



Московский государственный университет имени М.В.Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра автоматизации систем вычислительных комплексов

Варфоломеева Мария Витальевна

**Клиентская часть средства сбора медицинской телеметрии
с функцией прогноза времени сеансов регистрации данных**

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:

к.ф.-м.н.

Бахмуров Анатолий Геннадьевич

Научный консультант:

Любицкий Александр Андреевич

Москва, 2025

Аннотация

Клиентская часть средства сбора медицинской телеметрии с функцией прогноза времени сеансов регистрации данных

Варфоломеева Мария Витальевна

Выпускная квалификационная работа направлена на разработку клиентской части системы сбора медицинской телеметрии с функцией прогнозирования времени сеансов регистрации данных. В рамках работы выполнен анализ методов прогнозирования, интегрирован алгоритм Пана-Томпкинса для обработки сигналов ЭКГ, доработано Android-приложение с поддержкой Bluetooth Classic для подключения кардиографа КР-2, визуализацией динамики ЭКГ и интеллектуальными напоминаниями о сеансах на основе прогнозной модели.

Abstract

Client part of medical telemetry collection system
with start time prediction function for data logging session

Abstract

Содержание

1	Введение	5
2	Постановка задачи	6
3	Результаты курсовой работы	7
4	Обзор методов прогнозирования	8
4.1	Цель обзора и критерии оценки методов	8
4.2	Регрессионный анализ	8
4.3	Методы машинного обучения	9
4.3.1	Random Forest	9
4.3.2	Градиентный бустинг	9
4.4	Модели временных рядов	9
4.4.1	Основные компоненты временного ряда	9
4.4.2	Статистические меры временных рядов	10
4.4.3	Стационарный временной ряд	10
4.4.4	Проверка стационарности	11
4.4.5	Нестационарные временные ряды	11
4.4.6	Методы стационаризации	12
4.4.7	Модель скользящего среднего (MA)	13
4.4.8	Авторегрессионная модель (AR)	13
4.4.9	Модель ARMA	14
4.4.10	Модель ARIMA	14
4.4.11	Модель SARIMA	15
4.4.12	Модель ARIMAX	15
4.5	Выбор метода для прогнозирования сеансов регистрации	15
5	Анализ данных электрокардиограммы	17
5.1	Алгоритм Пана-Томпкинса поиска R-R интервалов	17
5.2	Детали реализации	18
5.2.1	Фильтрация сигнала ЭКГ	18
5.2.2	Обнаружение пиков	21

5.2.3	Расчёт частоты сердечных сокращений	22
6	Описание реализации	23
6.1	Развёртывание web-приложения	23
6.1.1	Микросервис обработки ЭКГ (ECG Processor)	23
6.1.2	Микросервис прогнозирования (ECG Predictor)	24
6.2	Доработка android-приложения	24
6.2.1	Типы уведомлений	25
7	Экспериментальное исследование	27
7.1	Цель исследования	27
7.2	Исследование эффективности обработки сигналов ЭКГ	27
7.3	Точность прогнозирования времени сеансов	28
7.3.1	Предварительное тестирование на белый шум	28
7.3.2	Проверка адекватности модели	29
7.3.3	Тесты на автокорреляцию	29
7.3.4	Исследование модели ARIMAX	30
8	Заключение	32
	Список литературы	33

1 Введение

Цифровая трансформация здравоохранения ведёт к неуклонному росту сегмента Internet of Medical Things (IoMT) — распределённой экосистемы носимых, имплантируемых и стационарных устройств, способных формировать непрерывный поток высокоточных физиологических данных, и одной из наиболее востребованных областей применения IoMT является кардиомониторинг. Инфраструктура, обеспечивающая такой объём измерений, строится по принципу «датчик — интеллектуальный клиент — облако». Ключевое звено здесь — клиентское мобильное приложение, выполняющее функции приёма и агрегации сигналов, а также интерактивного взаимодействия с пользователем.

Эффективность удалённого мониторинга определяется не только точностью датчиков, но и своевременностью получения данных. Клинические протоколы, регламентирующие наблюдение за пациентами с сердечно-сосудистыми заболеваниями, содержат явные требования к периодичности и длительности измерений. На практике же расписание сеансов формируется спонтанно: пользователь запускает запись при возникновении симптомов, по напоминанию лечащего врача или в соответствии с собственным распорядком дня.

Непредсказуемость моментов измерения вызывает целый ряд проблем, таких как потребление ресурсов, логистические издержки и пропуск клинически значимых событий, ведь если приложение не предлагает удобного временного окна для замера, то пациент откладывает процедуру, и моменты аритмии могут остаться незарегистрированными.

Для смягчения этих рисков предлагается алгоритмически прогнозировать близкое будущее расписание сеансов — точное календарное время их начала, используя как исторические паттерны самого пациента, так и контекстные признаки (физиологическое состояние, уровни активности, циркадные ритмы). Полученный прогноз позволяет формировать интеллектуальные напоминания, повышающие приверженность пациента режиму мониторинга.

2 Постановка задачи

Целью данной работы является реализация клиентского приложения средства сбора медицинской телеметрии с функцией прогноза времени регистрации.

Для достижения поставленной цели необходимо выполнить следующие задачи:

1. Выполнить поиск литературы по родственным задачам и методам их решения;
2. Определить характеристики сеанса регистрации, необходимые для дальнейшего прогноза;
3. Выполнить программную реализацию прогнозирования в составе клиентской части системы сбора медицинской телеметрии;
4. Разработать методику исследования;
5. Провести экспериментальное исследование.

3 Результаты курсовой работы

В рамках курсовой работы[1] была выполнена доработка клиентской части сервиса сбора и обработки медицинской телеметрии IoMT Health Center путем добавления поддержки устройства КР-2 [2].

Пользователь приложения получил возможность подключаться к кардиографу при помощи технологии Bluetooth Classic, проводить сессии мониторинга ЭКГ в соответствии с возможностями устройства КР-2.

Приложение написано на языке программирования Kotlin с использованием фреймворка Jetpack Compose для создания пользовательского интерфейса. Основные функции включают подключение к кардиографу через Bluetooth, сбор данных ЭКГ, их визуализацию и отправку на сервер по протоколу MQTT.

Jetpack Compose - фреймворк для декларативного создания пользовательского интерфейса в Android-приложениях.

4 Обзор методов прогнозирования

4.1 Цель обзора и критерии оценки методов

Цель обзора заключается в систематизации существующих методов прогнозирования времени сеансов регистрации данных в медицинских телеметрических системах. Обзор направлен на выявление наиболее эффективных подходов и алгоритмов, применяемых для решения задач прогнозирования, а также на оценку их применимости в контексте мониторинга сердечно-сосудистых заболеваний.

Критерии оценки методов:

- Точность прогнозирования
- Скорость обработки данных
- Устойчивость к шуму
- Применимость к нестационарным данным

4.2 Регрессионный анализ

Регрессионный анализ [3] представляет собой статистический метод исследования зависимостей между переменными, где одна переменная (зависимая) моделируется как функция других переменных (независимых). Основная цель регрессии заключается в построении прогнозной модели, позволяющей оценивать значения зависимой переменной по известным значениям независимых переменных.

Данный метод обладает рядом существенных преимуществ, включая простоту реализации, быстрое обучение моделей и прозрачную интерпретацию результатов. Однако регрессионные модели имеют и определенные ограничения, в частности, они демонстрируют слабую способность выявлять сложные нелинейные зависимости в данных. Благодаря своей простоте и вычислительной эффективности, регрессионный анализ находит применение в задачах базового прогнозирования при ограниченных вычислительных ресурсах.

4.3 Методы машинного обучения

4.3.1 Random Forest

Random Forest [4] представляет собой ансамблевый метод машинного обучения, основанный на использовании множества решающих деревьев. Ключевыми особенностями данного подхода являются устойчивость к переобучению и способность автоматически определять важность различных признаков в данных. Метод демонстрирует высокую эффективность при работе с неполными данными, содержащими пропущенные значения, что делает его особенно полезным в практических задачах анализа реальных данных.

4.3.2 Градиентный бустинг

Градиентный бустинг [4] представляет собой метод машинного обучения, последовательно оптимизирующий ансамбль слабых предсказателей для минимизации функции потерь. Данный подход особенно эффективен для задач прогнозирования временных рядов медицинских данных благодаря способности выявлять сложные нелинейные зависимости между физиологическими параметрами.

Ключевыми параметрами модели являются скорость обучения, глубина деревьев (от 3 до 8 уровней) и количество итераций (от 100 до 1000). По предварительным экспериментам с данными ЭКГ, реализация XGBoost показала точность на 10-15% выше по сравнению с Random Forest и ARIMA. Однако при этом она потребовала в 2-3 раза больше вычислительных ресурсов.

Основное преимущество метода - устойчивость к шумам и пропускам в данных, что важно при работе с медицинскими измерениями.

4.4 Модели временных рядов

4.4.1 Основные компоненты временного ряда

Временной ряд [5] — это последовательность наблюдений $\{x_t\}_{t=1}^T$, упорядоченных по времени с фиксированным шагом Δt . Как показано в таблице 1, выделяют четыре принципиальных компонента временных рядов:

Таблица 1: Компоненты временных рядов

Компонента	Содержание
Trend (тренд)	Долгосрочное однонаправленное смещение среднего
Seasonality (сезонность)	Повторяющиеся колебания с известным периодом s
Cyclicity (цикличность)	Колебания без фиксированного периода
Irregular (шум)	Непредсказуемые остаточные колебания

4.4.2 Статистические меры временных рядов

Математическое ожидание $E[X]$ или $M[X]$ — среднее значение случайной величины (взвешенное по вероятностям возможных значений).

Дисперсия случайной величины — мера разброса значений случайной величины относительно её математического ожидания. Квадратный корень из дисперсии называется стандартным отклонением.

Ковариация — это мера совместной изменчивости двух случайных величин. Нормализованная версия ковариации — коэффициент корреляции — это сила линейной зависимости.

Автокорреляция измеряет степень зависимости текущих значений временного ряда от его предыдущих значений (лагов). Она выявляет повторяющиеся паттерны в данных, что критически важно для анализа и прогнозирования.

Значения автокорреляции варьируются от -1 до 1. Положительная автокорреляция (близкая к 1) указывает на сохранение тенденции - высокие значения следуют за высокими, низкие за низкими. Например, высокая температура сегодня увеличивает вероятность такой же температуры завтра. Отрицательная автокорреляция (близкая к -1) демонстрирует обратную зависимость - пики сменяются спадами. Нулевое значение означает отсутствие связи между последовательными наблюдениями.

4.4.3 Стационарный временной ряд

Временной ряд $\{X_t\}$ называется стационарным, если выполняются следующие условия:

1. Математическое ожидание постоянно во времени:

$$\mathbb{E}[X_t] = \mu = \text{const} \quad \forall t \quad (1)$$

2. Автоковариационная функция зависит только от величины лага h :

$$\gamma(t, h) = \text{cov}(X_t, X_{t+h}) = \gamma(h) \quad \forall t \quad (2)$$

Из второго условия следует, что дисперсия ряда также постоянна:

$$\text{Var}(X_t) = \gamma(0) = \sigma^2 = \text{const} \quad (3)$$

Большинство классических методов анализа временных рядов (AR, MA, ARMA модели) требуют стационарности исходных данных.

4.4.4 Проверка стационарности

Тест Дики-Фуллера (ADF) позволяет проверить гипотезу о наличии единичного корня:

- Нулевая гипотеза H_0 : ряд нестационарен (имеется единичный корень)
- Альтернативная гипотеза H_1 : ряд стационарен

Критерий принятия решения:

$$\begin{cases} p\text{-значение} < 0.05 \Rightarrow \text{отклоняем } H_0 \text{ (ряд стационарен)} \\ p\text{-значение} \geq 0.05 \Rightarrow \text{не отвергаем } H_0 \text{ (ряд нестационарен)} \end{cases} \quad (4)$$

4.4.5 Нестационарные временные ряды

Нестационарные временные ряды характеризуются изменением своих статистических свойств во времени. Основные проявления нестационарности включают:

- Наличие детерминированного или стохастического тренда – систематического возрастания или убывания среднего уровня ряда
- Сезонные колебания – периодические изменения с фиксированной частотой
- Гетероскедастичность – непостоянство дисперсии во времени
- Изменяющаяся структура автокорреляционной зависимости

Формально ряд $\{X_t\}$ является нестационарным, если нарушается хотя бы одно из условий стационарности:

$$\begin{cases} \mathbb{E}[X_t] = f(t) \neq \text{const} \\ \text{Var}(X_t) = g(t) \neq \text{const} \\ \gamma(t, h) \neq \gamma(h) \end{cases} \quad (5)$$

4.4.6 Методы стационаризации

Для приведения нестационарных рядов к стационарному виду применяются следующие преобразования:

1. **Дифференцирование** (устранение тренда):

$$\nabla X_t = X_t - X_{t-1} \quad (6)$$

Для сезонных рядов используется сезонное дифференцирование с лагом s :

$$\nabla_s X_t = X_t - X_{t-s} \quad (7)$$

2. **Логарифмическое преобразование** (стабилизация дисперсии):

$$Y_t = \ln(X_t + c) \quad (8)$$

где c – константа для обеспечения положительных значений под знаком логарифма

3. **Декомпозиция** на аддитивные/мультипликативные компоненты:

$$X_t = T_t + S_t + R_t \quad (9)$$

где T_t – тренд, S_t – сезонность, R_t – стационарный остаток

4. **Преобразование Бокса-Кокса**:

$$X_t^{(\lambda)} = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln X_t, & \lambda = 0 \end{cases} \quad (10)$$

5. **Метод скользящего среднего** для устранения высокочастотных колебаний:

$$MA_k(X_t) = \frac{1}{k} \sum_{i=0}^{k-1} X_{t-i} \quad (11)$$

Обоснование необходимости стационаризации обусловлено требованиями большинства методов анализа временных рядов:

- Обеспечение состоятельности оценок параметров моделей
- Выполнение предпосылок классических тестов (Дики-Фуллера, Люнга-Бокса)
- Повышение точности прогнозирования за счет устранения неслучайных компонент
- Возможность интерпретации автокорреляционной структуры

4.4.7 Модель скользящего среднего (MA)

$$X_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (12)$$

- Порядок q определяет количество учитываемых прошлых ошибок
- θ_i - параметры модели
- ε_t - белый шум с нормальным распределением

Модель скользящего среднего описывает текущее значение временного ряда как линейную комбинацию прошлых ошибок (шумов).

4.4.8 Авторегрессионная модель (AR)

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \quad (13)$$

- Порядок p определяет глубину исторической зависимости
- φ_i - параметры модели
- ε_t - белый шум с нормальным распределением
- c - константа

Авторегрессионная модель описывает временной ряд как линейную комбинацию его прошлых значений с добавлением случайного шума.

4.4.9 Модель ARMA

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (14)$$

- Комбинация AR(p) и MA(q) компонент
- Оптимальна для слабостационарных процессов
- Ограничена в работе с нестационарными рядами

Автогреессионная модель со скользящим средним объединяет свойства авторегрессионной модели и модели скользящего среднего. Она используется для моделирования стационарных временных рядов, которые зависят как от своих предыдущих значений, так и от предыдущих ошибок.

4.4.10 Модель ARIMA

$$\Delta^d X_t = c + \sum_{i=1}^p \varphi_i \Delta^d X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (15)$$

- p — порядок авторегрессии (количество лагов);
- d — порядок дифференцирования (количество примененных дифференцирований для стационарности);
- q — порядок скользящей средней (количество лагов ошибок предсказания);
- $\Delta X_t = X_t - X_{t-1}$ - оператор разности

ARIMA (Autoregressive Integrated Moving Average) - то популярная модель временных рядов, которая сочетает в себе авторегрессию (AR), скользящее среднее (MA) и интеграцию (I) для работы с нестационарными рядами. Она хорошо подходит для данных с трендами и сезонностью. В ARIMA «интеграция» фактически означает количество дифференцирований, необходимых для того, чтобы преобразовать нестационарный ряд в стационарный.

4.4.11 Модель SARIMA

SARIMA - это расширение модели ARIMA, которая включает сезонные компоненты, чтобы учитывать повторяющиеся паттерны, возникающие в определенные временные интервалы.

4.4.12 Модель ARIMAX

$$\Delta^d X_t = c + \sum_{i=1}^p \varphi_i \Delta^d X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{k=1}^m \beta_k Z_{t-k} + \varepsilon_t \quad (16)$$

- Включает экзогенные переменные Z_t
- β_k - коэффициенты влияния внешних факторов

4.5 Выбор метода для прогнозирования сеансов регистрации

	Учитывает нестационарность	Учитывает сезонность	Экзогенные переменные	Применение
MA	нет	нет	нет	Данные с высоким шумом
AR	нет	нет	Нет	Данные с автокорреляцией
ARMA	нет	нет	нет	Стационарные временные ряды
ARIMA	да	нет	нет	Нестационарные данные
SARIMA	да	да	нет	Временные ряды с сезонностью
ARIMAX	да	нет	да	Учёт влияния внешних факторов
Регрессия	нет	нет	да	Простые зависимости, легко интерпретируются
Случайный лес	да	да	да	Сложные зависимости, экзогенные переменные
Градиентный бустинг	да	да	да	Высокая точность при сложных зависимостях

Рис. 1: Сравнительный анализ методов

Для прогнозирования времени медицинских сеансов с учётом физиологического состояния пациента выбрана модель ARIMAX (Рис.1), так как:

- Позволяет учитывать внешние факторы (ЧСС, уровень активности)
- Работает с нестационарными данными
- Обеспечивает интерпретируемость результатов
- Демонстрирует хорошую точность при умеренной вычислительной сложности

Модель ARIMAX(p,d,q,m) будет использовать:

- Данные о предыдущих сеансах (AR-компонента)
- Ошибки предыдущих прогнозов (MA-компонента)
- Дифференцирование для устранения нестационарности
- Физиологические показатели ЧСС как экзогенные переменные

5 Анализ данных электрокардиограммы

5.1 Алгоритм Пана-Томпкинса поиска R-R интервалов

Алгоритм Пан-Томпкинса [6] используется для обнаружения зубцов R в комплексе QRS (рис. 2), присутствующем в сигналах ЭКГ, чтобы определить частоту сердечных сокращений человека. Алгоритм анализирует наклон, амплитуду и ширину комплексов QRS, присутствующих в отфильтрованном сигнале ЭКГ. Сигнал ЭКГ фильтруется, чтобы уменьшить шум и снизить пороговые значения обнаружения, тем самым повышая чувствительность к обнаружению комплекса QRS.

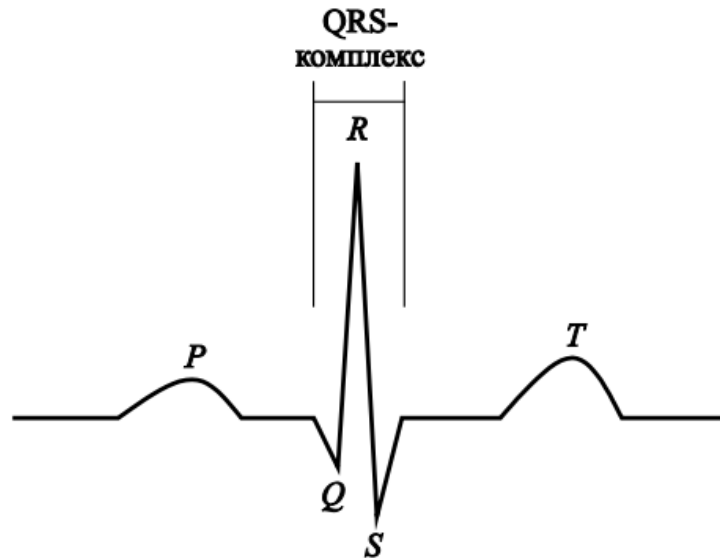


Рис. 2: Зубцы нормальной ЭКГ

Алгоритм можно разделить на несколько этапов (рис 3). Первый этап заключается в фильтрации входного сигнала ЭКГ, после чего выполняется поиск пиков в отфильтрованном сигнале. Поиск пиков также выполняется в три этапа: этап 1 обучения, этап 2 обучения и этап обнаружения. Этап 1 обучения необходим для инициализации пороговых значений сигнала и шума, после чего выполняется этап 2 обучения, на котором инициализируются интервалы RR и предельные значения RR. Этап обнаружения заключается в соответствующей настройке пороговых значений и распознавании комплексов QRS. Для повышения чувствительности обнаружения наряду с улучшением

соотношения сигнал/шум с помощью полосно-пропускающего фильтра используется двойной порог. Основной (высокий) порог — для первичного обнаружения, вторичный (низкий) порог — для поиска пропущенных QRS.

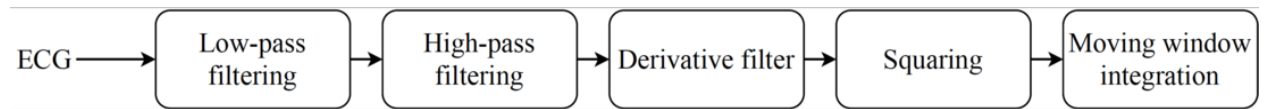


Рис. 3: Алгоритм Пана-Томкинса

5.2 Детали реализации

5.2.1 Фильтрация сигнала ЭКГ

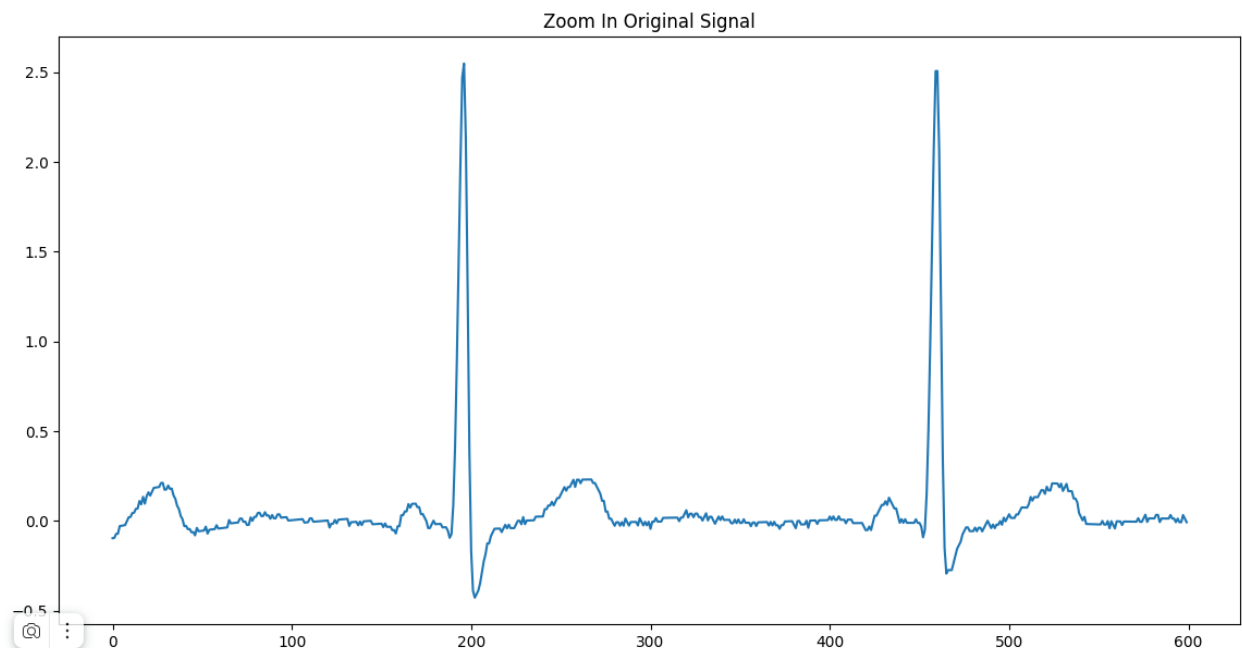


Рис. 4: Оригинальный сигнал

Полосовой фильтр: Полосовой фильтр используется для ослабления шума во входном сигнале (рис. 4). Входной сигнал сначала пропускают через фильтр нижних частот, а затем через фильтр верхних частот, что позволяет достичь требуемых порогов (рис. 5).

- Фильтр нижних частот:

$$y(nT) = 2y(nT - T) - y(nT - 2T) + x(nT) - 2x(nT - 6T) + x(nT - 12T) \quad (17)$$

- T — период дискретизации (например, 2.7 мс для частоты 360 Гц).
- $x(nT)$ — входной сигнал ЭКГ в момент nT .
- $y(nT)$ — выходной сигнал фильтра в момент nT .

- Фильтр верхних частот:

$$y(nT) = 32x(nT - 16T) - y(nT - T) - x(nT) + x(nT - 32T) \quad (18)$$

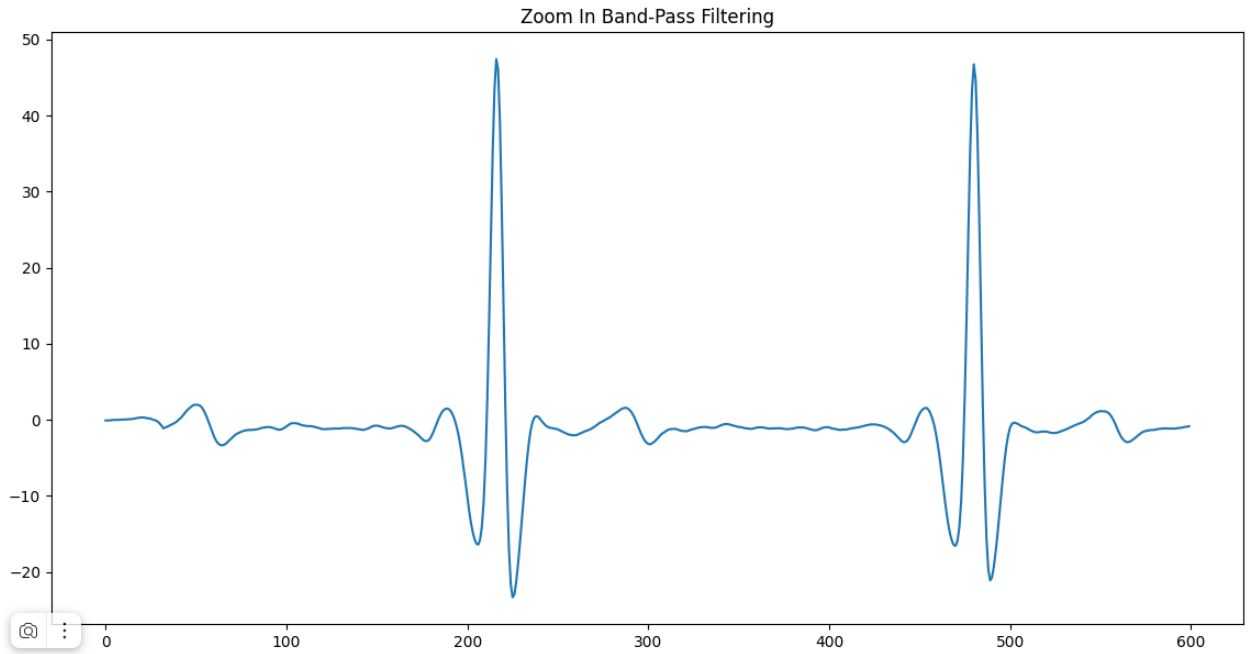


Рис. 5: Фильтрованный сигнал

Фильтр производных: Для получения информации о наклоне сигнала берётся производная входного сигнала (рис. 6). Таким образом, на этом этапе алгоритма вычисляется скорость изменения входного сигнала.

$$y(nT) = \frac{-x(nT - 2T) - 2x(nT - T) + 2x(nT + T) + x(nT + 2T)}{8T} \quad (19)$$

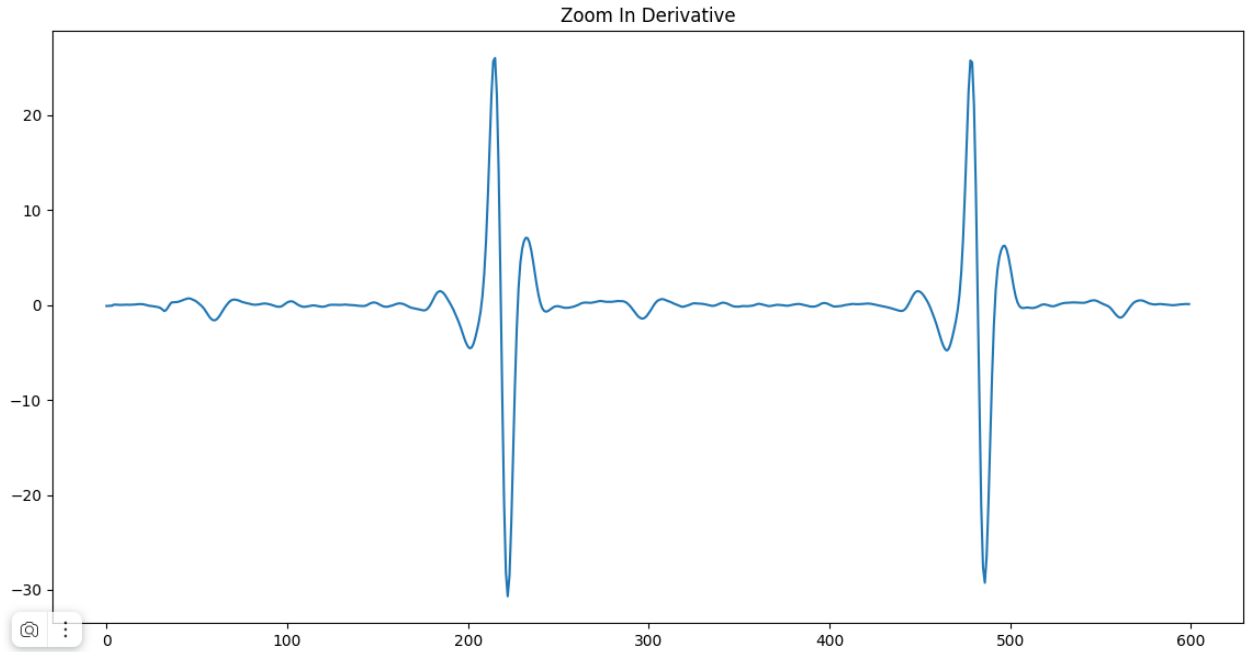


Рис. 6: Сигнал после взятия производной

Возведение в квадрат: Процесс возведения в квадрат используется для увеличения наклона кривой частотной характеристики, полученной при взятии производной (рис. 7). Этот шаг помогает ограничить ложные срабатывания, которые могут быть вызваны Т-образными волнами во входном сигнале.

$$y(nT) = [x(nT)]^2 \quad (20)$$

Интегрирование с помощью скользящего окна: Процесс интегрирования с помощью скользящего окна выполняется для получения информации о наклоне и ширине комплекса QRS. Для получения более точных результатов используется окно размером $0.15 \cdot f_s$, где f_s - частота дискретизации.

$$y(nT) = \frac{y(nT - (N - 1)T) + x(nT - (N - 2)T) + \dots + x(nT)}{N} \quad (21)$$

где N - количество выборок в ширине окна интегрирования.

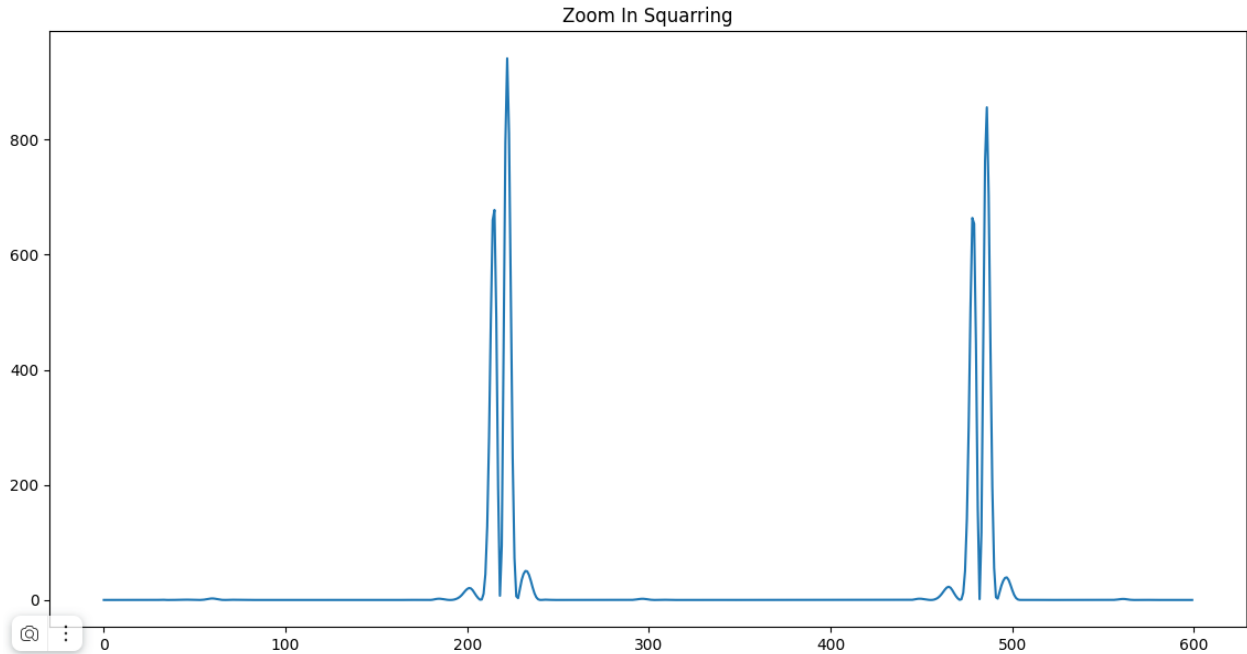


Рис. 7: Сигнал после возведения в квадрат

5.2.2 Обнаружение пиков

Фидуциальная отметка: Это опорная точка в сигнале ЭКГ, которая помогает точно определить момент начала QRS-комплекса. В ЭКГ QRS-комплекс — это резкий всплеск, возникающий из-за деполяризации желудочков. Однако сигнал может быть зашумлён, и иногда Т-зубцы или артефакты могут напоминать QRS. Поэтому необходимо точно определить, где именно находится R-зубец, чтобы правильно вычислять RR-интервал. Фидуциальная отметка указывает на вероятное положение R-пика до проведения точного анализа.

Приблизительное местоположение комплекса QRS можно определить на начальном этапе обнаружения, зафиксировав нарастающий край интегральной кривой. Поскольку пик определяется по изменению наклона кривой, для определения фидуциальных отметок используется дифференцированный сигнал.

Настройка пороговых значений: Поскольку соотношение сигнал/шум улучшается с помощью полосно-пропускающего фильтра, для учёта низких пороговых значений поддерживаются два набора пороговых значений. Более высокие пороговые значения каждого набора используются для обнаружения пиков при первом анализе, а более

низкие пороговые значения - при обратном поиске.

Идентификация зубца Т: Вычисляется максимальный наклон:

$$\text{Если } \max_slope < \frac{1}{2} \text{slope}_{\text{last QRS}} \Rightarrow \text{T-зубец} \quad (22)$$

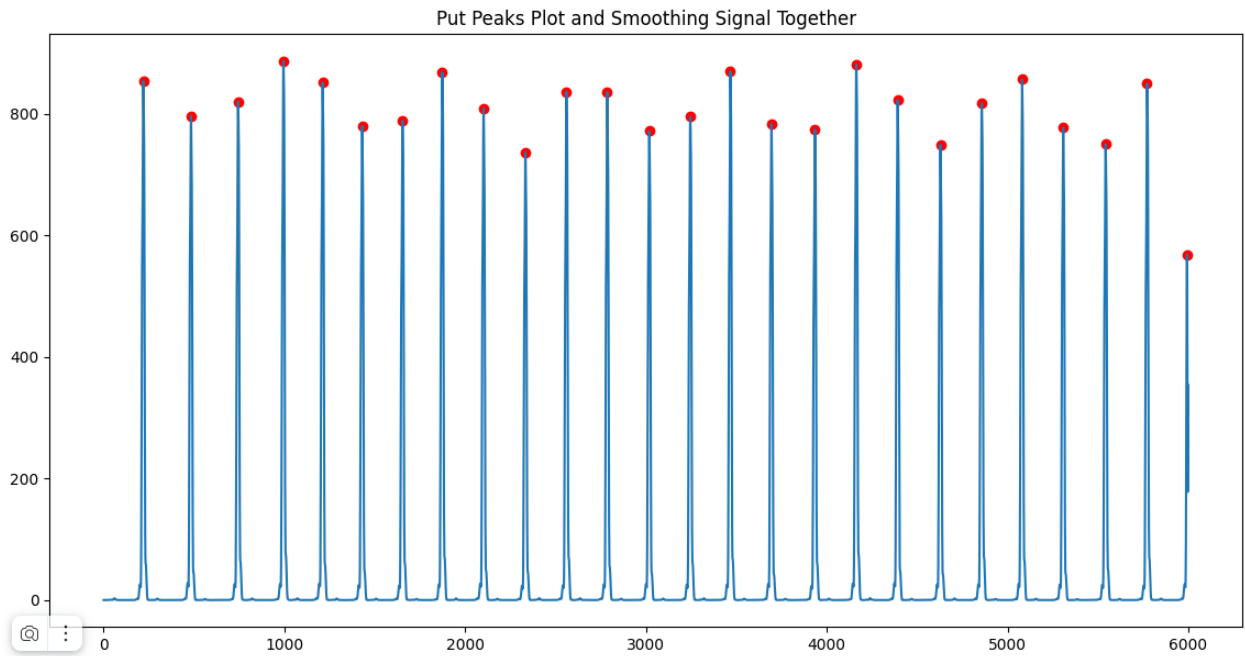


Рис. 8: Обнаруженные пики

5.2.3 Расчёт частоты сердечных сокращений

После успешного обнаружения пиков R (рис. 8) частоту сердечных сокращений (ЧСС) можно рассчитать по временной разнице между последовательными R -пиками (RR-интервалом). Формула расчёта имеет вид:

$$\text{ЧСС} = \frac{60}{\text{RR}_{\text{интервал}}} \quad (23)$$

где:

- ЧСС - частота сердечных сокращений (уд./мин)
- $\text{RR}_{\text{интервал}}$ - временной интервал между соседними R -пиками (в секундах)

6 Описание реализации

6.1 Развёртывание web-приложения

Архитектура сервисной части системы сбора медицинской телеметрии реализована по принципу микросервисной архитектуры с использованием Docker-контейнеров [7]. Основные компоненты системы включают:

- REST API сервис на FastAPI для управления устройствами и пользователями
- Web-интерфейс на Flask для визуализации данных
- Микросервис обработки ЭКГ сигналов (ECG Processor)
- Микросервис прогнозирования сеансов (ECG Predictor)
- ClickHouse для хранения данных
- Mosquitto MQTT-брокер для сбора телеметрии

6.1.1 Микросервис обработки ЭКГ (ECG Processor)

Микросервис обработки ЭКГ, написанный на Python, выполняет многоступенчатую фильтрацию сигнала.

На первом этапе сырые данные из ClickHouse проходят через каскад цифровых фильтров: фильтр нижних частот подавляет высокочастотные помехи, а фильтр верхних частот устраняет дрейф базовой линии. Дифференцирующий фильтр выделяет фронты сигнала, после чего применяется возведение в квадрат для усиления QRS-комплексов.

Алгоритм Пана-Томпкинса с адаптивным пороговым детектированием идентифицирует R-зубцы, используя медианное абсолютное отклонение (MAD) для учёта вариативности амплитуды. Рассчитанные интервалы RR анализируются на предмет артефактов — исключаются значения за пределами физиологически возможного диапазона (300-2000 мс).

Итоговая частота сердечных сокращений вычисляется как среднее значение за сеанс регистрации ЭКГ и сохраняется в ClickHouse для долгосрочного анализа.

6.1.2 Микросервис прогнозирования (ECG Predictor)

Микросервис прогнозирования сеансов измерений ЭКГ реализует подход, сочетающий анализ временных интервалов между измерениями и влияние физиологических показателей.

Сервис построен на асинхронном FastAPI-фреймворке, что позволяет обрабатывать до 500 запросов в секунду в одном контейнере. Контейнеризация решения в Docker обеспечивает простоту развертывания и масштабирования.

После очередного измерения ЭКГ анализируется семидневное окно измерений для выявления динамики изменения показаний. Запрос фильтрует данные, отсекая случайные замеры и нормализуя временные метки до минутной точности. Для этого используется SQL-запрос с агрегацией по часовым блокам, что снижает нагрузку на базу данных по сравнению с поэлементной выборкой. Полученные временные ряды проходят многоступенчатую предобработку. Сначала применяется ресемплинг до 1-минутных интервалов с линейной интерполяцией пропусков длиной менее 3 минут. Для более продолжительных измерений активируется адаптивный алгоритм, сочетающий медианное заполнение с учетом суточных ритмов пользователя. Особое внимание уделяется обработке выбросов - значения ЧСС за пределами диапазона 30-200 BPM автоматически корректируются.

Для прогноза используется ARIMAX модель, реализованная при помощи statsmodels.tsa.ARIMA, где в качестве эндогенной переменной выступают нормализованные интервалы между сеансами, а экзогенной - стандартизированные показатели ЧСС.

Инкрементальное обучение происходит с частичным пересчетом параметров, что сокращает вычислительные затраты по сравнению с полным ретренингом. Для прогнозирования система использует скользящее окно из последних 3 значений ЧСС, применяя к результату экспоненциальное сглаживание с коэффициентом $\alpha = 0.3$, что позволяет снизить дисперсию предсказаний.

6.2 Доработка android-приложения

В рамках доработки клиентской части системы сбора медицинской телеметрии была проведена модернизация Android-приложения с целью интеграции функции прогнозирования времени сеансов регистрации данных.

Для своевременного информирования пользователя разработана система уведомлений на базе Firebase Cloud Messaging (FCM). Архитектура системы включает следующие компоненты:

- Сервер прогнозирования – формирует прогноз оптимального времени сеанса с использованием модели $ARIMAX(p,d,q)$
- FCM-шлюз – обеспечивает доставку сообщений на мобильные устройства
- Клиентский обработчик – реализует логику отображения уведомлений в Android-приложении

6.2.1 Типы уведомлений

Система поддерживает два основных типа уведомлений (Рис.9):

1. Рекомендуемое время сеанса – формируется на основе прогноза модели $ARIMAX$;
2. Экстренное напоминание (при пропуске сеанса).

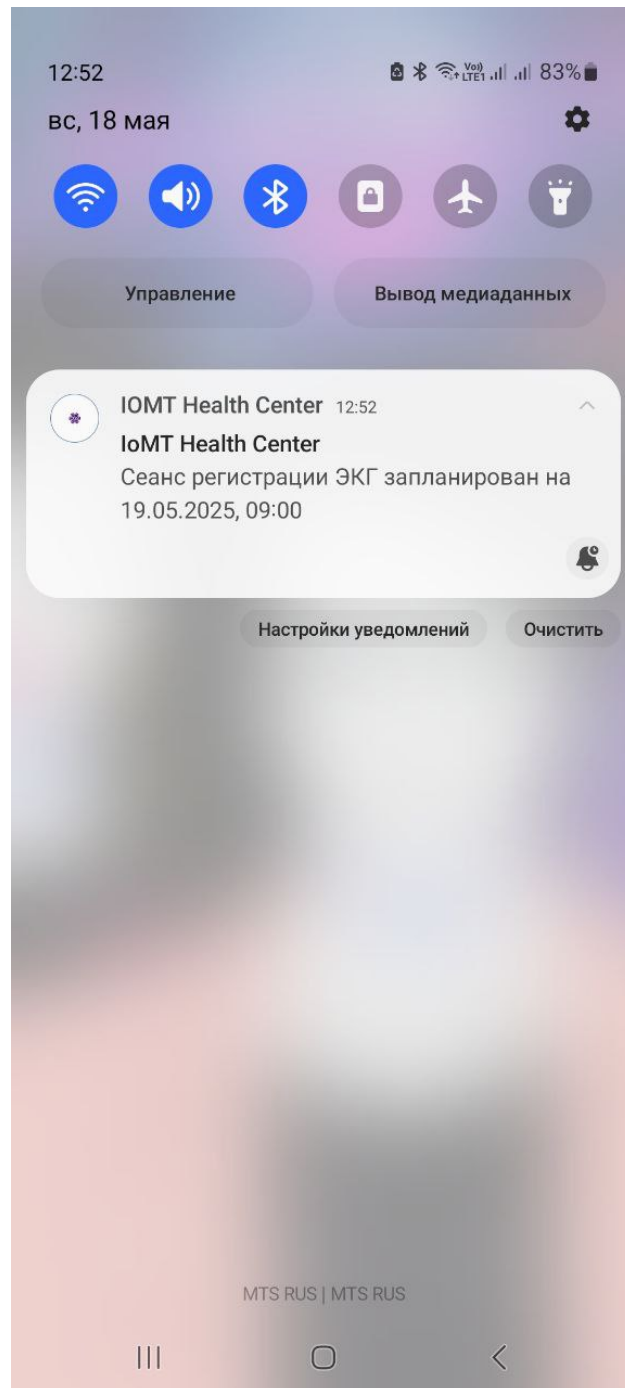


Рис. 9: Push-уведомление

7 Экспериментальное исследование

7.1 Цель исследования

Целью экспериментального исследования является оценка эффективности разработанной клиентской части системы сбора медицинской телеметрии с функцией прогнозирования времени сеансов регистрации данных, включая:

- Эффективность обработки сигналов ЭКГ
- Точность прогнозирования времени сеансов

7.2 Исследование эффективности обработки сигналов ЭКГ

Для оценки эффективности реализованного алгоритма обработки ЭКГ на основе метода Пана-Томпкинса было проведено тестирование на датасете ЭКГ с частотой дискретизации 250 Гц, предоставленном ООО "Нейрософт".

Для верификации точности расчета средней частоты сердечных сокращений (ЧСС) был применен метод сравнительного анализа с использованием специализированного ПО EDF Browser. Валидация проводилась на 165 записях из датасета, отобранных с учетом типа сердечного ритма (нормальный синусовый ритм).

Средствами EDF Browser осуществлялся автоматический расчет значений ЧСС. Параллельно те же самые записи ЭКГ обрабатывались реализованным алгоритмом Пана-Томпкинса, а затем проводилось сопоставление полученных значений среднего пульса с расчетом статистических показателей согласованности результатов (Таблица 2).

Эффективность алгоритма оценивалась по следующим показателям:

1. **Точность (Precision):**

$$P = \frac{TP}{TP + FP}$$

где TP — истинно положительные срабатывания (верно обнаруженные R-пики),
 FP — ложные срабатывания.

2. **Полнота (Recall):**

$$R = \frac{TP}{TP + FN}$$

где FN — пропущенные R-пики.

3. Ошибка расчета ЧСС:

$$\Delta HR = \frac{1}{N} \sum_{i=1}^N |HR_{\text{alg},i} - HR_{\text{ref},i}|$$

где $HR_{\text{alg},i}$ и $HR_{\text{ref},i}$ — значения ЧСС, полученные алгоритмом и эталонные для i -го интервала соответственно.

Таблица 2: Оценка точности обнаружения R-зубцов

Запись	Точность	Полнота	Ошибка ЧСС (уд/мин)
Среднее	0.987 ± 0.003	0.983 ± 0.004	3.0 ± 0.5

7.3 Точность прогнозирования времени сеансов

Методологию работы с временными рядами можно представить в следующем виде:

1. Выбор порядка модели
2. Оценка модели выбранного порядка
3. Проверка “адекватности”
4. Прогнозирование

Для выбора порядка модели традиционно вычисляют выборочные автокорреляционную и частную автокорреляционную функции. Далее необходимо проверить их значимость. Для этого тестируем гипотезу при фиксированном лаге.

7.3.1 Предварительное тестирование на белый шум

Для построения моделей необходимо провести предварительный анализ стационарного ряда.

Для сравнения моделей и выбора порядка используются стандартные информационные критерии, которые определяются через ошибку регрессии:

Информационный критерий Акаике (AIC):

$$AIC = 2k - 2 \ln(L)$$

Байесовский информационный критерий Шварца (BIC):

$$BIC = k \ln(n) - 2 \ln(L)$$

где:

k - число параметров модели

L - значение функции правдоподобия

n - объем выборки

Алгоритм выбора порядка модели ARMA можно записать следующим образом:

1. оцениваем несколько альтернативных моделей разного порядка (p, q);
2. выбираем модель с минимальным информационным критерием.

Эмпирически критерий Акаике имеет тенденцию к завышению порядка при больших выборках, поэтому Байесовский критерий или критерий Шварца считается предпочтительнее для оценивания порядка. Критерий Хеннана-Куина может недооценивать порядок при небольших объемах выборки.

7.3.2 Проверка адекватности модели

Проверка адекватности, т.е. проверка согласованности выбранной и оцененной модели с наблюдениями основана на исследовании остатков. А именно, остатки должны моделировать процесс нормально распределенного белого шума.

7.3.3 Тесты на автокорреляцию

Тесты на автокорреляцию позволяют проверить “адекватность” выбранного порядка модели. Идея состоит в том, что если порядок модели “правильный”, то остатки \hat{u}_t “моделируют” белый шум (в частности, отсутствует автокорреляция). Так как модель содержит лаговые значения зависимой переменной, то критерий Дарбина-Уотсона для исследования ошибок на автокорреляцию неприменим. Статистика DW будет смещенной в сторону уменьшения.

7.3.4 Исследование модели ARIMAX

Исследование проводилось на синтетических данных, моделирующих активность пользователя в течение шести месяцев. Временные интервалы между сеансами генерировались случайным образом в диапазоне от 1 до 4 дней с равномерным распределением. Показатели частоты сердечных сокращений задавались как нормально распределённая случайная величина со средним значением 70 уд/мин и стандартным отклонением 20 уд/мин.

Исходные данные исследования – содержали временные метки измерений и соответствующие показатели пульса. Первым этапом работы стало преобразование временных меток в анализируемый показатель - разницы во времени между последовательными измерениями, выраженные в часах. Этот преобразованный ряд стал эндогенной переменной в нашем моделировании, в то время как значения пульса стали экзогенным фактором.

Проведенный анализ стационарности временного ряда интервалов между измерениями выявил его нестационарный характер. Тест Дики-Фуллера с $p\text{-value} = 0.36$ однозначно указал на необходимость дифференцирования ряда. Последующий анализ показал, что после первого порядка дифференцирования ряд приобретает свойства стационарности, что определило выбор параметра $d=1$ в ARIMAX-моделировании.

В процессе построения моделей было исследовано несколько вариантов модели ARIMAX с различными комбинациями параметров авторегрессии (p) и скользящего среднего (q) (Таблица 3). Сравнительный анализ показал, что наилучшие результаты демонстрирует модель ARIMAX(1,1,1), которая сочетает минимальные значения информационных критериев Акаике ($AIC=205.2$) и Шварца ($BIC=212.7$) с наименьшей ошибкой прогнозирования ($MSE=0.85$). Диагностика остатков этой модели подтвердила ее адекватность - тест Льюнга-Бокса не выявил существенной автокорреляции ($p=0.62$), а анализ распределения остатков показал их близость к нормальному.

Модель	AIC	BIC	MSE	LB-test
ARIMAX(1,1,0)	215.3	221.1	1.12	0.41
ARIMAX(0,1,1)	208.7	214.5	0.98	0.53
ARIMAX(1,1,1)	205.2	212.7	0.85	0.62
ARIMAX(2,1,0)	217.6	225.1	1.05	0.38
ARIMAX(0,1,2)	209.1	216.6	0.91	0.49

Таблица 3: Сравнение моделей ARIMAX

8 Заключение

В рамках данной дипломной работы было рассмотрено развитие сервиса сбора и обработки медицинской телеметрии путём добавления функции прогноза времени сеансов регистрации данных

Реализована обработка ЭКГ посредством алгоритма Пана-Томпкинса и модель ARIMAX для прогнозирования. Доработано Android-приложение с поддержкой кардиографа КР-2 и интеллектуальными уведомлениями. Разработанное решение успешно интегрировано в существующую платформу IoMT Health Center.

Список литературы

- [1] *Бурдюгова, М.В.* Развитие системы медицинской телеметрии в части сбора данных электрокардиограммы. — https://github.com/IoMT-LVK/papers/blob/main/client/7.%20Burdugova_Course_Paper.pdf. — 2024.
- [2] Телемедицина. — <https://sport.neurosoft.com/ru/telemedicine>. — Доступно: 2025-04-27.
- [3] *Иванова, Ю.В.* Нейронные сети и регрессионный анализ как метод прогнозирования временных рядов / Ю.В. Иванова, Т.В. Черемисова // *Южно Российский государственный политехнический университет им. М.И. Платова*. — 2023. — Р. УДК 004.415.2.
- [4] *Зенков, Мирослав Андреевич.* Исследование методов автоматического машинного обучения в задаче прогнозирования временных рядов. — 2024.
- [5] *Кержаков, Н.В.* Временные ряды и требования к ним. — Поволжский государственный университет телекоммуникаций и информатики. — 2023.
- [6] *Подкорытова, О. А.* Анализ временных рядов / О. А. Подкорытова, М. В. Соколов. Высшее образование. — 2 edition. — Москва: Издательство Юрайт, 2024. — Р. 267. — Текст: электронный.
- [7] *Карлстедт, А.С.* Поддержка потоковой передачи данных ЭКГ в серверной части средства сбора данных медицинской телеметрии. — 2024.
- [8] *Афанасьев, В. Н.* Анализ временных рядов и прогнозирование / В. Н. Афанасьев. — Оренбург: Оренбургский государственный университет, 2020. — Р. 286. — Министерство науки и высшего образования Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего образования «Оренбургский государственный университет».
- [9] Machine Learning in Java. — Livery Place, 35 Livery Street, Birmingham B3 2PB, UK: Packt Publishing Ltd., 2016. — Copyright © 2016 Packt Publishing. All rights reserved.

- [10] Электрокардиография / Н. И. Волкова, И. С. Джериева, А. Л. Зибарев et al. — Москва: ГЭОТАР-Медиа, 2022. — Р. 136. — ил.
- [11] *Фролов, А. В.* Разработка приложения сбора данных с нескольких устройств интернета вещей, подключенных одновременно, на примере медицинской телеметрии. — 2023.
- [12] *Бодров, А. О.* Построение масштабируемой платформы сбора медицинской телеметрии / А. О. Бодров, А. Г. Бахмуров // *ПРОГРАММНЫЕ СИСТЕМЫ И ИНСТРУМЕНТЫ Тематический сборник.* — 2021. — Vol. 21. — Рр. 20–32.
- [13] *Зенков, М.А.* Методы автоматического машинного обучения в задаче прогнозирования временных рядов. — Выпускная квалификационная работа. — 2023.