


Министерство науки и высшего образования Российской Федерации  
Федеральное государственное автономное образовательное учреждение  
высшего образования

«Уральский федеральный университет  
имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий - РТФ  
Кафедра информационных технологий и систем управления

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

Зав. кафедрой ИТиСУ  
  
Е.В. Кислицын  
(подпись) (Ф.И.О.)  
« 04 » 06 2024 г.

### ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

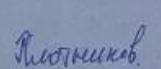
ИССЛЕДОВАНИЕ ИЗМЕНЕНИЯ СОСТОЯНИЯ ЗДОРОВЬЯ ЧЕЛОВЕКА  
НА ОСНОВЕ ДАННЫХ ЭКГ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО  
ОБУЧЕНИЯ

Научный руководитель: Борисов Василий Ильич  
к.т.н., доцент

  
подпись  
  
подпись

Нормоконтролер: Бредихина Наталья Сергеевна

Студент группы РИМ-220908: Плотников Григорий  
Александрович

  
подпись

Екатеринбург  
2024

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное автономное образовательное учреждение высшего  
образования  
«Уральский федеральный университет  
имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий - РТФ  
Кафедра информационных технологий и систем управления  
Направление подготовки 09.04.01 Информатика и вычислительная техника  
Образовательная программа Инженерия искусственного интеллекта

### ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студента Плотников Григорий Александрович группы РИМ-220908  
(фамилия, имя, отчество)

1. Тема выпускной квалификационной работы Исследование изменения состояния здоровья человека на основе данных ЭКГ с использованием машинного обучения

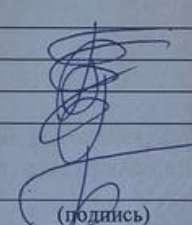
Утверждена распоряжением по институту от «4» декабря 2023 г. № 33.02-05/298

2. Научный руководитель Борисов Василий Ильич, к.т.н., доцент  
(Ф.И.О., должность, ученая степень, ученое звание)

3. Исходные данные к работе

4. Перечень демонстрационных материалов презентация в MS PowerPoint

5. Календарный план

№ п/п	Наименование этапов выполнения работы	Срок выполнения этапов работы	Отметка о выполнении
1.	Глава 1. Анализ предметной области	до 23.03.2024 г.	 (подпись)
2.	Глава 2. Методы и инструменты	до 29.04.2024 г.	
3.	ВКР в целом	до 20.05.2024 г.	

Научный руководитель

Борисов В.И.  
Ф.И.О.

(подпись)

Студент задание принял к исполнению

12.01.2024  
дата

Плотников  
(подпись)

6. Допустить Плотникова Григория Александровича к защите выпускной квалификационной работы в экзаменационной комиссии

Зав. кафедрой ИТиСУ

  
(подпись)

Е.В. Кислицын  
Ф.И.О.

## РЕФЕРАТ

Плотников Г.А. Исследование изменения состояния здоровья человека на основе данных ЭКГ с использованием машинного обучения.

Пояснительная записка содержит 55 страниц, 25 рисунков, 6 таблицы.  
МАШИННОЕ ОБУЧЕНИЕ, ЗДОРОВЬЕ ЧЕЛОВЕКА, ДАННЫЕ ЭКГ, АНАЛИЗ ДАННЫХ, ГЛУБОКОЕ ОБУЧЕНИЕ, НЕЙРОННЫЕ СЕТИ, МЕДИЦИНСКАЯ ДИАГНОСТИКА.

Объект исследования – электрокардиограммы (ЭКГ) как источник информации о состоянии здоровья человека.

Предмет исследования – алгоритмы машинного обучения для анализа ЭКГ и выявления изменений в состоянии здоровья человека.

Цель работы – изучить возможность использования данных ЭКГ для анализа состояния здоровья человека с помощью методов машинного обучения.

Решены следующие задачи:

- 1) проведен предметной области;
- 2) проведен анализ используемого датасета;
- 3) выбран инструмент для разработки системы оценки состояния здоровья человека;
- 4) проведена оценка работы модели;
- 5) выявлены направления для будущей работы.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	7
1. Анализ предметной области.....	10
1.1 Основы Электрокардиограммы (ЭКГ) .....	10
1.1.1 Физический принцип ЭКГ .....	10
1.1.2 ЭКГ. Системы отведений.....	10
1.1.3 Информативные зубцы ЭКГ .....	13
1.2 Виды аритмий.....	15
1.3 Обзор существующих методов машинного обучения, применяемых к данным ЭКГ .....	16
1.4 Выводы .....	19
2. Материалы и методы.....	20
2.1 Обзор датасета .....	20
2.2 Цикл работы кода для выявления данных здоровый или больных людей .....	29
2.2.1 Загрузка данных .....	30
2.2.2 Предобработка данных.....	33
2.2.3 Выбор классификатора и обучение.....	33
2.3 Цикл работы модели для оценки данных ЭКГ .....	34
2.3.1 Загрузка и предобработка данных .....	34
2.3.2 Создание модели .....	37
2.3.3 Обучение модели .....	39
2.3.4 Проверка модели на тестовом наборе данных.....	40
3. Обсуждение результатов.....	42
3.1 Результаты работы модели для оценки данных ЭКГ.....	42
3.2 Результаты работы модели для выявления данных здоровый или больных людей .....	43
ЗАКЛЮЧЕНИЕ .....	47

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ ..... 49

## СПИСОК СОКРАЩЕНИЙ

CNN (Convolutional Neural Network) – свёрточная нейронная  
RNN (Recurrent neural network) – рекуррентная нейронная сеть  
SVM (support vector machine) – метод опорных векторов  
ML (Machine Learning) – машинное обучение  
ЭКГ – электрокардиография  
СВПС – суправентрикулярное преждевременное сокращение  
СПБ – суправентрикулярное преждевременное биение  
ПСЖ – преждевременное сокращение желудочков  
XGBoost – eXtreme Gradient Boosting

## ВВЕДЕНИЕ

Во всем мире заболевания сердечно-сосудистой системы (ССС) человека до сих пор являются главной причиной смертности и инвалидности населения[3]. В последние два года наблюдается тенденция по увеличению частоты и тяжести таких заболеваний. Согласно данным Росстата, России ежегодно регистрируется около 900 тысяч случаев смерти, связанных с заболеваниями сердечно-сосудистой системы. Почти 46% всех смертей в России можно отнести к болезням сердца. В этой стране самый высокий уровень смертности от сердечно-сосудистых заболеваний по сравнению с любой другой развитой страной мира.

В настоящее время при оценке заболеваний сердечно-сосудистой системы по ЭКГ могут возникать ошибки, связанные с неправильным интерпретированием результатов, таких как:

1) Пропуск важных деталей, врач может пропустить важные изменения на ЭКГ, если он не обращает должного внимания на все аспекты записи.

2) Неправильная интерпретация: иногда изменения на ЭКГ могут быть сложными для интерпретации, особенно если они находятся вне нормы. В таких случаях врач может сделать неправильные выводы.

3) Плохая квалификация врача.

Эта проблема сердечно-сосудистых заболеваний стала важной для всей страны из-за высокого уровня заболеваемости, инвалидизации и смертности, а также из-за долгосрочного медикаментозного лечения, которое требует значительных затрат. В связи с этим растёт потребность обращать больше внимания на превентивные меры по предотвращению этих заболеваний. Отсюда вытекает объект исследования, электрокардиограммы как источник информации о состоянии здоровья человека.

Для сокращения риска возникновения опасных сердечных аритмий необходимо улучшение систем диагностики и обработки ЭКГ. Предоставление своевременной медицинской помощи играет решающую роль

в сохранении жизни и здоровья пациентов, а также в снижении инвалидности и смертности.

В последние годы машинное обучение стало важным инструментом для анализа данных и решения различных задач. Методы машинного обучения позволяют автоматически извлекать закономерности и шаблоны из данных. Это будет полезно для оценки когнитивных функций на основе данных ЭКГ.

На данный момент проведено несколько исследований, связанных с исследованием изменения состояния здоровья человека на основе данных ЭКГ с использованием методов машинного обучения. Некоторые из них являются крайне успешными, они показывают высокую точность и эффективность. Но существуют исследования, которые указывают на ограничения и сложность применения данного метода. Исследования существующих решений помогут определить проблемы, требующие дальнейшего изучения, а также предложить новые подходы и методы, применимые в данной области.

Отсюда следует актуальность исследования сигналов электрокардиографии (ЭКГ) для разработки системы исследования изменения состояния здоровья человека на основе методов машинного обучения (МО). Интеграция современных технологий в процесс выявления сердечно-сосудистых заболеваний по ЭКГ, поскольку оно может помочь автоматизировать процесс анализа данных и улучшить точность диагностики.

Предметом нашего исследования являются алгоритмы машинного обучения, которые могут анализировать большие объемы данных ЭКГ и выявлять определенные паттерны или аномалии, которые могут указывать на наличие сердечно-сосудистых заболеваний. Эти алгоритмы могут быть обучены на больших наборах данных ЭКГ, чтобы научиться распознавать различные типы заболеваний и определять их вероятность на основе характеристик ЭКГ.

Одно из преимуществ использования МО заключается в том, что оно может помочь врачам быстрее и точнее диагностировать сердечно-сосудистые заболевания. Вместо того, чтобы вручную анализировать каждую ЭКГ, врачи

могут использовать алгоритмы МО для автоматического анализа данных и предоставления им рекомендаций относительно возможных диагнозов, а также МО может помочь врачам лучше понять, какие характеристики ЭКГ наиболее важны для диагностики определенных типов сердечно-сосудистых заболеваний. Это может помочь врачам сосредоточить свое внимание на наиболее значимых аспектах ЭКГ и улучшить точность диагностики.

Целью данного исследования является изучить возможность использования данных ЭКГ для анализа состояния здоровья человека с помощью методов машинного обучения. Отсюда вытекают следующие задачи. Провести анализ существующих решений, выбрать набор данных ЭКГ для выполнения поставленной цели, выбрать инструмент для разработки системы оценки состояния здоровья человека, обучить и сравнить различные архитектуры моделей. Также необходимо провести анализ результатов.

## **1. Анализ предметной области**

### **1.1 Основы Электрокардиограммы (ЭКГ)**

#### **1.1.1 Физический принцип ЭКГ**

Электрокардиограмма (ЭКГ) – представляет собой графическое описание электрической активности сердца, возникающая в результате сокращения сердечной мышцы. Она является ключевым инструментом для диагностики и оценки работы сердца. ЭКГ регистрирует электрические сигналы, генерируемые сердцем, на поверхности тела с помощью электродов, помещенных на грудную клетку, запястье и голень. Источником электрической активности сердца служат работающие, сокращающиеся клетки миокарда, а также специальные клетки, обладающие автоматизмом[8].

Временной характеристикой ЭКГ является график, называемый электрокардиограммой. Он представляет собой запись отклонений величины электрического потенциала на поверхности тела в зависимости от времени. Он состоит из нескольких характеристических волн и комплексов, которые отражают различные стадии сердечного цикла[30].

#### **1.1.2 ЭКГ. Системы отведений**

Первоначальная система отведений, предложенная Эйнтховеном, основывается на предположениях о том, что тело — это гомогенный объемный проводник; все отведения симметричны; в центре объемного проводника располагается единственный эквивалентный диполь[17]. Стандартные отведения от конечностей (I, II, III) складываются из трех комбинаций электродов правой руки (ПР), левой руки (ЛР) и левой ноги (ЛН), как представлено на рисунке 1.

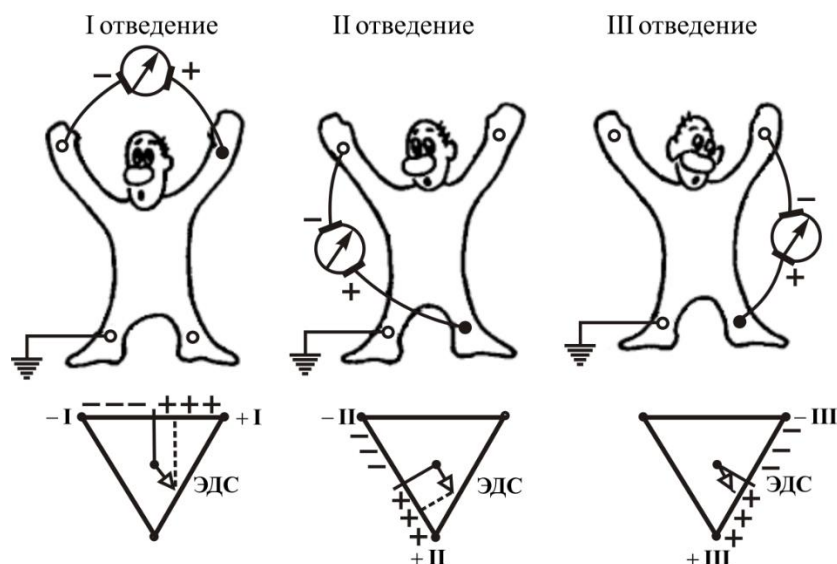


Рисунок 1 – Стандартные системы отведений

I отведение регистрирует разницу потенциалов между ЛР и ПР. При этом электрод на ЛР положительный, а на ПР – отрицательный[42].

II отведение отражает разность потенциалов между электродами ПР и ЛН, где положительным является электрод на ЛН[16].

III отведении фиксируется разность потенциалов между ЛР и ЛН, причем электрод ЛН также положительный. Видимо, Эйнтховен произвольно выбрал соотношения между положительными и отрицательными электродами в этих трех отведениях, руководствуясь лишь тем, чтобы комплекс QRS (см. ниже) был направлен вверх у большинства здоровых людей[18].

Отведения I, II и III позволяют оценить направление распространения электрического импульса в сердце и выявить возможные аномалии или изменения в работе сердца. Эти отведения являются основой для интерпретации ЭКГ и помогают врачам диагностировать различные сердечные состояния[34].

Также существует классическая система отведений, или иначе стандартных грудных отведений (V1-V6). Они представляют собой размещение электродов на грудной клетке пациента, как представлено на

рисунке 2. Каждое отведение позволяет регистрировать электрическую активность сердца с определенной точки зрения[39].

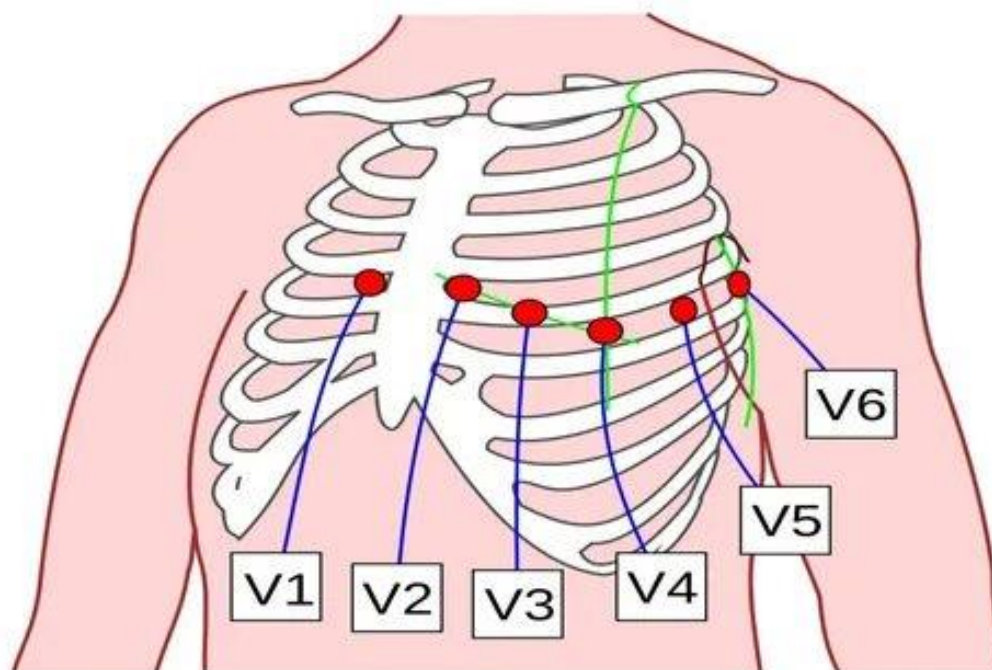


Рисунок 2 – Система грудных отведений[1]

Отведения V1-V6 располагаются следующим образом:

- 1) V1. Размещается на четвертом межреберье правого края грудины.
- 2) V2. Размещается на четвертом межреберье левого края грудины.
- 3) V3. Размещается между V2 и V4.
- 4) V4. Размещается на пятом межреберье по левой срединно-ключичной линии.
- 5) V5. Размещается на уровне V4 по передней подмышечной линии.
- 6) V6. Размещается на уровне V4 по передней среднеключичной линии.

Отведения V1-V6 позволяют оценить электрическую активность сердца с различных точек грудной клетки, что помогает врачам выявлять изменения в работе сердца и диагностировать сердечные заболевания[42].

### 1.1.3 Информативные зубцы ЭКГ

Основная функция сердца - перекачивать насыщенную кислородом кровь по всему организму. Эта перекачка осуществляется за счет сокращения и расслабления миокардиальной мышцы, находящейся в сердце. Синоатриальный (SA) узел генерирует электрические импульсы, которые отвечают за повторяющиеся сокращения и расслабления камер сердца[29]. Эти электрические импульсы можно обнаружить и измерить, поместив электроды в определенные места на нашем теле с помощью аппарата ЭКГ. Каждый импульс, регистрируемый аппаратом ЭКГ как волна P-QRS-T, имеет определенную амплитуду, продолжительность, форму и вид[6]. Последовательность волн P-QRS-T, собранных аппаратом ЭКГ, переводится в графическое представление и распечатывается на бумаге. Эти распечатанные данные могут быть интерпретированы врачом для выявления нерегулярных сердечных сокращений или аномалий сердца[45]. На рисунке 3 показан сигнал ЭКГ с различными компонентами.

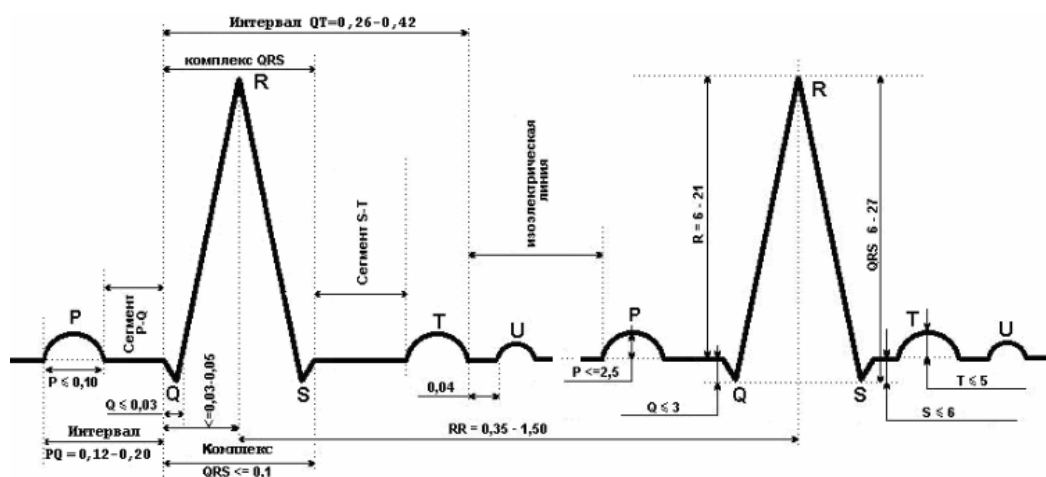


Рисунок 3 – Зубцы электрокардиограммы, интервалы и сегменты.

Волна P представляет собой деполяризацию предсердий. Благодаря этому артерии сокращаются, и в результате левое предсердие выталкивает

бедную кислородом кровь в левый желудочек; правое предсердие выталкивает богатую кислородом кровь, собранную из легких, в левый желудочек. Интервал PR обозначает прохождение электрического импульса от обеих артерий к АВ-узлу; сигнал замедляется и приостанавливается на короткий период, чтобы кровь из предсердий заполнила желудочки. - Комплекс QRS указывает на деполяризацию желудочков. Благодаря этому левый желудочек перекачивает бедную кислородом кровь в легкие, а правый желудочек перекачивает насыщенную кислородом кровь в части тела. - Сегмент ST отражает сокращение желудочка, но электричество по нему не проходит. - Волна T указывает на реполяризацию обоих желудочков, которая заставляет их расслабиться и подготовиться к следующему циклу перекачки крови[4].

Важные характеристики нормальной ЭКГ включают такие присущие ей характеристики, как амплитуда и продолжительность, которые приведены в таблице 1[32].

Таблица 1 – Важные характеристики нормальной ЭКГ

Характеристика ЭКГ	Описание
Р Волна	Представляет собой деполяризацию предсердий. Амплитуда: 0,1-0,12 мВ. Продолжительность: 0,08 - 0,1 с
Комплекс QRS	Представляет собой деполяризацию желудочков. Амплитуда: 1 мВ. Продолжительность: 0,08 - 0,12 с
Волна Т	Представляет собой реполяризацию желудочков. Амплитуда: 0,1 - 0,3 мВ. Продолжительность: 0,12 - 0,16 с
Р-QRS-Т	Указывает на один механический удар сердца. Продолжительность: 0,6 - 1 с
Интервал PR	Отслеживает импульс от предсердий через А-V узел. Продолжительность: 0.12-0.2 s

## Продолжение таблицы 1 – Важные характеристики нормальной ЭКГ

Интервал RR	Это время, прошедшее между двумя последовательными R-волнами. Продолжительность: 0,6 - 1 с
Интервал QT	Это время, необходимое для завершения всего желудочкового цикла. Продолжительность: 0,36 - 0,42 с
ST-сегмент	Представляет собой изоэлектрическую часть реполяризации желудочков. Продолжительность: 0,8 - 0,12 с

Оценка информативных зубцов на ЭКГ позволяет идентифицировать аномалии и нарушения в работе сердца. Однако, для точной диагностики и оценки сердечных состояний необходимо учитывать не только отдельные зубцы, но и их сочетания, отношения и динамику в течение времени. Поэтому важно проводить комплексный анализ всего ЭКГ с соблюдением стандартных протоколов и рекомендаций[2].

### 1.2 Виды аритмий

Различные виды аритмии возникают из-за сочетания аномальной частоты сердцебиения и аномального ритма. Ненормально повышенная частота сердечных сокращений может привести к обмороку и летальному исходу. Даже аномальный ритм сердца может привести к внезапной смерти. Поэтому определение правильного типа аритмии очень важно для повышения выживаемости. Тип аритмии можно определить по частоте сердечных сокращений, видимости Р-волны, соотношению между Р-волной и комплексом QRS, интервалу PR, ширине QRS и другим морфологическим характеристикам ЭКГ[41].

Некоторые виды аритмий являются опасными и требуют немедленного лечения, чтобы избежать внезапной смерти. Фибрилляция желудочков (ФЖ),

трепетание предсердий (ТП) и фибрилляция предсердий (ФП) являются наиболее часто встречающимися опасными для жизни аритмиями. Среди них большое количество смертей связано с фибрилляцией желудочков (ФЖ) и желудочковой тахикардией (ЖТ). Поскольку ЖТ и ФЖ визуализируются на ЭКГ одинаково, для их правильной идентификации необходим человеческий эксперт или модели машинного обучения. Трепетание предсердий (ТП) и фибрилляция предсердий (ФП) могут увеличить риск инсульта, если они сохраняются более нескольких недель без лечения[5]. Аритмия А-Фиб обычно встречается у пожилых людей из-за различных проблем со здоровьем. Она ассоциируется с инсультом, сердечной недостаточностью и смертностью. Блокада ветвей пучка (БВП) возникает из-за прерывания электрических импульсов в левой (LBBB) или правой (RBBB) ветвях пучка желудочков. LBBB или RBBB не лечатся напрямую, вместо этого необходимо лечить их причину, иначе они могут привести к другим серьезным заболеваниям сердца[19].

### **1.3 Обзор существующих методов машинного обучения, применяемых к данным ЭКГ**

Существует несколько методов машинного обучения, которые могут быть применены к данным ЭКГ для оценки когнитивных функций[31]. Некоторые из них включают:

#### **1) Классификация:**

– Метод опорных векторов (SVM): SVM использует гиперплоскость для разделения данных на разные классы[21]. Он может быть применен к данным ЭКГ для классификации испытуемых на основе их когнитивных состояний, например, наличия или отсутствия внимания или памяти.

– Случайный лес: Случайный лес — это ансамблевый метод, который комбинирует несколько деревьев решений для классификации данных. Он

может быть использован для классификации испытуемых на основе их ЭКГ-сигналов и предсказания когнитивных состояний[36].

## 2) Регрессия:

– Линейная регрессия: Линейная регрессия используется для предсказания непрерывных значений на основе набора входных данных[40]. Он может быть применен к данным ЭКГ для оценки уровня когнитивных функций, таких как внимание или память.

– Нейронные сети: Нейронные сети являются моделями, которые воспроизводят работу нейронов в мозге[35]. Они могут быть использованы для извлечения признаков из данных ЭКГ и предсказывать уровень когнитивных функций[22].

## 3) Кластеризация:

– Метод К-средних: Метод К-средних используется для разделения данных на заданное количество кластеров. Он может быть применен к данным ЭКГ для выявления групп испытуемых с похожими когнитивными характеристиками.

– DBSCAN: DBSCAN — это алгоритм кластеризации, который основывается на плотности данных. Он может быть применен к данным ЭКГ для выявления кластеров испытуемых с похожими когнитивными характеристиками[25].

## 4) Глубокое обучение:

– Сверточные нейронные сети (CNN): CNN являются типом нейронных сетей, которые могут эффективно анализировать данные, имеющие пространственную структуру, такие как изображения или сигналы[24]. Они могут быть применены к данным ЭКГ для извлечения признаков и предсказания когнитивных состояний[27].

– Рекуррентные нейронные сети (RNN): RNN — это тип нейронных сетей, которые способны обрабатывать последовательные данные. Они могут

быть применены к данным ЭКГ для анализа последовательности электрических сигналов и предсказания когнитивных состояний[15].

Рассмотрим некоторые используемые методы машинного обучения для классификации ЭКГ. В таблице 2 представлены работы, в которых происходит классификация ЭКГ, метод машинного обучения, и оценка F1 score[28].

Таблица 2 – Сравнения работ по классификации ЭКГ

Название статьи	Метод машинного обучения	F1 score
Application of Support Vector Machine Classifier on Developed Wireless ECG System	SVM	F1 score = 0.85
ECG-based pulse detection during cardiac arrest using random forest classifier	Random Forest	F1 score = 0.82
ECG Identification For Personal Authentication Using LSTM-Based Deep Recurrent Neural Networks	Neural Networks	F1 score = 0.88

Из вышеперечисленных работ можно сделать вывод, что:

1. SVM используется для разделения данных ЭКГ на различные классы, например, нормальное состояние сердца и аритмия. Он имеет хорошую точность и полноту при классификации различных типов аномалий на основе данных ЭКГ.

2. Случайный лес является эффективным методом для классификации данных ЭКГ и выявления аномалий. Он позволяет работать с большим количеством признаков и обеспечивает хорошую точность и полноту при классификации различных состояний сердца[38].

3. Нейронные сети позволяют обучать модели на больших объемах данных ЭКГ и выявлять сложные паттерны и аномалии. Они обладают высокой точностью и способностью к обобщению, что делает их эффективными в задачах классификации ЭКГ.

Это лишь некоторые из методов машинного обучения, которые могут быть применены к данным ЭКГ для оценки когнитивных функций. От выбора конкретного метода зависит от целей и характеристик исследования. Важно учитывать, что качество результатов будет зависеть от качества данных ЭКГ, а также от правильного выбора и настройки метода машинного обучения.

#### **1.4 Выводы**

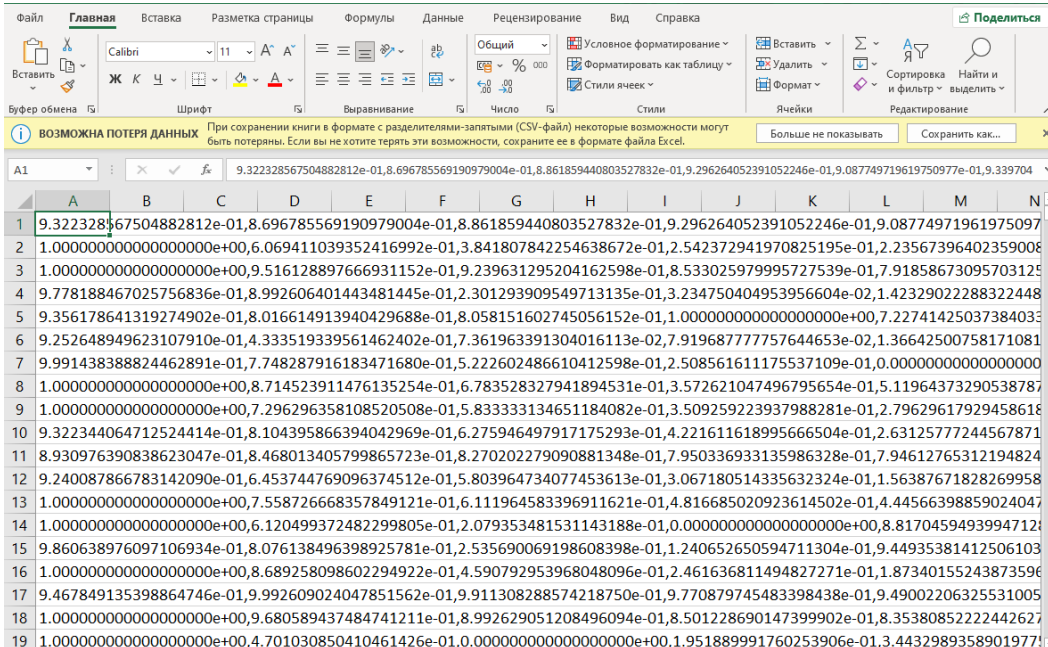
В данной главе дипломной была рассмотрена актуальность проблемы оценки сердечно-сосудистых заболеваний человека. Был сделан обзор на основную литературу по данной теме, а также проведен обзор используемых методов и подходов. Можно сделать вывод о важности изучения сердечно-сосудистых на основе данных ЭКГ и машинного обучения для предсказания и контроля этого состояния. В следующей главе будут рассмотрены данные ЭКГ и их применение в работе.

## 2. Материалы и методы

### 2.1 Обзор датасета

Набор данных состоит из двух коллекций сигналов сердцебиения, полученных из двух известных наборов данных по классификации сердцебиения: набора данных по аритмии MIT-BIH и базы данных диагностической ЭКГ РТВ[37]. Количество выборок в обеих коллекциях достаточно велико для обучения глубокой нейронной сети.

Графики ЭКГ были разбит на точки и сохранены в виде csv файла. Каждая строка соответствует отдельно взятому графику ЭКГ, в каждой строке через запятую перечислены числа. Первые 187 чисел соответствуют точке графика в определенный промежуток времени. Например, первое число в строке относится к первой секунде графика, второе ко второй и так далее. 188 число в каждой строке является пометкой к классу, к которому относится график. Пример записи графиков ЭКГ и их классов в csv файле представлен на рисунке 4.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	9.322328567504882812e-01,8.696785569190979004e-01,8.861859440803527832e-01,9.296264052391052246e-01,9.08774971961975097													
2	1.0000000000000000e+00,6.069411039352416992e-01,3.841807842254638672e-01,2.542372941970825195e-01,2.23567396402359008													
3	1.0000000000000000e+00,9.516128897666931152e-01,9.239631295204162598e-01,8.533025979995727539e-01,7.91858673095703125													
4	9.778188467025756836e-01,8.992606401443481445e-01,2.301293909549713135e-01,3.234750404953956604e-02,1.42329022288322448													
5	9.356178641319274902e-01,8.016614913940429688e-01,8.058151602745056152e-01,1.0000000000000000e+00,7.22741425037384033													
6	9.252648949623107910e-01,4.333519339561462402e-01,7.361963391304016113e-02,7.919687777757644653e-02,1.36642500758171081													
7	9.991438388824462891e-01,7.748287916183471680e-01,5.222602486610412598e-01,2.508561611175537109e-01,0.0000000000000000													
8	1.0000000000000000e+00,8.714523911476135254e-01,6.783528327941894531e-01,3.572621047496795654e-01,5.11964373290538787													
9	1.0000000000000000e+00,7.296296358108520508e-01,5.833333334651184082e-01,3.509259223937988281e-01,2.79629617929458618													
10	9.322344064712524414e-01,8.104395866394042969e-01,6.275946497917175293e-01,4.221611618995666504e-01,2.6312577244567871													
11	8.930976390838623047e-01,8.468013405799865723e-01,8.270202279090881348e-01,7.950336933135986328e-01,7.94612765312194824													
12	9.240087866783142090e-01,6.453744769096374512e-01,5.803964734077453613e-01,3.067180514335632324e-01,1.56387671828269958													
13	1.0000000000000000e+00,7.558726668357849121e-01,6.111964583396911621e-01,4.816685020923614502e-01,4.44566398859024047													
14	1.0000000000000000e+00,6.120499372482299805e-01,2.079353481531143188e-01,0.0000000000000000e+00,8.8170459493994712													
15	9.860638976097106934e-01,8.076138496398925781e-01,2.53560069198608398e-01,1.240652650594711304e-01,9.44935381412506103													
16	1.0000000000000000e+00,8.689258098602294922e-01,4.590729253968048096e-01,2.461636811494827271e-01,1.87340155243873596													
17	9.467849135398864746e-01,9.992609024047851562e-01,9.911308288574218750e-01,9.770879745483398438e-01,9.49002206325531005													
18	1.0000000000000000e+00,9.680589437484741211e-01,8.992629051208496094e-01,8.501228690147399902e-01,8.35380852222442627													
19	1.0000000000000000e+00,4.701030850410461426e-01,0.0000000000000000e+00,1.951889991760253906e-01,3.44329893589019771													

Рисунок 4 – Запись графиков ЭКГ и их классов в csv файле

База данных диагностических ЭКГ РТВ (Physionet Taiwan Big Data-ECGVB) сигналов используется для выявления здорового или больного человека. Она включает в себя записи ЭКГ от пациентов, которые были диагностированы с различными сердечно-сосудистыми заболеваниями.

База данных РТВ содержит 14552 записей ЭКГ, из них 10506 записей ЭКГ, принадлежащих здоровым людям (Normal), и 4046 записи ЭКГ людей с какими-либо заболеваниями (Abnormal). Каждая запись которых длится примерно 10 минут и состоит из 15 одновременно измеряемых сигналов, условные 12 отведений (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6 ) вместе с 3 ЭКГ в отведениях Франка (vx, vy, vz). Так же каждая запись в базе данных РТВ сопровождается метаданными, включая возраст пациента, пол, диагноз и другие клинические характеристики. Это позволяет исследователям анализировать данные ЭКГ в контексте клинических условий и характеристик пациентов. Примеры графиков Normal и Abnormal сигналов изображены на рисунке 5.

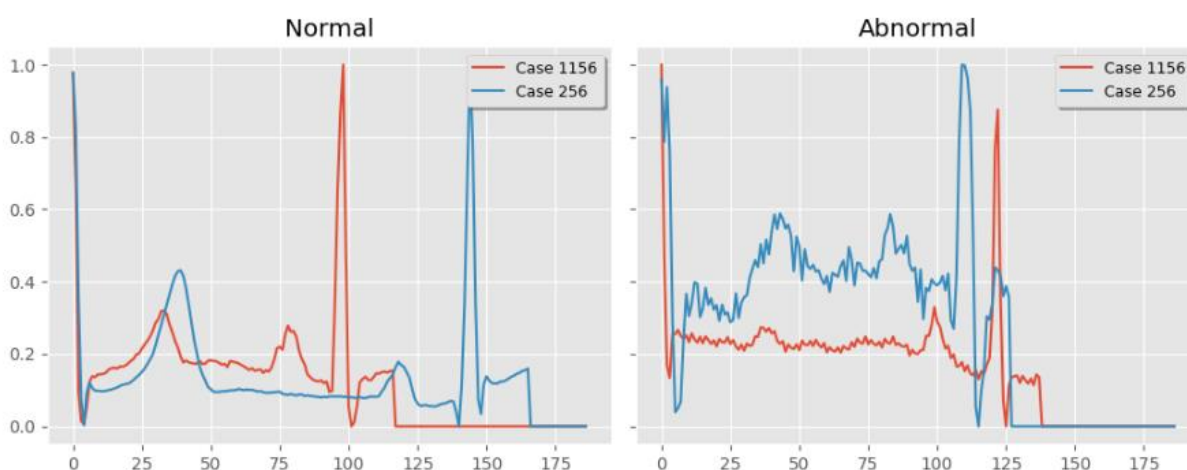


Рисунок 5 – Примеры случайно выбранных записей ЭКГ с пометкой Normal и Abnormal

База данных аритмии MIT-BIH (Boston's Beth Israel Hospital) — это коллекция записей ЭКГ, собранных в рамках проекта по изучению аритмий.

Она содержит 48-часовые записи ЭКГ от 23 пациентов, которые были госпитализированы из-за симптомов, связанных с аритмиями.

Записи были оцифрованы со скоростью 360 выборок в секунду на канал с 11-битным разрешением в диапазоне 10 мВ[44]. В итоге было получено около 110000 аннотаций собранных в единую базу данных. База данных включает в себя 2 канала ЭКГ (отведения I и II). Все записи в базе данных MIT-BIH сопровождаются метаданными, включая информацию о пациенте, такую как возраст, пол, диагноз и другие клинические характеристики. Кроме того, каждая запись также содержит метки времени, соответствующие каждому сердечному циклу, что позволяет исследователям анализировать данные ЭКГ в контексте времени.

110000 аннотаций были поделены на обучающую (train) и тестовую (test) выборки. На рисунке 6 представлены графики распределения данных в обеих выборках.

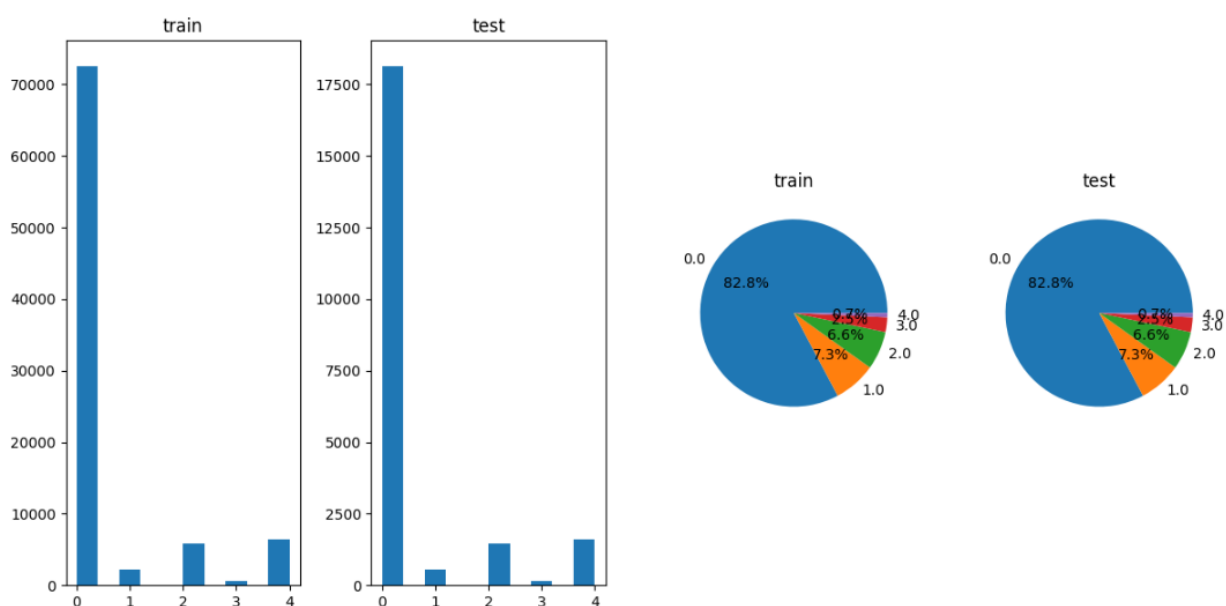


Рисунок 6 – Графики распределения данных на train и test выборках

База данных аритмии MIT-BIH содержит классы, представленные в таблице 3.

Таблица 3 – Классы датасета и их описание.

Класс	Обозначение класса	Описание класса
N	0	Нормальный ритм
S	1	Суправентрикулярное преждевременное биение
V	2	Преждевременное сокращение желудочков
F	3	Слияние желудочкового и нормального ритма
Q	4	Не классифицируемый ритм

Давайте более подробно рассмотрим каждый из них.

1) Нормальный ритм.

Нормальный ритм ЭКГ, также известный как синусовый ритм, это когда сердцебиение контролируется синусовым узлом, который является естественным кардиостимулятором сердца. Этот ритм характеризуется регулярностью и стабильной частотой сердечных сокращений. В норме частота сердечных сокращений составляет от 60 до 100 ударов в минуту. При синусовом ритме зубец P всегда предшествует комплексу QRS и имеет нормальную форму и продолжительность[26]. Комплекс QRS также имеет нормальную форму и продолжительность, без признаков патологических изменений, как представлено на рисунке 7[20].

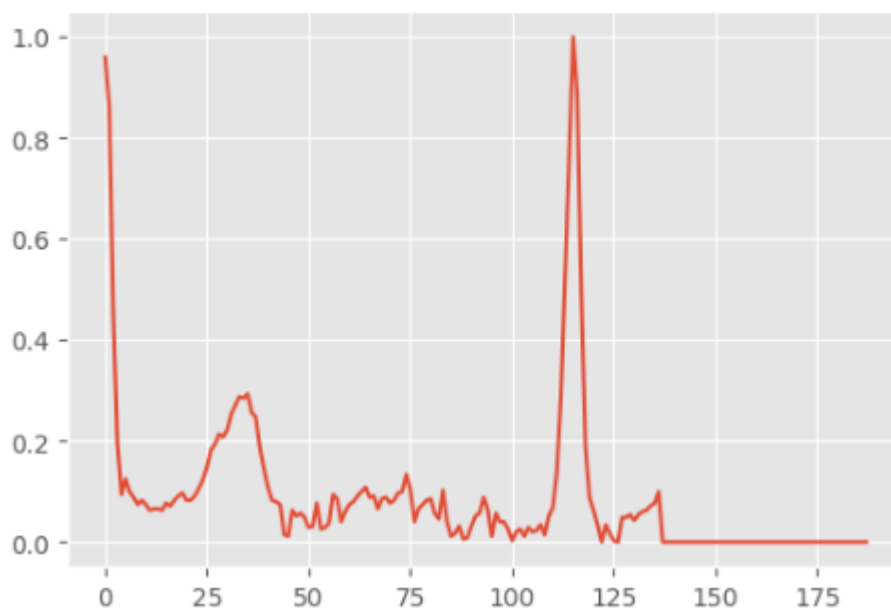


Рисунок 7 – Нормальный график ЭКГ

## 2) Суправентрикулярное преждевременное биение.

Суправентрикулярное преждевременное биение (СПБ) — это нарушение сердечного ритма, при котором возникают дополнительные электрические импульсы в предсердиях, вызывая дополнительные сокращения сердца[10].

Это явление может проявляться на электрокардиограмме (ЭКГ) следующим образом:

а) Ранний комплекс QRS: В случае СПБ, комплекс QRS (отражающий электрическую активность желудочков) появляется раньше ожидаемого, так как возникает дополнительное электрическое возбуждение в верхних камерах сердца перед нормальным сокращением[43].

б) Форма комплекса QRS: Комплекс QRS при СПБ часто имеет измененную форму по сравнению с нормальными комплексами QRS, так как импульс может пройти через альтернативный путь проведения, что может изменить форму QRS комплекса на ЭКГ.

в) Пауза после СПБ: После СПБ следует пауза, которая короче обычного интервала RR (времени между двумя последовательными зубцами R). Это связано с тем, что сердце пропускает сокращение, и следующему сокращению требуется больше времени, чтобы произойти.

d) Удлинение интервала PR: Интервал PR — это время между началом зубца P и началом комплекса QRS. При СПБ этот интервал может быть увеличен, так как дополнительный импульс заставляет сердце сокращаться раньше, чем ожидалось[13].

Пример графика с суправентрикулярным преждевременным биением можно наблюдать на рисунке 8.

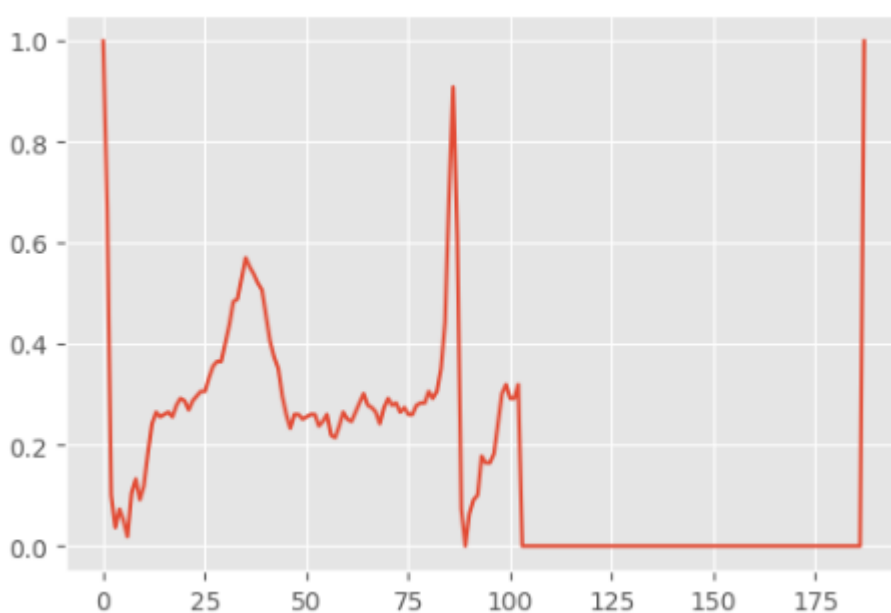


Рисунок 8 – График ЭКГ с суправентрикулярным преждевременным биением

### 3) Преждевременное сокращение желудочков.

Преждевременное сокращение желудочков — это состояние, при котором желудочки сердца начинают сокращаться раньше обычного. Это может привести к неправильной работе сердца, нарушению его ритма и, в конечном итоге, к сердечной недостаточности[11]. Преждевременные сокращения желудочков могут быть обнаружены при проведении электрокардиограммы (ЭКГ) или других методов кардиологического обследования[12].

Это явление может проявляться на электрокардиограмме (ЭКГ) следующим образом:

а) Ранний комплекс QRS: При ПСЖ на ЭКГ можно увидеть ранний комплекс QRS, который отражает раннюю активацию желудочков. Этот комплекс может иметь измененную форму или продолжительность по сравнению с нормальными комплексами QRS.

б) Пауза после ПСЖ: После ПСЖ следует пауза, которая длиннее обычного интервала RR (времени между двумя последовательными зубцами R). Это связано с тем, что сердце пропускает сокращение, и следующему сокращению требуется больше времени, чтобы произойти.

в) Короткая дистанция от предыдущего QRS: После ПСЖ обычно следует пауза до следующего нормального сокращения сердца. На ЭКГ это может отразиться как короткая дистанция между предыдущим QRS-комплексом и ранним комплексом QRS[33].

г) Укорочение интервала PR: Интервал PR - это время между началом зубца P и началом комплекса QRS. При ПСЖ этот интервал может быть уменьшен, так как дополнительный импульс заставляет сердце сокращаться раньше, чем ожидалось.

Пример графика с преждевременным сокращением желудочка можно наблюдать на рисунке 9.

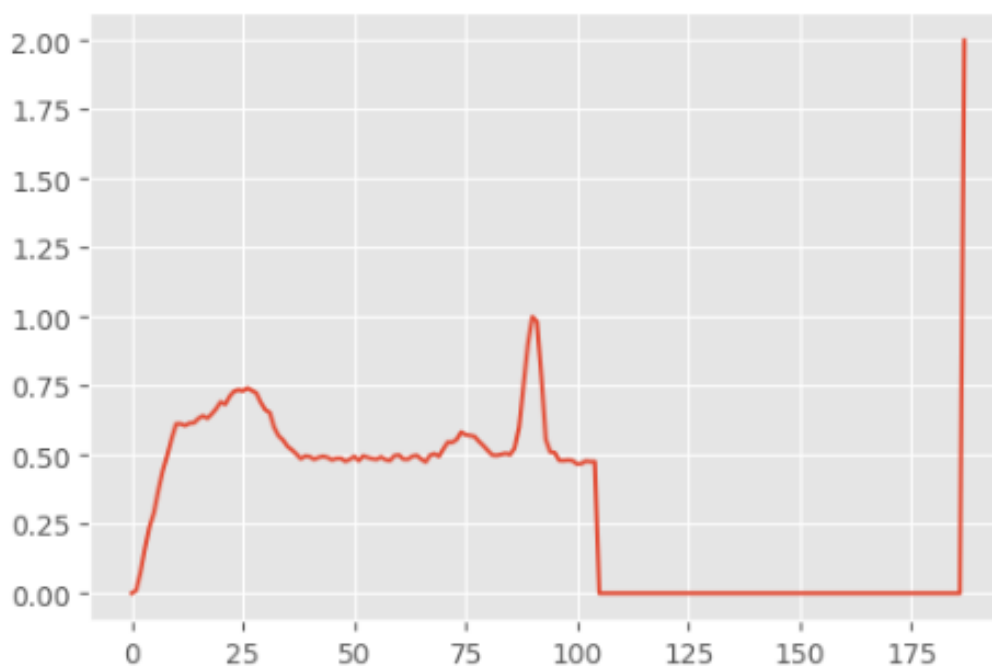


Рисунок 9 – График ЭКГ с преждевременным сокращением желудочка

#### 4) Слияние желудочкового и нормального ритма.

Слияние желудочкового и нормального ритма — это патологическое состояние, при котором одновременно регистрируются два различных ритма сердца: нормальный синусовый ритм и желудочковые экстрасистолы. Такое состояние называется парасистолией[7].

На электрокардиограмме (ЭКГ) слияние желудочкового и нормального ритма может проявляться следующим образом:

а) Комплекс QRS: Обычно на ЭКГ комплекс QRS представляет собой характерные колебания, отражающие электрическую активность желудочков сердца. При слиянии желудочкового и нормального ритма комплекс QRS может иметь необычную форму, которая сочетает черты как желудочкового, так и нормального ритма.

б) Интервал RR: Интервалы между последовательными комплексами QRS (интервалы RR) могут быть неправильной длины или иметь нерегулярный характер из-за наложения разных ритмов.

с) Частота сердечных сокращений: Слияние желудочкового и нормального ритма может привести к изменению частоты сердечных сокращений, что также может быть отражено на ЭКГ[23].

д) Пик Т: Форма пика Т на ЭКГ также может измениться при слиянии ритмов, хотя это не всегда является характерным признаком.

е) Другие признаки: в некоторых случаях на ЭКГ также могут наблюдаться другие аномалии, связанные с наложением различных ритмов.

Пример графика слияния желудочкового и нормального ритма можно наблюдать на рисунке 10.

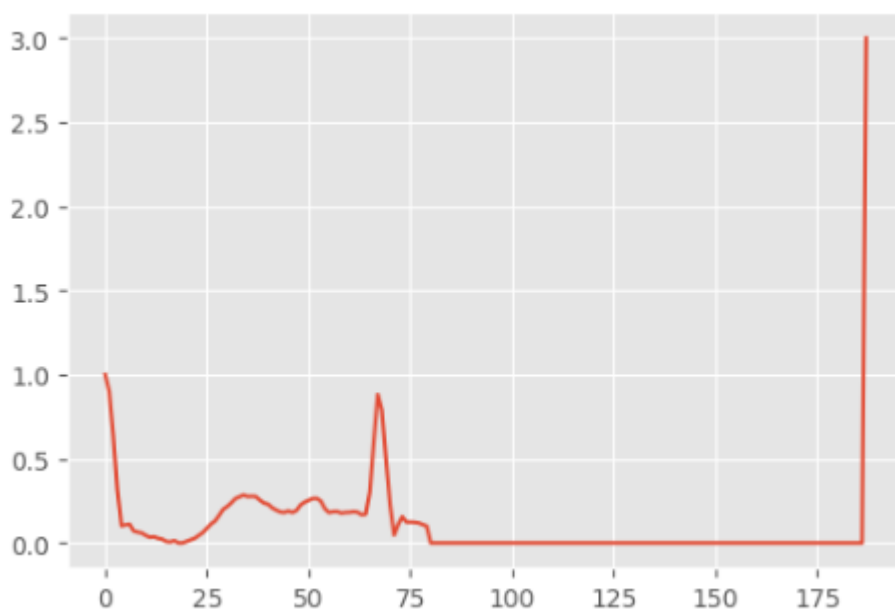


Рисунок 10 – График ЭКГ слияния желудочкового и нормального ритма

#### 5) Неклассифицируемый ритм[14].

Неклассифицируемый ритм на электрокардиограмме (ЭКГ) означает, что обнаруженный ритм не соответствует типичным шаблонам или классификациям сердечных ритмов. Это может быть вызвано различными факторами, такими как наличие аномалий в проведении сердечного импульса, электрической активности или другими необычными характеристиками сердечного ритма.

Неклассифицируемый ритм может быть вызван смешиванием нескольких типов аритмий, наложением разных ритмов, наличием электрических помех или другими факторами, усложняющими идентификацию конкретного типа ритма на ЭКГ[9].

Для точного определения и интерпретации неклассифицируемого ритма на ЭКГ может потребоваться более глубокий анализ и консультация специалиста по сердечным аритмиям или кардиолога. Такие случаи требуют дополнительных исследований и экспертизы для определения причин и последующего лечения, если необходимо.

Пример графика с неклассифицируемым ритмом можно наблюдать на рисунке 11.

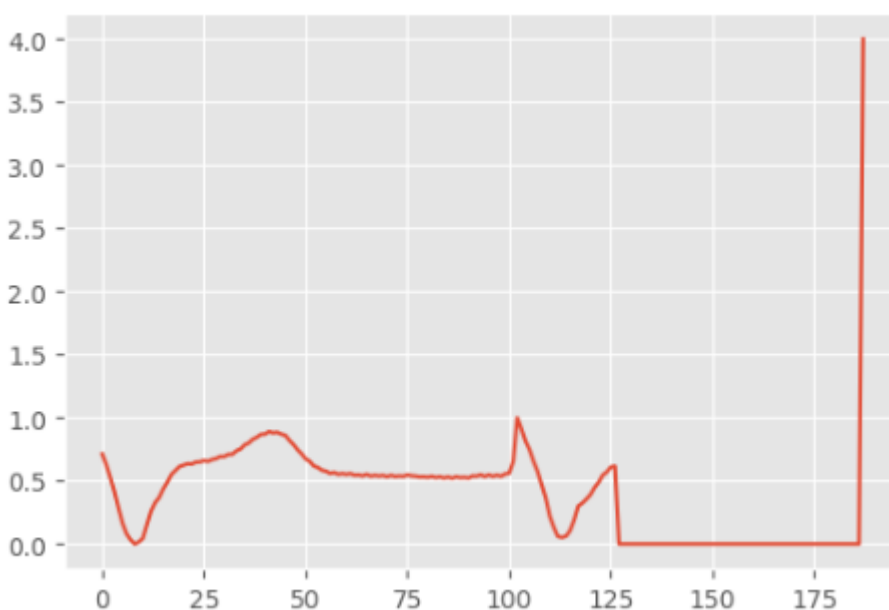


Рисунок 11 – График ЭКГ неклассифицируемым ритмом

## **2.2 Цикл работы кода для выявления данных здоровый или больных людей**

Цикл работы кода для выявления данных здоровых или больных людей по графику ЭКГ изображена на рисунке 12.

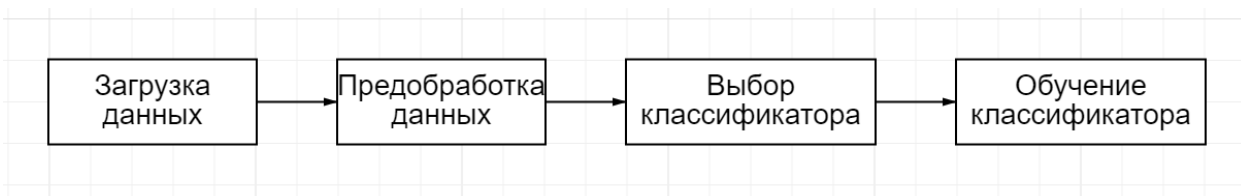


Рисунок 12 – Цикл работы кода для выявления данных здоровых или больных людей

Давайте более подробно рассмотрим каждый из этих этапов.

### 2.2.1 Загрузка данных

Как упоминалось ранее, была выбрана база данных диагностических ЭКГ РТВ (Physionet Taiwan Big Data-ECGBB). Он был заранее поделен на DataFrame `abnormal` и `normal`.

После загрузки данных, необходимо проверить, реально ли они соответствуют графикам ЭКГ. Для этого мы вывели графики для определенных строк из наших DataFrame `abnormal`. На графиках ось X показывает время, измеряемое в секундах, а ось Y – электрический потенциал, измеряемый в милливольтгах (мВ). То, что визуализировано из данных ЭКГ изображено на рисунке 13.

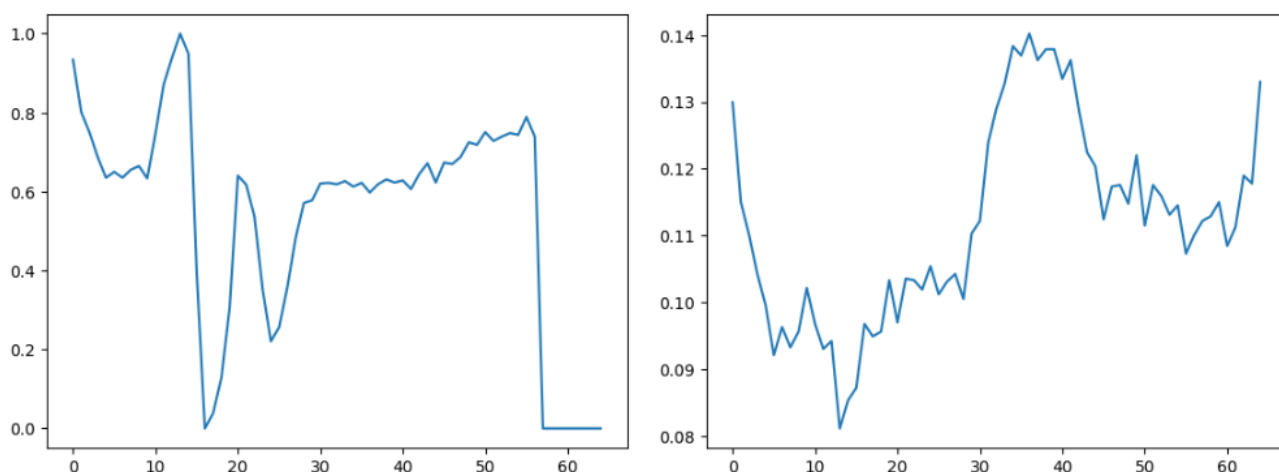


Рисунок 13 – Визуализированные данные ЭКГ `abnormal` DataFrame

Графики не особо похожи на нормальные графики ЭКГ, скорее на аномальные, которые не подлежат никакой оценке для пациента. Так же была построена цветовая гистограмма, изображенная на рисунке 14.

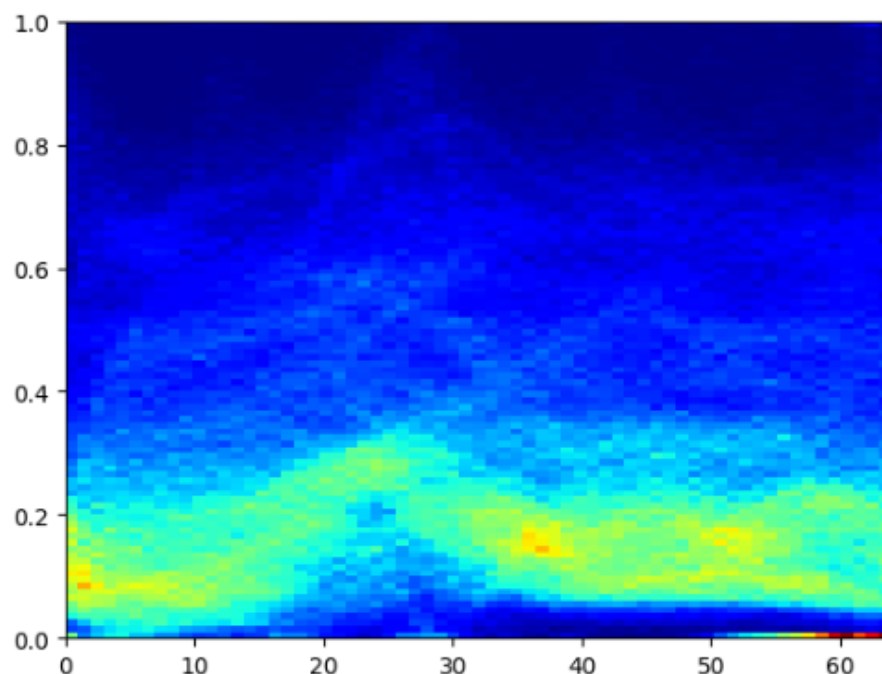


Рисунок 14 – Цветовая гистограмма распределения abnormal характеристик ЭКГ

Из приведенной выше цветовой карты гистограммы для данных РТВ, помеченных как аномальные, можно сделать вывод, что большинство характеристик ЭКГ широко распределены в диапазоне 0 - 0,4 и для этих данных не существует фиксированного шаблона.

Аналогичные действия были проведены с DataFrame normal.

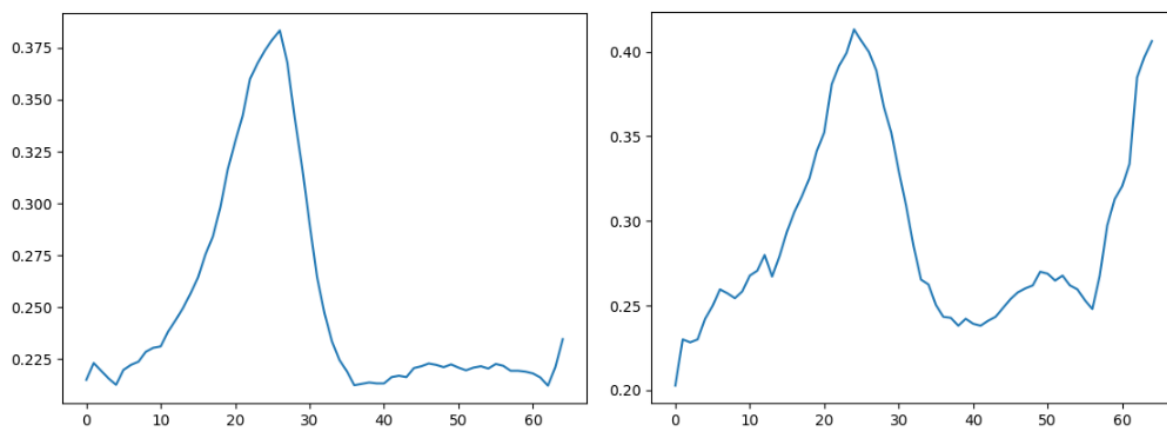


Рисунок 15 – Визуализированные данные ЭКГ normal DataFrame

Графики изображенные на рисунке 15 уже больше похожи на привычные на графики ЭКГ. Также рассмотрим цветовая гистограмма, изображенная на рисунке 16.

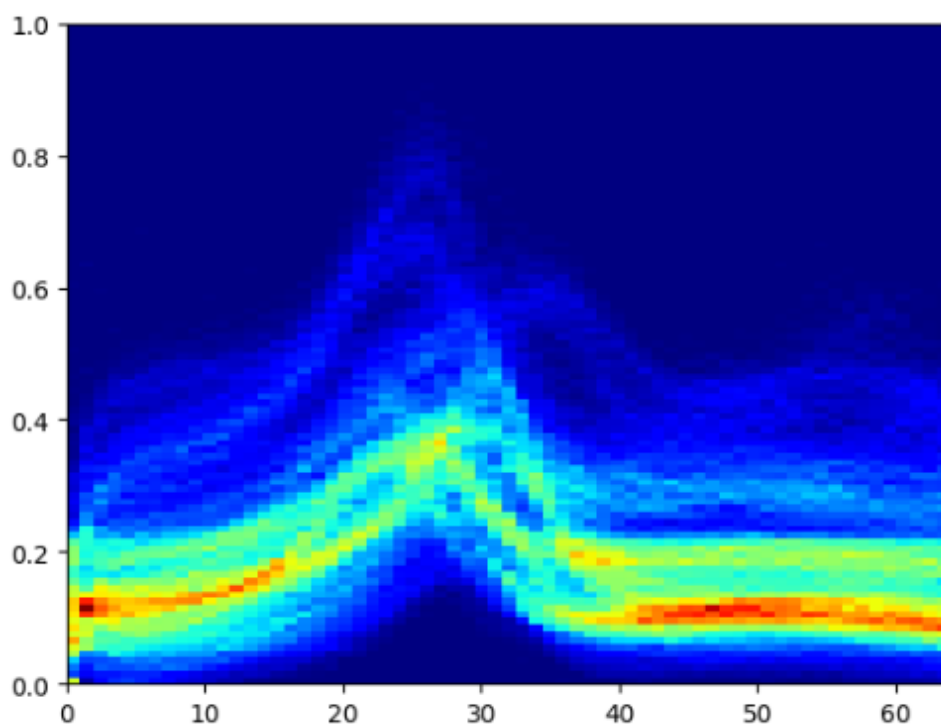


Рисунок 16 – Цветовая гистограмма распределения normal характеристик ЭКГ

Из приведенной выше цветовой карты гистограммы для данных РТВ, помеченных как нормальные, можно сделать вывод, что график признаков ЭКГ

имеет стандартную колоколообразную форму и достигает пика между признаками 20-30.

### 2.2.2 Предобработка данных

Для обучения модели, данные были преобразованы в Pandas-датафреймы. Были созданы 2 датафрейма, где один содержит единицы, а второй – нули.

Далее полученные датафреймы с помощью библиотеки scikit-learn были разделены на обучающую и тестовую выборки с соотношением 80% к 20%. Также полученные данные были проверены, не имеет ли какая-либо из функций значения null. Таким образом были подготовлены данные для дальнейшего использования в обучении модели.

### 2.2.3 Выбор классификатора и обучение

Для выбора модели был создан массив с различными классификаторами, такими как логистическая регрессия, ближайшие соседи, SVC, случайный лес, MLPClassifier и XGBoost. Результаты точности для каждой из модели представлены в таблице 4.

Таблица 4 – Результаты точности для каждой модели

Классификатор	F1_score
Логистическая регрессия	0.77
Ближайшие соседи	0.947
SVC	0.749
Случайный лес	0.979
MLPClassifier	0.868
XGBoost	0.985

Из таблицы 5 был сделан вывод, что XGBoost показывает результаты на обучающей выборке лучше, чем другие классификаторы. Следовательно, далее будем работать с данным классификатором. После обучения модели делаем предсказание на тестовой выборке.

Полученные результаты на тестовой выборке будут рассмотрены в следующей главе.

## **2.3 Цикл работы модели для оценки данных ЭКГ**

### **2.3.1 Загрузка и предобработка данных**

Изначально данные были поделены на обучающую и тестовую выборки. После загрузки обучающей выборки данные были поделены на 2 DataFrame. Первый (y\_df) это метка данных, описанная как категория от 0 до 4. Поэтому мы можем использовать их в качестве цели. Второй (X\_df) это точки данных ЭКГ, их будем использовать как признаки.

После разделения данных, необходимо вывести уникальные значения для каждой метки класса, чтобы отбросить повторяющиеся данные. Так же был составлен словарь для сопоставления целей с именами классов. Далее были построены первые 5 сигналов ЭКГ, для проверки данных (соответствуют ли они реально сигналам) и имен классов. Результат этих действий представлен на рисунке 17.

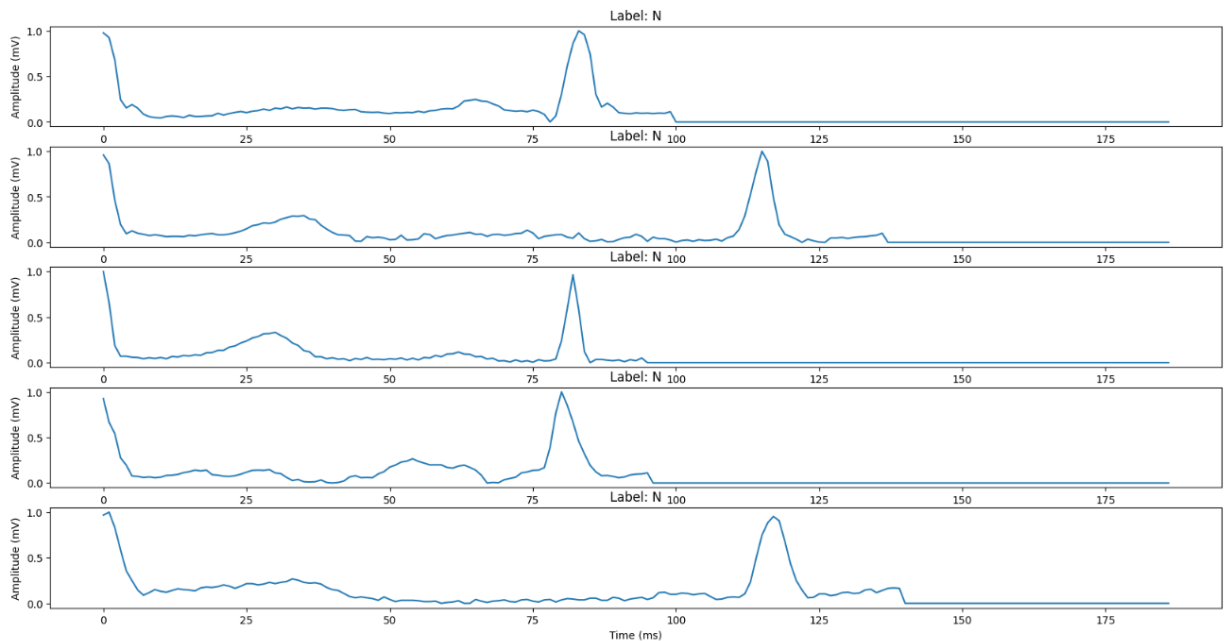


Рисунок 17 – Визуализированные данные ЭКГ обучающего DataFrame

Для работы нашей модели, необходимо преобразовать наши данные. Необходимо, чтобы вход был  $(batch\_size, 1, 187)$ , а выход -  $(batch\_size, 5)$ , потому что на входе один канал, а на выходе - 5 классов, так как именно такой формат будет принимать модель. Для этого мы добавили измерение канала к входным данными  $X\_df$  и выходным данным  $y\_df$ . Они были названы  $X\_pr$  и  $y\_pr$  соответственно.

Далее была проведена нормализация массива  $X\_pr$ . Сначала было вычислено среднее значение ( $X\_mean$ ), стандартное отклонение ( $X\_std$ ), максимальное значение ( $X\_max$ ) и минимальное значение ( $X\_min$ ) массива. Затем вычли среднее значение из каждого элемента массива, чтобы центрировать данные вокруг нуля. После этого полученный результат делят на стандартное отклонение, чтобы стандартизировать распределение данных.

После была построена гистограмма для нормализованного массива ( $X\_norm$ ), изображенная на рисунке 18. Он отображает распределение данных, где по оси  $X$  отложено время (в миллисекундах), а по оси  $Y$  - амплитуда (в милливольтгах).

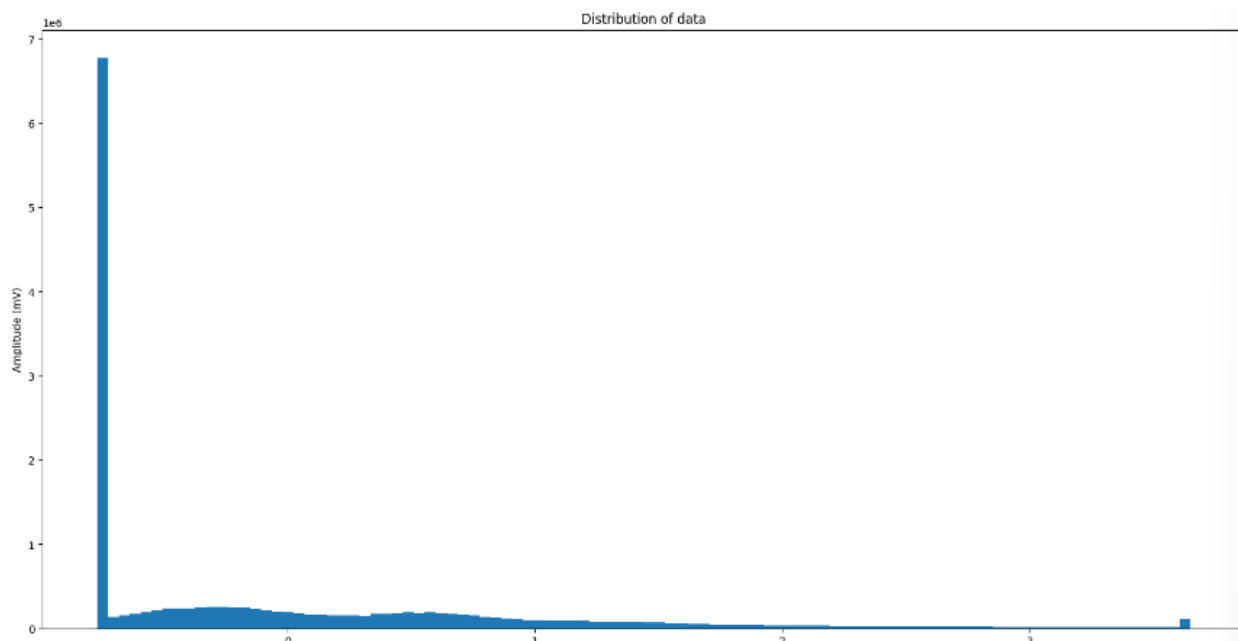


Рисунок 18 – Гистограмма распределения данных обучающего DataFrame

Когда данные были нормализованы, их сохранили в отдельный CSV файл. После был изменен тип данных на int. С помощью scikit-learn данные были разделены на обучающие и валидационные наборы X\_train, X\_val для точек данных ЭКГ и y\_train, y\_val для метки данных в соотношении 80 на 20. Далее вывели частоту встречаемости каждого класса в обучающем и проверочном наборах.

Далее для обучающей выборки были вычислены веса для классов, чтобы компенсировать дисбаланс в распределении классов. Для этого веса инвертируются относительно частоты классов и затем нормализуются, чтобы сумма всех весов была равна 1. Полученный результат изображен на рисунке 19.

```
(array([0.0058546 , 0.19148038, 0.07225367, 0.6639985 , 0.06641285]),
 {0: 'N', 1: 'S', 2: 'V', 3: 'F', 4: 'Q'})
```

Рисунок 19 – Веса для каждого класса

После к каждому примеру в обучающем наборе данных был присвоен вес в соответствии с их классами. Был создан массив нулей размером с обучающий набор данных, присваивая к каждому примеру вес соответствующий его классу. Далее создаются объекты DataLoader для обучающего и валидационного наборов данных. Для обучающего набора используется WeightedRandomSampler для обеспечения баланса классов при выборе мини-пакетов. Для валидационного набора используется обычный RandomSampler без весов, так как цель валидации - оценить общую производительность модели, а не корректировать дисбаланс классов. В результате было получено распределение классов, изображенное на рисунке 20.

```
torch.Size([1024, 1])
N: 196
S: 195
V: 189
F: 226
Q: 218
```

Рисунок 20 – Распределение данных по классам после балансировки классов

### 2.3.2 Создание модели

Сначала был создан словарь гиперпараметров для модели машинного обучения. В таблице 5 представлены ключи гиперпараметров, его описание и значение.

Таблица 5 – Гиперпараметры, их описание и значение

Ключ	Описание	Значение
bs	Гиперпараметр для размера мини-пакета	1024
lr	Начальная скорость обучения	1e-3
lr_decay	Коэффициент снижения скорости обучения	0.3
epochs	Количество эпох	50

Создается модель нейронной сети для классификации временных рядов. Модель начинается с нескольких блоков, которые включают в себя свёрточные слои (Nin), которые уменьшают размерность входных данных и увеличивают число каналов. Затем следуют слои MaxPool1d для дальнейшего уменьшения размерности. После нескольких дополнительных блоков Nin добавляется слой Dropout для предотвращения переобучения. Последний блок Nin используется для классификации на 5 классов, за которым следуют слои AdaptiveAvgPool1d и Flatten для преобразования выходных данных в формат, подходящий для задачи классификации. После чего были суммированы количество элементов в параметрах модели, которые требуют градиента и конвертированы в число в мегабайты, получилось 1.156 MB, а затем выведены общее количество тренируемых параметров, равное 304075.

Экспериментируя, было обнаружено, что, если задать некоторым классам весовые коэффициенты, модель работает лучше. Таким образом модель заставляют изучать классы, с которыми она не справилась. Для этого был определен критерий потерь, который измеряет ошибки модели во время обучения. Веса, передаваемые в функцию, помогают учесть дисбаланс классов, увеличивая потери для недопредставленных классов.

После, был выбран оптимизатор Adam, который будет использоваться для обновления весов модели во время обучения. Значение скорости обучения (lr) задается с использованием гиперпараметра, определенного ранее.

Используя, функция StepLR из библиотеки PyTorch, была определяет схему снижения скорости обучения, которая снижает скорость обучения на определенный коэффициент. В данном случае скорость обучения будет снижаться на 30% каждые 10 эпох.

### 2.3.3 Обучение модели

Была реализована функция обучения модели, которая принимает модель, функцию критерия, оптимизатор, обучающий и валидационный DataLoader, количество эпох и схему снижения скорости обучения (если необходимо). Во время обучения отслеживалась потеря обучения и валидации, а также точность валидации. Если модель достигала лучшей точности на валидационном наборе данных, код сохранял текущие параметры модели. По окончании обучения он возвращает список потерь обучения и валидации, лучшие параметры модели и наилучшую точность.

После обучения модель получила точность проверки 98,6%.

После обучения были загружены лучшие параметры модели и сохранены в отдельный файл. Так же была построена матрица ошибок для валидационного набора данных, представленная на рисунке 21.

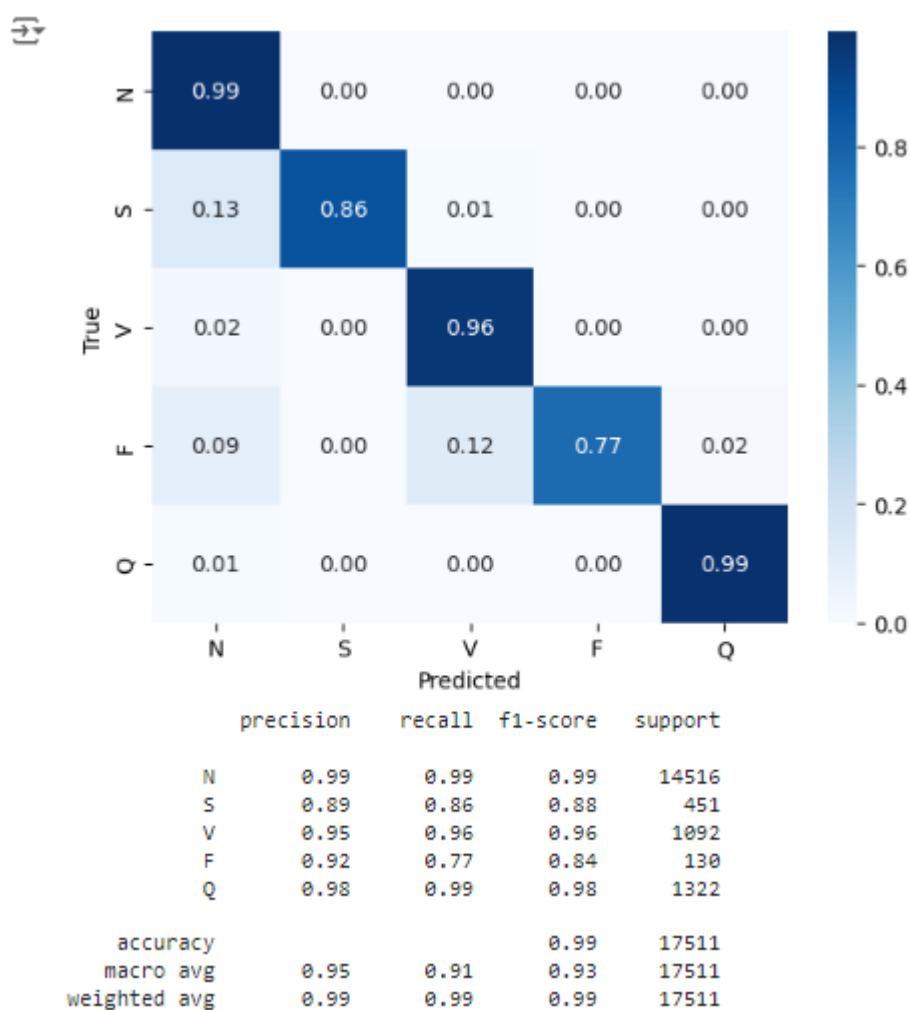


Рисунок 21 – Матрица ошибок для валидационного набора данных

Из матрицы ошибок можно сделать вывод, что модель отлично работает на вариационном наборе.

### 2.3.4 Проверка модели на тестовом наборе данных.

После загрузки тестовых данных, были проделаны те же действия, что и с обучающим набором данных:

- 1) Разделение данных на сигналы и метки.
- 2) Изменения формы данных, чтобы они имели вид (batch\_size, 1, 187) для сигналов и (batch\_size, 5) для меток, поскольку именно такой формы ожидает модель.

- 3) Проведена нормализация данных.
- 4) Выведено количество образцов в каждом классе
- 5) Создали тестовый набор данных и DataLoader для тестирования модели.

Далее была загружена ранее сохранённая модель и делается предсказание на тестовых данных. Полученные результаты на тестовых данных будут рассмотрены в следующей главе.

### 3. Обсуждение результатов

#### 3.1 Результаты работы модели для оценки данных ЭКГ

В результате проделанной работы, описанной в разделе 2.2, мы получили метрики качества, представленные на рисунке 22.

```
Precision: 0.9883936861652739  
Recall: 0.9906933457422057  
Accuracy: 0.9845413947097218  
f1: 0.9895421798745062
```

Рисунок 22 – Метрики качества полученные на тестовой выборке

Из полученных данных обученной модели можно сделать вывод, что:

1) Модель имеет высокую долю примеров, классифицированных как положительные, которые действительно являются положительными (Precision).

2) Высокая доля реально положительных примеров, которые были правильно определены моделью (Recall).

3) Высокая доля правильно классифицированных примеров (Accuracy).

4) Высокая среднее гармоническое между точностью и полнотой (F1-score).

Существуют другие методы классификации использующий такой же датасет. В приведенной ниже таблице 6 можно увидеть результаты других классификаторов.

Таблица 6 – Классификаторы и их результаты на нашем наборе данных

Классификатор	Precision	Recall	Accuracy	F1-score
CNN Autoencoders	0.55	0.89	0.77	0.65

Продолжение таблицы 6 – Классификаторы и их результаты на нашем наборе данных

Transformer	0.72	0.86	0.72	0.84
Logistic	0.67	0.92	0.65	0.77
Nearest Neighbors	0.95	0.89	0.94	0.94
RBF SVM	0.71	0.94	0.67	0.75
Random Forest	0.93	0.96	0.96	0.97
Neural Net	0.68	0.93	0.74	0.86

Из вышеперечисленных результатов можно сделать вывод, что наша модель работает лучше других. Больше всего к нашим показателям приблизились метод Nearest Neighbors (Ближайшие соседи) и Random Forest (Случайный лес).

### 3.2 Результаты работы модели для выявления данных здоровый или больных людей

В результате проделанной работы, описанной в разделе 2.3, мы получили метрики качества, представленные на рисунке 23.

	precision	recall	f1-score	support
N	0.99	0.99	0.99	18118
S	0.87	0.81	0.84	556
V	0.96	0.96	0.96	1448
F	0.81	0.77	0.79	162
Q	0.99	0.99	0.99	1608
accuracy			0.98	21892
macro avg	0.92	0.90	0.91	21892
weighted avg	0.98	0.98	0.98	21892

Рисунок 23 – Метрики качества полученные на тестовой выборке

На рисунке 23, видно, что построенная модель для мультиклассификации способна предсказывать несколько типов сердцебиений с точностью 98 %, этот вклад очень полезен; он предсказывает сразу 5 класса сердцебиений.

Из полученных данных обученной модели можно сделать вывод, что модель имеет высокие метрики качества, такие как точность (Accuracy), полнота (Recall), точность (Precision) и F1-мера (F1-score).

База данных MIT-BIH AR на сегодняшний день наиболее часто используется для оценки методов классификации аритмий. Однако провести справедливое сравнение между классификаторами сердцебиений — сложная задача. В приведенном ниже рисунке 24 можно увидеть результаты работы других классификаторов, обученных на том же датасете.

Decision Tree				
	precision	recall	f1-score	support
0.0	0.98	0.97	0.98	18118
1.0	0.64	0.64	0.64	556
2.0	0.86	0.86	0.86	1448
3.0	0.51	0.59	0.55	162
4.0	0.94	0.94	0.94	1608
accuracy			0.95	21892
macro avg	0.79	0.80	0.79	21892
weighted avg	0.95	0.95	0.95	21892
Random Forest				
	precision	recall	f1-score	support
0.0	0.98	0.97	0.98	18118
1.0	0.64	0.64	0.64	556
2.0	0.86	0.86	0.86	1448
3.0	0.51	0.59	0.55	162
4.0	0.94	0.94	0.94	1608
accuracy			0.95	21892
macro avg	0.79	0.80	0.79	21892
weighted avg	0.95	0.95	0.95	21892
XGBoost				
	precision	recall	f1-score	support
0.0	0.96	1.00	0.98	18118
1.0	0.95	0.53	0.68	556
2.0	0.95	0.79	0.86	1448
3.0	0.74	0.38	0.50	162
4.0	0.98	0.91	0.95	1608
accuracy			0.96	21892
macro avg	0.92	0.72	0.79	21892
weighted avg	0.96	0.96	0.96	21892

Рисунок 24 – Результаты работы классификаторов на нашем датасете

Сравнивая классификаторы, можно сделать вывод, что наша модель работает лучше, показывая результаты выше остальных.

Однако можно заметить, что у всех классификаторов возникают проблемы с классами 1.0 (S: Суправентрикулярное преждевременное биение) и 3.0 (F: Слияние желудочкового и нормального ритма). Несмотря на то, что в нашей модели эти классы определяются хуже, чем остальные, точность значительно выше, чем у других классификаторов представленных на рисунке 24. Матрица ошибок нашей модели представлена на рисунке 25.

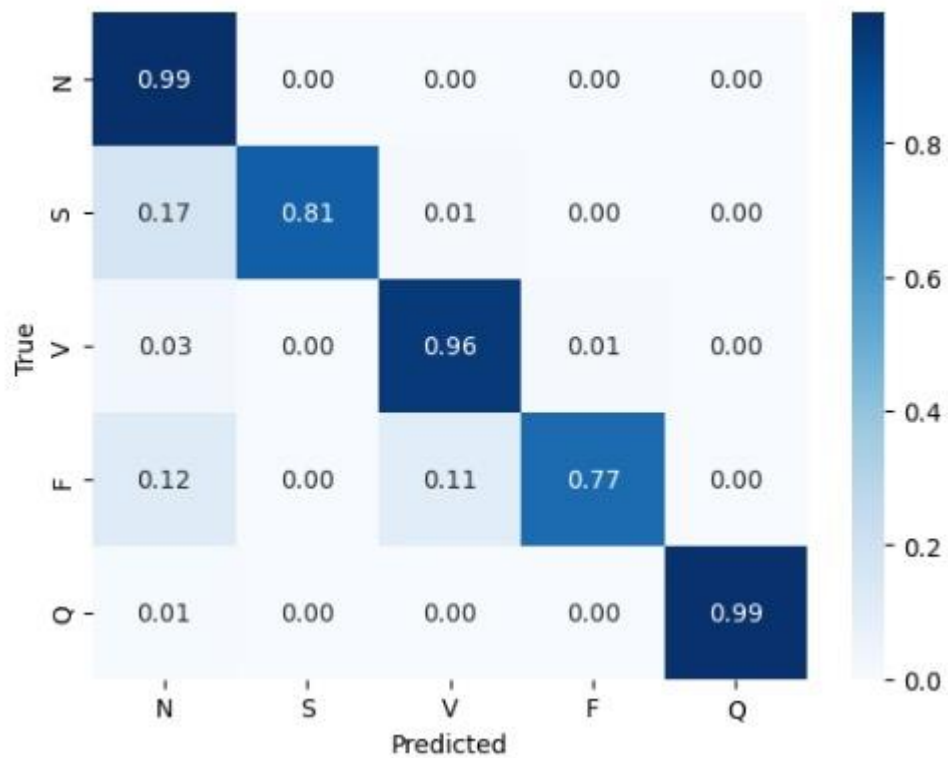


Рисунок 25 – Матрица ошибок для тестового набора данных

Давайте более подробно рассмотрим матрицу ошибок. Из матрицы видно, что модель выдает класс S за класс N, а также класс F за классы N и V. Возможно это связано с тем, что разные классы могли иметь в графике незначительные различия, однако при разметке графиков врачом они были помечены по-разному.

## ЗАКЛЮЧЕНИЕ

В данной выпускной квалификационной работе была поставлена цель – изучить возможность использования данных ЭКГ для анализа состояния здоровья человека с помощью методов машинного обучения. Для достижения этой цели были поставлены определенные задачи.

Вначале был проведен анализ литературы, который включал в себя изучение данных ЭКГ, существующие методы машинного обучения применяемых к данным ЭКГ. В результате чего, были выбраны алгоритмы, использующиеся в проделанной работе. После был проведен анализ выбранного датасета.

Были применены выбранные алгоритмы машинного обучения для исследования изменения состояния здоровья человека на основе данных ЭКГ с использованием машинного обучения.

Исходя из раздела “датасет” можно сделать вывод, что в данных присутствует огромный дисбаланс классов. Чтобы это избежать стоит добавить больше уникальных “аномальных” данных, вследствие чего модель будет более точно определять каждый класс.

В результате работы был проведен анализ результатов модели по выявлению здоровых или больных людей, и модели по оценке данных ЭКГ.

В будущем стоит не только увеличить уникальные данные для каждого класса, а также увеличить количество классов. В качестве будущей работы реализованные методы могут быть перестроены для работы со многими классами (например, более чем с 5 типами сердцебиений), также у некоторых болезней есть степень состояния, так, например использующийся в работе класс “Преждевременное сокращение желудочков”, можно заменить на преждевременное сокращение желудочков первой, второй и третьей степени. Это позволит узнать не только вид аритмии, а еще и её стадию. Работа может быть доработана для использования в режиме реального времени и постоянно

обучаться для ее улучшения и повышения точности, а также позволит быстрее принять меры для реабилитации пациента. Более того, весь процесс классификации можно использовать с другими типами данных, такими как данные о стрессе и клинические данные.

В результате проделанной работы можно сделать вывод, что понимание точной связи лежащей в основе физиологией и особенностями является потенциальным вопросом, требующим решения. Однако не существует окончательных рекомендаций относительно того, какие решения следует использовать для диагностики аритмий по ЭКГ с использованием машинного обучения.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. 70.pdf. – URL: <https://www.rzgm.ru/images/files/7/70.pdf> (дата обращения: 27.05.2024). – Текст : электронный.
2. 2023\_МД Мусилимов Динмухамед.pdf. – URL: [https://official.satbayev.university/download/document/33103/2023\\_%D0%9C%D0%94%20%D0%9C%D1%83%D1%81%D0%B8%D0%BB%D0%B8%D0%BC%D0%BE%D0%B2%20%D0%94%D0%B8%D0%BD%D0%BC%D1%83%D1%85%D0%B0%D0%BC%D0%B5%D0%B4.pdf](https://official.satbayev.university/download/document/33103/2023_%D0%9C%D0%94%20%D0%9C%D1%83%D1%81%D0%B8%D0%BB%D0%B8%D0%BC%D0%BE%D0%B2%20%D0%94%D0%B8%D0%BD%D0%BC%D1%83%D1%85%D0%B0%D0%BC%D0%B5%D0%B4.pdf) (дата обращения: 27.05.2024). – Текст : электронный.
3. Александрович, Г. В. Сердечно-сосудистая заболеваемость - одна из важнейших проблем здравоохранения / Г. В. Александрович, И. Е. Константиновна // Медицина и организация здравоохранения. – 2019. – Т. 4. – № 1. – С. 56-63.
4. Аритмия на ЭКГ: фото с расшифровкой признаков. – URL: <https://cardiograf.com/diagnostika/electro/aritmiya-na-ehkg.html> (дата обращения: 27.05.2024). – Текст : электронный.
5. В.в, К. Аритмия как случайный процесс. Анализ суточного мониторирования электрокардиограммы / К. В.в, К. Е.в // Российский кардиологический журнал. – 2019. – № 4S2. – С. 39-39.
6. Вестник Аритмологии :: АНАЛИЗ ВАРИАБЕЛЬНОСТИ СЕРДЕЧНОГО РИТМА ПРИ ИСПОЛЬЗОВАНИИ РАЗЛИЧНЫХ ЭЛЕКТРОКАРДИОГРАФИЧЕСКИХ СИСТЕМ (часть 1). – URL: <http://www.vestar.ru/article.jsp?id=1267> (дата обращения: 27.05.2024). – Текст : электронный.
7. Владимировна, Я. М. Желудочковые аритмии структурно нормального сердца / Я. М. Владимировна // Клиническая практика. – 2014. – № 4 (20). – С. 73-85.
8. Диагностика аритмии. – URL: <https://cardio.euroonco.ru/diagnostika-aritmii> (дата обращения: 27.05.2024). – Текст : электронный.

9. ИПСИЛАТЕРАЛЬНЫЙ БЛОК, ИЛИ МОЖНО ЛИ ПО ОДНОЙ ЭКГ ПОСТАВИТЬ ДИАГНОЗ? / К. А. Васильевич, С. Е. Викторова, Х. О. Васильевна [и др.] // Клиническая практика. – 2021. – Т. 12. – № 1. – С. 105-112.

10. ПАРОКСИЗМАЛЬНАЯ СУПРАВЕНТРИКУЛЯРНАЯ ТАХИКАРДИЯ И БЕРЕМЕННОСТЬ (КРАТКИЙ ОБЗОР ЛИТЕРАТУРЫ И КЛИНИЧЕСКОЕ НАБЛЮДЕНИЕ) / И. В. Анатольевич, Б. Д. Валерьевич, Р. В. Анатольевич, Р. О. Владимировна // Медицина: теория и практика. – 2021. – Т. 6. – № 2. – С. 32-37.

11. Признаки желудочковой экстрасистолии на ЭКГ. – URL: [https://meduniver.com/Medical/cardiologia/geludochkovaia\\_ekstrasistolia\\_na\\_ekg.html](https://meduniver.com/Medical/cardiologia/geludochkovaia_ekstrasistolia_na_ekg.html) (дата обращения: 16.05.2024). – Текст : электронный.

12. Принципы анализа ритма сердца - диагностика аритмий. – URL: [https://meduniver.com/Medical/cardiologia/dianostika\\_aritmii.html](https://meduniver.com/Medical/cardiologia/dianostika_aritmii.html) (дата обращения: 27.05.2024). – Текст : электронный.

13. Роль ЭКГ в диагностике жизнеугрожающих аритмий. / Статья на сайте Волынской больницы от 24 ноября 2015 г. – URL: <https://volynka.ru/Articles/Text/634> (дата обращения: 27.05.2024). – Текст : электронный.

14. СОВРЕМЕННЫЕ МЕТОДЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ФУНКЦИОНАЛЬНЫХ СОСТОЯНИЙ У ЗДОРОВЫХ И НЕКОТОРЫХ НАРУШЕНИЙ РИТМА СЕРДЦА У БОЛЬНЫХ - Научное обозрение. Реферативный журнал (научный журнал). – URL: <https://abstract.science-review.ru/ru/article/view?id=1798> (дата обращения: 27.05.2024). – Текст : электронный.

15. Х, С. КЛАССИФИКАЦИЯ АРИТМИЙ С ИСПОЛЬЗОВАНИЕМ ПРЕДВАРИТЕЛЬНО ОБУЧЕННОЙ МОДЕЛИ ГЛУБОКОГО ОБУЧЕНИЯ С БИНАРНЫМИ ИЗОБРАЖЕНИЯМИ СЕГМЕНТИРОВАННОЙ ЭКГ / С. Х, С. С // Известия высших учебных заведений России. Радиоэлектроника. – 2023. – Т. 26. – № 2. – С. 120-127.

16. ЭКГ-диагностика нарушений функции автоматизма синусового узла, замещающих комплексов и ритмов | Задионченко В.С., Ялымов А.А., Шехян Г.Г., Щикота А.М. | «РМЖ» №9 от 31.05.2016. – URL: [https://www.rmj.ru/articles/kardiologiya/EKG-diagnostika\\_narusheniy\\_funkcii\\_avtomatizmasinusovogo\\_uzla\\_zameschayuschih\\_kompleksov\\_i\\_ritmov/?utm\\_source=yandex.ru&utm\\_medium=organic&utm\\_campaign=yandex.ru&utm\\_referrer=yandex.ru](https://www.rmj.ru/articles/kardiologiya/EKG-diagnostika_narusheniy_funkcii_avtomatizmasinusovogo_uzla_zameschayuschih_kompleksov_i_ritmov/?utm_source=yandex.ru&utm_medium=organic&utm_campaign=yandex.ru&utm_referrer=yandex.ru) (дата обращения: 27.05.2024). – Текст : электронный.

17. Электрокардиография - ЭКГ ( отведения, нормальное экг ). – URL: <https://meduniver.com/Medical/Therapy/112.html> (дата обращения: 15.05.2024). – Текст : электронный.

18. Электрокардиография - ЭКГ (патология предсердий (P) и желудочков (QRS,ST). Гипертрофия желудочков. – URL: <https://meduniver.com/Medical/Therapy/113.html> (дата обращения: 27.05.2024). – Текст : электронный.

19. Ю.п, Г. Информационная оценка физиологических сигналов человека при изменении его психоэмоциональных состояний / Г. Ю.п // Pedagogics, psychology, medical-biological problems of physical training and sports. – 2007. – № 6. – С. 82-86.

20. Algoritmy\_EKG\_diagnostiki\_Polyakova\_Arabidze\_2020.pdf. – URL: [http://r967198x.beget.tech/wp-content/uploads/2022/10/Algoritmy\\_EKG\\_diagnostiki\\_Polyakova\\_Arabidze\\_2020.pdf](http://r967198x.beget.tech/wp-content/uploads/2022/10/Algoritmy_EKG_diagnostiki_Polyakova_Arabidze_2020.pdf) (дата обращения: 27.05.2024). – Текст : электронный.

21. Application of Support Vector Machine Classifier on Developed Wireless ECG System | Request PDF. – URL: [https://www.researchgate.net/publication/303681677\\_Application\\_of\\_Support\\_Vector\\_Machine\\_Classifier\\_on\\_Developed\\_Wireless\\_ECG\\_System](https://www.researchgate.net/publication/303681677_Application_of_Support_Vector_Machine_Classifier_on_Developed_Wireless_ECG_System) (дата обращения: 15.05.2024). – Текст : электронный.

22. Frontiers | A Study on Arrhythmia via ECG Signal Classification Using the Convolutional Neural Network. – URL:

[https://translated.turbopages.org/proxy\\_u/en-ru.ru.d6ba5815-66459c1b-024f6f24-74722d776562/https/www.frontiersin.org/articles/10.3389/fncom.2020.564015/full](https://translated.turbopages.org/proxy_u/en-ru.ru.d6ba5815-66459c1b-024f6f24-74722d776562/https/www.frontiersin.org/articles/10.3389/fncom.2020.564015/full) (дата обращения: 16.05.2024). – Текст : электронный.

23. heart-rate-variability-analysis.pdf. – URL: <https://s3.timeweb.com/cm94660-20b5ac41-549d-43b0-a32e-a074405a8023/rus/issues/no10-may-2017/heart-rate-variability-analysis.pdf> (дата обращения: 27.05.2024). – Текст : электронный.

24. Kim, В.-Н. ECG Identification For Personal Authentication Using LSTM-Based Deep Recurrent Neural Networks / В.-Н. Kim, J.-Y. Pyun // Sensors. – 2020. – Т. 20. – С. 3069.

25. m\_th\_a.p.lebedev\_2023.pdf. – URL: [https://elar.urfu.ru/bitstream/10995/129160/1/m\\_th\\_a.p.lebedev\\_2023.pdf](https://elar.urfu.ru/bitstream/10995/129160/1/m_th_a.p.lebedev_2023.pdf) (дата обращения: 16.05.2024). – Текст : электронный.

26. Osnovnye\_elementy\_normalnoj\_EKG\_2019g.pdf. – URL: [https://www.amursma.ru/upload/iblock/41f/Osnovnye\\_elementy\\_normalnoj\\_EKG\\_2019g.pdf](https://www.amursma.ru/upload/iblock/41f/Osnovnye_elementy_normalnoj_EKG_2019g.pdf) (дата обращения: 16.05.2024). – Текст : электронный.

27. (PDF) A comprehensive artificial intelligence-enabled electrocardiogram interpretation program. – URL: [https://www.researchgate.net/publication/346798497\\_A\\_comprehensive\\_artificial\\_intelligence-enabled\\_electrocardiogram\\_interpretation\\_program](https://www.researchgate.net/publication/346798497_A_comprehensive_artificial_intelligence-enabled_electrocardiogram_interpretation_program) (дата обращения: 15.05.2024). – Текст : электронный.

28. (PDF) A STUDY ON ECG SIGNALS FOR EARLY DETECTION OF HEART DISEASES USING MACHINE LEARNING TECHNIQUES. – URL: [https://www.researchgate.net/publication/356084524\\_A\\_STUDY\\_ON\\_ECG\\_SIGNALS\\_FOR\\_EARLY\\_DETECTION\\_OF\\_HEART\\_DISEASES\\_USING\\_MACHINE\\_LEARNING\\_TECHNIQUES](https://www.researchgate.net/publication/356084524_A_STUDY_ON_ECG_SIGNALS_FOR_EARLY_DETECTION_OF_HEART_DISEASES_USING_MACHINE_LEARNING_TECHNIQUES) (дата обращения: 15.05.2024). – Текст : электронный.

29. (PDF) Updates on the Electrocardiogram in Acute Coronary Syndromes. – URL: [https://www.researchgate.net/publication/257809517\\_Updates\\_on\\_the\\_Electrocard](https://www.researchgate.net/publication/257809517_Updates_on_the_Electrocard)

iogram\_in\_Acute\_Coronary\_Syndromes (дата обращения: 15.05.2024). – Текст : электронный.

30. Shvyrkova\_Aleksandra\_Alekseevna\_Bachelor\_Report\_2020\_text.pdf. – URL: [https://se.math.spbu.ru/thesis/texts/Shvyrkova\\_Aleksandra\\_Alekseevna\\_Bachelor\\_Report\\_2020\\_text.pdf](https://se.math.spbu.ru/thesis/texts/Shvyrkova_Aleksandra_Alekseevna_Bachelor_Report_2020_text.pdf) (дата обращения: 27.05.2024). – Текст : электронный.

31. Trends and applications of ECG analysis and classification. – URL: [https://www.researchgate.net/publication/357841798\\_Trends\\_and\\_applications\\_of\\_ECG\\_analysis\\_and\\_classification](https://www.researchgate.net/publication/357841798_Trends_and_applications_of_ECG_analysis_and_classification) (дата обращения: 15.05.2024). – Текст : электронный.

32. Tseng, K. ECG Identification System Using Neural Network with Global and Local Features. / K. Tseng, D. Lee, C. B. Chen. – Текст : электронный // International Association for Development of the Information Society. – 2016. – URL: <https://www.semanticscholar.org/paper/ECG-Identification-System-Using-Neural-Network-with-Tseng-Lee/447bdb1714c0fd8dc8315feffdd27f31940c87a6> (дата обращения: 27.05.2024).

33. UP\_BFSk2\_BTS.pdf. – URL: [https://op.vlsu.ru/fileadmin/Programmy/Magistratura/12.04.04/Metod\\_doc/UP\\_BFSk2\\_BTS.pdf](https://op.vlsu.ru/fileadmin/Programmy/Magistratura/12.04.04/Metod_doc/UP_BFSk2_BTS.pdf) (дата обращения: 27.05.2024). – Текст : электронный.

34. VKR\_GubajdullinBA.pdf. – URL: [https://dspace.spbu.ru/bitstream/11701/26512/1/VKR\\_GubajdullinBA.pdf](https://dspace.spbu.ru/bitstream/11701/26512/1/VKR_GubajdullinBA.pdf) (дата обращения: 27.05.2024). – Текст : электронный.

35. М.Д, Л. АНАЛИЗ ЭЛЕКТРОКАРДИОГРАММ ПОСРЕДСТВОМ НЕЙРОННЫХ СЕТЕЙ / Л. М.Д. – Text : electronic // Международный научно-исследовательский журнал. – 2015. – № № 4 (35). – URL: <https://research-journal.org/en/archive/4-35-2015-may/analiz-elektrokardiogramm-posredstvom-nejronnyh-setej> (дата обращения: 27.05.2024).

36. Alarsan, F. I. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms / F. I. Alarsan, M. Younes // Journal of Big Data. – 2019. – Vol. 6. – № 1. – P. 1-15.

37. ECG Heartbeat Categorization Dataset. – URL: <https://www.kaggle.com/datasets/shayanfazeli/heartbeat> (дата обращения: 16.05.2024). – Текст : электронный.

38. ECG-based pulse detection during cardiac arrest using random forest classifier / A. Elola, E. Aramendi, U. Irusta [et al.] // *Medical & Biological Engineering & Computing*. – 2019. – Vol. 57. – № 2. – P. 453-462.

39. El-Khafif, S. H. Artificial Neural Network-Based Automated ECG Signal Classifier / S. H. El-Khafif, M. A. El-Brawany // *International Scholarly Research Notices*. – 2013. – Vol. 2013. – P. e261917.

40. Localization of Ventricular Activation Origin from the 12-Lead ECG: A Comparison of Linear Regression with Non-Linear Methods of Machine Learning / S. Zhou, A. AbdelWahab, J. L. Sapp [и др.] // *Annals of biomedical engineering*. – 2019. – Т. 47. – Localization of Ventricular Activation Origin from the 12-Lead ECG. – № 2. – С. 403-412.

41. Аритмия: виды, диагностика и лечение. – URL: <https://mag.103.by/medicinskie-stati/99683-aritmija-vidy-diagnostika-i-lechenije/> (дата обращения: 16.05.2024). – Текст : электронный.

42. Гурьянова, В. Современные методы классификации одноканальных электрокардиограмм / В. Гурьянова.

43. ПроДокторов – сайт отзывов пациентов о врачах №1 в России. – URL: <https://probolezny.ru/aritmija/> (дата обращения: 27.05.2024). – Текст : электронный.

44. Methods of Signal Analysis for Automatic Diagnosis of Shockable Cardiac Arrhythmias: A Review / D. A. Lipchak, A. A. Chupov, A. E. Zhdanov, V. I. Borisov. – Текст : электронный // 2022 Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT) 2022 Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT). – Yekaterinburg, Russian Federation : IEEE, 2022. – Methods of Signal Analysis for Automatic Diagnosis of Shockable Cardiac

Arrhythmias. – С. 008-011. – URL: <https://ieeexplore.ieee.org/document/9923383/>  
(дата обращения: 16.05.2024).

45. Volkova, N. I. *Electrocardiography. Study Guide* / N. I. Volkova, I. S. Jeriyeva, A. L. Zibarev. – Текст : электронный // *Electrocardiography. Study Guide*. – ООО «GEOTAR-Media» Publishing Group, 2021. – URL: <http://www.studmedlib.ru/book/ISBN9785970464434.html> (дата обращения: 16.05.2024).