


«УТВЕРЖДАЮ»

Директор Федерального
государственного учреждения
«Федеральный исследовательский центр
«Информатика и управление»
Российской академии наук»,
член-корреспондент РАН

М.А. Посышкин


«17» марта 2025 г.**ОТЗЫВ**

ведущей организации – Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» – на диссертацию Новикова Романа Сергеевича «Методы машинного обучения по спектрам кардиологических данных для проблемно-ориентированных цифровых продуктов», представленную на соискание ученой степени кандидата технических наук по специальности 2.3.5 – математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

Диссертационная работа посвящена разработке новых инструментов для создания телемедицинских систем дистанционного мониторинга кардиологических данных с использованием методов машинного обучения (классификации) на основе амплитудно-частотных спектров кардиологических данных. В работе предложены методы бинарной и порядковой классификации и архитектура фреймворка машинного обучения, направленного на формирование моделей удаленного выявления состояний пациента по спектрам кардиологических данных.

Актуальность темы.

Развитие телемедицинских систем дистанционного мониторинга кардиологических данных открывает новые возможности для пациентов и медицинских учреждений. Благодаря современным информационным технологиям, пациенты могут получать качественную медицинскую помощь без частых посещений поликлиники, а медучреждения разгружаются от избыточного потока пациентов.

Существующие портативные кардиографы, такие как смарт-часы с функцией ЭКГ, позволяют неинвазивно собирать кардиологические данные в комфортных для пациента условиях, преобразовывать в цифровую форму и передавать по Интернету в центр обработки, где их возможно анализировать, в т.ч. с помощью моделей машинного обучения. При этом анализ амплитудно-частотных спектров кардиологических сигналов помогает выявлять показатели состояния здоровья, которые могут быть скрыты в исходных данных из-за искажений при их самостоятельном сборе.

Однако разработчики рассматриваемых систем сталкиваются с серьезной проблемой: создание эффективных модулей классификации различных состояний пациента по кардиологическим данным является наукоемким и длительным процессом, особенно когда для обучения моделей не хватает подходящих данных и их нужно собирать отдельно. Это затруднение подчеркивает необходимость в специализированном инструменте, который ускорит разработку таких модулей как проблемно-ориентированных цифровых продуктов.

В связи с этим диссертационное исследование, направленное на разработку такого специализированного инструмента, основанного на методах машинного обучения и учитывающего особенности кардиологических данных, представляется актуальным и своевременным.

Анализ содержания диссертационной работы.

Диссертационная работа имеет общий объем 138 страниц и состоит из введения, четырех глав, заключения, списка литературы из 65 наименований, 15 рисунков, 9 таблиц и одного приложения.

В введении содержится обоснование актуальности исследования, ставятся цели и задачи работы, формулируется научная новизна и практическая значимость работы, а также приводятся основные положения, выносимые на защиту.

В первой главе проводится анализ архитектуры телемедицинских систем дистанционного мониторинга кардиологических данных. Особое внимание уделяется важности создания и внедрения модулей для классификации состояния пациентов. Глава также раскрывает значимость спектрального анализа кардиологических сигналов в подобных системах. Глава заканчивается рассмотрением прикладной задачи разработки телемедицинской системы дистанционного мониторинга кардиологических данных для выявления признаков нарушения углеводного обмена (кардиопризнаков НУО).

Во второй главе разработаны методы построения моделей бинарной и порядковой классификации на основе амплитудно-частотных спектров кардиологических данных. Предложена методика формирования выборки кардиологических данных для построения моделей.

В третьей главе представлены результаты валидации созданных моделей классификации в рамках приведенной прикладной задачи с помощью предложенных методов. Валидация проводилась в ходе двух клинических исследований с использованием разработанных методов и методики.

В четвертой главе описана архитектура фреймворка для разработки модулей классификации (состояния пациента) по кардиологическим данным как цифровых продуктов. Приводится реализация данного фреймворка и демонстрируется его практическое применение при разработке телемедицинской системы удаленного мониторинга кардиологических данных, решающей описанную ранее прикладную задачу.

В заключении перечислены основные результаты диссертационной работы.

Автореферат диссертации полно и корректно отражает ее содержание.

Основные результаты диссертационной работы.

В диссертационной работе Р.С. Новикова получены следующие основные результаты.

1. Разработан метод построения моделей машинного обучения с учителем для бинарной классификации объектов для телемедицинских систем дистанционного мониторинга кардиологических данных на основе машинного обучения при помощи анализа амплитудно-частотных спектров (Фурье-спектров) кардиологических данных.
Метод позволяет построить модель машинного обучения для внедрения в модуль классификации как проблемно-ориентированного цифрового продукта для телемедицинских систем дистанционного мониторинга кардиологических данных.
2. Разработан метод построения моделей машинного обучения с учителем для порядковой классификации объектов при помощи имитационного моделирования выборки по результатам работы бинарного классификатора кардиологических данных.
Метод применим в условиях, когда для использования методов машинного обучения на исходной выборке достаточно размеченных данных по объектам только «крайних» классов – классов, чей диапазон характеризующего значения строго больше, либо строго меньше диапазона аналогичных значений у других классов. Метод основан на генерации недостающей выборки с помощью имитационного моделирования методом Монте-Карло.
3. Предложена методика формирования выборки данных для создания моделей классификации по спектрам кардиологических данных.
Для данной методики сформулированы требования к исходному набору данных (ЭКГ и метаданные о пациенте) и рекомендуемые алгоритмы предварительной обработки сигналов ЭКГ и преобразования их в спектры кардиологических данных, используемые при дальнейшем анализе.
4. Предложена архитектура фреймворка для разработки программных модулей классификации в качестве проблемно-ориентированных цифровых продуктов для телемедицинских систем дистанционного мониторинга кардиологических данных по разработанным методам построения моделей машинного обучения с учителем для бинарной и порядковой классификаций по спектрам кардиологических данных.
Описана реализация данного фреймворка на основе программной платформы Peraspera Cardio с использованием инструмента Node-RED.
5. При помощи реализованного фреймворка для разработки моделей классификации разработаны программные модули выявления кардиопримарных нарушений углеводного обмена (НУО) по Фурье-спектрам кардиологических данных и модель выявления подозрения на НУО по серии Фурье-спектров кардиологических данных. Модели машинного обучения, реализованные в этих модулях, защищены патентами: №2728869 (РФ, также подтвержден как патент Японии №7562338) и №2751817 (РФ) – и применены при разработке телемедицинской системы дистанционного мониторинга кардиологических данных на основе машинного обучения – медицинской системы самостоятельного управления здоровьем за счет скрининга состояния пациента по последовательности электрокардиограмм сердца первого отведения (ЭС СД2).

Теоретическая и практическая значимость.

Теоретическая значимость диссертации заключается в разработке новых методов и архитектуры фреймворка разработки моделей машинного обучения для телемедицинских систем дистанционного мониторинга кардиологических данных. Разработанный метод построения моделей бинарной классификации позволяет учитывать спектры кардиологических сигналов как исходные данные для принятия решения по объекту по типу «есть исследуемое состояние»/«нет». Разработанный метод построения моделей порядковой классификации позволяет формировать итоговые модели принятия решения о состоянии объекта с тремя и более состояниями (например, «здоров»/ «предиабет»/ «диабет»), что расширяет возможности построения таких моделей в условиях ограниченного размера исходной выборки данных.

Предложенная архитектура фреймворка, содержащая разработанные методы, направлена на автоматизацию формирования программных модулей для телемедицинских систем дистанционного мониторинга кардиологических данных, что расширяет инструментарий разработки таких систем и обеспечивает повышение производительности труда разработчиков без повышения требований к их квалификации.

Практическая значимость полученных результатов состоит в наличии разработанной версии телемедицинской системы дистанционного мониторинга кардиологических данных (ЭС СД2), нацеленной на выявление признаков НУО по спектрам кардиологических данных с помощью разработанных моделей машинного обучения. Окончательное внедрение данной системы в рамках федерального проекта «Здоровье здоровых» по сравнению с применением существующих рекомендованных аналогов (лабораторных анализов крови), 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 сортировки участников на тех, у кого было обнаружено подозрение на предиабет или сахарный диабет 2 типа (и кому в итоге требуется диагностика по крови) на серии спектров кардиологических данных, и остальных.

Результаты работы внедрены организацией ООО «КардиоАрм» (акт внедрения от 05.11.2024) в процесс разработки телемедицинской системы дистанционного мониторинга кардиологических данных.

Новизна исследования и полученных результатов и выводов, сформулированных в диссертации.

В диссертации представлены новые методы построения моделей бинарной и порядковой классификации состояния пациента по спектрам кардиологических данных (на примере решения задачи выявления подозрения на сахарный диабет 2 типа и предиабет), полученных удаленным относительно медицинских учреждений и неинвазивным способом.

Научная новизна исследования заключается в следующем.

1. За основу разработанного метода построения модели бинарной классификации взят анализ явления автовозврата Ферми-Паста-Улама, позволяющий анализировать энергию гармоник спектра кардиологических данных пациентов с характерными заболеваниями.

2. Проведена работа по поиску закономерностей в спектрах кардиологических данных у пациента с признаками заболевания и без на примере нарушений углеводного обмена, по которой сделаны выводы о наличии характерных энергетических свойств гармоник спектра, свойственных пациентам с сахарным диабетом 2 типа и здоровым пациентам. Выявлена уникальная возможность неинвазивного и удаленного определения признаков нарушений углеводного обмена по спектрам кардиологических данных.
3. Разработаны математические модели бинарной классификации для выявления кардиопризнаков нарушений углеводного обмена с высокой чувствительностью (70% для пациентов с диабетом) и специфичностью (86% для здоровых пациентов) по спектрам ЭКГ.
4. За основу разработанного метода построения модели порядковой классификации взято имитационное моделирование выборки, позволяющее строить модели классификации по 3 и более классов на выборке ограниченного размера.
5. В ходе проведенных клинических исследований сделан вывод, что по одной ЭКГ определить подозрение на предиабет или сахарный диабет 2 типа невозможно. Для получения выводов о состоянии пациента необходима серия ЭКГ с определением, сколько должно быть ЭКГ в серии.
6. Разработана новая модель выявления подозрений на нарушения углеводного обмена (порядковой классификации) с высокой полнотой распознавания: 75% для класса “Здоров”, 77% для класса “Предиабет” и 100% для класса “Сахарный диабет 2 типа” при анализе серии из 11 ЭКГ.

Практическая новизна работы подтверждается получением двух патентов РФ на изобретение (№2728869 и №2751817), один из которых также защищен патентом Японии (№7562338).

Достоверность полученных результатов.

Достоверность полученных результатов подтверждается экспериментальной проверкой работоспособности системы ЭС СД2, которая реализована с помощью разработанных методов построения моделей классификации и методики формирования выборки данных, а также с применением фреймворка. Проверка проводилась в ходе двух клинических исследований, посвященных выявлению кардиопризнаков НУО на спектрах ЭКГ пациентов.

Рекомендации по использованию результатов.

Рекомендуется продолжить теоретическую и практическую работу по развитию фреймворка для создания модулей классификации по кардиологическим данным – например, путем добавления других видов моделей классификации на выбор или обеспечения возможности построения моделей на других представлениях кардиологических данных (например, фотоплетизмограмм).

Результаты работы могут быть внедрены в коммерческих и государственных организациях, занимающихся централизованным анализом кардиологических данных и разработкой телемедицинских систем удаленного мониторинга кардиологических данных, реабилитационных центрах.

Кроме того, результаты диссертации могут быть использованы в учебном процессе студентов ФГАОУ ВО Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский

Университет), ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России (Пироговский Университет), других медицинских университетов, осуществляющих подготовку по специальности «Кардиология», в ходе повышения квалификации врачей профильных специальностей (кардиология, эндокринология) в ФГБОУ ДПО РМАНПО Минздрава России и других организациях.

Оценка диссертационной работы

Основные результаты диссертации опубликованы в 5 работах автора, включая 3 статьи в журналах из перечня ВАК и 2 патента РФ на изобретение, а также представлялись на 7 конференциях.

Следует отметить ряд замечаний к тексту диссертации.

1. В работе недостаточно подробно раскрыта новизна предлагаемой методики формирования выборки спектров кардиологических данных. Имело смысл уточнить новизну путем выделения практической направленности на рассматриваемую автором предметную область.
2. В диссертационной работе не приведен пример применения разработанных соискателем методов на открытых выборках кардиологических данных для обоснования их полезности вне проведенных при участии автора клинических исследований.
3. В тексте диссертации выбор используемых моделей классификации обосновывается требованием к интерпретируемости, однако отсутствуют пояснения, каким образом предлагается осуществлять интерпретацию результаты работы модели бинарной классификации, и примеры интерпретации результатов.
4. В главе 1 представлен сравнительный анализ существующих моделей выявления сахарного диабета второго типа по параметрам variability сердечного ритма по публикациям лишь до 2019 года.

Также в работе имеется ряд ошибок редакционного характера. В частности, список литературы не приведен к единому стилю; имеются погрешности в оформлении рисунков, разделов и подразделов, выражения, приведенные как формулы (6) и (8), формулами не являются и т.д.

Вместе с тем, отмеченные недостатки не снижают в целом положительного впечатления от рецензируемой диссертационной работы, которая выполнена на достаточно высоком уровне и является законченным научным исследованием.

Заключение.

Диссертация Р.С. Новикова является законченной научно-квалификационной работой, в которой решена задача исследования и разработки методов машинного обучения с учителем для построения моделей классификации объектов по спектрам их кардиологических данных. Основные результаты диссертации в достаточной мере апробированы на профильных конференциях, опубликованы, в том числе в журналах из перечня ВАК, и корректно отражены в автореферате диссертации. Полученные результаты являются новыми, степень их обоснованности и достоверности является достаточной.

Таким образом, диссертационная работа «Методы машинного обучения по спектрам кардиологических данных для проблемно-ориентированных цифровых продуктов»

отвечает критериям Положения о присуждении ученых степеней, утвержденного Постановлением Правительства РФ от 24.09.2013 г. №842, а ее автор Р.С. Новиков заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.5 - математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Отзыв ведущей организации обсужден и принят на заседании секции Ученого совета Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» 27 марта 2025 г., протокол № 4. Присутствовало 14 человек, результаты голосования: «за»-12 чел., «против» - 2 чел., воздержались – 0 чел.

Главный научный сотрудник Отдела 63 «Методы и программные средства накопления и обработки данных» ФИЦ ИУ РАН, доктор технических наук, специальность 05.13.13 «Телекоммуникационные системы и компьютерные сети»

Будзко Владимир Игоревич

«17» марта 2025 г.

Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН)
Адрес: 119333, Москва, Вавилова, д.44, кор. 2
<http://www.frccsc.ru/>
Тел: + 7 (499) 135-62-60
E-mail: frccsc@frccsc.ru