

Автономная некоммерческая  
образовательная организация высшего образования  
**«НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»**

Научный центр информационных технологий и искусственного интеллекта  
направление «Математическое моделирование в биомедицине и геофизике»

К ЗАЩИТЕ ДОПУСТИТЬ  
Руководитель направления  
«Математическое моделирование в  
биомедицине и геофизике»  
к.т.н., д.э.н.  
\_\_\_\_\_ М. В. Ширяев  
« \_\_\_\_ » \_\_\_\_\_ 2025 г.

**«АНАЛИЗ ФОРМЫ И СКОРОСТИ ПУЛЬСОВОЙ ВОЛНЫ ДЛЯ  
ДИАГНОСТИКИ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ»**

Магистерская диссертация  
по направлению подготовки 01.04.02 «Прикладная математика и  
информатика»,  
направленность (профиль) «Математическое моделирование процессов и  
материалов»

Студент гр. М01ММ-23  
\_\_\_\_\_ Е.К. Удалова  
« \_\_\_\_ » \_\_\_\_\_ 2025 г.

Руководитель  
к.ф.-м.н., доцент направления  
«Математическое моделирование в  
биомедицине и геофизике»  
\_\_\_\_\_ Т.М. Гамилов  
« \_\_\_\_ » \_\_\_\_\_ 2025 г.

Автономная некоммерческая  
образовательная организация высшего образования  
**«НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»**

Научный центр информационных технологий и искусственного интеллекта  
направление «Математическое моделирование в биомедицине и геофизике»

УТВЕРДИТЬ

Руководитель направления  
«Математическое моделирование в  
биомедицине и геофизике»

к.т.н., д.э.н.

\_\_\_\_\_ М. В. Ширяев  
« \_\_\_\_ » \_\_\_\_\_ 2025 г.

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ**

на выполнение магистерской диссертации  
по направлению подготовки 01.04.02 «Прикладная математика и  
информатика»,  
направленность (профиль) «Математическое моделирование процессов и  
материалов»

**Удалова Екатерина Константиновна**

1. Тема: «Анализ формы и скорости пульсовой волны для диагностики сердечно-сосудистых заболеваний».
2. Цель: повысить эффективность ранней диагностики и скрининга сердечно-сосудистых заболеваний с помощью недорогих и доступных неинвазивных методов.
3. Задачи:
  - провести сбор, предобработку и анализ базы данных реальных пациентов;
  - построить и оценить модели машинного обучения для прогнозирования повышенного сосудистого старения;

- создать синтетическую базу данных ПВ виртуальных пациентов с использованием одномерной модели кровотока;
- разработать алгоритм анализа формы ПВ: определить характерные точки сигнала и на их основе вычислить количественные показатели ПВ, потенциально чувствительные к наличию АР;
- провести последовательное испытание разработанных моделей.

4. Рабочий график (план) выполнения магистерской диссертации:

№	Перечень заданий	Сроки выполнения
1.	Анализ результатов, полученных в результате НИР, проводимых по теме Анализ результатов, полученных в результате НИР, проводимых по теме «Анализ пульсовых волн для диагностики сердечно-сосудистых заболеваний».	18.02.2025 - 18.03.2025
2.	Исследование устойчивости моделей машинного обучения к шумам.	19.03.2025
3.	Финальный анализ и подготовка заключений.	20.05.2025

Дата выдачи: «18» февраля 2025 г.

Руководитель:

\_\_\_\_\_ Т.М. Гамилов

Задание принял к исполнению:

Студент группы М01ММ–23

\_\_\_\_\_ Е.К. Удалова

«18» февраля 2025 г.

## Реферат

Выпускная квалификационная работа, 74 страницы, 27 рисунков, 11 таблиц, 30 источников, 1 приложение.

Ключевые слова: МАШИННОЕ ОБУЧЕНИЕ, АОРТАЛЬНАЯ РЕГУРГИТАЦИЯ, АРТЕРИАЛЬНАЯ ЖЕСТКОСТЬ, ПУЛЬСОВАЯ ВОЛНА, ГЕМОДИНАМИКА, ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ.

Настоящее исследование посвящено разработке и оценке методов прогнозирования аортальной регургитации и сердечно-сосудистых заболеваний, связанных с повышенной сосудистой жесткостью, на основе анализа клинических и синтетических данных с использованием методов машинного обучения. Цель работы – повысить эффективность ранней диагностики и скрининга сердечно-сосудистых заболеваний с помощью недорогих и доступных неинвазивных методов.

В ходе исследования были решены следующие задачи:

- проведен анализ клинических данных пациентов, включая антропометрию, лабораторные показатели и измерения САВИ (сердечно-лодыжечного сосудистого индекса);
- разработаны и протестированы модели машинного обучения для прогнозирования повышенного сосудистого старения;
- создана синтетическая база данных пульсовых волн с использованием одномерной модели гемодинамики для моделирования различных степеней аортальной регургитации;
- разработан алгоритм анализа формы пульсовой волны, позволяющий выделять диагностически значимые параметры;
- проведено тестирование моделей машинного обучения (логистическая регрессия, случайный лес, наивный байесовский классификатор, многослойный перцептрон) на синтетических и реальных данных.

Исследование подтверждает потенциал использования пульсовой волны и методов машинного обучения для неинвазивной диагностики АР и оценки сосудистого старения. Полученные результаты открывают возможности для развития цифровой кардиологии и внедрения новых подходов к профилактике и раннему выявлению сердечно-сосудистых заболеваний.

## The Abstract

Cardiovascular diseases (CVD) are still one of the leading causes of death worldwide. Early and low-cost screening tools are therefore very important. This thesis explores whether the shape and speed of the arterial pulse wave—a pressure wave that travels along the arteries with every heartbeat—can be used to spot early signs of vascular ageing and aortic regurgitation (AR) without expensive imaging.

We combined two kinds of data. First, we analysed clinical records from 600 real patients that included basic body measurements, blood tests and Cardio-Ankle Vascular Index (CAVI) readings, a standard marker of arterial stiffness. Second, we created a large synthetic database of pulse waves for 1 458 “virtual” older adults using a one-dimensional blood-flow model. By tweaking key heart and vessel parameters, we simulated healthy cases and different grades of AR.

From each pulse wave we automatically extracted clear, physiologically meaningful points—such as the main systolic peak and the dicrotic notch—and turned them into time and amplitude features. Four machine-learning algorithms (logistic regression, naïve Bayes, random forest and a small multilayer perceptron) were trained to predict abnormal arterial stiffness, and classify pulse waves as healthy or AR-like.

On the synthetic data all models reached high accuracy (up to 99 %). When tested on real patients, performance dropped, but the neural network kept the best balance between sensitivity and specificity. Adding noise during training made the models slightly more robust.

The results show that pulse-wave analysis, boosted by machine learning and supported by realistic simulations, can become a simple, non-invasive aid for early CVD screening. Future work should enlarge the clinical dataset and refine the models for use in primary care devices and wearable sensors.

## Сокращения, обозначения, термины и определения

В настоящей работе применяют следующие термины с соответствующими определениями:

- $\mathbb{R}$  – множество вещественных чисел.

В настоящей работе применяют следующие сокращения и обозначения:

- ССЗ – сердечно-сосудистые заболевания;
- АР – аортальная регургитация;
- ПВ – пульсовая волна;
- ЧСС – частота сердечных сокращений;
- УО – ударный объем;
- ЧСС – частота сердечных сокращений;
- СВ – сердечный выброс;
- САVI – Cardio-Ankle Vascular Index / сердечно-лодыжечный сосудистый индекс;
- ЛЖ – левый желудочек;
- ЛР – логистическая регрессия;
- НБ – наивный байесовский классификатор;
- MLP – Multilayer Perceptron / многослойный перцептрон;
- БД – база данных.

## Оглавление

Введение.....	10
1 ОБЗОР ЛИТЕРАТУРЫ .....	13
1.1 Сердечно-сосудистая система.....	13
1.2 Артериальная пульсовая волна.....	16
1.3 Сердечно-лодыжечный сосудистый индекс.....	17
1.4 Аортальная регургитация.....	18
1.5 Одномерная модель кровотока .....	20
1.6 Синтетические базы данных .....	21
1.7 Методы машинного обучения .....	22
2 ИССЛЕДОВАНИЕ АРТЕРИАЛЬНОЙ ЖЕСТКОСТИ .....	24
3 МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ .....	28
3.1 Логистическая регрессия.....	28
3.2 Случайный лес.....	30
3.3 Наивный байесовский классификатор.....	33
3.4 Многослойный перцептрон.....	34
3.5 Метрики .....	37
4 ОДНОМЕРНАЯ МОДЕЛЬ КРОВОТОКА .....	41
5 АНАЛИЗ ФОРМЫ ПУЛЬСОВОЙ ВОЛНЫ .....	44
6 РЕЗУЛЬТАТЫ.....	47
6.1 Модели для определения сосудистой жесткости.....	47
6.2 Расчет вероятности ускоренного сосудистого старения.....	54
6.3 База данных виртуальной популяции .....	56
6.4 Модели машинного обучения для классификации пульсовой волны	59
6.5 Исследование влияния шума на метрики моделей машинного обучения для классификации пульсовой волны .....	61
6.5.1 Логистическая регрессия .....	61
6.5.2 Наивный байесовский классификатор.....	62
6.5.3 Случайный лес.....	62
6.5.4 Многослойный перцептрон .....	62
Заключение.....	64
Список использованных источников.....	67



Приложение А (обязательное) Антропометрические параметры исследуемой группы пациентов.....	71
--	----

## Введение

Сердечно-сосудистые заболевания (ССЗ) остаются одной из главных причин смертности и инвалидизации населения во всём мире. Одним из ключевых факторов риска развития ССЗ является сосудистое старение – структурнофункциональное изменение артерий, связанное с возрастом, гипертензией, атеросклерозом и другими патологиями.

Важным маркером сосудистого старения является сердечно-лодыжечный сосудистый индекс (Cardio-Ankle Vascular Index, CAVI), который отражает жесткость артерий независимо от уровня артериального давления. Повышение CAVI коррелирует с риском развития различных ССЗ и их осложнений.

Еще одной важной проблемой кардиологии является аортальная регургитация (АР) – это патология клапана сердца, при которой происходит обратный ток крови из аорты в левый желудочек во время диастолы из-за неполного смыкания или деформации створок клапана.

При АР у людей появляются жалобы на отдышку при небольшой физической нагрузке, быструю утомляемость, учащенное сердцебиение и т. д., а длительная объемная перегрузка желудочка приводит к систолической дисфункции и может вызывать жизнеугрожающие осложнения: фибрилляцию предсердий, желудочковую аритмию, кардиогенный шок. На ранних же стадиях АР часто протекает бессимптомно, что затрудняет своевременную диагностику и лечение.

Высокая распространенность традиционных факторов риска ССЗ среди внешне здоровых пациентов без ССЗ создаёт предпосылки для будущего роста заболеваемости.

Традиционные методы диагностики АР (ЭхоКГ, МРТ) требуют дорогостоящего оборудования и высокой квалификации врачей, поэтому разработка неинвазивных методов скрининга на основе машинного обучения представляет значительный интерес.

Одним из перспективных направлений является анализ пульсовых волн (ПВ). ПВ содержит значительный объем информации о состоянии сердечно-сосудистой системы. На ее характеристики влияют сердечные факторы, включая частоту сердечных сокращений (ЧСС) и ударный объем (УО), которые определяют продолжительность и морфологию ПВ, а также сосудистые факторы, такие как артериальная жесткость и волновые отражения. Благодаря этому анализ ПВ позволяет оценивать различные физиологические параметры, имеющие важное значение для диагностики, мониторинга состояния пациентов и принятия клинических решений.

Несмотря на диагностический потенциал анализа ПВ, его широкое внедрение в клиническую практику сталкивается с существенными ограничениями, связанными с доступностью данных. Основными проблемами являются: строгие требования к защите персональных данных пациентов, сложность стандартизации методов измерения между различными медицинскими учреждениями, а также высокая стоимость проведения масштабных клинических исследований. В этом контексте особую ценность приобретают альтернативные подходы, включая использование синтетических данных, генерируемых математическими моделями гемодинамики. Разработка методов анализа таких данных представляет собой перспективное направление в цифровой кардиологии.

Особенностью и научной новизной этой работы является комплексное сочетание современных технологий анализа сердечно-сосудистых параметров: математического моделирования гемодинамики для генерации синтетических данных, машинного обучения для выявления значимых диагностических маркеров, а также использования неинвазивных клинических измерений. Комбинация этих методов позволяет разрабатывать универсальные алгоритмы оценки состояния сердечно-сосудистой системы, которые могут быть применены как для врачебной диагностики, так и для профилактического мониторинга в повседневной жизни.

Целью данного исследования является разработка методов прогнозирования сосудистого старения и АР на основе анализа клинических и синтетических данных с использованием машинного обучения.

Для достижения этой цели поставлены следующие задачи:

1. Провести сбор, предобработку и анализ базы данных реальных пациентов, включающей антропометрические сведения и результаты лабораторных анализов крови, а также САVI.

2. Построить и оценить модели машинного обучения для прогнозирования повышенного сосудистого старения.

3. Создать синтетическую базу данных ПВ виртуальных пациентов с использованием одномерной модели кровотока, варьируя ключевые физиологические параметры для имитации разных степеней АР.

4. Разработать алгоритм анализа формы ПВ: определить характерные точки сигнала и на их основе вычислить количественные показатели ПВ, потенциально чувствительные к наличию АР. Выявить, какие из этих показателей существенно различаются при патологии клапана.

5. Провести последовательное испытание моделей: во-первых, на синтетических данных и во-вторых, на реальных данных пациентов. Сравнить качество классификации разных алгоритмов и их устойчивость к шумам.

# 1 ОБЗОР ЛИТЕРАТУРЫ

## 1.1 Сердечно-сосудистая система

Сердечно-сосудистая система обеспечивает циркуляцию крови в организме и состоит из сердца и кровеносных сосудов. Человеческое сердце располагается в грудной полости, по срединной линии между лёгкими, в анатомическом пространстве, называемом средостением. К его верхней поверхности (основанию сердца) прикрепляются крупные вены - верхняя и нижняя полая - и крупные артерии - аорта и лёгочной ствол. Само сердце состоит из четырёх камер: двух предсердий и двух желудочков (рисунок 1.1). Правое предсердие принимает венозную (дезоксигенированную) кровь из большого круга кровообращения, а левое предсердие – артериальную (оксигенированную) кровь из лёгких [1]. Сокращаясь, предсердия проталкивают кровь в нижележащие камеры – правый и левый желудочки. Правый желудочек выбрасывает кровь в лёгкие, а левый – главный насос, проталкивающий кровь по всему организму. С каждой систолой кровь поступает и по коронарным артериям. Левая коронарная артерия кровоснабжает левую сторону сердца, правая – правое предсердие, части обоих желудочков и элементы проводящей системы.

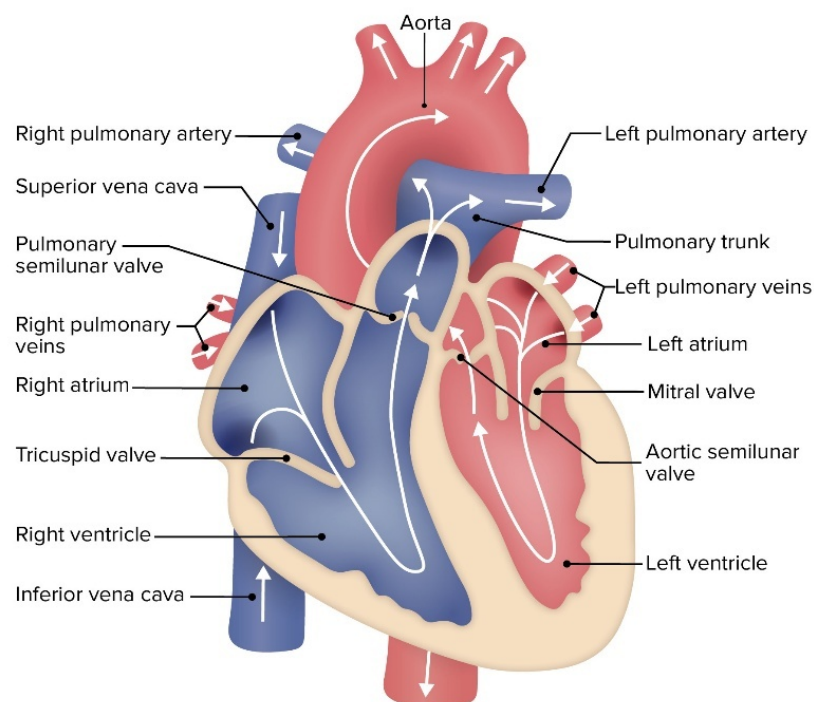


Рисунок 1.1 – Строение сердца человека

В системе кровообращения выделяют два взаимосвязанных круга – лёгочный и системный. Лёгочный круг переносит кровь к лёгким и обратно, где она насыщается кислородом и отдаёт углекислый газ. Системный круг доставляет артериальную кровь почти ко всем тканям организма и возвращает к сердцу венозную кровь, обогащённую  $\text{CO}_2$ , для последующей отправки в лёгкие.

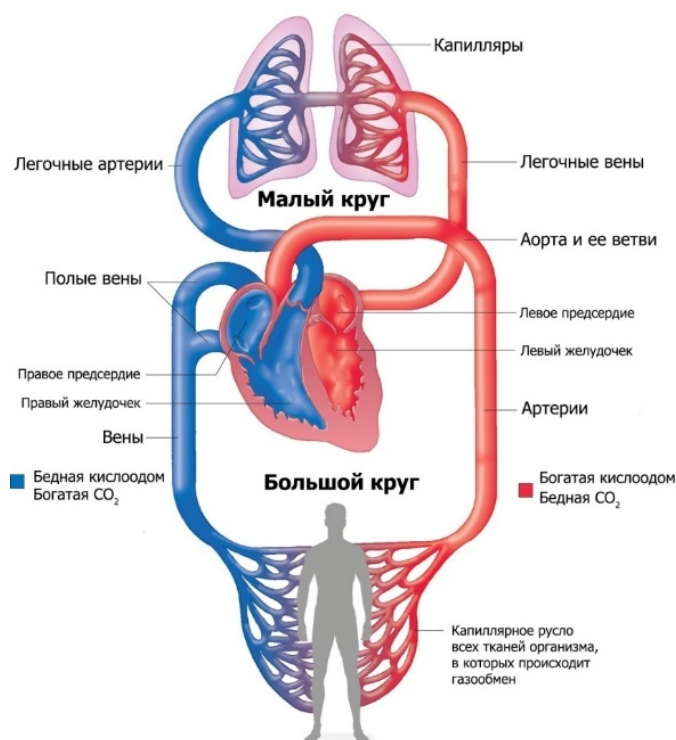


Рисунок 1.2 – Схема кровообращения человека

Интервал от начала сокращения предсердий до окончания расслабления желудочков называют сердечным циклом. Период сокращения камер, когда кровь выбрасывается в сосуды, называется систолой; период их заполнения кровью – диастолой. Всего выделяют четыре фазы сердечного цикла [1]:

1. Общая диастола: и предсердия, и желудочки расслаблены. Кровь свободно поступает в правое предсердие через верхнюю и нижнюю полые вены и в левое предсердие – через четыре лёгочные вены.

2. Систола предсердий: мышцы предсердий сокращаются, давление в них растёт, и кровь поступает в желудочки. В этот момент объём крови в каждом желудочке у взрослого в покое  $\approx 130$  мл – это конечный диастолический объём (preload).

3. Систола желудочков: давление в желудочках быстро увеличивается, створки атриовентрикулярных клапанов закрываются, и начинается фаза изгнания, когда кровь выталкивается в аорту и лёгочный ствол. Количество этой крови называется ударным объёмом (УО).

4. Диастола желудочков: давление падает, полулунные клапаны закрываются, и камеры начинают заполняться кровью, подготавливаясь к новому циклу.

5. Сердечный выброс (СВ) – это объём крови, выбрасываемый каждым желудочком за одну минуту. Он рассчитывается по формуле  $СВ = ЧСС \times УО$ , где ЧСС – частота сердечных сокращений (уд/мин).

## 1.2 Артериальная пульсовая волна

Артериальная пульсовая волна представляет собой распространение колебаний давления и кровотока, возникающих при каждом ударе сердца. При выбросе крови в аорту, волна давления распространяется вдоль артериального русла. Анализ формы этой волны позволяет оценить важные параметры сердечно-сосудистой системы [2].

В проксимальном отделе аорты нормальная форма волны давления характеризуется быстрым повышением систолического давления до пика, последующим спадом с формированием инцизуры (дикротической выемки, отражающей закрытие аортального клапана), а затем медленным снижением давления в диастолическую фазу.

На морфологию ПВ влияют несколько ключевых факторов: сократительная функция левого желудочка, скорость выброса крови, эластичность артериальной стенки и явление отражения волн от периферических сосудов. По мере распространения импульса по артерии происходит уменьшение диаметра артерий, увеличение мышечного компонента сосудистой стенки, повышение периферического сопротивления и формирование отраженных волн.

В эластичных артериях отраженная волна возвращается в диастолу, повышая диастолическое давление. При увеличении жесткости артерий (например, при старении или гипертонии) скорость ПВ увеличивается, и



отражения возвращаются раньше, часто в конце систолы, что приводит к повышению систолического и пульсового давления [3].

Ключевыми параметрами анализа ПВ являются индекс аугментации (соотношение прямой и отраженной волн), время возврата отраженной волны и общая форма кривой артериального давления.

### **1.3 Сердечно-лодыжечный сосудистый индекс**

Важным показателем артериальной жесткости является кардио-лодыжечный сосудистый индекс (CAVI). Этот относительно новый параметр отражает общую жесткость артерий от аорты до лодыжки и разработан таким образом, чтобы минимизировать зависимость от уровня артериального давления во время измерения – в отличие от скорости ПВ, на которую давление оказывает значительное влияние [4]. CAVI рассчитывается с использованием скорости ПВ между аортой и лодыжкой, кровяного давления и константы, что дает безразмерное значение параметра жесткости. Чем выше показатель CAVI, тем более выражена артериальная жесткость. На практике CAVI измеряется с помощью устройств, которые надувают манжеты регистрируют время распространения ПВ с помощью манжет, накладываемых на лодыжки и руки, синхронизируя данные с ЭКГ.

Преимущество CAVI заключается в том, что он учитывает жесткость центральной и периферической артерий и сохраняет стабильность даже при значительных колебаниях артериального давления.

С клинической точки зрения CAVI рассматривается как предиктор сердечно-сосудистых исходов. Крупное японское когортное исследование (CAVI-J study) продемонстрировало, что CAVI независимо предсказывает сердечно-сосудистые события у пациентов из группы риска [5]. В этом исследовании было выявлено, что повышенные исходные значения CAVI коррелируют с увеличением частоты неблагоприятных событий даже после учета традиционных факторов риска. Другое исследование, включившее

пациентов без явных ССЗ, показало, что  $CAVI > 9$  ассоциирован с увеличением риска серьезных осложнений примерно в 2,3 раза. Это означает, что у таких пациентов вероятность развития событий более чем вдвое выше по сравнению с теми, у кого жесткость артерий ниже [6].

#### 1.4 Аортальная регургитация

АР характеризуется диастолическим обратным током крови из аорты в левый желудочек (ЛЖ) из-за неполного смыкания аортального клапана. В результате возникает объемная перегрузка ЛЖ, поскольку УО увеличивается за счет прямого выброса и дополнительного объема регургитации [7]. По мере прогрессирования заболевания возрастает конечный диастолический объем ЛЖ, увеличивается УО, а также наблюдается повышение систолического артериального давления на фоне снижения диастолического, что ведет к росту пульсового давления [7]. Однако со временем продолжающаяся перегрузка приводит к повышению конечного диастолического давления в ЛЖ и, в конечном итоге, к систолической дисфункции со снижением фракции выброса [7].

АР является относительно распространенным пороком сердца, и его частота увеличивается с возрастом. Признаки недостаточности аортального клапана (легкая АР) часто наблюдаются у пожилых – например, эпидемиологические исследования сообщают о легкой или умеренной АР примерно у 8–13% лиц старше 65 лет [8]. Однако клинически значимая АР (умеренной или тяжелой степени) встречается реже, поражая примерно 4–5% пожилых людей [9].

Исследования в странах с высоким уровнем дохода показывают, что АР занимает третье место по распространенности среди приобретенных пороков сердца у пожилых, уступая лишь аортальному стенозу и митральной регургитации. Этот диагноз выявляется примерно у 1,1–1,8% людей старше 60 лет [10]. Данные Фрамингемского исследования подтверждают, что АР в той

или иной степени присутствует у 4,9% участников, что указывает на значительную долю бессимптомных случаев. С учетом старения населения и роста продолжительности жизни ожидается дальнейшее увеличение распространенности этого заболевания [11].

Хроническая АР характеризуется длительным бессимптомным течением, при котором компенсаторные механизмы длительное время поддерживают адекватный сердечный выброс. Симптомы, такие как одышка при нагрузке, утомляемость, сердцебиение или стенокардия, обычно появляются лишь на стадии декомпенсации ЛЖ или при значительной дилатации его полости [7].

К моменту возникновения выраженной симптоматики или дисфункции ЛЖ заболевание часто достигает поздних стадий, что осложняет выбор оптимального времени для хирургического вмешательства – слишком долгое ожидание может привести к необратимому поражению миокарда, в то время как у пациентов с компенсированной сердечной недостаточностью операция на ранней стадии может оказаться преждевременной. Современные рекомендации обычно советуют протезирование аортального клапана при тяжелой АР при появлении симптомов или при признаках систолической дисфункции ЛЖ либо выраженной дилатации (например, конечный систолический размер ЛЖ > 50 мм или фракция выброса < 50%) [12]. Однако новые данные свидетельствуют, что некоторым пациентам может быть полезно более раннее вмешательство, даже до появления традиционных показаний [12].

Применение современных технологий, таких как машинное обучение, позволило идентифицировать пациентов из группы высокого риска еще до появления симптомов или снижения фракции выброса [12]. Тем не менее, определение оптимальных сроков операции остается сложной задачей, требующей индивидуального подхода.

В целом, медленное прогрессирование хронической АР в сочетании с трудностями диагностики приводит к тому, что многие пациенты

направляются на операцию лишь на поздних стадиях. Это стимулирует исследования по улучшению диагностических критериев и предикторов риска для более раннего выявления пациентов с высоким риском [12].

## 1.5 Одномерная модель кровотока

Одномерные модели артериального кровообращения основаны на фундаментальных принципах гидродинамики и применяются в условиях, характерных для крупных сосудов. В рамках этих моделей кровотоков вдоль каждого артериального сегмента рассматривается как непрерывный процесс, описываемый усредненными по поперечному сечению параметрами, такими как скорость потока, давление и площадь поперечного сечения, которые изменяются в зависимости от осевой координаты и времени.

Интегрирование уравнений Навье–Стокса по поперечному сечению с учетом осесимметричности и тонких упругих стенок сосуда позволяет получить систему нелинейных дифференциальных уравнений в частных. Эта система включает уравнения сохранения массы (непрерывности) и импульса для упругой трубки [13]. Данные уравнения аналогичны уравнениям несжимаемого потока и описывают волновые явления, где пульсовая волна представляет собой движущееся возмущение.

Классическая формулировка включает в себя слагаемые для конвективного ускорения, силы градиента давления и упругости стенок, которая часто выражается через зависимость давления от площади сосуда [13]. Пример такой системы включает уравнение неразрывности:

$$\frac{\partial A}{\partial t} + \frac{\partial(Au)}{\partial x} = 0, \quad (1.1)$$

где  $A$  – площадь поперечного сечения;

$u$  – скорость потока.

А также уравнение импульса:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + \frac{1}{\rho} \frac{\partial P}{\partial x} = - \frac{8\pi\mu u}{\rho A}, \quad (1.2)$$

учитывающее вязкостные потери [13].

Эти уравнения дополняются реологическим соотношением, связывающим площадь поперечного сечения  $A$  с местным давлением  $P$ , например, через закон упругой трубки.

Исследования показали, что одномерные модели кровотока способны фиксировать основные гемодинамические явления, такие как пик систолического давления, диастолическая выемка и влияние артериального давления на форму ПВ [14].

В работе [15] авторы провели оригинальное сравнительное исследование одномерных моделей артерий и обнаружили, что многочисленные численные схемы, применяемые к одномерным формулировкам, точно воспроизводят основные характеристики волн давления и потока в крупных артериях [15].

Сопоставление теоретических результатов, численного моделирования и экспериментальных данных подтвердило, что одномерные модели, несмотря на приближенный характер, адекватно отражают основные физические процессы распространения и отражения ПВ в артериальной системе [15].

## 1.6 Синтетические базы данных

Одним из приложений одномерного моделирования артериальной системы является создание виртуальных когорт – больших коллекций моделируемых пациентов – для изучения сердечно-сосудистых процессов. Использование численных моделей позволяет проводить эксперименты, которые были бы непрактичными или неэтичными в реальности, обеспечивая при этом данные с известными параметрами.

В исследовании [13] авторы создали синтетический набор данных ПВ для изучения сосудистого старения, сфокусировавшись на возрастной группе 60–70 лет [13]. В их работе использовалась связанная модель: 1-D модель системного кровообращения в сочетании с параметрической моделью сердца [13].

Особый интерес представляет применение машинного обучения к таким синтетическим данным. В упомянутом исследовании виртуальная когорта послужила основой для обучения алгоритма прогнозирования САVI по легко измеряемым параметрам. Последующая проверка на реальных пациентах показала высокую точность прогноза со средней ошибкой около 6,5%, что подтверждает практическую ценность подхода [13].

Ключевое преимущество подхода – возможность создания детализированных смоделированных наборов данных, где точно известны артериальная жесткость, артериальное давление и ПВ. На таких данных можно обучать алгоритмы и затем применять их в клинике.

Кроме того, искусственные данные позволяют моделировать редкие или экстремальные сценарии (например, нестандартные комбинации параметров), что важно для проверки устойчивости диагностических методов.

Этот пример демонстрирует потенциал виртуальных популяций в разработке и проверке диагностических инструментов.

## **1.7 Методы машинного обучения**

Сердечно-сосудистая диагностика активно использует методы машинного обучения, начиная с классических подходов, таких как логистическая регрессия (ЛР), и заканчивая более сложными алгоритмами, включая нейронные сети. Эти методы помогают решать разнообразные задачи, от прогнозирования исходов у пациентов до классификации тяжести заболеваний на основе медицинских изображений или сигналов, а также выделения специфических подгрупп патологий.

ЛР – простая интерпретируемая модель, которая часто используется в медицине для бинарной классификации, например, в шкалах риска. Однако она плохо учитывает нелинейные взаимодействия без явного задания.

В исследовании 2023 года у пациентов с тяжелой АР модель на основе случайного леса предсказывала 1- и 2-летнюю выживаемость, используя такие факторы как возраст, конечно-систолический размер ЛЖ, диастолическое давление [12]. Эта модель превзошла более простые подходы [12].

Наивный байесовский классификатор (НБ) реже используется, но эффективен при малых данных. Например, в задаче классификации клапанных пороков НБ в сочетании с SVM достиг точности 88.5% [16]. В других работах НБ показывал сопоставимую с более сложными методами производительность и был устойчив к переобучению [16].

Многослойный перцептрон (MLP) обучает нелинейные зависимости и применяется в анализе ЭКГ, классификации шумов. В исследовании кровотока в виллизиевом круге MLP показал точность 79% в условиях зашумленных данных, превзойдя глубокие архитектуры [12].

## 2 ИССЛЕДОВАНИЕ АРТЕРИАЛЬНОЙ ЖЕСТКОСТИ

В исследование были включены 600 пациентов (43% мужчин и 57% женщин, средний возраст  $39,8 \pm 18,3$  года), проходивших ежегодное клиническое обследование в Университетской клинической больнице № 4 им. Сеченова. Критериями включения в исследование являлись: возраст старше 18 лет, отсутствие тяжелых хронических заболеваний и наличие письменного информированного согласия.

Всем участникам проводили антропометрические измерения, которые включали такие параметры как окружность талии и индекс массы тела, а также оценивали биохимические показатели крови. Измеряли уровни общего холестерина, липопротеидов низкой и высокой плотности, триглицеридов, глюкозы и креатинина. На основании этих данных рассчитывали дополнительные индексы: продукт накопления липидов, индекс висцерального ожирения и процентное содержание жира в организме. Средние значения всех перечисленных выше показателей в разрезе возрастных групп вынесены в Приложение А.

Артериальная жесткость измерялась сфигмоманометрическим методом на аппарате VaSera VS-1000 (Fukuda Denshi, Япония, 2010). Перед исследованием пациенты соблюдали стандартные условия подготовки:

- измерение проводилось утром, чтобы минимизировать влияние суточных колебаний сосудистого тонуса;
- во время измерения пациенты находились в положении лежа на спине, в состоянии физического и эмоционального покоя на протяжении не менее чем за 10 минут до начала процедуры;
- исключалось употребление алкоголя и напитков, содержащих кофеин, а также курение любых содержащих никотин устройств;
- измерения проводились в комфортных условиях при комнатной температуре, чтобы избежать повышения или понижения сосудистого сопротивления за счёт перепадов температуры.



Для интерпретации результатов использовали универсальные ( $CAVI \geq 9,0$ ) и возраст-зависимые пороговые значения, установленные российским экспертным консенсусом (2016):

Таблица 2.1 – Верхние границы CAVI для определения жесткости сосудов в зависимости от возраста

Возраст, год	<19	21-30	31-40	41-50	51-60	61-70	>70
CAVI	6.7	7.2	7.4	7.55	8.0	8.5	9.8

Проводились измерения отдельно левой и правой частей тела. Таким образом для каждого пациента получилось два показателя CAVI: RCAVI (правая часть тела) и LCAVI (левая часть тела). Жесткость сосудов считалась повышенной, если хоть один из этих показателей превышал границы из таблицы 2.1.

В таблице 2.2 представлены средние значения рассматриваемых параметров пациентов, участвовавших в исследовании. Пациенты разделены на условно здоровых и больных с помощью модели с возраст-зависимыми пороговыми значениями CAVI.

Таблица 2.2 – Клинические и демографические характеристики пациентов

Параметр	Все пациенты	Здоровые пациенты	Пациенты с повышенным CAVI	p-value
Возраст, год	39.8 ± 18.3	35.02 ± 16.9	51.11 ± 16.6	<0.001
Рост, м	1.71 ± 0.1	1.71 ± 0.09	1.7 ± 0.09	0.001
Вес, кг	77.1 ± 19.1	75.91 ± 19.9	79.88 ± 16.96	0.001
ИМТ, кг/м <sup>2</sup>	26.4 ± 5.9	25.91 ± 5.87	27.55 ± 5.73	<0.001
Окружность талии, см	80.79 ± 15.51	80.12 ± 15.66	83.47 ± 14.68	0.025
Окружность бедер, см	100.56 ± 9.44	100.47 ± 9.8	100.91 ± 7.89	0.06
Систолическое давление, мм рт. ст.	128.4 ± 15.3	126.67 ± 13.54	132.6 ± 18.3	<0.001

Продолжение таблицы 2.2

Диастолическое давление, мм рт. ст.	79.19 ± 9.32	78.07 ± 8.52	81.83 ± 10.53	0.001
Холестерин, ммоль/л	4.8 ± 1.2	4.61 ± 1.04	5.27 ± 1.43	<0.001
Глюкоза, ммоль/л	5.3 ± 1.89	5.05 ± 1.55	5.95 ± 2.47	<0.001
Креатинин, нмоль/л	84.13 ± 18.06	82.57 ± 16.36	88.53 ± 21.65	0.003
Курение, %	29.32	30.9	25.8	0.044
Гипертензия, %	31.47	20.9	55.4	<0.001
Триглицериды, ммоль/л	1.82 ± 1.1	1.85 ± 1.17	1.77 ± 1.19	0,791

Основываясь на данных таблицы 2.2 можно сделать вывод, что люди с повышенным САVI в среднем старше, обладают более высокими антропометрическими показателями, а также демонстрируют менее благоприятный кардиометаболический профиль: у них выше уровень артериального давления, холестерина, глюкозы и креатинина, а также чаще встречается гипертензия.

Также наблюдается парадокс, связанный с процентом курящих людей в обеих категориях – курение реже встречается в группе с повышенным САVI [17]. В данном случае это объяснимо спецификой исследуемой когорты – к категории курящих относились пациенты, использующие различные электронные системы доставки никотина, а их возраст обычно не превышает 35 лет, поэтому эффекты старения сосудов у них не наблюдались массово.

Статистический анализ и корреляционный анализ данных проводили с использованием языка программирования python (библиотеки pandas, scipy, numpy, matplotlib и др.).

Коэффициент корреляции Спирмена был рассчитан для каждой пары переменных, включая биномиальные переменные (курение, гипертензия, старение сосудов и т.д.).

Коэффициент корреляции Спирмена – это статистический показатель, показывающий степени монотонной связи между двумя переменными [18]. Этот коэффициент рассчитывается по формуле:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (2.1)$$

где  $d_i$  – разница между рангами соответствующих значений двух переменных;  
 $n$  – количество наблюдений.

Разработка прогностической модели осуществлялась с использованием методов машинного обучения на основе клинической базы данных пациентов. В качестве основного алгоритма была выбрана ЛР, что обусловлено возможностью получения четкой математической формулы для расчета вероятности развития артериальной жесткости.

Для более точной индивидуальной оценки был выполнен переход к непрерывной шкале нормативных значений САVI путем математической аппроксимации представленных в таблице 2.1 возрастных границ.

### 3 МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

В данной работе использовались классические и современные методы машинного обучения, позволяющие решать задачи классификации с различной степенью сложности. Основное внимание было уделено четырем ключевым алгоритмам.

#### 3.1 Логистическая регрессия

Логистическая регрессия представляет собой один из базовых методов классификации, широко применяемый для решения задач бинарного разделения объектов на два класса. Несмотря на свою простоту, данный метод обладает высокой интерпретируемостью, хорошей устойчивостью к переобучению (особенно при введении регуляризации), а также эффективностью как с вычислительной, так и с практической точек зрения. В данной работе логистическая регрессия использовалась в качестве базовой модели для диагностики аортальной регургитации по признакам ПВ.

Основная идея логистической регрессии заключается в том, чтобы смоделировать вероятность принадлежности объекта к одному из классов как функцию от линейной комбинации признаков. Модель вычисляет линейную функцию от входных признаков (плюс смещение) [19]:

$$z = w_0 + w_1x_1 + \dots + w_nx_n \quad (3.1)$$

где  $w$  – веса;

$x$  – признаки;

$w_0$  – смещение.

Результат  $z$  преобразуется в вероятность принадлежности к положительному классу с помощью сигмоидальной функции [20]:

$$P(y = 1) = \frac{1}{1 + e^{-z}}. \quad (3.2)$$

Таким образом, выход модели находится в интервале  $[0,1]$  и может интерпретироваться как вероятность. Если  $P(y = 1) \geq 0.5$  (по умолчанию), объект относят к классу 1, иначе – к классу 0 [20].

Обучение логистической регрессии сводится к подбору таких весов  $w$ , которые максимизируют правдоподобие наблюдаемых данных. При условии независимости наблюдений функция правдоподобия для  $m$  экземпляров имеет вид:

$$\mathcal{L}(w) = \prod_{i=1}^m P(y_i|x_i) = \prod_{i=1}^m \sigma(z_i)^{y_i} (1 - \sigma(z_i))^{1-y_i}, \quad (3.3)$$

где  $\sigma(z_i)^{y_i}$  – значение сигмоиды для объекта  $i$ .

Так как на практике произведение является неудобным, используют логарифм правдоподобия, благодаря которому произведение становится суммой:

$$\ell(w) = \sum_{i=1}^m [y_i \ln \sigma(z_i) + (1 - y_i) \ln(1 - \sigma(z_i))] \quad (3.4)$$

Максимизация этой функции производится численно – с использованием градиентного спуска. Производная логарифма правдоподобия по параметрам имеет аналитическую форму, что позволяет эффективно вычислять шаги оптимизации.

Для предотвращения переобучения, особенно в случае большого количества признаков, к функции потерь добавляют регуляризующее слагаемое. Часто применяют L1 или L2 регуляризацию. В случае первой добавляется  $\lambda \|w\|_1$  (сумма модулей весов), а во второй добавляется  $\lambda \|w\|^2$ .

Одним из ключевых достоинств логистической регрессии является интерпретируемость её параметров. Каждый коэффициент  $w_j$  отражает

влияние соответствующего признака  $x_j$  на логарифм отношения шансов принадлежности объекта к положительному классу:

$$\log\left(\frac{P(y = 1|x)}{P(y = 0|x)}\right) = z = w_0 + w_1x_1 + \dots + w_nx_n \quad (3.5)$$

Таким образом, коэффициент  $w_j$  задаёт изменение логарифма отношения шансов при увеличении признака  $x_j$  на единицу.

Логистическая регрессия быстро обучается, требует мало вычислительных ресурсов, хорошо работает с линейно разделимыми и разреженными данными. Модель легко реализуется, устойчива численно и позволяет интерпретировать вклад каждого признака в итоговое решение. Однако у нее есть и ограничения. ЛВ предполагает линейную зависимость между признаками и логарифмом шансов, плохо переносит мультиколлинеарность (когда есть линейная зависимость между признаками) и не справляется с ярко выраженной нелинейностью без дополнительной обработки. Также логистическая регрессия чувствительна к выбросам и может давать смещённые результаты при несбалансированных классах [21].

### 3.2 Случайный лес

Случайный лес (Random Forest) – это ансамблевый метод машинного обучения, основанный на объединении большого числа решающих деревьев. Он был включён в работу как более гибкий алгоритм, способный выявлять сложные, в том числе нелинейные зависимости в данных, а также устойчиво работать при наличии шумов и коррелированных признаков. Случайный лес широко используется как для задач классификации, так и регрессии.

Метод основан на построении ансамбля  $M$  случайно построенных деревьев каждое из которых обучается на случайной подвыборке обучающей выборки. Этот приём известен как бутстрэппинг (bootstrap sampling) и

предполагает случайную выборку с возвращением. Помимо случайного выбора наблюдений, на каждом шаге построения дерева случайно отбирается подмножество признаков, среди которых выбирается оптимальное разбиение. Такая двойная случайность – в выборке данных и выборе признаков – способствует снижению корреляции между деревьями и повышению устойчивости модели.

Обозначим через  $x \in \mathbb{R}^d$  входной вектор признаков, через  $D^n$  – обучающую выборку объёма  $n$ , а через  $\Theta_j$  – случайный вектор, определяющий выбор подвыборки и признаков при обучении  $j$ -го дерева [22]. Тогда предсказание этого дерева в точке  $x$  обозначается как  $m_n(x; \Theta_j, D_n)$ . Итоговое предсказание ансамбля из  $M$  деревьев вычисляется усреднением:

$$m_{M,n}(x) = \frac{1}{M} \sum_{j=1}^M m_n(x; \Theta_j, D_n). \quad (3.6)$$

При  $M \rightarrow \infty$  оценка приближается к математическому ожиданию по случайности построения деревьев:

$$m_{\infty,n}(x) = E_{\Theta}[m_n(x; \Theta_j, D_n)]. \quad (3.7)$$

Схема работы алгоритма случайного леса представлена на рисунке 3.1. Каждый элемент схемы иллюстрирует процесс построения дерева на своей случайной подвыборке и последующего агрегирования предсказаний с использованием механизма голосования или усреднения, в зависимости от постановки задачи.

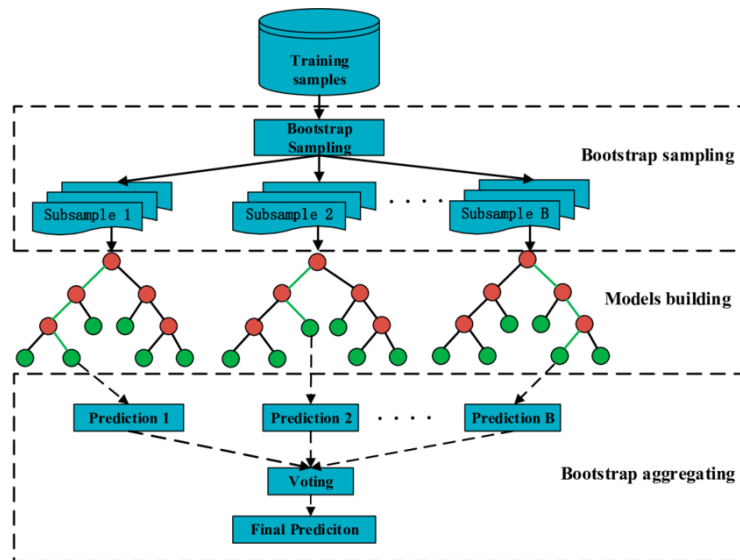


Рисунок 3.1 – Схема работы алгоритма случайного леса

С точки зрения обучения, случайный лес не требует тонкой настройки гиперпараметров и поэтому хорошо работает без глубокой настройки. Основными настраиваемыми параметрами являются количество деревьев в ансамбле, глубина деревьев, минимальное количество объектов в листе, а также количество признаков, отбираемых при каждом разбиении.

Случайный лес обладает рядом практических достоинств. Он устойчив к переобучению, особенно при большом количестве деревьев, и способен моделировать сложные зависимости [23], включая взаимодействия между признаками. Метод хорошо работает как с числовыми, так и с категориальными переменными, не требует масштабирования признаков и способен обрабатывать пропущенные значения. Благодаря бутстрэппингу и случайному отбору признаков, случайный лес устойчив к шуму в данных и может сохранять высокое качество предсказания даже при наличии нерелевантных переменных.

Дополнительным преимуществом является возможность оценки важности признаков: по степени влияния признака на снижение ошибки (например, по среднему уменьшению энтропии или точности при его перестановке).



### 3.3 Наивный байесовский классификатор

Для сравнения результатов использовался наивный байесовский классификатор, основанный на теореме Байеса:

$$P(y|x) = \frac{P(X|x) \cdot P(y)}{P(X)}, \quad (3.8)$$

где  $P(y|x)$  – апостериорная вероятность класса  $y$  при данных признаках  $X$ ;

$P(X|y)$  – вероятность признаков  $X$  при условии класса  $y$ ;

$P(y)$  – априорная вероятность класса  $y$ ;

$P(X)$  – нормирующая константа.

Все признаки  $x_1, x_2, \dots, x_n$  условно независимы при данном классе:

$$P(X|y) = P(x_1|y) \cdot P(x_2|y) \cdot \dots \cdot P(x_n|y). \quad (3.9)$$

Для объекта  $X$  выбирается класс  $y$ , максимизирующий [24]:

$$y = \arg \max P(y) \cdot \prod_{i=1}^n P(x_i|y). \quad (3.10)$$

НБ удобен, потому что очень быстро обучается и почти не требует настройки. Он хорошо работает на небольших выборках и относительно устойчив к шуму, что полезно в задачах, где качество данных неидеальное – как, например, при работе с пульсовыми волнами. В этой работе он показал достаточно стабильные результаты даже без сложных параметров или предварительной оптимизации.

Основной минус НБ в предположении, что все признаки независимы друг от друга. В реальности это почти никогда не так, особенно при анализе физиологических сигналов, где параметры пульсовой волны часто коррелируют. Из-за этого модель может ошибаться в оценке вероятностей и терять в точности. Кроме того, выбор предполагаемого распределения признаков (чаще всего – нормального) может не соответствовать реальному

характеру данных. Например, параметры ПВ не всегда следуют нормальному распределению, и это также ограничивает точность предсказания [24].

Тем не менее, несмотря на эти ограничения, наивный байесовский классификатор остаётся мощным инструментом в условиях, когда требуется быстрое и устойчивое решение без необходимости тонкой настройки параметров модели.

### 3.4 Многослойный перцептрон

MLP представляет собой последовательность узлов (нейронов), соединенных линейными операциями [25].

В нейронной сети, обозначаемой как  $N$ , есть  $D$  слоев, из которых  $D - 2$  являются скрытыми слоями. В  $k$ -м слое содержится  $n_k$  нейронов. Каждый нейрон получает на вход вектор  $x_k - 1$  из предыдущего слоя, который преобразуется линейно:

$$y = w[0]x[0] + w[1]x[1] + \dots + w[p]x[p] + b, \quad (3.11)$$

где  $y$  – это взвешенная сумма входных признаков  $x[0], \dots, x[p]$ .

Входные признаки взвешены по вычисленным в ходе обучения коэффициентам  $w[0], \dots, w[p]$ . Их можно представить графически (рисунок 3.2).

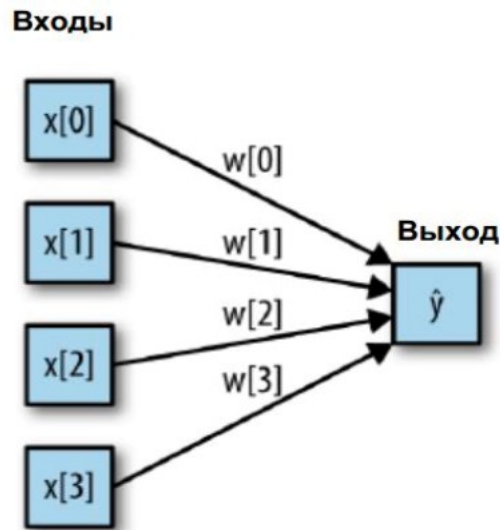


Рисунок 3.2 – Визуализация логистической регрессии, в которой входные признаки и прогнозы показаны в виде узлов, а коэффициенты – в виде соединений между узлами [26]

Далее результат подаётся на нелинейную функцию активации, что позволяет модели обучать сложные зависимости и адаптироваться к разнообразным типам входных данных. Эта операция позволяет сети формировать нелинейные границы решений, что критично для задач, где признаки имеют сложную структуру взаимосвязей.

В MLP процесс вычисления взвешенных сумм повторяется несколько раз. Сначала вычисляются скрытые элементы (hidden units), которые представляют собой промежуточный этап обработки. Они вновь объединяются с помощью взвешенных сумм для получения конечного результата (Рисунок 3.3):

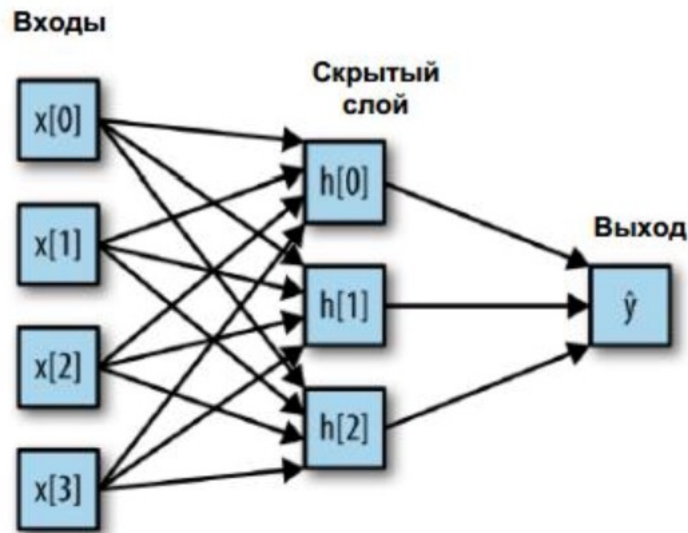


Рисунок 3.3 – Многослойный перцептрон с одним скрытым слоем [26]

Многослойный перцептрон способен находить сложные, в том числе нелинейные зависимости между признаками. В отличие от более простых моделей, он не ограничен линейной границей между классами и может хорошо справляться с запутанными структурами в данных. В задачах вроде анализа пульсовой волны это даёт ему определённое преимущество, особенно когда зависимости между параметрами выражены неявно. Кроме того, модель гибкая и может адаптироваться под разные объёмы и типы входных данных. С другой стороны, MLP требует больше данных и тщательной настройки количества слоёв, нейронов, функции активации, шага обучения и других параметров, которые будут сильно влиять на результат. При недостаточной настройке он может легко переобучиться, особенно если данные шумные или объём выборки ограничен.

Кроме того, обучение MLP требует значительно больших вычислительных ресурсов по сравнению с классическими моделями, такими как логистическая регрессия или наивный байесовский классификатор, что может быть критическим фактором при ограничениях вычислительной среды.

Также сама модель плохо интерпретируема: в отличие от, например, логистической регрессии, здесь невозможно только по коэффициентам понять, какие признаки важнее. Поэтому, несмотря на высокую

потенциальную точность, нейросеть не всегда оказывается лучшим выбором, особенно если требуется интерпретируемость модели.

### 3.5 Метрики

Одним из наиболее популярных способов, позволяющих оценить качество бинарной классификации, является использование матрицы ошибок. Это таблица размером 2x2, где строки соответствуют фактическим классам, а столбцы соответствуют спрогнозированным классам [27].

Элементы главной диагонали матрицы ошибок соответствуют правильным прогнозам (результатам классификации), тогда как остальные элементы показывают, сколько примеров, относящихся к одному классу, были ошибочно классифицированы как другой класс. Правильно классифицированные положительные примеры называются истинно положительными (true positive), а правильно классифицированные отрицательные примеры – истинно отрицательными (true negative). Эти термины, как правило, записывают в сокращенном виде как FP, FN, TP и TN и приводят к следующей интерпретации матрицы ошибок (Рисунок 3.4):

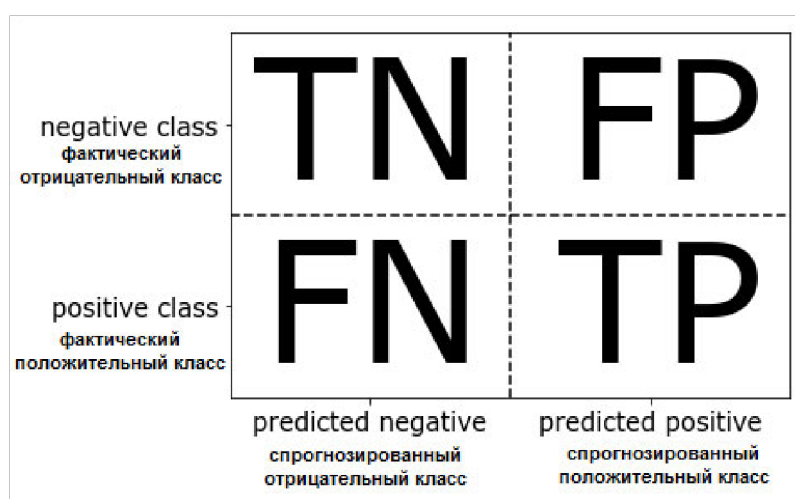


Рисунок 3.4 – Матрица ошибок

Точность (Accuracy) показывает долю верных предсказаний среди всех объектов:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3.12)$$

где  $TP$  – истинно положительные;

$TN$  – истинно отрицательные;

$FP$  – ложноположительные;

$FN$  – ложноотрицательные случаи.

Однако при дисбалансе классов эта метрика может быть необъективной.

Точность (Precision) показывает, сколько из предсказанных положительных примеров оказались действительно положительными:

$$Precision = \frac{TP}{TP + FP}. \quad (3.13)$$

Чувствительность (Recall, TPR) измеряет, какую долю положительных объектов модель верно обнаружила:

$$Recall = \frac{TP}{TP + FN}. \quad (3.14)$$

Специфичность (Specificity) оценивает способность модели корректно идентифицировать отрицательные классы:

$$Specificity = \frac{TN}{TN + FP}. \quad (3.15)$$

Для баланса между точностью предсказаний положительного класса (Precision) и чувствительностью используется F1-score – гармоническое среднее:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}. \quad (3.16)$$

AUC-ROC оценивает качество разделения классов при разных порогах, вычисляя площадь под ROC-кривой, которая строится на основе TPR и FPR (False Positive Rate) [28]. Чем ближе AUC к 1, тем лучше модель различает классы:

$$FPR = \frac{FP}{FP + TN}, \quad (3.17)$$

$$AUC = \int_0^1 TPR(FPR)dFPR. \quad (3.18)$$

Таким образом, каждая из упомянутых метрик даёт свою, частичную оценку качества модели и полезна в разных контекстах. Например, точность (Precision) важна там, где особенно критичны ложноположительные ошибки. Это может быть задача диагностики заболевания, при которой ложноположительное срабатывание приведёт к ненужным дополнительным обследованиям у пациента.

Чувствительность (Recall), напротив, показывает, насколько хорошо модель умеет находить именно нужный класс - например, больных среди всех пациентов. Если Recall низкий, значит, модель часто "пропускает" нужные случаи, что может быть критично в задачах скрининга или предупреждения заболеваний.

Специфичность (Specificity) полезна в дополнение к чувствительности и отражает способность модели правильно распознавать отрицательный класс – например, здоровых пациентов.

F1-мера объединяет Precision и Recall в одну метрику, используя гармоническое среднее. Его особенность в том, что такое среднее чувствительно к малым значениям. В отличие от обычного среднего арифметического, гармоническое среднее не будет высоким, если хотя бы один из компонентов (например, Recall или Precision) очень мал. Это делает F1 полезным показателем в тех случаях, когда необходимо одновременно

контролировать и полноту, и точность модели. Например, если F1 низкий, значит, модель либо плохо угадывает положительный класс, либо часто ошибается, принимая отрицательные объекты за положительные.

ROC-кривая строится при постепенном изменении порога классификации, и AUC фактически измеряет «среднюю» способность модели принимать правильное решение по всей шкале вероятностей. Это полезно при сравнении нескольких моделей между собой.

В совокупности эти метрики позволяют оценить качество модели не только по количеству верных предсказаний, но и по характеру ошибок, что важно при работе с медицинскими или несбалансированными данными, где последствия разных типов ошибок могут быть неравнозначны.



## 4 ОДНОМЕРНАЯ МОДЕЛЬ КРОВОТОКА

Кровоток в крупных системных артериях представлен одномерной моделью нестационарного течения вязкой несжимаемой ньютоновской жидкости в сети соединенных эластичных трубок. Для каждого отдельного сосуда с индексом  $k$  должны выполняться законы сохранения массы и импульса [13]:

$$\frac{\partial S_k}{\partial t} + \frac{\partial(S_k u_k)}{\partial x} = 0, \quad (4.1)$$

$$\frac{\partial u_k}{\partial t} + \frac{\partial(u_k^2/2 + p_k/\rho)}{\partial x} = \psi_k, \quad (4.2)$$

где  $x$  – координата по сосуду;

$t$  – время;

$\rho = 1,04$  г/см<sup>3</sup> – постоянная плотность крови;

$S_k$  – площадь поперечного сечения по сосуду;

$u_k$  – усредненная по поперечному сечению скорость потока;

$p_k$  – усредненное по поперечному сечению кровяное давление;

$\psi_k$  – сила трения на единицу массы крови:

$$\psi_k = -\frac{8\pi\mu u_k}{\rho S_k}, \quad (4.3)$$

где  $\mu$  – динамическая вязкость крови.

На входе в модель артериального русла применяется граничное условие на основе сосредоточенной модели левого сердца с динамикой клапанов. Для каждого сердечного отдела (левое предсердие, левый желудочек) и клапанов (митральный, аортальный) решается система дифференциальных уравнений, учитывающая инерцию, сопротивление, эластичность и активность миокарда.

Контракция миокарда задается функцией времени с учётом физиологической активации и расслабления [13].

Особенностью модели является включение уравнений движения для углов раскрытия клапанов, описываемых аналогом второго закона Ньютона:

$$\frac{d^2\theta_{av}}{dt^2} = (P_{lv} - P_{aorta})K_{av}^p \cos \theta_{av} - K_{av}^p \frac{d\theta_{av}}{dt} - F(\theta_{av}), \quad (4.4)$$

где  $F(\theta)$  – регуляризирующая функция, предотвращающая перераскрытие клапана.

AP моделируется с помощью эффективной площади поперечного сечения ЛЖ, которая может быть рассчитана исходя из максимальной площади поперечного сечения  $A_{max}$  и угла раскрытия клапана  $\theta$ , где  $g(\theta)$  – это гладкая функция, которая равна 1, когда клапан полностью открыт ( $\theta = \theta_{max}$ ), и 0, когда клапан полностью закрыт ( $\theta = \theta_{min}$ ):

$$g(\theta) = \begin{cases} \frac{(1 - \cos\theta_{min})^2}{(1 - \cos\theta_{max})^2}, & \theta < \theta_{min} \\ \frac{(1 - \cos\theta)^2}{(1 - \cos\theta_{max})^2}, & \theta_{min} \leq \theta \leq \theta_{max} \\ 1, & \theta > \theta_{max} \end{cases}. \quad (4.5)$$

Граничные условия оттока реализованы через трехэлементные модели Виндкесселя, включающие сопротивление ( $R_1$  и  $R_2$ ), податливость ( $C$ ) и давление покоя (обычно 7 мм рт. ст.). Они позволяют учитывать инерционные и емкостные эффекты периферического сосудистого русла:

$$\frac{dQ}{dt} = \frac{1}{R_1} \left( \frac{dp_k}{dt} - \frac{dP}{dt} \right), \quad (4.6)$$

$$\frac{dP}{dt} = \frac{Q}{C} \left( 1 + \frac{R_1}{R_2} \right) - \frac{p_k - p}{R_2 C}, \quad (4.7)$$

где  $Q$  – поток на выходе сосуда;

$p$  – давление в виндкессель-камере.

Следует отметить, что в рамках настоящего исследования использованная одномерная модель кровообращения рассматривалась как внешняя расчетная компонента, применяемая без внесения изменений в её математическую структуру или численную реализацию. Модель функционировала в режиме автоматизированной генерации физиологически обоснованных пульсовых волн, и её внутренняя динамика не являлась объектом анализа или адаптации. Такой подход позволил сосредоточиться на задачах обработки выходных данных модели и разработке алгоритмов машинного обучения.

## 5 АНАЛИЗ ФОРМЫ ПУЛЬСОВОЙ ВОЛНЫ

В рамках исследования анализируются пульсовые волны, смоделированные в плечевой артерии.

Основная обработка и анализ сигналов проведены с помощью языка Python и следующих библиотек: NumPy, SciPy, а также Matplotlib – для визуализации сигналов и проверки корректности нахождения характерных точек.

Каждая записанная пульсовая волна представляет собой последовательность значений артериального давления в зависимости от времени. Перед извлечением количественных характеристик сигнала была проведена предварительная обработка, включающая выравнивание начала цикла, сглаживание и нормализацию по времени.

Для устранения высокочастотного шума и мелких артефактов, которые могут мешать точному определению экстремумов, применялся фильтр Савицкого – Голея [29]. Этот метод производит полиномиальное сглаживание скользящим окном, позволяя выравнивать кривую, сохраняя при этом форму и положения основных максимумов и минимумов. В работе использовались параметры окна сглаживания длиной 15 точек и полиномом 2-й степени, что экспериментально обеспечило баланс между устранением шума и сохранением ключевых особенностей формы волны.

Анализ формы пульсовой волны включает поиск ключевых фазовых точек на сглаженном сигнале давления. Эти характерные точки соответствуют физиологически значимым моментам сердечного цикла:

- систолический пик (SBP) – максимальное давление в цикле, достигаемое в фазе систолы. Эта точка определяется как глобальный максимум сглаженной кривой давления;
- дикротический спад (дикротический вырез, d<sub>ic</sub>) – локальный минимум давления, следующий за систолическим пиком. Он соответствует кратковременному падению давления после закрытия аортального клапана.

Эта точка определяется как первый локальный минимум после глобального максимума;

- дикротическая волна (второй пик, dia) – вторичный максимум давления в диастолической фазе, возникающий вслед за дикротическим вырезом из-за отраженной волны от периферических сосудов. Определяется как первый локальный максимум, следующий за дикротическим минимумом;
- конец диастолы (DBP) – глобальный минимум давления в конце рассматриваемого цикла;

На основе выявленных ключевых точек и общей формы пульсовой волны рассчитывались диагностически значимые количественные признаки:

- пульсовое давление (PP) – разница между систолическим и диастолическим давлением:  $PP = SBP - DBP$ . Этот показатель характеризует амплитуду пульсовой волны;
- среднее давление за цикл – среднее значение давления по времени за весь сердечный цикл;
- время подъёма (CT) – интервал времени от начала цикла (минимума давления) до достижения систолического пика. Характеризует длительность фазы систолического выброса и скорость нарастания давления;
- время до дикротического пика – интервал от начала цикла до появления второго (дикротического) пика;
- интервал между пиками (PTR) – разница времен между систолическим пиком и дикротическим вторым пиком. Этот признак отражает время распространения и возврата отраженной волны от периферии;
- длительность диастолы (Tdia) – время от дикротического выреза до конца цикла, то есть продолжительность фазы диастолического спадания давления после закрытия клапана;
- индекс отражения (RI) – отношение высоты основного систолического пика к высоте дикротического второго пика. Этот индекс характеризует относительную выраженность отраженной волны и может

свидетельствовать об изменении тонуса периферических сосудов или жесткости артерий;

- индекс жесткости (SI, Stiffness Index) – показатель, связанный со скоростью распространения пульсовой волны и жесткостью артериальной стенки. В нашем случае вычислялся обратным значением интервала между пиками [30].

На рисунке 5.1 наглядно отмечены перечисленные выше параметры ПВ:

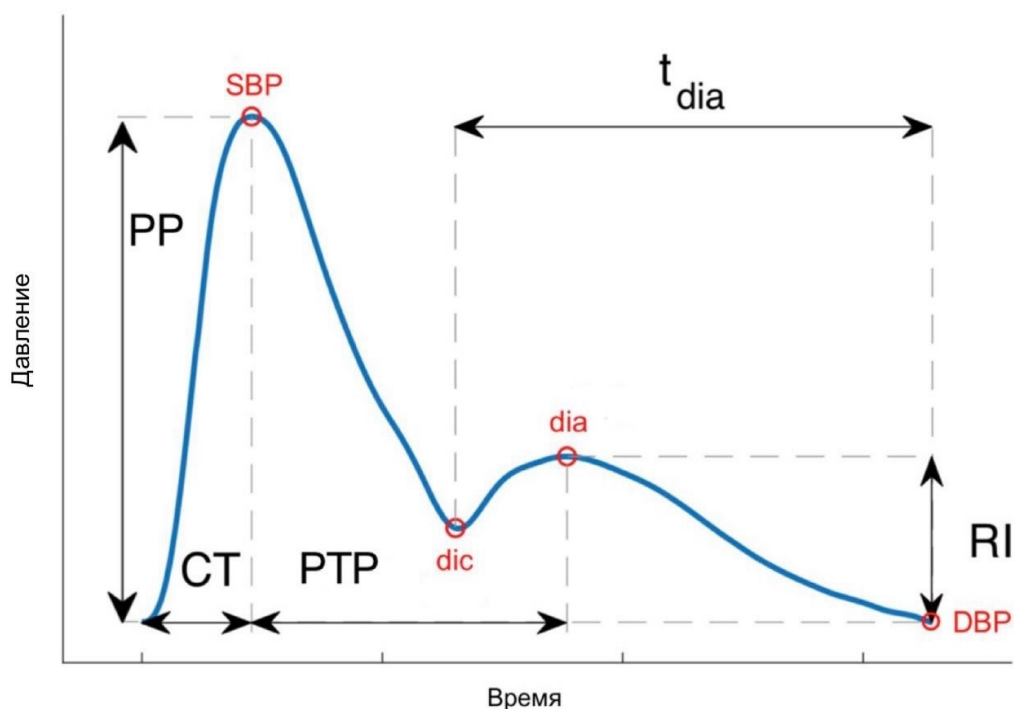


Рисунок 5.1 – ПВ и её характеристики

Помимо абсолютных значений, для устранения межиндивидуальных различий в амплитудах была выполнена нормализация амплитудных признаков. Каждый из амплитудных параметров нормировался относительно систолического пика, а временные характеристики нормировались на длительность цикла, что позволило сравнивать формы волн при разной частоте сердечных сокращений.

## 6 РЕЗУЛЬТАТЫ

### 6.1 Модели для определения сосудистой жесткости

В исследовании использовалась клиническая БД, включающая 28 признаков, охватывающих демографические характеристики, антропометрию, биохимию крови и параметры артериального давления.

Целевыми переменными являлись два варианта бинарных меток сосудистого старения, рассчитанные на основе показателей RCAVI и LCAVI:

- Модель 1 (жёсткий порог  $<9$ ): соответствует отсутствию сосудистого старения при  $RCAVI \leq 9$  и  $LCAVI \leq 9$  независимо от возраста;
- Модель 2 (возраст-зависимый верхний порог): критерии старения адаптированы под возрастную группу пациента.

Распределение значений CAVI по половому признаку в разрезе возрастных групп представлено на рисунках 6.1 и 6.2:

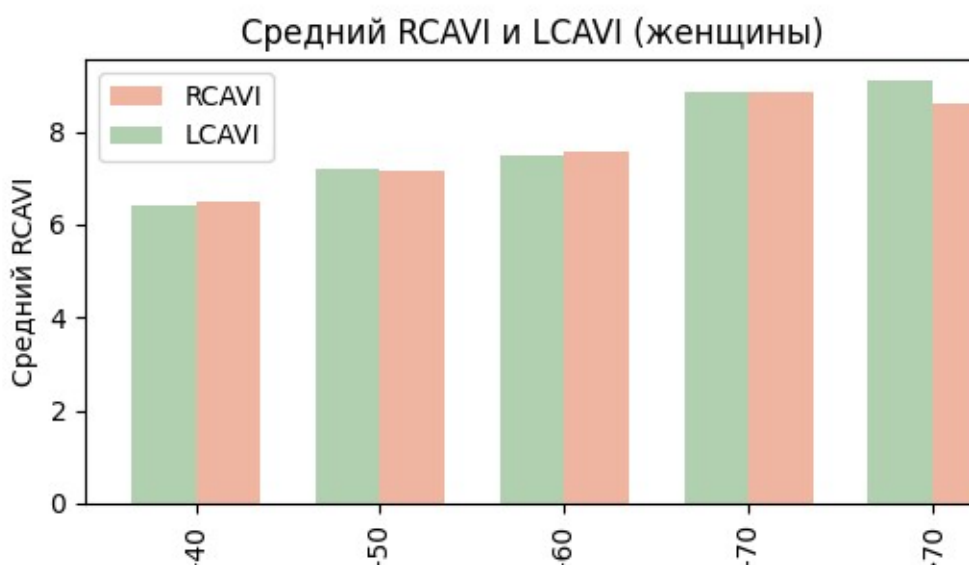


Рисунок 6.1 – распределение CAVI среди женщин.

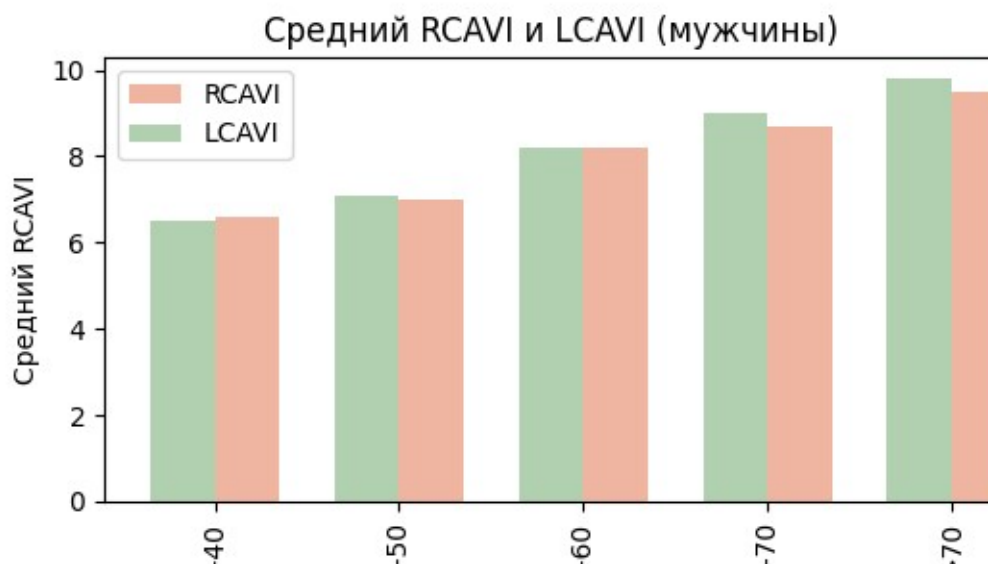


Рисунок 6.2 – распределение CAVI среди женщин

Также был выполнен корреляционный анализ исследуемых параметров с построением матрицы корреляций, что позволило выявить степень взаимосвязи между переменными.



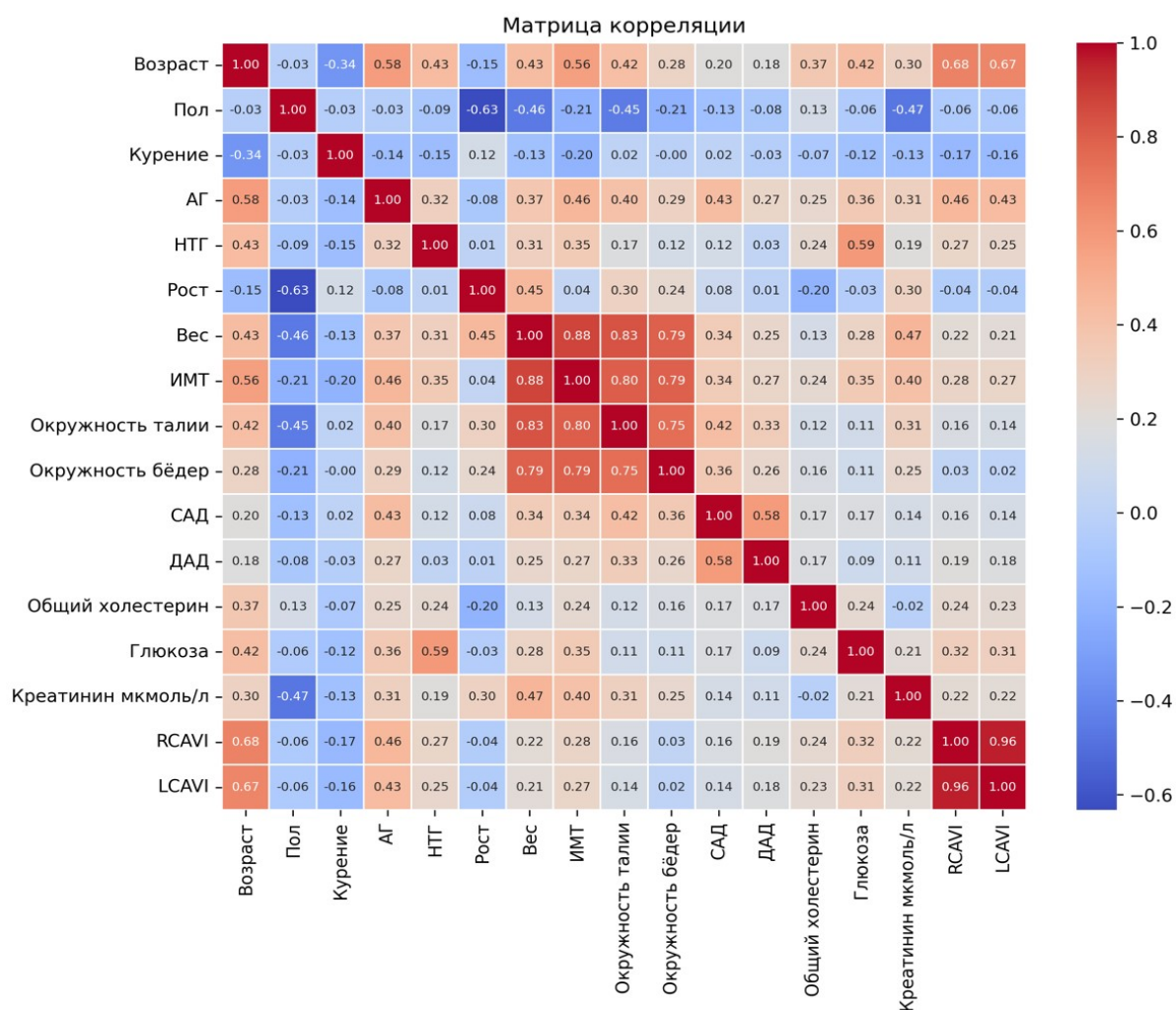


Рисунок 6.3 – матрица корреляций параметров пациентов

Исходя из результатов корреляционного анализа можно предположить, что наибольшее значение имеют такие параметры как возраст и наличие у человека артериальной гипертензии (АГ). Так же есть заметная зависимость между САVI и показателем глюкозы.

На рисунке 6.4 видно, что модель с индивидуальными возрастными границами выявляет больше случаев повышенного САVI по сравнению с моделью, использующей единый порог практически для всех возрастных групп. Это особенно заметно в средних возрастных категориях, где индивидуальные референсные значения, вероятно, учитывают естественные возрастные изменения сосудистой жесткости. Результаты демонстрируют, что персонализированный подход повышает чувствительность диагностики, что

может быть критически важно для раннего выявления сердечно-сосудистых рисков.

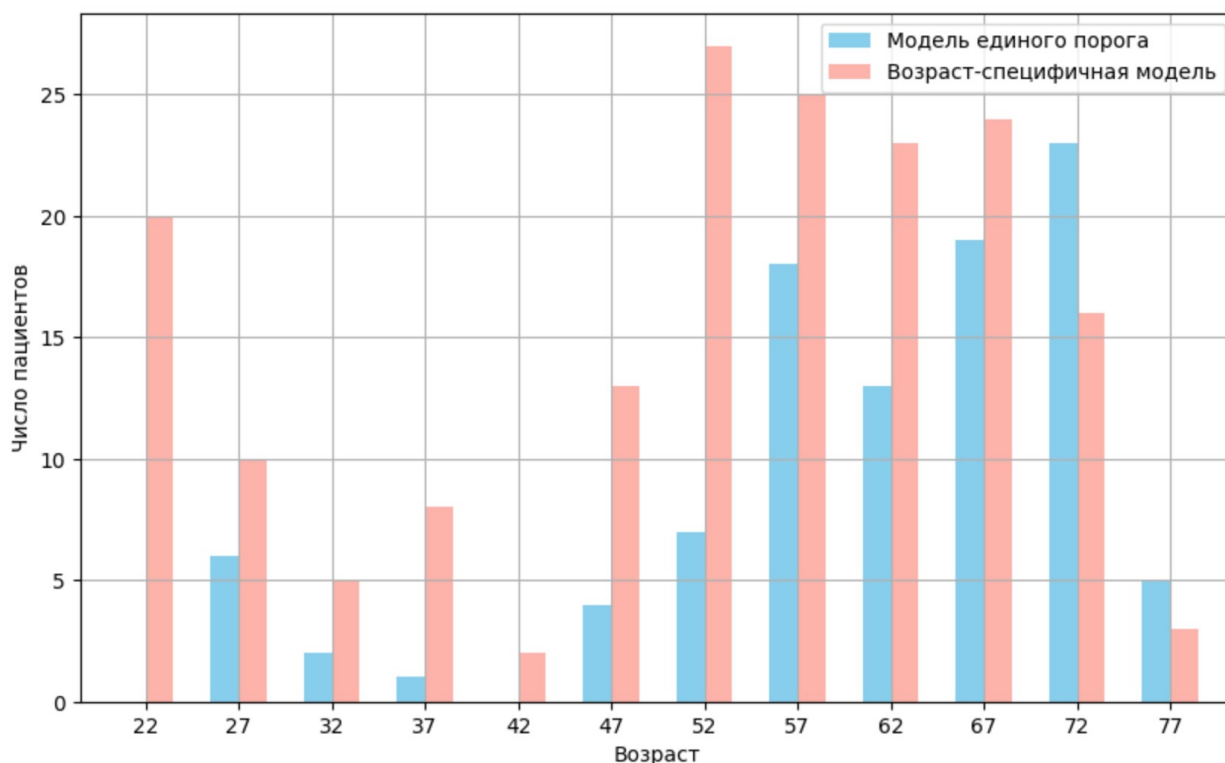


Рисунок 6.4 – Распределение числа пациентов с повышенным значением CAVI по возрастным группам

Для каждой из целевых переменных была обучена ЛР. Чтобы оценить устойчивость и воспроизводимость результатов, весь процесс обучения и валидации повторялся 100 раз с новым случайным разбиением на тренировочную и тестовую выборки. В каждой итерации рассчитывались следующие метрики:

- чувствительность – способность модели распознавать случаи сосудистого старения;
- специфичность – способность модели корректно выявлять отсутствие старения;
- общая точность;
- площадь под ROC-кривой (AUC ROC).

Таблица 6.1 – Сравнение метрик моделей машинного обучения

	Чувствительность	Специфичность	Точность	AUC ROC
Модель 1	0.73 [0.69, 0.77]	0.70 [0.71, 0.75]	0.72 [0.71, 0.73]	0.63, [0.61, 0.65]
Модель 2	0.78 [(0.76, 0.80]	0.73 [0.71, 0.75]	0.76 [0.74, 0.78]	0.67, [0.65, 0.68]

Видно, что модель 2 демонстрирует более высокие показатели диагностической эффективности по всем ключевым метрикам по сравнению с моделью 1. Ее чувствительность и специфичность указывают на лучшую способность как правильно выявлять случаи повышенного САVI, так и минимизировать ложноположительные результаты.

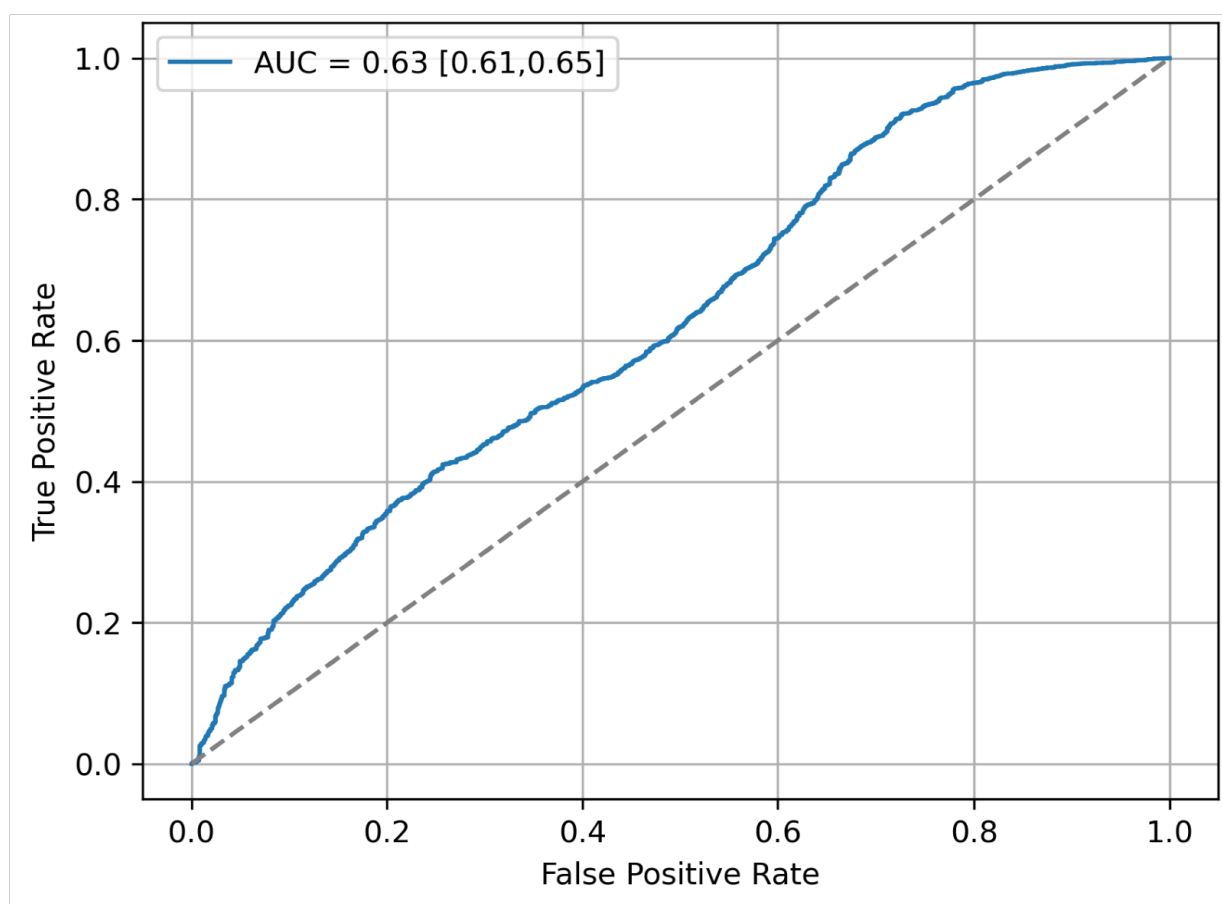


Рисунок 6.5 – ROC-кривая для ЛР модели 1

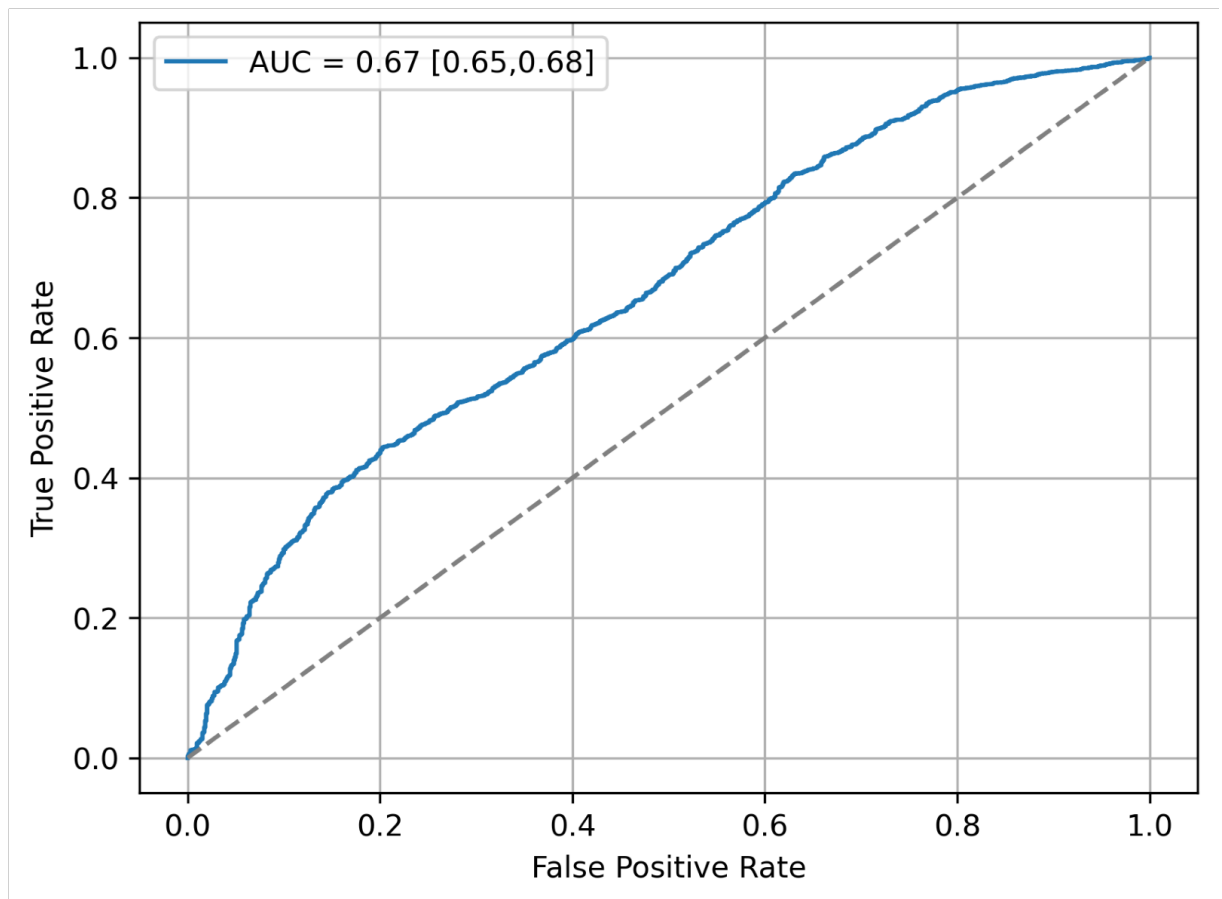


Рисунок 6.6 – ROC-кривая для ЛР модели 2

Преимущество модели с индивидуальными возрастными границами наблюдается и в значении AUC ROC, что свидетельствует о лучшей дискриминационной способности.

Также был осуществлен переход к непрерывной шкале нормативных значений САVI для каждого возраста:

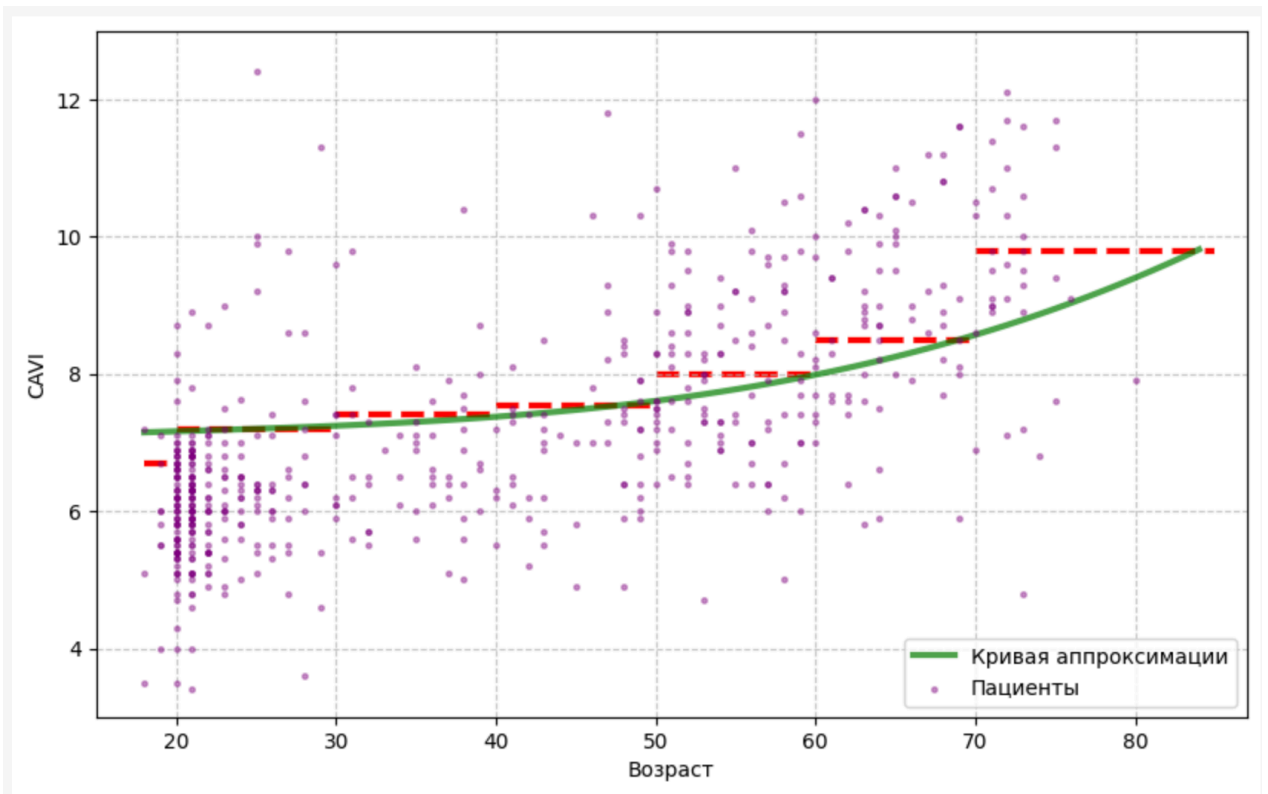


Рисунок 6.7 – кривая, аппроксимирующая значения САVI по возрастам

При переходе к непрерывной шкале нормативных значений САVI путем математической аппроксимации выросли метрики модели ЛР:

Таблица 6.2 – Сравнение метрик моделей машинного обучения модели, рассматривающий индивидуальные пороговые значения для каждого возраста

Чувствительность	Специфичность	Точность	AUC ROC
0.76 [0.73, 0.79]	0.76 [0.74, 0.78]	0.76 [0.75, 0.77]	0.84, [0.82, 0.86]

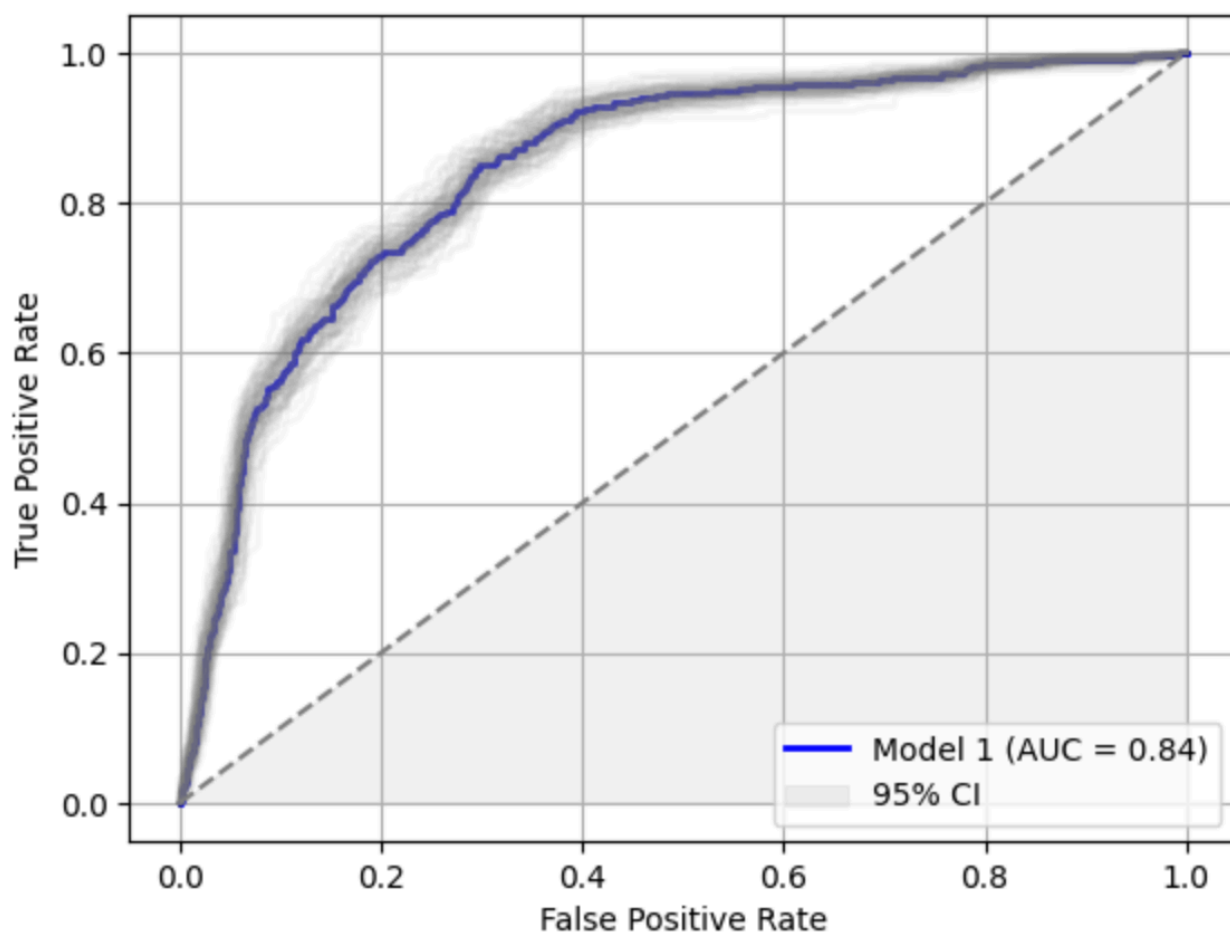


Рисунок 6.8 – ROC-кривая для ЛР модели индивидуальных пороговых значений САVI

Из рисунка 6.8 видно, что значительно возросла прогностическая способность метода: площадь под ROC-кривой (AUC ROC) достигла 0.84. Также немного улучшились и другие показатели.

Таким образом, переход к непрерывной шкале нормативных значений САVI показал себя наиболее эффективным и перспективным методом в данном исследовании.

## 6.2 Расчет вероятности ускоренного сосудистого старения

Формула (6.1) рассчитывает вероятность наличия ускоренного сосудистого старения (раздел 3.1).

$$\alpha = \left[ 1 + \exp(-\beta_0 - \sum_{i=1}^8 \beta_i x_i) \right]^{-1}, \quad (6.1)$$

где параметры  $\beta$  являются фиксированными;

$x$  – характеристики конкретного пациента.

В результате работы алгоритма ЛР были получены веса  $w$  всех признаков:

Таблица 6.3 – Коэффициенты ЛР для прогнозирования САVI

Описание характеристики $x_i$	Значение $\beta_i$	Ед. измерения	Standart error	p-value
-	-2.15966078	-	0.202428	<0.001
Возраст	0.05668087	год	0.305197	0.004
Пол	-0.36457506	1 – муж, 2 – жен	0.222258	0.384
Стаж АГ	0.0176793	год	0.234792	0.559
НТГ	0.34788669	1 – есть, 0 – нет	0.217079	0.495
Вес	-0.01905415	кг	0.255772	0.230
Общий холестерин	0.10361634	ммоль/л	0.187401	0.452
Глюкоза	0.05716038	ммоль/л	0.218373	0.461
Креатинин	-0.00385309	мкмоль/л	0.207095	0.489

Полученные коэффициенты модели были применены для проверки работоспособности формулы 6.1 на независимой выборке из 153 пациентов, не участвовавших в обучении. Модель продемонстрировала высокую чувствительность – 0.87, что указывает на её способность правильно идентифицировать подавляющее большинство истинных случаев повышенного САVI (75 из 86 случаев). Однако специфичность оказалась крайне низкой 0.015, что означает отсутствие способности корректно исключать пациентов без патологии. Точность модели равна 0.5 и соответствует случайному угадыванию. При высокой детекции истинно положительных случаев модель ошибочно присваивает повышенный САVI почти всем пациентам (66 ложноположительных результатов).

Полученные результаты демонстрируют ограниченную диагностическую ценность, что в первую очередь связано с недостаточной репрезентативностью исследовательской выборки. Основная проблема заключается в неравномерном распределении пациентов по ключевым демографическим и клиническим характеристикам, что привело к систематическим ошибкам в оценке диагностических параметров.

Полученные результаты также показали, что использование синтетических смоделированных данных в исследовании представляется особенно ценным подходом, поскольку это позволяет преодолеть ключевые ограничения реальных клинических данных. Генерация искусственных, но статистически достоверных данных дает возможность существенно расширить объем анализируемой информации, включая редкие клинические случаи и экстремальные возрастные группы, которые часто недостаточно представлены в реальной выборке. Такой подход помогает сбалансировать распределение пациентов по различным категориям, что критически важно для построения надежных диагностических моделей.

### **6.3 База данных виртуальной популяции**

БД состоит из 1458 записей, каждая из которых соответствует одному виртуальному пациенту. Все пациенты были сгенерированы с использованием физиологической модели, воспроизводящей основные закономерности кровообращения у лиц пожилого возраста. В выборке представлены два класса:

- 729 здоровых пациентов;
- 729 пациентов с АР различной степени (15-20%).

Таким образом, выборка сбалансирована, что обеспечивает корректность обучения и валидации алгоритмов машинного обучения, исключая смещение по классу.



Каждый пациент представлен уникальным идентификатором, а целевая переменная, отображающая отсутствие/наличие у пациента АР, принимает значение 0 или 1, соответственно.

Перед анализом данные были преобработаны и проверены на наличие аномалий. Дополнительно были проведены следующие шаги:

- проверка выбросов по каждому параметру;
- визуализация распределений;
- предварительный анализ корреляций между признаками.

При изучении амплитудных характеристик пульсовых волн особое внимание уделялось ключевым физиологическим параметрам. В БД вошли следующие характеристики: частота сердечных сокращений, систолическое давление, диастолическое давление, давление на диастолическом пике, давление в дикротической выемке, пульсовое давление, среднее давление за период, время достижения систолического пика ПВ, время достижения диастолического пика ПВ, время диастолического спада, время между систолическим и диастолическим пиками, индекс отражения ПВ, индекс жесткости сосудов, а также их нормализованные значения.

В таблице 6.4 представлены физиологические характеристики получившегося виртуального населения.

Таблица 6.4 – Характеристики ПВ виртуальных пациентов

Параметр	Все пациенты	Больные	Здоровые	p-value
Систолическое давление, мм рт. ст.	118.25 ± 6.61	115.67 ± 6.45	119.16 ± 6.25	<0.001
Диастолическое давление, мм рт. ст.	63.65 ± 9.71	54.92 ± 5.59	71.38 ± 4.91	<0.001
Дикротическая выемка, мм рт. ст.	100.08 ± 6.82	97.58 ± 7.00	102.27 ± 5.76	<0.001
Пик отраженной волны, мм рт. ст.	101.92 ± 7.18	98.56 ± 7.12	104.69 ± 5.84	<0.001
Падение давления, мм рт. ст.	53.84 ± 10.28	60.49 ± 8.93	48.34 ± 7.22	<0.001

Продолжение таблицы 6.4

Среднее давление, мм рт. ст.	88.55 ± 6.70	83.30 ± 4.56	93.30 ± 4.30	<0.001
Время достижения 1го пика ПВ, с.	0.15 ± 0.02	0.16 ± 0.01	0.14 ± 0.01	<0.001
Время запаздывания отраженной волны, с	0.12 ± 0.02	0.12 ± 0.02	0.11 ± 0.02	<0.001
Время диастолического спада, с.	0.54 ± 0.08	0.52 ± 0.08	0.55 ± 0.08	<0.001
Время достижения 2го пика ПВ, с.	0.26 ± 0.03	0.28 ± 0.02	0.25 ± 0.02	<0.001
Индекс отражения	1.15 ± 0.04	1.17 ± 0.04	1.14 ± 0.02	<0.001
Индекс жесткости стенки сосуда, м/с	8.62 ± 7.21	8.20 ± 7.47	8.93 ± 6.95	<0.001

На основании проведенного анализа выявлены статистически значимые различия ( $p < 0.05$ ) между группами больных и здоровых по всем исследуемым параметрам. У больных наблюдается значительное снижение систолического, диастолического и среднего артериального давления по сравнению со здоровыми, а также более выраженное падение давления. Повышенный индекс отражения у больных подтверждает усиленную рефлексию пульсовой волны.

Для сравнения в таблице 6.5 представлены те же характеристики реальных пациентов, данные которых были предоставлены для исследования врачами-кардиологами УКБ №3 Первого МГМУ им. И.М. Сеченова.

Таблица 6.5 – Характеристики ПВ реальных пациентов

Параметр	Все пациенты	Больные	Здоровые	p-value
Систолическое давление, мм рт. ст.	131.59 ± 17.48	138.61 ± 21.68	129.94 ± 15.68	0.26
Диастолическое давление, мм рт. ст.	84.10 ± 11.23	84.28 ± 9.10	83.79 ± 11.94	0.81
Дикротическая выемка, мм рт. ст.	117.13 ± 17.99	128.42 ± 21.07	115.04 ± 16.37	0.07
Пик отраженной волны, мм рт. ст.	111.47 ± 16.31	114.46 ± 18.12	110.54 ± 15.71	0.55

## Продолжение таблицы 6.5

Падение давления, мм рт. ст.	45.12 ± 13.21	52.05 ± 19.47	44.21 ± 9.88	0.15
Среднее давление, мм рт. ст.	108.86 ± 13.32	109.33 ± 13.93	108.70 ± 13.16	0.51
Время достижения 1го пика ПВ, с.	0.14 ± 0.04	0.18 ± 0.04	0.16 ± 0.04	0.08
Время запаздывания отраженной волны, с	0.23 ± 0.10	0.18 ± 0.12	0.24 ± 0.09	0.65
Время диастолического спада, с.	0.58 ± 0.15	0.60 ± 0.16	0.57 ± 0.14	0.17
Время достижения 2го пика ПВ, с.	0.39 ± 0.11	0.37 ± 0.13	0.39 ± 0.10	0.87
Индекс отражения	1.16 ± 0.11	1.16 ± 0.10	1.15 ± 0.11	0.44
Индекс жесткости стенки сосуда, м/с	4.42 ± 1.94	5.47 ± 2.08	4.23 ± 1.89	0.32

Анализ представленных данных показывает, что значения параметров выглядят реалистично и соответствуют диапазонам, характерным для клинической практики. У больных наблюдается тенденция к более высоким значениям систолического давления и индекса жесткости стенки сосуда. При этом отсутствие значимых различий в диастолическом давлении, среднем давлении и временных параметрах пульсовой волны говорит о сохранности основных гемодинамических показателей.

### **6.4 Модели машинного обучения для классификации пульсовой волны**

В первой серии экспериментов все четыре модели были протестированы на синтетически сгенерированном наборе данных без добавления шума. Результаты показывают исключительно высокую эффективность всех алгоритмов:

Таблица 6.6 – Сравнение метрик моделей, протестированных на синтетических данных

Модель	Точность	Чувствительность	Специфичность	F1-score	AUC ROC
ЛР	0.93	0.93	0.93	0.93	0.99
RF	0.99	1.00	0.97	0.99	0.99
НБ	0.90	0.90	0.89	0.90	0.97
MLP	0.99	0.99	0.99	0.99	0.99

Во второй серии экспериментов те же алгоритмы были протестированы на реальных клинических данных, что позволило оценить их применимость в условиях практического использования. Как видно из таблицы 6.7, наблюдается резкое снижение качества классификации для всех моделей:

Таблица 6.7 – Сравнение метрик моделей, протестированных на данных реальных пациентов

Модель	Точность	Чувствительность	Специфичность	F1-score	AUC ROC
ЛР	0.75	0.16	0.96	0.25	0.67
RF	0.53	0.44	0.56	0.33	0.52
НБ	0.67	0.40	0.76	0.38	0.54
MLP	0.46	0.56	0.40	0.35	0.51

Эти результаты указывают на значительное снижение обобщающей способности моделей при переходе от синтетических к реальным данным, что, скорее всего, связано с несовпадением распределений признаков, наличием шума и разнообразием клинических сценариев.

Снижение качества классификации на реальных данных не является неожиданным: оно отражает сложность перехода от контролируемой симуляции к реальной клинической практике. Это важный шаг в процессе разработки – он выявляет реальные вызовы и области, требующие доработки.

Модель MLP показала наивысшую чувствительность среди всех алгоритмов, несмотря на общее снижение других метрик. Это может свидетельствовать о потенциальной пригодности таких нейросетевых

архитектур в задачах, где важно минимизировать упущенные положительные случаи (например, при первичном скрининге).

## 6.5 Исследование влияния шума на метрики моделей машинного обучения для классификации пульсовой волны

Для оценки устойчивости моделей классификации к шумам были проведены эксперименты обучения вышеописанных моделей на синтетической БД виртуальной популяции с добавлением шума 20%.

Ниже в таблицах 6.8 и 6.9 представлены значения метрик исследуемых моделей:

Таблица 6.8 – Сравнение метрик моделей, протестированных на синтетических данных и обученных на данных с добавлением шума 20%

Модель	Точность	Чувствительность	Специфичность	F1-score	AUC ROC
ЛР	0.92	0.93	0.92	0.92	0.98
RF	0.99	1.00	0.99	0.99	0.99
НБ	0.90	0.90	0.89	0.90	0.97
MLP	0.99	0.99	0.99	0.99	0.99

Таблица 6.9 – Сравнение метрик моделей, протестированных на данных реальных пациентов и обученных на данных с добавлением шума 20%

Модель	Точность	Чувствительность	Специфичность	F1-score	AUC ROC
ЛР	0.76	0.16	0.94	0.24	0.66
RF	0.55	0.44	0.59	0.34	0.54
НБ	0.68	0.40	0.78	0.39	0.55
MLP	0.65	0.44	0.72	0.39	0.56

### 6.5.1 Логистическая регрессия

На синтетических данных модель работает стабильно даже с шумом – падения показателей метрик практически нет. При этом специфичность после

тестирования на реальных пациентах сохраняется на высоком уровне (0.94 – 0.96). Это говорит о том, что модель практически не распознаёт заболевших пациентов (Recall  $\approx$  16%), но успешно предсказывает здоровых. Добавление шума в S4 не помогает улучшить обобщающую способность.

ЛР не адаптируется к шуму и плохо переносится на реальные данные.

### ***6.5.2 Наивный байесовский классификатор***

НБ демонстрирует стабильную производительность на синтетической БД с шумом. При валидации на реальных данных чувствительность сохраняется на уровне 0.40, что показывает, что модель распознаёт часть больных. Добавление шума незначительно улучшает показатели при валидации на выборке реальных пациентов.

### ***6.5.3 Случайный лес***

Как и остальные модели случайный лес показывает крайне высокие результаты на синтетических данных. При переходе на реальные данные чувствительность остаётся на уровне 0.44, что больше, чем у ЛР и сравнимо с НБ. Добавление шума помогает улучшить общую устойчивость, но не значительно. В реальных условиях сохраняет приемлемую чувствительность, поэтому подходит для задач, где Recall важнее Precision.

### ***6.5.4 Многослойный перцептрон***

MLP устойчив к шуму на синтетической БД. На реальных данных демонстрирует наивысшую специфичность среди всех моделей, при обучении как с шумом, так и без. Однако точность и AUC ROC остаются низкими. После обучения на зашумлённых данных, точность и специфичность существенно возрастают.

Можно сделать вывод о том, что MLP лучше всего адаптируется к реальным данным, особенно с шумом. Это делает MLP самым перспективным методом среди исследованных в данной работе

## Заключение

Процесс скрининга ССЗ на сегодняшний день требует доступных и недорогих методик обнаружения риска наличия заболевания.

В данной работе решалась проблема недостаточной изученности методов классификации антропометрических данных человека и данных анализа крови с целью прогнозирования развития АР и ССЗ, связанных с повышенной сосудистой жесткостью, и как следствие недостаточного обеспечения процесса скрининга недорогими методами прогнозирования наличия заболевания.

Проведенное исследование демонстрирует потенциал использования пульсовой волны давления и методов машинного обучения для диагностики АР и оценки сосудистого старения.

В ходе исследования решены следующие задачи:

1. Проведен анализ клинических данных пациентов, включая антропометрические показатели, лабораторные анализы и измерения САVI.

2. Разработаны и оценены модели машинного обучения для прогнозирования повышенного сосудистого старения, продемонстрировавшие свою эффективность в выявлении пациентов с высоким риском.

3. Создана синтетическая база данных ПВ с использованием математической модели гемодинамики, что позволило смоделировать различные степени АР и изучить их влияние на форму пульсовой волны.

4. Разработан алгоритм анализа ПВ, включающий выделение ключевых характеристик сигнала и вычисление диагностически значимых параметров, чувствительных к наличию АР.

5. Проведено тестирование моделей машинного обучения на синтетических и реальных данных.

Основные результаты работы:



Модели, обученные на синтетических данных, показывают высокие метрики даже в условиях зашумленных входов. Особенно хорошо проявляют себя ансамблевые методы (Случайный лес) и MLP, обеспечивая точность и AUC ROC выше 0.99.

Однако при валидации на реальных клинических данных наблюдается резкое падение качества классификации. Логистическая регрессия показывает крайне низкую чувствительность (~16%), практически полностью игнорируя случаи с патологией. MLP, напротив, демонстрирует наивысшую способность адаптироваться к шуму и реальным данным, хоть и уступает по точности моделям на синтетике. Более низкое качество классификации в данном случае может быть обусловлено склонностью моделей к переобучению.

Добавление шума на этапе обучения позволяет незначительно повысить обобщающую способность моделей, особенно для MLP и НБ, что подтверждает актуальность использования техник искусственного расширения данных в медицинских задачах с ограниченными объемами реальных данных.

Перспективы дальнейшего развития работы связаны, в первую очередь, с расширением и углублением экспериментальной базы. Возможным направлением является включение в анализ других патологий, оказывающих влияние на форму пульсовой волны, что позволит расширить область применения разработанных моделей. Также представляет интерес интеграция более сложных методов машинного обучения, включая градиентный бустинг и сверточные нейронные сети. Дополнительное внимание следует уделить адаптации моделей к шумным и неполным клиническим данным, что повысит их устойчивость в реальных условиях. Наконец, дальнейшая валидация на большем числе реальных пациентов и использование мультицентровых данных позволит оценить обобщающую способность подхода и приблизить его к возможной клинической применимости.

Проведенное исследование вносит вклад в развитие цифровой кардиологии, предлагая новые подходы к раннему выявлению сосудистого

старения и АР, что может способствовать снижению риска развития тяжелых сердечно-сосудистых осложнений.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Мухамеджанов А.Х. СЕРДЕЧНО-СОСУДИСТАЯ СИСТЕМА / Мухамеджанов А.Х., Резяпова Д.Р., Зокиров Ф.К. // Modern Science and Research. – 2024. – Т. 3, № 9. - С. 534-537.
2. X Peter H. Charlton. Modeling arterial pulse waves in healthy aging: a database for in silico evaluation of hemodynamics and pulse wave indexes / X Peter H. Charlton,1 X Jorge Mariscal Harana,1 Samuel Vennin [и др.] // American Journal of Physiology - Heart and Circulatory Physiology. – 2019. – Т. 317. – С. H1062–H1085.
3. S. Laurent. Повреждение крупных артерий при гипертензии / S. Laurent, H. Beaussier, C. Collin [и др.] // Современные проблемы науки и образования. – 2010. – Т. 16, № 2.
4. Anna Bragina. Age-Specific Approach to Arterial Stiffness Prediction in Apparently Healthy Patients / Anna Braginaa, Yulia Rodionovaa, Natalia Druzhinina [и др.] // Journal of Clinical Medicine Research. – 2024. – Т. 16, № 9. – С. 423-435.
5. Toru Miyoshi. Predictive Value of the Cardio-Ankle Vascular Index for Cardiovascular Events in Patients at Cardiovascular Risk / Toru Miyoshi, Hiroshi Ito, Kohji Shirai [и др.] // Journal of the American Heart Association. – 2012. – Т. 10, № 16.
6. Yuko Okamoto. Cardio-Ankle Vascular Index as an Arterial Stiffness Marker Improves the Prediction of Cardiovascular Events in Patients without Cardiovascular Diseases / Yuko Okamoto, Toru Miyoshi, Keishi Ichikawa [и др.] // Journal of Cardiovascular Development and Disease. – 2022. – Т. 9, № 11. – С. 368.
7. Gerald Maurer. Aortic Regurgitation / d Maurer // Heart. – 2006. – Т. 92. – С. 994-1000.
8. Roberta Ancona. Epidemiology of aortic valve stenosis (AS) and of aortic valve incompetence (AI): is the prevalence of AS/AI similar in different parts of the

world? / Roberta Ancona, Salvatore Comenale Pinto. // e-Journal of Cardiology Practice. – 2020. – Т. 18, № 10.

9. Mark Lebehn. Contemporary Evaluation and Clinical Treatment Options for Aortic Regurgitation / Mark Lebehn, Torsten Vahl, Polydoros Kampaktsis, Rebecca T Hahn // Journal of Cardiovascular Development and Disease. – 2023. – Т. 10, № 9. - С. 364.

10. Shinnosuke Sawano. Deep learning model to detect significant aortic regurgitation using electrocardiography / Shinnosuke Sawano, Satoshi Kodera , Susumu Katsushika // Journal of cardiology. – 2022. – Т. 79, № 3, - С. 334341.

11. Joanna Bartkowiak. Advances and Challenges in Aortic Regurgitation / Joanna Bartkowiak, MD; Jonas Lanz, MD; Thomas Pilgrim, MD; and Fabien Praz // JACC: Cardiovascular Interventions. – 2025. – Т. 19, № 2, - С. 41-44.

12. Vidhu Anand. Machine learning-based risk stratification for mortality in patients with severe aortic regurgitation / Vidhu Anand, Hanwen Hu, Alexander D Weston [и др.] // European heart journal. Digital health. – 2023. – Т. 4, № 3, - С. 188-195.

13. Artem Rogov. Synthetic pulse wave dataset for analysis of vascular ageing in elderly patients / Artem Rogov, Timur Gamilov, Anna Bragina [и др.] // Mathematical Modelling of Natural Phenomena. – 2024. – Т. 19, - С. 20.

14. Karim Kadry. Biomechanics of diastolic dysfunction: a one-dimensional computational modeling approach / Karim Kadry, Stamatia Pagoulatou, Quentin Mercier [и др.] // Heart and Circulatory Physiology. – 2020. – Т. 319, № 4. – С. H882-H892.

15. Etienne Boileau. A benchmark study of numerical schemes for one-dimensional arterial blood flow modelling / Etienne Boileau<sup>1</sup>, Perumal Nithiarasu<sup>1</sup>, Pablo J. Blanco [и др.] // INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN BIOMEDICAL ENGINEERING. – 2015. – Т. 31, № 10.

16. Muhammad Usman Aslam. Machine learning-based classification of valvular heart disease using cardiovascular risk factors / Muhammad Usman Aslam, Songhua Xu, Sajid Hussain [и др.] // Nature: Scientific Reports. – 2024. – Т. 14.

17. Обрезан А.А. Взаимосвязь факторов кардиоваскулярного риска и маркеров сердечно-сосудистого старения / Обрезан А.А., Сережина Е.К., Кантемирова Р.К. [и др.] // Кардиология: новости, мнения, обучение. – 2023. – Т. 11, № 3. – С. 30-42.
18. Philip Sedgwick. Spearman's rank correlation coefficient / Philip Sedgwick // BMJ. – 2014. – Т. 349.
19. Hosmer, D.W. and Lemeshow, S. (1989), Applied Logistic Regression. New York:Wiley.
20. Kleinbaum, D.G. Logistic regression. Part of the Statistics for Biology and Health book series (SBH). Springer, New York; 2002.
21. Harrell FE. Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis. Springer, New York; 2001.
22. G. Biau. A random forest guided tour / Gérard Biau, Erwan Scornet // TEST. – 2016. – Т. 25. – С. 197–227.
23. Yile Ao. The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling / Yile Ao, Hongqi Li, Liping Zhu [и др.] // Journal of Petroleum Science and Engineering. – 2019. – Т. 174. – С. 776-789.
24. Eibe Frank. Technical Note: Naive Bayes for Regression / E. Frank, L. Trigg, G. Holmes, Ian H. Witten // Machine Learning. – 2000. – Т. 41. – С. 5-25.
25. Abbasi J., Andersen P. Physical activation functions (pafs): An approach for more efficient induction of physics into physics-informed neural networks (pinns) // arXiv preprint arXiv:2205.14630. – 2022.
26. Muller A, Guido S. Introduction to Machine Learning With " Python. 1st ed. Schanafelt D, editor. Sebastopol: O'Reilly Media; 2016.
27. Diogo V. Carvalho. Machine Learning Interpretability: A Survey on Methods and Metrics / Diogo V. Carvalho, Eduardo M. Pereira, Jaime S. Cardoso // Electronics. – 2019. – Т. 8, № 8. – С. 832.

28. Xiaofeng Dou. Determination of Tumor Marker Screening for Lung Cancer Using ROC Curves / Xiaofeng Dou, Jiachen Lu, Yingying Yu [и др.] // Disease Markers. – 2024. – №. 1. ID 4782618.

29. Sahar Jahani. Motion artifact detection and correction in functional near-infrared spectroscopy: a new hybrid method based on spline interpolation method and Savitzky–Golay filtering / Sahar Jahani, Seyed K. Setarehdan, David A. Boas, Meryem A. Yücel // Neurophotonics. – 2018. – Т. 5, №1.

30. Кубарко А. И. Пульсовые волны. Роль в осуществлении и регуляции гемодинамики. Часть 1. Виды пульсовых волн, происхождение, распространение, регистрация и анализ; связь параметров с работой сердца и гемодинамикой. / Кубарко А. И., Гуринович И. М. // Медицинский журнал. – 2025. – № 1.

## Приложение А

(обязательное)

### Антропометрические параметры исследуемой группы пациентов

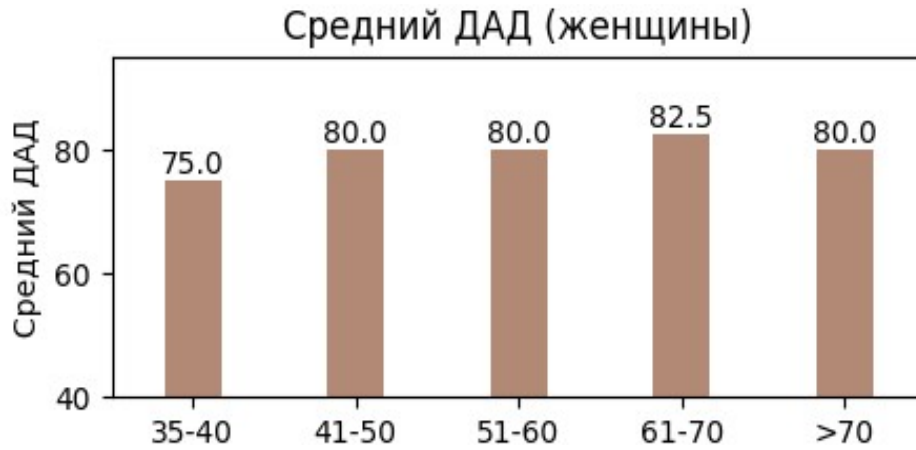


Рисунок А.1 – средний ДАД у женщин

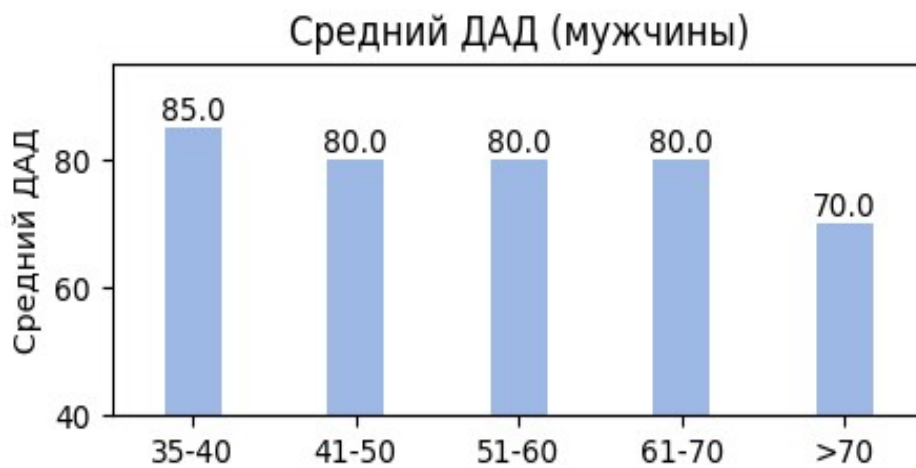


Рисунок А.2 – средний ДАД у мужчин



Рисунок А.3 – средний ИМТ у женщин

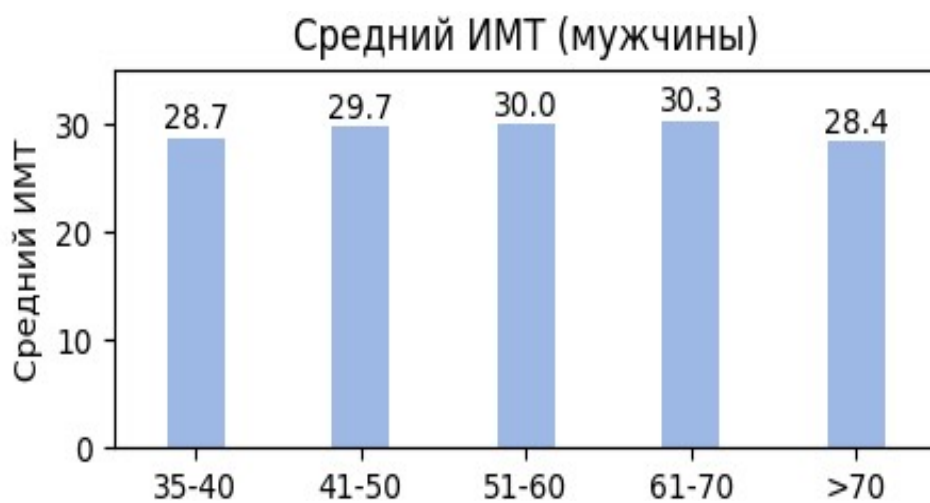


Рисунок А.4 – средний ИМТ у мужчин



Рисунок А.5 – средний САД у женщин

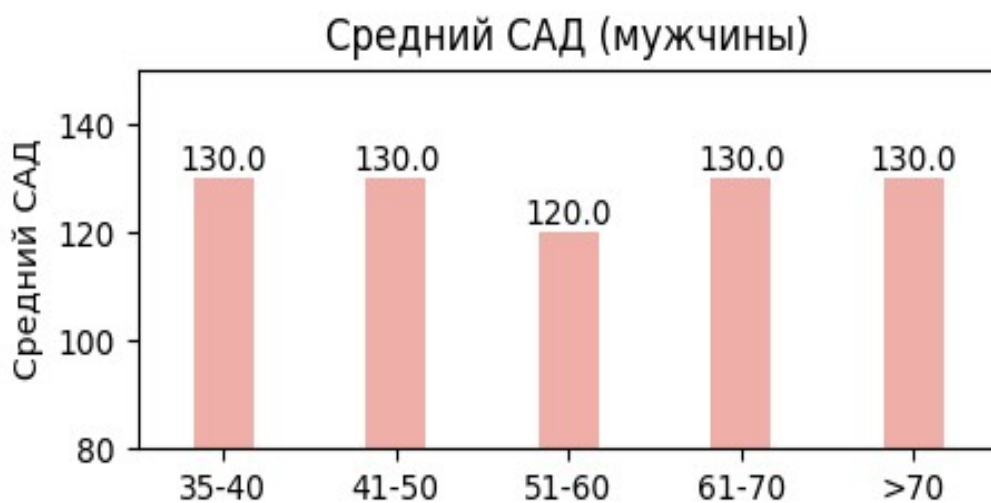




Рисунок А.6 – средний САД у мужчин



Рисунок А.7 – средний уровень триглицеридов у женщин



Рисунок А.8 – средний уровень триглицеридов у мужчин

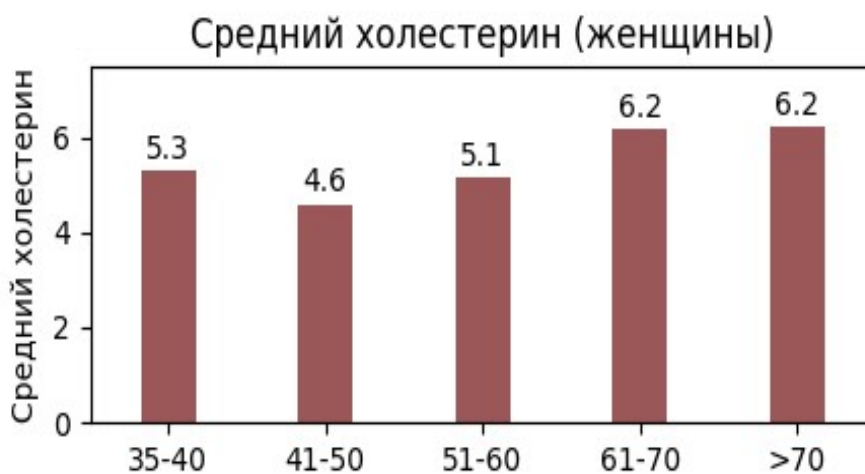


Рисунок А.9 – средний уровень холестерина у женщин



Рисунок А.10 – средний уровень холестерина у мужчин

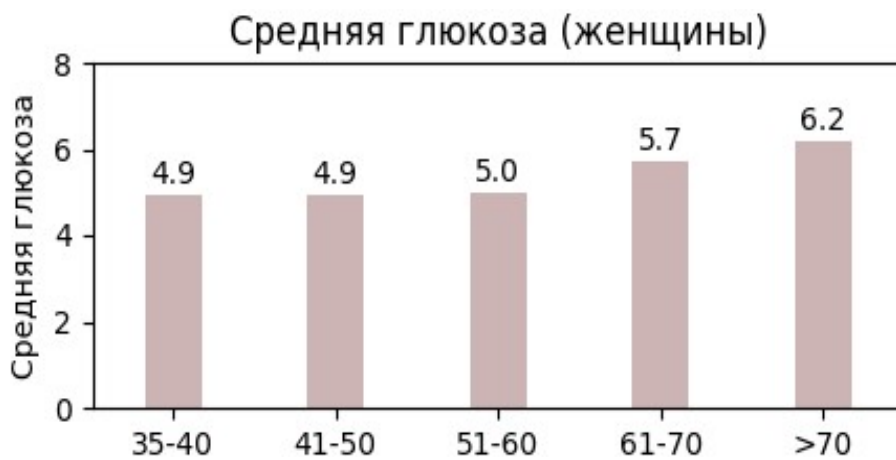


Рисунок А.11 – средний уровень глюкозы у женщин

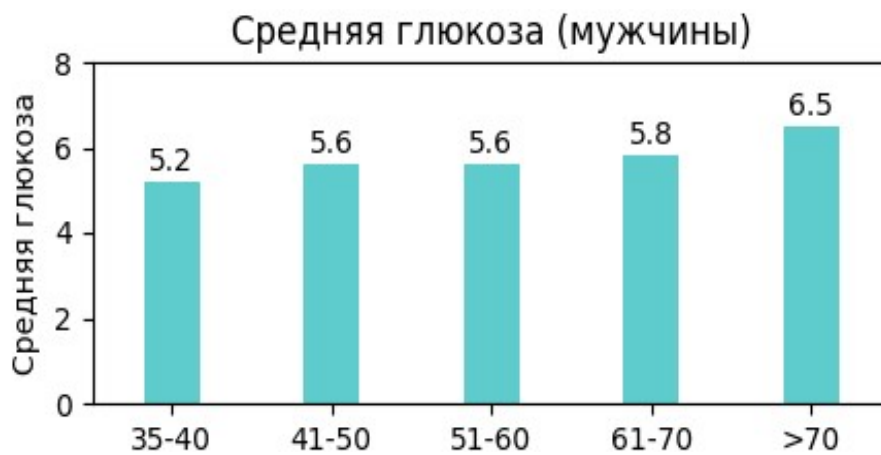


Рисунок А.12 – средний уровень глюкозы у мужчин