

ЗАО «ЕС-лизинг»

На правах рукописи

Новиков Роман Сергеевич

**МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ПО СПЕКТРАМ
КАРДИОЛОГИЧЕСКИХ
ДАННЫХ ДЛЯ ПРОБЛЕМНО-ОРИЕНТИРОВАННЫХ
ЦИФРОВЫХ ПРОДУКТОВ**

Специальность 2.3.5 –
«Математическое и программное обеспечение вычислительных
систем, комплексов и компьютерных сетей»

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель:
д.т.н., проф.
Позин Борис Аронович

Москва – 2025

Оглавление

Введение	4
Глава 1. Описание возможностей применения моделей машинного обучения в телемедицинских системах дистанционного биомониторинга на основе спектров кардиологических данных	14
1.1. Структура телемедицинских систем дистанционного мониторинга. Роль машинного обучения в данных системах.	14
1.2. Понятие кардиологических данных и применение над ними спектрального анализа. Анализ кардиологических сигналов в телемедицинских системах дистанционного мониторинга.....	17
1.3. Фреймворки машинного обучения (классификации) при анализе кардиологических данных.....	22
1.4. Бинарная и порядковая классификация в телемедицинских системах.	24
1.5. Рассмотрение прикладной проблемы создания медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по кардиологическим данным.....	26
1.6. Выводы по 1 главе.....	34
Глава 2. Описание разработанных методов построения моделей машинного обучения	37
2.1. Метод построения моделей машинного обучения с учителем для бинарной классификации объектов по амплитудно-частотным спектрам кардиологических данных.....	38
2.2. Методика формирования выборки спектров кардиологических данных для решения задач классификации	44
2.3. Метод построения моделей машинного обучения с учителем для порядковой классификации объектов с повторными измерениями	51

2.4. Применение методов построения моделей машинного обучения для классификации	62
2.5. Выводы по 2 главе.....	75
Глава 3. Экспериментальная проверка разработанных моделей классификации на примере задачи обнаружения кардиопризнаков НУО по кардиологическим данным.....	76
3.1 Экспериментальная оценка возможности выявления кардиопризнака НУО на кардиологических данных пациентов с СД2.....	77
3.2 Валидация методов неинвазивного выявления НУО по ЭКГ	89
3.3 Выводы по 3 главе.....	105
Глава 4. Разработка фреймворка для формирования модулей классификации по спектрам кардиологических данных как цифровых продуктов и его применение для построения медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по кардиологическим данным.....	107
4.1 Требования к телемедицинским системам дистанционного мониторинга по кардиологическим данным.	107
4.2 Архитектура фреймворка для формирования модулей по спектрам кардиологических данных и его реализация.....	112
4.3 Пример реализации телемедицинской системы дистанционного мониторинга по кардиологическим данным с применением фреймворка – ЭС СД2	118
4.4 Выводы по 4 главе.....	125
Заключение.....	126
Список литературы	128
Приложение.....	137

Введение

В современном мире телемедицина и дистанционные методы мониторинга здоровья становятся все более актуальными и востребованными. Развитие телемедицинских систем дистанционного мониторинга по кардиологическим данным позволяет своевременно выявлять и предотвращать кардиологические заболевания, обеспечивая удаленный контроль за состоянием пациентов.

Под «кардиологическими данными» или «кардиологическими сигналами» в работе понимаются электрокардиограммы (ЭКГ), выраженные в виде амплитудно-временного ряда значений, и ряды, вычисляемые из ЭКГ (например, кардиоинтервалограммы).

В России телемедицина получила развитие в правовом поле в 2017 году, когда на законодательном уровне стало возможно вести оказывать медицинскую помощь с применением телемедицинских технологий, включая дистанционное наблюдение (мониторинг). Фактически телемедицинские системы дистанционного мониторинга получили стимул к развитию в 2020 году с пандемией коронавирусной инфекции COVID-19, когда появилась потребность в дистанционном контроле врачами пациентов, которым не требовалось находиться в стационаре, для снижения рисков заражения.

Особенности этих систем заключаются в автоматизированном удаленном сборе сигналов от измерительных приборов (устройств регистрации ЭКГ и других кардиологических данных), где пациент обычно снимает данные самостоятельно, передаче данных в центр обработки и организации предоставления информации врачу для медицинского контроля или пациентом для самоконтроля.

Актуально применение в телемедицинских системах программных модулей классификации состояния здоровья пациента, использующих технологии машинного обучения, по амплитудно-частотным спектрам

кардиологических данных (сигналов) пациентов, т.к. такие сигналы возможно получить снимая пациентами вне медучреждений с помощью портативных электрокардиографов, а по их спектрам возможно выделять информативные признаки, которые могут быть незаметны на амплитудно-временной форме сигнала из-за шумов или других форм искажений. Однако разработка данных модулей является трудоемким и наукоемким процессом из-за необходимости строить модели машинного обучения, особенно, когда ставится задача распознавания по кардиологическим данным таких состояний пациента, для которых не хватает доступных выборок данных. Таким образом, возникает потребность в инструменте, который позволит автоматизировать формирование данных модулей при разработке телемедицинских системы дистанционного мониторинга.

Под «амплитудно-частотными спектрами» (или просто «спектрами») кардиологических данных или сигналов понимаются результаты преобразования этих данных алгоритмом быстрого преобразования Фурье.

Под «цифровыми продуктами» в целом понимается программное обеспечение, онлайн-сервис или приложение, предлагающие пользователям определенную ценность или функциональность. Закладываемое поведение модулей классификации состояния здоровья пациента внутри телемедицинской системы дистанционного мониторинга зависит от того, какие состояния пациента данная система должна классифицировать, поэтому модули классификации внутри данных систем являются проблемно-ориентированными цифровыми продуктами

Примером актуальности разработки телемедицинских системы дистанционного мониторинга, содержащего модули классификации состояния здоровья пациента по спектрам кардиологических данных, является потребность в удаленном выявлении нарушений углеводного обмена (НУО) при скрининге населения. В России за 2016 год у 5,4% человек 20–79 лет есть диагноз СД2 (ранее диагностированный - 2,5%; ранее не диагностированный -

2,9%), у 19,3% человек есть предиабет. На начало 2020 года в России зарегистрировано около 4 млн 800 тыс. больных СД, из них 92% — больные СД2. [1]

Скрининг на наличие НУО в масштабах страны по рекомендованным тестам по анализам крови не проводился в связи с высокой стоимостью из-за расходов на медицинских работников и лабораторное оборудование, необходимость наличия лабораторий и квалифицированного медицинского персонала в первичном звене. Скрининг по анализам крови сложно провести там, где нет нужного персонала/оборудования (например, в фельдшерско-акушерских пунктах в сельской местности).

В свою очередь, создание телемедицинской системы дистанционного мониторинга с модулем классификации состояния здоровья пациента машинного обучения по спектрам кардиологических данных позволило бы предложить альтернативный способ скрининга населения: более дешевый, доступный, удаленный и неинвазивный.

Степень разработанности темы исследования.

При наличии большого количества известных инструментов формирования моделей машинного обучения (Tensorflow, Scikit-learn и т.д.) для модулей классификации не было обнаружено инструмента, который позволял бы построить модель, учитывающую специфику кардиологических сигналов: учет требований к исходным данным, алгоритмы предварительной обработки и извлечения информативных признаков из спектров сигналов.

Основные работы, в которых рассказывается о применении анализа амплитудно-частотных спектров ЭКГ при выявлении разных состояний пациентов, опубликованы Березиным А.А. (2017), Новопашиным М.А. (2017).

Основные работы, посвященные применению анализа variability сердечного ритма (включая анализ амплитудно-частотных спектров кардиоинтервалограмм), опубликованы следующими учеными: Парин В.В. (1968), Баевский Р.М. (1968, 2001), Каплан А.Я. (1999). По variability

сердечного ритма обычно оценивают функциональные состояния пациентов в т.ч. спортсменов. Работы, посвященные применению анализа вариабельности сердечного ритма при тестировании спортсменов с целью управления тренировочным процессом, опубликованы Сивоховым В.Л. (2016).

Основные работы, в которых приведены модели классификации наличия/отсутствия сахарного диабета 2 типа (СД2) по показателям, рассчитанным по ЭКГ, опубликованы следующими учеными: Acharya U.R. (2013, 2015), Swapna G.(2018), Yildirim O. (2019). Однако упомянутые модели не разрабатывались под короткие записи ЭКГ (до 5 минут), сделанные на портативных электрокардиографах. Также по данным работам нет доступной выборки данных, поэтому невозможно валидировать данные или построить и оценить на них собственные модели машинного обучения.

Объектом исследования являются инструменты и методы формирования моделей машинного обучения.

Предметом исследования являются инструменты и методы формирования модулей классификации состояния здоровья, содержащего модели машинного «неглубокого» обучения по амплитудно-частотным спектрам ЭКГ и кардиоинтервалограмм, для телемедицинских систем дистанционного мониторинга.

Целью данной работы является разработка и реализация новых методов формирования модулей классификации, использующих модели машинного обучения по амплитудно-частотным спектрам кардиологических данных. В ходе исследования для достижения поставленной цели в работе поставлены следующие задачи:

- Разработать метод построения моделей машинного обучения с учителем для бинарной классификации объектов при помощи анализа амплитудно-частотного спектра (Фурье-спектра) кардиологических данных;

- Разработать метод построения моделей порядковой классификации объектов в условиях ограниченной выборки при помощи имитационного моделирования выборки по результатам работы бинарного классификатора;
- Разработать методику формирования выборки данных для создания моделей классификации по спектрам кардиологических данных.
- Разработать и оценить модели обнаружения кардиопризнаков НУО на спектрах кардиологических данных в двух вариантах с помощью разработанных методов: на основе анализа Фурье-спектра ЭКГ и на основе анализа параметров ВСР (включая сформированные из Фурье-спектра кардиоинтервалограмм);
- Разработать с помощью разработанных методов и оценить интегральную модель оценки риска наличия предиабета или СД2 по серии ЭКГ;
- Разработать фреймворк для формирования модулей классификации состояния здоровья пациента с использованием разработанных методов построения моделей машинного обучения с учителем для бинарной и порядковой классификации по спектрам кардиологических данных.
- Разработать телемедицинскую систему проведения неинвазивного удалённого скрининга НУО по ЭКГ на основе разработанного фреймворка.

Научная новизна. Результатами диссертационной работы стали следующие новые научные положения:

1. Разработан метод построения моделей машинного обучения с учителем для бинарной классификации объектов, отличающихся от остальных методов применением анализа амплитудно-частотного спектра кардиологических данных при построении признакового описания модели;

2. Разработан метод построения моделей порядковой классификации объектов, от остальных методов применимостью в условиях ограниченной выборки при помощи имитационного моделирования выборки;
3. Разработана методика формирования выборки данных для создания моделей классификации по спектрам кардиологических данных, отличающаяся от остальных учетом требований к исходным кардиологическим данным;
4. Разработана программная архитектура фреймворка для разработки модулей классификации с использованием разработанных методов построения моделей машинного обучения с учителем для бинарной и порядковой классификации по спектрам кардиологических данных.

Данные положения частично изложены в патентах РФ № 2751817 «Компьютеризированный способ неинвазивного выявления нарушений углеводного обмена по вариабельности сердечного ритма и носимое автономное устройство для его реализации» (дата начала отсчета срока действия патента – 31.08.2020) и № 2728869 «Компьютеризированный способ неинвазивного выявления нарушений углеводного обмена по электрокардиограмме» (дата начала отсчета срока действия патента – 30.08.2019), разработанных с участием автора, где впервые предложено по форме спектров ЭКГ проводить классификацию состояний (в частности, определение нарушения углеводного обмена) пациента путем сопоставления параметров и форм спектров кардиологических данных с эталонными параметрами и формами спектров кардиологических данных.

Практическая значимость работы.

Применение предложенных методов, методики и фреймворка позволили реализовать медицинскую систему самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности электрокардиограмм сердца первого отведения, позволяющую неинвазивно и

удаленно проводить выявление по серии спектров ЭКГ признаков нарушения углеводного обмена с классификацией пациентов на здоровых, с высокой вероятностью страдающих предиабетом и страдающих диабетом. Это в свою очередь, по сравнению с применением текущих рекомендованных скрининговых тестов, позволит:

- увеличить охват взрослого населения, которое может принять участие в скрининге, не менее чем на 33% за счет охвата сельского населения, слабо обеспеченного лабораториями для анализа крови и поликлиниками;
- снизить затраты на анализы крови всех участников скрининга не менее чем на 61% за счет раннего обнаружения участников с подозрением на НУО на основе анализа ЭКГ.

Методы исследования. Для решения поставленных задач использовались методы машинного обучения, теория цифровой обработки сигналов, анализ вариабельности сердечного ритма.

Основные положения, выносимые на защиту:

1. Метод машинного обучения с учителем для построения моделей бинарной классификации объектов при помощи анализа амплитудно-частотного спектра кардиологических данных;
2. Метод машинного обучения с учителем для построения моделей порядковой классификации объектов при помощи анализа амплитудно-частотного спектра кардиологических данных на повторных измерениях объекта;
3. Методика формирования выборки данных для создания моделей классификации по спектрам кардиологических данных.
4. Программная архитектура фреймворка для разработки модулей классификации по разработанным методам построения моделей машинного обучения с учителем;

Апробация работы. Достоверность результатов обеспечена теоретическим обоснованием и экспериментальным подтверждением выдвинутых положений систематической проверкой оригинальных данных, полученных в результате проведенных клинических исследований:

1. в Городской клинической больнице №52 и Московском Клиническом Научном Центре им. А. С. Логинова (исследование проводилось на базе кафедры эндокринологии и диабетологии МГМСУ им. А.И.Евдокимова) в 2019 году;
2. в поликлинике № 6 Мытищинской городской клинической больницы (исследование проводилось на базе НМИЦК им. ак. Е.И. Чазова) в 2021 году.

Автор являлся участником данных клинических исследований: оказывал консультации персоналу лечебно-профилактических учреждений по применению использованных в исследованиях электрокардиографов CardioQVARK; сопровождал информационную систему проведения неинвазивного удаленного скрининга по ЭКГ (ЭС СД2), разработанную автором и задействованную в исследованиях; проводил обработку статистики по материалам исследований; участвовал в написании заключительных отчетов по исследованиям. Отчет по исследованию 2019 года (руководитель исследования - заведующий кафедрой эндокринологии и диабетологии МГМСУ им. А.И.Евдокимова, д.м.н., профессор А.М. Мкртумян) утвержден в МГМСУ им. А.И. Евдокимова. Отчет по исследованию 2021 года утвержден директором НМИЦК им. ак. Е.И. Чазова, академиком РАН, д.м.н., профессором С.А. Бойцовым.

Результаты данной работы докладывались на конференциях:

1. Actual Problems of System and Software Engineering, 2019, 2021, Москва;

2. Ежегодная межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В.Арменского, 2019, 2020, Москва);
3. SYRCoSE Software Engineering Colloquium, 2020, 2022, Москва;
4. XVIII Международная научно-техническая конференция «Новые информационные технологии и системы», 2021, Пенза;

Публикации и личный вклад автора. По теме диссертации опубликовано 5 печатных работ, в том числе в изданиях, рекомендованных ВАК [2,3,4], а также 2 патента РФ [5,6].

В статье [2] совместно с соавторами поставлена задача, автору принадлежит основная часть: разработка и описание процесса проведения неинвазивного удаленного скрининга по ЭКГ и сравнение с рекомендованными процессами проведения скрининга (скрининговыми тестами).

В статье [3] вместе с соавторами поставлена задача. Реализация ролевой модели пациента, личного кабинета пациента медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности электрокардиограмм сердца первого отведения, разработка самой системы и финальное редактирование текста выполнены автором.

В статье [4] вместе с соавторами поставлена задача. Автором выполнена разработка метода построения модели порядковой классификации на повторных измерениях объекта и применения его при построении модели выявления подозрения на СД2/предиабет по серии ЭКГ.

В патентах [5,6] вклад автора заключается в разработке моделей выявления кардиопризнаков НУО по спектрам кардиологических данных.

Внедрение результатов. Результаты, полученные в рамках данной работы, внедрены в медицинскую систему самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности

электрокардиограмм сердца первого отведения (ЭС СД2). Работы по развитию и внедрению системы включены в федеральный проект «Здоровье здоровых».

Объем и структура работы. Диссертация состоит из введения, четырёх глав, заключения. Полный объем диссертации составляет 138 страниц текста, включая 15 рисунков и 19 таблиц. Список литературы содержит 65 наименований.

Глава 1. Описание возможностей применения моделей машинного обучения в телемедицинских системах дистанционного биомониторинга на основе спектров кардиологических данных

В данной главе:

- Рассмотрена структура телемедицинских систем дистанционного мониторинга и указана актуальность разработки и применения внутри них модулей классификации состояния здоровья пациента как проблемно-ориентированных цифровых продуктов;
- Рассмотрена актуальность применения спектрального анализа кардиологических сигналов в телемедицинских системах дистанционного мониторинга. Рассмотрены существующие фреймворки для разработки модулей с моделями классификации и указана актуальность создания специализированного фреймворка, рассматривающего спектры кардиологических данных в качестве выборки для формирования моделей, и методику формирования выборки кардиологических данных для фреймворка;
- Рассмотрена одна из прикладных проблем создания медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по кардиологическим данным, использующей модели машинного обучения;
- Поставлены цель и задачи текущей работы.

1.1. Структура телемедицинских систем дистанционного мониторинга. Роль машинного обучения в данных системах.

Телемедицинские системы дистанционного мониторинга представляют собой специализированные мониторные системы, ориентированные на

конкретные заболевания с передачей информации о состоянии больного в лечебно-профилактические учреждения и лечащему врачу. [7]

Особенность данных систем заключается в обеспечении возможности проводить диагностику и мониторинг состояния здоровья пациентов на расстоянии. В таких системах сбор данных о физиологических параметрах пациента осуществляется пациентом самостоятельно с помощью специального устройства сбора. Примерами измеряемых данных являются: электрокардиограмма, давление крови, температура тела, газовый состав крови; энцефалограмма, частота сердечных сокращений и содержание сахара в крови.

Данные системы могут быть предназначены как для отслеживания состояния здоровья работников на производстве или другой профессиональной деятельности (машинисты, операторы атомных станций), так и для мониторинга пациента в домашних условиях.

Стандартизированной структуры телемедицинских систем дистанционного мониторинга найдено не было, однако обычно выделяют следующие компоненты телемедицинских систем дистанционного мониторинга [7-9]:

- Портативное устройство сбора данных о физиологических параметрах пациента;
- Интерфейс передачи данных на удаленную обработку (специальный терминал или смартфон);
- Подсистема обработки и анализа данных, в т. ч. средствами искусственного интеллекта (в частности, машинного обучения);
- Подсистема хранения исходных и обработанных данных;
- Рабочее место врача, где предоставляются данные, полученные от пациента, и результаты их обработки и анализа.

На рисунке 1 представлен вариант схемы данных систем.

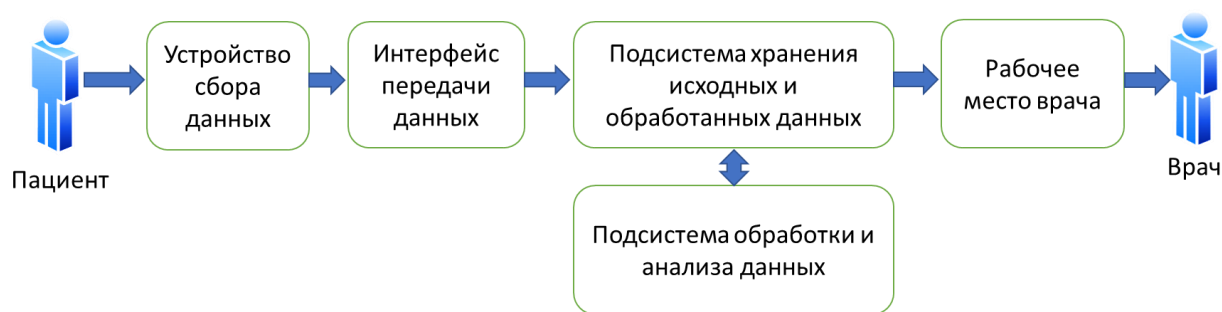


Рисунок 1 – Вариант общей схемы телемедицинских систем дистанционного мониторинга

В таких системах активно используются модули классификации состояния здоровья пациента, использующие методы и модели машинного обучения, в качестве проблемно-ориентированных цифровых продуктов для анализа данных и поддержки принятия решений врачами.

Под методом машинного обучения понимается подход, при котором прошлые данные или примеры используются для первоначального формирования и совершенствования схемы предсказания (модели машинного обучения). [10] В работе рассматривается, в основном, категория машинного обучения с учителем (классификация), когда модель принудительно обучается с помощью примеров входных и эталонных выходных данных. Примерами моделей машинного обучения являются логистическая регрессия, метод опорных векторов, байесовский классификатор, дерево решений. Однако в работе также упоминается машинное обучение без учителя (в частности, кластеризация), когда ставится задача обнаружение закономерностей в данных и их группировка без наличия эталонных выходных данных.

Значимым периодом активного интереса для внедрения по всему миру моделей машинного обучения в телемедицинские системы можно обозначить период пандемии коронавирусной инфекции (COVID-19) [11,12] в 2020 году, когда появилась потребность в дистанционном контроле врачами пациентов, которым не требовалось находиться в стационаре, для снижения рисков заражения.

Модели машинного обучения позволяют автоматически анализировать большие объёмы данных о состоянии здоровья пациента, выявлять закономерности и тенденции, а также прогнозировать развитие заболеваний. Это помогает врачам своевременно принимать меры и корректировать лечение, что повышает качество медицинской помощи и улучшает результаты лечения.

Особняком стоит глубокое обучение - разновидность машинного обучения с использованием более сложных моделей, где признаки объектов, по которым модель принимает решение, формируются не человеком, а средствами машинного обучения. Примером модели глубокого обучения является многослойный перцептрон. [13]

Однако, несмотря на превосходство моделей глубокого обучения перед остальными моделями машинного обучения (вроде логистической регрессии) по точности прогноза, в данной работе они не будут рассматриваться, поскольку в телемедицинских системах при автоматизированном принятии решения по пациенту обычно предъявляется требование интерпретируемости: способности объяснить пациенту простыми словами, почему система выдала ему именно «тот» результат модели, а не другой. [14,15], а модели глубокого обучения имеют слабую интерпретируемость, т.к. даже разработчик модели не знает, на основе каких признаков такая модель будет принимать решение после обучения. Особенно это требование важно, когда пациент не находится под непосредственным наблюдением врача, который мог бы сам все объяснить.

1.2. Понятие кардиологических данных и применение над ними спектрального анализа. Анализ кардиологических сигналов в телемедицинских системах дистанционного мониторинга.

Под «кардиологическими данными» или «кардиологическими сигналами» в работе понимаются электрокардиограммы (ЭКГ), выраженные в

виде амплитудно-временного ряда значений, и ряды, вычисляемые из ЭКГ (например, кардиоинтервалограммы).

ЭКГ записывают с помощью электродов с поверхности тела пациента: с грудной клетки и/или верхних и нижних конечностей. Электроды накладывают на определенные точки тела человека и соединяют проводами с электрокардиографом – устройством съема ЭКГ. Соединение двух точек тела человека, имеющих разные потенциалы, называется отведением. При наложении электродов на груди регистрируют грудные отведения, на конечностях – отведения от конечностей (стандартные и усиленные). [16]

По ЭКГ обычно определяют:

- частоту сердечных сокращений, ритма и проведения электрических импульсов в сердце;
- наличие состояний, связанных с работой сердца. Например, аритмию, ишемическую болезнь сердца, сужение коронарных артерий, инфаркт миокарда и т.д.

Оценкой ЭКГ занимается кардиолог или врач функциональной диагностики. Обычно ЭКГ оценивают путем анализа зубцов (P,Q,R,S,T-зубцы) и отрезков между зубцами (QT-интервал, QRS-комплекс, ST-сегмент и т.д.). Схема ЭКГ в пределах отдельного сердечного цикла изображена на рисунке 2. Сердечным циклом (кардиоциклом) называют интервал времени, за который происходит одно сокращение и расслабление сердца. Горизонтальная черта, на базе которой появляются зубцы ЭКГ, называется изолинией.

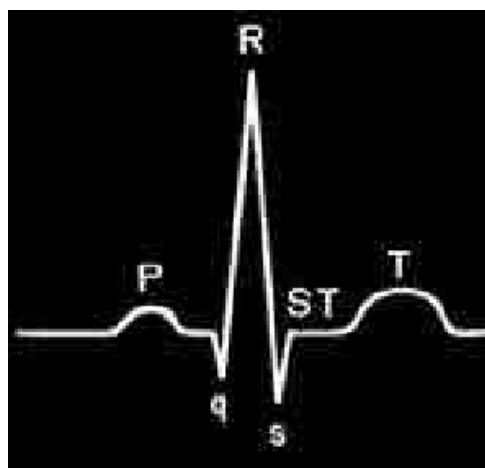


Рисунок 2 – Схематическое изображение сердечного цикла (кардиоцикла) ЭКГ [16]

Перед анализом ЭКГ обычно проводится предварительная обработка ЭКГ, включающая в себя удаление артефактов: фильтрация шума и обрезка амплитудных выбросов. [17]

Спектральный анализ ЭКГ является способом получения дополнительной информации об исходной ЭКГ. Данный анализ проводится путем перевода исходного или предобработанного сигнала в амплитудно-частотный спектр (например, алгоритмом быстрого преобразования Фурье) и рассмотрением мощности спектра в разных частотных диапазонах. С помощью спектрального анализа проводят как предварительную обработку ЭКГ (фильтрация частот для подавления шума, дрейфа изолинии и сетевых помех на ЭКГ), так и исследование характеристик кардиоцикла: анализ Р-зубца, Т-зубца и QRS-комплекса.

Спектральный анализ ЭКГ находит свое применение при обнаружении гипертрофии правого предсердия [18], ишемической болезни сердца [19], нарушений расстройств сна [20].

В работе [19] спектр ЭКГ, полученный быстрым преобразованием Фурье, представлен в качестве модели сердца как нелинейной автоколебательной системы. Авторами была выдвинута гипотеза о том, что динамика электрической активности нормально функционирующего сердца может быть

интерпретирована феноменом автовозврата (повторением во времени картины состояния нелинейной системы) Ферми–Паста–Улама (ФПУ), а нарушения в сердце приводят к нарушению автовозврата ФПУ. Предполагается, что на кардиограмме здорового человека можно обнаружить перекачку энергии между низкочастотным (1-5 Гц) и высокочастотным (20-35 Гц) диапазонам в виде шума на спектре ЭКГ. Данная гипотеза была проверена при визуальном сравнении спектров ЭКГ пациентов с ишемической болезнью сердца и без.

Кардиоинтервалограммой (КИГ) называется динамический ряд RR-интервалов (интервалов между пиками соседних R-зубцов). Анализ кардиоинтервалограмм в литературе называется анализом вариабельности сердечного ритма (ВСР)[21,22].

С помощью анализа ВСР обычно проводят оценку:

- функционального состояния организма и его изменений;
- выраженности адаптационного ответа организма при воздействии различных стрессоров;
- общей активности механизмов регуляции сердечного ритма, соотношения между активностью симпатического и парасимпатического отделов вегетативной нервной системы.

Перед анализом ВСР обычно проводится предварительная обработка КИГ, включающая в себя обрезку аномальных RR-интервалов. [23]

Анализ ВСР используется как в клинической (например, для уточнения диагноза автономной невропатии при диабете), так и в космической и спортивной медицине [24] – для оценки адаптации систем организма к физическим и психоэмоциональным нагрузкам. В частности, спектральный анализ КИГ дает информацию о распределении мощности колебаний сердечного ритма в зависимости от частоты колебаний. Применение спектрального анализа позволяет количественно оценить различные частотные составляющие колебаний ритма сердца и наглядно графически

представить соотношения разных компонентов сердечного ритма, отражающих активность определенных звеньев регуляторного механизма.

Рассмотрим некоторые телемедицинские системы дистанционного мониторинга, работающие с кардиологическими данными.

В качестве примеров таких систем можно привести систему мониторинга ЭКГ «Кардиометр-МТ» [25, 26], принимающую ЭКГ в 12 стандартных отведениях, отображающую ее на приемных и передающих пунктах и формирующую автоматическое заключение по результатам анализа ЭКГ. Данные результаты при необходимости корректирует врач-кардиолог и высылает в удаленный пункт установленный диагноз со своими рекомендациями. В анализ ЭКГ входит: автоматическое выделение элементов электрокардиограммы, имеющих диагностическую значимость, измерение их амплитудно-временных параметров; анализ сердечного ритма пациента на основе измеренных амплитудно-временных параметров зубцов и интервалов электрокардиограммы [27].

Другим примером является система мониторинга ЭКГ пациентов, состоящая из портативного электрокардиографа «ЭКГ-ЭКСПРЕСС» и веб-ориентированного программного комплекса удаленного мониторинга ЭКГ. [28, 29] Данная система принимает ЭКГ с портативного кардиографа ЭКГ и анализирует ЭКГ на предмет выявления нарушений ритма, оценки общего уровня стресса и выявления ишемических синдромов.

Также в качестве примера телемедицинские системы дистанционного мониторинга можно привести программно-аппаратный комплекс «Киберсердце» (№6471 в реестре рос. ПО), работающий с ЭКГ в 12 стандартных отведениях и использующий при анализе ЭКГ такие модели машинного обучения как метод опорных векторов, дерево решений, искусственные нейронные сети, методы квадратичного и линейного дискриминантного анализа, метод случайных подпространств, AdaBoost, случайный лес, логистическая регрессия [30].

Было отмечено, что при изучении телемедицинских систем дистанционного мониторинга, работающих с кардиологическими данными, были найдены системы, работающие со спектрами КИГ при анализе ВСР, но не было обнаружено систем, работающих со спектрами ЭКГ при принятии решения по пациенту. Однако был обнаружен патент РФ на медицинскую систему удаленного мониторинга, анализа и прогнозирования состояния пациента, принимающей на вход сигналы ЭКГ, преобразующую в амплитудно-частотный спектр, а после - в «дискретизированную ЭКГ» (ДЭКГ), представляющую собой «точку с координатами, значения которых несут информацию о форме исходной ЭКГ». [31] Эти ДЭКГ, согласно патенту, далее участвуют в качестве выборки при формировании моделей машинного обучения без учителя (кластеризации).

Итого, несмотря на наличие практического смысла в проведении спектрального анализа ЭКГ с целью выявления разных заболеваний или других состояний пациента, в телемедицинских системах дистанционного мониторинга спектральный анализ ЭКГ в моделях машинного обучения используется довольно редко.

1.3. Фреймворки машинного обучения (классификации) при анализе кардиологических данных

Под фреймворками машинного обучения понимается среда с набором инструментов для обеспечения быстрой разработки программных продуктов, содержащих модели машинного обучения.

Среди самых популярных фреймворков машинного обучения можно выделить TensorFlow, PyTorch, OpenCV и Scikit-learn.

TensorFlow и PyTorch предназначены для глубокого обучения, поддерживают создание и обучение сложных нейронных сетей и включают в себя инструменты для работы как с изображениями, так и временными рядами, коими являются кардиологические данные.

OpenCV представляет собой библиотеку для компьютерного зрения, которую можно использовать для анализа медицинских изображений, в т.ч. изображений кардиологических данных, что также относится к глубокому обучению.

Из фреймворков, нацеленных на построение «простых» моделей машинного обучения вроде логистической регрессии или дерева решений, можно выделить библиотеку Scikit-learn на языке Python, которые возможно применять для анализа кардиологических данных в виде рядов значений. Однако кардиологические данные требуют предварительной обработки, чего не обеспечивает Scikit-learn, а также требуют предварительной оценки, соответствуют ли кардиологические данные требованиям, изложенным в стандартах или рекомендациях. Также Scikit-learn не имеет специальной функциональности проведения спектрального анализа кардиологических данных и извлечения признаков из спектров данных.

Итого, для ускорения разработки модулей классификации в телемедицинских системах дистанционного мониторинга, использующих спектральный анализ кардиологических данных, необходим специализированный фреймворк, который позволит провести:

- проверку кардиологических данных на соответствие требованиям;
- предварительную обработку кардиологических данных, включая преобразование в амплитудно-частотный спектр и выделение признаков из спектров;
- формирование модулей классификации состояния здоровья пациента на основе моделей машинного обучения, обладающих свойством интерпретируемости, по признакам, полученных в т.ч. из спектров кардиологических данных.

В силу специализированности данного фреймворка, для него требуется разработать отдельно методику формирования выборки кардиологических данных (для подготовки к формированию моделей) и отдельно методы

построения моделей машинного обучения по кардиологическим данным. В рамках исследования рассматриваются модели бинарной и порядковой классификации.

1.4. Бинарная и порядковая классификация в телемедицинских системах.

Под задачей бинарной классификации понимается задача разделения множества объектов заданной выборки на две группы (класса).

Типичный процесс построения формирования модели машинного обучения [32], в т.ч. при классификации состоит из этапов:

- Сбор данных в установленной форме;
- Нормализация данных: очистка от выбросов и пропусков в данных, перевод данных в формат, который может использовать модель машинного обучения;
- Конструирование признаков [33] – индивидуальных измеримых характеристик объекта, которые, как предполагает разработчик модели, могут иметь связь с принадлежностью объекта к своему классу;
- Обучение и оценка модели;
- Внедрение модели в автоматизированную систему в виде программного модуля (в нашем случае - в телемедицинскую систему дистанционного мониторинга);

Под задачей порядковой классификации [34] понимается разделение множества объектов выборки на заданное число упорядоченных классов по характеристике объекта. Для получения характеристики объекта необходимо провести одно или несколько измерений объекта. В случае, когда все объекты разделены на непересекающиеся непустые классы, такая характеристика позволяет однозначно отнести объект из общего множества к одному из классов.

Порядок между классами может задаваться порядком между диапазонами характеристики объектов, свойственных определённым классам. В таком случае, если один класс «больше» по порядку другого, то это значит, что каждый объект первого класса будет иметь большее значение характеристики, чем каждый объект второго класса.

В текущей работе при рассмотрении задачи порядковой классификации рассматриваются только те классы, между которыми отношение порядка выглядит как "класс 1" < "класс 2" < ... < "класс n", где n – общее количество классов. Отношение порядка можно представить в виде числовой прямой, лучом или отрезком (в зависимости от области допустимых значений характеристики объекта), разделённой на отдельные непересекающиеся диапазоны, свойственные каждому классу (Рисунок 3). Характеристика объектов задаётся в соответствии с поставленной проблемой.

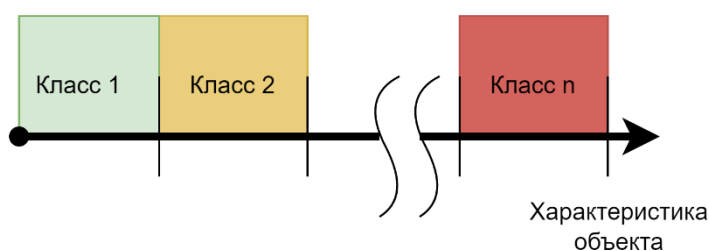


Рисунок 3 – Схема отношения порядка между классами по заданной характеристике объекта («Класс 1» < «Класс 2» < «Класс 3»)

Бинарная и порядковая классификация широко применяются в телемедицинских системах дистанционного мониторинга для анализа кардиологических данных. Бинарная классификация используется при установлении одного из двух состояний пациента (например, наличия или отсутствия аритмии). Порядковая классификация используется при установлении одного из нескольких упорядоченных между собой состояний пациента (например, определение стадии сердечной недостаточности).

Нередки случаи, когда в результате исследования полученная размеченная по классам выборка оказывается довольно мала в силу ограниченного времени на полноценный сбор данных или ограниченного

доступа к данным в целом, а для проведения нового исследования нет достаточных средств. Например, в медицине, когда исследуются редкие заболевания или уникальные хирургические операции.

В случае наличия малой выборки, когда для построения порядкового классификатора размеченная по классам выборка обладает достаточным количеством данных (объектов и измерений объектов) «крайних» классов (диапазоны характеристики этих классов находятся по краям на прямой/луче/отрезке характеристик объектов из рисунка 3), но малым количеством данных остальных классов, представляется возможным построить на имеющихся данных модель бинарной классификации, способную распознавать «крайние» классы. Наличие серии повторных измерений одного и того же объекта на выборке позволяют построить бинарный классификатор, принимающий во внимание каждое измерение объекта. Проблему недостаточной размеченной выборки возможно решить генерацией достаточной по размеру выборки имитационным моделированием. Однако предварительно требуется разработать методику генерации имитированных данных (имитационного моделирования выборки), позволяющую учесть проблему нехватки размеченных данных, а затем разработать способ формирования порядковой модели классификации по имитированным данным.

1.5. Рассмотрение прикладной проблемы создания медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по кардиологическим данным.

Скрининг в здравоохранении - активное выявление болезни или предболезненного состояния у лиц, считающихся или считающих себя здоровыми. [35] Особенностью скрининга является его массовость, скрининговые программы обычно реализуют во многих развитых странах в том числе при государственной поддержке.

Скрининг населения необходимо проводить, когда заболевание является важной проблемой здоровья, коими являются такие нарушения углеводного обмена (НУО), как сахарный диабет 2 типа (СД2) и предиабет.[36]

Без проведения скрининга СД2 и предиабета пациент, как правило, узнает о своем диагнозе СД2 спустя 7-9 лет от начала заболевания случайно или в ходе диспансеризации, когда у него уже есть заболевания, сочетающиеся с СД2: ожирение, ишемическая болезнь сердца, артериальная гипертензия и т.д. [37]

Для массового проведения скрининга СД2 и предиабета важно подобрать такой скрининговый тест, который был бы точным, валидизированным [35], безопасным, дешевым и доступным для применения по всей стране.

Рекомендованными тестами являются тест глюкозы плазмы натощак, пероральный глюкозотолерантный тест (ПГТТ) с 75 г глюкозы и анализ крови на гликированный гемоглобин. [38]

Однако скрининг на наличие НУО в масштабах страны по рекомендованным тестам еще не проводился. Это можно объяснить принципиальными недостатками рекомендуемых скрининговых тестов по крови: высокая стоимость из-за расходов на медицинских работников и лабораторное оборудование, инвазивность, необходимость наличия лабораторий и квалифицированного мед. персонала в первичном звене. Скрининг по данным рекомендованным тестам сложно провести там, где нет нужного персонала/оборудования (например, в фельдшерско-акушерских пунктах в сельской местности).

Для решения задачи проведения по стране сортировки пациентов на «здоровых», кому не требуется наблюдение эндокринолога, и тех, кому оно нужно, необходима новая методы выявления признаков НУО: дешевая, доступная, неинвазивная. Под «признаком НУО» имеется в виду признак,

присутствующий в анализируемых данных пациента с НУО и отсутствующий у пациента без НУО.

Из-за необходимости в адекватных, дешевых и технологичных способах проведения скрининга ведется поиск в области других физических принципов, на которых могут быть основаны способы проведения скрининга на наличие НУО. Обнаружение НУО по ЭКГ очень привлекательно, поскольку в таком случае пациенту для прохождения скрининга понадобится лишь мобильный электрокардиограф для съема и передачи ЭКГ на анализ. Учитывая, что в России доступ к Интернету есть у большинства населения (на начало 2023 года в России насчитывалось 127,6 млн интернет-пользователей – 88,2% от общего числа населения), анализ ЭКГ возможно проводить удаленно от мест снятия ЭКГ в едином центре обработке данных.

Вероятными способами обнаружения признака НУО являются способы, включающие спектральный анализ ЭКГ и анализ variability сердечного ритма (включая спектральный анализ КИГ)

Основой построения медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности электрокардиограмм сердца первого отведения может стать платформа Pegaspera Cardio, способная принимать ЭКГ по Интернету, проводить предварительную обработку, формировать КИГ и выдавать результаты анализа ЭКГ и КИГ (включая спектральный анализ).

Известно, что СД2 является признанным фактором риска сердечно-сосудистых заболеваний, и что ещё задолго до начала СД2 на фоне формирующегося предиабета начинают развиваться макро- и микрососудистые осложнения [39], прогрессируя по мере утяжеления отклонений уровня глюкозы от нормы. Это позволяет построить гипотезу о том, что у пациентов с нормально функционирующим сердцем, и у пациентов сердечными патологиями, вызванным СД2/предиабетом, будут отличаться ЭКГ, в т.ч. на уровне форм спектров, как была приведена в работе [40]

гипотеза о том, что динамика электрической активности нормально функционирующего сердца может быть интерпретирована феноменом автовозврата ФПУ, а нарушения в сердце приводят к нарушению автовозврата ФПУ.

Предпосылками возможности определения НУО по ВСР и по спектру КИГ, в частности, являются следующее: у людей с дисфункцией вегетативной нервной системы - высокой частотой сердечных сокращений (ЧСС) и низкой ВСР - имеется повышенный риск развития диабета [41]. Одной из причин, по которой НУО влияет на параметры ВСР, приводится то, что нарушение вегетативной функции связано с повышенными концентрациями инсулина в сыворотке крови и снижением чувствительности к инсулину. Высокое содержание глюкозы в крови у лиц с сахарным диабетом вызывает деградацию микроциркуляторного русла, что приводит к вегетативной нейропатии.

Также у лиц, страдающих предиабетом, наблюдалось снижение некоторых параметров ВСР по сравнению с пациентами без НУО – например, снижение стандартного отклонения нормальных (без внеочередных преждевременных сердечных сокращений) RR-интервалов. [42] В качестве причины этому приводилось то, что предиабет связан со снижением парасимпатической активности при регуляции ритма сердца и сдвигом в сторону повышения симпатического тонуса.

Таким образом, теоретически возможно определить, какой набор значений параметров ВСР больше свойственен пациенту с НУО (СД2 или предиабет), а какой набор – пациенту без НУО.

Существует множество исследований, где предпринимались попытки построить математическую модель выявления НУО (СД2) по параметрам ВСР. В таблице 1 приведены работы, в которых описаны и оценены данные модели.

Таблица 1 – Модели выявления НУО (СД2) по параметрам ВСР

Ссылка на работу, год публикации	Методы анализа ВСР	Качество моделей выявления СД2 (классификации)
[43], 2013	Анализ параметров, полученных из КИГ: скатерограмма, recurrence plots, correlation dimension Модель классификации: Adaboost	Точность - 86%, Чувствительность - 87.5 Специфичность - 84.6%.
[44], 2013	Анализ параметров, полученных из КИГ: Recurrence plots, Approximate Entropy, Largest Lyapunov exponent, Detrended fluctuation analysis Модель классификации Adaboost	Точность - 90%, Чувствительность - 92.5 Специфичность - 88.7%.
[45], 2015	Анализ параметров, полученных путем дискретного вейвлет-преобразования КИГ Модель классификации: дерево решений	Точность: 92% Чувствительность: 93% Специфичность: 91%
[46], 2013	Анализ параметров, полученных путем разложения КИГ на «спектры высшего порядка» (Higher Order Spectra) Модель классификации: метод опорных векторов (SVM)	Точность: 80% Чувствительность: 71% Специфичность: 89%
[47], 2016	Анализ параметров, полученных путем разложения КИГ на «эмпирические моды» Модель классификации: метод наименьших квадратов опорных векторов (LS-SVM)	Точность: 95.63% Чувствительность: 97.5% Специфичность: 93.7%
[48], 2018	Анализ изображений КИГ Модель классификации: Сверточная нейронная сеть + сеть с долгой краткосрочной памятью	Точность: 95.1%
[49], 2019	Анализ изображений спектрограмм, полученных из КИГ Модель классификации: Сверточная нейронная сеть	Точность: 97.62% Чувствительность: 100% Специфичность: 96.72%

Под «точностью» (accuracy) в таблице 1 понимается общая доля долю правильных классификаций.

Под «чувствительностью» (sensitivity) в целом понимается «число лиц, точно классифицированных по результатам исследования, как находящихся в определенном состоянии, деленное на число всех лиц в этом состоянии». [50] Конкретно для описанной задачи выявления НУО это означает долю пациентов с СД2, кого модель отнесла к классу «СД2». В терминах машинного обучения «чувствительность» является «полнотой» (recall) по классу «СД2».

Под «специфичностью» (specificity) в целом понимается «число лиц, правильно классифицированных по результатам исследования, как не находящихся в определенном состоянии, деленное на число всех лиц, не находящихся в определенном состоянии». [50] Конкретно для описанной задачи выявления НУО это означает долю пациентов без СД2, кого модель отнесла к классу «Здоров». В терминах машинного обучения «чувствительность» является «полнотой» по классу «Здоров».

Несмотря на предъявляемую высокую точность/чувствительность/специфичность моделей, данные модели сложно интерпретировать: в одних исследованиях используются параметры, полученные методами нелинейной динамики (Approximate Entropy, Detrended Fluctuation Analysis, correlation dimension) и которым сложно дать физиологическую интерпретацию, в других исследованиях используются нейронные сети. Сложность интерпретации затрудняет применение методов в клинической практике. Также по данным исследованиям нет информации, проходили ли полученные модели в дальнейшем клиническую валидацию, и в данных работах рассматриваются только выборки, к которым доступ ограничен.

Публикаций с разработками и оценкой моделей обнаружения признаков предиабета на ЭКГ по анализу ВСР, обнаружено не было.

По результатам рассмотрения работ, касающийся связи НУО и параметров ВСР (в частности, параметров спектров КИГ), а также публикаций, указанных в таблице 1, сделан вывод, что:

- значения некоторых параметров ВСР имеют статистическую связь с наличием у пациента НУО (СД2 или предиабета), что позволяет предварительно судить о принципиальной возможности построить модель выявления признаков (кардиопризнаков) НУО у пациента по параметрам ВСР, применение которого позволит решить *задачу раннего обнаружения предиабета и СД2* у населения в группе риска в масштабах России;
- при повторных исследованиях ВСР невозможно получить полностью идентичные результаты в связи с тем, что вегетативная нервная система постоянно реагирует на условия окружающей среды. Поскольку при проведении удаленного скрининга НУО неизбежно несоблюдение пациентом строгой методики записи ЭКГ, однократного анализа ВСР у пациента может быть недостаточно для устойчивой работы модели выявления кардиопризнаков НУО у пациента по параметрам ВСР. При разработке такой модели необходимо предусмотреть возможность интегральной оценки риска наличия НУО (СД2 или предиабета) по нескольким ЭКГ, снятым пациентом.
- существующие модели выявления наличия НУО (только СД2) по ЭКГ, несмотря на высокое качество, не проходили клиническую валидацию. Поэтому на текущий момент использовать их как часть метода выявления признаков наличия НУО у пациента при скрининге затруднительно.

По результатам рассмотрения работ, касающийся связи НУО и параметров спектра ЭКГ, сделаны выводы, что:

- Отсутствуют публикации с результатами разработки и клинической валидации моделей выявления наличия НУО у пациента по данным спектра ЭКГ;
- Теоретически, руководствуясь гипотезой о наличии различного характера автовозврата ФПУ между пациентами с нормально функционирующим сердцем и пациентами с сердечными патологиями, вызванным СД2/предиабетом, возможно построить метод выявления наличия НУО у пациента по данным спектра ЭКГ, который позволит решить *задачу раннего обнаружения предиабета и СД2* у населения в группе риска в масштабах России;
- В силу различия степени негативного влияния предиабета и СД2 на сердечно-сосудистую систему, при разработке метода выявления наличия НУО по Фурье-спектру ЭКГ необходимо предусмотреть возможность интегральной оценки риска наличия НУО для категоризации «низкий риск НУО»/«подозрение на предиабет»/«подозрение на СД2» на основе серии ЭКГ.

Для проведения скрининга населения по ЭКГ с использованием моделей машинного обучения необходимо разработать и задействовать телемедицинскую систему дистанционного мониторинга, действующей модули классификации по кардиологическим данным на основе машинного обучения. Также для облегчения процесса формирования модулей классификации по спектрам кардиологических данных, встраиваемых в систему, разработчику системы необходимо предоставить инструмент (фреймворк) формирования данных модулей.

Согласно данным Росстата, в России число жителей в группе риска по возрасту (от 45 лет) на момент 1 января 2022 года – около 16,3 млн человек (26%) сельского населения и 45,7 млн человек (74%) – городского. Учитывая, что лаборатории для анализа крови, в основном, есть только в городской среде, покрытие населения в группе риска НУО при скрининге по ЭКГ увеличится

на 33% за счет включения сельского населения, малообеспеченного лабораториями для анализа крови, но обеспеченного доступом в Интернет.

Объектом автоматизации при создании данной системы являются процессы:

- сбора и обработки ЭКГ, включая формирование КИГ и преобразования ЭКГ и КИГ в амплитудно-частотные спектры;
- предоставления пользователям результатов анализа собственных ЭКГ.

1.6. Выводы по 1 главе

В рамках данной главы дано общее представление о телемедицинских системах дистанционного биомониторинга и о роли в них модулей классификации состояния здоровья пациента на основе машинного обучения при анализе кардиологических данных.

Анализ существующих телемедицинских систем дистанционного биомониторинга, фреймворков машинного обучения, и возможностей спектрального анализа кардиологических данных показал, что:

- анализ спектров кардиологических данных (в частности, ЭКГ) при принятии решения о пациенте используется редко, однако имеет свою полезность и может быть успешно внедрен в рассматриваемые системы в виде моделей машинного обучения;
- для ускорения разработки телемедицинских системах дистанционного мониторинга, использующих спектральный анализ кардиологических данных, необходим специализированный фреймворк, для которого также требуется разработать отдельно методику формирования выборки кардиологических данных и отдельно метод построения моделей классификации (бинарной и порядковой) по кардиологическим данным.

Также рассмотрена необходимость в создании медицинскую систему самостоятельного управления здоровьем за счет скрининга состояния

пациента по последовательности электрокардиограмм сердца первого отведения для выявления по серии спектров ЭКГ признаков нарушения углеводного обмена по кардиологическим данным с помощью моделей машинного обучения.

Целью данной работы является разработка и реализация новых методов формирования модулей классификации, использующих модели машинного обучения по амплитудно-частотным спектрам кардиологических данных. В ходе исследования для достижения поставленной цели в работе поставлены следующие задачи:

- Разработать метод построения моделей машинного обучения с учителем для бинарной классификации объектов при помощи анализа амплитудно-частотного спектра (Фурье-спектра) кардиологических данных;
- Разработать метод построения моделей порядковой классификации объектов в условиях ограниченной выборки при помощи имитационного моделирования выборки по результатам работы бинарного классификатора;
- Разработать методику формирования выборки данных для создания моделей классификации по спектрам кардиологических данных.
- Разработать и оценить модели обнаружения кардиопризнаков НУО на спектрах кардиологических данных в двух вариантах с помощью разработанных методов: на основе анализа Фурье-спектра ЭКГ и на основе анализа параметров ВСР (включая сформированные из Фурье-спектра кардиоинтервалограмм);
- Разработать с помощью с помощью разработанных методов и оценить интегральную модель оценки риска наличия предиабета или СД2 по серии ЭКГ;
- Разработать фреймворк для формирования модулей классификации состояния здоровья пациента с использованием разработанных методов

построения моделей машинного обучения с учителем для бинарной и порядковой классификации по спектрам кардиологических данных.

- Разработать телемедицинскую систему проведения неинвазивного удалённого скрининга НУО по ЭКГ на основе разработанного фреймворка.

Глава 2. Описание разработанных методов построения моделей машинного обучения

В данной главе описываются разработанные в рамках работы:

- Метод построения моделей машинного обучения с учителем для бинарной классификации объектов по амплитудно-частотным спектрам кардиологических данных;
- Методика формирования выборки данных для создания моделей классификации по спектрам кардиологических данных;
- Метод построения моделей машинного обучения с учителем для порядковой классификации объектов с повторными измерениями;
- Прикладное применение разработанных методов при формировании моделей выявления кардиопризнака НУО на отдельной ЭКГ как бинарных классификаторов и модели интегральной оценки риска наличия НУО по серии ЭКГ как порядкового классификатора.

При разработке моделей машинного обучения, основанных на анализе спектров кардиологических данных, стоит следующий набор задач классификации:

- Определить принципиальную возможность выявления крайних случаев состояния пациента по кардиологическим данным. Например, «болен»/«здоров» или «истощение»/«превосходное функциональное состояние».
- Определить возможность выявления промежуточных случаев по кардиологическим данным. Например, «заболевание в хронической стадии»/«заболевание в зарождении»/«отсутствие болезни» или «истощение»/«неполное восстановление»/«превосходное функциональное состояние».

В случае необходимости выявления одного из двух крайних состояний предлагается формировать модели машинного обучения как бинарные

классификаторы по амплитудно-частотному спектру произвольной ЭКГ (что чаще имеет место в случаях определения болезни), либо спектру КИГ (что чаще имеет место в случаях определения функционального состояния).

В случае необходимости выявления одного из множества промежуточных состояний построение модели машинного обучения как бинарного классификатора является недостаточным, т.к. такая модель принципиально не способна определять по объекту промежуточные состояния. В таком случае предлагается строить порядковый классификатор как вариант «обогащения» бинарного классификатора, позволяющий принимать решение по объекту (пациенту), основываясь не на одном измерении (спектре кардиологических данных), а на серии измерений.

Далее методы построения моделей машинного обучения описаны в п. 2.1 (бинарный классификатор) и 2.3 (порядковый классификатор). Методика подготовка выборки кардиологических данных, включая преобразование в амплитудно-частотные спектры для приведена в п. 2.2.

2.1. Метод построения моделей машинного обучения с учителем для бинарной классификации объектов по амплитудно-частотным спектрам кардиологических данных

На рисунке 4 представлена схема работы метода построения моделей машинного обучения с учителем для бинарной классификации объектов по амплитудно-частотным спектрам кардиологических данных.

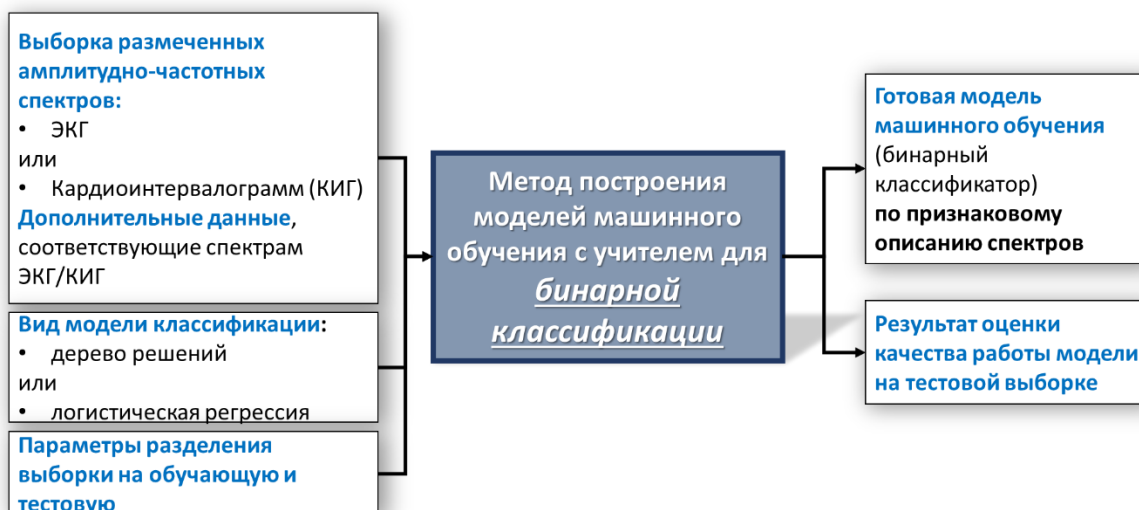


Рисунок 4 – Схема работы метода построения моделей машинного обучения с учителем для бинарной классификации объектов по амплитудно-частотным спектрам кардиологических данных

Для построения модели машинного обучения (бинарной классификации) необходимо подать следующие данные на входе:

- Выборку амплитудно-частотных спектров кардиологических данных: спектров ЭКГ или спектров КИГ, получаемых из ЭКГ. Каждый спектр должны быть размечен как принадлежащий к одному или другому классу. Спектры должны быть представлены как числовой массив амплитуд спектра и числом, выражающим шаг спектра в частной области. Спектры должны содержать информативные частотные области. Для спектров ЭКГ — это области от 0 до 40 Гц [51], для спектров КИГ — области от 0,015 до 0,15 Гц [21].
- Вид данных, из которых получены спектры: ЭКГ, либо КИГ;
- (опционально) Дополнительные данные, соответствующие каждому спектру ЭКГ/КИГ, касающиеся анализируемого пациента: пол, масса, возраст, статистические параметры variability сердечного ритма, рассчитанные из КИГ;
- Вид модели бинарной классификации, которую ожидается получить в результате работы метода. Поскольку при анализе кардиологических

данных ожидается наличие возможности интерпретировать результаты модели, ожидаемыми видами модели являются дерево решений, логистическая регрессия, либо отдельная модель оценки по степени хаотизации гармоник (только для спектров ЭКГ, общий вид модели и принимаемых параметров и описан в п. 2.1.1).

- Параметры разделения выборки на обучающую и тестовую выборки - выражается в процентном соотношении.

Метод работает следующим образом:

- 1) Спектры данных распределяются случайным образом на обучающую и тестовую выборки в соответствии с заданными параметрами распределения;
- 2) Над каждым спектром выполняется задание интересующих частотных областей:
 - a. Если вид данных, из которых получены спектры, - ЭКГ, то выполняется поиск гармоник путем построения огибающей спектра путем сглаживания и вычисления локальных максимумов на огибающей.

Первая гармоника называется базовой. В качестве интересующих частотных областей спектра выделяются окрестности с 1-й по 20-ю гармоник путем задания частотного диапазона ± 0.5 базовой частоты относительно частоты каждой гармоники.
 - b. Если вид данных, из которых получены спектры, - КИГ, то в качестве интересующих частотных областей спектра выделяются диапазоны 0,015–0,04 Гц (очень низкочастотный), 0,04–0,15 Гц (низкочастотный), 0,15–0,4 Гц (высокочастотный) и 0,015–0,4 Гц (общий). Данные диапазоны имеют клиническое значение и используются при анализе variability сердечного ритма [21]
- 3) Над каждым спектром выполняется расчет признакового описания спектра:

- a. Если вид данных, из которых получены спектры, - ЭКГ, то для каждой частотной области гармоника выполняется расчет степени хаотизации как отношение медианной к максимальной амплитуде внутри частотной области.
 - b. Если вид данных, из которых получены спектры, - кардиоинтервалограмма, то для каждой частотной области выполняется расчет мощности спектра: VLF- мощность спектра в очень низкочастотном диапазоне, LF- мощность спектра в низкочастотном диапазоне, HF - мощность спектра в высокочастотном диапазоне, TP – общая мощность спектра.
- 4) На наборе полученных признаков описаний спектров из обучающей выборки выполняется обучение модели классификации в соответствии с выбранным видом модели и полученными на входе гиперпараметрами
- a. Обучение модели классификации как логистической регрессии или как дерева решения предлагается проводить с помощью существующих алгоритмов обучения [52-55]
 - b. Обучение модели классификации как модели оценки по степени хаотизации гармоник описано отдельно в п. 2.1.1.
- 5) На наборе полученных признаков описаний спектров из тестовой выборки выполняется оценка качества модели классификации - вычисление точности и полноты модели относительно каждого класса.
- Точность (precision - P) определяется по формуле $P = TP / (TP + FP)$, где TP — количество истинно положительных результатов, а FP — количество ложноположительных результатов.
- Полнота (recall - R) определяется по формуле $R = TP / (TP + FN)$, где TP — количество истинно положительных результатов, а FN — количество ложноотрицательных результатов.

- 6) Обученная модель классификации преобразуется в вид пригодный для внедрения в программный модуль. Для логистической регрессии модель представляется в виде функции, выражающей соответствующую математическую формулу, для дерева решений - в виде функции, выполняющей проход вниз по дереву.

Выходными данными метода являются:

- 1) Модель машинного обучения (бинарной классификации) в пригодном виде для внедрения в программный модуль.
- 2) Результат оценки качества модели классификации - точность и полнота модели относительно каждого класса.

2.1.1. Модель оценки по степени хаотизации гармоник

Предлагаемая модель предназначена для обучения по параметрам, рассчитываемым из гармоник амплитудно-частотных спектров ЭКГ. [5]

Данная модель основана на работе [19], где ЭКГ представлена как модель сердца в виде нелинейной автоколебательной системы, где на амплитудно-частотном спектре ЭКГ наблюдается явление автовозврата (повторения во времени картины состояния нелинейной системы) Ферми–Паста–Улама (ФПУ), а нарушения в сердце приводят к нарушению автовозврата ФПУ. Предполагается, что на кардиограмме здорового человека можно обнаружить перекачку энергии сигнала между низкочастотным (1-5 Гц) и высокочастотным (20-35 Гц) диапазонами в виде шума на спектре ЭКГ.

Для сокращения шанса переобучения модели параметрами (признаковым описанием) модели поставлены не все точки спектра ЭКГ, а параметры отдельных гармоник спектра, где в качестве параметра гармоника, иллюстрирующего энергию сигнала на частотном диапазоне в окрестности гармоника, представлена степень хаотизации ch_i , вычисляемая по формуле (1):

$$ch_i = \frac{\text{Медианная амплитуда в области гармоника } i}{\text{Максимальная амплитуда в области гармоника } i} \quad (1)$$

где i – номер гармоники

Частотный диапазон каждой гармоники представляет собой окрестность вокруг гармоники с радиусом в половину базовой частоты спектра (частота первой гармоники).

Для оценки общей хаотизации спектра предлагается вычислить общую статистику из всех степеней хаотизации интересующих гармоник – степень хаотизации n -ного номера гармоники после упорядочивания гармоник по возрастанию хаотизации, что является вариацией квантиля степени хаотизации.

Для выполнения задач классификации, модель оценки по степени хаотизации гармоник представляет собой сравнение оценки общей хаотизации спектра с пороговой величиной:

$$Y = \begin{cases} 1, & \text{если } \sum_i^I f(ch_i) > \mathbf{num}, \text{ при } i \in I, I \subseteq \{1, 2 \dots 20\} \\ 0, & \text{иначе} \end{cases} \quad (2)$$

$$f(ch_i) = \begin{cases} 1, & \text{если } ch_i < \mathbf{ch}^* \\ 0, & \text{иначе} \end{cases} \quad (3)$$

$$ch_i = \frac{\text{Медианная амплитуда в области гармоники } i}{\text{Максимальная амплитуда в области гармоники } i} \quad (4)$$

где Y - логическое значение, обозначающее наличие ($Y=1$) или отсутствие ($Y=0$) положительного результата модели;

i – номер гармоники спектра;

ch_i – уровень хаотизации i -й гармоники;

I - множество номеров гармоник, учитываемых моделью. $|I|$ - мощность множества I (количество номеров гармоник, учитываемых моделью);

\mathbf{num} - пороговое количество гармоник из множества I , уровень хаотизации которых меньше пороговой величины \mathbf{ch}^* ;

\mathbf{ch}^* - пороговая величина степени хаотизации.

Обучение данной модели проводится следующим образом:

- Выборка из признаков описаний, полученных из обучающей выборки спектров ЭКГ, случайным образом делится 4 к 1 на обучающую и валидационную подвыборки.
- На обучающей подвыборке проводится полный перебор следующих гиперпараметров:
 - I – одно из сочетаний множества номеров гармоник $\{1, 2 \dots 20\}$;
 - num – целое число от 1 до мощности множества I .

Далее при каждом варианте набора значений гиперпараметров проводится перебор пороговой величины степени хаотизации от 0 до 1 с шагом 0,01 и для данного варианта сохраняется промежуточная модель с текущим набором значений гиперпараметров и такой пороговой величиной, при которой сумма полноты каждого класса максимальна.

- На валидационной подвыборке проводится оценка всех получивших промежуточных моделей и выбирается наилучшей та, при которой при которой сумма полноты каждого класса максимальна.
- Далее обучающую и валидационную подвыборки заново объединяются в выборку. На ней наилучшая модель повторно обучается при имеющихся гиперпараметрах: заново проводится перебор пороговой величины степени хаотизации от 0 до 1 с шагом 0,01, и в модели сохраняется такая пороговая величина, при которой сумма полноты каждого класса максимальна.

2.2. Методика формирования выборки спектров кардиологических данных для решения задач классификации

Приведем общую схему формирования выборки спектров кардиологических данных для решения задач классификации:

- Сформировать требования к выборке данных (метаданных о пациенте и снятых пациентом кардиологических данных). Отдельно

сформировать требования к исходным кардиологическим данным и их спектрам;

- Провести сбор/поиск выборки данных согласно требованиям;
- Провести предобработку кардиологических данных в соответствии с их спецификой и их преобразование в спектры.

2.2.1. Формирование требований к выборке данных и к исходным кардиологическим данным и их спектрам

Для работы предлагаемого метода построения моделей классификации в п. 2.1 необходимо сформировать выборку спектров данных, чтобы с помощью нее было возможно получить достаточно качественную модель классификации с хорошей обобщающей способностью (способностью давать правильные результаты на новых данных, которые не принимают участия в процессе обучения модели). Особенностью предлагаемой методики формирования выборки спектров является необходимость учета требований к исходным сигналам и спектрам как к результатам преобразования сигналов.

Предварительно необходимо подготовить 2 размеченные выборки с исходными сигналами ЭКГ: на которой будет формироваться и обучаться модель машинного обучения («обучающая»), и на которой модель будет оцениваться («тестовая»). Распространенной практикой является установка соотношения размера между обучающей и тестовыми выборками 3 к 1.

Также необходимо определить какой объект предполагается классифицировать. Если предполагается классифицировать отдельную ЭКГ по её спектру или пациента по спектру отдельной ЭКГ (например, для классификации состояния пациента в момент времени), то объектом является спектр ЭКГ. Если предполагается классифицировать пациента по серии спектров снятых им ЭКГ (например, для классификации состояния пациента на протяжении некоторого интервала времени), то объектом является пациент, а спектры ЭКГ пациента являются измерениями объекта.

Каждая запись из выборки должна включать следующие данные:

- Метаданные о пациенте – идентификатор, класс и данные, подтверждающие класс с помощью «золотого стандарта». Опционально – вес, рост, возраст и т.д.
- Сигнал ЭКГ с метаданными сигнала – идентификатор, длительность сигнала, частота дискретизации. Опционально – серийный номер, марка устройства съема и т.д.

Минимальное количество требуемых объектов предполагается определять:

- по согласованию с экспертами при планировании исследования;
- на основе допустимой ошибки метрик качества.

Рассмотрим второй вариант. В качестве примера данной метрики можно взять полноту по каждому классу, поскольку полнота по определенному классу рассчитывается целиком на объектах этого класса и является долей правильно «угаданных» моделью объектов этого класса, а это означает, что возможно рассчитать минимальное количество пациентов или ЭКГ пациентов каждого класса ($n_{\text{класс}}$) на основе формулы (5) минимально необходимого объема выборка на основе расчета доверительного интервала [56]:

$$n_{\text{класс}} = \hat{p}_{\text{класс}} * (1 - \hat{p}_{\text{класс}}) * \left(\frac{z_{\alpha}}{\Delta}\right)^2 \quad (5)$$

где $\hat{p}_{\text{класс}}$ – ожидаемая доля исследуемого признака (в данном случае - полноты по конкретному классу). Если исследование проводится впервые и/или ожидаемой доли не получить из прошлых исследований, то $\hat{p}_{\text{класс}}$ присваивают значение 0.5

z_{α} – квантиль стандартного нормального распределения уровня α - уровня доверительности (обычно α берут равным 95% или 99%, z_{α} соответственно равен 1,96 или 2,58)

Δ – допустимая ошибка ожидаемой доли исследуемого признака (полноты по каждому классу). Определяется исследователем, заказчиком исследования, или другим лицом, принимающим решение.

То есть в случае, когда ранее не было исследований, на основе которых оценивались схожие модели классификации, и ожидаемую полноту по каждому классу получить неоткуда, то при допустимой ошибке полноты в 0,05 и уровне доверительности 95% для каждого класса необходимо получить не менее $n_{\text{класс}} = 0,5 * 0,5 * \left(\frac{1,96}{0,05}\right)^2 \approx 385$ объектов каждого класса. Стоит отметить, что на тестовой выборке желательно иметь репрезентативную выборку по классам относительно генеральной совокупности, иначе оценки метрик качества, зависящих от распределения классов, например, точность (precision) – доля верно классифицированных объектов среди всех объектов, которые к этому классу отнесла модель – могут получиться смещенными. Это означает, что если на тестовой выборке имеется $n_{\text{класс}}$ объектов самого редкого класса на генеральной совокупности, то количество объектов остальных классов должно быть пропорционально распределению классов на генеральной совокупности.

На обучающей выборке возможно не предъявлять требования к репрезентативности, поскольку методы формирования моделей машинного обучения позволяют проводить присвоение разных весов каждому классу в зависимости от его доли в данных (взвешивание классов). Однако, поскольку на практике обучающая выборка больше тестовой больше в 3 раза, минимальный размер обучающей выборки должен быть увеличен соответствующим образом.

Требования к спектрам сигналов напрямую зависят от соответствия требованиям исходного сигнала, которые по возможности необходимо искать в стандартах, либо (при отсутствии стандартов) других публикаций.

Далее приведены требования к исходным сигналам ЭКГ для построения из них спектров ЭКГ и отдельно – для построения спектров кардиоинтервалограмм (КИГ) – , которые можно использовать в методе построения моделей, описанном в 1.7:

- 1) Для спектра ЭКГ есть требование к тому, что ширина полосы спектра должна составлять от 0,01 до 40 Гц с минимальным шагом спектра по частоте 0,01 Гц (при частоте дискретизации ЭКГ – 1000 Гц). [51]. Если для преобразования ЭКГ в спектр используется алгоритм быстрого преобразования Фурье, такое достижимо, если сигнал самой ЭКГ состоит из минимум 65536 точек. Это значит, что, при частоте дискретизации 1000 Гц минимальная длительность ЭКГ должна составлять 66 секунд, при частоте дискретизации 500 Гц – 132 секунд.

Однако необходимо учесть, что сигналы ЭКГ на выборке могут быть сформированы пациентами самостоятельно, без привлечения медицинского персонала (на пример, с помощью «умных» часов), из-за чего следует увеличить минимальную длительность ЭКГ до 180 секунд, чтобы при её обработке была возможность для удалять дефекты и оставалась возможность построить спектр ЭКГ с требуемыми характеристиками. Длительность ЭКГ, большая чем 3 минуты может быть слишком неудобной для пациента, поэтому рассматриваются ЭКГ с частотой дискретизации от 500 Гц и выше.

- 2) Спектр КИГ используется для расчёта спектральных параметров ВСР, а именно для вычисления мощностей спектра в отдельных частотных диапазонах. Ширина полосы спектра КИГ должна охватывать диапазон 0,015-0,4 Гц с шагом по частотной области до 0,001 Гц. Для расчета спектральных параметров ВСР необходима исходная ЭКГ длительностью не менее 2 минут, согласно [57]. Для расчета параметров ВСР на практике задействуют ЭКГ, на которых есть не менее 100 кардиоциклов [58], т.е. не менее 80 секунд при средней длительности кардиоцикла в 0,8 секунд.

Однако также из учета того, что пациент может снимать ЭКГ самостоятельно, минимальную длительность ЭКГ следует увеличить

до 3 минут, чтобы при автоматизированной обработке была возможность удалять из КИГ аномальные RR-интервалы (отличающиеся по длительности от соседних RR-интервалов более чем на 20%) и оставалась возможность вычисления мощностей спектра КИГ. Оптимальной частотой дискретизации считается 250-500 Гц (допускается выше) [57], так как низкое значение частоты дискретизации ЭКГ может вызывать серьезные погрешности в определении времени появления R-пиков. Длительность ЭКГ от 1 часа не рассматривается, поскольку для ЭКГ такой приводятся отдельные требования и проводятся другие исследования. [21]

2.2.2. Обработка кардиологических данных и их преобразование в спектры

Способы предварительной обработки отличаются в зависимости от вида исследуемых кардиологических данных.

Если исследуемыми данными являются исходные сигналы ЭКГ, то схема обработки состоит из:

- Устранение дрейфа изолинии на сигнале, вызванных дыханием или движением пациента;
- Устранение амплитудных выбросов на сигнале, вызванных потерей контакта электродов с кожей пациента;
- Преобразование сигнала в спектр алгоритмом быстрого преобразования Фурье.

Поскольку итоговыми данными являются спектры ЭКГ, необходимо использовать методы предобработки, не влияющие напрямую на спектр в исследуемой ширине полосы частот (от 0,01 до 40 Гц). Например, для устранения дрейфа изолинии не подойдут фильтры частот, но подойдут алгоритмы вычитания скользящего среднего [59] или их вариации, а примером устранения амплитудных выбросов подойдет на основе анализа статистических характеристик кардиоцикла [60].

Если исследуемыми данными являются сигналы кардиоинтервалограмм, то схема предварительной обработки состоит из следующих шагов:

- Устранение артефактов на исходной ЭКГ: удаление дрейфа изолинии и сглаживание;
- Поиск R-пиков и измерение длительностей между соседними R-пиками (RR-интервалов). Примером алгоритма поиска R-пиков является алгоритм Пана-Томпкинса [61];
- Формирование кардиоинтервалограммы (КИГ) из ряда длительностей RR-интервалов и удаление аномальных RR-интервалов. Обычно RR-интервал считается аномальным, если его длительность отличается от соседних интервалов более чем на 20%;
- Расчет статистических параметров variability сердечного ритма на КИГ [21] (перечень рассчитываемых параметров VCP вместе с описанием указан в Приложении);
- Интерполяция КИГ из оставшихся RR-интервалов (NN-интервалов) для преобразования в ряд интерполированных дискретных значений КИГ как непрерывной функции величин NN-интервалов, определенных на моментах времени их появления;
- Преобразование полученного ряда интерполированных дискретных значений КИГ в спектр алгоритмом быстрого преобразования Фурье (спектр КИГ);
- Расчет спектральных параметров variability сердечного ритма на спектре КИГ [21].

Итого после преобразования кардиологических сигналов в спектры, каждая запись из выборки будет включать следующие данные:

- Метаданные о пациенте – идентификатор, класс и данные, подтверждающие класс с помощью «золотого стандарта». Опционально – вес, рост, возраст и т.д.

- Метаданные сигнала – идентификатор, длительность сигнала, частота дискретизации. Опционально – серийный номер, марка устройства съема и т.д.
- Спектр кардиологического сигнала (ЭКГ или КИГ, в зависимости от выбора исследователя).

На рисунке 5 представлен пример спектра ЭКГ: по оси абсцисс откладывается частота в Герцах, по оси ординат - амплитудная характеристика исходного сигнала ЭКГ, проявляющаяся на данной частоте. Основная информационная составляющая ЭКГ находится в части спектра в диапазоне от 0 до 40 Гц [41], поскольку именно на этих частотах наблюдается явление автовозврата Ферми-Паста-Улама, рассматриваемое в п. 1.2.

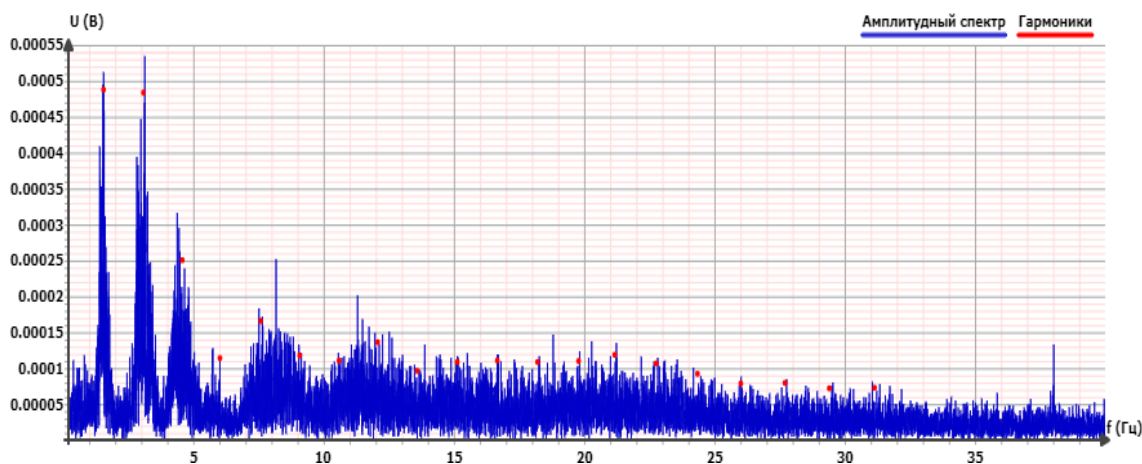


Рисунок 5 – Пример амплитудного спектра ЭКГ в диапазоне частот от 0 до 40 Гц (красным цветом выделены гармоники).

2.3. Метод построения моделей машинного обучения с учителем для порядковой классификации объектов с повторными измерениями

На рисунке 6 представлена схема работы метода построения моделей машинного обучения с учителем для порядковой классификации объектов с повторными измерениями по оценке работы бинарного классификатора на

порядковых классах путем имитационного моделирования (метод Монте-Карло).

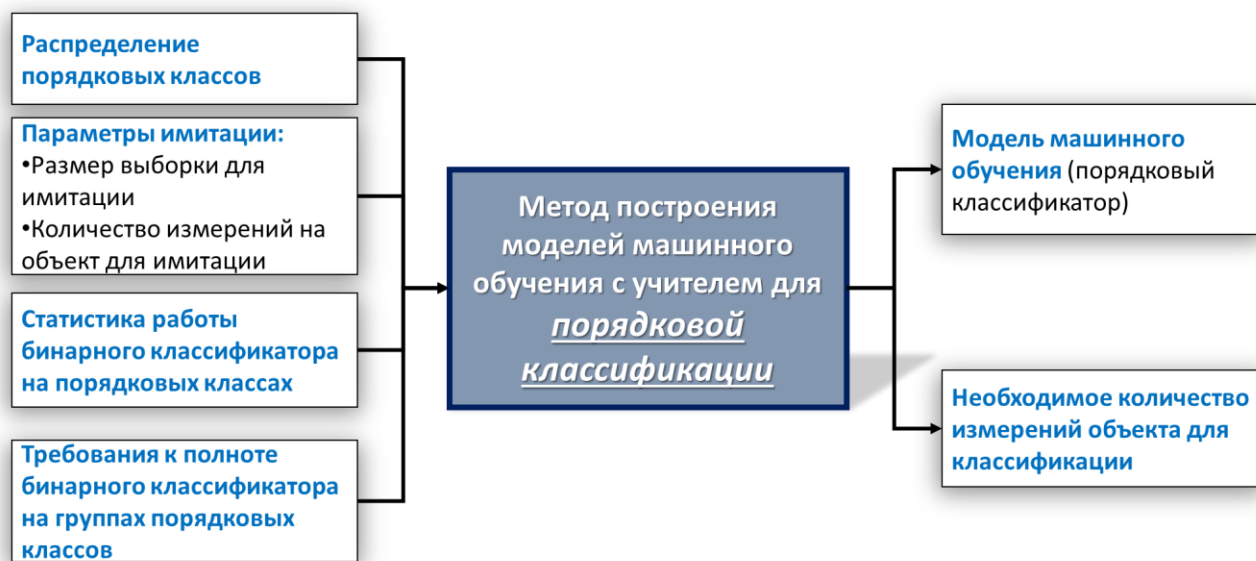


Рисунок 6 – Схема работы метода построения моделей машинного обучения с учителем для порядковой классификации объектов с повторными измерениями

Особенность данного метода заключается в том, что он позволяет учесть ситуацию, когда для построения порядкового классификатора размеченная по классам выборка обладает достаточным количеством данных (объектов и измерений объектов) «крайних» классов, но малым количеством данных остальных классов.

Предполагается, что на имеющейся обучающей выборке уже построена модель бинарной классификации (бинарный классификатор), способная распознавать «крайние» классы.

Предлагается проблему недостаточной размеченной выборки решить генерацией достаточной по размеру выборки методом Монте-Карло (имитационным моделированием). Такая выборка будет состоять из:

- выборки объектов со случайным распределением классов, соответствующим *известному* распределению классов на генеральной совокупности;

- серии измерений каждого объекта. Измерения генерируются как случайные результаты работы бинарного классификатора («положительный»/«отрицательный результат»), соответствующие известному распределению результатов его работы на измерениях объектов каждого класса на тестовой части выборки.

Построение данной модели проводится в 2 этапа:

- 1) Имитационное моделирование выборки;
- 2) Построение порядкового классификатора по результатам имитационного моделирования.

Суть имитационного моделирования выборки данных заключается в следующем: сгенерировать для формирования порядкового классификатора размеченную выборку объектов случайного класса и для каждого объекта сгенерировать случайным образом измерения - результаты работы бинарного классификатора («положительный»/ «отрицательный результат»).

Случайная генерация класса объекта должна соответствовать информации о распределении классов (соотношении количества объектов каждого класса) в генеральной совокупности. Случайная генерация измерений объекта должна соответствовать распределению положительных/отрицательных результатов на измерениях объектов соответствующего класса в тестовой части реальной выборки. Случайная генерация проводится с использованием генератора псевдослучайных чисел.

Для построения модели порядковой классификации необходимо выполнить подать следующие данные на входе:

- Распределение порядковых классов среди анализируемых объектов в процентном соотношении;

При наличии n классов должна быть известна доля ($prev$) объектов каждого класса на генеральной совокупности. Требование приведено как формула (6).

$$prev_1, prev_2 \dots prev_n, \forall prev_i \in (0; 1), i \in \{1, 2 \dots n\} \quad (6)$$

Несоблюдение данного требования не позволит провести имитацию репрезентативной выборки – следовательно, по данной выборке не удастся изначально построить адекватную модель порядковой классификации.

- Параметры имитации выборки– количество объектов (N) и измерений (m) на объект

Данные параметры напрямую задают размер имитированной выборки.

В силу того, что имитационное моделирование выборки проводится методом Монте-Карло, от количества объектов зависит ошибка всех статистических вычислений, полученных из выборки, и скорость моделирования. В рамках оценки качества классификации рассматривается метрика полноты (доля истинно положительных классификаций) по каждому классу по формуле (7).

$$R = \frac{TP}{TP+FN}, \quad (7)$$

где R (recall) – полнота по определенному классу на анализируемой выборке, TP (true positive)-количество истинно-положительных случаев работы модели классификации по определенному классу на анализируемой выборке, FN (false negative)- количество ложно-отрицательных случаев работы модели классификации по определенному классу на анализируемой выборке.

Чем больше будет число требуемых объектов для имитации, тем больше будет длительность моделирования выборки (линейная зависимость), но тем ниже будет ошибка при вычислении качества классификации (обратная квадратичная зависимость). Имеет смысл учитывать свойство периода генератора псевдослучайных чисел, используемого программистом, который будет моделировать выборку – количество моделируемых объектов и измерений не должно превышать период генератора в целях сохранения свойства независимости случайных величин.

От количества измерений на объект зависит как скорость моделирования, так и ожидаемая полнота бинарного классификатора по парам

групп классов с сохранением порядка. Пары групп классов с сохранением порядка в данной работе определяются как всевозможные разбиения множества классов на 2 непустые непересекающиеся группы при котором каждый класс из первой группы «меньше» по порядку каждого класса из второй группы. Например, если порядок между классами определен как «класс 1» < «класс 2» < ... «класс n», то рассматриваемыми парами групп являются пары ($\{\text{«класс 1»}\}, \{\text{«класс 2»} \dots \text{«класс n»}\}$), ($\{\text{«класс 1»}, \text{«класс 2»}\}, \{\text{«класс 3»} \dots \text{«класс n»}\}$) и т.д. до ($\{\text{«класс 1»}, \text{«класс 2»} \dots \text{«класс n-1»}\}, \{\text{«класс n»}\}$).

- Статистика работы бинарного классификатора на порядковых классах

Должна быть известна информация о доле положительных результатов (*res*) у существующей модели бинарной классификации по измерениям объектов каждого класса на тестовой части реальной выборки, требование приведено как формула (8). Без данной информации невозможно будет разметить измерения объектов имитационной выборки по признакам «положительный/отрицательный результат бинарного классификатора».

$$res_1, res_2 \dots res_n, \forall res_i \in [0;1], i \in \{1, 2 \dots n\} \quad (8)$$

Отдельным требованием является не только наличие информации о статистике работы бинарного классификатора на порядковых классах, но и наличие соответствия порядка между частотой возникновения положительного результата по каждому классу и порядком между самими классами, требование приведено как формула (9). Без соблюдения данного требования не будет увеличения ожидаемой полноты по парам групп классов при увеличении количества измерений на объект.

$$\begin{aligned} (\text{класс } 1 < \text{класс } 2 < \dots < \text{класс } n) &\rightarrow \\ &\rightarrow (res_1 < res_2 < \dots < res_n) \end{aligned} \quad (9)$$

Алгоритм имитационного моделирования выборки следующий:

1. Моделируются N объектов, каждый объект размечается путём присваивания класса случайным образом по заданному распределению $(prev_1, prev_2 \dots prev_n)$;
2. Для каждого объекта случайным образом моделируются m измерений - результатов работы бинарного классификатора - в соответствии с классом объекта и заданной в качестве входных данных доли возникновения положительного результата у бинарной модели классификации по соответствующему классу объекта;
3. Формируется статистика работы бинарного классификатора на парах групп классов с сохранением порядка – рассчитывается полнота по каждой группе классов относительно каждой пары при принятиях решения:
 - 3.1. по самому первому результату бинарного классификатора. Условием положительного результата по объекту является положительный результат классификатора по измерению;
 - 3.2. по первым двум результатам бинарного классификатора. В качестве условиями положительного результата по объекту отдельно рассматриваются:
 - 3.2.1. хотя бы один положительный результат классификатора по 2 измерениям;
 - 3.2.2. оба положительных результата классификатора по 2 измерениям;
 - 3.3. и т.д. до рассмотрения всех m измерений по каждому объекту и рассмотрения соответствующих условий положительного результата по объекту.

Полнота в паре групп классов рассчитывается по формуле (10).

$$\{R_{ij}^{\leq class}; R_{ij}^{> class}\} = \left\{ \frac{TP_{ij}^{\leq class}}{TP_{ij}^{\leq class} + FN_{ij}^{\leq class}}; \frac{TP_{ij}^{> class}}{TP_{ij}^{> class} + FN_{ij}^{> class}} \right\}, \quad (10)$$

где $R_{ij}^{\leq class}, R_{ij}^{>class}$ - полнота по группе классов, соответствующих условию « $\leq class$ »/« $> class$ », при принятии положительного решения по объекту в случае наличия хотя бы i положительных результатов бинарного классификатора на j измерений. Условие « $\leq class$ » означает, что в группе находятся все объекты класса $class$, либо «меньшего» класса по отношению порядка. Условие « $> class$ » означает, что в группе находятся все объекты «большого» класса, чем $class$ по отношению порядка.

$TP_{ij}^{\leq class}, TP_{ij}^{>class}$ – количество истинно-положительных случаев работы модели бинарной классификации по объектам группы классов, удовлетворяющих условию « $\leq class$ »/« $> class$ » на имитационной выборке.

$FN_{ij}^{\leq class}, FN_{ij}^{>class}$ – количество ложно-отрицательных случаев работы модели бинарной классификации по объектам группы классов, удовлетворяющих условию « $\leq class$ »/« $> class$ » на имитационной выборке.

Словосочетание «случайным образом» подразумевает применение генератора псевдослучайных чисел.

Итоговым результатом моделирования являются:

- выборка из N объектов, размеченных по классам, и m измерений объектов, размеченных по результату работы бинарного классификатора
- таблица с полнотой бинарного классификатора по объектам каждой группы классов относительно всех пар при количестве измерений каждого объекта от 1 до m (Таблица 2). Обозначение полноты по группе классов в таблице соответствует обозначению полноты в формуле (10).

Таблица 2 - Полнота бинарного классификатора по каждой группе классов при разной серии измерений каждого объекта

Измерений	Условие принятия положительного решения по объекту («хотя бы ... положительных из всех ... результатов»)	"Класс 1"/"Больше класса 1"		"Не меньше класса 2"/"Больше класса 2"		"Не меньше класса n-1"/"Класс n"	
		Полнота "Класс 1"	Полнота "Больше класса 1"	Полнота "Не меньше класса 2",	Полнота "Больше класса 2"	Полнота "Не меньше класса n-1"	Полнота "Класс n"
1	1 из 1	$R_{11}^{\leq 1}$	$R_{11}^{> 1}$	$R_{11}^{\leq 2}$	$R_{11}^{> 2}$	$R_{11}^{\leq n-1}$	$R_{11}^{> n}$
2	от 1 из 2	$R_{12}^{\leq 1}$	$R_{12}^{> 1}$	$R_{12}^{\leq 2}$	$R_{12}^{> 2}$	$R_{12}^{\leq n-1}$	$R_{12}^{> n}$
	2 из 2	$R_{22}^{\leq 1}$	$R_{22}^{> 1}$	$R_{22}^{\leq 2}$	$R_{22}^{> 2}$	$R_{22}^{\leq n-1}$	$R_{22}^{> n}$
...							
m	от 1 из m	$R_{1m}^{\leq 1}$	$R_{1m}^{> 1}$	$R_{1m}^{\leq 2}$	$R_{1m}^{> 2}$	$R_{1m}^{\leq n-1}$	$R_{1m}^{> n}$
	от 2 из m	$R_{2m}^{\leq 1}$	$R_{2m}^{> 1}$	$R_{2m}^{\leq 2}$	$R_{2m}^{> 2}$	$R_{2m}^{\leq n-1}$	$R_{2m}^{> n}$

	m из m	$R_{mm}^{\leq 1}$	$R_{mm}^{> 1}$	$R_{mm}^{\leq 2}$	$R_{mm}^{> 2}$	$R_{mm}^{\leq n-1}$	$R_{mm}^{> n}$

Полученная таблица с полнотой бинарного классификатора по каждой группе классов относительно всех пар при количестве измерений каждого объекта от 1 до m является основой для построения модели порядковой классификации по имитационным данным.

Суть принятия решения моделью по объекту следующая:

- Изначально присваиваем объекту класс 1.
- Если доля положительных срабатываний бинарного классификатора среди измерений объекта превышает определенный порог, то «повышаем» класс объекта до номера, соответствующего количеству превышений порогов. Если превышен 1 порог – то присваиваем объекту класс 2, если 2 порога – то класс 3 и т.д.

Характеристикой объекта для данной модели является доля положительных результатов работы бинарного классификатора среди всех измерений объекта (*ratio*).

Общий вид модели порядковой классификации представлен формулой (11):

$$Y = \begin{cases} \text{класс 1, если } ratio < cond_1 \\ \text{класс 2, если } cond_1 \leq ratio < cond_2 \\ \dots \\ \text{класс } n, \text{ если } cond_{n-1} \geq ratio \end{cases}, \quad (11)$$

$$\forall cond_{i \in [1, n-1]} = \frac{c_i}{a} \in [0; 1],$$

где Y - результат работы порядкового классификатора,

$cond_i$ - пороговая доля положительных результатов бинарного классификатора среди всех измерений объекта для присвоения объекту класса выше i ,

a - необходимое количество измерений на объект для принятия решения по объекту.

c_i - количество положительных результатов бинарного классификатора среди всех результатов для присвоения объекту класса выше i ,

Задача, которую надо решить для построения модели, сводится к поиску пороговых долей, удовлетворяющим образом разделяющих объекты на классы.

Входными данными для алгоритма построения порядкового классификатора являются:

- Таблица с полнотой бинарного классификатора по объектам каждой группы классов относительно всех пар при разном количестве измерений, построенная в результате имитационного моделирования выборки;
- Требования к полноте бинарного классификатора по объектам каждой группы классов относительно всех пар.

Задаются для обозначения, какое качество классификации порядкового классификатора будет удовлетворительным. Требования задаются как набор пар со значениями минимальной полноты бинарного классификатора по каждой группе классов по формуле (12).

$$\{R_{min}^{\leq 1}; R_{min}^{> 1}\}, \{R_{min}^{\leq 2}; R_{min}^{> 2}\}, \dots \{R_{min}^{\leq n-1}; R_{min}^{> n-1}\} \quad (12)$$

где $R_{ij}^{\leq class}, R_{ij}^{> class}$ - полнота по группе классов, соответствующих условию « $\leq class$ »/« $> class$ », при принятии положительного решения по объекту в случае наличия хотя бы i положительных результатов бинарного классификатора на j измерений. Условие « $\leq class$ » означает, что в группе находятся все объекты класса $class$, либо «меньшего» класса по отношению порядка. Условие « $> class$ » означает, что в группе находятся все объекты «большого» класса, чем $class$ по отношению порядка.

Стоит учитывать, что при увеличении требуемой полноты бинарного классификатора по любой из групп классов может увеличиваться необходимое количество измерений на объект, чтобы бинарный классификатор удовлетворял требованиям к полноте, поэтому возможны случаи, когда при слишком низком заданном количестве измерений m на объект при имитационном моделировании выборки модель не будет сформирована в силу отсутствия варианта модели, удовлетворяющей всем требованиям к полноте бинарного классификатора.

Алгоритм формирования порядкового классификатора, следующий:

1. Рассматриваем только первые a измерений всех объектов имитационной выборки (изначально $a=n-1$, меньше $n-1$ измерений на объект рассматривать не имеет смысла, т.к. не будет достаточного количества порогов для классификации по n классам)
2. Смотрим на значения полноты по группам классов из пары «класс b »/«больше класса b » из Таблицы 2 (изначально $b=1$).
3. Смотрим на значения полноты по группам классов при условии принятия положительного решения бинарным классификатором по объекту «если от c положительных результатов по измерению объекта из a измерений» (изначально $c=1$) – значения $R_{ca}^{\leq b}$ и $R_{ca}^{> b}$

3.1. Если они соответствуют требованиям к полноте бинарного классификатора ($R_{ca}^{\leq b} \geq R_{min}^{\leq b}$ И $R_{ca}^{> b} \geq R_{min}^{> b}$), то:

3.1.1. присваиваем $cond_b$ значение $\frac{c}{a}$;

3.1.2. Если $b=n$, то заканчиваем алгоритм с успешным результатом, заполняем модель порядковой классификации значениями $cond_1, cond_2 \dots cond_n$ с указанием, что необходимо не менее a измерений на каждый объект.

3.1.3. Иначе увеличиваем b и c на 1 и повторяем п. 2;

3.2. Иначе увеличиваем c на 1;

3.2.1. Если $c=a+1$, то увеличиваем a на 1, c присваиваем 1

3.2.1.1. Если $a>t$, то заканчиваем алгоритм с результатом, что построить модель порядковой классификации невозможно, т.к. требуется больше t измерений на объект.

3.2.1.2. Иначе повторяем п. 1;

3.2.2. Иначе повторяем п. 3;

В идеальной ситуации оценка итоговой модели порядковой классификации должна происходить на тестовой размеченной по классам выборке путем подсчета полноты и точности итоговой модели по каждому отдельному классу. Однако, поскольку изначально ставилась проблема нехватки размеченных данных при построении модели, предлагается оценивать полноту итоговой модели по каждому отдельному классу на имитационной выборке.

Выходными данными метода являются:

- 1) Модель порядковой классификации в виде формулы для внедрения в программный модуль;
- 2) Оценка полноты порядкового классификатора на имитационной выборке
- 3) Минимальное количества измерений (срабатываний бинарного классификатора) для работы модели.

Либо сообщение о невозможности построить модель порядковой классификации.

2.4. Применение методов построения моделей машинного обучения для классификации

В текущем разделе приведены варианты применения метода построения моделей машинного обучения для определения признаков наличия/отсутствия нарушения углеводного обмена (кардиопризнаков НУО) на спектрах кардиологических данных.

Первый вариант (п. 2.4.1) основан на анализе спектра ЭКГ как явления автовозврата Ферми-Паста-Улама.

Второй вариант (п. 2.4.2) основан на анализе вариабельности сердечного ритма, включающий в себя спектральный анализ кардиоинтервалограммы – динамического ряда RR-интервалов.

2.4.1. Выявление кардиопризнака НУО на спектре ЭКГ

В текущем разделе описана суть выявления кардиопризнаков НУО на амплитудно-частотном спектре (Фурье-спектре) отдельной ЭКГ. Принципиальная возможность выявления кардиопризнака НУО у пациента на основе спектрального анализа рассмотрена в п. 1.5.

Кардиопризнаком НУО является наличие сходства формы амплитудного спектра Фурье ЭКГ с формой спектра ЭКГ пациентов с НУО и отсутствие сходства с формой спектра ЭКГ у здоровых пациентов. В частности, кардиопризнак НУО выражается в сниженной хаотизации гармоник спектра, пояснение о понятии хаотизации спектра в п. 2.1.1.

На рисунке 7 показано предварительное сравнение ЭКГ и спектра, полученного из ЭКГ, одного пациента без НУО и одного пациента с СД2. Между формой самой ЭКГ пациента с СД2 и пациента без НУО принципиальных различий найдено не было, однако между формами ЭКГ пациентов были обнаружены различия.

На спектре ЭКГ «здорового» пациента (без НУО), видно перекачку энергии между у низкочастотным (1-5 Гц) и высокочастотным (20-35 Гц)

диапазонам в виде шума на спектре ЭКГ. Однако у пациента с СД2 форма спектра в областях гармоник была слабо зашумлена, можно увидеть четкую выраженность гармоник на спектре, форма спектра ЭКГ напоминает «расческу». Это может означать нарушение синхронизации процесса перетекания энергии между низкочастотным и высокочастотным диапазоном.

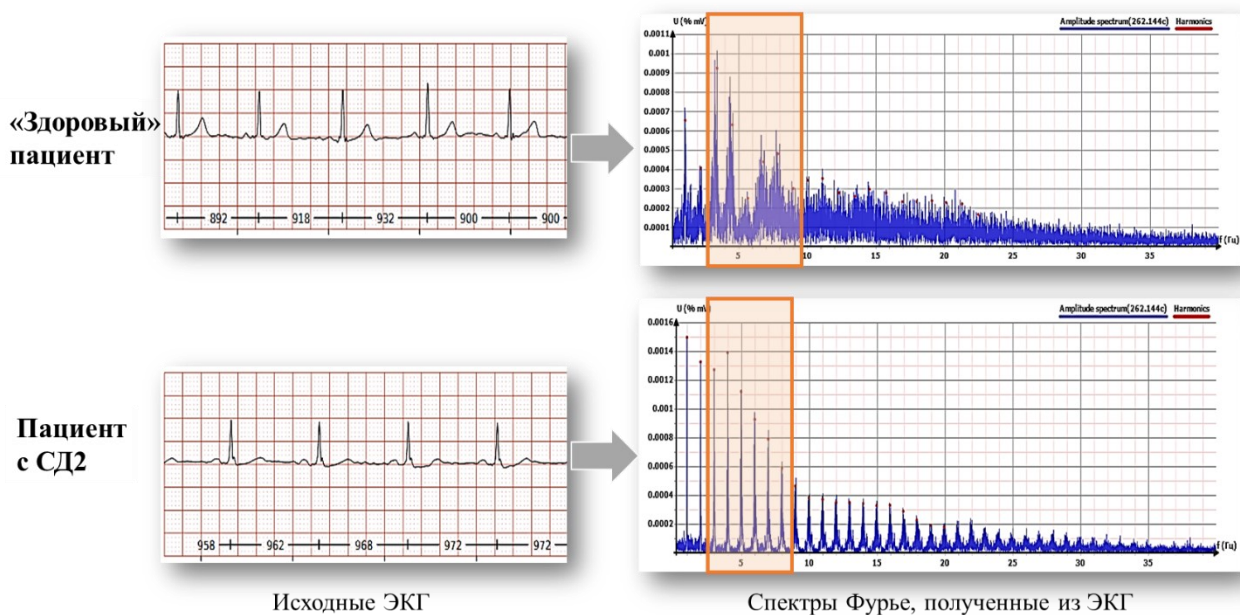


Рисунок 7 - Отличия между спектрами ЭКГ у пациента без НУО («здорового») и у пациента с СД2

В сердце как автоколебательной системе присутствуют хаотические колебания в отдельных областях гармоник. Эти характеристики хаотизации колебаний в областях гармоник решено использовать в качестве параметров модели выявления кардиопризнака НУО на ЭКГ. Причем чем больше форма спектра ЭКГ в области гармоник напоминает импульс с максимальной амплитудой в точке самой гармоники и нулевой амплитудой в окрестности, тем меньше хаотизация колебаний.

Построение модели выявления кардиопризнака НУО по спектру ЭКГ, согласно методу, описанному в п. 2.1, выполнялось следующим основным шагом:

- 1) Задание параметров метода - выборки спектров ЭКГ и параметров обучения в подсистеме машинного обучения для построения модели машинного обучения – модели выявления кардиопризнака НУО;
- 2) Построение модели выявления кардиопризнака НУО в соответствии с методом, ее обучение и предоставление результатов: итоговая модель и её оценка на тестовой выборке.

На момент построения модели выявления кардиопризнаков НУО отсутствовало достаточное количество исходных размеченных спектров ЭКГ. В связи с этим было решено провести клиническое исследование с целью определения принципиальной возможности определять по ЭКГ статистические значимые отклонения в кардиологических данных пациентов с НУО (сахарный диабет 2 типа) и пациентов без НУО.

Требуемое количество ЭКГ пациентов с НУО и без для построения модели изложено в дизайне исследования в п. 3.1.1, итоговая выборка ЭКГ и пациентов приведена в п. 3.1.2

Сбор ЭКГ и предварительная обработка ЭКГ велась согласно методике формирования выборки, описанной в п. 2.2.

В рамках решаемой задачи заданы следующие параметры метода построения моделей машинного обучения для бинарной классификации:

- Выборка – размеченные ЭКГ пациентов с НУО и без;
- Вид модели – модель оценки по степени хаотизации гармоник (описание модели и ее процесс обучения приведен в п. 2.1.1);
- Разделение выборки на обучающую и тестовую – в соотношении 4 к 1.

Процесс построения модели выявления кардиопризнака НУО ведется в соответствии с методом, описанным в п. 2.1, и указанными параметрами метода.

Общий вид модели выглядит следующим образом:

$$Y = \begin{cases} 1, & \text{если } \sum_i^I f(ch_i) > \mathbf{num}, \text{ при } i \in I, I \subseteq \{1, 2 \dots 20\} \\ 0, & \text{иначе} \end{cases} \quad (13)$$

$$f(ch_i) = \begin{cases} 1, & \text{если } ch_i < \mathbf{ch}^* \\ 0, & \text{иначе} \end{cases} \quad (14)$$

$$ch_i = \frac{\text{Медианная амплитуда в области гармоника } i}{\text{Максимальная амплитуда в области гармоника } i}, \quad (15)$$

где Y - логическое значение, обозначающее наличие ($Y=1$) или отсутствие ($Y=0$) признака НУО;

i – номер гармоники спектра, первая ($i=1$) гармоника находится на частоте, равной частоте сердечных сокращений (ЧСС), остальные гармоники – на частоте, равной $\text{ЧСС} \cdot i$;

ch_i – уровень хаотизации i -й гармоники, где медиану и максимум определяем среди набора значений амплитуд, расположенных в диапазоне частот ($\text{ЧСС} \cdot (i-1/2)$; $\text{ЧСС} \cdot (i+1/2)$);

I - множество номеров гармоник, учитываемых моделью. $|I|$ - мощность множества I (количество номеров гармоник, учитываемых моделью);

\mathbf{num} - пороговое количество гармоник из множества I , уровень хаотизации которых меньше пороговой величины \mathbf{ch}^* ;

\mathbf{ch}^* - пороговая величина степени хаотизации.

Итоговый вид модели после обучения на экспериментальных данных и результаты оценки модели приведены в п. 3.1.3.

2.4.2. Выявление кардиопризнаков НУО на КИГ с помощью параметров variability сердечного ритма (ВСР) и спектра КИГ

Кардиопризнаком НУО на отдельной ЭКГ для данного метода является наличие сходства анализируемых параметров ВСР с аналогичными параметрами ВСР пациентов с НУО и отсутствие сходства с аналогичными параметрами ВСР здоровых пациентов.

На рисунке 8 показано предварительное сравнение КИГ и спектра КИГ одного пациента без НУО и одного пациента с СД2. У пациента с СД2 был снижен разброс RR-интервалов, повышена средняя длительность RR-интервалов, и снижены мощности спектра, полученного из кардиоинтервалограммы (как в отдельных частотных диапазонах, так и целом).

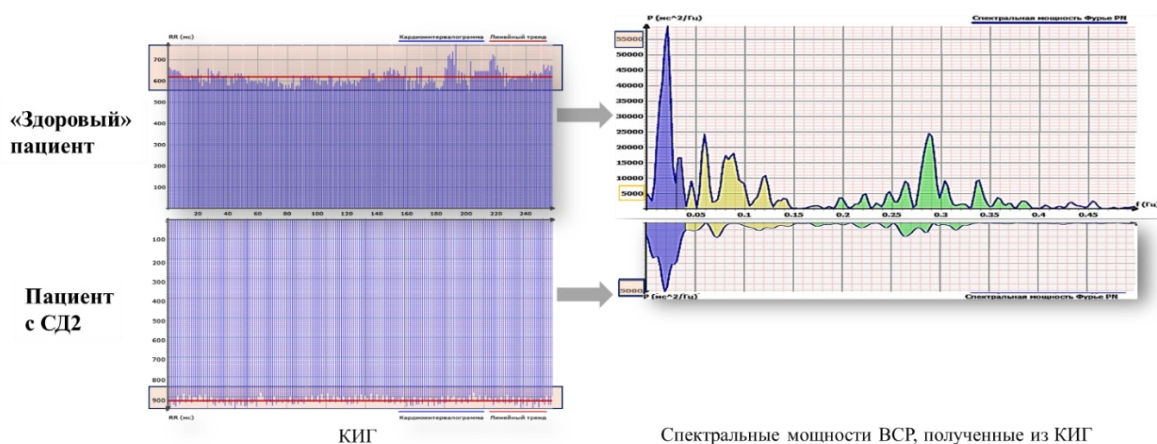


Рисунок 8 - Отличия между КИГ и спектральной мощностью ВСР у пациента без НУО («здорового») и у пациента с СД2

Таким образом, в этом случае, кардиопризнак НУО выразился в сниженной вариабельности RR-интервалов и сниженной спектральной мощности ВСР, т.е. выразился в сниженном уровне активности симпатической и парасимпатической отделов вегетативной нервной системы.

Дальнейшее клиническое исследование, описанное в п. 3.1, подтвердило указанное понятие кардиопризнака НУО и данные различия в кардиоинтервалограммах и спектральных параметрах ВСР между пациентами без НУО и пациентами с СД2.

Перечень рассчитываемых параметров ВСР вместе с описанием указан в Приложении. Расчеты каждого параметра в методе основываются на методических рекомендациях по анализу ВСР. [21, 62.]

Выявление кардиопризнака НУО на ЭКГ по анализу ВСР выполняется по следующим основным шагам:

- 1) Задание параметров метода - выборки спектральных и статистических параметров ВСР с данными о пациентах и параметров обучения в подсистеме машинного обучения для построения модели машинного обучения – модели выявления кардиопризнака НУО;
- 2) Построение модели выявления кардиопризнака НУО в соответствии с методом, ее обучение и предоставление результатов: итоговая модель и её оценка на тестовой выборке.

На момент построения модели выявления кардиопризнаков НУО отсутствовало достаточное количество исходных размеченных спектров ЭКГ внутри Pegaspera Cardio. В связи с этим было решено провести клиническое исследование с целью определения принципиальной возможности определять по ЭКГ статистически значимые отклонения в кардиологических данных пациентов с НУО (сахарный диабет 2 типа) и пациентов без НУО.

Требуемое количество ЭКГ пациентов с НУО и без для построения модели изложено в дизайне исследования в п. 3.1.1, итоговая выборка ЭКГ и пациентов приведена в п. 3.1.2

Сбор ЭКГ и предварительная обработка ЭКГ велась согласно методике формирования выборки, описанной в п. 2.2. Полный список рассматриваемых параметров ВСР приведен в Приложении.

В рамках решаемой задачи заданы следующие параметры метода построения моделей машинного обучения для бинарной классификации:

- Выборка – размеченные записи о спектральных и статистических параметрах ВСР пациентов с НУО и без, вес и рост пациентов с НУО и без;
- Вид модели – модель логистической регрессии;
- Разделение выборки на обучающую и тестовую – в соотношении 4 к 1.

После получения требуемой выборки данных процесс построения модели выявления кардиопризнака НУО ведется в соответствии с методом, описанным в п. 2.1, и указанными параметрами метода.

Модель выявления кардиопризнаков НУО представлена в виде логистической регрессии.

Общий вид модели выглядит следующим образом:

$$\begin{aligned}
 y = & a_1 * RRNN + a_2 * SDNN + a_3 * RMSSD + \\
 & + a_4 * \text{Эксцесс} + a_5 * \text{Ассиметрия} + a_6 * CV + a_7 * NN50 + \\
 & + a_8 * pNN50 + a_9 * Mo + a_{10} * Amo + a_{11} * NN50 + \\
 & + a_{12} * dRR + a_{13} * SATI + a_{14} * SI + a_{15} * SWAI + \\
 & + a_{16} * LF + a_{17} * HF + a_{18} * VLF + a_{19} * \frac{LF}{HF} + \\
 & + a_{20} * LFp + a_{21} * HFp + a_{22} * VLFp + a_{23} * TP + \\
 & + a_{24} * \text{Вес} + a_{25} * \text{Рост} + a_0,
 \end{aligned} \tag{16}$$

$$P = \frac{1}{1+e^{-y}} \tag{17}$$

$$Y = \begin{cases} 1, & \text{если } P < P^* \\ 0, & \text{если } P \geq P^* \end{cases} \tag{18}$$

где Y - логическое значение, обозначающее наличие ($Y=1$) или отсутствие ($Y=0$) признака НУО;

P^* - пороговое значение вероятности наличия признака НУО при данных параметрах ВСП;

a_0 – свободный член регрессии;

$a_1, a_2 \dots a_{25}$ – коэффициенты регрессии;

$RRNN, SDNN, RMSSD$ и пр. - параметры ВСП, расшифровка приведена в Приложении.

Общий вид модели выявления кардиопризнаков НУО требует обучения на экспериментальной выборке ЭКГ пациентов без НУО и с НУО.

Итоговый вид модели после обучения на экспериментальных данных и результаты оценки модели приведены в п. 3.1.3.

2.4.3. Выявления подозрений на НУО по серии спектров кардиологических данных

Каждая из предложенных в п. 2.4.1 и 2.4.2 моделей предназначена для выявления наличия/отсутствия кардиопризнака НУО на спектре отдельной ЭКГ или КИГ. Однако, данные методы дают бинарный ответ («есть кардиопризнак НУО на ЭКГ»/«нет»), поэтому по отдельному спектру ЭКГ нельзя дать ответ, есть ли у пациента высокий риск СД2, высокий риск предиабета или низкий риск НУО.

Поскольку СД2 влияет негативно на сердечно-сосудистую систему сильнее, чем предиабет, а предиабет – сильнее, чем отсутствие НУО, требуется построить отдельную модель интегральной оценки риска наличия предиабета или СД2 по серии спектров кардиологических данных.

Схема построения модели интегральной оценки риска наличия предиабета или СД2 по серии спектров с помощью метода построения модели порядковой классификации приведена на рисунке 9.



Рисунок 9 - Схема построения модели интегральной оценки риска наличия предиабета или СД2 по серии спектров кардиологических данных.

Построения модели интегральной оценки риска наличия предиабета или СД2 выполняется с помощью метода, описанного в п. 2.3, по следующим основным шагам:

- 1) Задание параметров метода:
 - распределение порядковых классов – распространенности в России сахарного диабета 2 типа (СД2) и предиабета;
 - параметры имитации – сколько объектов (пациентов) и измерений на объект (результатов работы модели выявления кардиопризнака НУО на ЭКГ по каждому пациенту) предлагается имитировать при имитационном моделировании
 - статистика работы бинарного классификатора на порядковых классах – доли выявления кардиопризнака НУО на спектрах ЭКГ или КИГ пациентов с СД2/предиабетом/здоровых
 - требования к полноте бинарного классификатора – с какой полнотой при множественных измерениях объекта бинарный классификатор должен определять пациентов класса «СД2», «Предиабет или Здоров», «СД2 или Предиабет», «Здоров» (слово «или» означает объединение классов)
- 2) Построение модели интегральной оценки риска наличия предиабета или СД2 и предоставление результатов: итоговая модель и необходимое количество ЭКГ для получения результата, соответствующим требованиям к полноте бинарного классификатора .

В рамках решаемой задачи, в подсистему машинного обучения заданы следующие параметры метода построения моделей порядковой классификации:

- распределения порядковых классов – класс «СД2» - 5,4%, класс «Предиабет» - 19,3%, класс «Здоров» - 75,3%. Данное распределение обусловлено распространенностью СД2 и предиабета в России среди населения в возрасте от 20 до 79 лет [1]

$$prev_1, prev_2 \dots prev_n, \forall prev_i \in (0; 1), i \in \{1, 2 \dots n\}$$

- параметр имитации – N=10 млн. пациентов, по m=11 измерений на пациента;
- статистика работы бинарного классификатора на порядковых классах – на момент постановки задачи неизвестна. Статистика работы моделей выявления кардиопризнака НУО на спектре ЭКГ/на основе анализа ВСР после клинической валидации приведена в п. 3.2.3.
- требования к полноте бинарного классификатора – требуемая полнота моделей выявления кардиопризнака НУО на спектре ЭКГ/на основе анализа ВСР при множественных измерениях пациента по классам «Здоров», «Предиабет или СД2», «Здоров или предиабет», «СД2» составила соответственно:

$$\begin{aligned} Rmin^{\text{Здоров}} &= 75\%; Rmin^{\text{Предиабет или СД2}} = 80\%; \\ Rmin^{\text{Здоров или предиабет}} &= 80\%; Rmin^{\text{СД2}} = 80\%; \end{aligned} \quad (19)$$

Получив заданные параметры, подсистема машинного обучения начинает построение модели порядковой классификации - модели интегральной оценки риска наличия предиабета или СД2 по серии ЭКГ

Модель интегральной оценки риска наличия предиабета или СД2 по серии спектров в общем виде представлена следующим образом:

$$DiabetesRisk = \begin{cases} 2, & \text{если } \frac{n_{Y=1}}{n} \geq p_{DM}^* \\ 1, & \text{если } p_{PreDM}^* \geq \frac{n_{Y=1}}{n} > p_{DM}^*, \quad n \geq n_{min} \\ 0, & \text{иначе} \end{cases} \quad (20)$$

где *DiabetesRisk* - значение, обозначающее высокий риск СД2 (2)/высокий риск предиабета(1)/низкий риск НУО (0);

n – количество ЭКГ, снятых пациентом за определенный интервал времени (например, в течение месяца или 6 недель) – длина серии ЭКГ;

n_{Y=1} - количество ЭКГ с выявленным кардиопризнаком НУО по выбранному методу, снятых пациентом за определенный интервал времени;

n_{min} – минимальная длина серии ЭКГ;

p_{DM}^* - пороговая доля ЭКГ пациента с выявленным кардиопризнаком НУО от всех ЭКГ пациента, после которой признается у пациента высокий риск СД2;

p_{PreDM}^* - пороговая доля ЭКГ пациента с выявленным кардиопризнаком НУО от всех ЭКГ пациента, после которой признается у пациента высокий риск предиабета.

n_{min} , p_{DM}^* и p_{PreDM}^* являются параметрами модели оценки риска наличия НУО, настраиваемыми в ходе обучения моделей выявления кардиопризнаков НУО на спектре отдельной ЭКГ или КИГ, и рассматриваемыми при валидации соответствующих моделей.

Данная модель построена на основе допущения, что у пациента с предиабетом кардиопризнак НУО на ЭКГ будет выявляться чаще, чем у пациента без НУО, а у пациента с СД2 – чаще, чем у пациента с предиабетом:

Итоговая модель интегральной оценки риска наличия предиабета или СД2 после проведения клинической валидации моделей выявления кардиопризнаков НУО на спектре ЭКГ или КИГ и минимальной длины серии спектров для работы итоговой модели приведена п. 3.2.4.

2.4.4. Планирование проверки качества моделей классификации

Для того, чтобы убедиться, что полученные модели выявления кардиопризнаков НУО на спектре ЭКГ/КИГ и модель интегральной оценки риска наличия предиабета или СД2 по серии спектров достаточно качественны для использования в телемедицинской системе дистанционного мониторинга при скрининге населения, необходимо оценить качество предложенных моделей путем вычисления:

- метрик качества моделей выявления кардиопризнака НУО на спектре отдельной ЭКГ/КИГ: чувствительность (полнота по классу «есть НУО»), специфичность (полнота по классу «нет НУО»), точность по классам «Нет НУО»/«Есть НУО».
- метрик качества моделей интегральной оценки риска подозрения на НУО у пациента по серии спектров: полнота и точность по классам «СД2»/«Предиабет»/«Здоров».

Полноту (R) моделей по каждому классу предлагается определять двумя способами :

- для моделей выявления кардиопризнака НУО на спектре отдельной ЭКГ или КИГ - как долю спектров ЭКГ или КИГ пациентов своего класса, которые модель предсказала верно, формулы (21), (22);
- для модели интегральной оценки риска подозрения на НУО у пациента по серии спектров— как долю пациентов своего класса, которые модель предсказала верно, формулы (23)-(25).

$$R_{\text{Нет НУО}} = \frac{\text{Спектров без кардиопризнака НУО у пациентов без НУО}}{\text{Спектров пациентов без НУО}} * 100\% \quad (21)$$

$$R_{\text{Есть НУО}} = \frac{\text{Спектров с кардиопризнаком НУО у пациентов с НУО}}{\text{Спектров пациентов с НУО}} * 100 \quad (22)$$

$$R_{\text{СД2}} = \frac{\text{Пациентов с СД2 и с подозрением на СД2}}{\text{Пациентов с СД2}} * 100\% \quad (23)$$

$$R_{\text{Предиабет}} = \frac{\text{Пациентов с предиабетом и с подозрением на предиабет}}{\text{Пациентов с предиабетом}} * 100\% \quad (24)$$

$$R_{\text{Здоров}} = \frac{\text{Здоровых пациентов без подозрений}}{\text{Здоровых паицентов}} * 100\% \quad (25)$$

Точность (P) моделей по каждому классу предлагается определять двумя способами :

- для моделей выявления кардиопризнака НУО на спектре отдельной ЭКГ или КИГ - как долю спектров ЭКГ или КИГ с присвоенным моделью классом, которые модель предсказала верно, формулы (26), (27);
- для модели интегральной оценки риска подозрения на НУО у пациента про серии спектров – как долю пациентов с присвоенным моделью классом, которые модель предсказала верно, формулы (28)- (30);

$$P_{\text{Нет НУО}} = \frac{\text{Спектров без кардиопризнака НУО у пациентов без НУО}}{\text{Спектров без кардиопризнака НУО}} * 100\% \quad (26)$$

$$P_{\text{Есть НУО}} = \frac{\text{Спектров с кардиопризнаком НУО у пациентов с НУО}}{\text{Спектров с кардиопризнаком НУО}} * 100 \quad (27)$$

$$P_{\text{СД2}} = \frac{\text{Пациентов с СД2 и с подозрением на СД2}}{\text{Пациентов с подозрением на СД2}} * 100\% \quad (28)$$

$$P_{\text{Преиабет}} = \frac{\text{Пациентов с преиабетом и с подозрением на преиабет}}{\text{Пациентов с подозрением на преиабет}} * 100\% \quad (29)$$

$$R_{\text{Здоров}} = \frac{\text{Здоровых пациентов без подозрений}}{\text{Пациентов без подозрений}} * 100\% \quad (30)$$

У «бесполезной» модели классификации, сравнимой с подбросом монетки, полнота по каждому классу равна 50%, а точность по своему классу будет равна доле содержания объектов (спектров или пациентов) этого класса на выборке

Для проверки полезности данных моделей поставим нулевые гипотезы о бесполезности модели («оцениваемая модель имеет ту же полноту/точность, что и бесполезная модель»). Гипотезы предлагается проверять оценкой значений полноты и точности каждой модели по каждому классу и построении от них 95% доверительных интервалов (95% ДИ) для долей. Если точность и

полнота «бесполезных» моделей не попадают в 95% ДИ соответствующей точности и полноты оцениваемых моделей, то нулевые гипотезы отвергаются и принимаются альтернативные. [63]

2.5. Выводы по 2 главе

Предложены два новых метода построения моделей классификации по спектрам кардиологических данных (для бинарной и порядковой классификации) и новая методика формирования выборки кардиологических данных. Данные методы и методику предлагается использовать при разработке моделей для телемедицинских систем дистанционного мониторинга по кардиологическим данным.

Дополнительно в рамках главы приведено прикладное применение:

- метода построения моделей бинарной классификации объектов по амплитудно-частотным спектрам кардиологических данных - для построения моделей выявления признаков нарушений углеводного обмена (кардиопризнаков НУО) на спектре ЭКГ и на параметрах variability сердечного ритма (ВСР) со спектрами КИГ.
- метода построения моделей для порядковой классификации объектов с повторными измерениями – для построения модели интегральной оценки риска наличия НУО по серии спектров кардиологических данных, позволяющей классифицировать пациента по классам «СД2»/«Предиабет»/«Здоров».

Данные модели требуют обучения и оценки на наборах ЭКГ пациентов с подтвержденным НУО и без в ходе отдельных клинических исследований.

Глава 3. Экспериментальная проверка разработанных моделей классификации на примере задачи обнаружения кардиопризнаков НУО по кардиологическим данным.

В данной главе:

- описаны результаты клинических исследований, проведенных с целью оценки применимости моделей классификации по спектрам кардиологических данных, построенных на основе разработанных методов построения моделей классификации. Оценены модели выявления кардиопризнака НУО на отдельной ЭКГ и модель интегральной оценки риска подозрения на НУО по серии ЭКГ) при скрининге:
 - Первое исследование было проведено в 2019 году на выборке пациентов с диагностированным сахарным диабетом, находящихся на госпитализации, и здоровых пациентов (НИР 2019).
 - Второе исследование было проведено в 2021 году на выборке пациентов без диагностированного сахарного диабета (НИР 2021), однако у некоторых пациентов был обнаружен сахарный диабет 2 типа и предиабет по результатам анализа крови на гликированный гемоглобин.
- Проведены количественные оценки потенциального снижения затрат на анализы крови при внедрении в процесс скрининг НУО медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности электрокардиограмм сердца первого отведения (ЭС СД2).

3.1 Экспериментальная оценка возможности выявления кардиопризнака НУО на кардиологических данных пациентов с СД2

Целью первого исследования (НИР 2019) была оценка принципиальной возможности устойчиво отличать кардиологические данные, принадлежащие больному НУО (СД2), от кардиологических данных, принадлежащих здоровому от НУО.

В рамках НИР 2019 стояли задачи:

- Собрать выборку ЭКГ пациентов с НУО (СД2) и без НУО согласно дизайну исследования и разработанной методике формирования выборки;
- обучить модели выявления кардиопризнаков НУО на обучающей подвыборке собранной в ходе НИР выборке кардиологических данных, описанные в п. 2.4.1 и 2.4.2 - формулы (13)-(15) и (16)-(18);
- оценить качество обученных моделей на тестовой подвыборке кардиологических данных.

3.1.1 Дизайн исследования

Исследование проводилось как одноцентровое проспективное наблюдательное исследование на базе центра - кафедры эндокринологии и диабетологии МГМСУ им. А.И. Евдокимова. Научный руководитель: Ашот Мусаелович Мкртумян, д.м.н., профессор, заведующий кафедрой эндокринологии и диабетологии МГМСУ им. А.И.Евдокимова.

Критерии включения и исключения пациентов приведены в таблице 3.

Таблица 3 - Критерии включения и исключения (НИР 2019)

	Пациенты с СД2	Здоровые пациенты
Критерии включения	<ul style="list-style-type: none"> Наличие диагноза «сахарный диабет 2 типа»; Период пребывания в эндокринологическом стационаре ГБУЗ «Городская Клиническая Больница №52» ДЗМ или в ГБУЗ «Московский Клинический Научный Центр имени А.С. Логинова» ДЗМ - более 5 дней; Возраст старше 18 лет; Подписанное информированное согласие пациента. 	<ul style="list-style-type: none"> Отсутствие аннотированного сахарного диабета 2-го типа; Нормальный уровень гликированного гемоглобина, измеренный не позднее 2 месяцев (для пациентов в возрасте от 45 лет); Возраст от 18 до 64 лет; Подписанное информированное согласие пациента;
Критерии невключения/исключения:	<ul style="list-style-type: none"> Наличие психических и неврологических расстройств, Неспособность к регистрации ЭКГ; Отсутствие комплаенса. 	<ul style="list-style-type: none"> Наличие психических и неврологических расстройств, Неспособность к регистрации ЭКГ; Отсутствие комплаенса.

Участие пациентов в исследовании проходило по следующей схеме:

- Пациенты с СД2, пребывающие в стационаре:

Пациенты наблюдаются во время пребывания в стационаре не менее 5 дней.

В первый день после ознакомления с информацией об исследовании и подписания информированного добровольного согласия, медперсонал регистрирует пациента в приложении CardioQVARK и вносит в приложение о следующие данные о пациенте:

- Вес, рост, дата рождения, пол;
- Окружность талии в см;
- Длительность диабета в годах;
- Уровень гликированного гемоглобина в процентах;
- Сопутствующие заболевания по классификатору МКБ 10;
- Принимаемые лекарственные препараты;
- Длительность гипертонической болезни в годах;
- Наличие/отсутствие инфаркта миокарда (при наличии с датой события);
- Наличие/отсутствие нарушения мозгового кровообращения (при наличии с датой события);
- Наличие/отсутствие курения;

После регистрации профиля, в первый, третий и пятый дни пациент под контролем медперсонала проходит следующие мероприятия:

- Регистрирует ЭКГ 1 стандартного (ст.) отведения при помощи устройства CardioQVARK в положении сидя, с опорой для спины, руки на коленях или на столе. Длительность измерения - 3 минуты.
- При помощи портативного глюкометра определяет уровень глюкозы в крови, затем вносит значение уровня глюкозы рядом с соответствующей ЭКГ в приложении CardioQVARK.
- Отправляет ЭКГ вместе с данными уровня глюкозы на обработку.

Данные мероприятия необходимо выполнять 6 раз в день: натощак, через 2 часа после завтрака, до обеда, через 2 часа после обеда, до ужина, через 2 часа после ужина. Итого на одного пациента с НУО ожидалось 18 ЭКГ и 18 соответствующих замеров уровня глюкозы. Схема участия пациента с НУО в НИР 2019 изложена в таблице 4.

Таблица 4 – схема участия пациента с НУО в НИР 2019

Дни наблюдения	День 1	День 3	День 5
Регистрация пациента через приложение CardioQVARK, внесение краткой истории болезни.	1		
Анализ крови на гликированный гемоглобин	1		
Снятие ЭКГ 1 ст. отведения.	6	6	6
Определение уровня глюкозы в крови с помощью глюкометра	6	6	6

Все данные отправлялись на хранение и обработку в автоматизированную систему проведения неинвазивного удалённого скрининга НУО по серии ЭКГ ЭС СД2 (подробности об ЭС СД2 в п. 4.3).

- Пациенты без НУО (здоровые пациенты):

Пациенты наблюдаются вне стационара не менее 7 дней.

В первый день после ознакомления с информацией об исследовании и подписания информированного добровольного согласия, медперсонал регистрирует пациента в приложении CardioQVARK и вносит в приложение о следующие данные о пациенте:

- Вес, рост, дата рождения, пол;
- Окружность талии в см;

- Уровень гликированного гемоглобина в процентах (для пациентов в возрасте от 45 лет);
- Сопутствующие заболевания по классификатору МКБ 10;
- Принимаемые лекарственные препараты;
- Наличие/отсутствие инфаркта миокарда (при наличии с датой события);
- Наличие/отсутствие нарушения мозгового кровообращения (при наличии с датой события);
- Наличие/отсутствие курения.

После регистрации профиля пациент проходит следующие мероприятия:

- Регистрирует ЭКГ 1 ст. отведения при помощи устройства CardioQVARK в положении сидя, с опорой для спины, руки на коленях или на столе. Длительность измерения - 3 минуты.
- При помощи портативного глюкометра определяет уровень глюкозы в крови, затем вносит значение уровня глюкозы рядом с соответствующей ЭКГ в приложении CardioQVARK.
- Отправляет ЭКГ вместе с данными уровня глюкозы на обработку.

Данные мероприятия необходимо выполнить всего 10 раз без регламентированного времени проведения наблюдений. Итого на одного здорового пациента ожидалось 10 ЭКГ и 10 соответствующих замеров уровня глюкозы.

Ожидаемая выборка для исследования согласно протоколу исследования:

- 1 000 наблюдений 100 пациентов различного пола и возраста старше 18 лет с СД2;

- 300 наблюдений 30 здоровых пациентов различного пола и возраста. Из них 15 пациентов в возрастной группе от 18 до 44 лет, 15 пациентов – от 45 до 64 лет.

Все данные отправлялись на хранение и обработку в телемедицинскую информационную систему ЭС СД2 (подробности об ЭС СД2 в главе 4). Итоговая выборка спектров кардиологических данных (ЭКГ и КИГ) формировалась на основе методики из п. 2.2.

Перед обучением моделей выявления кардиопризнака НУО выборка спектров кардиологических данных проходила следующую обработку.

- из всей выборки спектров рассчитывались параметры, необходимые для моделей с помощью ЭС СД2. Список параметров приведен в п. 2.4.1 и 2.4.2, формулы (13)-(15) и (16)-(18);
- выборка спектров с рассчитанными параметрами разделялась случайным образом на обучающую и тестовую подвыборки в соотношении 80:20 соответственно таким образом, чтобы спектры ЭКГ/КИГ одного и того же пациента распределялись строго в обучающей, либо строго в тестовой подвыборке.

Обучение моделей велось согласно методу построения моделей, описанному в п. 2.1.

3.1.2 Выборка

На площадках ГКБ №52 и МКНЦ им. Логинова получены аннотированные выборки ЭКГ пациентов с НУО (всего 1997 ЭКГ у 128 человек) и у пациентов без НУО (всего 741 ЭКГ у 59 человек). Соотношение числа собранных ЭКГ пациентов с НУО и ЭКГ пациентов без НУО – 73%/27%.

Регистрировали ЭКГ 1 отведения с помощью портативного электрокардиографа CardioQVARK длительностью 3 минуты с частотой дискретизации 1000 Гц.

Сводка по полученным экспериментальным данным приведена в таблице 5. Структура полученной выборки приведена в таблицах 5 и 6. Данные структурированы по половозрастному составу.

Таблица 5 - Сводка экспериментальных данных, НИР 2019

	МКНЦ им. Логинава			ГКБ №52		
	СД2	Без НУО	Итого	СД2	Без НУО	Итого
Пациентов	55	22	77	73	35	108
Снято ЭКГ	925	318	1243	1062	317	1379
Итого пациентов	128	59				
Итого ЭКГ	1997	741				
ЭКГ на пациента в среднем	16,82	14,45		14,55	9,06	

Таблица 6 - Структура полученной выборки, НИР 2019

С НУО

Пол	18-44 лет		45-64 лет		65+ лет		ЭКГ	Пац.	ЭКГ на пац.
	Пац.	ЭКГ	Пац.	ЭКГ	Пац.	ЭКГ			
Мужчины	3	40	31	507	19	264	811	53	15,30
Женщины	1	17	36	574	38	595	1186	75	15,81
Итого							1997	128	15,60

Без НУО

Пол	18-44 лет		45-64 лет		65+ лет		ЭКГ	Пац.	ЭКГ на пац.
	Пац.	ЭКГ	Пац.	ЭКГ	Пац.	ЭКГ			
Мужчины	10	140	1	12	0	0	152	11	13,82
Женщины	29	405	18	174	1	10	589	48	12,27
Итого							741	59	12,56
Всего							2738	187	14,64

3.1.3 Полученные модели классификации

Результатом обучения модели выявления кардиопризнака НУО по спектру КИГ и параметрам ВСР стала следующая модель:

$$Y = \begin{cases} 0, & \text{если } P < 0,61578 \\ 1, & \text{если } P \geq 0,61578 \end{cases} \quad (31)$$

$$P = \frac{1}{1 + e^{-y}} \quad (32)$$

$$y = (0,065101 * \text{Вес}) + (0,236656 * \text{Эксцесс}) \\ + (0,006246 * RRNN) + (0,003499 * SI) \\ - 11,724946 \quad (33)$$

где Y - логическое значение, обозначающее наличие ($Y=1$) или отсутствие ($Y=0$) кардиопризнака НУО;

Эксцесс - Коэффициент эксцесса NN-интервалов (RR-интервалов без артефактов и экстрасистол);

RRNN - Средняя длительность NN-интервалов, мс;

SI – Индекс напряжения.

Результатом обучения методов выявления кардиопризнаков НУО по спектру ЭКГ стала следующая модель:

$$Y = \begin{cases} 1, & \text{если } \sum_{i \in \{4,5,6,7,8,9,10\}} f(ch_i) > 3, \\ 0, & \text{иначе} \end{cases} \quad (34)$$

$$f(ch_i) = \begin{cases} 1, & \text{если } ch_i < 0,26 \\ 0, & \text{иначе} \end{cases} \quad (35)$$

$$ch_i = \frac{\text{Медианная амплитуда в области гармоники } i}{\text{Максимальная амплитуда в области гармоники } i} \quad (36)$$

где Y – логическое значение, обозначающее наличие ($Y=1$) или отсутствие

($Y = 0$) кардиопризнака НУО;

i – номер гармоники спектра, первая ($i=1$) гармоника находится на частоте, равной частоте сердечных сокращений (ЧСС), остальные гармоники – на частоте, равной $ЧСС \cdot i$;

ch_i – уровень хаотизации i -й гармоники, где медиану и максимум определяем среди набора значений амплитуд, расположенных в диапазоне частот ($ЧСС \cdot (i-1/2)$; $ЧСС \cdot (i+1/2)$)

Результаты обработки ЭКГ показали, что у пациентов с СД2 устойчиво определяются на ЭКГ кардиопризнаки НУО.

Рассматривались значения полноты и точности моделей по отдельным спектрам ЭКГ/КИГ пациента из тестовой подвыборки (таблица 7).

Таблица 7 – Полнота и точность моделей выявления кардиопризнаков НУО, отдельные ЭКГ, НИР 2019

	Модель по спектрам ЭКГ	Модель по параметрам ВСП и спектру КИГ анализа ВСП
Чувствительность, полнота по классу «Есть НУО», (95% дов. интервал)	77% (73% - 81%)	81% (77% - 85%)
Специфичность, полнота по классу «Нет НУО», (95% дов. интервал)	92% (86% - 95%)	93% (88% - 96%)
Точность по классу «Есть НУО», (95% дов. интервал)	96% (93% - 98%)	97% (94% - 98%)
Точность по классу «Нет НУО», (95% дов. интервал)	60% (53% - 66%)	64% (58% - 71%)

Поскольку по обоим моделям доверительные интервалы полноты не содержат неинформативного значения «50%», доверительные интервалы

точности по классу «Есть НУО» не содержат неинформативного значения «73%» (доля ЭКГ пациентов с НУО от ЭКГ всех пациентов выборки), а доверительные интервалы точности по классу «Нет НУО» не содержат неинформативного значения «27%» (доля ЭКГ пациентов без НУО от ЭКГ всех пациентов выборки), можно судить о статистически значимом различии качества построенных моделей выявления кардиопризнака НУО от «бесполезных» моделей вроде броска монетки.

Также на выборке рассматривались значения полноты модели интегральной оценки риска подозрения на НУО по первым трем снятым ЭКГ пациентов (таблица 8). Всего пациентов из выборки, снявших от 3 ЭКГ: 127 пациентов с НУО и 59 пациентов без НУО, соотношение – 68%/32%. Поскольку в выборке не было пациентов с предиабетом, модель интегральной оценки риска наличия НУО рассматривалась в усеченном виде (полный вид модели в п. 2.4.3).

$$DiabetesRisk = \begin{cases} 1, & \text{если } \frac{n_{Y=1}}{n} \geq p_{DM}^*, \quad n \geq n_{min} \\ 0, & \text{иначе} \end{cases} \quad (33)$$

где *DiabetesRisk* - значение, обозначающее высокий риск НУО(1)/низкий риск НУО (0);

n – количество ЭКГ, снятых пациентом– длина серии ЭКГ;

n_{Y=1} - количество ЭКГ с выявленным кардиопризнаком НУО по выбранному модели, снятых пациентом;

n_{min} – минимальная длина серии ЭКГ;

p_{DM}^{}* - пороговая доля ЭКГ пациента с выявленным кардиопризнаком НУО от всех ЭКГ пациента, после которой признается у пациента высокий риск НУО;

Таблица 8 – Полнота по классу пациентов «Есть НУО» (чувствительность) и полнота по классу пациентов «Нет НУО» (специфичность) модели интегральной оценки риска подозрения на НУО, первые 3 ЭКГ пациентов, НИР 2019 [64]

Длина серии ЭКГ (n_{min})	Условие подозрения на НУО ($n_{Y=1}$)	Модель по спектрам ЭКГ		Модель по параметрам ВСП и спектру КИГ анализа ВСП	
		Чув-ть (95% ДИ)	Спец-ть (95% ДИ)	Чув-ть (95% ДИ)	Спец-ть (95% ДИ)
1	1	70% (62%–77%)	86% (75%–93%)	83% (75%–88%)	84% (73%–92%)
2	1	83% (76%–89%)	84% (72%–91%)	87% (81%–92%)	80% (68%–88%)
	2	58% (50%–66%)	89% (80%–95%)	74% (66%–81%)	88% (77%–94%)
3	1	88% (81%–93%)	74% (60%–83%)	90% (83%–94%)	78% (66%–87%)
	2	76% (69%–84%)	86% (75%–93%)	82% (74%–88%)	81% (70%–89%)
	3	48% (40%–57%)	91% (82%–96%)	69% (61%–77%)	90% (80%–95%)

По сумме полноты:

- лучше всего модель интегральной оценки риска НУО с использованием модели выявления кардиопризнака НУО по параметрам ВСП и спектру КИГ анализа ВСП показала себя при выявлении подозрения на НУО по первой снятой ЭКГ;
- лучше всего модель интегральной оценки риска подозрения на НУО с использованием модели выявления кардиопризнака НУО по

спектрам ЭКГ показала себя при выявлении подозрения на НУО при выявлении хотя бы одного кардиопризнака НУО на первых двух снятых ЭКГ;

Доверительные интервалы полноты модели интегральной оценки риска подозрения на НУО по первым трем ЭКГ моделей (за исключением условия «На 3 ЭКГ из 3 выявлен кардиопризнак НУО» при использовании модели выявления кардиопризнака НУО по спектрам ЭКГ) также не содержат неинформативного значения «50%», что позволяет судить о статистически значимом различии качества построенных моделей выявления кардиопризнака НУО по первым 3 ЭКГ от «бесполезных» моделей.

3.1.4 Выводы исследования

- Собрана выборка ЭКГ пациентов с НУО (СД2) и без НУО. Всего собрано 1997 ЭКГ у 128 пациентов с НУО и 741 ЭКГ у 59 пациентов без НУО.
- На собранной выборке ЭКГ было проведено обучение и валидация моделей выявления кардиопризнаков НУО, описанных в п. 2.4.1 и 2.4.2.
- По результатам валидации моделей было выяснено наличие принципиальной возможности:
 - устойчиво отличать кардиологические данные, принадлежащие больному НУО, от кардиологических данных, принадлежащих здоровому от НУО;
 - устойчиво по серии из 3 первых ЭКГ отличать пациентов с НУО (СД2) от здоровых от НУО.

3.2 Валидация методов неинвазивного выявления НУО по ЭКГ

Целью НИР 2021 была валидация возможности применения методов неинвазивного выявления НУО по ЭКГ и реализующей его медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности электрокардиограмм сердца первого отведения (ЭС СД2) в первичном звене здравоохранения РФ.

В рамках НИР 2021 стояли задачи:

- Собрать выборку ЭКГ пациентов с НУО (СД2 или предиабет) и без НУО согласно дизайну исследования;
- оценить качество обученных моделей выявления кардиопризнаков НУО в ходе НИР 2019 на собранной выборке ЭКГ в ходе текущего исследования.

Отчет о НИР 2021 был утвержден генеральным директором НМИЦК им. ак. Е.И. Чазова, академиком РАН Бойцовым С.А.

3.2.1 Дизайн исследования

Исследование проводилось одноцентровое проспективное наблюдательное исследование на базе центра - НМИЦК им. ак. Е.И. Чазова. Данные собирались в поликлинике № 6 ГБУЗ МО Мытищинской Городской Клинической Больницы по методике, описанной в п. 2.2).

Критерии включения и исключения пациентов приведены в таблице 9.

Таблица 9 - Критерии включения и исключения (НИР 2021)

Критерии включения	<ul style="list-style-type: none"> • Мужчины и женщины в возрасте 40-79 лет с отсутствием в анамнезе диагноза «сахарный диабет»; • Подписано информированное согласие.
Критерии невключения/исключения:	<ul style="list-style-type: none"> • Возраст менее 40 и более 79 лет; • Наличие в анамнезе диагноза «сахарный диабет»; • Беременность; • Психические и неврологические расстройства; • Неспособность к регистрации ЭКГ; • Прием препаратов из группы бета-адреноблокаторов или группы бета-адреномиметиков; • Прием препаратов из группы тиазидных диуретиков или глюкокортикоидов в любой форме.

При отборе каждый пациент заполнял анкету, предназначенную для документирования показателей для оценки больного по критериям включения или невключения пациента в клиническое исследование.

Участие пациентов в исследовании проходило в течение одного дня по следующей схеме:

- 1) После ознакомления пациентом с информацией об исследовании и подписания информированного добровольного согласия, медперсонал регистрирует пациента в приложении CardioQVARK и вносит в приложение о следующие данные о пациенте: вес, рост, дата рождения, пол;
- 2) Производится съём первой ЭКГ 1 стандартного (ст.) отведения в течение 3 минут в положении сидя с помощью устройства CardioQVARK;

- 3) Производится забор крови для анализа на гликированный гемоглобин;
- 4) Производится забор капиллярной крови натошак с помощью глюкометров Ассu-Сhek;
- 5) Производится съем второй ЭКГ первого отведения в течение 3 минут в положении сидя с помощью устройства CardioQVARK;

Кровь для анализа на гликированный гемоглобин передавалась в лабораторию.

Регистрировали ЭКГ 1 ст. отведения в процедурном кабинете поликлиники с помощью портативного электрокардиографа CardioQVARK длительностью 3 минуты с частотой дискретизации 1000 Гц.

Обработка снятых ЭКГ осуществлялась в удаленном вычислительном центре с помощью ЭС СД2 с протоколированием выявления кардиопризнака НУО и записью ЭКГ в базу данных, расчётом КИГ, преобразованием в амплитудно-частотный спектр ЭКГ и КИГ, и расчетом параметров variability сердечного ритма.

Схема участия пациента с НУО в НИР 2019 изложена в таблице 10.

Ожидаемая выборка для исследования согласно протоколу исследования не менее 100 пациентов и 200 ЭКГ.

Таблица 10 – схема участия пациента в НИР 2021

Дни наблюдения	День 1
Регистрация пациента через приложение CardioQVARK, внесение данных о пациенте: дата рождения, пол, вес, рост.	1
Анализ крови на гликированный гемоглобин	1
Снятие ЭКГ 1 ст. отведения.	2
Определение уровня глюкозы в крови с помощью глюкометра	1

Перед валидацией моделей выявления кардиопризнака НУО выборка ЭКГ проходила следующую обработку - из всей выборки ЭКГ рассчитывались параметры, необходимые для моделей, (параметры) с помощью ЭС СД2. Список параметров моделей приведен в п. 2.4.1 и 2.4.2.

Валидация моделей выявления кардиопризнака НУО по методам анализа ВСП и анализа спектра ЭКГ проводилась на всей выборке с рассчитанными из ЭКГ параметрами следующим образом:

- Рассчитывались точность и полнота моделей выявления кардиопризнака НУО на спектре ЭКГ или КИГ с 95% доверительным интервалом для долей.
- Рассчитывались точность и полнота модели интегральной оценки риска подозрения на НУО у пациента по серии спектров первой/первых двух ЭКГ/КИГ с 95% доверительным интервалом для долей.

Стоит отметить, дизайн НИР 2019 и НИР 2021 в силу поставленных целей имеют свои различия между собой. Основные различия изложены в таблице 11.

Таблица 11 –различия в дизайне исследований между НИР 2019 и НИР 2021

	НИР 2019	НИР 2021
Краткая цель исследования	Оценка принципиальной возможности выявлять кардиопризнак НУО у больных СД2 и отсутствие кардиопризнака НУО у пациентов без НУО	Валидация методов выявления кардиопризнака НУО у пациентов с НУО в первичном звене здравоохранения путем проведения скрининга
Группы пациентов	Больные СД2 – пациенты, лежащие в стационаре.	Все пациенты были без диагноза «сахарный диабет».

	<p>Пациенты без НУО в стационаре не лежат.</p> <p>Две возможные группы: СД2/без НУО</p>	<p>Наличие/отсутствие НУО у пациентов неизвестно до начала исследования</p> <p>Три возможные группы: СД2/предиабет/без НУО</p>
Обучение мат. моделей	+	-
Валидация мат. моделей	+	+
Ожидаемая выборка	<p>СД2: 100 пациентов старше 18 лет и не менее 1000 ЭКГ (не менее 10 ЭКГ на пациента);</p> <p>Без НУО: 30 пациентов от 18 до 64 лет и не менее 300 ЭКГ (не менее 10 ЭКГ на пациента).</p>	<p>Всего не менее 100 пациентов от 40 до 79 лет и 200 ЭКГ (по 2 ЭКГ на пациента)</p>

3.2.2 Выборка

На площадке поликлиники № 6 ГБУЗ МО Мытищинской Городской Клинической Больницы (МГКБ) получены аннотированная выборка из 456 ЭКГ, снятых 220 пациентами.

Регистрировали ЭКГ 1 отведения с помощью портативного электрокардиографа CardioQVARK длительностью 3 минуты с частотой дискретизации 1000 Гц

Каждый пациент снял минимум 2 ЭКГ, 11 пациентов попросили переснять ЭКГ еще один раз (5 пациентов – единожды, 6 пациентов - дважды) в связи с обнаружением в ходе записи ЭКГ большого количества артефактов,

из-за чего такие ЭКГ переставали соответствовать требованиям к длительности ЭКГ (см п. 2.2.1) после удаления артефактов. Данные артефакты были вызваны спонтанным отходом датчиков кардиографа от тела пациента и периодическими электропомехами от щитовой, расположенной непосредственно за стенкой процедурного кабинета – местом снятия ЭКГ.

ЭКГ, несоответствующие требованиям к длительности, решено исключить из рассмотрения. Пациентов, у которых не осталось ЭКГ, неисключенных из рассмотрения, решено также исключить.

Наличие СД2/предиабета предполагалось устанавливать по результатам анализа крови на гликированный гемоглобин и по результатам определения уровня глюкозы в крови натощак с помощью глюкометра.

Сводка по полученным экспериментальным данным до и после исключения, не соответствующих требованиям к длительности ЭКГ после предварительной обработки, приведена в таблице 12.

Таблица 12 - Сводка экспериментальных данных, НИР 2021

	№ 6 ГБУЗ МО МГКБ			
	С СД2	С предиабетом	Без НУО	Итого
До исключения ЭКГ, не соответствующих требованиям				
Пациентов	8	41	171	220
Снято ЭКГ	18	83	355	456
ЭКГ на пациента в среднем	2,25	2,02	2,08	2,07
После исключения ЭКГ, не соответствующих требованиям				
Пациентов	5	35	151	191
Снято ЭКГ	9	59	261	329
ЭКГ на пациента в среднем	1,8	1,69	1,73	1,72

Структура полученной выборки приведена в таблице 13. Данные структурированы по половозрастному составу.

Таблица 13 - Структура полученной выборки, НИР 2021

СД2

Пол	18-44 лет		45-64 лет		65+ лет		ЭКГ	Пац.	ЭКГ на пац.
	Пац.	ЭКГ	Пац.	ЭКГ	Пац.	ЭКГ			
Мужчины	0	0	1	1	1	2	3	2	1,5
Женщины	1	2	1	2	1	2	6	3	2
Итого							9	5	1,8

Предиабет

Пол	18-44 лет		45-64 лет		65+ лет		ЭКГ	Пац.	ЭКГ на пац.
	Пац.	ЭКГ	Пац.	ЭКГ	Пац.	ЭКГ			
Мужчины		3		12		5	23	14	1,6
Женщины	0	0		26		10	36	21	1,7
Итого							59	35	1,7

Без НУО (здоров)

Пол	18-44 лет		45-64 лет		65+ лет		ЭКГ	Пац.	ЭКГ на пац.
	Пац.	ЭКГ	Пац.	ЭКГ	Пац.	ЭКГ			
Мужчины	6	9	17	28	4	8	45	27	1,7
Женщины	18	29	81	142	25	45	216	124	1,7
Итого							261	151	1,7
Всего							329	191	1,7

Сравнение структур выборок пациентов и ЭКГ между прошлым и текущим исследованием приведено в таблице 14.

Таблица 14 - Сравнение структур выборок пациентов и ЭКГ (НИР 2019 и НИР 2021)

	НИР 2019			НИР 2021		
	СД2	Предиабет	Здоров	СД2	Предиабет	Здоров
Всего пациентов	128	-	59	5	35	151
Средний индекс массы тела пациента	33	-	25	32	30	29
Средний возраст пациента, лет	63	-	36	56	56	55
Распределение пациентов по полу («М/Ж»)	53/75	-	11/48	3/6	23/36	45/216
Всего ЭКГ	1997	-	741	9	59	261

В выборках исследований наблюдаются следующие различия:

- В выборке НИР 2019 преобладают пациенты с СД2, в выборке НИР 2021 преобладают пациенты без НУО;
- В выборке НИР 2019 было снято больше ЭКГ (в целом всего и на одного пациента), чем в выборке НИР 2021;
- В выборке НИР 2019 больше различия между классами пациентов (СД2/без НУО) по возрасту и индексу массы тела, чем в выборке НИР 2021.

3.2.3 Оценка качества моделей классификации

Результаты применения модели выявления кардиопризнаков НУО по спектру ЭКГ и модели по спектру ЭКГ приведены в таблице 15. Поскольку изначально модели обучались на спектрах кардиологических данных пациентов с СД2 и пациентов без НУО, здесь приведена оценка моделей на выборке данных пациентов с СД2 и пациентов без НУО (т.е. без предиабета). Соотношение количества исходных ЭКГ пациентов с СД2 и пациентов без НУО составляет 3%/97%

Таблица 15 – Полнота и точность моделей выявления кардиопризнаков НУО на отдельной ЭКГ, НИР 2021

	Модель по спектру ЭКГ	Модель по параметрам ВСР и спектру КИГ
Чувствительность, полнота по классу «Есть НУО» (СД2) (95% дов. интервал)	100% (65% - 100%)	78% (44%- 95%)
Специфичность, полнота по классу «Нет НУО», (95% дов. интервал)	57% (51% - 62%)	59% (53%-65%)
Точность по классу «Есть НУО» (СД2), (95% дов. интервал)	7% (4%-14%)	6% (3%-12%)
Точность по классу «Нет НУО», (95% дов. интервал)	100% (97,5%- 100%)	99% (95%-100%)

На выборке данных НИР 2021 наблюдается снижение качества моделей выявления кардиопризнака НУО, и в доверительных интервалах всех метрик

модель по параметрам ВСП и спектру КИГ содержала неинформативные значения, из-за чего далее данная модель не рассматривалась.

Однако у модели выявления кардиопризнаков НУО по спектру ЭКГ доверительные интервалы метрик качества все также не содержат неинформативных значений: «50%» для полноты по каждому классу, «3%» - для точности по классу «Есть НУО» (СД2) и «97%», что все еще позволяет судить о статистически значимом различии между оцениваемой моделью и «бесполезной» моделью вроде броска монетки или другого случайного угадывания. Однако стоит учесть, что данные метрики качества не учитывают наличие на выборке кардиологических данных пациентов с предиабетом.

Рассматривались значения полноты и точности модели интегральной оценки риска подозрения на НУО (СД2) по серии из первой/первых двух ЭКГ пациента (таблица 16), данная модель приведена в п. 2.4.3. Соотношение количества исходных пациентов с СД2 и без НУО на выборке составляет 3%/97%

Таблица 16 – Полнота и точность модели интегральной оценки риска подозрения на НУО, первые 2 ЭКГ пациентов, НИР 2021

Длина серии ЭКГ (n_{min})	Условие подозрения на НУО ($n_{Y=1}$)	Метод на основе анализа спектра ЭКГ			
		Чув-ть, Полнота по классу «Есть НУО» (95% ДИ)	Спец-ть, Полнота по классу «Есть НУО» (95% ДИ)	Точность, класс «Есть НУО» (95% ДИ)	Точность, класс «Есть НУО» (95% ДИ)
1	1	100% (60%-100%)	48% (40%-57%)	6% (2%-14%)	100% (96%-100%)
2	1	100%	25%	9% (4%-20%)	100% (84%-100%)

		<i>(54%– 100%)</i>	<i>(17%– 38%)</i>		
	2	<i>100% (54%– 100%)</i>	<i>57% (47%– 66%)</i>	<i>15% (6%–31%)</i>	<i>100% (84%–100%)</i>

Поскольку доверительные интервалы полноты модели интегральной оценки риска подозрения на НУО содержат неинформативные значения, сделан вывод что по одной или двум ЭКГ устойчиво выявить подозрение (или отсутствие подозрения) на НУО у пациента нельзя, необходимо снимать у пациента больше 2 ЭКГ.

Оцениваемая модель при данных длинах серии ЭКГ не соответствуют ожидаемой полноте и точности. Вероятные причины:

- Принципиальные различия между выборками НИР 2019 и НИР 2021:
 - В выборке НИР 2019 преобладают пациенты с СД2, в выборке НИР 2021 преобладают пациенты без НУО;
 - В выборке НИР 2019 было снято больше ЭКГ (в целом всего и на одного пациента), чем в выборке НИР 2021;
 - В выборке НИР 2019 больше различия между классами пациентов (СД2/без НУО) по возрасту, чем в выборке НИР 2021.
- В НИР 2019 исследовались только две группы пациентов: аннотированные СД2, аннотировано здоровые;
- Пациенты с СД2 в НИР 2019 уже употребляли лекарства от диабета и сопутствующих заболеваний, в то время как пациенты НИР 2021 впервые узнали о своем диагнозе в ходе эксперимента и не принимали никаких лекарств;

- Недостаточно длинный интервал между регистрацией 1-й и 2-й ЭКГ – около 4 минут вместо 2 часов в НИР 2019.

Все эти причины являются неучтенными факторами при расчетах минимального количества ЭКГ на одного человека.

Статистика работы моделей выявления кардиопризнаков НУО по всей выборке спектров ЭКГ пациентов (включая пациентов с предиабетом) приведена в таблице 17.

Таблица 17 – Статистика параметров модели выявления кардиопризнаков НУО по спектру ЭКГ, НИР 2021

	Доля ЭКГ пациентов с СД2 с выявленным кардиопризнаком НУО	Доля ЭКГ пациентов с предиабетом с выявленным кардиопризнаком НУО	Доля ЭКГ пациентов без НУО с выявленным кардиопризнаком НУО
Модель по спектру ЭКГ	100% (9/9)	70% (35/159)	49% (80/161)

Основными выводами НИР 2021 стали следующие:

- СД2/предиабет нельзя выявить по одной или двум ЭКГ;
- Из двух моделей выявления кардиопризнака НУО на выборке пациентов НИР 2021 лучше себя показала модель по спектру ЭКГ в плане качества классификации;
- Необходимо найти формы проведения скрининга НУО, которые не сводятся к однократному или двукратному снятию ЭКГ;
- Необходимо провести дополнительные исследования для определения границ областей «СД2»/«Предиабет»/«Здоров» по доле ЭКГ с выявленным кардиопризнаком НУО. Выборка пациентов должна быть репрезентативной по доле людей с сахарным диабетом 2

типа/предиабетом/здоровых относительно генеральной совокупности, но с большей длиной серии ЭКГ на пациента.

Для достижения целевой полноты по всем классам от 75% при различении классов пациентов «СД2»/«Предиабет»/«Здоров» требуется увеличить необходимое кол-во ЭКГ на одного пациента – длину серии.

3.2.4 Имитационное моделирование скрининга НУО по ЭКГ

Поскольку в данном исследовании максимальная длина серии ЭКГ состояла только из 2 штук, а требовалось оценить, при какой длине серии ЭКГ и разработанных методах выявления кардиопризнака НУО на отдельной ЭКГ теоретически возможно достичь целевой полноты по классам «СД2»/«Предиабет»/«Здоров» модели интегральной оценки риска подозрения на НУО, решено провести имитационное моделирование скрининга НУО по ЭКГ в масштабе Москвы.

Целью моделирования является:

- определение минимальной длины серии ЭКГ на пациента для достижения целевой чувствительности и специфичности;
- оценка доли пациентов, у которых будет по результатам скрининга по ЭКГ выявлен высокий риск СД2/предиабета и которым будет дана рекомендация обратиться к эндокринологу на диагностику;

Для расчета минимальной длины серии проведено имитационное моделирование выборки на основе выборки из НИР 2021 по методу из п. 2.3. Выборка моделировалась на основе:

- размера имитационной выборки пациентов N – 10 млн. пациентов, по m 11 измерений на пациента;
- априорной вероятности наличия НУО (доля пациентов из выборки с СД2/диабетом/здоровых в России) : СД2 ($prev_{СД2}$) – 3%, предиабет ($prev_{Предиабет}$) – 18%, здоровые ($prev_{Здоров}$) – 79%;

- статистика работы моделей выявления кардиопризнаков НУО на выборке спектров ЭКГ из НИР 2021:
 - с СД2 ($res_{СД}$) – 100%;
 - с предиабетом ($res_{Предиабет}$) - 70%;
 - здоровых ($res_{Здоров}$) - 49%.

Результатом имитационного моделирования стали следующие расчетные характеристики чувствительности модели по пациентам с СД2 или предиабетом и специфичности по здоровым пациентам (таблица 18).

Таблица 18 – Полнота по разным группам классов

«Здоров»/«Предиабет»/«СД2» при разной серии измерений каждого объекта (в сокращённом виде)

Измерений	Условия принятия положительного решения по объекту	"Здоров"/"Предиабет или СД2"		"Здоров или предиабет"/"СД2"	
		Полнота "Здоров"	Полнота "Предиабет или СД2"	Полнота "Здоров или предиабет"	Полнота "СД2"
1	1 из 1	51%	77%	47%	100%
2	от 1 из 2	26%	93%	22%	100%
	2 из 2	76%	60%	71%	100%
...					
11	от 1 из 11	0%	100%	0%	100%
	от 2 из 11	1%	100%	1%	100%
	от 3 из 11	4%	100%	3%	100%
	от 4 из 11	13%	100%	10%	100%
	от 5 из 11	30%	98%	24%	100%
	от 6 из 11	53%	94%	44%	100%
	от 7 из 11	75%	84%	64%	100%
	от 8 из 11	90%	66%	80%	100%
	от 9 из 11	97%	46%	91%	100%
	от 10 из 11	100%	31%	97%	100%
	11 из 11	100%	23%	100%	100%

Если указать требования к полноте что модели обнаружения кардиопризнака НУО по следующим группам классов: $R_{min}^{\leq \text{Здоров}}=75\%$;

$Rmin^{>Здоров}=75\%$; $Rmin^{\leq\text{Здоров или предиабет}}=75\%$; $Rmin^{>\text{Здоров или предиабет}}=75\%$, то на таблице 18 видно, что модель обнаружения кардиопризнака НУО на серии спектров ЭКГ будет соответствовать требованиям к полноте при 11 ЭКГ на пациента. Для этого необходимо установить порог в 7 положительных результатов модели из 11 между группами классов «Здоров» и «Предиабет или СД2» и порог в 11 положительных результатов из 11 между группами классов «Здоров или Предиабет» и «СД2».

По результатам имитационного моделирования, модель интегральной оценки риска подозрения на НУО по серии спектров ЭКГ выглядит следующим образом:

$$DiabetesRisk = \begin{cases} \text{"Здоров", если } ratio < \frac{7}{11} \\ \text{"Предиабет", если } \frac{7}{11} \leq ratio < \frac{11}{11} \\ \text{"СД2", если } ratio = \frac{11}{11} \end{cases} \quad (37)$$

где $DiabetesRisk$ – результат работы модели;

n – количество ЭКГ, снятых пациентом – длина серии ЭКГ;

$n_{Y=1}$ - количество ЭКГ с выявленным кардиопризнаком НУО на спектрах ЭКГ.

Полнота модели интегральной оценки риска подозрения на НУО составила на имитационной выборке на классе «Здоров» составила 75%, на классе «Предиабет» - 77%, на классе «СД2» – 100%

Данная модель была использована на базе ЭКГ собранной с помощью ЭС СД2. Всего в базе сохранено 3293 пациента, снявших от 11 ЭКГ и больше, и 353880 ЭКГ всего. На таблице 19 представлена сравнительная таблица с итоговым распределением пациентов выборки из базы ЭКГ ЭС СД2 по классам «Здоров»/«Предиабет»/«СД2», и распределением классов на генеральной совокупности в России, полученной в результате исследования распространенности СД2 и предиабета в России от 20 до 79 лет с применением анализов крови на гликированный гемоглобин. [1].

Таблица 19 - Сравнение распределений классов пациентов
«Здоров»/«Предиабет»/«СД2»

	Выборка пациентов из базы ЭС СД2, согласно работе модели интегральной оценки риска подозрения на НУО	Выборка пациентов из исследования NATION, согласно анализам крови [1]
Здоров ($prev_{\text{здоров}}$)	80%	75,3%
Предиабет ($prev_{\text{предиабет}}$)	15,9%	19,3%
СД2 ($prev_{\text{СД2}}$)	4,1%	5,4%

При проведении скрининга с использованием модели интегральной оценки риска подозрения на НУО по серии спектров ЭКГ из 11 штук внутри телемедицинской системы ЭС СД2 из всей группы риска только у $N_{\text{positive}}=39\%$ пациентов (при ожидаемой полноте $R_{\text{здоров}}=75\%/R_{\text{предиабет}}=77\%/R_{\text{СД2}}=100\%$ по классам «Здоров»/«Предиабет»/«СД2») будет выявлено подозрение на НУО, и только им будет дана рекомендация обратиться к врачу-эндокринологу.

$$N_{\text{positive}} = (100\% - R_{\text{здоров}}) * prev_{\text{здоров}} + R_{\text{предиабет}} * prev_{\text{предиабет}} + R_{\text{СД2}} * prev_{\text{СД2}} = 25\% * 0,753 + 77\% * 0,193 + 100\% * 0,054 = 39\% \quad (38)$$

При использовании ЭС СД2 количество требуемых анализов крови сократится на 61% по сравнению с применением рекомендуемых тестов на всю группу риска.

3.2.5 Выводы исследования

- СД2/предиабет нельзя выявить по одной ЭКГ;

- Из двух моделей выявления кардиопризнака НУО на выборке спектров кардиологических данных из НИР 2021 лучше себя показала модель по спектрам ЭКГ в плане качества классификации «СД2/здоров»;
- Для выявления СД2/предиабета необходимо снять не менее 11 ЭКГ на пациента;
- Необходимо найти формы проведения скрининга НУО, которые не сводятся к однократному или двукратному снятию ЭКГ;
- Необходимо провести дополнительные исследования для определения границ областей «СД2»/«Предиабет»/«Здоров» по доле ЭКГ с выявленным кардиопризнаком НУО.

3.3 Выводы по 3 главе

С целью оценки качества, разработанной моделей классификации (модели выявления кардиопризнака НУО на выборке спектров кардиологических данных и модель интегральной оценки риска подозрения на НУО по серии спектров) проведены два клинических исследования. По результатам проведенных исследований сделаны следующие выводы:

- Предлагаемые модели выявления кардиопризнака НУО на выборке спектров кардиологических данных устойчиво выявляют кардиопризнаки НУО у пациентов с СД2 в условиях стационара;
- В поликлинических условиях при малой длине серии ЭКГ на пациента (1-2 ЭКГ) выявить наличие или отсутствие подозрения СД2 или предиабет нельзя;
- Модель выявления кардиопризнаков НУО на спектре ЭКГ прошла клиническую валидацию на базе НМИЦК им. ак. Е.И. Чазова. На выборке ЭКГ, полученной при клинической валидации, из двух методов точнее себя показал метод выявления кардиопризнаков НУО по ВСР.

- Результаты имитационного моделирования скрининга на примере населения Москвы показали, что разработанная технология проведения скрининга и методы выявления кардиопризнаков НУО по ЭКГ позволят снизить затраты на анализы крови для всей группы риска НУО не менее чем на 61% по сравнению с рекомендованными скрининговыми тестами.
- Необходимо продолжение исследования методов выявления кардиопризнаков НУО на более длинных сериях ЭКГ (от 11 ЭКГ на пациента) для анализа возможности устойчивого выявления пациентов с НУО и для определения границ областей «СД2»/«Предиабет»/«Здоров» по доле ЭКГ с выявленным кардиопризнаком НУО.

Глава 4. Разработка фреймворка для формирования модулей классификации по спектрам кардиологических данных как цифровых продуктов и его применение для построения медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по кардиологическим данным

В данной главе описаны:

- Требования к телемедицинским системам дистанционного мониторинга по спектрам кардиологическим данным, использующим модули классификации на основе машинного обучения как проблемно-ориентированные цифровые продукты;
- Разработка архитектуры фреймворка для формирования модулей бинарной и порядковой классификации по спектрам кардиологических данных и его реализация;
- Реализация телемедицинской системы дистанционного мониторинга по кардиологическим данным - медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности электрокардиограмм сердца первого отведения для выявления по серии спектров ЭКГ признаков нарушения углеводного обмена (ЭС СД2) – при использовании разработанного фреймворка.

4.1 Требования к телемедицинским системам дистанционного мониторинга по кардиологическим данным.

Для формирования требований к архитектуре телемедицинских систем дистанционного мониторинга (далее - Систем) по кардиологическим данным (сигналам), использующих модули классификации по спектрам

кардиологических данных на основе машинного обучения, рассмотрим общую схему данных Систем (Рисунок 10).



Рисунок 10 –Общая схема телемедицинских систем дистанционного мониторинга по кардиологическим данным (Систем)

Данные Системы имеют следующие общие ключевые функции:

- прием кардиологических сигналов (далее - сигналов) с устройства съёма
- прием метаданных о пациенте (например, пол, дата рождения и т.д.) с GUI пациента;
- хранение сигналов и метаданных о пациенте в подсистеме хранения исходных и обработанных данных;
- преобразование сигналов в амплитудно-частотный спектр (Фурье-спектр) с помощью модуля преобразования сигнала в спектр;
- анализ сигналов, спектров сигналов и метаданных о пациенте с помощью модели машинного обучения для классификации состояния пациента;
- отправка исходных сигналов/спектров сигналов/метаданных о пациенте/результатов анализа на GUI конечного пользователя (врача и пациента).

Изменяемыми компонентами Системы в зависимости от группы интересующих состояний (заболевание или функциональное состояние)

пациента, является модуль классификации состояния пациента по спектрам кардиологических данных машинного обучения и GUI конечного пользователя.

Из выделенных общих ключевых функций выделим следующие функциональные требования к архитектуре:

- Должен быть обеспечен интерфейс графического взаимодействия (GUI) Системы с конечным пользователем – пациентом и врачом.
- Должен быть обеспечен децентрализованный сбор сигналов с устройств съема и метаданных о пациенте с GUI пациента на централизованное хранение;
- Должна быть обеспечена централизованная обработка сигналов, включающая преобразование в Фурье-спектр и хранение результатов обработки.
- Должна быть обеспечена возможность использования моделей классификации на обработанных сигналах для их анализа и хранения результатов анализа.

При реализации Системы необходимо учесть следующие требования к отдельным компонентам:

- Устройство съема сигналов должно обеспечивать съём и передачу сигналов, удовлетворяющие требованиям, при которых возможно получить Фурье-спектр с необходимыми для моделей машинного обучения шириной полосы частот и шагом спектра в частотной области. Пример требований к кардиологическим данным для моделей машинного обучения (бинарной классификации), построенных по методу, описанном в п. 2.1, приведен в п. 2.2.1;
- GUI пациента должен обеспечивать прием и передачу метаданных о пациенте в формате, пригодном для приема на хранение подсистемой хранения исходных и обработанных данных;

- Модуль классификации состояний пациента должен удовлетворять требованиям к качеству, задаваемыми клиническими стандартами обнаружения изучаемого заболевания или функционального состояния. При отсутствии стандартов требования к качеству должны устанавливаться экспертом;
- GUI пациента и GUI врача должны предоставлять информацию пользователю в соответствии его с ролевой моделью.

Поскольку конечными пользователями Системы являются пациент (Пациент) и лечащий его врач (Врач), рассмотрим их ролевые модели.[3]

Особенности Пациента следующие:

- Пациент не владеет медицинской терминологией;
- Пациент может неправильно обращаться с устройством съема сигналов. Если пациент будет пользоваться данным устройством самостоятельно, то из-за различных факторов (сетевая наводка, лишние движения пациента и т. д.) сигналы в результате могут оказаться непригодными для дальнейшего анализа;
- Пациент может регистрировать сигналы не обязательно в медучреждении, поэтому рядом может не быть квалифицированного медработника, который проконтролирует, правильно ли пациент снимает сигналы.

Из этого следует, что при представлении результатов обработки сигналов Пациенту нужно отобразить следующую информацию:

- «В порядке ли все с ним?» - должно быть сказано, есть ли у него признаки наличия исследуемого заболевания или плохого функционального состояния (в зависимости от используемой модели машинного обучения).
- «Ему надо к врачу?» - должна быть представлена рекомендация пойти к врачу определенной специализации при выявлении признаков исследуемого состояния.

- «Качественно ли он снял ЭКГ?» - если снятые сигналы получились недостаточно качественными для соответствия требованиям к сигналам, должна быть представлена рекомендация «переснять» сигналы.

Исходя из ролевой модели, предъявляются следующие требования к тому, что должен отображать GUI Пациента:

- значения показателей исследуемого состояния, которые можно легко объяснить без врача.
- информацию о наличии/отсутствии возможного исследуемого состояния по результатам обработки данных. При наличии исследуемого состояния – рекомендацию, к врачу какой специализации стоит обратиться за консультацией;
- рекомендацию «переснять» снятые сигналы при их плохом качестве.

Информация должна предоставляться в упрощённом виде, например, в виде инфографики.

Особенности Врача следующие:

- Врач владеет медицинской терминологией, относящейся к своей специализации;
- Врач обладает исключительным правом принимать решение по Пациенту, однако может принять во внимание результаты работы модели машинного обучения;
- Врачу важно знать, когда пациент снимал сигналы.

Из этого следует, что при представлении результатов обработки сигналов Врачу нужно отобразить следующую информацию:

- «Что за Пациент и когда снял свои данные?» - должна быть представлена информация о Пациенте (метаданные) и о событиях съема новых сигналов.
- «Качественно ли Пациент снял сигнал?» - если снятый сигнал Пациентом получился недостаточно качественным для соответствия

требованиям, должна быть представлена информация об этом, чтобы Врач мог связаться с Пациентом по этому поводу.

- «Как выглядят исходные сигналы и их результаты обработки?» - должна быть представлена информация об исходных снятых пациентом сигналах и их результатах обработки в терминах специализации Врача, чтобы он имел возможность самостоятельно принимать решение по пациенту.
- «В порядке ли все с Пациентом и почему?» - должна быть представлена информация, есть ли у Пациента признаки наличия исследуемого состояния с объяснением, какие показатели, рассчитанные из сигналов, повлияли на решение.

Исходя из ролевой модели, предъявляются следующие требования к тому, что должен отображать GUI Врача:

- Информация о событии снятия новых сигналов Пациентом, и об их качестве;
- Снятые сигналы в исходном виде и их результаты обработки, включая расчетные показатели, относящиеся к исследуемому состоянию.
- Информацию о наличии/отсутствии возможного исследуемого состояния по результатам обработки данных с объяснением результата;

Также GUI должен обеспечить возможность сохранения информации о решении Врача по Пациенту на основании сигналов или метаданных о Пациенте.

4.2 Архитектура фреймворка для формирования модулей по спектрам кардиологических данных и его реализация

На рисунке 11 представлена схема архитектуры фреймворка для формирования модулей бинарной и порядковой классификации по спектрам

кардиологических данных как проблем-ориентированных цифровых продуктов для Систем.



Рисунок 11 - Схема архитектуры фреймворка для формирования модуля классификации по спектрам кардиологических данных. Серым цветом выделены модули, логика которых описана во 2й главе. Синим выделен целевой результат работы фреймворка – готовый модуль классификации.

Внешне для пользователя (разработчика моделей) фреймворк представляет собой набор следующих доступных команд:

- Запуск формирования модели бинарной классификации. Команда должна включать в себя набор значений параметров метода, описанного в п. 2.1;
- Запуск формирования модели порядковой классификации. Команда должна включать в себя набор значений параметров метода, описанного в п. 2.3;
- Развертывание сформированной модели в выделенный модуль классификации. Команда должна содержать идентификатор модели и путь к выделенному модулю классификации.

Внутри фреймворк представляет собой набор следующих компонент:

- 1) Модуль приема кардиологических сигналов и метаданных о пациенте через подготовленные интерфейсы передачи данных;

- 2) Модуль формирования выборки спектров кардиологических данных, работающий согласно соответствующей методике, описанной в п. 2.2. В частности, процесс обработки кардиологических сигналов с преобразованием в спектры данных описан в п. 2.2.2;
- 3) База с выборкой спектров кардиологических сигналов, пополняемая модулем 2) при поступлении новых данных с модуля 1);
- 4) Модуль формирования модели бинарной классификации по спектру кардиологических сигналов, начинающий работать по команде пользователя. Данный модуль запрашивает выборку из базы 3), затем выполняет формирование модели бинарной классификации по методу, описанному в п. 2.1. Результат (сериализованная модель и метрики качества модели) направляется в базу 6);
- 5) Модуль формирования модели порядковой классификации по спектру кардиологических сигналов, начинающий работать по команде пользователя. Данный модуль запрашивает выборку из базы 3), и модель бинарной классификации, на основе которой будет строиться новая модель, из базы 6), затем выполняет формирование модели порядковой классификации по методу, описанному в п. 2.3. Результат (сериализованная модель и метрики качества модели) направляется в базу 6);
- 6) База моделей классификации с их оценками качества;
- 7) Модуль развертывания модели классификации, начинающий работать по команде. Данный модуль запрашивает модель из базы 6) в сериализованном виде, а затем десериализует и развёртывает ее в выделенный модуль классификации.

Под «сериализованным» понимается вид модели, переведенный в битовую последовательность для обеспечения возможности сохранить ее в виде файла. Под «десериализацией» модели понимается обратный перевод

модели из битовой последовательности в формат, пригодный для использования – например, в виде вызываемой функции предсказания.

На рисунке 12 представлена схема реализованного фреймворка для формирования модулей классификации по спектрам кардиологических данных как цифровых продуктов для Систем на основе программной платформы Peraspera Cardio с помощью инструмента Node-RED на языке JavaScript. В Node-RED функциональные блоки обозначаются своими терминами. [65] Например, узел представляет собой базовый функциональный блок, который запускает свою работу при получении сообщения от другого, связанного с ним узла, либо при возникновении внешнего события (например, входящего HTTP-запроса или таймера). Узел обрабатывает сообщение или событие, а затем отправляет сообщение следующим другим связанным с ним узлам. Подпоток представляет собой набор узлов, которые логически объединены в один узел.

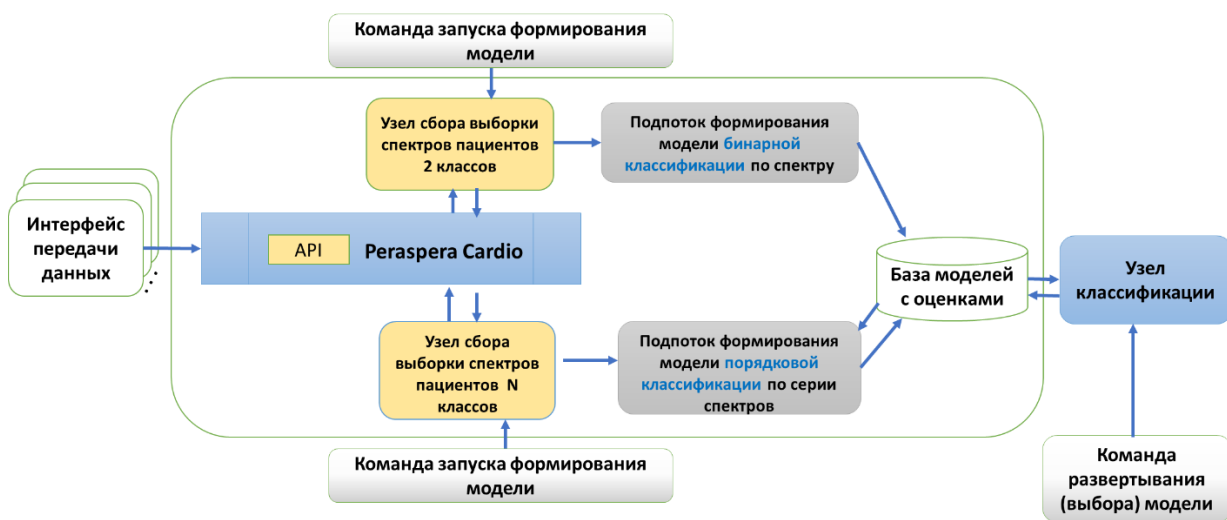


Рисунок 12 - Схема реализованного фреймворка для формирования моделей классификации по спектрам кардиологических данных.

Структура реализации фреймворка состоит из следующих компонент:

- 1) Программная платформа Peraspera Cardio - данный компонент объединяет в себе модуль приема кардиологических сигналов и метаданных о пациенте, модуль формирования выборки спектров

кардиологических данных и базу с выборкой спектров кардиологических сигналов

2) Модуль формирования модели бинарной классификации по спектру кардиологических сигналов реализован через Node-RED как две следующие сущности:

- a. Узел сбора выборки спектров сигналов пациентов двух классов. По команде запуска формирования модели бинарной классификации узел запрашивает выборку спектров сигналов из Peraspera Cardio по API в соответствии с параметрами из полученной команды, затем отправляет выборку далее по схеме в виде сообщения;
- b. Подпоток формирования модели бинарной классификации по спектру. По приему сообщения от узла 2a) подпоток начинает формировать модель классификации согласно методу, описанному в п. 2.1. Результат (сериализованная модель и метрики качество модели) направляется в базу 4);

3) Модуль формирования модели порядковой классификации по серии спектров кардиологических сигналов реализован через Node-RED как две следующие сущности:

- a. Узел сбора выборки спектров сигналов пациентов N классов. По команде запуска формирования модели порядковой классификации узел запрашивает выборку спектров сигналов из Peraspera Cardio по API в соответствии с параметрами из полученной команды, затем запрашивает из базы 4) модель бинарной классификации в соответствии с параметрами из полученной команды, затем отправляет выборку далее по схеме в виде сообщения.
- b. Подпоток формирования модели бинарной классификации по спектру. По приему сообщения от узла 3a) подпоток начинает

формировать модель классификации согласно методу, описанному в п. 2.3. Результат (сериализованная модель и метрики качества модели) направляется в базу 4).

- 4) База моделей классификации с их оценками качества – хранится в файловой системе на вычислительной машине, на которой поставлен фреймворк. Содержит в себе файлы модели классификации в сериализованном виде и файлы с информацией о качестве моделей.
- 5) Узел (модуль) классификации представляет собой узел, который по команде развертывания запрашивает из базы необходимую модель классификации, а затем становится пригодным для внедрения в Систему в качестве вычислительного узла на Node-RED.

Процесс формирования моделей классификации (модели выявления кардиопризнаков НУО и модели выявления подозрения на НУО по серии ЭКГ) выполняется при взаимодействии разработчика Системы следующим образом:

- 1) Разработчик указывает дает команду фреймворку на запуск формирования модели бинарной классификации, по спектрам каких кардиологических данных (ЭКГ или кардиоинтервалограмма) пациентов какого из 2 классов требуется сформировать модель (с указанием параметров классификации: вида модели и доли выборки, отводимой на обучение).
- 2) Фреймворк внутри себя запрашивает по API у Perapsara Cardio выборку спектров кардиологических данных, по которым пользователь запросил сформировать модель бинарной классификации.
- 3) Получив спектры кардиологических данных, фреймворк приступает к формированию и оценке модели.
- 4) По итогам формирования и оценки модели бинарной классификации фреймворк сохраняет модель в сериализованном виде вместе с оценкой качества модели на тестовой выборке.

- 5) Разработчик дает команду фреймворку на формирования модели порядковой классификации с указанием параметров:
 - a. распределения классов на генеральной совокупности;
 - b. размера имитируемой выборки;
 - c. названия модели бинарной классификации и выборки спектров кардиологических данных, на основе которых будет проводить подсчет статистики работы на указанных классах (долю положительных результатов модели по каждому классу) ;
 - d. требования к полноте модели бинарной классификации при повторных измерениях.
- 6) Фреймворк подсчитывает статистику работы модели бинарной классификации, затем на основе полученной команды приступает к формированию и оценке модели порядковой классификации.
- 7) По итогам формирования и оценки модели по серии спектров кардиологических данных фреймворк сохраняет модель в сериализованном виде вместе с оценкой качества модели.
- 8) Далее средствами инструмента Node-RED разработчик развертывает сохраненную модель в выделенный модуль (узел) классификации внутри Системы.

4.3 Пример реализации телемедицинской системы дистанционного мониторинга по кардиологическим данным с применением фреймворка – ЭС СД2

В данном разделе описывается пример реализации телемедицинской системы дистанционного мониторинга по кардиологическим данным (Системы), где модули классификации реализованы с помощью фреймворка, описанного в п. 4.2, - медицинской систему самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности

электрокардиограмм сердца первого отведения, позволяющую проводить выявление по серии спектров ЭКГ признаков нарушения углеводного обмена с классификацией пациентов на здоровых, с высокой вероятностью страдающих предиабетом и страдающих диабетом (ЭС СД2). Назначением ЭС СД2 является обеспечение проведения удаленного от медицинских учреждений скрининга пациентов по серии ЭКГ.

Данная система соответствует схеме Системы и требованиям, описанным в п. 4.1, за исключением того, что пользователями ЭС СД2 являются только пациенты скрининга (далее – Пациент), лечащий врач как роль отсутствует.

Общая схема ЭС СД2 показана на рисунке 13.

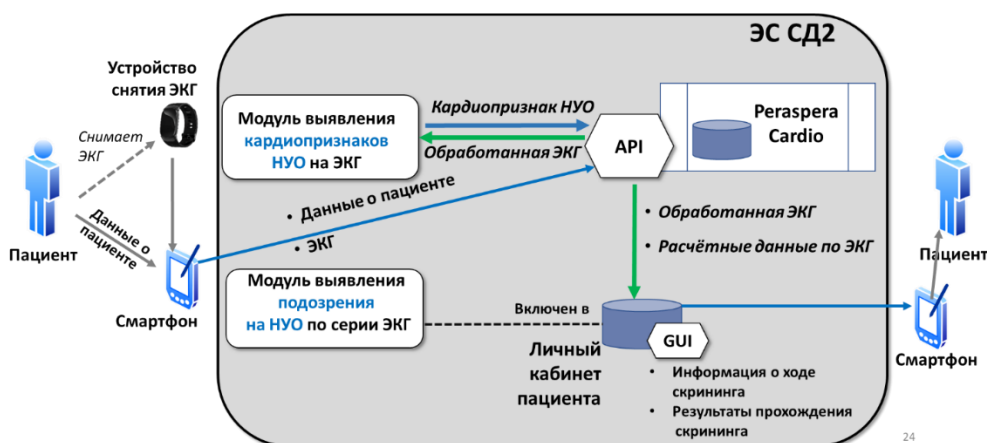


Рисунок 13 – Схема ЭС СД2

ЭС СД2 состоит из следующих компонентов:

- Устройство съема ЭКГ;
- Программная платформа «Peraspera Cardio», №10705 в реестре российского ПО - в качестве подсистемы хранения исходных и обработанных данных и частично в качестве подсистемы обработки и анализа данных;
- Личный кабинет (ЛК) Пациента на смартфоне (в качестве GUI пациента)

- Модуль выявления кардиопризнаков нарушений углеводного обмена (НУО) на ЭКГ – частично в качестве подсистемы обработки и анализа данных;
- Модуль выявления подозрения на НУО по серии ЭКГ - работает внутри ЛК пациента.

Назначением устройства съема ЭКГ в рамках ЭС СД2 является:

- Съем ЭКГ первого стандартного отведения у Пациента;
- Отправка ЭКГ в ЛК Пациента на дальнейшую передачу в Peraspera Cardio на хранение и обработку.

Устройство съема ЭКГ выполняет следующие функции:

- Обеспечение съема ЭКГ у пациента и передачи в ЛК пациента на хранение и обработку.

Взаимодействие портативного кардиографа с ЛК пациента осуществляется через Bluetooth.

Примерами портативных кардиографов, которые можно задействовать внутри ЭС СД2 являются устройства CardioQVARK и «умные» часы с функцией съема и отправки ЭКГ через Bluetooth.

Назначением платформы Peraspera Cardio в рамках ЭС СД2 является:

- Сбор и хранение ЭКГ и метаданных о пациентах;
- Предварительная обработка ЭКГ;
- Выполнение аналитических расчетов по ЭКГ;
- Предоставление результатов расчетов через API в смежные компоненты.

Накопление собранных ЭКГ, а также результатов их предварительной обработки и аналитических расчётов, осуществляется в базе данных СУБД PostgreSQL.

Программа "Peraspera Cardio" выполняет следующие функции предварительной обработки ЭКГ:

- Удаление дрейфа изолинии на сигнале ЭКГ;

- Сглаживание ЭКГ с использованием частотного сглаживающего фильтра;
- Удаление амплитудных выбросов.

Программа "Peraspera Cardio" выполняет следующие аналитические расчеты по ЭКГ:

- Расчет Фурье-спектра ЭКГ
- Обнаружение R-пикув на ЭКГ;
- Построение КИГ;
- Удаление аномальных RR-интервалов;
- Удаление линейного тренда RR-интервалов;
- Расчет Фурье-спектра КИГ и параметров ВСР.

Предоставление результатов аналитических расчетов ЭКГ в смежные системы (ЛК пациента и модуль выявления кардиопризнаков НУО по ЭКГ), а также обмен данными с смежными системами, осуществляется с помощью API.

API представляет собой набор ссылок URL с общим корнем, принимающих запросы по протоколу HTTPS. Для отправки данных используются POST-запросы, для выгрузки данных используются GET-запросы. Формат обмена данными - JSON.

Назначением ЛК пациента в рамках ЭС СД2 является:

- Обеспечение возможности отправки ЭКГ пациента на обработку;
- Предоставление пациенту доступа к информации о ходе и результате скрининга НУО по ЭКГ.

ЛК пациента выполняет следующие функции:

- Регистрация пациента в системе и сохранение данных о пациенте – вес, рост, пол, дата рождения через смартфон;
- Прием ЭКГ из портативных электрокардиографов через смартфон и передача их на обработку вместе с данными о пациенте в Peraspera Cardio;

- Прием результатов расчетов по ЭКГ (наличие/отсутствие кардиопризнака НУО) из Peraspera Cardio;
- Хранение и предоставление пациенту информации о ходе скрининга НУО, о результатах прохождения скрининга;
- Предоставление пациенту рекомендаций по результатам скрининга на основе модели интегральной оценки риска наличия НУО, описанной в п. 2.4.3.

Реализация модели интегральной оценки риска наличия НУО располагается внутри модуля выявления подозрения на НУО по серии ЭКГ.

Взаимодействие с портативным электрокардиографом (прием ЭКГ) осуществляется через Bluetooth. Взаимодействие с платформой Peraspera Cardio (передача ЭКГ с данными о пациенте, прием результатов обработки ЭКГ) осуществляется через Интернет по API Peraspera Cardio.

Доступ к ЛК пациента реализован через приложение на смартфоне на Android. Интерфейс регистрации пациента в личном кабинете отображен на рисунке 14.

The screenshot shows a registration form on a dark background. At the top, there is a back arrow and the title 'Регистрация'. Below the title are two input fields: 'E-Mail' and 'Пароль'. Under the password field, there is a red note: 'Пароль должен содержать не менее 8 символов'. Below this is a section titled 'Рекомендуем заполнить: (необязательные поля)'. This section contains a 'Пол' field with three radio buttons: 'Не указан', 'Мужской', and 'Женский' (which is selected). Below the gender field are three more input fields: 'Дата рождения' (with a calendar icon and a note 'В формате ДД.ММ.ГГГГ'), 'Рост, см', and 'Вес, кг'. At the bottom of the form is a large button with a person icon and the text 'Зарегистрироваться'.

Рисунок 14 - Интерфейс регистрации пациента в личном кабинете (ЛК)

Примеры отображения результатов прохождения скрининга в личном кабинете указаны на рисунке 15. Слева направо на рисунке представлены варианты экраны с состоянием пациента и рекомендациями.

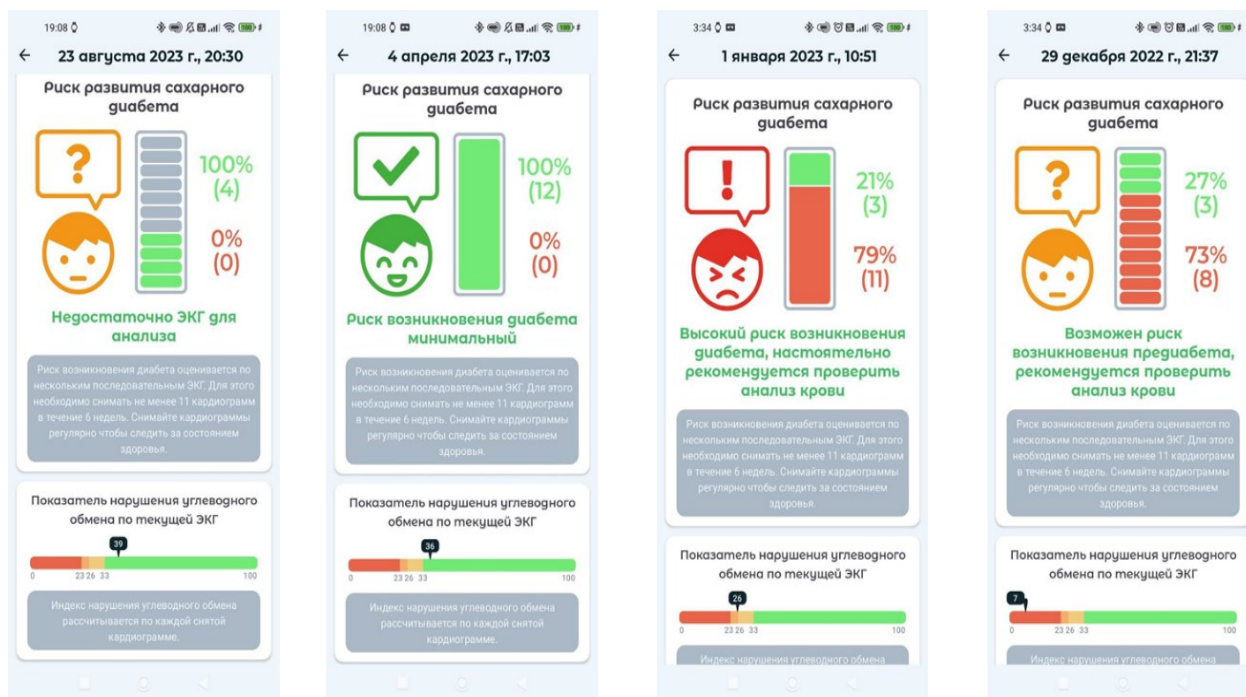


Рисунок 15 - Вид личного кабинета Пациента на смартфоне (результаты прохождения скрининга)

Назначением модуля выявления кардиопризнаков НУО на ЭКГ в рамках ЭС СД2 является:

- Оценка наличия кардиопризнака НУО на ЭКГ по модели бинарной классификации, полученной в результате клинического исследования, и описанной в п. 3.1.3.

Модуль выявления кардиопризнаков НУО на ЭКГ выполняет следующие функции:

- Прием ЭКГ и данных о пациенте от Peraspera Cardio на обработку;
- Оценка наличия кардиопризнака НУО по ЭКГ по методу на основе анализа спектра ЭКГ (формулы 24, 25);
- Передача результатов оценки наличия кардиопризнака НУО на ЭКГ в Peraspera Cardio.

Взаимодействие с платформой Peraspera Cardio (передача ЭКГ с данными о пациенте, прием результатов обработки ЭКГ) осуществляется через Интернет по API Peraspera Cardio.

Физически данный модуль расположен внутри контура ЗАО «ЕС-лизинг» на расчетных выделенных виртуальных машинах и реализован в среде Node-RED (Node.js) на языке Javascript с помощью фреймворка для формирования модулей классификации по спектрам кардиологических данных.

Взаимодействие Пациента с ЭС СД2 в режиме использования проходит по следующему процессу скрининга:

- 1) Пациент регистрируется с помощью ЛК пациента на смартфоне и заносит следующие данные о себе: вес, рост, дата рождения, пол. ЛК пациента передает данные о пациенте в Peraspera Cardio;
- 2) Пациент подключает устройство съема ЭКГ к смартфону с ЛК пациента через Bluetooth;
- 3) Пациент снимает 3-минутную ЭКГ первого стандартного отведения с помощью устройства съёма ЭКГ и отправляет ее в ЛК пациента. ЛК пациента отправляет ЭКГ с данными о пациенте в Peraspera Cardio;
- 4) После приема Peraspera Cardio осуществляет предварительную обработку ЭКГ и отправляет результаты предобработки и аналитических расчётов (спектры кардиологических данных и данных о пациенте) в модуль выявления кардиопризнаков НУО по ЭКГ;
- 5) Модуль выявления кардиопризнаков НУО по ЭКГ проводит оценку наличия кардиопризнака НУО, затем отправляет результаты оценки в Peraspera Cardio;
- 6) ЛК пациента периодически запрашивает у Peraspera Cardio результаты обработки ЭКГ. Как только обработка заканчивается, Peraspera Cardio передает в ЛК пациента результаты обработки ЭКГ (оценку наличия/отсутствия кардиопризнака НУО);

- 7) Пациент через ЛК получает информацию о ходе скрининга (сколько осталось снять еще ЭКГ), затем повторяет шаги 3-6) каждый день примерно в одно и то же время пока не будет снято и обработано не менее 11 ЭКГ;
- 8) Пациент через ЛК получает информацию о результатах прохождения скрининга, рассчитанного с помощью модуля выявления подозрения на НУО по серии ЭКГ: высокий риск СД2, риск предиабета, низкий риск наличия НУО. При высоком риске СД2 или риске предиабета пациент получит через ЛК рекомендацию проверить анализ крови.

Пользователями ЭС СД2 являются только пациенты, врачи-эндокринологи задействованы на этапе диагностики заболевания, согласно способу скрининга, описанному в [5].

4.4 Выводы по 4 главе

В данной главе приведено подробное описание требований к телемедицинским системам дистанционного мониторинга по спектрам кардиологическим данным.

С учетом требований сформирована разработана архитектура фреймворка для формирования модулей бинарной и порядковой классификации по спектрам кардиологических данных и приведена его реализация;

В качестве примера реализации телемедицинской системы дистанционного мониторинга по спектрам кардиологическим данным с использованием фреймворка для формирования модулей классификации состояния пациента приведена автоматизированная система проведения неинвазивного удалённого скрининга НУО по серии ЭКГ (ЭС СД2). ЭС СД2 прошла испытания в ходе двух клинических исследований, описанных в п. 3.1 и 3.2.

Заключение

Основные результаты работы заключаются в следующем:

1. Разработан метод построения моделей машинного обучения с учителем для бинарной классификации объектов по амплитудно-частотным спектрам (Фурье-спектрам) кардиологических данных.
Метод позволяет построить интеллектуальную модель машинного обучения (классификации) в качестве мат. обеспечения цифровых продуктов, основанных на анализе спектров кардиологических данных.
Метод применен при формировании моделей выявления наличия кардиопризнаков нарушений углеводного обмена (НУО) на ЭКГ по:
 - a. степени хаотизации областей окрестностей гармоник Фурье-спектра ЭКГ путем анализа явления автовозврата Ферми-Паста-Улама;
 - b. спектральным и статистическим параметрам variability сердечного ритма.
2. Разработан метод построения моделей машинного обучения с учителем для порядковой классификации объектов с повторными измерениями на основе статистики работы бинарного классификатора. Данный метод применим в условиях, когда исходной выборки размеченных данных достаточно только для построения бинарного классификатора, способного распознавать «крайние» по порядку классы; есть информация о распределении всех классов и о частоте возникновения положительного результата у существующей модели классификации по каждому классу. Данный метод основан на генерации недостающей выборки с помощью имитационного моделирования методом Монте-Карло.
3. Разработана методика формирования выборки данных для создания моделей классификации по спектрам кардиологических данных.
По данной методике изложены требования к исходному набору данных (ЭКГ и метаданные о пациенте) и рекомендуемые алгоритмы предварительной обработки сигналов ЭКГ и преобразованию в данные,

используемые при анализе кардиологических данных: Фурье-спектр ЭКГ; кардиоинтервалограмма (КИГ) и Фурье-спектр КИГ; параметры variability сердечного ритма.

4. Предложена архитектура фреймворка для разработки модулей классификации в качестве проблемно-ориентированных цифровых продуктов по разработанным методам построения моделей машинного обучения с учителем для бинарной классификации и порядковой классификации по спектрам кардиологических данных.

Данный фреймворк может иметь применение при разработке телемедицинских систем дистанционного мониторинга по кардиологическим данным в части автоматизации разработки модулей классификации по спектрам кардиологическим данным.

Приведена реализация данного фреймворка на основе программной платформы Pegasus Cardio с использованием инструмента Node-RED.

5. При помощи реализованного фреймворка для разработки моделей классификации разработаны модели выявления кардиопризнаков нарушений углеводного обмена (НУО) по Фурье-спектрам кардиологическим данным и модель выявления подозрения на НУО по серии Фурье-спектров кардиологических данных. Данные модели защищены патентами: RU2728869C1 (РФ)/JP2021037286A (Япония) и RU2751817C1 (РФ) - и применены при разработке телемедицинской системы - медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности электрокардиограмм сердца первого отведения, позволяющую проводить выявление по серии спектров ЭКГ признаков нарушения углеводного обмена (ЭС СД2).

Данная система валидирована в ходе двух клинических исследований, посвященных определению кардиопризнаков НУО на ЭКГ у пациентов.

Список литературы

1. «Дедов И. И. и др. Распространенность сахарного диабета 2 типа у взрослого населения России (исследование NATION) // Сахарный диабет. 2016. №2.
2. Березин А.А., Новиков Р.С., Новопащин М.А., Позин Б.А., Шмид А.В. Применение метода неинвазивного оценивания нарушений углеводного обмена при скрининге населения. Труды Института системного программирования РАН, том 32, вып. 5, 2020, стр. 121–130. DOI: 10.15514/ISPRAS-2020-32(5)-9
3. Бусенков А.А., Холодов Э.А., Новиков Р.С., Позин Б.А. Подход к отображению данных для пользователей телемедицинских систем раннего обнаружения заболеваний по ЭКГ. Труды Института системного программирования РАН. 2022;34(2):201-208. [https://doi.org/10.15514/ISPRAS-2022-34\(2\)-16](https://doi.org/10.15514/ISPRAS-2022-34(2)-16)
4. Новиков Р.С., Новопащин М.А., Позин Б.А. Применение имитационного моделирования недостающей выборки для создания модели классификации по трем и более классам на примере задачи обнаружения степени нарушения углеводного обмена//Программная инженерия. – 2025. – Т. 16
5. Компьютеризированный способ неинвазивного выявления нарушений углеводного обмена по электрокардиограмме: Пат. 2728869 Рос. Федерация. МПК А61В 5/0402, G06F 17/14, G16H 50/20, G16H 50/70 / А.В. Шмид, А.А. Березин, М.А. Новопащин, Р.С. Новиков, Б.А. Позин, А.М. Мкртумян, Т.Н. Маркова № 2019127394; заявл. 30.08.2019; опубл. 31.07.2020, Бюл. № 22
6. Компьютеризированный способ неинвазивного выявления нарушений углеводного обмена по вариабельности сердечного ритма и носимое автономное устройство для его реализации: Пат. 2751817 Рос.

- Федерация. МПК А61В 5/024, А61В 5/0295, А61В 5/346, А61В 5/1455, G06F 17/14, G16Н 50/20, G16Н 50/70 / А.В. Шмид, М.А. Новопашин, А.А. Березин, Р.С. Новиков, А.М. Мкртумян, Б.А. Позин, № 2020128752, заявл. 31.08.2020; опубл. 19.07.2021 Бюл. № 20
7. Гельман В. Я., Дохов М. А. Проблемы развития домашнего мониторинга состояния здоровья. Медицина 2020; 8(2): 50-60.
 8. Булдаков Н. С. Особенности телемедицинских систем дистанционного мониторинга состояния человека //Современные тенденции развития науки и технологий. – 2016. – №. 9-1. – С. 7-10.
 9. Ляпина И. Н., Зверева Т. Н., Помешкина С. А. Современные способы дистанционного наблюдения и реабилитации пациентов с сердечно-сосудистыми заболеваниями //Комплексные проблемы сердечно-сосудистых заболеваний. – 2022. – Т. 11. – №. 1. – С. 112-123.
 10. Вьюгин В.В. Математические основы теории машинного обучения и прогнозирования. М.: 2013. - 387 с.
 11. Schünke L. C. et al. A rapid review of machine learning approaches for telemedicine in the scope of COVID-19 //Artificial Intelligence in Medicine. – 2022. – Т. 129. – С. 102312.
 12. Лебедев Г.С., Шепетовская Н.Л., Решетников В.А. Телемедицина и механизмы ее интеграции. Национальное здравоохранение. 2021;2(2):21-27. <https://doi.org/10.47093/2713-069X.2021.2.2.21-27>
 13. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМКПресс, 2018. – 652 с.: цв. ил
 14. Къяра Галлезе-нобиле Правовые аспекты использования искусственного интеллекта в телемедицине // Journal of Digital Technologies and Law. 2023. №2
 15. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.

16. Руководство по электрокардиографии / В.Н. Орлов. — 9-е изд., испр. — Москва : ООО Медицинское информационное агентство», 2017. — 560 с. : ил
17. Aziz, S., Ahmed, S. & Alouini, MS. ECG-based machine-learning algorithms for heartbeat classification. *Sci Rep* 11, 18738 (2021). <https://doi.org/10.1038/s41598-021-97118-5>
18. Ляшенко Н.И., Усенко Ю.Д., Хвастухин М.Ю., Чайковский И.А. Применение спектрального анализа ЭКГ к проблеме диагностики гипертрофии правого предсердия при ХНЗЛ. *Пульмонология*. 1994;(3):65-69.
19. Novopashin M. A., Shmid A. V., Berezin A. A. Fermi-Pasta-Ulam auto recurrence in the description of the electrical activity of the heart //Medical hypotheses. – 2017. – Т. 101. – С. 12-16.
20. Захаров С. М., Знайко Г. Г. Спектральный анализ электрокардиосигналов //Вопросы радиоэлектроники. – 2019. – №. 3. – С. 110-115.
21. Баевский Р. М. и др. Анализ variability сердечного ритма при использовании различных электрокардиографических систем (методические рекомендации) //Вестник аритмологии. – 2001. – Т. 24. – №. 6. – С. 86.
22. Парин В. В., Баевский Р. М. (ред.). Математические методы анализа сердечного ритма. – Наука, 1968
23. Champseix R, Ribiere L, Le Couedic C 2021 A Python Package for Heart Rate Variability Analysis and SignalPreprocessing. *Journal of OpenResearch Software*, 9: 28.DOI: <https://doi.org/10.5334/jors.305>
24. В. Л. Сивохов, Е. Л. Сивохова, О. А. Иванова, Ю. К. Кириллов. Использование инновационных технологий в медико-биологическом обеспечении физической культуры и спорта. Монография.– Иркутск, ООО «Призма, 2016. – 210 с.

25. Методы и алгоритмы принятия решений на основе механизмов визуализации и нечеткой логики: монография / И. А. Кочеткова, В. Г. Рубанов. – Белгород: Изд-во БГТУ, 2015. – 112 с
26. Пат. 2138982 Российская Федерация, МПК А61В5/02. Устройство для оценки функционального состояния сердечно-сосудистой системы «Кардиометр» / К.М. Матус; заявитель и патентообладатель ЗАО «МИКАРД-ЛАНА» (RU). – № 98108969/14, заявл. 07.05.1998; опубл. 10.10.1999.
27. Свидетельство о государственной регистрации программы для ЭВМ № 2017613677 Российская Федерация. Кардиометр-МТ. Сервис автоматической интерпретации 3.0 : № 2017611241 : заявл. 15.02.2017 : опубл. 27.03.2017 / М. Н. Антипов, А. М. Конобасов, А. А. Жердер [и др.]; заявитель Акционерное общество «МИКАРД-ЛАНА».
28. Лежнина И. А. и др. Портативный беспроводной кардиоанализатор индивидуального применения // Медицинская техника. – 2019. – №. 6. – С. 14.
29. Lezhnina I.A., Boyakhchyan A.A., Overchuk K.V., Uvarov A.A. Clinical trials of a personal electrocardiograph / JOP Conf. Ser.: Mater. Sci. Eng. 2017. Vol. 881. 1 01204. PP. 1-6.
30. Москаленко В. А. и др. Программный комплекс "Киберсердце-Диагностика" для автоматического анализа электрокардиограмм с применением методов машинного обучения // Современные технологии в медицине. 2019. №2.
31. Медицинская система удаленного мониторинга, анализа и прогнозирования состояния пациента по последовательности электрокардиограмм сердца первого отведения и компьютеризированный способ мониторинга, анализа и прогнозирования состояния пациента Пат. 2752707 Рос. Федерация. МПК А61В 5/349, G06F 16/906, G06F 17/17, G06F 17/18, G16H 50/20,

- G16H 50/70 / А.В. Шмид, А.А. Березин, М.А. Новопашин, , Е.Ю. Зимина, № 2020116375, заявл. 2018.01.31; опубл. 2021.07.30 Бюл. № 22
32. А.Мюллер, С.Гвидо. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. Вильямс, 2017, 480 с. ISBN 978-5-99089-108-1.
33. Bishop C. M., Nasrabadi N. M. Pattern recognition and machine learning. – New York : springer, 2006. – Т. 4. – №. 4. – С. 738.
34. Фуремс Е. М. Многокритериальная порядковая классификация на основе метода STEPCLASS // Искусственный интеллект и принятие решений. 2012. №. 4. С. 104-115. DOI: 10.3103/S0147688215060064
35. Погосова Н. В. и др. Профилактический скрининг: все за и против //Кардиоваскулярная терапия и профилактика. – 2016. – Т. 15. – №. 3. – С. 4-13
36. Федеральный проект "Борьба с сахарным диабетом" призван переломить ситуацию с этим заболеванием // Российская Газета URL: <https://rg.ru/2023/12/20/sladkaia-epidemiia-dolzha-byt-ostanovlena.html> (дата обращения: 10.02.2025).
37. Дедов И. И. и др. Сахарный диабет 2 типа у взрослых //Сахарный диабет. – 2020. – Т. 23. – №. 2S. – С. 4-102.
38. «Алгоритмы специализированной медицинской помощи больным сахарным диабетом» Под редакцией И.И. Дедова, М.В. Шестаковой, А.Ю. Майорова 9-й выпуск. Сахарный диабет. 2019;22(1S1):1-144
39. Барбараш О. Л. и др. Предиабет как междисциплинарная проблема: определение, риски, подходы к диагностике и профилактике сахарного диабета 2 типа и сердечно-сосудистых осложнений //Российский кардиологический журнал. – 2019. – №. 4. – С. 83-91.
40. Novopashin M. A., Shmid A. V., Berezin A. A. Fermi-Pasta-Ulam auto recurrence in the description of the electrical activity of the heart //Medical hypotheses. – 2017. – Т. 101. – С. 12-16.

41. Carnethon, M. R.. (2003). Prospective Investigation of Autonomic Nervous System Function and the Development of Type 2 Diabetes: The Atherosclerosis Risk In Communities Study, 1987-1998. *Circulation*, 107(17), 2190–2195. doi:10.1161/01.CIR.0000066324.74807.95
42. Santhanalakshmi, D.; Gautam, S.; Gandhi, A.; Chaudhury, D.; Goswami, B.; Mondal, S. Heart Rate Variability(HRV) in prediabetics—A cross sectional comparative study in north India. *Indian J. Physiol. Pharmacol.* 2019, 63, 275–282
43. Acharya, U Rajendra et al. “An integrated diabetic index using heart rate variability signal features for diagnosis of diabetes.” *Computer methods in biomechanics and biomedical engineering* vol. 16,2 (2013): 222-34. doi:10.1080/10255842.2011.616945
44. Rajendra Acharya, U et al. “Automated identification of normal and diabetes heart rate signals using nonlinear measures.” *Computers in biology and medicine* vol. 43,10 (2013): 1523-9. doi:10.1016/j.compbiomed.2013.05.024
45. Rajendra Acharya, U.; Vidya, K. Sudarshan; Ghista, Dhanjoo N.; Lim, Wei Jie Eugene; Molinari, Filippo; Sankaranarayanan, Meena . (2015). Computer-aided diagnosis of diabetic subjects by heart rate variability signals using discrete wavelet transform method. *Knowledge-Based Systems*, 81, 56–64. doi:10.1016/j.knosys.2015.02.005
46. Jian, Lee Wei; Lim, Teik-Cheng . (2013). Automated Detection of Diabetes by Means of Higher Order Spectral Features Obtained from Heart Rate Signals. *Journal of Medical Imaging and Health Informatics*, 3(3), 440–447. doi:10.1166/jmihi.2013.1178
47. Pachori, Ram Bilas; Kumar, Mohit; Avinash, Pakala; Shashank, Kora; Acharya, U. Rajendra . (2016). an improved online paradigm for screening of diabetic patients using rr-interval signals. *Journal of Mechanics in Medicine and Biology*, 16(1), 1640003. doi:10.1142/S0219519416400030

48. G, Swapna; Kp, Soman; R, Vinayakumar . (2018). Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. *Procedia Computer Science*, 132, 1253–1262. doi:10.1016/j.procs.2018.05.041
49. Yildirim, Ozal et al. “Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals.” *Computers in biology and medicine* vol. 113 (2019): 103387. doi:10.1016/j.compbiomed.2019.103387
50. ГОСТ Р 53022.2–2008 Технологии лабораторные клинические. Требования к качеству клинических лабораторных исследований. Часть 3. Правила оценки клинической информативности лабораторных тестов
51. Zimina E., Novopashin M., Shmid A. Cloud technologies in the problems of mathematical analysis of cardiological information //IV International Conference on “Information Technology and Nanotechnology”(ITNT-2018). – 2018. – pp. 112-118.
52. Defazio A., Bach F., Lacoste-Julien S. SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives //Advances in neural information processing systems. – 2014. – Vol. 27.
53. Schmidt M., Le Roux N., Bach F. Minimizing finite sums with the stochastic average gradient //Mathematical Programming. – 2017. – Vol. 162. – pp. 83-112.
54. Fletcher R. Practical methods of optimization. – John Wiley & Sons, 2000.
55. Breiman L. Classification and regression trees. – Routledge, 2017.
56. Наркевич А.Н., Виноградов К.А. Методы определения минимально необходимого объема выборки в медицинских исследованиях. *Социальные аспекты здоровья населения* [сетевое издание] 2019; 65(6):10. URL: <http://vestnik.mednet.ru/content/view/1123/30/lang,ru/> (дата обращения: 10.02.2025). DOI:10.21045/2071-5021-2019-65-6-10

57. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). *Circulation*, 93(5), 1043–1065.
58. В. Л. Сивохов, Е. Л. Сивохова, О. А. Иванова, Ю. К. Кириллов. Использование инновационных технологий в медико-биологическом обеспечении физической культуры и спорта. Монография.– Иркутск, ООО «Призма, 2016. – 210 с.
59. Полосин В. Г. и др. Коррекция дрейфа изолинии при цифровой обработке электрокардиосигнала // *Медицинская техника*. – 2016. – №. 2. – С. 33-36.
60. J. Lee, D. D. McManus, S. Merchant and K. H. Chon, "Automatic Motion and Noise Artifact Detection in Holter ECG Data Using Empirical Mode Decomposition and Statistical Approaches," in *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1499-1506, June 2012, doi: 10.1109/TBME.2011.2175729
61. J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," in *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230-236, March 1985, doi: 10.1109/TBME.1985.325532.
62. Каплан А. Я. Вариабельность ритма сердца и характер обратной связи по результату операторской деятельности у человека // *Журнал высшей нервной деятельности*. – 1999. – Т. 48. – №. 6. – С. 345-350
63. Тишков А. В. и др. Статистический контроль качества диагностических тестов с бинарными исходами // *СПб.: Издательство СПбГМУ.-2012.-17 с.*
64. Schmid A., Mkrumyan A. Remote noninvasive detection of carbohydrate metabolism disorders by first-lead ECG screening in CardioQVARK project // *2019 Actual Problems of Systems and Software Engineering (APSSE)*. – IEEE, 2019. – pp. 139-145.

65. Node-RED Concepts : Node-RED. — Текст : электронный // Node-RED : [сайт]. — URL: <https://nodered.org/docs/user-guide/concepts> (дата обращения: 10.02.2025).

Приложение

Таблица рассматриваемых параметров variability сердечного ритма
(BCP)

Название	Описание
Статистические параметры BCP	
RRNN	Средняя длительность RR-интервалов
SDNN	Стандартное отклонение RR-интервалов
RMSSD	Среднеквадратичное различие между продолжительностью соседних RR-интервалов
Эксцесс	Коэффициент эксцесса RR-интервалов
Асимметрия	Коэффициент Асимметрии RR-интервалов
CV	Коэффициент вариации RR-интервалов
NN50	Число пар последовательных RR-интервалов, разность которых больше 50 мс
pNN50	Доля пар RR-интервалов, разность которых больше 50 мс к общему количеству RR-интервалов.
Геометрические параметры BCP	
Mo	Мода – середина диапазона на гистограмме, ордината которого максимальна
Amo	Амплитуда моды - значение ординаты гистограммы в %, соответствующее Mo .
dRR	Вариационный размах - разница между наименьшим и наибольшим значениями ряда RR-интервалов
SATI	Индекс симпатоадреналового комплекса - отношение Amo к половинному отношению RMSSD к RRNN , выраженное в процентах
SI	Стресс индекс - отношение Amo к удвоенному произведению Mo на dRR

SWAI	Индекс медленноволновой аритмии
Спектральные параметры ВСР	
TP	Средняя мощность спектра на всех частотных диапазонах
LF	Средняя мощность спектра на частотном диапазоне 0,04-0,15 Гц
HF	Средняя мощность спектра на частотном диапазоне 0,15-0,4 Гц
VLF	Средняя мощность спектра на частотном диапазоне 0,015-0,04 Гц
LF/HF	Отношение LF к HF
LFp	Мощность LF в процентном отношении к TP
HFp	Мощность HF в процентном отношении к TP
VLFp	Мощность VLF в процентном отношении к TP