Все выявленные опасные факторы относятся к электробезопасности:

- поражение электрическим током;

Причины поражения электрическим током: прикосновение к незаземленным проводам, к металлическим частям компьютера, которые в результате повреждения изоляции находятся в напряжении; использование неисправных электроприборов; одновременное прикосновение к двум проводам, которые находятся под напряжением. В зависимости от силы тока реакция на организм может быть различной: от слабо ощутимых воздействий, судорог до глубоких ожогов. В случае повреждения электрическим током необходимо прекратить действия тока на пострадавшего, оценить состояние дыхательной и сердечно-сосудистой системы, при отсутствии пульса принять меры первичной реанимации и при любых повреждениях вызвать скорую помощь.

- повышенный уровень статического электричества;

Источниками статического электричества в помещении являются компьютеры, оргтехника, электроприборы. У компьютера основным источником является экран монитора. Согласно СанПин 2.2.2/2.4.1340-03 допустимый уровень электростатического потенциала монитора – 500 В.

Длительное воздействие статического электричества влияет: на сердечно-сосудистую систему организма человека, центральную нервную систему. Помимо этого у человека может наблюдаться раздражительность, головная боль, нарушение сна.

К средствам защиты от повышенного уровня статического электричества относятся: заземляющие и увлажняющие устройства, нейтрализаторы, антиэлектростатические вещества и экранирующие устройства [42]. Чтобы защитить себя от воздействия статического поля необходимо делать влажную уборку в кабинете, увлажнять воздух, проветривать помещение, обязательно необходимо заземлять всю технику.

- короткое замыкание – это соединения двух точек электрической цепи, в результате которого возникает сильный ток, что приводит к возгоранию или выходу электрического прибора из строя. Во избежание повреждения изоляции проводов и возникновения коротких замыканий не следует вешать что-либо на провода, закрашивать и белить шнуры и провода, закладывать провода и шнуры за батареи, выдергивать штепсельную вилку из розетки за шнур, усилие должно быть приложено к корпусу вилки.

Мероприятия защиты при электробезопасности следующие:

- перед началом работы проверить исправность электропроводки, розеток и вилок компьютера, заземление компьютера;
- чистка компьютера должна производиться без напряжения;
- работать за компьютером следует с сухими руками;
- при аварийных ситуациях необходимо сразу отсоединить компьютер от сети;
- не вступать в контакт с оголенным проводом;
- проведение инструктажа;
- проверка знаний правил безопасности и инструкций в соответствии с выполняемой работой.

## **5.2 Экологическая безопасность**

Вопрос экологической безопасности является важным при исследовании. Ведь состояние защищенности природной среды и возможного негативного воздействия на неё зависит от самого человека.

Анализ негативного воздействия на окружающую среду включает в себя утилизацию люминесцентных ламп, компьютерной техники и оргтехники. Каждая люминесцентная лампа содержит в себе 3-5 мг ртути, которая находится в агрегатном состоянии в виде паров. Поврежденная лампа, высвобождая пары ртути, может вызвать тяжелое отравление. При вдыхании ее паров, организм человека получает поражение на нервную систему, печень, почки, желудочно-кишечный тракт. По гигиенической классификации ртуть относится к первому классу опасности (чрезвычайно опасное химическое вещество). Предельно допустимая концентрация ртути в атмосферном воздухе и воздухе жилых, общественных помещений составляет 0,0003 мг. В ходе повреждения одной лампы, в помещении без проветривания в течение нескольких часов концентрации ртути в воздухе может достигнуть до 0,05 мг и больше, что выше допустимого значения. Так же опасны выбросы и осадения ртути с осадками в воду, поскольку в результате деятельности

микроорганизмов происходит образование растворимой в воде и токсичной метилртути. Органические соединения ртути более токсичны, чем неорганические.

Поэтому должен быть организован централизованный сбор и утилизация люминесцентных ламп, но обычно отработанные лампы выбрасываются вместе с обычным мусором с последующим размещением на полигонах твердых бытовых отходов, что недопустимо.

Компьютеры и оргтехнику в случае выхода из строя, либо по причине морального износа необходимо утилизировать. В том случае, если оргтехнику просто выбрасывают на мусорку, на старое оборудование могут воздействовать атмосферные осадки, что может привести к возникновению различных химических реакций с выделением вредных веществ. Утилизация компьютерной техники производится через специализированные организации, имеющие на это лицензию. В любой оргтехнике имеется определенное количество драгоценных металлов. Согласно российскому законодательству все организации в обязательном порядке должны проводить учет и движение драгоценных металлов. После проведения комплексной утилизации все чистые металлы сдаются в Госфонд России, что значительно улучшает экологическую ситуацию в стране.

### **5.3 Безопасность в чрезвычайных ситуациях**

По виду чрезвычайные ситуации могут быть: техногенные, экологические и природные. В ходе анализа была выявлена техногенная чрезвычайная ситуация, такая как пожар. Пожар в помещении может привести к порче имущества, потере ценной информации и даже к угрозе жизни и здоровья людей. В помещении находятся электропровода напряжением 220В, розетки и выключатели.

Основными источниками возгорания могут быть:

- короткое замыкание в электропроводке;
- неработоспособное оборудование;
- неисправности в проводке, розетках и выключателях;
- возгорание ламп искусственного освещения;
- возгорание мебели в помещении из-за нарушения правил пожарной безопасности.

Согласно своду правил 12.13130.2009 помещение относится к категории В[43].

В случае возникновения пожара необходимо:

- сообщить по телефону «01» в пожарную охрану;

- дать сигнал о пожаре, тем самым оповестить всех людей, находящихся в здании;
- принять меры по организации эвакуации людей;
- самостоятельно имеющимися средствами пожаротушения приступить к тушению пожара;
- отключить электроэнергию.

Мероприятия по обеспечению пожарной безопасности:

- проведение противопожарного инструктажа;
- обучение правилам техники безопасности;
- проверка проводки;
- наличие в помещении планов эвакуации;
- соблюдение эксплуатационных норм оборудования.

В каждой учебной аудитории назначается ответственное лицо за соблюдения правил пожарной безопасности. В коридоре помещения расположен пожарный щит, установлен рубильник, имеется порошковый огнетушитель типа ОП, предназначенные для пожаров классов А,В,С,Е, на каждом этаже приведен план эвакуации в случае возникновения пожара. В учебных кабинетах имеется порошковый огнетушитель типа ОП и углекислотный огнетушитель типа ОУ. Помимо этого в здании должны быть освобождены проходы к эвакуационным и аварийным выходам.

Обязательной частью системы пожарной безопасности является внутренний противопожарный водопровод. Краны должны быть оборудованы рукавами и стволами, помещенными в шкафы, которые пломбируются. В шкафу должен находиться рычаг для облегчения открытия крана. На дверях шкафа по правилам должны быть указан индекс и порядковый номер пожарного крана. Так же они должны периодически проходить техническое обслуживание и проверяться на работоспособность. О результатах проверок должны составляться акты.

Соблюдение правил пожарной безопасности позволит избежать возникновения пожара, а в случае его наступления, сохранить жизнь себе и спасти других людей.

#### **5.4 Правовые и организационные вопросы обеспечения безопасности**

Рассмотрим организацию рабочих мест. По СанПин 2.2.2/2.4.1340-03 при установке рабочих мест с компьютерами расстояние между рабочими столами с видеомониторами должны быть не менее 2 м, между боковыми поверхностями видеомониторов - не менее

1,2 м. Расположение экрана монитора должны быть 600-700 мм. Оптимальное расположение клавиатуры 100-300 мм от края стола. Сиденье должно быть удобным, регулироваться по высоте и наклону спины, это необходимо для снижения нагрузки мышц, под ноги можно устанавливать по соответствующим нормам подставку для ног.

Что касается режима труда, длительность работы не должна превышать 8 часов, в зависимости от времени работы устанавливаются перерывы в пределах от 20 до 30 минут, на обед предусмотрено от 40 до 60 минут. Каждый студент, руководитель или сотрудник должен быть ознакомлен с техникой безопасности, инструкцией по охране труда. После ознакомления он должен расписаться в специальном журнале.

Рассмотрим правовые вопросы обеспечения безопасности. Работа за компьютером и разработка программного обеспечения относятся к категории не тяжелых работ, не предусматривает ненормированный рабочий график. Поэтому в данной работе не предусмотрено применение режима сокращённого рабочего дня, запрещение использования труда женщин и подростков, компенсаций за вредный труд, ночные смены.

Все условия труда должны соответствовать допустимым нормам, установленные Российским законодательством. Правовую основу обеспечения безопасности жизнедеятельности составляют соответствующие законы и постановления, принятые представительными органами Российской Федерации. Основным закон государства – Конституция, остальные законы и правовые акты не должны противоречить ей. Задачами экологической безопасности является охрана природы, предупреждение вредного воздействия. В трудовой деятельностью важным является охрана труда. Каждый человек имеет право на труд, рабочее место, защищенное от воздействия вредных или опасных производственных факторов. Все эти аспекты регулируются соответствующими органами и законами.

#### **Выводы по разделу:**

В данном разделе был выполнен анализ опасных и вредных факторов, приведено описание каждого фактора, его действие на организм человека и рекомендуемые меры по соблюдению норм показателей. Был выполнен расчет искусственного освещения помещения, рассмотрено вредное воздействие на окружающую среду, описана чрезвычайная ситуации и рассмотрены правовые и организационные вопросы безопасности. Рабочее место соответствует требованиям действующих нормативов.

## Заключение

В результате выполнения магистерской диссертации была разработана алгоритмическая база системы поддержки принятия врачебных решений при выборе тактики лечения детей с эндокринопатиями.

Решены следующие задачи выпускной квалификационной работы:

- был проведен обзор литературных источников и публикаций, изучены отечественные и зарубежные практики решения задач обработки многомерных данных и получения новых знаний;
- был проведен обзор инструментария работы с многомерными данными, выбраны инструменты для решения задач, поставленных в ВКР, проведено обоснование выбора инструментария;
- была проведена первичная обработка подготовка данных – анализ выбросов и пропущенных значений, сформирован массив данных, готовый к дальнейшему исследованию;
- была решена задача снижения размерности многомерных данных за счет выявления и отбора информативных признаков;
- была построена база знаний на основе деревьев решений, определяющих выбор траектории лечения, была проведена оценка построенной модели;
- сформирован алгоритм принятия решения о выборе траектории лечения для детей с эндокринопатиями.

Построенная в рамках работы над ВКР система поддержки принятия решений проходит опытную эксплуатацию на базе НИИ Курортологии и физиотерапии.

Результаты работы были доложены:

(перечень конференций)

Публикации по теме исследования:

1. Емельянова Ю. А. Оценка информативности многомерных данных//Молодёжь и современные информационные технологии: сборник трудов XV Международной научно-практической конференции студентов, аспирантов и молодых ученых. Томск, 04-07 декабря 2017 г. – 2018 – Томск: Д-Принт. – С.356-357. (Выступление с докладом, диплом 3 степени).

2. Емельянова Ю. А. Информационная технология оценки информативности показателей основных физиологических систем организма в задаче диагностики степени ожирения у детей и подростков// Информационные технологии и технологии

коммуникаций: современные достижения. Сборник материалов I Молодежной международной конференции. – Астрахань: Издатель: Сорокин Роман Васильевич, 2017. – С. 235-237.

3. Емельянова Ю. А. Оценка информативности показателей основных физиологических систем организма в задаче диагностики степени ожирения у детей и подростков//Современные технологии медицинской реабилитации и санаторно-курортного лечения: материалы Всероссийской научно-практической конференции, посвященной 95-летию со дня основания Томского научно-исследовательского института курортологии и физиотерапии. Томск, 2-3 ноября 2017 г. – С.24-28. (Выступление с докладом).

4. Емельянова Ю. А. Применение математико-статистических методов в задаче исследования значений клиничко-лабораторных показателей пациентов до и после лечений//Информационные технологии в науке, управлении, социальной сфере и медицине: сборник научных трудов IV Международной научной конференции. В 2 частях, Томск, 5-8 декабря 2017. – Томск: ТПУ, 2017 – Ч.1 – С.331-335. (Выступление с докладом, диплом 2 степени).

5. Marukhina O.V., Berestneva O.G., Emelyanova Yu.A. , Romanchukov S.V., Petrova L., Lombardo C., Kozlova N.V. Basic physiological systems indicator's informative assessment for children and adolescents obesity diagnosis tasks//The open access Journal of Physics: Conference Series (InPress).

6. Емельянова Ю.А. Система поддержки принятия врачебных решений при выборе тактики лечения детей с эндокринопатиями// VIII Международный молодежный научно-культурный форум «Новые форматы транснациональной научно-образовательной деятельности». (Выступление с докладом, диплом 2 степени, в печати).

## Список источников:

1. Карпов О.Э., Клименко Г.С., Лебедев Г.С. Применение интеллектуальных систем в здравоохранении// Современные наукоемкие технологии. – 2016. –№ 7 (часть 1) – С. 38–43.
2. Чеснокова И.В., Чернов В.И., Чернов А.В. Система поддержки принятия врачебных решений на догоспитальном этапе в терапии артериальной гипертензии//Системный анализ и управление в биомедицинских системах. – 2007. –№ 3 – С.675–679.
3. Никитина М.И. Трехуровневая система поддержки принятия решений перинатальному мониторингу//Образовательные ресурсы и технологии. – 2014. – С. 151–156.
4. Атьков О., Кудряшов Ю., Прохоров А., Касимов О. Система поддержки принятия врачебных решений// Научная электронная библиотека «КиберЛенинка». – 2013.
5. Найданов Ч.А. Система поддержки принятия решений для предупреждения рисков возникновения критических состояний//Альманах современной науки и образования. – 2015. –№ 8(98). – С.92–95.
6. Купеева И.А., Разнатовский К.И., Раводин Р.А., Карелин В.В. Оценка эффективности интеллектуальной системы поддержки принятия врачебных решений// Vestnik Sankt-Peterburgskogo Universiteta. – 2016. – Seriia 10, p62–68.
7. Преферанский Н.Г. Экспертная составляющая индивидуальной комбинированной фармакотерапии как элемента системы поддержки принятия врачебных решений// Информационно-измерительные и управляющие системы. –2011. –№ 9(12) – С.50–52.
8. Усов В.Ю. Структура медицинской информационной системы поддержки решений, использующей алгоритмы интеллектуального анализа данных// Вестник Воронежского института высоких технологий. – 2017. - № 2 – С.122–126.
9. Спирычин А.А., Бурковский В.Л., Воропаев А.П. Анализ эффективности использования метода нечеткой классификации и генетических алгоритмов в интеллектуальной системе поддержке принятия решений//Вестник Воронежского государственного технического университета. – 2015.
10. Судок О.В., Петрова Т.Н., Алексеев Н.Ю., Фурсова Е.А. Построение информационного комплекса поддержки принятия врачебных решений в лечебно-диагностическом процессе больных сахарным диабетом в сочетании с артериальной гипертензией//Прикладные информационные аспекты медицины. – 2015.- №6 – С.4–9.



11. Чеснокова И.В. Использование системы поддержки принятия врачебных решений для диагностики и терапии артериальной гипертензии// Научно-медицинский вестник Центрального Черноземья. – 2014. – №58 – С.162–167.
12. Кнышов Г.В., Настенко Е.А., Яковенко, А.В. Особенности проектирования медицинской информационной системы поддержки принятия решений, основанной на интеллектуальном анализе данных// Кибернетика и вычислительная техника. – 2014. –№ 3–С.79–87.
13. Никитяев В., Нагуманова Ю., Проничев А, Чистов К. Высокотехнологичная система поддержки принятия врачебных решений при диагнозе острых лейкозов//// Научная электронная библиотека «Киберленинка». – 2011. – С.70–73.
14. Евдокимова Е.Ю., Попова У.Ю. Ожирение у детей. Маркеры метаболического синдрома у детей// Вестник Совета молодых учёных и специалистов Челябинской области. -2017. - №1(2) – С.16–19.
15. Dietz WH. Health consequences of obesity in youth: childhood predictors of adult disease// The causes and health consequences of obesity in children and adolescents Pediatrics (Evanston). – № 101(3). – p.518–525.
16. Мрясева В.В., Узунова А.Н. Ожирение у детей. Проблемы и способы их решения// Непрерывное медицинское образование и наука. – 2016.- № 11(3) – С. 20–23.
17. Картелишев А., Румянцев А., Смирнова Н. Актуальные проблемы ожирения у детей и подростков// Медпрактика. – 2010. – 280 с.
18. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. [Электронный ресурс]. URL: [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(17\)32129-3/fulltext?elsca1=tlpr](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)32129-3/fulltext?elsca1=tlpr).
19. Серебрякова В., Каждый десятый юноша-третьекурсник в Томске имеет избыточную массу тела. Официальный интернет-портал Администрации Томской области. [Электронный ресурс]. URL: <https://tomsk.gov.ru/news/front/view/id/13229>.
20. Gomez–Ambrosi J, Silva C, Galofre JC, Escalada J, Santos S, Millán D, Vila N, Ibanez P, Gil MJ, Valenti V, Rotellar F, Ramírez B, Salvador J, Fruhbeck G. Body mass index classification misses subjects with increased cardiometabolic risk factors related to elevated adiposity//International Journal of Obesity. –2012, Vol. 36 Issue 2, p286, 9 p.
21. Сигел Э. Практическая бизнес–статистика. – М.: Издательский дом "Вильямс". – 2008. — 1056 с.

22. Калинина В.Н., Соловьев В.И. Введение в многомерный статистический анализ: Учебное пособие/ ГУУ. – М., 2003. – 66с.
23. Колесникова С.И. Методы анализа информативности разнотипных признаков. [Электронный ресурс]. URL: <http://cyberleninka.ru/article/n/metody-analiza-informativnosti-raznotipnyh-priznakov> (дата обращения: 17.05.2017).
24. Курносков Ю.В. Алгебра аналитика. Секреты мастерства в аналитической работе. Базы знаний. – М: РУСАКИ. – 2015. – 227с.
25. Гаврилова Т.А., Хорошевский В.Ф.. Базы знаний интеллектуальных систем – СПб: Питер, 2000 – 384 с.
26. Коробова И.Л. Методы представления знаний. – Тамбов: Изд-во Тамб. гос. техн. ун-та. – 2003. – 24 с. [Электронный ресурс]. URL: <http://www.tstu.ru/book/elib/pdf/2003/korobova.pdf> (дата обращения: 14.11.2017)
27. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – СПб.: БХВ Петербург, 2005. – 736 с.
28. Панфилов С.А., Язенин А.В.. Генетический алгоритм оптимизации структуры лингвистических переменных при построении баз знаний нечетких систем// Программные продукты и системы. – 2004. – №1(65).
29. Сидоров А.В. Алгоритмы создания дерева принятия решений. [Электронный ресурс]. URL: <http://econf.rae.ru/pdf/2014/03/3245.pdf> (дата обращения: 25.01.2018)
30. Шитиков В.К. Базовые графические возможности R: диаграммы размахов. [Электронный ресурс]. URL: [http://ranalytics.blogspot.ru/2011/11/r\\_08.html#.WoP3m4PFK70](http://ranalytics.blogspot.ru/2011/11/r_08.html#.WoP3m4PFK70) (дата обращения: 14.02.2018).
31. Поворознюк А.И., Гуторова Т.В. Оценка информативности медицинских показателей в компьютерных системах медицинской диагностики // Вестник НТУ "ХПИ". Сборник научных трудов.–2004. – № 8.– С. 45–48.
32. И.Б. Богер, Н.В. Шаланов. Методический подход к статистической оценке информативности системы показателей//Сибирская финансовая школа. – 2012. - №6(95) – С. 158–160.
33. С.И. Колесникова. Методы анализа информативности разнотипных признаков/Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2009. –№ 1(6). – С. 69–80.
34. И.С. Голованова. Выбор информативных признаков. Оценка информативности. [Электронный ресурс]. URL: <http://ime.tpu.ru/study/disciplinary/INF-PR.pdf> (дата обращения: 20.02.2018).

35. Е.В. Гублер. Вычислительные методы анализа и распознавания патологических процессов, 1978, 269 с.
36. СанПин 2.2.1/2.1.1.1278-03. Гигиенические требования к естественному, искусственному и совместному освещению жилых и общественных зданий. – М.: Госкомсанэпиднадзор, 2003.
37. О.Б. Назаренко. Расчёт искусственного освещения. Методические указания к выполнению индивидуальных заданий для студентов дневного и заочного обучения всех направлений и специальностей ТПУ. – Томск: Изд. ТПУ, 2008. – 20 с.
38. СанПин 2.2.2/2.4.1340-03. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы. – М: Минздрав России, 2003.
39. СанПин 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений. – М: Минздрав России, 1997.
40. СанПин 2.2.4.3359-16. Санитарно-эпидемиологические требования к физическим факторам на рабочих местах, 2016.
41. ТОИ Р-45-084-01. Типовая инструкция при работе на персональном компьютере. – 2001.
42. ГОСТ 12.4.011-89. Система стандартов безопасности труда. Средства защиты работающих. Общие требования и классификация. – М: ИПК издательство стандартов, 2001.
43. СП 12.13130.2009. Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности. – М: 2009.

**Приложение А**  
**(справочное)**

**Разделы:**

Введение

Глава 1. Обзор литературы

Глава 2. Поиск методов решения задачи

Глава 3. Реализация проекта

Студент:

Группа	ФИО	Подпись	Дата
8КМ61	Емельянова Юлия Андреевна		

Консультант отделения ИТ:

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

Консультант – лингвист отделения ИЯ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Краснова Т.И.			

## **Introduction**

The healthcare computerization creates opportunities to the clinical decision support system development. In the course of diagnosis, doctor manipulates a considerable amount of data and makes a decision in the context of uncertainty basing upon the first-hand experience and knowledge. The situation is exacerbated by the fact that the knowledge scope in medicine is incrementally growing, but the decision-making time does not increase. The amount of medical malpractice is growing and it leads to various negative effects, even the mortality rate increase. IT-solution development for clinical purposes is one of the most promising and efficient ways to prevent these effects. That is why the efforts of many IT specialists are directed to the doctor's heuristics simulating software or expert-based medical decision-making algorithms development. Thus, the objective of this study is to develop clinical decision support system in the choice of tactics of treatment of children with endocrinopathy.

The object of the study is the problem of overweight in children and teenagers. The subject of the study is clinical and laboratorial parameters of patients.

Tasks of final qualifying work are:

- review of literature sources and publications for solutions of domestic and foreign practices of solutions for processing multidimensional data and obtaining new knowledge;
- overview of the tools for working with multidimensional data;
- initial processing of data preparation;
- estimation of informative value assessment of clinical and laboratorial parameters;
- formation of a decision-making algorithm for choosing the trajectory of treatment for children with endocrinopathy.

The topicality of the research consists of creation of clinical decision support system to avoid many diagnostic mistakes.

The topicality is based on following aspects:

- the “Modernization of medical health care organization processes with use of information technologies” project is started by the Russian Federation government;
- there is an important goal to stabilize the part of Russian population with obesity;
- there are a lot of medical data that should be processed promptly.

The novelty of the research is creation of decision-making algorithm for choosing the trajectory of treatment for children with endocrinopathy based on clinical and laboratorial

parameters analysis. Clinical decision support system is being operated on the basis of Tomsk Research Institute of Health Resort Study and Physical Therapy.

## **Chapter 1.Literature review**

Endocrinopathy is a generalizing concept for diseases associated with disorders of the endocrine glands. Symptoms of endocrinopathy are:

- increase in growth that exceeds the norm,
- increase of separate body parts,
- muscle weakness and fast fatigue from usual physical activities,
- weight gain,
- changes in blood serum,
- heart disruption.

One of the most common pathologies of endocrinopathy is obesity. Obesity is a chronic metabolic disorder, accompanied with excessive deposition of adipose tissue.

In 1998 the World Health Organization (WHO) mentioned obesity as a global epidemic. Nowadays it is one of the most widespread chronic conditions in the world: by the beginning of the XXI century, over 30% of the world population was overweight [1]. The major medical consequences of obesity such as the non-insulin dependent diabetes and cardiovascular diseases lead to the high disability and premature mortality rate [2-4]. Over 60% of adults has an obesity that appeared in their childhood, which is only worsening and causes the major complications [5-7]. The pediatric and adolescent obesity that is prolonged in adult life has a more severe disease course and is accompanied by a more significant weight gain and concurrent disease incidence rate than the obesity appeared at a mature age [8,9].

In 2004, the World Health Organization put forward an initiative: “The Global Strategy on Diet, Physical Activity and Health (DPAS)”. The key element of this initiative is in-patient program supplemented by professional and therapeutic training aimed at the body weight reduction. Therefore, in Russia up to the present moment, there has been no targeted obesity detection programs and obesity is diagnosed lately. According to A.V. Kartelishev [10] 65% of all obese children falls on the first-degree obesity children but only 5,5% of them attend the doctor.

In 2014, the Government of the Russian Federation declared the list of top-priority research missions [11] that require engaging all available resources. This list includes a socially significant and orphan endocrine system diseases personalized treatment development (mission

№7). It is necessary to reduce the amount of obese people in Russia down to rate of 5-10%, number of non-insulin dependent diabetes cases to 3% and disability diagnosis rate resulting from endocrine system disease complications by 30-50%.

Currently, medical problems can be solved with help of actively developing information technologies. The technologies are covering almost all aspects of medical organizations activities, including management of resources, treatment and medical care. One of the most perspective direction is the development of clinical decision support systems (CDSS). CDSS serve as a basis for fundamentally new approaches creation. The approach solves traditional healthcare tasks such as monitoring a patient, consulting, deciding on further treatment.

The medical research support systems came into common use not so long ago even though the theoretical studies devoted to these systems were carried out at the end of 1950s. In general, the term “decision support system” (DSS) means a computer system that can influence the decision-making processes in different branches of human activities by information acquisition and analysis. In healthcare, these systems are named as “clinical decision support systems”. Studies in this field have been carried out in various directions [12-23] for 30 years. According to e-Library, the amount of publications on this subject is constantly growing in our country, mostly in the last two years.

Such authors as Atkov O., Kudryashov Yu., Prokhorov A develop a system for cardiovascular disease treatment. The system helps a doctor to choose clinical decisions for eight cardiovascular diseases (acute coronary syndrome, hypertension, etc.) In the article “Deformations and degenerative spine diseases” of Naidanov Ch.A. is describing DSS for decreasing of critical state in the field.

Nowadays there are several main approaches to medical decisions support that are aimed to increase their efficiency, there is an on-line analytical processing toolkit, an algorithmic approach (pattern recognition techniques, artificial intelligence, fuzzy logic, applied mathematical statistics etc.), knowledge-based systems (datamining, expert systems) and external resource-based systems (evidence-based resource databases, doctor and patient forums). There is also a set of knowledge retrieval systems and some new software products.

Among the homegrown technologies, it is possible to mention EXNA, ACCOD, OTEKS systems (Novosibirsk), IMSLOG integrated tool set (Yankovskaya A.E., Tomsk), DeepDataDriver system (Dyuk V.A. and Aseev M.G., St. Petersburg) and General-purpose Qualifier (Yudin V.Sch., Moscow). WIZWHY system (WizSoft, USA) based on the limited search algorithms to find out some Boolean data patterns is widely used among foreign programs. There are also some widely used decision tree systems such as See5/C5.0 (RuleQuest,

Australia) and statistics packages SAS (SAS Institute Company), SPSS, STATGRAPICS, STADIA, STATISTICA and etc. that have DataMining elements. BrainMaker (CSS) and OWL (HyperLogic) are popular for neural networks design. CART (USA) and PatternRecognitionWorkbench (Unica, USA) were developed for classification and regression tree system creation. Stefanyuk V.L., Osipov G.S., Averkin A.N., Fominykh I.B., Popov E.V., Pospelov D.A., Gavrilova T.A., Finn V.K., Khoroshevsky V.F., Borisov A.N., E. Mamdani and L. Zadeh, G. Klir and other scientists carried out researches on computer expert systems. N. A. Korenevsky, E. S. Podvalny, B. A. Korbinsky, O. V. Rodionov, E. N. Korovin, V. N. Frolov, D. Ferruccio, E. S. Boerner and others devoted their articles to the computer-assisted diagnosis issues. Meanwhile, the decision-making process efficiency improvement tasks during the obesity diagnosis and treatment are still relevant.

For example, Behnam Malmir published the article “A medical decision support system for disease diagnosis under uncertainty”. This paper presents a decision support system (DSS) modeled by a fuzzy expert system (FES) for medical diagnosis to help physicians make better decisions. The proposed system collects comprehensive information about a disease from a group of experts. A fuzzy rule based system is then formed based on this information, which contains a set of significant symptoms relevant to the suspected disease. Two case studies on kidney stone and kidney infection were conducted to demonstrate how the proposed method could be used. The results show that the proposed fuzzy expert system is capable of diagnosing diseases with a high degree of accuracy and precision comparing to a couple of machine learning methods [24].

Chernaya, KS, Tymchenko, BI, Komleva, N.O. described in their article a decision support system for automated medical diagnostics. For diagnosing automation, the state of bronchopulmonary system DSS DiaSpectrEx was developed. The analysis of changes in qualitative and quantitative composition of air exhaled by the patient is the source of information about the damage of the respiratory tract, inflammatory processes and the effectiveness of the treatment. DiaSpectrEx system allows improving efficiency of pulmonary diseases diagnosing methods through the use of modern computerized equipment, machine learning algorithms and data processing.

According to the articles, the systems are created for different medical directions. One of the main aspects while working with decision support system data is to decrease the n-dimension attribute space. To solve this problem, the indexes must be selected with respect to their informative value assessment.



## Chapter 2 Choice of the task completion

### 2.1 Informative value assessment approaches

To set a diagnosis means to determine the medical condition or to make sure that it lacks. It can be done if the indexes common to this research target (patient) are obtained and analyzed properly. These indexes must be relevant and informative.

So, first of all, it is necessary to assess the index informative value and exclude the less informative indexes in order to reduce the attribute space. The index informative value depends entirely on how it helps to differentiate the condition we are concerned with. If the index can be regularly observed under different conditions, it is likely that it will not help the diagnostics and prediction. The more informative index corresponds to the greater the distance between random variables [25]. There are at least two approaches to assess the index informative value (energy and information approaches) [26].

The energy approach is based on the fact that the index informative value is assessed by the index value itself. The indexes are ordered by the magnitude and the index with the highest magnitude is considered as the most informative. For instance, during the amplitude-time analysis of ECG, the R wave amplitude is considered as the most informative index. Therefore, this approach to the informative value assessment can be inappropriate for the object recognition tasks. In fact, if an index has high magnitude, but is almost equal for various class objects, it is almost useless for class identification tasks.

On the other hand, index with low magnitude, which is different for various object classes, it can be easily used as a classifier. It means that the information approach is more appropriate for object recognition. If the object should be ranged in class during the recognition, it is possible to use the index probability distributions difference as a significant mark. Informative value assessment  $I(x_j)$  is an area of  $x_j$  index distribution that is not common with the other distribution areas of the same index.

### 2.2 Information approach

**Cumulative frequency method (CFM).** If there are two index samples (x) typical for two different classes, the x index empirical distributions and the cumulative frequencies (the total of frequencies from the initial one to the immediate distribution interval) can be calculated and built within the same axis. The maximum cumulative frequencies difference can be used to assess the informative value of the x index.

**Shannon method** suggests assessing the informative value as an average-weighted amount of data accounted for various index gradations. It describes index's information as the eliminated

entropy value.

So, the informative value  $j$ -index is:

$$I(x_i) = 1 + \sum_{i=1}^G (P_i \sum_{k=1}^K P_{i,k} \cdot \log_k P_{i,k}), \quad (1)$$

where  $G$  – number of index gradations;

$K$  – amount of classes;

$P_i$  – probability of the  $i$ -th index gradation.

$$P_i = \frac{\sum_k m_{i,k}}{N} \quad (2)$$

$m_{i,k}$  – rate of the  $i$ -th gradation in the  $K$ -class;

$N$  – total amount of observations;

$P_{i,k}$  – probability of the  $i$ -th index gradation in  $K$ -class.

$$P_{i,k} = \frac{m_{i,k}}{\sum_{k=1}^K m_{i,k}} \quad (3)$$

**Kullback method** is used to assess the informative value. In medicine, it is possible to deal with the diagnosis issues, establishing diagnosis and medical condition definition only when the informative indexes typical to the patient are obtained and analyzed. Thus, it is necessary to determine the most informative indexes that characterize the patients psychophysical condition. The diagnosis depends on a kind of datamining based on these indexes. There are various techniques to assess the informative value, e.g. Shannon method and the cumulative frequency method, but let us concentrate on Kullback method of informative value assessment.

Indexes informational content depends entirely on how much this index helps to differentiate the object state. The more informative index corresponds to the greater distance between the random variables. The most widespread informative value for these distances is Kullback measure [25].

In 1948, a logarithmic measure was proposed by N. Wiener and K. Shannon as a way to define the amount of information. They offered the formula that was recognized as information quantitative measure [25]. There is one measure that was introduced by Jeffreus in 1964 and was studied in detail by Kullback as an informative value  $J(1,2)$ , a measure of discrepancies between

statistical distributions 1 and 2. For discrete distributions, this formula is as follows:

$$J(x_i / A_1, x_i / A_2) = \sum_j \lg \frac{P(x_{ij} / A_1)}{P(x_{ij} / A_2)} [P(x_{ij} / A_1) - P(x_{ij} / A_2)] \quad (4)$$

where  $A_1, A_2$  – state classes;  $i$  – sign number;  $j$  – number of  $i$ -th sign range;  $P(x_{ij} / A_k)$  – probability of object hitting belonging to  $A_k$  class in the  $j$  sign range. Kullback method to determine the informative value can be presented with the following algorithm (fig. 1).

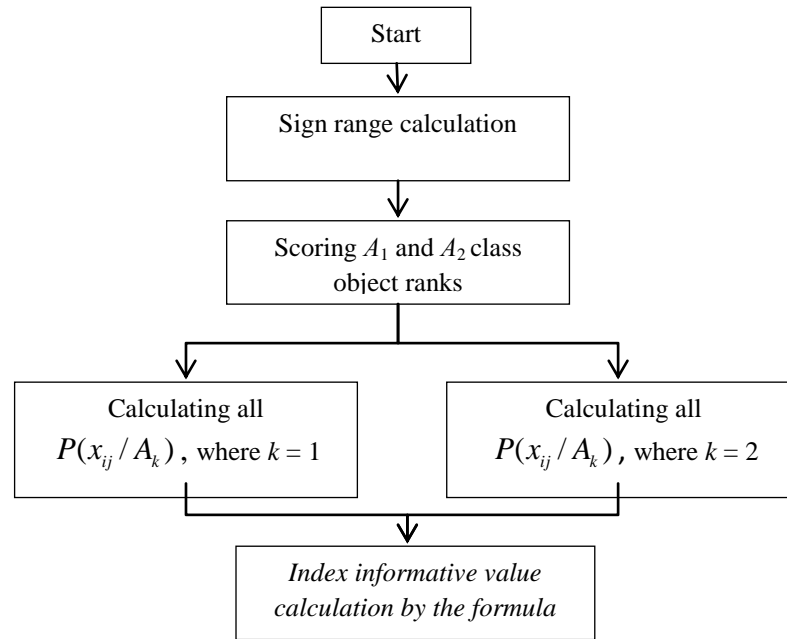


Figure 1. Kullback's informative value assessing algorithm

This criterion allows coming to conclusion on the differences of empirical aspects without any special restrictions on random distributions that form an empirical aspect. The indexes where the low informative value ranks are dominant, will have a low informative value, in other words, they will provide “correct” and “incorrect” diagnostic criteria with a close frequency. Thus, it is worth placing the indexes in a diagnostic table in an order of decreasing the informative value. Herein, the time rate of the right boundary reaching will be the highest upon the average, and the amount of mistakes will be the lowest.

To identify what contribution this index rank makes to reach the right diagnostic boundary and this is informative value  $J(x)$ , it is necessary to know not only the likelihood ratio (or its logarithms) but also to take into account the probability of  $A_1$  or  $A_2$  class people to get to this rank.

The probability for some people will be determined by the number of diagnostic criteria that will bring the answer to the boundary that is correct for the majority of observations of this rank. The probability for other people will be determined by the number of diagnostic criteria that will

pull the answer away from the boundary.

Table 1 shows criteria that determine the more appropriate technique to detect the index informative value for the given study:

Table 1. Comparative analysis of the index informative value assessing methods

	Index encoding degree	Number of classes	Dependence on sample size
CFM	Is sufficient	only two classes	sample size should be equal for both classes
Shannon method	Is insufficient	random number of classes	sample size can be different for different
Kullback method	Is insufficient	only two classes	sample size can be different for different

This study uses two sample types, the sample size is different for two classes and the selected technique does not depend on encoding, so it is possible to conclude that Kullback method is more appropriate to determine the index informative value.

The authors developed the NPP – a software product that uses Kullback method for informative value assessment. Programm interface is whown at figure 2 (but we have to mention that this software solution has only Russian localisation now).

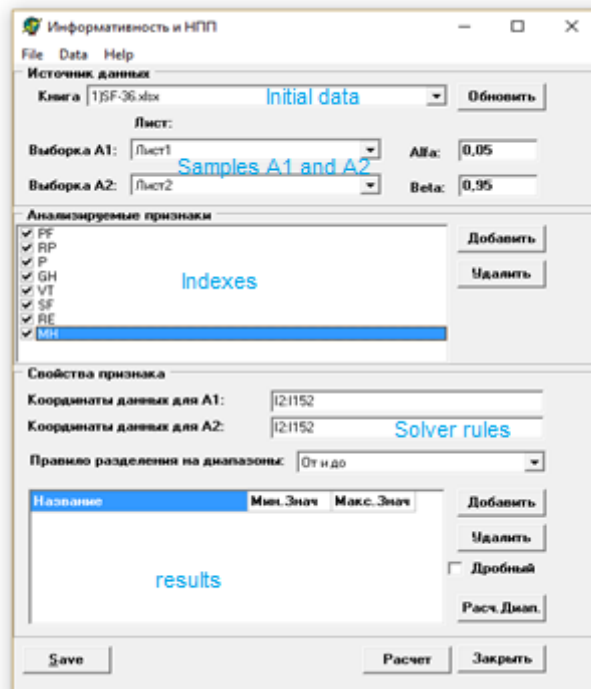


Figure 2. NPP interface

## Chapter 3. Project implementation

### 3.1 Research task description and solution

Tomsk Research Institute of Health Resort Study and Physical Therapy provided the following data sets: clinical findings, cardiovascular system indicators, physical efficiency, lipid metabolism, blood biochemistry, hormonal state, immunological status, blood plasma oxidation capacity and kininogenase system status (fig. 3). The data were received during the clinical studies from 2006 to 2013 and describes 464 children at the ages from 10 to 15 years suffering from obesity.

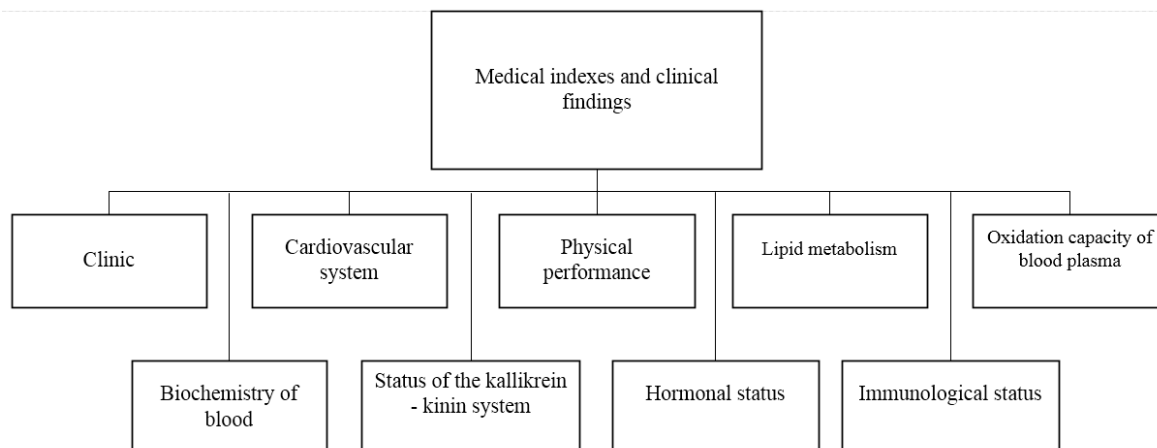


Figure 3. Sets of medical indexes

The problem can be solved based on the given data that contain the following basic indexes:

- WAIST – waist measurement;
- THIGHS – measurement round the hips;
- WEIGHT – body weight;
- EX\_MASS – excess weight in %;
- TMT - lean body mass;
- IMT – body-weight index;
- SAD – systolic blood pressure;
- DAD – diastolic blood pressure;
- HOMA – insulin resistance homeostasis model assessment;
- TFN – exercise tolerance;
- WORK – performance efficiency;
- DV.PR –myocardium double oxygen saturation;
- OL – LDL-cholesterol lipids;
- TAG – triacylglyceride content in blood serum;
- OXC – total cholesterol content in blood serum;
- aXC – alpha cholesterol content in blood serum;

- LPONP – very low-density lipoproteins;
- LPNP - low density lipoproteins;
- IA – atherogenic index;
- glukoza – glycaemic level;
- T3 (1,0-2,8) – triiodothyronine content in blood serum;
- T4 (53-158) – tetra-iodothyronine content;
- TTG (0,23-3,4) – thyroid-stimulating hormone content in blood serum;
- Insylin – insulin content in blood serum;
- Kortizol – cortisone content in blood serum;
- FNO (ne >2,5 pg/ml) – tumour necrosis factor content in blood serum;
- IgA – immunoglobulin A concentration in blood serum;
- IgG – immunoglobulin G concentration;
- IgM – immunoglobulin M concentration;
- CIK – circulating immune complexes in blood serum;
- Lizocim – lysozyme activity in blood serum;
- T-lim, Txel, T-syp, B-lim, – circulating immune complexes in blood serum;
- KK – kininogenase level in blood serum;
- PI – prothrombin consumption in blood serum;
- MG – macroglobulin in blood serum;
- APF – angiotensin converting enzyme in blood serum;
- NO - the content of nitric oxide in the blood.

Mentioned indexes were studied twice: before and after the treatment, as well as their differences and changes were described. The index informative value was assessed with NPP program. The informativity was calculated for all physiological groups in each treatment group. There are 5 treatment groups. For example, the informative value of group 1 components is displayed in table 2.

Table 2. Group 1 indexes informative value

<b>Physiologicalgroup</b>	<b>Index</b>	<b>Informativevalue</b>
Clinic	TMT	7,91
	IMT	0,6
	Mass	0,23
	EX_MASS	0,16
Cardiovascularsystem	DAD	1,89
	CAD	0,8
Physicalperformance	TFN	9,35
	HOMA	0,43

	DV.PR	0,2
Lipidmetabolism	LPONP	1,7
	TAG	1,29
	LPNP	0,61
	Cholesterol	0,47
	IA	0,42
	aXC	0,4
	Commonlipids	0,11
Biochemistryofblood	Ceruloplasmin	0,86
	MDA	0,83
	Catalase	0,63
Immunologicalstatus	IgM	0,84
	Lysozyme	0,51
	IgG	0,49
	CIK	0,47
	IgA	0,4
	T-syp	0,4
	B-lim	0,28
	Txel	0,14
	T-lim	0,04
Hormonalstatus	IL-6	6,32
	Leptin	2,58
	Insulin	0,69
	TTG (0,23-3,4)	0,4
	Cortisol	0,27
	T3 (1,0-2,8)	0,24
	FNO (ne >2,5 pg/ml)	0,1
	AT TPO (upto 30)	0,03
	T4 (53-158)	0
Status of the kallikrein - kinin system	PI	1,82
	APF	1,05
	KK	1,03
	MG	0,44
Oxidationcapacityofbloodpla sma	NO	9,56
	okLPNP	1,72

Treatment complexes influence on physiological systems of the organism differently, which resulted in different informative values in each group of treatment. The graphs of the most and the least informative index were built for every set. An A1 sample includes the values-before-treatment; an A2 sample consists of the values after the treatment.

Let us consider informative index diagram examples. There are index NO of the treatment group 1 from “Oxidation capacity of blood plasma” and index APF of the treatment group 4

from “Status of the kallikrein - kinin system”. There are index NO and index APF on figures 4 and 5.

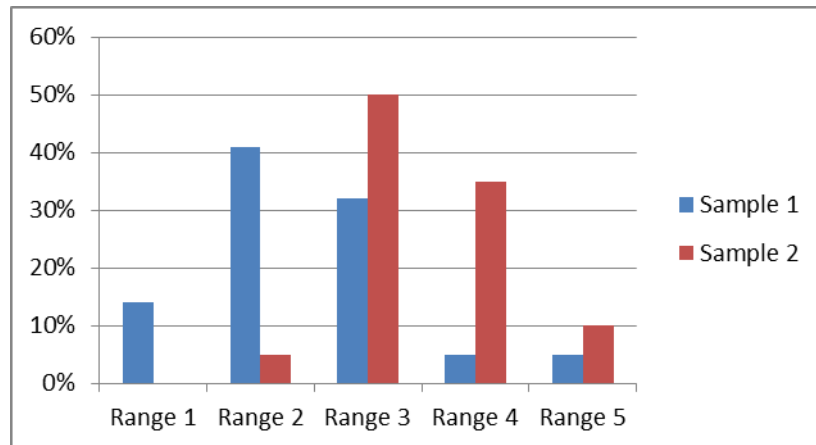


Figure 4. “NO” index

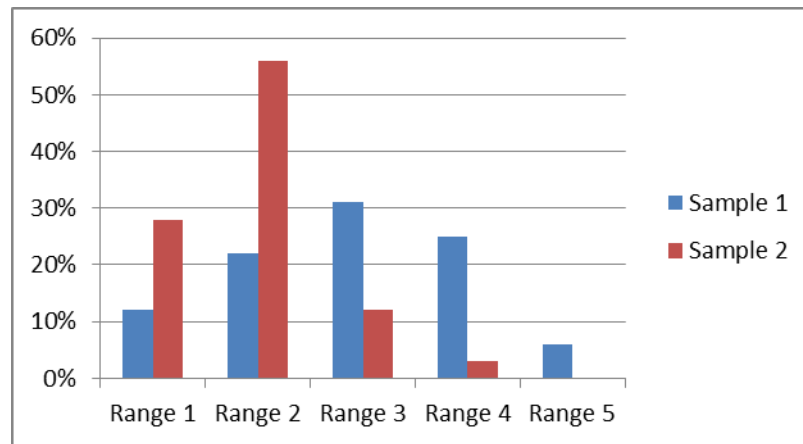


Figure 5. “APF” index

The uninformative index diagram examples are made on figures 6 and 7. There are index Mass of the treatment group 1 and index PI of the treatment group 2.

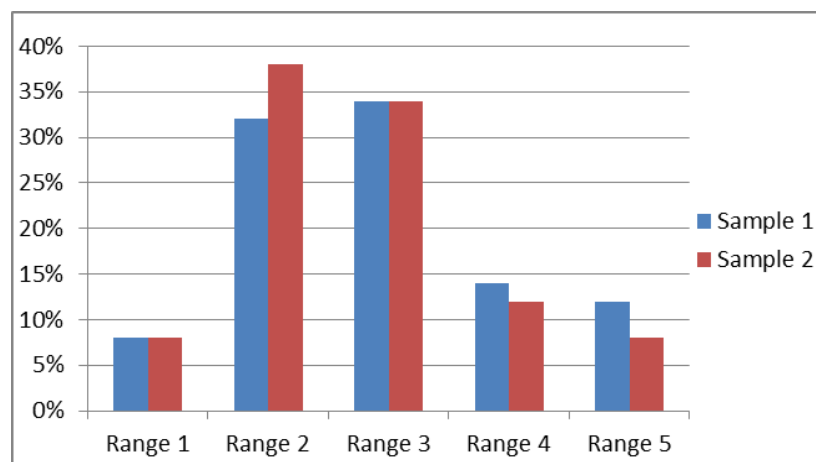


Figure 6. “Mass” index



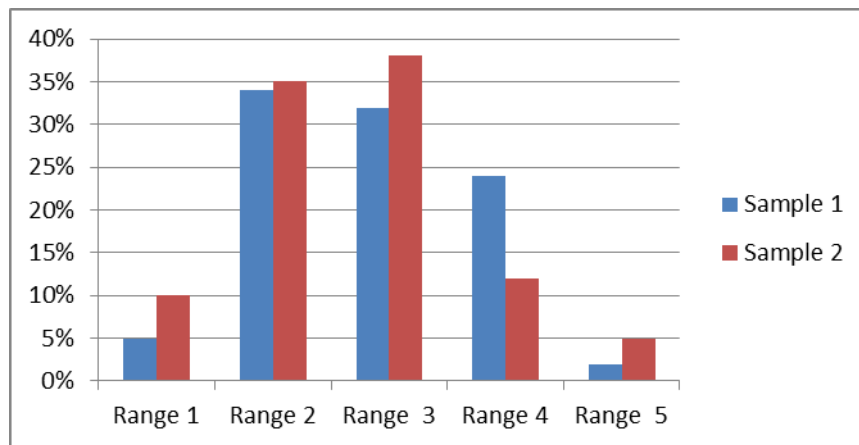


Figure7. "PI"index

The bigger difference in displacement the more informative index is. Figure 4 shows that A1 sample distribution is displaced to the left side, A2 sample distribution is displaced to the right and the difference is evident. Figures 6,7 show that the difference between the samples is no is nonsignificant, so the indexes is less informative.

### Conclusion

Thus, the informative indexes were selected based on the given sample. The information value was calculated for 5 treatment groups. As a result dimension of multidimensional data was decreased from 55 to 27. The data include useful information for treatment.

The research data were used to develop the Tomsk Research Institute of Health Resort Study and Physical Therapy knowledge database and an intelligent clinical decision support system.

## References

- [1] Kartelishvili A, Romyantsev A and Smirnova N 2010 *Burning obesity issues of children and adolescents* (Moscow: Medpactice)
- [2] Raitakari O T, Juonala M, Kähönen M, Taittonen L, Laitinen T, Mäki-Torkko N, Jarvisalo M J, Uhari M, Jokinen E, Ronnemaa T, Akerblom H K and Viikari J S 2003 Cardiovascular risk factors in childhood and carotid artery intima-media thickness in adulthood: the Cardiovascular Risk in Young Finns Study *JAMA* 290(17) 2277-83
- [3] Hebebrand J, Holm J C, Woodward E, Baker J L, Blaak E, Durrer-Schutz D, Farpour-Lambert N J, Frühbeck G, Halford J G C, Lissner L, Micic D, Mullerova D, Roman G, Schindler K, Toplak H, Visscher T L S and Yumuk V 2017 A Proposal of the European Association for the Study of Obesity to Improve the ICD-11 Diagnostic Criteria for Obesity Based on the Three Dimension Etiology, Degree of Adiposity and Health Risk *Obesity Facts* 10(4) 284-307
- [4] Gómez-Ambrosi J, Silva C, Galofré J C, Escalada J, Santos S, Millán D, Vila N, Ibañez P, Gil M J, Valentí V, Rotellar F, Ramírez B, Salvador J and Frühbeck G 2012 Body mass index classification misses subjects with increased cardiometabolic risk factors related to elevated adiposity *International Journal of Obesity* 36(2) 286-94
- [5] Davidson K K, Birch L L 2000 Predictors of change in girls' Body Mass Index between ages 5 to 7 *Obesity Research* 8 (Suppl1) 18
- [6] Dietz W H 1998 Health consequences of obesity in youth: Childhood predictors of adult diseases *Pediatrics* 101 518-525
- [7] Avogaro A, Giaccari A, Fioretto P, Genovese S, Purrello F, Giorgino F & Del Prato S 2017 Consensus Statement for clinical use of the renal sodium-glucose co-transporter-2 inhibitor Dapagliflozin in patients with type 2 Diabetes Mellitus *Expert Review of Clinical Pharmacology*
- [8] Triosi R L, Weiss S T, Paker D R, Sparrow D, Young J B and Landsberg L 1991 Relation of obesity and diet to sympathetic nervous system activity *Hypertension* 14 669-677
- [9] Snyder E E, Walts B, Perusse L, Chagnon Y C, Weisnagel S J, Rankinen T and Bouchard C 2004 The human obesity gene map: 2003 update *Obesity Research* 12 369-439
- [10] Kartelishvili A V 2006 Issues of preliminary diagnosis of children underlying risks to constitutive exogenous obesity *Pediatrics* 4 7-11.
- [11] Website of the Government of the Russian Federation <http://government.ru/orders/3908/> [Accessed 27 February 2018]
- [12] Ravodin R A, Rezvantsev R V 2013 Intelligent decision support systems as a medical malpractice preventive measure *Medical and biological and psychosocial security issues in cases of emergency* 1 27–31.
- [13] Zaripova G R, Bogdanova Yu A, Kataev V A and Khanov V O 2016 Advanced models of expert decision support systems to predict the operative risk during the surgery *Tavrisheskiy Medical and Biological Bulletin* 19(4) 140–145
- [14] Berestneva O G, Volovodenko V A, Gerget O M, Sharopin K A and Osadchaya I A 2013 Multidimensional Medical Data Visualization Methods Based on Generalized Graphic Images *World Applied Sciences Journal* 24 18-23
- [15] Karpov O E, Klimenko G S and Lebedev G S 2016 Application of intelligent systems in healthcare *Modern knowledge intensive technologies* 7-1 38-43
- [16] Berner E S 2007 *Clinical Decision Support Systems* (N. Y.: Springer)
- [17] Garg A X, Adhikari N K, McDonald H, Rosas-Arellano M P, Devereaux P J, Beyene J, Sam J and Haynes R B 2005 Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review *Journal of the American Medical Association* 293(10) 1223-1238
- [18] Kawamoto K, Houlihan C A, Balas E A and Lobach D F 2005 Improving Clinical Practice Using

- Clinical Decision Support Systems: A Systematic Review of Trials to Identify Features Critical to Success *British Medical Journal* 330(7494) 765-768
- [19] Atkov O Yu, Kudryashov Yu Yu, Prokhorov A A and Kasimov O V 2013 Clinical decision support systems *Clinical decision support systems* 6 67-75
- [20] Zharkova O S, Berestneva O G, Moiseenko A V and Marukhina O V 2013 Psychological computer testing based on multitest portal *World Applied Sciences Journal* 24(24) 220-224
- [21] Andersson B 2011 Prediction of severe acute pancreatitis at admission to hospital using artificial neural networks *Pancreatology* 11(3) 328 -335
- [22] Mofidi R 2007 Identification of severe acute pancreatitis using an artificial neural network *Surgery* 141(1) 59-66
- [23] Polat K, Akdemir B and Gbne S 2008 Computer aided diagnosis of ECG data on the least square support vector machine *Digit Signal Process* 18(1) 25-32
- [24] Behnam Malmir A medical decision support system for disease diagnosis under uncertainty *Expert Systems with Applications* Volume 88 95-108
- [25] Gubler E V 1978 *Computational techniques of analysis and recognition of pathological conditions* (Leningrad: Nauka)
- [26] Marukhina O V, Rogozhkina I D, Bobrova M V, Kozlova N V and Berestneva E V 2016 Informative predict system of educational activity of the first-year students *Advances in Computer Science Research* 51 116-119
- [27] Mastitsky S E, Shitikov V A Statistic analysis and visualization of data by R <http://r-analytics.blogspot.ru/> [Accessed 27 February 2018]
- [28] Razumnikov S V, Kremnyova M S 2015 Decision support system of transition IT-applications in the cloud enviroment Proceeding of International Siberian Conference on Control and Communications (Omsk, May 21-23, SIBCON) (Electronic Materials <http://dx.doi.org/10.1109/SIBCON.2015.7147186>) [Accessed 27 February 2018]