

Инженерная школа информационных технологий и робототехники  
 Прикладная математика и информатика 01.03.02  
 Отделение информационных технологий

### БАКАЛАВРСКАЯ РАБОТА

Тема работы
<b>Методы оценки и прогноз функционального состояния развивающихся систем</b>

УДК 518.876

Студент

Группа	ФИО	Подпись	Дата
8Б61	Борисова Алена Александровна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор	Гергет Ольга Михайловна	д.т.н.		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Подопригора Игнат Валерьевич	к.э.н		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Матвиенко Владимир Владиславович	-		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Шевелев Геннадий Ефимович	к.ф.-м.н		

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
 Прикладная математика и информатика 01.03.02  
 Отделение информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП

\_\_\_\_\_  
 (Подпись)    (Дата)    (Ф.И.О.)

**ЗАДАНИЕ**  
**на выполнение выпускной квалификационной работы**

В форме:

бакалаврской работы
---------------------

Студенту:

Группа	ФИО
8Б61	Борисовой Алене Александровне

Тема работы:

<b>Методы оценки и прогноз функционального состояния развивающихся систем</b>	
Утверждена приказом директора (дата, номер)	

Срок сдачи студентом выполненной работы:

--	--

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<p><b>Исходные данные к работе</b></p> <p><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>	<ul style="list-style-type: none"> <li>• Факторы риска развития СД;</li> <li>• Характеристики тяжести и длительности СД;</li> <li>• Сопутствующие осложнения СД;</li> <li>• Диабетические осложнения.</li> </ul>
---	--

<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b></p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ul style="list-style-type: none"> <li>• Аналитический обзор литературы по обработке медицинской информации;</li> <li>• Выбор языка и среды программирования для проведения исследования;</li> <li>• Предварительная обработка данных;</li> <li>• Реализация методов оценки и прогноза функционального состояния пациентов;</li> <li>• Процедура анализа результатов;</li> <li>• Заключение по работе.</li> </ul>
<p><b>Перечень графического материала</b></p> <p><i>(с точным указанием обязательных чертежей)</i></p>	

<p><b>Консультанты по разделам выпускной квалификационной работы</b></p> <p><i>(с указанием разделов)</i></p>	
<p><b>Раздел</b></p>	<p><b>Консультант</b></p>
<p>1. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение</p>	<p>Подопригора Игнат Валерьевич</p>
<p>2. Социальная ответственность</p>	<p>Матвиенко Владимир Владиславович</p>

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

**Задание выдал руководитель:**

<p>Должность</p>	<p>ФИО</p>	<p>Ученая степень, звание</p>	<p>Подпись</p>	<p>Дата</p>
<p>Профессор</p>	<p>Гергет Ольга Михайловна</p>	<p>д.т.н</p>		

**Задание принял к исполнению студент:**

<p>Группа</p>	<p>ФИО</p>	<p>Подпись</p>	<p>Дата</p>
<p>8Б61</p>	<p>Борисова Алена Александровна</p>		

## Планируемые результаты обучения по направлению 01.03.02

### «Прикладная математика и информатика»

Код Результатов	Результат обучения (выпускник должен быть готов)	Требования ФГОС, Критерии АИОР
<i>Профессиональные компетенции</i>		
Р1	Применять <i>глубокие математические и профессиональные знания</i> для решения задач научно-исследовательской, проектной, производственной и технологической деятельности в области системного и прикладного программирования.	Требования ФГОС (ОК-11, 12, ПК-3, 10), Критерий 5 АИОР (п. 5.2.1), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий Требования работодателей: ФГУП «РФЯЦ-ВНИИТФ им. академика Е.И. Забабахина», ООО НАЦ «Недра», ИХН СО РАН
Р2	Умение использовать знания по естественнонаучным дисциплинам при определении задач математического моделирования объектов и явлений в различных предметных областях	Требования ФГОС (ПК-3,9) Критерий 5 АИОР (п.5.2.3), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> Требования работодателей: ФГУП «РФЯЦ-ВНИИТФ им. академика Е.И. Забабахина», ООО «НАПО им. В.П. Чкалова», ИХН СО РАН
Р3	Демонстрировать понимание сущности и значения информации в развитии современного общества, владение основными методами, способами и средствами получения, хранения, переработки информации; использование для решения коммуникативных задач современных технических средств и информационных технологий.	Требования ФГОС (ОК-5, 11, 12,14,15, ПК-2, 6), Критерий 5 АИОР (п. 5.2.2), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий Требования работодателей: Контек, ОАО «Газпром переработка», ООО Нижневартовскэнергонефть».
Р4	Выполнять <i>инновационные проекты с применением глубоких профессиональных знаний и эффективных методов проектирования</i> для достижения <i>новых результатов</i> , обеспечивающих <i>конкурентные</i>	Требования ФГОС (ОК-14, ПК- 7, 9,14), Критерий 5 АИОР (п. 5.2.4), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий.

	<i>преимущества</i> в условиях экономических, экологических, социальных и других ограничений.	Требования работодателей: Контек, ОАО «Газпром переработка», ИХН СО РАН.
P5	<i>Демонстрировать знание о формах организации образовательной и научной деятельности в высших учебных заведениях, иметь навыки преподавательской работы.</i>	Требования ФГОС (ОК-1, 10, 16, ПК-1, 14, 15), Критерий 5 АИОР (п. 5.2.1), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i>
P6	Способность осуществлять организационно-управленческую и социально-ориентированную деятельность с соблюдением профессиональной этики	Требования ФГОС (ОК-5,13,16, ПК-11-13,16) Критерий 5 АИОР (п.5.2.12-13) согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i>
<i>Универсальные компетенции</i>		
P7	<i>Активно владеть иностранным языком</i> на уровне, позволяющем работать в интернациональной среде, включая разработку документации и представление результатов инновационной деятельности. Толерантность в восприятии социальных и культурных различий.	Требования ФГОС (ОК-2, 3,4, 7, ПК-8). Критерий 5 АИОР (п. 5.2.11), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> и Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий
P8	Эффективно работать индивидуально, в качестве <i>члена и руководителя группы</i> , состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность <i>следовать корпоративной культуре</i> организации	Требования ФГОС (ОК-1,4, 6, ПК-8,11,12), Критерий 5 АИОР (пп. 5.2.9,5.2.13), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> и Требования профессиональных стандартов Ассоциации предприятий компьютерных и информационных технологий. Требования работодателей: Контек, ОАО «Газпром переработка», ООО Нижневартовскэнергонефть».
P9	<i>Самостоятельно учиться и непрерывно повышать квалификацию</i> в течение всего периода профессиональной деятельности. Способность к интеллектуальному, культурному, нравственному и профессиональному саморазвитию.	Требования ФГОС (ОК-8,9,16, ПК-5, 11), Критерий 5 АИОР (5.2.14), согласованный с требованиями международных стандартов <i>EUR-ACE</i> и <i>FEANI</i> . Требования работодателей: Контек, ОАО «Газпром переработка», ООО Нижневартовскэнергонефть».

## РЕФЕРАТ

Выпускная квалификационная работа 98 с., 7 рис., 30 табл., 43 источника, 4 прил.

Ключевые слова: машинное обучение, Random Forest, интегральная оценка состояния биосистем, дисперсионный анализ, сахарный диабет

Объектом исследования являются информационные процессы, протекающие в организме человека, у которого диагностирован сахарный диабет

Цель работы – оценка и прогноз функционального состояния больных сахарным диабетом

В работе проводились следующие исследования: обзор и анализ литературы, выявление закономерностей в исследуемых экспериментальных данных, выбор и реализация методов дисперсионного анализа для оценки влияющих показателей на риск развития когнитивных дисфункций и диабетических осложнений, реализован метод машинного обучения случайный лес для классификации и прогнозирования типа сахарного диабета, развития диабетических осложнений и когнитивных дисфункций.

В результате исследования получены оценки функционального состояния пациентов с сахарным диабетом, выявлены закономерности в исследуемых данных, реализован и обучен классификатор, позволяющий с высоким процентов точности классифицировать типа СД, возможность развития когнитивных дисфункций и диабетических осложнений.

Степень внедрения: Центр МРТ г. Томска

Область применения: здравоохранение

Экономическая эффективность/значимость работы: применение реализованных методов и алгоритмов анализа данных в актуальных задачах центра МРТ направлено на решение задач качественной и своевременной диагностики осложнений и развития когнитивных дисфункций у больных сахарным диабетом

В будущем планируется провести исследование влияния диабетических осложнений и когнитивных дисфункций на изменения объемов секторов головного мозга; провести сравнительный анализ методов, применяющихся для исследования головного мозга больных СД, оценить их значимость и информативность; также планируется создать графический интерфейс для удобства пользователя в применении исследованных методов и алгоритмов.

## Нормативные ссылки

1. ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация.
2. СанПиН 2.2.4.548–96 Гигиенические требования к микроклимату производственных помещений
3. СП 52.13330.2016 Нормы естественного и искусственного освещения предприятий, (Актуализированная редакция СНиП 23-05-95).
4. ГОСТ 12.1.003-2014 ССБТ Шум. Общие требования безопасности.
5. СанПиН 2.2.2/2.4.1340–03 Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».
6. ГОСТ 12.1.030-81 ССБТ Электробезопасность, Защитное заземление, зануление.
7. СН 2.2.4/2.1.8.562–96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории застройки.
8. ГОСТ 12.1.019-2017 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты.
9. Федеральный классификационный каталог отходов [Электронный ресурс]. – 2013. – Режим доступа: <http://www.ecoguild.ru/faq/fedwastecatalog.htm>, свободный
10. Федеральный закон от 22.07.2013 г. №123 – ФЗ, Технический регламент о требованиях пожарной безопасности
11. ГОСТ Р 12.2.143-2002 ССБТ. Системы фотолюминесцентные эвакуационные. Элементы систем. Классификация. Общие технические требования. Методы контроля.

Глоссарий .....	10
Введение .....	11
1 Обзор литературы .....	13
2 Объект и методы исследования.....	14
2.1 Описание объекта исследования .....	14
2.2 Методы исследования .....	16
2.2.1 Интегральный критерий оценки состояния .....	16
2.2.2 Дисперсионный анализ .....	18
2.2.3 Решающие деревья и случайный лес .....	24
3 Расчеты и аналитика .....	30
3.1 Пропущенные значения в исходных данных.....	30
3.2 Проверка на нормальность распределения.....	31
3.3 Проверка наличия корреляции между исходными переменными .....	32
4 Результаты проведенного исследования.....	33
5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.....	43
5.1 Организация и планирование работы .....	43
5.2 Расчет сметы затрат на выполнение проекта.....	54
5.3 Оценка научно-технического эффекта.....	60
Выводы по разделу .....	63
6. Социальная ответственность.....	67
Аннотация.....	67
Введение .....	67
6.1. Правовые и организационные вопросы обеспечения безопасности .....	68
6.1.1. Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства .....	68
6.1.2. Организационные мероприятия при компоновке рабочей зоны .....	70
6.2. Профессиональная социальная безопасность.....	72
6.2.1. Анализ вредных и опасных факторов, которые может создать объект исследования .....	72
6.2.2. Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте при проведении исследований .....	73

6.2.3. Обоснование мероприятий по защите исследователя от действия опасных и вредных факторов (техника безопасности и производственная санитария).....	74
6.3. Экологическая безопасность .....	80
6.4. Безопасность в чрезвычайных ситуациях .....	81
6.4.1. Анализ вероятных ЧС, которые может инициировать объект исследований.....	81
6.4.2. Анализ вероятных ЧС, которые могут возникнуть на рабочем месте при проведении исследований .....	82
6.4.3. Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС .....	83
Выводы и рекомендации по разделу.....	84
Заключение .....	86
Список источников .....	87
Приложение А .....	91
Приложение Б.....	92
Приложение В .....	93
Приложение Г. Листинг .....	95

## ГЛОССАРИЙ

СД – сахарный диабет

КЖ – качество жизни

АГ – артериальная гипертензия

ИМТ – индекс массы тела

КН – когнитивные нарушения

ГВ – грудное вскармливание

НbA1c – гликированный гемоглобин

ДА – дисперсионный анализ

RF – Random Forest, или случайный лес

## **ВВЕДЕНИЕ**

Сахарный диабет сегодня является мировой проблемой первого порядка, стоящей перед медицинской наукой и здравоохранением. По разным данным, 7% мирового населения страдает сахарным диабетом. По данным ВОЗ, заболеваемость приобретает характер пандемии и достигает таких масштабов, что был принят ряд нормативных актов, направленных на борьбу с заболеванием, характеризующимся тяжелыми исходами, ранней инвалидизацией и смертностью больных [1].

Термин «сахарный диабет» (СД) означает группу метаболических заболеваний, характеризующихся хронической гипергликемией, которая является результатом нарушения секреции инсулина, действия инсулина или обоих этих факторов. Хроническая гипергликемия при СД сопровождается повреждением, дисфункцией и недостаточностью различных органов, в особенности глаз, почек, нервов, сердца и кровеносных сосудов [2].

Существует ряд классификаций сахарного диабета. При недостаточности инсулина (сахарный диабет 1-го типа) или нарушении механизма взаимодействия инсулина с клетками организма (сахарный диабет 2-го типа) глюкоза накапливается в крови в больших количествах (гипергликемия), а клетки организма (за исключением инсулиннезависимых органов) лишаются основного источника энергии [3].

Этиология сахарного диабета различна для каждого типа. Так, СД 1-го типа является следствием мутаций, которые определяют предрасположенность к заболеванию [4]. СД 2-го типа характеризуется наличием не столько причин развития, сколько факторов риска развития. СД 2-го типа имеет определенную наследственную предрасположенность [5], факторами риска его развития являются артериальная гипертензия, избыточная масса тела, стрессы, курение, избыточное питание, малоподвижный образ жизни [4].

Сахарный диабет оказывает негативное влияние на все аспекты качества жизни (КЖ) пациентов [5-7]. Прогрессирование диабетических осложнений

связывают с ухудшением КЖ в таких аспектах, как трудовая жизнь, финансы, уверенность способности делать что-то и даже возможности передвижения на различные расстояния. Своевременная оценка функционального состояния пациента с наличием СД позволит врачу диагностировать возможность возникновения осложнений и назначить необходимые рекомендации по предупреждению заболеваний и лечению, которое снизит тяжесть протекания осложнения.

Таким образом, целью данной работы является оценка функционального состояния пациентов с СД и прогноз наличия осложнений.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- исследовать методы оценки и прогнозирования функционального состояния человека;
- получить интегральную оценку состояния пациентов;
- провести дисперсионный анализ с целью поиска зависимостей экспериментальных данных, в частности, факторов риска развития СД, характеристик тяжести и длительности протекания заболевания, социальных факторов;
- провести дисперсионный анализ наличия КН;
- провести анализ возможных осложнений СД;
- реализовать метод машинного обучения для классификации и прогноза типа СД и наличия осложнений.

Актуальность данной работы обуславливается необходимостью повышения точности оценки и прогнозирования функционального состояния больного СД и сокращением времени между началом развития у пациента возможных осложнений и когнитивных нарушений и принятием врачебного решения о назначении превентивных мероприятий по предупреждению развития КН и осложнений СД.

Практическая значимость исследования состоит в подборе и анализе статистических методов дисперсионного анализа и машинного обучения, которые позволили бы врачу проводить качественную и своевременную диагностику осложнений и развития когнитивных дисфункций у больных СД.

## 1 ОБЗОР ЛИТЕРАТУРЫ

Все биологические процессы, происходящие в человеческом организме, выражаются некоторыми сигналами. В медицине сигналами могут быть сведения о состоянии человека, его весе, росте, составе крови и других биологических жидкостей [8]. С ростом базы таких сигналов о состоянии человеческого организма возникает необходимость переработки данных, вычленение закономерностей и связей, применение методов статистической обработки для возможного применения результатов исследований в диагностике заболевания и предупреждения их развития.

Возможность использования множества математических методов и вычислительной техники для описания функционального состояния биосистем не всегда гарантирует эффективность применения с точки зрения прикладных задач исследования: попытки точного описания состояния приводят к сложным для анализа моделям, а данные, полученные в отдельных исследованиях, могут привести к разбросу в величине и надежности оценки состояния биосистемы из-за наличия внутри- и межгрупповой вариабельности биообъектов [9]. В работе [10] интегральный критерий рассматривается как решение, при котором необходимо получить оценку общего состояния системы, учитывая весь комплекс показателей, в совокупности отображающих свойства состояния, и сократить размерность задачи путем выбора признаков, характеризующих состояние биосистемы, путем функционального отображения их значений в одну шкалу.

Основной целью медицины будущего является не столько эффективное лечение заболевания, сколько его предотвращение. В этом случае, одной из первостепенных задач обработки медицинских данных является выявление

закономерностей и исследование наличия существенного влияния какого-либо фактора или их совокупности на изменение резульативного признака.

Методы машинного обучения применяются во многих областях, в частности, перспективным приложением является медицинская диагностика [11 - 14].

Для решения задачи классификации используются как классические статистические методы [12], так и методы, ориентированные под машинное обучение (метод опорных векторов, нейронные сети, композиционные методы (бэггинг, бустинг) и другие [13]. В [14] рассматривается применение методов ансамблевого обучения (бэггинг, адаптивный бустинг) для решения задачи прогнозирования стадий сахарного диабета.

В частности, применение методов дисперсионного анализа и машинного обучения решают задачу потребности в обработке и анализе большого количества накопленных данных и получения новых знаний [8, 14, 15, 26].

## **2 ОБЪЕКТ И МЕТОДЫ ИССЛЕДОВАНИЯ**

### **2.1 Описание объекта исследования**

Объектом исследования в данной работе являются информационные процессы, протекающие в организме человека, у которого диагностирован сахарный диабет. Предметом исследования являются методы оценки и прогнозирования функционального состояния.

Функциональное состояние – характеристика уровня функционирования систем организма в определенный период времени, отражающая особенности гомеостаза и процесса адаптации [16].

Организм человека, больного СД, подвержен существенным изменениям, которые могут быть вызваны осложнениями, приобретенными пациентом в случае неправильного «ведения» образа жизни с наличием СД. Функциональное состояние больного СД значительно отличается от состояния здорового человека, меняется во времени и зависит от многих факторов риска развития СД.

В данном исследовании функциональное состояние пациентов больных СД, описывается следующим набором медицинских показателей:

1. Факторы риска развития: возраст, стаж, уровень глюкозы, инсулин;
2. Характеристики тяжести и длительности: холестерин, ИМТ, HbA1c, наличие КН;
3. Социальные факторы: образование, курение, алкоголь, кофе, ГВ;
4. Сопутствующие осложнения: АГ
5. Осложнения: диабетическая нейропатия, диабетическая ретинопатия, диабетическая ретинопатия.

Границы норм исследуемых показателей представлены в Приложении А.

В настоящее время, оценка функционального состояния больного СД не представляется возможной универсальным образом. В данном исследовании, для получения наиболее приближенных к возможным результатам стала необходимой классификация пациентов на группы. Проведено деление на группы с учетом типа СД и КН следующим образом:

1. Группа контроля. К данной группе относятся здоровые пациенты, для которых показатели 1-3 групп факторов получены в границах нормы;
2. Группа пациентов с СД 1 типа без когнитивных нарушений;
3. Группа пациентов с СД 1 типа с когнитивными нарушениями;
4. Группа пациентов с СД 2 типа без когнитивных нарушений;
5. Группа пациентов с СД 2 типа с когнитивными нарушениями.

Для каждой группы пациентов был проведен ряд исследований, позволяющих совместно с врачами сформировать рекомендации и выявить скрытые закономерности в исследуемых показателях. Значения медицинских показателей пациента представлены в Приложении Б.

## 2.2 Методы исследования

### 2.2.1 Интегральный критерий оценки состояния

При решении задачи оценки состояния, поскольку в области медицины существует необходимость работы с разнотипной информацией, возникает необходимость получения количественной интегральной оценки, которая дала бы возможность оценивать состояние системы в динамике и отражала всю информацию, которая заложена в значениях информативных показателей. Что касается оценки, то неоднократно делались попытки формирования интегральных оценок состояния здоровья человека, так как обобщенные оценки унифицируют и ускоряют процесс первичной диагностики организма (работы Л. Е. Полякова и Д. М. Малинского, С. А. Гаспаряна, Ю. М. Комарова, О.Г. Берестневой, Р.М. Баевского, П.К. Анохина, Я.С. Пеккер, Б.А. Кобринского, А.В. Шафиркина, Агаджян, Мамасахлисов, В.А. Фокина, И.А. Ходашинского и других ученых). Но для всех рассмотренных методов вычисления интегрального показателя ограничения по их применению оказываются идентичными:

- ограничение на сферы применения оценки (учет при оценке дисфункции конкретных систем и органов человека);

- необходимость использования дополнительного дорогостоящего диагностического оборудования;

- не учтена динамика изменения показателей системы.

В связи с этим получена и программно реализована методика оценивания функционального состояния на основе меры близости наблюдаемого и предпочтительного состояний, где мера близости нормирована в метрике Махаланобиса на внутримножественное расстояние области предпочтительного состояния.

Пусть  $S_0$ ,  $S$  – заданные референтное и оцениваемое состояния системы, характеризуемые множествами объектов  $\{\vec{x}_i | i \in N_{S_0}\}$  и  $\{\vec{x}_j | j \in N_S\}$  соответственно. Каждое из состояний можно представить множествами точек, занимающими некоторые области в одном и том же пространстве признаков. Количественно оценка состояния некоторого объекта  $\vec{x}_i \in S$  может быть оценена его близостью

к референтному состоянию  $S_0$ . На ее величину будет влиять конфигурация области референтного состояния, взаимное расположение объектов референтного состояния внутри этой области, а также их расположение относительно оцениваемого объекта  $\vec{x}_i$ .

Определим интегральный критерий оценки близости объекта  $\vec{x}_i$  к состоянию  $S_0$  следующим образом [17]:

$$I_{S_0}(\vec{x}_i) = \frac{d(\vec{x}_i, S_0)}{D_{S_0}} \quad (1),$$

где  $d(\vec{x}_i, S_0)$  – некоторая мера близости объекта  $\vec{x}_i$  к множеству  $S_0$ ;  $D_{S_0}$  – мера компактности области занимаемой в пространстве признаков, объектами относящимися к состоянию  $S$ .

Для разработки метода, позволяющего получить интегральную оценку состояния пациента с СД, сформулирован следующий ряд требований:

- интегральная оценка должна включать весь комплекс показателей, характеризующих определенный уровень в иерархическом описании системы;
- математический аппарат, на основе которого планируется вычисление интегральной оценки, должен учитывать вариабельность и наличие взаимосвязей между анализируемыми показателями;
- метод должен обеспечивать единство оценочного алгоритма для различных состояний биосистемы;
- интегральный критерий должен позволять проводить индивидуализированную оценку состояния объекта исследования.

В качестве меры близости объекта  $\vec{x}_i$  к эталонному состоянию  $S_0$  в данной работе используется усредненное расстояние Махаланобиса от этого объекта до всех объектов состояния системы. Расстояние Махаланобиса широко применяется в биомедицинских задачах, в частности, в задачах кластеризации и классификации: вычисленное расстояние не зависит от масштаба переменных, а потому является мерой эффекта этих переменных, что характерно для такого рода медицинских задач.

В [17] показано, что формула (1) может быть преобразована к следующему виду интегральной оценки близости объекта  $\vec{x}_i$  к эталонному состоянию  $S_0$ :

$$I_{S_0}(\vec{x}_i) = \frac{1}{2mN_{S_0}} d(\vec{x}_i, S_0), \quad (2)$$

где в качестве  $d(\vec{x}_i, S_0)$  используется усреднено расстояние Махаланобиса объекта  $\vec{x}_i$  до референтного состояния  $S_0$ :

$$d(\vec{x}_i, S_0) = \frac{1}{N_{S_0}} \sum_{i=1}^{N_{S_0}} d(\vec{x}_j, \vec{x}_i). \quad (3)$$

Ниже изображено расстояние от объектов до референтного состояния  $S_0$ .

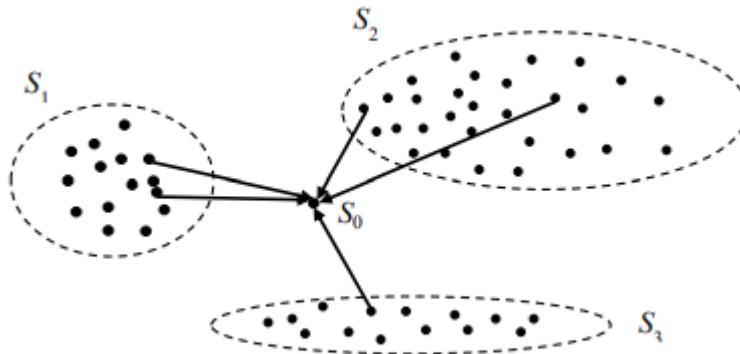


Рис. 1 Расстояние от объектов биосистемы до референтного состояния

В биомедицинских задачах переменные, характеризующие состояние объекта, являются числовыми, например, в данной работе, такими переменными являются значения уровня глюкозы, возраст, стаж болезни. В качестве оценки состояния биосистемы часто выступает ограниченный набор признаков, в идеальном случае, один. Интегральная оценка, вычисленная на основе усредненного расстояния Махаланобиса, позволяет учитывать вариабельность признаков, более чувствительна к малым изменениям состояния биосистем [18], производит уменьшение размерности задачи с сохранением информации, содержащейся в пространстве признаков.

### 2.2.2 Дисперсионный анализ

Дисперсионный анализ [19] – статистический метод, позволяющий проводить исследование наличия или отсутствия существенного влияния какого-либо фактора на изменение исследуемого результативного признака. Важным условием проведения дисперсионного анализа является значение переменных: независимая переменная должна быть категориальной, зависимая – метрической.

Подробнее, суть дисперсионного анализа (ANOVA, analysis of variance) заключается в разложении дисперсии измеряемого признака на независимые слагаемые, каждое из которых характеризует величину влияния того или иного фактора или их взаимодействия. Апостериорное сравнение таких слагаемых (в случае, если  $n > 2$ , где  $n$  – количество слагаемых), позволяет оценить значимость каждого исследуемого фактора или их комбинации.

Дисперсионный анализ основан на расчете F-статистики, которая предполагает исследование отношений двух дисперсий: межгрупповой и внутригрупповой. Полученное отношение сравнивается с критическим значением отношения Фишера и, если фактическое значение больше критического, то средние классов отличаются друг от друга, и исследуемый фактор оказывает существенное влияние на изменение результативного признака. Аналогией F-теста Фишера в случае двух исследуемых выборок является t-тест Стьюдента.

Преимуществом дисперсионного анализа является возможность его применения для оценки и проверки гипотез в моделях факторного анализа, регрессии, классификации объектов.

Классификация моделей дисперсионного анализа основана на количестве, распределении и зависимости групп оцениваемых факторов. Например, если в эксперименте необходимо оценить различие уровня глюкозы для нескольких групп пациентов, такие группы фактора будут независимыми. При исследовании влияния лечения на изменение уровня глюкозы, у исследуемой группы пациентов берут два анализа: в начале и в конце лечебного процесса, и такие группы будут зависимыми. В настоящей работе исследуются только независимые между собой группы факторов.

### **Параметрический однофакторный дисперсионный анализ**

Однофакторный дисперсионный анализ применяется, когда имеется три или более независимые выборки (группы), полученные из одной генеральной совокупности при изменении какого-либо фактора  $\tau$ , имеющего, соответственно, равное количеству выборок уровня  $\tau_i, i = 1, \dots, k$ . Параметрический дисперсионный анализ предполагает, что  $k$ -выборок подчиняются нормальному

закону и имеют разные выборочные средние и равные выборочные дисперсии. В результате, применение параметрического дисперсионного анализа позволяет ответить на вопрос, оказал ли фактор существенное влияние на разброс выборочных средних. Есть определить  $\mu_k$  как выборочное среднее  $k$ -выборки, существенность влияния можно определить проверкой следующих гипотез:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k, \text{ где } k - \text{ число выборок (групп)} \quad (4)$$

$$H_1: \text{ не все } \mu_k \text{ равны}$$

Основная идея однофакторного дисперсионного анализа состоит в том, что сумма квадратов отклонений статистического комплекса объясняется двумя компонентами:

$$SS = SS_k + SS_e, \quad (5)$$

где  $SS_k$ - объясненная влиянием  $k$ -фактора сумма квадратов отклонений,  $SS_e$  - необъяснённая сумма квадратов отклонений или сумма квадратов отклонений ошибки.

Пусть  $n_i$  – число наблюдений в каждой группе, тогда  $\sum_{i=1}^k n_i$  (6) – общее число наблюдений. Принимая во внимание (4) и (5), получаем следующие формулы:

$$\text{Общее число квадратов отклонений: } SS = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2, \quad (7)$$

Объясненная влиянием фактора  $k$  сумма квадратов отклонений:  $SS_k$

$$= \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2, \quad (8)$$

Необъясненная сумма квадратов отклонений или сумма отклонений ошибки:

$$SS_e = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2, \quad (9)$$

где  $\bar{X} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}$  (10) – общее среднее наблюдений;  $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$  (11) – среднее наблюдение в каждой группе. Кроме того, сумму отклонений ошибки можно получить через дисперсию группы  $s_i$ :

$$\begin{aligned}
SS_e &= \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 = \sum_{i=1}^k (n_i - 1)s_i^2 \\
&= (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2 \quad (12)
\end{aligned}$$

Проведение дисперсионного анализа подразумевает анализ не вычисленных выше сумм квадратов отклонений (7-9), а так называемых средних квадратов, которые получается путем деления сумм квадратов отклонений на соответствующе число степеней свободы. Степенью свободы называют число наблюдений минус число связывающих их уравнений. Так, для среднего квадрата  $MS_k$ , являющегося несмещенной оценкой межгрупповой дисперсии, число степеней свободы будет равно  $v_k = k - 1$ , так как при его расчете используются  $k$ -групповых средних, связанных между собой одним уравнением (10). Средний квадрат  $MS_e$ , которые является несмещенной оценкой внутригрупповой дисперсии, имеет число степеней свободы  $v_e = n - k$ , так как при его расчете используются  $n$  (5) наблюдений, связанных между собой  $k$ -уравнениями (11). Таким образом:

$$MS_k = \frac{SS_k}{k - 1}; MS_e = \frac{SS_e}{k - 1}. \quad (13)$$

Для проверки гипотезы (4) о том, что все дифференциальные эффекты равны нулю, необходимо отношение среднего межгруппового квадрата отношения к среднему внутригрупповому, то есть вычислить  $F$ -значение:

$$F_H = \frac{MS_k}{MS_e}, \quad (14)$$

Для того, чтобы сделать окончательный вывод, необходимо найти критическое значение  $F_{кр}$  по таблице Фишера, учитывая вычисленные ранее степени свободы ( $v_k$  и  $v_e$ ) и соответствующий уровень значимости (обычно и в данном исследовании равный 0.05). Если наблюдаемое значение критерия  $F_H$ , вычисленное по данным выборки, принадлежит критической области ( $F_H > F_{кр}$ ), то гипотезу отвергают; если наблюдаемое значение не принадлежит критической области, то нет оснований отвергать гипотезу. Принятие гипотезы о равенстве всех  $k$ -групповых средних генеральному среднему свидетельствует о

незначимом различии групповых и генерального среднего. Результат считается значимым, если гипотезу отвергают, поскольку это говорит о существенном влиянии фактора на разброс выборочных средних.

## **Непараметрический дисперсионный анализ для сравнения независимых групп**

Однако, при работе с реальными данными, их распределение не подчиняется нормальному закону. В этом случае, применяется один из непараметрических аналогов однофакторного дисперсионного анализа – ранговый дисперсионный анализ Краскела–Уоллиса. Критерий Краскела–Уоллиса основан на рангах, а не на исходных наблюдениях, и предполагает, что рассматриваемая переменная непрерывна и измерена, как минимум, в порядковой шкале. Критерий проверяет гипотезу: сравниваемые выборки имеют одно и то же распределение или распределения с одной и той же медианой. Таким образом, интерпретация критерия схожа с интерпретацией параметрической однофакторной ANOVA, за исключением того, что этот критерий основан на рангах, а не на средних значениях.

На примере следующей задачи рассмотрим основы непараметрического критерия Краскела-Уоллиса [20]. Пусть данные включают  $N = \sum_{j=1}^k n_j$  наблюдений, по  $n_j$  наблюдений на  $j$ -ю выборку (группу),  $j = 1, \dots, k$ . Математическая модель непараметрического однофакторного анализа имеет вид:

$$x_{ij} = \mu + \tau_j + \varepsilon_{ij}, i = 1, \dots, n, j = 1, \dots, k, (15),$$

где  $\mu$  – неизвестное общее среднее;  $\tau_j$  – неизвестный эффект  $j$ -ой выборки ( $\sum_{j=1}^k \tau_j = 0$ ). Все случайные ошибки  $\varepsilon_{ij}$  взаимно независимы и извлечены из одной и той же непрерывной совокупности параметра. Математическая модель проверяет гипотезу  $H_0: \tau_1 = \tau_2 = \dots = \tau_k$  в следующем порядке:

1. Ранжируются все  $N$  наблюдений вместе от меньшего к большему. Ранг обозначается как  $r_{ij}$ .
2. Предполагается для  $j = 1, \dots, k$ :

$$R_j = \sum_{i=1}^{n_j} r_{ij}, R_{*j} = \frac{R_j}{n_j}, R_{**} = \frac{N+1}{2}. \quad (16)$$

Например,  $R_1$  – сумма рангов, которые были присвоены для значений 1-ой выборки,  $R_{*1}$  – средний ранг 1-ой выборки.

3. Вычисляется статистика

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k n_j (R_{*j} - R_{**})^2 = \left[ \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} \right] - 3(N+1) \quad (17)$$

4. На уровне значимости  $\alpha$  гипотеза  $H_0$ :

— отклоняется, если  $H_0 \geq h(\alpha, k, (n_1, \dots, n_k))$

— принимается, если  $H_0 < h(\alpha, k, (n_1, \dots, n_k))$

где постоянная  $h(\alpha, k, (n_1, \dots, n_k))$  удовлетворяет уровню  $P\{H \geq h(\alpha, k, (n_1, \dots, n_k))\} = \alpha$ . Если гипотеза  $H_0$  верна, то статистика  $H$  имеет асимптотическое (при  $\min(n_1, \dots, n_k) \rightarrow \infty$ ) распределение  $\chi^2$  с  $k-1$  степенями свободы.

### Апостериорный анализ

Однофакторный дисперсионный анализ Крускала-Уоллиса является расширением критерия Манна-Уитни. Критерий Манна-Уитни считается апостериорный анализом исследуемых переменных, и применяется в случае наличия более, чем двух независимых выборок и принятия альтернативной гипотезы  $H_1$  анализа Крускала-Уоллиса.

Таким образом, непараметрический  $U$  критерий Манна-Уитни применяется для сравнения средних значений двух независимых выборок. Эмпирическое значение непараметрического  $U$  критерия Манна-Уитни определяет насколько два независимых распределения совпадают. Анализ пересечения двух выборок основывается на последовательном ранжировании значений. Условия применения  $U$ -критерия аналогичны  $H$ -критерию Крускала-Уоллиса:

— сравниваемые значения выборок не коррелируют между собой;

— распределение признаков может не соответствовать нормальному.

В случае соответствия выборок нормальному распределению, альтернативой  $U$ -критерия является  $t$ -тест Стьюдента.

Нулевую гипотезу U-критерий Манна-Уитни можно сформулировать следующим образом: распределение признака в первой выборке соответствует распределению признака во второй выборке. Пусть сравниваются две выборки объемом  $n_1$  и  $n_2$  соответственно. Математическая модель должна вычислять U-значение в следующем порядке:

1. Обе выборки объединяют в один ряд, помечая данные, принадлежащие разным выборкам.
2. Значения ранжируются, приписывая наименьший ранг меньшему значению. Количество рангов должно быть равно  $(n_1 + n_2)$
3. Для каждой выборки подсчитывается сумма рангов и определяется та, чья ранговая сумма наибольшая.
4. Значение U определяется по формуле:

$$U = n_1 n_2 + \frac{n_x + 1}{2} - T_x, (18)$$

где,  $n_1$  и  $n_2$  объемы выборок соответственно;  $T_x$  – большая из двух ранговых сумм;  $n_x$  – объем максимальной выборки,  $n_x = \max(n_1, n_2)$ .

5. Принимается уровень значимости  $\alpha$ , по таблице критических значений критерия U Манна-Уитни определяется критическое значение  $U_{кр}(\alpha)$ . Если  $U_{эмп} > U_{кр}(\alpha)$ .  $H_0$  принимается. Если  $U_{эмп} \leq U_{кр}(\alpha)$   $H_0$  отвергается. Чем меньше значения U, тем достоверность различий выше.

### 2.2.3 Решающие деревья и случайный лес

Сегодня задача классификации в машинном обучении – отнесение образца выборки к какому-либо конечному классу – может быть решена различными методами, основными из которых являются логистическая регрессия, метод k-ближайших соседей, методы опорных векторов, дискриминантный анализ, случайный лес, композиционные методы (бэггинг, бустинг) [14]. В [13] описано применение композиции бэггинга и деревьев решений, которое дает более точный результат, чем применение отдельного алгоритма на конкретном наборе данных.

Random Forest, или случайный лес основан на построении ансамбля деревьев решений, каждое из которых обучается на выборке, получаемой из исходной тренировочной выборки, с помощью бэггинг (т.е. выборки с обращением) [15]. Для определения алгоритма работы случайного леса целесообразно рассмотреть следующие важные понятия и определения.

### Дерево решений

Пусть  $G = (V, E)$  – граф, состоящий из конечного непустого множества  $V$ , элементами которого являются вершины, множества пар которых  $E$ , называемые ребрами. Последовательность ребер вида  $(v_1, v_2), (v_2, v_3) \dots (v_{m-1}, v_m)$  называется путем графа. Если пара вершин  $v, w$  образующая ребро  $(v, w)$ , является упорядоченной, то такое ребро является ориентированным; если все ребра графа являются ориентированными, то такой граф является ориентированным. Если дерево содержит одну выделенную вершину, то такое дерево называется корневым, а вершина является корнем дерева. В данной работе рассматриваются только ориентированные корневые деревья, в которых ребра направлены по направлению от корня.

Пусть имеется независимая выборка  $D = \{x_i, y_i\}_{i=1}^l$  из некоторого независимого распределения  $P(x, y) = P(x)P(y|x)$ . Такая выборка является обучающей выборкой.  $x_i, i = 1, 2, \dots, l$ , – векторы признаков (прецеденты), координаты которых представляют значения  $n$  признаков (независимых переменных), измеренных на некотором объекте (образе). Множество всех возможных значений векторов признаков  $\chi$  называется пространством образов. Соответствующие  $y_i$  являются значениями зависимой переменной. В данной работе  $y_i$  имеют конечное число значений, т.е.  $y_i \in \{w_1, w_2, \dots, w_c\}, c \geq 2$ , поэтому решается задача классификации.

Деревом решений является дерево, с каждой вершиной  $t$  которого связаны [21]:

1. Некоторое подмножество  $\chi_t \subset \chi$ ; с корневой вершиной связывается все пространство образов  $\chi$ ;
2. Подвыборка  $D_t \subset D$  обучающей выборки  $D$ , такая, что  $D_t = \{(x, y) \in D : x \in \chi_t\}$ ; таким образом с корневой вершиной связывается вся выборка  $D$ ;
3. Некоторая функция (правило)  $f_t: \chi \rightarrow \{0, 1, \dots, k_t - 1\}$  (здесь  $k_t \geq 2$  – количество потомков вершины  $t$ ), определяющая разбиение множества  $\chi$  на  $k$  непересекающихся подмножеств. С терминальными вершинами не связывается никакая функция.

Если обозначить  $t_{i(t)}, i = 0, 1, \dots, k - 1$  вершину являющуюся  $i$ -м потомком вершины  $t$ , то множество  $\chi_t$  и правило  $f_t$  определяют множество  $\chi_{t_{i(t)}}$  следующим образом:

$$\chi_{t_{i(t)}} = \chi_t \cap \{x \in \chi : f_t(x) = i\}. \quad (19)$$

Цель построения дерева решений в данной работе состоит в классификации векторов  $x$  из распределения  $P(x)$ . Так как метод дерева решений часто интерпретируют с методом принятия интуитивных человеческих решений, процесс принятия решений начинается с корневой вершины и состоит в последовательном применении правил, которые определены в вершинах дерева. Определение терминальной вершины  $t$  такой, что  $x \in \chi_t$  является результатом процесса.

### **Загрязненность**

Понятие загрязненности является противоположным понятию меры однородности. Если  $t$  – некоторая вершина дерева решений,  $D(t)$  – подвыборка, связанная с этой вершиной, то  $i(t)$  является загрязненностью вершины, которая стремится к 0, если  $D(t)$  содержит прецеденты только одного класса и к 1, если  $D(t)$  содержит одинаковое число прецедентов каждого класса. В работе, при обучении и настройке (tuning) классификатора рассматриваются две наиболее популярные меры загрязненности вершин:

— Энтропия:

$$i(t) = - \sum_{i=1}^c P(w_i) \log_2 P(w_i), \quad (20)$$

где  $P(w_i)$  – доля примеров класса  $(w_i)$  в подвыборке  $D(t)$ .

— Индекс Джини, который представляет собой частоту ошибочной классификации при случайном назначении меток класса наблюдениям подвыборки  $D(t)$ :

$$i(t) = 1 - \sum_{i=1}^c P^2(w_i), \quad (21)$$

### Оптимальное расщепление вершин

Расщеплением множества признаков  $\chi$  является правило разбиения данного множества. Бинарное расщепление вершины  $t$  рассматривается как функция  $f_t: \chi \rightarrow \{0,1\}$ ,  $x \in \chi$ , где в случае  $f(x)=0$  вектор  $x$  принадлежит левому потомку, а в случае  $f(x)=1$  – правому. Расщепление подвыборки осуществляется таким образом, чтобы свести параметр загрязненности к минимальному значению. Для бинарных деревьев уменьшение загрязненности вершины  $t$  можно определить как

$$\Delta i(t) = i(t) - P_L i(t_L) - P_R i(t_R), \quad (22)$$

где  $P_L$  и  $P_R$  доли примеров подвыборки  $D(t)$ , соответствующие левому и правому потомку ( $t_L$  и  $t_R$ ). Наилучшим расщеплением вершины  $t$  естественно считать разбиение, которое максимизирует величину  $\Delta i(t)$  [15]. Однако, непосредственное использование формулы (22) приводит к тому, что чаще для расщепления выбираются признаки, имеющие больше уровней, по сравнению с остальными. Поэтому, в алгоритмах деревьев решений для расщепления множества признаков корректно использовать следующую формулу:

$$\Delta i_B(t) = \frac{\Delta i(t)}{-\sum_{k=1}^B P_k \log_2 P_k}, \quad (23)$$

и выбирать расщепление, максимизирующее величину  $\Delta i_B(t)$ .

Процесс расщепления осуществляется до тех пор, пока возможно уменьшить загрязненность очередной вершины. В этом случае дерево считается полным, а каждая терминальная вершина содержит примеры только одного

класса. Полные деревья обладают низкой точностью, что описывается в [22]. Проблема переπροгонки (overfitting) заключается в том, что фактически, полученная статистическая модель описывает лишь рассмотренную ей выборку и непригодна для генеральной совокупности. В данной работе применяется один из популярных способов принятия решения остановки расщепления – применения кросс-проверки.

### **Кросс-валидация**

Для оценки работы алгоритма, необходимо иметь три подвыборки из исходных данных: обучающая, валидационная (проверочная) и тестовая. На так называемом «наборе проверки» выполняется оценка работы классификатора, который обучился на тестовой выборке. После того, как экспериментальная оценка превысит минимальную, эксперимент считается удачным, и окончательная оценка может быть сделана на тестовом наборе. Такой подход значительно улучшает работу алгоритма, однако приводит к резкому сокращению количества выборок, используемых для изучения модели. Понятие кросс-валидации связано с перекрестной проверкой обучаемой модели. В базовом подходе, называемом  $k$ -кратной кросс-валидацией, или CV, обучающая выборка делится на  $k$  наборов. Алгоритм CV может быть описан следующим образом и применяется для каждого из  $k$  наборов [23]:

1. Модель обучается с использованием  $k-1$  набором в качестве тренировочных данных;
2. Полученная модель проверяется на оставшейся части данных (т. е. она используется в качестве тестового набора для вычисления показателя производительности, такого как точность (score)).

### **Случайный лес**

В работе используется алгоритм индуктивного построения случайного леса, описанного Брейманом и Катлер [24]. Понятие случайного леса может быть определено как ансамбль деревьев решений, каждое из которых строится на основе бутстреп выборки из исходной обучающей выборки [25], при этом для

расщепления вершин используется доля случайно отбираемых признаков из  $\chi$ . При построении случайного леса, строятся полные деревья решений, а классификация деревьев в ансамбле осуществляется большинством голосов. Алгоритм индукции случайного леса [26]:

1. Для  $i = 1, 2, \dots, B$  ( $B$  – количество деревьев в ансамбле) выполнить:
  - a. Сформировать бутстреп выборку  $S$  размера  $l$  по исходной обучающей выборке  $D = \{x_i, y_i\}_{i=1}^l$ ;
  - b. По бутстреп выборке  $S$  индуцировать неусеченное дерево решений  $T_i$  с минимальным количеством наблюдений в терминальных вершинах равным  $n_{min}$ , рекурсивно следуя следующему подалгоритму:
    - i. Из исходного набора  $n$  признаков случайно выбрать  $p$  признаков;
    - ii. Из  $p$  признаков выбрать признак, который обеспечивает наилучшее расщепление;
    - iii. Расщепить выборку, соответствующую обрабатываемой вершине на две подвыборки;
2. В результате выполнения шага 1 получаем ансамбль деревьев решений  $\{T_i\}_{i=1}^B$ ;
3. Предсказание новых наблюдений для классификации осуществляется следующим образом: пусть  $\hat{\omega}_i(x) \in \{\omega_1, \omega_2, \dots, \omega_c\}$  – класс, предсказанный деревом решений  $T_i$ , т.е.  $T_i(x) = \hat{\omega}_i(x)$ ; тогда  $\hat{\omega}_i^B(x)$  – класс, наиболее часто встречающийся во множестве  $\{\hat{\omega}_i(x)\}_{i=1}^B$ .

Метод Out-Of-Bag (OOB), предложенный Брейманом [27], обеспечивает получение естественной оценки вероятности ошибочной классификации случайных лесов на основе наблюдений, не входящих в обучающие бутстреп выборки, используемые для построения деревьев (эти наблюдения называются OOB выборками).

### 3 РАСЧЕТЫ И АНАЛИТИКА

Предварительная обработка экспериментальных данных проводится с целью качественного и корректного проведения исследования. Предобработка включает в себя проверку статистических характеристик исследуемого набора данных, а также исключение грубых погрешностей и данных, которые могли быть неверно введены непосредственно в компьютер. В данном исследовании все вычисления проводились на высокоуровневом языке Python. Язык имеет множество библиотек с открытым кодом, одной из сфер его применения является медицина. Построение модели бэггинг-классификатора осуществлялось с применением библиотеки Scikit-learn [28, 29]

#### 3.1 Пропущенные значения в исходных данных

При выполнении выпускной квалификационной работы использованы данные, полученные в МРТ центре г. Томска. Непосредственный ввод данных в компьютер не исключает возможные пропуски или неверно введенные значения по причине ошибки измерений. Также наличие пропусков обуславливается невозможностью сбора данных. Алгоритмы, используемые в исследовании, не предполагают способность адекватных вычислений при работе с пропущенными значениями, ввиду этого существует необходимость удаления объекта или признака, содержащего пропущенные значения, или их доопределение.

В данном исследовании для восстановления пропущенных численных значений использована линейная интерполяция. Линейная интерполяция предполагает приближенное вычисление значений функции  $y = f(x)$  для внутренних точек отрезка измерения  $[a, b]$  по известным значениям  $f(a)$  и  $f(b)$  вычисляемой по следующей формуле:

$$|f(x) - \varphi(x)| \leq \frac{M}{2}(x - a)(b - x), \quad (24)$$

где  $M = \sup_{x \in [a, b]} f''(x)$ .

В исследовании также присутствует большое количество категориальных признаков. Наличие пропуска в категориальном признаке свидетельствует об

отсутствии этого признака у пациента или отсутствие необходимости его определения медицинским методом. Следовательно, такие пропуски были восстановлены нулями.

### 3.2 Проверка на нормальность распределения

Важным аспектом предобработки экспериментальных данных является проверка соответствия их нормальному распределению.

Проверим гипотезу  $H_0$ : «случайная величина  $X$  распределена в соответствии с нормальным законом распределения», применяя один из наиболее эффективных критериев для проверки нормальности – критерий Шапиро-Уилка [30]. Пусть имеется выборка  $(X_1, \dots, X_n)$ . Статистика критерия Шапиро-Уилка может быть вычислена по следующей формуле [31]:

$$W = b^2 / S^2, (25)$$

где  $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$  и  $b^2 = \sum_{i=1}^k a_{i,n} (X_{n-i+1} - X_i)$ . Значение  $k$  определяется следующим образом:  $k = n/2$ , если  $n$  – четное;  $k = n - 1/2$ , иначе. Коэффициенты  $\{a_{i,n}, i = 1, \dots, k; n = 3, \dots, N\}$ , где  $N$  – численность выборки, являются известными константами и определяется по таблице. Есть  $W$ , полученное в (13), ниже критического значения  $W(p)$ , где  $p$  – критический уровень значимости, гипотеза о нормальности распределения отклоняется.

Вычисление достигнутого уровня значимости осуществляется с помощью нормальной аппроксимации:

$$1 - \Phi_{0,1}(|b_n + c_n \ln((W - d_n)/(1 - W))|), (26)$$

где  $\Phi_{0,1}$  – стандартное нормальное распределение,  $b_n, c_n, d_n$  – константы, для которых известны табличные значения.

В результате исследования гипотеза о нормальности распределения исходных данных была отклонена для всех имеющихся выборок. Такой результат позволяет принять решение об использовании в исследовании методов непараметрического дисперсионного анализа, которые не требуют от данных подчинения нормальному закону распределения.

Полученные значения интегральных оценок функционального состояния пациентов также были проанализированы на соответствие нормальному закону распределения. В результате гипотеза о нормальности была отклонена.

### 3.3 Проверка наличия корреляции между исходными переменными

Статистическая взаимосвязь между случайными величинами, распределение которых подчиняется нормальному закону, определяется коэффициентом корреляции Пирсона по формуле:

$$r_{XY} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2(y-\bar{y})^2}}. \quad (27)$$

Однако, часто распределение анализируемых данных не соответствует нормальному. В таком случае, применение статистических методов анализа неприемлемо. Для оценки меры линейной связи между случайными переменными, распределение которых отлично от нормального, применяют коэффициент корреляции Кендалла, которая является ранговой [32].

Коэффициент корреляции Кендалла вычисляется по следующей формуле:

$$\tau = 1 - \frac{4}{n(n-1)}R, \quad (28)$$

где  $R = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [[x_i < x_j] \neq [y_i < y_j]]$  – количество инверсий, образованных величинами  $y_i$  расположенными в порядке возрастания соответствующих  $x_i$ .

Существование статистической взаимосвязи крайне важно при определении предикторов для зависимой переменной для реализации случайного леса. Сильнокоррелированные независимые факторы свидетельствуют о том, что при изменении одного влечет значительное изменение другого. Высокий коэффициент корреляции независимых факторов предполагает исключение одного из них из выборки предикторов.

В Приложении В представлена таблица значений коэффициентов корреляции для исходных значений переменных. Зависимыми переменными в

данном исследовании: тип сахарного диабета; наличие когнитивных нарушений (степень КН); три вида осложнений: ретинопатия, нейропатия, нефропатия.

#### 4 РЕЗУЛЬТАТЫ ПРОВЕДЕННОГО ИССЛЕДОВАНИЯ

В результате проведенного исследования были выполнены все поставленные задачи. Построение моделей машинного обучения классификации и прогнозирования типа СД, наличия осложнений и развития КН осуществлено методом композиции бэггинга и случайного леса. В работе проведена настройка гиперпараметров, которые используются для получения лучшего показателя перекрестной проверки

В таблице 1 представлены медианные значения интегрального показателя, основанного на мере близости Махаланобиса для 5-ти исследуемых групп пациентов. В таблице 2 представлены результаты непараметрического дисперсионного анализа для проверки гипотезы  $H_0$ : «медианы исследуемых выборок равны».

Таблица 1. Медианные значения интегрального критерия

Группа СД	Медианное значение интегрального критерия
СД 1 типа без КН	8,53
СД 1 типа с КН	9,96
СД 2 типа без КН	6,24
СД 2 типа с КН	7,82
Группа контроля	5,9

Таблица 2. U-значения теста Манна-Уитни

Группа пациентов	Группа пациентов	U-значение теста Манна-Уитни	pvalue
СД 1 типа без КН	СД 1 типа с КН	0	0.00
СД 1 типа без КН	СД 2 типа без КН	84	0.00

СД 1 типа без КН	СД 2 типа с КН	60	0.00
СД 1 типа без КН	Группа контроля	48	0.00
СД 1 типа с КН	СД 2 типа без КН	84	0.00
СД 1 типа с КН	СД 2 типа с КН	20	0.040
СД 1 типа с КН	Группа контроля	6	0.006
СД 2 типа без КН	СД 2 типа с КН	58	0.004
СД 2 типа без КН	Группа контроля	126	0.001
СД 2 типа с КН	Группа контроля	0	0.046

Анализ Таблицы 1 показал, что медианные значения интегральной оценки для каждой группы различны. Следовательно, отклонение обобщенного показателя, характеризующего состояние организма в целом, от нормы (референтного состояния), свидетельствует о разных состояниях в разных группах. Значение интегральной оценки референтного состояния (группы контроля) равно 5,9. Сравнительная характеристика значений, отраженных в таблице 1, подтверждена врачами.

Для проверки гипотезы о различии медиан выборок интегральной оценки был проведен непараметрический дисперсионный анализ Манна-Уитни, который позволил попарно сравнить медианы исследуемых выборок. Значения U-критерия и соответствующий уровень значимости p Манна-Уитни, приведенные в Таблице 2, помогают сделать вывод о статистически значимом различии интегральных оценок для каждой группы пациентов.

Для прогнозирования типа СД был проведен дисперсионный анализ факторов, оказывающих влияние на определения типа СД. Критерий Крускала-Уоллиса, рассматривающий гипотезу  $H_0$  о равенстве медиан выборок был принят

для факторов Холестерин (p-value равно 0.52) и Образование (p-value равно 0.71). Такой результат свидетельствует о том, что значения этих факторов статистически незначимы для разного типа СД.

Реализованный в работе метод машинного обучения случайный лес позволил с точностью равной 97% определить тип сахарного диабета. Величина AUC ROC равна 1. Величина отображает соотношение между долей объектов от общего количества носителей признака, верно классифицированных как несущие признак и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущие признак при варьировании порога решающего правила. Значение 1 свидетельствует о высоком качестве классификатора.

Как отмечалось выше, Random Forest также может применяется в решении задач выявления наиболее информативных признаков для обучения классификатора. На Рис. 2 представлены веса факторов (importance), оказывающих наибольшее влияние на работу классификатора при определении типа СД. Сумма весов факторов равна 1, в случае, если все деревья не являются деревьями с одним узлом, состоящим только из корневого узла, иначе, значения весов были бы равны нулю.

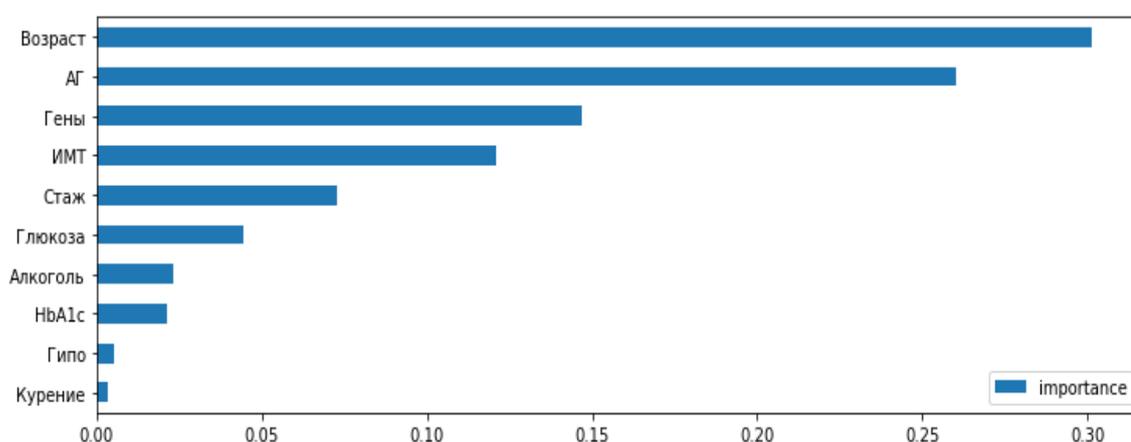


Рис. 2 Значимость факторов при работе классификатора при определении типа СД

Значения весов факторов, представленные на Рис. 2, помогают найти наиболее важные переменные для построения более точных моделей для

классификации. Таким образом, можно сделать вывод, что, для прогнозирования типа СД с точностью равной 97% наиболее важными характеристиками, описывающих функциональное состояния пациента, являются первые четыре признака: возраст (0,31), наличие сопутствующего заболевания АГ (0,26), наличие генетической предрасположенности (0,25) и ИМТ (0,125).

Рассмотренные в исследовании группы СД определялись по наличию или отсутствию у пациента когнитивных нарушений. В рамках исследования стояла задача определить факторы, предшествующие развитию когнитивных нарушений. Так, на основании отклоненной гипотезы о равенстве медиан выборок, были определены показатели, которые были включены для обучения классификатора случайного леса. В результате исследования, с 80% точностью можно предсказать наличие когнитивных нарушений по факторам, веса важность которых представлены на Рис 3.

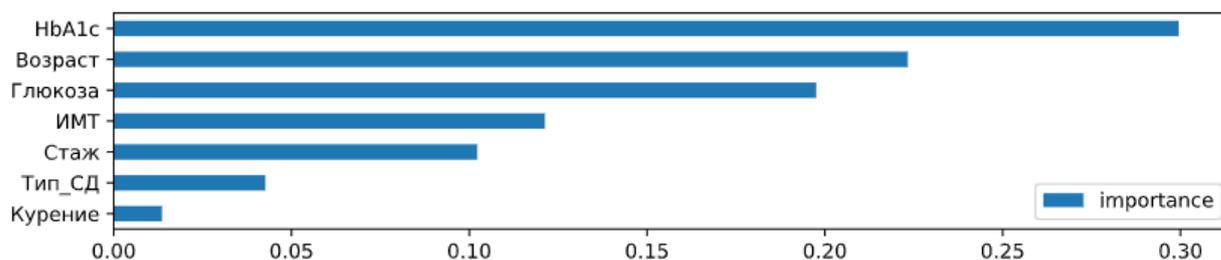


Рис. 3 Значимость факторов при работе классификатора при определении наличия КН

Исследование медианных значений влияющих факторов, представленное в таблице 3, позволяет сделать вывод о причинах развития когнитивных нарушений: высокие значения НbA1c и глюкозы характерны для факта развития КН.

Таблица. 3 Результаты исследования факторов, предшествующих развитию КН

Наличие КН	Тип СД	НbA1c	Возраст	Глюкоза	ИМТ	Стаж	Курение
Отсутствие	Группа контроля	5,0	36	4,75	24,00	0,0	Нет
	1	6,9	26	7,95	21,83	12,5	Нет
	2	5,6	59	5,70	31,00	10,0	Да

Наличие	1	7,6	27	9,10	22,66	11,0	Нет
	2	8,0	59	8,10	32,00	8,0	Да

В рамках данного исследования проведен дисперсионный анализ возможных осложнения СД. В таблицах 4-6 отражены значения корреляции, значения критерия Крускала-Уолисса и соответствующий уровень значимости для факта наличия осложнений Ретинопатии, Нефропатии и Нейропатии соответственно.

Таблица 4. Результаты дисперсионного анализа осложнения Ретинопатия

Фактор риска (ФР), характеристики тяжести и длительности СД (Х), социальные факторы (СФ)	Коэффициент корреляции	Критерий Крускала-Уолисса	
		Статистика	Уровень значимости
Группа (ФР)	-0,67	141,52	,00
Холестерин (Х)	-0,64	9,86	,002
Возраст (ФР)	-0,48	206,27	,00
ИМТ (Х)	-0,46	206,25	,00
Алкоголь (СФ)	-0,24	5,73	0,017
Глюкоза (ФР)	0,21	206,26	,00
Стаж (ФР)	0,26	156,71	,00
ГВ (СФ)	0,56	14,39	,00

Таблица 5. Результаты дисперсионного анализа осложнения Нейропатия

Фактор риска (ФР), характеристики тяжести и длительности СД (Х), социальные факторы (СФ)	Коэффициент корреляции	Критерий Крускала-Уолисса	
		Статистика	Уровень значимости
Группа (ФР)	-0,16	135,92	,00
Холестерин (Х)	-0,18	6,37	,012

ИМТ (X)	-0,22	208,37	,00
---------	-------	--------	-----

Таблица 6. Результаты дисперсионного анализа осложнения Нефропатия

Фактор риска (ФР), характеристики тяжести и длительности СД (X), социальные факторы (СФ)	Коэффициент корреляции	Критерий Крускала-Уолисса	
		Статистика	Уровень значимости
Группа (ФР)	-0,35	132,52	,00
Холестерин (X)	-0,22	4,35	,037
ИМТ (X)	-0,18	210,65	,00
Глюкоза (ФР)	0,23	210,67	,00
Инсулин (ФР)	-0,19	10,84	,00
ГВ (СФ)	0,21	32,23	,00
КН (X)	0,26	8,45	,004
Курение (СФ)	-0,08	7,22	,007
Кофе (СФ)	0,35	5,07	,024

Результаты, полученные в рамках дисперсионного анализа, позволили выявить статистически значимые факторы развития осложнения СД.

Показатели качества построенных моделей машинного обучения для прогнозирования наличия заболеваний представлены в таблице 7.

Таблица 7. Показатели качества построенных моделей машинного обучения для прогнозирования наличия заболеваний

Осложнение	Score	AUC-ROC	Точность обобщения
Ретинопатия	0,85	0,88	0,79
Нефропатия	0,75	0,77	0,85
Нейропатия	0,76	0,71	0,81

На рис 4-6 представлены признаки, которые наиболее информативны в задаче классификации наличия осложнений.

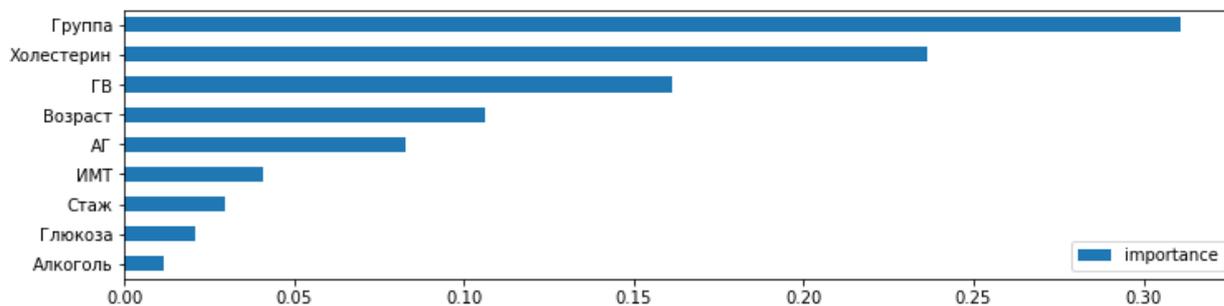


Рис 4. Оценка важности факторов в задачи классификации наличия ретинопатии

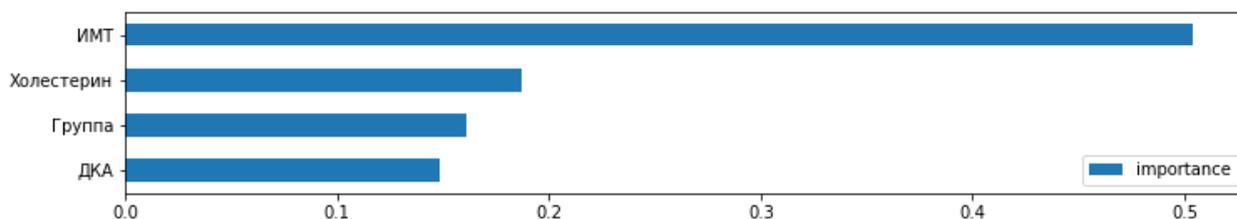


Рис 5. Оценка важности факторов в задачи классификации наличия нейропатии

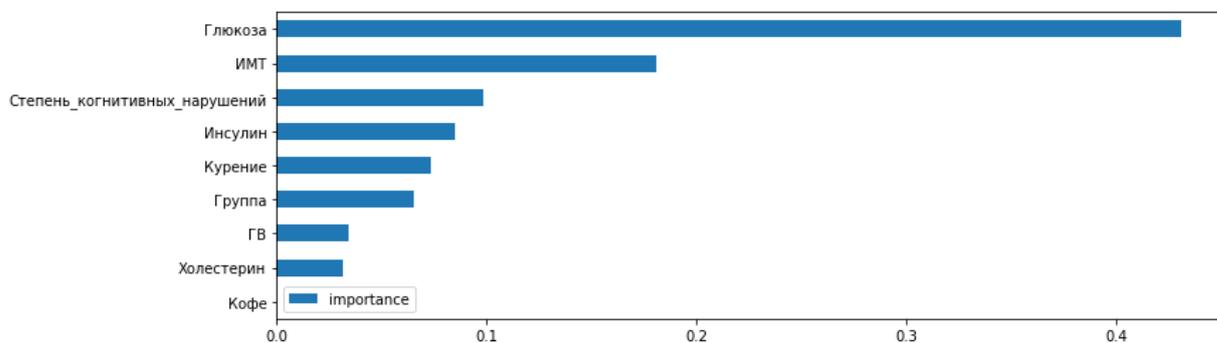


Рис 6. Оценка важности факторов в задачи классификации наличия нефропатии

Таблица 8. Процент наличия осложнений по группам

Группа	Осложнение СД		
	Нейропатия	Нефропатия	Ретинопатия
СД 1 типа без КН	73	93	100
СД 1 типа с КН	85	65	84
СД 2 типа без КН	71	81	33
СД 2 типа с КН	65	60	30

Анализ исследований, проведенных для задачи прогнозирования развития осложнений СД показал значимые результаты для врачебной диагностики заболевания. Принимая во внимание значения корреляции осложнения и признаков с наибольшим весом значимости для работы классификатора, (на Рис. 4-5 первые 4 весомых признака), а также Таблицу 8 можно сделать следующие заключения: диагностика развития ретинопатии характерна для больных СД 1 типа, факторами риска развития являются высокий уровень холестерина, некомпенсированная АГ, ГВ или беременность, высокий ИМТ. Факторами риска развития нейропатии являются высокий ИМТ и высокий уровень холестерина; в группе риска на стадии определения типа СД больные СД 1 типа (73-85%). Факторами риска развития нейропатии являются высокий уровень глюкозы и ИМТ, в группе риска больные СД 1 и 2 типа.

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

Студенту:

Группа	ФИО
8Б61	Борисовой Алене Александровне

Школа	ИЯТШ	Отделение школы (НОЦ)	ЯТЦ
Уровень образования	Бакалавр	Направление/специальность	01.03.02 Прикладная математика и информатика

<b>Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:</b>	
1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	1.Стоимость расходных материалов. 2.Стоимость расхода электроэнергии. 3.Норматив заработной платы.
2. Нормы и нормативы расходования ресурсов	1. Тариф на электроэнергию. 2. Коэффициенты для расчёта заработной платы.
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	1. Отчисления во внебюджетные фонды (30%)
<b>Перечень вопросов, подлежащих исследованию, проектированию и разработке:</b>	
1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения	1.Потенциальные потребители результатов исследования. 2.Анализ конкурентных технических решений. 3.SWOT – анализ.
2. Планирование и формирование бюджета научных исследований	1.Разработка структуры работы в рамках научного исследования; 2.Определение трудоемкости выполнения работ и разработка графика проведения научного исследования; 3. Бюджет научно – технического исследования.
3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	1. Определение показателей ресурсоэффективности разработки
<b>Перечень графического материала (с точным указанием обязательных чертежей):</b>	
1. Оценка конкурентоспособности технических решений 2. Матрица SWOT	

3. Альтернативы проведения НИ
4. График проведения и бюджет НИ
5. Оценка ресурсной, финансовой и экономической эффективности НИ

**Дата выдачи задания для раздела по линейному графику**

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Подопригора Игнат Валерьевич	к.э.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8Б61	Борисова Алена Александровна		

## **5. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ**

Для эффективного использования научного потенциала научно-исследовательской работы (НИР) необходимо прилагать усилия не только к непосредственно её разработке, но и к проведению её анализа с точки зрения экономических требований. Таким образом, целью данного раздела является комплексное описание и анализ финансово-экономических аспектов проведения исследования методов оценки и прогноза функционального состояния больных СД на языке Python.

В разделе «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» части выпускной квалификационной работы рассмотрены следующие вопросы:

- провести SWOT-анализ;
- определить эффективность исследования
- провести планирование научно-исследовательской работы;
- произвести расчёт бюджета научно-исследовательской работы;
- составить оценку научно-технического эффекта.

### **5.1 Организация и планирование работы**

При организации процесса реализации данного исследования необходимо рационально планировать занятость каждого из его участников и сроки проведения отдельных работ.

В данном пункте составляется полный перечень проводимых работ, определяются их исполнители и рациональная продолжительность. Так как число исполнителей редко превышает двух в большинстве случаев, то для наглядного результата чаще пользуются линейным графиком. Для построения такого графика приведем в Таблице 9 перечень работ и занятость исполнителей.

Таблица 9. Перечень работ и продолжительность их выполнения

№ Этапа	Этапы работы	Исполнители	Загрузка исполнителей
1	Постановка целей и задач, получение исходных данных	Научный руководитель	НР – 100%
2	Составление и утверждение ТЗ	Научный руководитель, студент	НР – 100% С – 10%
3	Подбор и изучение материалов по тематике	Научный руководитель, студент	НР – 50% С – 100%
4	Разработка календарного плана	Научный руководитель, студент	НР – 100% С – 10%
5	Обсуждение литературы	Научный руководитель, студент	НР – 30% С – 100%
6	Написание программы	Студент	С – 100%
7	Тестирование программы	Студент	С – 100%
8	Оформление расчетно-пояснительной записки	Студент	С – 100%
9	Оформление графического материала	Студент	С – 100%
10	Анализ полученных результатов	Научный руководитель, студент	НР – 60% С – 100%

### 5.1.1 Продолжительность этапов работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Трудоемкость выполнения проекта оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов. Для определения ожидаемого (среднего) значения трудоемкости  $t_{ож}$  используется следующая формула:

$$t_{ож} = \frac{3 \cdot t_{min} + 2 \cdot t_{max}}{5}, \quad (19)$$

где  $t_{ож}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы чел.-дн.;

$t_{min}$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы (оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств), чел.-дн.;

$t_{max}$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн..

Для выполнения перечисленных в *таблице 8* работ, требуется группа специалистов из следующего состава:

- Студент (С), соискатель степени бакалавра;
- Научный руководитель (НР).

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях  $T_{рд}$ , учитывающая параллельность выполнения работ несколькими исполнителями. Так, для построения линейного графика необходимо рассчитать длительность этапов в рабочих днях, а затем перевести ее в календарные дни. Расчет продолжительности выполнения каждого этапа в рабочих днях ( $T_{рд}$  ведется по формуле:

$$T_{рд} = \frac{t_{ож}}{K_{вн}} \cdot K_{д}, \quad (30)$$

где  $t_{ож}$  – продолжительность работы, дн.;

$K_{вн}$  – коэффициент выполнения работ, учитывающий влияние внешних факторов на соблюдение предварительно определенных длительностей, в частности, возможно  $K_{вн} = 1$ ;

$K_{д}$  – коэффициент, учитывающий дополнительное время на компенсацию непредвиденных задержек и согласование работ ( $K_{д} = 1-1,2$ ; в этих границах конкретное значение принимает сам исполнитель). Возьмем значение  $K_{д} = 1$ .

Продолжительность этапов работ и их трудоемкости по исполнителям, занятым на каждом этапе представлена в Таблице 10.

Таблица 10. Временные показатели проведения научного исследования

№ Этапа	Исполнители	Продолжительность работ, дни			Трудоемкость работ, дни			
		$t_{min}$	$t_{max}$	$t_{ож}$	$T_{рд}$		$T_{кд}$	
					НР	С	НР	С
1	Научный руководитель	1	2	1,6	1,6	-	1,92	-
2	Научный руководитель, студент	5	10	7	7	0,7	8,4	0,84
3	Научный руководитель, студент	10	15	12	6	12	7,2	14,4
4	Научный руководитель, студент	5	10	7	7	0,7	8,4	0,84
5	Научный руководитель, студент	1	2	1,6	1,6	0,48	1,92	0,58
6	Студент	15	20	17	-	17	-	20,4
7	Студент	3	5	3,8	-	3,8	-	4,56
8	Студент	10	20	14	-	14	-	16,8
9	Студент	1	2	1,6	-	1,6	-	1,92
10	Научный руководитель, студент	5	10	7	4,2	7	5,04	8,4
Итого:				<b>72,6</b>	<b>27,4</b>	<b>57,28</b>	<b>32,88</b>	<b>68,74</b>

### 5.1.2 Разработка графика проведения научного исследования

Наиболее удобным и наглядным способом отслеживания выполнения проектной работы является диаграмма Ганта.

Диаграмма Ганта – горизонтальный ленточный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{ki} = T_{pi} \cdot k_{\text{кал}}, \quad (21)$$

где  $T_{ki}$  – продолжительность выполнения этапа в календарных днях;

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

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

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}}, \quad (32)$$

где  $T_{\text{кал}}$  – календарные дни ( $T_{\text{КАЛ}} = 365$ );

$T_{\text{вых}}$  – выходные дни ( $T_{\text{ВД}} = 52$  для при шестидневной рабочей недели);

$T_{\text{пр}}$  – праздничные дни ( $T_{\text{ПД}} = 10$ ).

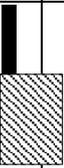
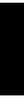
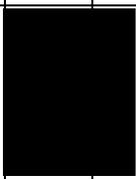
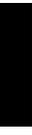
$$k_{\text{кал}} = \frac{365}{365 - 52 - 10} = 1,20.$$

Таким образом, коэффициент календарности  $k_{\text{кал}}$  равен 1,20.

Величины трудоемкости этапов по исполнителям  $T_{ki}$  (данные столбцов 8 и 9 кроме итогов) позволяют построить линейный график осуществления проекта.

Пример построения линейного графика приведен в Таблице 11.

Таблица 11. Линейный график работ

Этап	Вид работ	НР	С	Продолжительность выполнения работ															
				март			апрель			май									
				10	20	30	10	20	30	10	20	30							
1	Постановка целей и задач, получение исходных данных	1,92	-																
2	Составление и утверждение ТЗ	8,4	0,84																
3	Подбор и изучение материалов по тематике	7,2	14,4																
4	Разработка календарного плана	8,4	0,84																
5	Обсуждение литературы	1,92	0,58																
6	Написание программы	-	20,4																
7	Тестирование программы	-	4,56																
8	Оформление расчетно-пояснительной записки	-	16,8																
9	Оформление графического материала	-	1,92																
10	Анализ полученных результатов	5,04	8,4																

 — научный руководитель;

 — Студент.

### 5.1.3 SWOT-анализ

SWOT-анализ представляет собой сводную таблицу, иллюстрирующую связь между внутренними и внешними факторами компании. Целью SWOT-анализа является предоставление возможности оценки риска и конкурентоспособности компании или товара в данной отрасли производства.

Методика SWOT-анализа необходима, для того, чтобы определить наиболее прозрачное положение компании, продукции или услуги в данной отрасли.

Приведем матрицу SWOT-анализа для программы оценки и прогноза функционального состояния развивающихся биосистем на языке Python.

Таблица 12. Матрица SWOT - анализа

	<b>Сильные стороны</b>	<b>Слабые стороны</b>
	<p>С1. Проверка научного потенциала используемых методов для дальнейшего использования в прогнозировании.</p> <p>С2. Проверка прогнозируемых состояний с реальными на основе статистического анализа в рамках доверительных интервалов.</p> <p>С3. Создание адекватных прогнозов функционального состояния на основе малозатратных анализов и тестов.</p>	<p>Сл1. Необходимость использования большого количества данных для обучения и составления адекватного прогнозирования.</p> <p>Сл2. Высокий риск переобучения модели и необходимость подбора гиперпараметров.</p> <p>Сл3. Слабая возможность учета «человеческого фактора», резких выбросов в показателях, неверного ввода информации в компьютер.</p>

<p><b>Возможности</b></p> <p>В. Использование методов для исключения ошибок 1-го и 2-го рода в принятии решений.</p> <p>В2. Поддержка врачей в принятии решений о необходимости проведения дорогостоящих тестов.</p> <p>В3. Повышение осведомленности населения о факторах, свидетельствующих о причинах и последствиях типа диабета.</p>	<p><b>В2С3</b></p> <p>Реализация алгоритма, позволяющего сделать вывод о необходимости для пациента проведения дорогостоящих тестов (МРТ, ФМРТ)</p> <p><b>В1С1</b></p> <p>Реализация возможности верного отнесения пациента к той или иной группе с высоким процентом точности и полноты.</p> <p><b>В3С1</b></p> <p>Формирование особого слоя населения, отличающегося высоким уровнем осведомленности о возможных причинах и осложнениях типа диабета.</p>	<p><b>В1Сл1Сл2</b></p> <p>Возможность улучшения представленного метода ввиду недостатка данных для обучения (например, включение кросс-валидации)</p> <p><b>В1Сл3</b></p> <p>Реализация повышенной восприимчивости метода к выбросам и возможным ошибкам при считывании входных данных.</p>
<p><b>Угрозы</b></p> <p>У1. Наличие малой накопленной базы пациентов</p> <p>У2. Возникновение новых факторов, влияющих на</p>	<p><b>У1С1С2</b></p> <p>Реализация алгоритма, допускающего обучение на небольшой выборке.</p> <p><b>У3С3</b></p> <p>Обеспечение конкуренции за счет</p>	<p><b>У1Сл1Сл2</b></p> <p>Включение в алгоритм методов, позволяющих с высокой точностью обеспечивать качественный прогноз</p> <p><b>У2Сл2</b></p>

определение диабета	типа	малозатратных исследований.	Возможность изменения концепции алгоритма, допускающее учет новых параметров
УЗ. исследования за рубежом	Аналогичные		

Таким образом, можно сделать вывод о том, что наиболее эффективными в сложившейся ситуации представляются следующие стратегии:

- Необходимо повышать уровень осведомленности населения о возможных причинах и осложнениях типа диабета. Это в свою очередь, обеспечит рост заинтересованности к состоянию здоровья и приведет к увлечению базы больных СД
- Необходимо изучить новейшие методы для оценки и прогнозирования функционального состояния пациента с СД при составлении алгоритма работы и планировании прогноза.

#### 5.1.4 Анализ конкурентных решений

Основной целью любого инвестора является сохранение и приумножение своего капитала. Такая цель может достигаться путем вложения денег в банк на банковский депозит, использование услуг брокерских компаний, либо при помощи создания портфеля финансовых инструментов. Каждый из них имеет свои особые признаки, обусловленные принципами организации их деятельности и особенностями функционирования, анализ которых представлен в Таблице 13. Для вышеперечисленных групп инвесторов приведем следующие критерии оценки. Для оценочной карты были выбраны следующие критерии:

- доходность. Данный показатель является одним из основополагающих критериев, так как инвестор выберет путь, приносящий ему наибольшую доходность;
- диверсификация риска позволяет инвестору вложить свои средства в различные группы активов, тем самым снизив риск;
- надежность влияет на психологию инвесторов. Старая и крупная компания, в которую деньги вкладывались десятками лет обладает большим

запасом надежности, чем молодые и неизвестные компании;

- ликвидность показывает способность активов быть быстро проданными по цене, близкой к рыночной;
- ограничения по минимальному вкладу позволят инвесторам мудро распоряжаться своими средствами;
- государственное регулирование обеспечит защиту инвесторов;
- наиболее выгодное предложение будет определяться стоимостью услуг.

Таблица 13. Оценочная карта для сравнения конкурентных решений

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		К <sub>1</sub> (Банк)	Ф (инвестиционный портфель)	К <sub>1</sub> (Брокерская компания)	К <sub>1</sub> Банк	Ф инвестиционный портфель	К <sub>2</sub> Брокерская компания
<b>Экономические критерии оценки эффективности</b>							
1. Доходность	0,45	2	4	4	0,9	1,8	1,8
2. Диверсификация риска	0,25	3	5	2	0,76	1,2 5	0,5
3. Надежность	0,15	5	5	5	0,75	0,7 5	0,7 5
4. Ликвидность	0,05	5	5	3	0,25	0,2 5	0,1 5
5. Ограничения по минимальному вкладу	0,04	5	4	2	0,2	0,1 8	0,0 8
6. Государственное регулирование	0,03	5	5	2	0,15	0,1 6	0,0 8
7. Стоимость услуг	0,03	5	5	4	0,15	0,1 5	0,1 1
<b>Итого</b>	1	3 0	3 3	2 2	3,16	<b>4,5</b> <b>4</b>	3,4 7

Позиция разработки оценивается по каждому показателю экспертным путем по пятибалльной шкале, где 1 – наиболее слабая позиция, а 5 – наиболее сильная. Анализ конкурентных решений определяется по формуле:

$$K = \sum B_i * B_i, (33)$$

где  $K$  – конкурентоспособность решения или конкурента,  $B_i$  – вес показателя (в долях единицы),  $B_i$  – балл  $i$ -го показателя.

Таким образом, можно сделать вывод, что инвестиционный портфель по многим показателям является более предпочтительным, чем другие механизмы инвестирования средств (значение 4,54 является максимальным).

### **5.1.5 Потенциальные потребители результатов исследований**

Для анализа потребителей результатов исследования необходимо рассмотреть целевой рынок и провести его сегментирование.

Целевой рынок – сегменты рынка, на котором будет продаваться в будущем разработка. В свою очередь, сегмент рынка – это особым образом выделенная часть рынка, группы потребителей, обладающих определенными общими признаками.

Сегментирование – это разделение покупателей на однородные группы, для каждой из которых может потребоваться определенный товар (услуга). Можно применять географический, демографический, поведенческий и иные критерии сегментирования рынка потребителей, возможно применение их комбинаций с использованием таких характеристик, как возраст, пол, национальность, образование, любимые занятия, стиль жизни, социальная принадлежность, профессия, уровень дохода.

В зависимости от категории потребителей (коммерческие организации, физические лица) необходимо использовать соответствующие критерии сегментирования. Например, для коммерческих организаций критериями сегментирования могут быть: месторасположение; отрасль; выпускаемая продукция; размер и др.

Для физических лиц критериями сегментирования могут быть: возраст; пол; национальность; образование; уровень дохода; социальная принадлежность; профессия и др.

Из выявленных критериев целесообразно выбрать два наиболее значимых для рынка. На основании этих критериев строится карта сегментирования рынка, которая представлена в Таблице 14.

Таблица 14. Сегментация

	Программа оценки и прогнозирования функционального состояния	Изучение закономерностей	Продолжение исследования
Медицинский персонал			
Научное сообщество, заинтересованное в продолжении исследования			

Очевидно, наиболее перспективным направлением является изучение закономерностей и продолжение текущей работы, что позволит повысить качество программной реализации для оценки и прогнозирования функционального состояния больных СД.

## 5.2 Расчет сметы затрат на выполнение проекта

В состав затрат на создание проекта включается величина всех расходов, необходимых для реализации комплекса работ, составляющих содержание данной разработки. Расчет сметной стоимости ее выполнения производится по следующим статьям затрат:

- Материалы и покупные изделия;
- Заработная плата;
- Социальный налог;
- Расходы на электроэнергию (без освещения);
- Амортизационные отчисления;
- Оплата услуг связи;
- Прочие (накладные расходы) расходы.

### 5.2.1 Расчет материальных затрат

К данной статье расходов относится стоимость материалов, покупных изделий, расходуемых непосредственно в процессе выполнения работ над объектом исследования.

Покажем отражение стоимости всех материалов, используемых при работе над проектом, включая расходы на их приобретение и, при необходимости, доставку. Расчет затрат на материалы производится по форме, приведенной в Таблице 15.

Таблица 15. Материальные затраты

Наименование	Единица измерения	Количество	Цена за ед, руб.	Сумма, руб
Бумага	Пачка	1	300	300
Канцелярские принадлежности	шт.	5	100	500
Картридж для принтера	шт.	1	3000	3000
Итого:				<b>3800</b>

Допустим, что ТЗР составляют 5 % от отпускной цены материалов, тогда расходы на материалы с учетом ТЗР равны:

$$C_{\text{мат}} = 3\,800 * 1,05 = 3\,990 \text{ руб.}$$

### 5.2.2 Расчет заработной платы для исполнителей

Данная статья расходов включает заработную плату научного руководителя и студента (в его роли выступает исполнитель проекта), а также премии, входящие в фонд заработной платы.

Расчет основной заработной платы выполняется на основе трудоемкости выполнения каждого этапа и величины месячного оклада исполнителя.

Среднедневная тарифная заработная плата ( $ЗП_{\text{дн-т}}$ ) рассчитывается по формуле:

$$ЗП_{\text{дн-т}} = \frac{МО}{25,083} \quad (33)$$

Учитывающей, что в году 301 рабочий день и, следовательно, в месяце в среднем 25,083 рабочих дня (при шестидневной рабочей неделе).

Пример расчета затрат на полную заработную плату приведены в таблице 16. Затраты времени по каждому исполнителю в рабочих днях с округлением до целого взяты из таблицы – 9. Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется следующий ряд коэффициентов:  $K_{ГР} = 1,1$ ;  $K_{доп.ЗП} = 1,188$ ;  $K_p = 1,3$ . Таким образом, для перехода от тарифной (базовой) суммы заработка исполнителя, связанной с участием в проекте, к соответствующему полному заработку (зарплатной части сметы) необходимо первую умножить на интегральный коэффициент  $K_{и} = 1,1 * 1,188 * 1,3 = 1,699$ . Вышеуказанное значение  $K_{доп.ЗП}$  применяется при шестидневной рабочей неделе, при пятидневной оно равно 1,113, соответственно в этом случае  $K_{и} = 1,62$ .

Таблица 16. Затраты на заработную плату

Исполнитель	Оклад, руб./мес.	Среднедневная тарифная ставка руб./раб.день	Затраты времени, раб.дни	Коэффициент	Фонд з/платы, руб.
НР	33664	1342,09	28	1,699	<b>63 845,9</b>
С	15470	616,75	58	1,62	<b>57 949,83</b>
Итого					<b>121795,73</b>

### 5.2.3 Расчет затрат на социальный налог

Затраты на единый социальный налог (ЕСН), включающий в себя отчисления в пенсионный фонд, на социальное и медицинское страхование, составляют 30 % от полной заработной платы по проекту, т.е.  $S_{соц} = S_{зп} * 0,3$ .

Итак, в нашем случае:

$$S_{соц} = 121\,795,73 * 0,3 = 36\,538,72 \text{ руб.}$$

### 5.2.4 Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную в ходе выполнения проекта на работу используемого оборудования, рассчитываемые по формуле:

$$C_{\text{эл.об}} = P_{\text{об}} * t_{\text{об}} * \text{ЦЭ}, \quad (34)$$

где  $P_{\text{об}}$  – мощность, потребляемая оборудованием, кВт;

ЦЭ – тариф на 1 кВт·час;

$t_{\text{об}}$  – время работы оборудования, час.

Для ТПУ ЦЭ = 5,748 руб./кВт·час (с НДС).

Время работы оборудования вычисляется на основе итоговых данных таблицы – 9 для студента (ТРД) из расчета, что продолжительность рабочего дня равна 8 часов.

$$t_{\text{об}} = T_{\text{рд}} * K_t, \quad (35)$$

где  $K_t \leq 1$  – коэффициент использования оборудования по времени. Возьмем его равным 1.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{\text{об}} = P_{\text{ном}} * K_C, \quad (36)$$

где  $P_{\text{ном}}$  – номинальная мощность оборудования, кВт;

$K_C \leq 1$  – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности  $K_C = 1$ .

Пример расчета затраты на электроэнергию для технологических целей приведен в Таблице 17.

Таблица 17. Затраты на электроэнергию технологическую

Наименование оборудования	Время работы оборудования $t_{\text{об}}$ , час	Потребляемая мощность $P_{\text{об}}$ , кВт	Затраты Эоб, руб.
Персональный компьютер	464	0,3	800,12
Струйный принтер	2	0,1	1,15
Итого:			<b>801,27</b>

### 5.2.5 Расчет амортизационных расходов

В статье «Амортизационные отчисления» рассчитывается амортизация используемого оборудования за время выполнения проекта.

Используется формула:

$$C_{AM} = \frac{N_A * Ц_{ОБ} * t_{рф} * n}{F_D}, \quad (37)$$

где  $N_A$  – годовая норма амортизации единицы оборудования;

$Ц_{ОБ}$  – балансовая стоимость единицы оборудования с учетом ТЗР;

$t_{рф}$  – фактическое время работы оборудования в ходе выполнения проекта, учитывается исполнителем проекта;

$n$  – число задействованных однотипных единиц оборудования.

Например, для ПК в 2019 г. (299 рабочих дней при шестидневной рабочей неделе) можно принять  $F_D = 299 * 8 = 2392$  часа.

Для принтера из справочника  $F_D = 500$  часов.

При использовании нескольких типов оборудования расчет по формуле делается соответствующее число раз, затем результаты суммируются.

Для ПК найдем  $N_A = 0,4$ . Для принтера  $N_A = 0,5$ .

Стоимость ПК = 20 000 рублей. Время использования 304 часа, тогда для него:

$$C_{AM}(ПК) = \frac{0,4 * 20\,000 * 304 * 1}{2392} = 1551,84 \text{ руб.}$$

Стоимость принтера 5000 руб. Время использования 2 часа, тогда для него:

$$C_{AM}(ПР) = \frac{0,5 * 5\,000 * 2 * 1}{500} = 10 \text{ руб.}$$

Итого начислено амортизации 1 561,84 руб.

### 5.2.6 Расчет прочих расходов

В статье «Прочие расходы» отражены расходы на выполнение проекта, которые не учтены в предыдущих статьях, их следует принять равными 10% от суммы всех предыдущих расходов, т.е.

$$C_{\text{проч}} = (C_{\text{мат}} + C_{\text{зд}} + C_{\text{соц}} + C_{\text{эл.об}} + C_{\text{ам}}) * 0,1 \quad (38)$$

Для нашего примера это:

$$C_{\text{проч}} = (3\,990 + 121\,795,73 + 36\,538,72 + 801,12 + 1\,561,84) * 0,1 \\ = 16\,468,74 \text{ руб.}$$

### 5.2.7 Расчет общей себестоимости разработки

Проведя расчет по всем статьям сметы затрат на разработку, можно определить общую себестоимость проекта. Данные результаты можно посмотреть в Таблице 18.

Таблица 18. Смета затрат на разработку проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	$C_{\text{мат}}$	3 990
Основная заработная плата	$C_{\text{зп}}$	121 795,73
Отчисления в социальные фонды	$C_{\text{соц}}$	36 538,72
Расходы на электроэнергию	$C_{\text{эл.}}$	801,12
Амортизационные отчисления	$C_{\text{ам}}$	1 561,84
Прочие расходы	$C_{\text{проч}}$	16 468,74
<b>Итого:</b>		<b>181 156,15</b>

Таким образом, затраты на разработку составили  $C = 181\,156,15$ руб.

### 5.2.8 Расчет прибыли

Прибыль примем в размере 10 % от полной себестоимости проекта. В нашем примере она составляет 18 115,61 руб. (10 %) от расходов на разработку проекта.

### 5.2.9 Расчет НДС

НДС составляет 20% от суммы затрат на разработку и прибыли. В нашем случае:

$$\text{НДС} = (181\,156,15 + 18\,115,61) * 0,2 = 39\,854,35 \text{ руб.}$$

### 5.2.10 Цена разработки НИР

Цена равна сумме полной себестоимости, прибыли и НДС:

$$C_{\text{НИР(КР)}} = 181\,156,15 + 18\,115,61 + 39\,854,35 = 239\,126,11 \text{ руб.}$$

### 5.3 Оценка научно-технического эффекта

Социально-научный эффект проявляется в росте числа открытий, изобретений, увеличении суммарного объема научно-технической информации, полученной в результате выполнения выпускной квалификационной работы, создании научного «задела», являющегося необходимой предпосылкой для проведения в будущем прикладных исследований и выполнения работа по модернизации конструкций выпускаемых изделий.

За последние годы появились предложения не только по качественной характеристике социального эффекта, но и по системе количественных показателей.

Элементом количественной оценки социально-научного эффекта следует считать определение научно-технического эффекта бакалаврской работы по следующей методике. Сущность этой методики состоит в том, что на основе оценок признаков работы определяется коэффициент научно-технического эффекта ВКР:

$$H_T = \sum_{i=1}^3 r_i * k_i, (40)$$

где  $r_i$  – весовой коэффициент  $i$ -го признака (определяющийся по Таблице 19);  $k_i$  – количественная оценка  $i$ -го признака.

Проведем расчет коэффициента научно-технического эффекта ВКР для программной реализации оценки и прогнозирования функционального состояния больных СД.

Таблица 19. Определение весового коэффициента

Признак научно технического эффекта ВКР(i)	Применение значения весового коэффициента (r)
Уровень новизны	0,5
Теоретический уровень	0,3
Возможность реализации	0,2

Количественная оценка уровня новизны ВКР определяется на основе значений Таблицы 20.

Таблица 20. Количественная оценка уровня новизны ВКР

Уровень новизны разработки	Характеристика уровня новизны	Баллы
Принципиально новая	Результаты исследований открывают новое направление в данной области науки и техники	8-10
Новая	По-новому или впервые объяснены известные факты, закономерности	5-7
Относительно новая	Результаты исследований систематизируют и обобщают имеющиеся сведения, определяют пути дальнейших исследований	2-4
Традиционная	Работа выполнена по традиционной методике, результаты исследования носят информационный характер	1
Не обладающая новизной	Получен результат, который ранее был известен	0

Для данной выпускной квалификационной работы уровень новизны – новая, баллы – 5.

Теоретический уровень полученных результатов выпускной квалификационной работы определяется на основе значения баллов, приведенных в Таблице 21.

Таблица 21. Теоретический уровень полученных результатов в ВКР

Теоретический уровень полученных результатов	Баллы
Установления закона, разработка новой теории	10
Глубокая разработка проблемы: многоаспектный анализ связей, взаимозависимости между фактами с наличием объяснения	8

Разработка способа (алгоритм, программ мероприятий, устройство, и т.д.)	6
Элементарный анализ связей между фактами с наличием гипотезы, симплексного прогноза, классификации, объясняющей версии, или практических рекомендаций частного характера	2
Описание отдельных элементарных фактов (вещей, свойств, отношений); изложение опыта, наблюдений, результатов измерений	0,5

В данной выпускной квалификационной работе была разработана программная реализация оценки и прогнозирования состояния пациента, проведены многоаспектные выводы и определены связи влияющих на состояние факторов, следовательно, теоретический уровень полученных результатов равен 8 баллам.

Возможность реализации научных результатов определяется на основе значения баллов из Таблицы 22.

Таблица 22. Время и масштабы реализации проекта

Время реализации	Баллы
В течение первых лет	10
От 5 до 10 лет	4
Более 10 лет	2
Масштабы реализации	Баллы
Одно или несколько предприятий	2
Отрасль(министерство)	4
Народное хозяйство	10
Примечание: Баллы по времени и масштабам реализации складываются	

Реализация программы оценки и прогнозирования возможна в течение первых лет (10 баллов), реализация возможна при наличии данных о состоянии, как минимум, 30 пациентов для каждого типа СД (10 баллов).

Рассчитаем коэффициент научно-технического эффекта:

$$Нт = 0,5 * 5 + 0,3 * 8 + 0,2 * 20 = 8,9$$

Приведем таблицу оценок уровня научно-технического эффекта.

Таблица 23. Оценка уровня научно-технического эффекта

Уровень научно-технического эффекта	Коэффициент научно-технического эффекта
Низкий	1-4
Средний	5-7
Сравнительно высокий	8-10
Высокий	11-14

В соответствии с Таблицей 23, уровень научно-технического эффекта – сравнительно высокий.

### Выводы по разделу

1. Проведено планирование НИР, а именно: определена структура и календарный план работы, трудоемкость, составлена ленточная диаграмма Гантта, и определен бюджет научно-исследовательской работы. В ходе планирования научно-исследовательских работ определен перечень работ, выполняемый рабочей группой. В данном случае рабочая группа состоит из двух человек: руководитель и инженер. Результаты соответствуют требованиям ВКР по срокам и иным параметрам.

2. Бюджет научно-технического исследования составил 239126 руб. Бюджет НТИ состоит из материальных затрат (3990 рублей), амортизационных отчислений (1561 рублей), затрат на оплаты труда (121795 рублей), отчислений во внебюджетные фонды (36538 рубля) и накладных расходов (16468 рубля).

3. В ходе выполнения раздела «Финансовый менеджмент» с помощью SWOT-анализа были выведены наиболее эффективные в сложившейся ситуации стратегии. После формирования бюджета затрат на проектирование суммарные капиталовложения составили 239126,11 рублей. Уровень научно-технического эффекта – высокий. Проект экономически целесообразен.

4. Капиталовложения в размере 239126,11 рублей позволят реализовать разработанный проект по программной реализации оценки и методов прогнозирования функционального состояния пациентов с сахарным диабетом 1 и 2 типов.

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

<b>Группа</b>	<b>ФИО</b>
8Б61	Борисовой Алене Александровне

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение (НОЦ)</b>	<b>ОАР</b>
Инженерная школа информационных технологий и робототехники	Бакалавриат	Отделение информационных технологий	01.03.02 Прикладная математика и информатика

Тема ВКР: Методы оценки и прогноза функционального состояния развивающихся биосистем	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объект исследования: методы оценки и прогноза функционального состояния развивающихся биосистем.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> <ul style="list-style-type: none"> <li>– специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>– организационные мероприятия при компоновке рабочей зоны.</li> </ul>	Рабочее место должно соответствовать требованиям ГОСТ 12.2.032-78. В соответствии с СН-245-71 в помещении должен быть организован воздухообмен. В соответствии с СН-181-70 рекомендуются следующие цвета окраски помещений: потолок – белый или светлый цветной; стены – сплошные, светло-голубые; пол – темно-серый, темно-красный или коричневый.
<b>2. Производственная безопасность:</b> <p>2.1. Анализ вредных и опасных факторов, которые может создать объект исследования.</p> <p>2.2. Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте при проведении исследований.</p> <p>2.3. Обоснование мероприятий по защите исследователя от действия опасных и вредных факторов.</p>	1. Отклонение показателей микроклимата 2. Превышение уровня шума 3. Отсутствие или недостаток естественного света 4. Недостаточная освещенность рабочей зоны 5. Электромагнитное излучение
<b>3. Экологическая безопасность:</b> <ul style="list-style-type: none"> <li>– защита селитебной зоны</li> <li>– анализ воздействия объекта на атмосферу (выбросы);</li> <li>– анализ воздействия объекта на гидросферу (сбросы);</li> <li>– анализ воздействия объекта на литосферу (отходы);</li> <li>– разработать решения по обеспечению экологической безопасности со ссылками на НТД по охране окружающей среды.</li> </ul>	Анализ негативного воздействия на окружающую природную среду: утилизация люминесцентных ламп, компьютеров и другой оргтехники
<b>4. Безопасность в чрезвычайных ситуациях:</b>	Наиболее типичная ЧС – пожар. Для повышения устойчивости объекта к

<ul style="list-style-type: none"> <li>– перечень возможных ЧС при разработке и эксплуатации проектируемого решения;</li> <li>– выбор наиболее типичной ЧС;</li> <li>– разработка превентивных мер по предупреждению ЧС;</li> <li>– разработка действий в результате возникшей ЧС и мер по ликвидации ее последствий.</li> </ul>	<p>пожарам необходимо использовать огнеупорные материалы, а также ознакомить персонал с режимом работы объекта в случае возникновения ЧС и обучить выполнению конкретных работ по ликвидации очагов поражения.</p> <p>Предусмотренные средства пожаротушения (согласно требованиям противопожарной безопасности СНиП 2.01.02-85): огнетушитель ручной углекислотный ОУ-5, пожарный кран с рукавом и ящик с песком (в коридоре). Кроме того, каждое помещение оборудовано системой противопожарной сигнализации</p>
--	--

Дата выдачи задания для раздела по линейному графику	
--	--

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Матвиенко Владимир Владиславович	-		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8Б61	Борисова Алена Александровна		

## **6. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ**

### **Аннотация**

Представление понятия «Социальная ответственность» сформулировано в международном стандарте (МС) IS CSR-08260008000: 2011 «Социальная ответственность организации».

В соответствии с МС - Социальная ответственность - ответственность организации за воздействие ее решений и деятельности на общество и окружающую среду через прозрачное и этическое поведение, которое:

- содействует устойчивому развитию, включая здоровье и благосостояние общества;
- учитывает ожидания заинтересованных сторон;
- соответствует применяемому законодательству и согласуется с международными нормами поведения (включая промышленную безопасность и условия труда, экологическую безопасность);
- интегрировано в деятельность всей организации и применяется во всех ее взаимоотношениях (включая промышленную безопасность и условия труда, экологическую безопасность).

### **Введение**

Объект исследования – программная реализация методов оценки и прогноза функционального состояния развивающихся биосистем. Научно-исследовательская работа заключалась в разработке и программной реализации методов оценки и прогноза состояния пациентов с развивающимся сахарным диабетом (СД) 1 и 2 типов. Работа выполнялась с использованием ЭВМ. В разделе будут рассмотрены опасные и вредные факторы, оказывающие влияние на производственную деятельность технологического персонала, работающего с автоматизированной системой управления технологическим процессом, рассмотрены воздействия разрабатываемой системы на окружающую среду, правовые и организационные вопросы, а также мероприятия в чрезвычайных ситуациях.

## **6.1. Правовые и организационные вопросы обеспечения безопасности**

### **6.1.1. Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства**

Нормы трудового права – это правила трудовых отношений, установленные или санкционированные государством посредством законодательных актов.

Нормы трудового права регулируют любые отношения, связанные с использованием личного труда.

Формы их реализации разнообразны, в этом разделе будет рассмотрено регулирование коллективных отношений.

Настоящий коллективный договор является правовым актом, регулирующим социально-трудовые отношения работников с работодателем.

Основной задачей коллективного договора является создание необходимых организационно-правовых условий для достижения оптимального согласования интересов сторон трудовых отношений.

По заключенному коллективному договору работодатель обязан:

- соблюдать трудовое законодательство и иные нормативные правовые акты, содержащие нормы трудового права, локальные нормативные акты, условия коллективного договора, соглашений и трудовых договоров;
- обеспечивать безопасность и условия труда, соответствующие государственным нормативным требованиям охраны труда;
- обеспечивать работников оборудованием, инструментами, технической документацией и иными средствами, необходимыми для исполнения ими трудовых обязанностей;
- обеспечивать работникам равную оплату за труд равной ценности, постоянно совершенствовать организацию оплаты и стимулирования труда, обеспечить материальную заинтересованность работников в результатах их труда;
- выплачивать в полном размере причитающуюся работникам заработную плату в сроки, установленные в соответствии с ТК РФ,

- коллективным договором, настоящими Правилами, трудовыми договорами;
- вести коллективные переговоры, а также заключать коллективный договор в порядке, установленном ТК РФ;
  - знакомить работников под роспись с принимаемыми локальными нормативными актами, непосредственно связанными с их трудовой деятельностью;
  - осуществлять обязательное социальное страхование работников в порядке, установленном федеральными законами;
  - возмещать вред, причиненный работникам в связи с исполнением ими трудовых обязанностей, а также компенсировать моральный вред в порядке и на условиях, которые установлены ТК РФ, федеральными законами и иными нормативными правовыми актами РФ;
  - принимать необходимые меры по профилактике производственного травматизма, профессиональных или других заболеваний работников, своевременно предоставлять льготы и компенсации в связи с вредными (опасными, тяжелыми) условиями труда (сокращенный рабочий день, дополнительные отпуска и др.), обеспечивать в соответствии с действующими нормами и положениями, специальной одеждой и обувью, другими средствами индивидуальной защиты;
  - постоянно контролировать знание и соблюдение работниками всех требований инструкций по охране труда, производственной санитарии и гигиене труда, противопожарной безопасности.

Работодатель обязуется проводить аттестацию и сертификацию рабочих мест один раз в пять лет с участием представителя профкома.

Ежегодно издавать приказ о мероприятиях по охране труда и промышленной безопасности, считать эти мероприятия соглашением по охране труда на год.

Порядок обеспечения работников спецодеждой, спецобувью и средствами индивидуальной защиты, стирки и дезинфекции устанавливается локальными

нормативными актами работодателя, принимаемыми по согласованию с профкомом.

### 6.1.2. Организационные мероприятия при компоновке рабочей зоны

Проектирование рабочих мест, снабженных видеотерминалами, относится к числу важных проблем эргономического проектирования в области вычислительной техники.

Организация рабочего места программиста или оператора регламентируется следующими нормативными документами: ГОСТ 12.2.032-78 ССБТ, ГОСТ 12.2.033-78 ССБТ, СанПиН 2.2.2/2.4.1340-03 и рядом других.

Главными элементами рабочего места программиста или оператора являются стол и кресло. Основным рабочим положением является положение сидя.

Рациональная планировка рабочего места предусматривает четкий порядок и постоянство размещения предметов, средств труда и документации. То, что требуется для выполнения работ чаще, расположено в зоне легкой досягаемости рабочего пространства.

Моторное поле - пространство рабочего места, в котором могут осуществляться двигательные действия человека.

Максимальная зона досягаемости рук - это часть моторного поля рабочего места, ограниченного дугами, описываемыми максимально вытянутыми руками при движении их в плечевом суставе.

Оптимальная зона - часть моторного поля рабочего места, ограниченного дугами, описываемыми предплечьями при движении в локтевых суставах с опорой в точке локтя и с относительно неподвижным плечом.

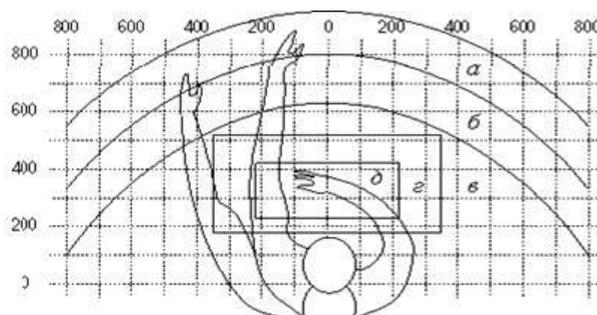


Рис. 7 Зоны досягаемости рук в горизонтальной зоне

- а – зона максимальной досягаемости;
- б – зона досягаемости пальцев при вытянутой руке;
- в – зона легкой досягаемости ладони;
- г – оптимальное пространство для грубой ручной работы;
- д – оптимальное пространство для тонкой ручной работы.

Оптимальное размещение предметов труда и документации в зонах досягаемости:

- дисплей размещается в зоне "а" (в центре);
- системный блок размещается в предусмотренной нише стола;
- клавиатура - в зоне "г"/"д";
- манипулятор "мышь" - в зоне "в" справа;
- документация: необходимая при работе - в зоне легкой досягаемости ладони – "в", а в выдвижных ящиках стола - литература, неиспользуемая постоянно.

Для комфортной работы стол должен удовлетворять следующим условиям:

- высота стола должна быть выбрана с учетом возможности сидеть свободно, в удобной позе, при необходимости опираясь на подлокотники;
- нижняя часть стола должна быть сконструирована так, чтобы программист мог удобно сидеть, не был вынужден поджимать ноги;
- поверхность стола должна обладать свойствами, исключающими появление бликов в поле зрения программиста;
- конструкция стола должна предусматривать наличие выдвижных ящиков (не менее 3 для хранения документации, листингов, канцелярских принадлежностей);
- высота рабочей поверхности рекомендуется в пределах 680- 760мм. Высота поверхности, на которую устанавливается клавиатура, должна быть около 650мм.

Большое значение придается характеристикам рабочего стула (кресла).

Рабочий стул (кресло) должен быть подъемно-поворотным и регулируемым по высоте и углам наклона сиденья и спинки, а также регулируемым по расстоянию спинки от переднего края сиденья. Конструкция стула должна обеспечивать:

- ширину и глубину поверхности сиденья не менее 400 мм; - поверхность сиденья с закругленным передним краем;
- регулировку высоты поверхности сиденья в пределах 400 - 550 мм и углов наклона вперед до 15° и назад до 5°;
- высоту опорной поверхности спинки  $300 \pm 20$  мм, ширину - не менее 380 мм и радиус кривизны горизонтальной плоскости - 400 мм;
- угол наклона спинки в вертикальной плоскости в пределах  $0 \pm 30^\circ$ ;
- регулировку расстояния спинки от переднего края сиденья в пределах 260-400 мм; - стационарные или съемные подлокотники длиной не менее 250 мм и шириной - 50-70 мм;
- регулировку подлокотников по высоте над сиденьем в пределах  $230 \pm 30$  мм и внутреннего расстояния между подлокотниками в пределах 350 - 500 мм

Поверхность сиденья, спинки и других элементов стула (кресла) должна быть полумягкой с нескользящим, неэлектризующимся и воздухопроницаемым покрытием, обеспечивающим легкую очистку от загрязнения.

Создание благоприятных условий труда и правильное эстетическое оформление рабочих мест на производстве имеет большое значение как для облегчения труда, так и для повышения его привлекательности, положительно влияющей на производительность труда.

## **6.2. Профессиональная социальная безопасность**

### **6.2.1. Анализ вредных и опасных факторов, которые может создать объект исследования**

Согласно номенклатуре, опасные и вредные факторы по ГОСТ 12.0.003-2015 ССБТ [33] делятся на следующие группы:

- физические;

- химические;
- психофизиологические;
- биологические.

Перечень опасных и вредных факторов, влияющих на персонал в заданных условиях деятельности, представлен в Таблице 24

Таблица 24 - Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Разра- ботка	Нормативные документы
2. Превышение уровня шума	+	
3.Отсутствие или недостаток естественного света	+	
4.Недостаточная освещенность рабочей зоны	+	
5.Электромагнитное излучение	+	

Эти факторы могут влиять на состояние здоровья, привести травмоопасной или аварийной ситуации, поэтому следует установить эффективных контроль за соблюдением норм и требований, предъявленных к их параметрам.

### 6.2.2. Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте при проведении исследований

При внедрении усовершенствованной системы управления технологическим процессом важную роль играет планировка рабочего места. Она должна соответствовать правилам охраны труда и удовлетворять

требованиям удобства выполнения работы, экономии энергии и времени оператора.

Основным документом, определяющим условия труда на персональных ЭВМ, являются Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы», СанПиН 2.2.2/2.4.1340–03 [37].

Основным опасным фактором является опасность поражения электрическим током. Исходя из анализа состояния помещения, учебных аудиторий кафедры Программной Инженерии НИ ТПУ, по степени опасности поражения электрическим током можно отнести к классу помещений без повышенной опасности (согласно ПУЭ).

Основным опасным производственным фактором на рабочем месте является высокое напряжение в сети, от которой запитана система управления.

### **6.2.3. Обоснование мероприятий по защите исследователя от действия опасных и вредных факторов (техника безопасности и производственная санитария).**

#### **6.2.3.1 Отклонение показателей микроклимата**

Значимым физическим фактором является микроклимат рабочей зоны (температура, влажность и скорость движения воздуха).

Температура, относительная влажность и скорость движения воздуха влияют на теплообмен и необходимо учитывать их комплексное воздействие. Нарушение теплообмена вызывает тепловую гипертермию, или перегрев.

Оптимальные нормы температуры, относительной влажности и скорости движения воздуха производственных помещений для работ, производимых сидя и не требующих систематического физического напряжения (категория Ia), приведены в Таблице 25, в соответствии с СанПиН 2.2.2/2.4.1340-03 [37] и СанПиН 2.2.4.548-96 [34].

Таблица 25. Нормы температуры, относительной влажности и скорости движения воздуха

Период года	Категория работы	Температура, °С	Относительная влаж. воздуха, %	Скорость движения воздуха, не более м/с
Холодный	1а	22-24	40-60	0.1
Теплый	1а	23-25	40-60	0.1

Допустимые микроклиматические условия установлены по критериям допустимого теплового и функционального состояния человека на период 8-часовой рабочей смены. Они устанавливаются в случаях, когда по технологическим требованиям, техническим и экономически обоснованным причинам не могут быть обеспечены оптимальные величины.

Допустимые величины показателей микроклимата на рабочих местах представлены в Таблице 26.

Таблица 26. Допустимые величины показателей микроклимата

Период года	Категория работы	Температура, °С	Относительная влаж. воздуха, %	Скорость движения воздуха, не более м/с
Холодный	1а	20-25	15-75	0.1
Теплый	1а	21-28	15-75	0.1-0.2

Для обеспечения установленных норм микроклиматических параметров и чистоты воздуха на рабочих местах и в помещениях применяют вентиляцию. Общеобменная вентиляция используется для обеспечения в помещениях соответствующего микроклимата. Периодически должен вестись контроль влажностью воздуха. В летнее время при высокой уличной температуре должны использоваться системы кондиционирования.

В холодное время года предусматривается система отопления. Для отопления помещений используются водяные системы центрального отопления. При недостаточной эффективности центрального отопления должны быть использованы масляные электрические нагреватели.

### **6.2.3.3 Отсутствие или недостаток естественного света**

Освещение рабочего места – важнейший фактор создания нормальных условий труда. Освещению следует уделять особое внимание, так как при работе наибольшее напряжение получают глаза.

Освещение делится на естественное, искусственное и совмещенное. Совмещенное сочетает оба вида освещения.

На посту управления, где расположено рабочее место оператора, используется совмещенное освещение.

Для определения приемлемого уровня освещенности в помещении необходимо:

- определить требуемый для операторов уровень освещенности внешними источниками света;
- если требуемый уровень освещенности не приемлем для других операторов, работающих в данном помещении, надо найти способ сохранения требуемого контраста изображения другими средствами.

Рекомендуемые соотношения яркостей в поле зрения следующие:

- между рабочими поверхностями не должно превышать 1:3 – 1:5;
- между рабочими поверхностями и поверхностями стен и оборудования – 1:10.

Освещенность на рабочем месте должна соответствовать характеру зрительной работы, который определяется наименьшим размером объекта различения, контрастом объекта с фоном и характеристикой фона.

Рабочие столы следует размещать таким образом, чтобы видеодисплейные терминалы были ориентированы боковой стороной к световым проемам, чтобы естественный свет падал преимущественно слева.

Освещенность на поверхности стола в зоне размещения рабочего документа должна быть 300 - 500 лк (СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95\* [35], СанПиН 2.2.2/2.4.1340- 03 [37]). Освещение не должно создавать бликов на поверхности экрана. Освещенность поверхности экрана не должна быть более 300 лк. Следует

ограничивать прямую блескость от источников освещения, при этом яркость светящихся поверхностей (окна, светильники и др.), находящихся в поле зрения, должна быть не более 200 кд/м<sup>2</sup>. Показатель ослепленности для источников общего искусственного освещения в производственных помещениях должен быть не более 20.

Согласно СНИП 23-05-95 [35] нормы на освещение для оператора поста управления берутся для производственных помещений. Эти нормы представлены в Таблице 27.

Таблица 27. Нормы на освещение для оператора

Характер зрительной работы	Разряд зрительной работы	Подразряд зрительной работы	Искусственное освещение		Естественное освещение КЕО $e_n$ , % при боковом
			Освещенность при системе общего освещения, лк	Коэффициент пульсации, кп, %	
Различие объектов высокой точности	Б	1	300	15	1.0

#### 6.2.3.4 Превышение уровня шума

В производственных условиях имеют место шумы различной интенсивности и частотного спектра, которые генерируются источниками шумов.

Для исследуемого объекта (производство и пункт управления) основными источниками шумов являются производственное оборудование (внешние источники) и оборудование поста управления (внутренние источники).

ПДУ шума для объектов типа поста управления нормируются ГОСТ 12.1.003-2014 [36] и СН 2.2.4/2.1.8.562–96 [39]. Значения ПДУ согласно этим документам представлены в Таблице 28 (для постоянных шумов)

Таблица 28. Значение ПДУ для постоянных шумов

Рабочие места	Уровни звукового давления (ДБ) в октавных полосах со среднегеометрическими частотами, Гц								Уровни звука и эквивалентные уровни звука, дБА
	63	125	250	500	1000	2000	4000	8000	
ПУ	83	74	68	63	60	78	55	54	65

Для оценки соблюдения ПДУ шума необходим производственный контроль (измерения и оценка). В случае превышения уровней необходимы организационно-технические мероприятия по защите от действия шума (защита временем, расстоянием, экранирование источника, либо рабочей зоны, замена оборудования, использование СИЗ).

#### 6.2.3.5 Электромагнитные излучения

Электромагнитным излучением называется излучение, прямо или косвенно вызывающее ионизацию среды. Контакт с электромагнитными излучениями представляет серьезную опасность для человека, по сравнению с другими вредными производственными факторами (повышенное зрительное напряжение, психологическая перегрузка, сохранение длительное время неизменной рабочей позы).

Когда все устройства персонального компьютера включены, в районе рабочего места программиста, формируется сложное по структуре электромагнитное поле. Реальную угрозу для пользователя компьютера представляют электромагнитные поля. Влияние их на организм человека не обходится без последствий. Исследования показали, что в организме человека под влиянием электромагнитного излучения монитора происходят значительные изменения гормонального состояния, специфические изменения биотоков головного мозга, изменение обмена веществ. Пыль, притягиваемая электростатическим полем монитора, иногда становится причиной дерматитов лица, обострения астматических симптомов, раздражения слизистых оболочек.

Для снижения воздействия электромагнитного излучения следует применять мониторы с пониженным уровнем излучения, также устанавливать

защитные экраны, придерживаться регламентированного режима труда и отдыха, а также проводить регулярную гигиеническую уборку помещения.

Нормы электромагнитных полей, создаваемых ПЭВМ приведены в Таблице 29 и Таблице 30, в соответствии с СанПиН 2.2.2/2.4.1340-03 [37]

Таблица 29. Временные допустимые ЭМП, создаваемые ПЭВМ

Наименование параметров		ВДУ ЭМП
Напряженность электрического поля	В диапазоне частот 5 Гц – 2 кГц	25 В/м
	В диапазоне частот 2 кГц – 400 кГц	2,5 В/м
Плотность магнитного потока	В диапазоне частот 5 Гц – 2 кГц	250 нТл
	В диапазоне частот 2 кГц – 400 кГц	25 нТл
Электростатический потенциал экрана видеомонитора		500 В

Таблица 30. Временные допустимые уровни ЭМП, создаваемые ПЭВМ на рабочих местах

Наименование параметров		ВДУ ЭМП
Напряженность электрического поля	В диапазоне частот 5 Гц – 2 кГц	25 В/м
	В диапазоне частот 2 кГц – 400 кГц	2,5 В/м
Плотность магнитного потока	В диапазоне частот 5 Гц – 2 кГц	250 нТл
	В диапазоне частот 2 кГц – 400 кГц	25 нТл
Электростатический потенциал экрана видеомонитора		500 В

Для оценки соблюдения уровней необходим производственный контроль (измерения). В случае превышения уровней необходимы организационно-технические мероприятия (защита временем, расстоянием, экранирование источника, либо рабочей зоны, замена оборудования, использование СИЗ).

### 6.2.3.6 Электрический ток

Степень опасного воздействия на человека электрического тока зависит от:

- рода и величины напряжения и тока;
- частоты электрического тока;
- пути прохождения тока через тело человека;
- продолжительности воздействия на организм человека;
- условий внешней среды.

Согласно ПУЭ аудиторию КЦ–105 НИ ТПУ по степени опасности поражения электрическим током можно отнести к классу помещений без повышенной опасности.

Основными мероприятиями по защите от электропоражения являются:

- обеспечение недоступности токоведущих частей путем использования изоляции в корпусах оборудования;
- применение средств коллективной защиты от поражения электрическим током;
- защитного заземления, зануления (ГОСТ 12.1.030–81 ССБТ [38]);
- защитного отключения;
- использование устройств бесперебойного питания.

Технические способы и средства применяют отдельно или в сочетании друг с другом так, чтобы обеспечивалась оптимальная защита.

Электробезопасность должна обеспечиваться (ГОСТ Р 12.1.019-2009 ССБТ [40]):

- конструкцией электроустановок;
- техническими способами и средствами защиты;
- организационными и техническими мероприятиями.

### **6.3. Экологическая безопасность**

Рассмотрим загрязнения литосферы в результате исследовательской деятельности бытовым мусором, на примере люминесцентных ламп. Их эксплуатация требует осторожности и четкого выполнения инструкции по обращению с данным отходом (код отхода 35330100 13 01 1, класс опасности – 1[41]). В данной лампе содержится опасное вещество ртуть в газообразном состоянии. При не правильной утилизации, лампа может разбиться и пары ртути могут попасть в окружающую среду. Вдыхание паров ртути может привести к тяжелому повреждению здоровья.

При перегорании ртутьсодержащей лампы (выходе из строя) ее замену осуществляет лицо, ответственное за сбор и хранение ламп (обученное по электробезопасности и правилам обращения с отходом). Отработанные

люминесцентные лампы сдаются только на полигон токсичных отходов для захоронения. Запрещается сваливать отработанные люминесцентные лампы с мусором [41].

Бытовой мусор помещений организаций несортированный, образованный в результате деятельности работников предприятия (код отхода 91200400 01 00 4). Агрегатное состояние отхода твердое; основные компоненты: бумага и 67 древесина, металлы, пластмассы и др [41]. Для сбора мусора рабочее место оснащается урной. При заполнении урны, мусор выносится в контейнер бытовых отходов. Предприятие заключает договор с коммунальным хозяйством по вывозу и размещению мусора на организованных свалках.

## **6.4. Безопасность в чрезвычайных ситуациях**

### **6.4.1. Анализ вероятных ЧС, которые может инициировать объект исследований**

Перечень возможных ЧС на объекте исследования может быть достаточно широк. Ограничиваясь местоположением объекта и условиями его эксплуатации, его можно представить следующим (ориентировочным) вариантом:

- наводнение;
- удар молнии;
- пожар на объекте;
- взрыв.

В этом разделе наиболее актуальным будет рассмотрение вида ЧС – пожар, определение категории помещения по пожаровзрывобезопасности в котором происходит управление технологическим процессом, то есть аудитория КЦ–105 НИ ТПУ и регламентирование мер противопожарной безопасности.

Рабочее место оператора поста управления, должно соответствовать требованиям Ф3 Технический регламент по ПБ о требованиях пожарной безопасности [42].

По пожарной, взрывной, взрывопожарной опасности помещение относится к категории Д, т.е. к помещению, в котором находятся негорючие вещества и материалы в холодном состоянии.

Основным поражающим фактором пожара для помещений данной категории является наличие открытого огня и отравление ядовитыми продуктами сгорания оборудования

#### **6.4.2. Анализ вероятных ЧС, которые могут возникнуть на рабочем месте при проведении исследований**

Пожар в помещении оператора может возникнуть вследствие причин неэлектрического и электрического характера.

К причинам неэлектрического характера относятся халатное и неосторожное обращение с огнем (курение, оставление без присмотра нагревательных приборов). К причинам электрического характера относятся:

- короткое замыкание;
- перегрузка проводов;
- большое переходное сопротивление;
- искрение;
- статическое электричество.

Режим короткого замыкания – появление в результате резкого возрастания силы тока, электрических искр, частиц расплавленного металла, электрической дуги, открытого огня, воспламенившейся изоляции.

Причины возникновения короткого замыкания:

- ошибки при проектировании
- старение изоляции
- увлажнение изоляции
- механические перегрузки.

Пожарная опасность при перегрузках – чрезмерное нагревание отдельных элементов, которое может происходить при ошибках проектирования в случае длительного прохождения тока, превышающего номинальное значение.

Пожарная опасность переходных сопротивлений – возможность воспламенения изоляции или других близлежащих горючих материалов от тепла, возникающего в месте аварийного сопротивления (в переходных клеммах, переключателях и др.).

#### **6.4.3. Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС**

Пожарная безопасность объекта должна обеспечиваться системами предотвращения пожара и противопожарной защиты, в том числе организационно-техническими мероприятиями.

Пожарная защита должна обеспечиваться применением средств пожаротушения, а также применением автоматических установок пожарной сигнализации.

Должны быть приняты следующие меры противопожарной безопасности:

- обеспечение эффективного удаления дыма, т.к. в помещениях, имеющих оргтехнику, содержится большое количество пластиковых веществ, выделяющих при горении летучие ядовитые вещества и едкий дым;
- обеспечение правильных путей эвакуации;
- наличие огнетушителей и пожарной сигнализации;
- соблюдение всех противопожарных требований к системам отопления и кондиционирования воздуха.

Для тушения пожаров на участке производства необходимо применять углекислотные (ОУ-5 или ОУ-10) и порошковые огнетушители (например, типа ОП-10), которые обладают высокой скоростью тушения, большим временем действия, возможностью тушения электроустановок, высокой эффективностью борьбы с огнем.

Помещение (КЦ НИ ТПУ) оборудовано пожарными извещателями, которые позволяют оповестить дежурный персонал о пожаре. В качестве пожарных извещателей в помещении устанавливаются дымовые фотоэлектрические извещатели типа ИДФ-1 или ДИП-1.

Выведение людей из зоны пожара должно производиться по плану эвакуации. План эвакуации представляет собой заранее разработанный план (схему), в которой указаны пути эвакуации, эвакуационные и аварийные выходы, установлены правила поведения людей, порядок и последовательность действий в условиях чрезвычайной ситуации по п. 3.14 ГОСТ Р 12.2.143-2002 [43].

Согласно Правилам пожарной безопасности, в Российской Федерации ППБ 01-2003 (п. 16) в зданиях и сооружениях (кроме жилых домов) при одновременном нахождении на этаже более 10 человек должны быть разработаны и на видных местах вывешены планы (схемы) эвакуации людей в случае пожара.

Ответственность за нарушение Правил пожарной безопасности, согласно действующему федеральному законодательству, несет руководитель объекта.

### **Выводы и рекомендации по разделу**

Проанализировав условия труда на рабочем месте, где было проведено исследование в рамках данной выпускной квалификационной работы, сделаны выводы о том, что помещение удовлетворяет необходимым нормам и, в случае соблюдения техники безопасности и правил пользования компьютером, работа в данном помещении не приведет к ухудшению состояния работника, выполнявшего исследование.

Само помещение и рабочее место в помещении удовлетворяет нормативным требованиям. Кроме того, действие вредных и опасных факторов сведено к минимуму, т.е. микроклимат, освещение и электробезопасность соответствуют требованиям, предъявленным в соответствующих нормативных документах.

Относительно рассмотренного вопроса об экологической безопасности можно сказать, что деятельность, выполненная в рамках данной работы не представляет опасности окружающей среде.

Важно добавить, что монитор компьютера может служить источником ЭМП – вредного фактора, который негативно влияет на здоровье работника при продолжительной непрерывной работе и приводит к снижению

работоспособности. Поэтому, необходимо делать перерывы при работе с ЭВМ во избежание негативного влияния на здоровье.

## ЗАКЛЮЧЕНИЕ

Исследование, выполненное в рамках данной работы, показало принципиальную возможность анализа большого количества накопленных данных методами информационного, статистического анализа и машинного обучения. В основе информационного подхода, применяемого с целью обобщенного оценивания функционального состояния, реализована методика оценивания функционального состояния на основе меры близости наблюдаемого и предпочтительного состояний, где мера близости нормирована в метрике Махаланобиса на внутримножественное расстояние области предпочтительного состояния. Построенный таким образом интегральный показатель функционального состояния пациентов инвариантен к количеству и малому изменению факторов риска СД и характеристикам тяжести и длительности протекания заболевания.

Применение в исследовании метода машинного обучения - случайного леса позволило решить задачу по прогнозированию типа сахарного диабета (СД) и наличию осложнений у пациентов с более, чем 80% и 75% точностью соответственно. Алгоритмы исследования направлены на увеличение качества оценки функционального состояния пациентов, автоматизацию процесса анализа медицинских данных, раннее выявление осложнений у больных с сахарным диабетом и возможность своевременного прогноза развития когнитивных нарушений. Результаты, полученные в ходе исследования, подтверждены врачами центра МРТ г. Томска и сходятся с ведущими исследованиями развития СД и гликемических осложнений.

Были разработаны разделы «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» и «Социальная ответственность».

## СПИСОК ИСТОЧНИКОВ

1. Современное состояние проблем профилактики и лечения сахарного диабета: обзор / С.С. Жестовский, Л.В. Петрова, А.С. Аметов[Текст]: // Терапевтический архив . – 2007. – Т. 79, № 10 – с. 46 – 50
2. Алгоритмы специализированной медицинской помощи больным сахарным диабетом / Под редакцией И.И. Дедова, М.В. Шестаковой, А.Ю. Майорова. – 9-й выпуск (дополненный). – М.; 2019.
3. Потемкин В.В.. Эндокринология: Учебник. – 3-е изд., перераб. и доп. – В.В. Потемкин. [Текст] М.: Медицина, 1999. – 640 с.
4. Приказ № 262 от 7 апреля 2005г. Министерства здравоохранения и социального развития Российской Федерации «Об утверждении стандарта медицинской помощи больным сахарным диабетом». – М ., 2005.
5. Современное состояние проблем профилактики и лечения сахарного диабета: обзор / С.С. Жестовский, Л.В. Петрова, А.С. Аметов[Текст]: // Терапевтический архив . – 2007. – Т. 79, № 10 – с. 46 – 50
6. Ионова Т.И., Один В.И., Никитина Т.В, Курбатова К.А. "Изучение качества жизни и симптомов гипогликемии у больных сахарным диабетом 2-го типа, получающих интенсифицированную инсулинотерапию, как модель пациент-ориентированного подхода" Сахарный диабет, vol. 18, no. 4, 2015, pp. 48-58
7. Францева А.П., Карпов С.М., Белякова Н.А., Штемберг Л.В., Муравьев К.А., Осипова Н.А.. "Оценка качества жизни при сахарном диабете 1-го типа" Кубанский научный медицинский вестник, no. 2, 2014, pp. 114-118.
8. Субботин С. А. Построение деревьев решений для случая малоинформативных признаков // Радіоелектроніка, інформатика, управління. 2019. №1 (48)
9. Фокин В.А, Пеккер Я. С., Берестнева О.Г., Гергет О.М. Интегральные методы оценки состояния сложных систем // Известия Томского политехнического университета, - 2012, - №5. С. 120-124

10. Фокин Василий Александрович. "Интегральная оценка состояния сложных биомедицинских систем" Векторы благополучия: экономика и социум, № 1 (1), 2011, С. 656-667.
11. Зубов Н.Н. Методы многомерного статистического анализа данных в медицине: учебное пособие/ Н.Н. Зубов, В.И. Кувакин / под общ ред. доцента Зубова Н.Н.. – СПб.: Изд-во ООО "Литография Принт", 2017. – 348с., илл.
12. Теория и практика машинного обучения: учеб. пособие / В.В. Воронина, А.В. Михеев, Н.Г. Ярушкина, К.В. Святлов. Ульяновск: УлГТУ, 2017. 290с.
13. Клячкин В.Н., Кувайскова Ю.Е., Жуков Д.А. "Выбор метода бинарной классификации при технической диагностике с применением машинного обучения" Известия Самарского научного центра Российской академии наук, vol. 20, no. 4-3, 2018, pp
14. Кротова О.С., Пиянзин А.И., Хворова Л.А., Жариков А.В. Некоторые математические подходы в построении моделей прогнозирования стадий компенсации и декомпенсации сахарного диабета у детей и подростков // Известия АлтГУ. 2018. №4 (102)
15. Чистяков Сергей Павлович. "Случайные леса: обзор" Труды Карельского научного центра Российской академии наук, no. 1, 2013, pp. 117-136
16. Анохин П.К. Очерки по физиологии функциональных систем. М., 1975
17. Фокин В. А. Системный подход к интегральной оценке состояния биосистем. / В кн.: Современные методы представления и обработки биомедицинской информации. Под ред. Ю. В. Кистенева, Я. С. Пеккера. – Томск: Изд-во ТПУ, 2004. – С. 51-123.
18. А.Л. Филинов, Л.Б. Брагина. Функциональное состояние вегетативной нервной системы при нормально протекающем гестационном процессе // Медицинский альманах, -2012, - №5(24). С.44-45.
19. Шеффе Г. Дисперсионный анализ. – М.: Наука, 1980. – 512 с
20. Холлендер М., Вульф Д.А. Непараметрические методы статистики, пер. с англ. – М.: Мир, 1983. – 518 с.

21. Айвазян С. А., Бухштабер В. Н., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерностей. Справочное издание / Ред. Айвазян С. А.// Финансы и статистика. М., 1989. С. 607
22. Breiman L., Friedman R., Olshen R., Stone C. Classification and Regression Trees. Belmont, California: Wadsworth International, 1984. 342 p
23. Kohavi, Ron (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection". *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann. 2 (12): 1137–1143
24. Cutler A., Breiman L. RAFT: Random Forest Tool. URL: <http://www.stat.berkeley.edu/users/breiman/RandomForests/>
25. Breiman L. Bagging predictors // *Machine Learning*. 1996. Vol. 24, N 2. P. 123–140.
26. Siroky D. Navigating Random Forests and related advances in algorithmic modeling // *Statistics Surveys*. 2009. Vol. 3. P. 147–163
27. Breiman L. Out-of-bag estimation // Technical report, Statistics Department University of California, Berkeley. 1996. P. 1–13. URL: <ftp://ftp.stat.berkeley.edu/pub/users/breiman/OOBest>
28. Рашка, С. Python и машинное обучение [Текст] / С. Рашка. – М. : ДМК Пресс, 2017. – 418 с.
29. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Текст] / П. Флах. – М.: ДМК Пресс, 2015. – 400 с.]
30. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. II. Непараметрические критерии. – М. : Изд-во стандартов, 2002. – 64 с
31. Shapiro, S. S. & Wilk, M.B (1965). An analysis of variance test for normality (complete samples), *Biometrika*, Vol. 52, pp. 591-611
32. Кобзарь А. И. Прикладная математическая статистика. — М.: Физматлит, 2006. — 624-626 с

33. ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация.
34. СанПиН 2.2.4.548–96 Гигиенические требования к микроклимату производственных помещений
35. СП 52.13330.2016 Нормы естественного и искусственного освещения предприятий, (Актуализированная редакция СНиП 23-05-95).
36. ГОСТ 12.1.003-2014 ССБТ Шум. Общие требования безопасности.
37. СанПиН 2.2.2/2.4.1340–03 Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».
38. ГОСТ 12.1.030-81 ССБТ Электробезопасность, Защитное заземление, зануление.
39. СН 2.2.4/2.1.8.562–96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории застройки.
40. ГОСТ 12.1.019-2017 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты.
41. Федеральный классификационный каталог отходов [Электронный ресурс]. – 2013. – Режим доступа: <http://www.ecoguild.ru/faq/fedwastecatalog.htm>, свободный
42. Федеральный закон от 22.07.2013 г. №123 – ФЗ, Технический регламент о требованиях пожарной безопасности
43. ГОСТ Р 12.2.143-2002 ССБТ. Системы фотолюминесцентные эвакуационные. Элементы систем. Классификация. Общие технические требования. Методы контроля.

## ПРИЛОЖЕНИЕ А

Наименование и границы нормы исследуемых лабораторных показателей.

Наименование показателя	Нижняя граница нормы	Верхняя граница нормы
Глюкоза, ммоль/л	3,3	5,5
Общий холестерин, ммоль/л	3,6	7,8
ИМТ (индекс массы тела)	18,5	25
НbA1c (гликированный гемоглобин), %	4,27	6,07

## ПРИЛОЖЕНИЕ Б

### Значения медицинских показателей больного СД

Показатели	Значение
Индекс массы тела	Величина, позволяющая оценить степень соответствия массы человека и его роста и тем самым, косвенно, оценить, является ли масса недостаточной, нормальной, избыточной (ожирение).
Гликированный гемоглобин, HbA1c	Форма гемоглобина, образующаяся в результате его взаимодействия с глюкозой. Гликированный гемоглобин отражает повышение сахара крови, имевшее место на протяжении периода жизни эритроцитов (до 120 суток). Эритроциты, циркулирующие в крови, имеют разный возраст, поэтому ориентируются на срок - 60 суток. Данный лабораторный тест широко используется для оценки степени компенсации сахарного диабета
Артериальная гипертензия, АГ	Заболевание сердечно-сосудистой системы, при котором давление крови в артериях системного (большого) круга кровообращения стабильно повышено.
Диабетическая нейропатия	Комплекс клинических и субклинических синдромов, каждый из которых характеризуется диффузным или очаговым поражением периферических и/или автономных нервных волокон в результате СД. В группе риска заболевания ДН больные СД 1 типа с недостижением целевых уровней гликемического контроля спустя 5 лет от дебюта заболевания, больные СД 2 типа с момента диагностики заболевания.
Диабетическая ретинопатия	Микрососудистое осложнение СД, характеризующееся поражением сетчатки в результате ишемии, повышения проницаемости и эндотелиальной дисфункции сосудов, приводящее к значительному снижению зрения вплоть до его полной потери. Основными факторами риска являются гликемический уровень, контроль артериального давления.
Диабетическая нефропатия	Специфическое поражение почек при СД, сопровождающееся формированием узелкового гломерулосклероза, приводящего к развитию терминальной почечной недостаточности, требующей проведения заместительной почечной терапии (диализ, трансплантация). По данным скрининга, ДН наблюдается у пациентов с СД 1 типа с длительностью заболевания >5, а также у всех пациентов с СД 2 типа; характерно наличие диабетической ретинопатии

## ПРИЛОЖЕНИЕ В

Образов	ИМТ	Степень_	Мока_ба	НбА1с	Холестер	Глюкоза	Стаж	Возраст	Тип_СД	Группа	
										1	Группа
									1	0,909987	Тип_СД
								1	0,875732	0,780814	Возраст
							1	-0,117455	-0,311177	-0,271255	Стаж
						1	-0,007343	-0,258607	-0,256718	-0,373032	Глюкоза
					1	-0,204211	-0,303887	0,845066	0,949649	0,845585	Холестер
				1	0,143346	0,509998	-0,14164	0,027422	0,048174	-0,115205	НбА1с
			1	0,3982	0,033756	0,352167	-0,030225	-0,012372	-0,027202	-0,431532	Мока_ба
		1	0,828131	0,313632	0,139301	0,242236	-0,064993	0,150411	0,111929	-0,234861	Степень_
	1	0,152137	0,050392	0,150302	0,672732	-0,06391	-0,174605	0,557905	0,705625	0,619749	ИМТ
1	0,060929	0,057145	-0,024377	0,104802	0,00753	0,026443	-0,052712	-0,004741	0,025487	0,038964	Образов
-0,239023	-0,093703	0,022618	0,02915	0,064062	0,007289	0,016883	0,11333	0,044246	-0,051322	-0,055727	Гипо
-0,176825	0,291006	0,175314	0,10714	0,127235	0,392234	-0,028512	-0,06213	0,361185	0,365663	0,295235	ДКА
-0,188386	0,30511	0,075991	0,057539	0,079386	0,470532	-0,05257	-0,097849	0,447692	0,477984	0,407235	Курение
-0,020337	0,320649	0,08809	0,025403	-0,011021	0,430297	-0,089108	-0,074232	0,400589	0,437363	0,370702	Алкоголь
0,035189	-0,587618	-0,150903	-0,008711	-0,044505	-0,767393	0,193558	0,273626	-0,686709	-0,808081	-0,722693	ГВ
0,096817	-0,442835	0,073939	0,14683	-0,008629	-0,628552	0,245258	0,211168	-0,536571	-0,661878	-0,655394	Кофе
-0,034759	0,563495	0,18892	0,035265	0,089352	0,838091	-0,153882	-0,282358	0,77058	0,871678	0,766078	Инсулин
-0,024852	0,667334	0,122313	-0,008119	0,011146	0,880555	-0,243115	-0,30242	0,842789	0,932899	0,836342	АГ
-0,105358	-0,523513	0,061877	0,220166	-0,021234	-0,612115	0,196347	0,179287	-0,527744	-0,665226	-0,691625	Ретинопа
0,042487	-0,184909	0,259121	0,312899	0,132467	-0,224684	0,227276	0,095193	-0,171434	-0,247808	-0,348293	Нефропа
-0,066523	-0,224338	-0,051519	-0,033349	-0,038616	-0,184007	0,147057	0,103935	-0,143778	-0,195019	-0,157247	Нейропат
0,001614	0,597434	0,062991	-0,071764	-0,024995	0,782344	-0,194277	-0,286818	0,749631	0,88195	0,814231	Гены

	Ген	Нейропат	Нефропа	Ретинопа	АГ	Инсулин	Кофе	ГВ	Алкоголь	Курение	ДКА	Гипо
Группа												
Тип_СД												
Возрас												
Стаж												
Глюкоз												
Холест												
НbA1с												
Мока_б												
Степен												
ИМТ												
Образо												
Гипо												1
ДКА											1	0,502298
Курени										1	0,361858	0,246848
Алкого									1	0,54855	0,388733	0,332061
ГВ								1	-0,346381	-0,390021	-0,297817	0,073125
Кофе							1	0,549043	-0,183133	-0,214102	-0,17425	0,02933
Инсули						1	-0,576944	-0,704386	0,404441	0,431322	0,396934	0,071657
АГ					1	0,831258	-0,617466	-0,753858	0,492853	0,5132	0,41369	0,050385
Ретино				1	-0,572755	-0,558299	0,463827	0,551788	-0,330888	-0,331186	-0,251753	0,027288
Нефро			1	0,269796	-0,211731	-0,189249	0,355401	0,206411	-0,057588	-0,088406	0,015745	0,124978
Нейроп		1	0,302855	0,084413	-0,167295	-0,119971	0,178681	0,114519	-0,028175	-0,019743	0,087033	0,148046
Гены	1	-0,09446	-0,207034	-0,579	0,822063	0,794	-0,583744	-0,712687	0,428666	0,422223	0,268562	-0,09232

## ПРИЛОЖЕНИЕ Г. Листинг

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import seaborn as sns
from scipy.stats import kruskal

# Линейная интерполяция
data = data.fillna(patient_info.interpolate(method='linear'))
# Проверка на нормальность, визуализация
for factor in data:
    stats.probplot(data [factor], dist='norm', plot=plt)
    print('Shapiro-Wilk test for normality:', data [factor].name,
          stats.shapiro(data [factor]))
    plt.show()
# Корреляция между переменными
corrMatrix = data.corr()
sns.heatmap(corrMatrix, annot=True)
plt.show()
# Функция расчета расстояния Махаланобиса
def Mahalanobis(x, covariance_matrix, mean):
    x = np.array(x)
    mean = np.array(mean)
    covariance_matrix = np.array(covariance_matrix)
    t1 = (x-mean)
    icov = np.linalg.inv(covariance_matrix)
    t2 = (x.transpose()-mean.transpose())
    mahal = t1.dot(icov.dot(t2))
    return mahal

# H-критерий Краскела-Уоллиса для категориальной переменной Тип_СД
(аналогично Наличия_КН и осложнений)
for factor in data:
    if (kruskal(data[factor], data['Тип_СД']).pvalue < 0.05):
```

```

print(data[factor].name, 'Гипотеза о равенстве медиан
выборок отвергается', '\n',
'Статистика Крускала-Уолисса: ', kruskal(data[factor],
data['Тип_СД']).statistic, '\n', 'p-value:',
kruskal(data[factor], data['Тип_СД']).pvalue)
else:
print(data[factor].name, 'Гипотеза о равенстве медиан
выборок НЕ ОТВЕРГАЕТСЯ при', '\n', 'p-value: ',
kruskal(data[factor], data['Тип_СД']).pvalue)
# Статистика Манна-Уитни для исследования медиан средние (аналогично
для аналогично Наличия_КН и осложнений)
for n in range(0,3):
for m in range(0,3):
dfnm = data.loc[(data['Тип_СД'] == n) | (data['Тип_СД'] == m)]
if (n!=m):
print(dfnm['Степень_когнитивных_нарушений'].name,
'Тип СД:', n, m, '\n' 'Статистика Манна-Уитни:',
mannwhitneyu(dfnm['Степень_когнитивных_нарушений'],
dfnm['Тип_СД']).statistic, '\n' 'p-value:',
mannwhitneyu(dfnm['Степень_когнитивных_нарушений'],
dfnm['Тип_СД']).pvalue)
# Рандомный лес с настройкой гиперпараметров
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import GridSearchCV
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size
= 0.33)
# n_estimators - количество деревьев
# max_depth - максимальная глубина деревьев
# min_samples_leaf - минимальное количество образцов
# должно быть в листовом узле
# min_samples_split - минимальное количество выборок,
# необходимое для разделения внутреннего узла
params = {'n_estimators': range(4,20,2),
'max_depth':range(4,12,2),
'min_samples_leaf':range(2,7),

```

```

        'min_samples_split':range(2,9,2),
        'criterion':['gini', 'entropy']}]
# Классификатор, для которого будут подобраны гиперпараметры -
случайный лес
clf = RandomForestClassifier(random_state=42)
# Классификатор, для которого будут подобраны гиперпараметры -
случайный лес
clf = RandomForestClassifier(random_state=42)
# Используем перекрестную проверку cv = 4
Grid_Search_CV_RF = GridSearchCV(estimator = clf,
                                param_grid = params,
                                n_jobs=-1,cv = 4)
# Подбор классификатора с лучшими гиперпараметрами для модели на
текущем наборе данных
Grid_Search_CV_RF.fit(X_train, y_train)
# Классификатор random forest с лучшими параметрами для данного
набора данных
best_clf = Grid_Search_CV_RF.best_estimator_
# Тренировка классификатора с наилучшими гиперпараметрами
best_clf.fit(X_train, y_train)
best_clf.score(X_test, y_test)
# Тренировка классификатора с наилучшими гиперпараметрами
best_clf.fit(X_train, y_train)
best_clf.score(X_test, y_test)
# Действующая классификация
best_clf_predictions = best_clf.predict(X_test)
# Вероятности для каждого класса
best_clf_probs = best_clf.predict_proba(X_test)
# Расчет прогнозов классификации (predict) наряду
# с прогностической вероятностью (predict_proba),
# чтобы вычислить ROC AUC
from sklearn.metrics import roc_auc_score

# Рассчитываем roc auc
roc_value = roc_auc_score(y_true = y_test,
                          y_score = best_clf_probs,

```

```
multi_class="ovr")

roc_value

# Визуализация лучших определяющих параметров случайного леса
imp = pd.DataFrame(best_clf.feature_importances_,
index=X_train.columns, columns=['importance'])
imp.sort_values('importance').plot(kind='barh', figsize=(12, 4))
```