

ТОМСКИЙ
ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ



На правах рукописи



Лызин Иван Александрович

**ОБРАБОТКА ВРЕМЕННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ
ДИАГНОСТИЧЕСКИХ ПРЕДИКТОРОВ ДЛЯ
АНАЛИЗА ТРАЕКТОРИЙ ЛЕЧЕНИЯ**

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

2.3.1 Системный анализ, управление и обработка информации,
статистика

Томск – 2025

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования "Национальный исследовательский Томский политехнический университет"

Научный руководитель:

Аксёнов Сергей Владимирович

кандидат технических наук, доцент Национальный исследовательский Томский политехнический университет

Официальные оппоненты:

Ходашинский Илья Александрович

доктор технических наук, профессор, ФГБОУ ВО "Томский государственный университет систем управления и радиоэлектроники", профессор кафедры компьютерных систем в управлении и проектировании

Часовских Наталия Юрьевна

доктор медицинских наук, доцент, ФГБОУ ВО "Сибирский государственный медицинский университет" Министерства здравоохранения Российской Федерации, заведующая кафедрой медицинской и биологической кибернетики

Защита состоится 25 июня 2025 г. в __ часов __ мин на заседании диссертационного совета ДС.ТПУ.15 Национального исследовательского Томского политехнического университета по адресу: Россия, 634034, г. Томск, ул. Советская, 84/3, ауд. 214.



С диссертацией можно ознакомиться в научно-технической библиотеке Томского политехнического университета и на сайте dis.tpu.ru при помощи QR-кода.

Автореферат разослан " __ " _____ 2025 г.

Ученый секретарь
диссертационного совета ДС.ТПУ.15
д.т.н.

A handwritten signature in blue ink, consisting of stylized, overlapping strokes.

А.Я. Пак

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Современная медицина активно развивается в направлении персонализированной медицины, где лечение подбирается индивидуально для каждого пациента на основе его клинико-лабораторных показателей, анамнеза болезни и доступных терапевтических комплексов, формирующих траекторию лечения. Траектории лечения – это последовательность событий и вмешательств, через которые проходит пациент в процессе получения медицинской помощи. Правильно выбранная траектория лечения позволяет повысить эффективность терапии, снизить риск побочных эффектов и нежелательных реакций. Таким образом, подбор траекторий лечения пациентов способствует обеспечению наиболее эффективному, безопасному и экономически обоснованному медицинскому обслуживанию, улучшая при этом качество жизни пациентов.

Исходя из этого, возникает необходимость в создании систем поддержки принятия врачебных решений, направленных на анализ траекторий лечения пациентов. Система поддержки принятия решений (СППР) – это интеллектуальная компьютерная программа, которая может проводить анализ, консультировать, ставить диагноз, прогнозировать процесс лечения и его результат. Целью СППР является улучшение оказания медицинской помощи путем усиления медицинских решений с помощью целенаправленных клинических знаний, данных о состоянии здоровья и другой информации о пациентах. Современные СППР, помимо выдачи результатов моделирования и рекомендаций, должны предоставлять возможность интерпретаций и объяснений их получения.

Использование СППР в клинической практике обеспечивает следующие преимущества. Во-первых, врач получает подсказки по вероятным диагностическим решениям, которые можно принять на основании собранных данных. Во-вторых, база данных, включенная в СППР, учитывает современные представления по проблеме, что позволяет ей реализовать обучающую функцию, подсказывая врачу не только необходимый объем клинического, лабораторного и инструментального обследования пациента, но также и алгоритмы принятия решений, обобщающие знания экспертов по данной проблеме. В-третьих, появляется возможность оперативной обработки возрастающего объема медицинских данных (в том числе динамических значений клинико-лабораторных показателей).

Целью диссертационной работы является повышение качества принятия врачебных решений при выборе траектории лечения пациента путём анализа разнородных диагностических предикторов и их временных последовательностей, получаемых во время пребывания пациента в стационаре с использованием методов интеллектуального анализа данных и машинного обучения.

Для достижения цели исследования необходимо решить следующие **задачи**:

1. Обзор и анализ современного состояния предметной области, методов и алгоритмов построения систем поддержки принятия врачебных решений и выбора траектории лечения;
2. Разработка алгоритмического решения для анализа и прогнозирования траектории лечения, обработки аномальных значений и построения кластеров эффективности терапевтических мероприятий;

3. Разработка программного обеспечения системы выбора траектории лечения с учётом диагностических предикторов и их временных последовательностей для пациентов, проходящих лечение в условиях стационара;
4. Тестирование и апробация предлагаемых методов по данным диагностических предикторов и их временных последовательностей для пациентов в стационаре на примере лечения рожистого заболевания, хронического пиелонефрита и ожирения.

Методы исследования. В работе использованы методы анализа данных и машинного обучения, методы непараметрической и параметрической статистики для предварительной обработки и анализа медицинских данных (в том числе, кластерный анализ, методы очистки данных и поиска аномальных наблюдений), методы разработки программного обеспечения, технологии разработки баз данных и баз знаний, методы визуализации многомерных данных. Исследования и разработка программного обеспечения производились с использованием языка программирования Python.

Научная новизна

1. Разработана интерактивная интеллектуальная система поддержки принятия врачебных решений, позволяющая формировать наилучшую траекторию лечения пациента, на основе анализа диагностических предикторов, в условиях стационара с учетом динамики лечения и особенностей протекания заболевания; анализ и прогноз траектории лечения осуществляется на основе аддитивных объяснений Шепли, заключающихся в оценке вклада каждого диагностического предиктора в предсказание исхода лечения.
2. Предложен и исследован метод анализа аномальных значений предикторов на основе ансамбля, полученного агрегированием предсказаний индивидуальных методов: одноклассового метода опорных векторов, изолированного леса, фактора локального выброса и эллиптической оболочки, что обеспечивает синергетический эффект более точного выявления аномальных значений по сравнению с индивидуальными методами.
3. Предложен и исследован метод выделения кластеров значимых диагностических признаков для определения и двумерной визуализации групп пациентов с наиболее эффективным комплексом лечения на основе минимизации дивергенции Кульбака-Лейблера распределений оригинальных значений диагностических предикторов и их векторных вложений в двумерном пространстве.

Практическая ценность работы подтверждается использованием полученных в ней результатов для решения практических задач и заключается в разработке технологии анализа данных для выбора траектории лечения пациентов, проходящих лечение в условиях стационара с последующим включением в основу системы поддержки принятия врачебных решений. Технология включает в себя:

- предварительную очистку и визуализацию многомерных данных;
- выявление значимых предикторов и их графическая интерпретация в двухмерном и трехмерном виде;

- проведение комплексной оценки состояния пациента на примере лечения пациентов с хроническим пиелонефритом, ожирением и рожистым воспалением.

Разработанная система может быть использована в работе медицинских учреждений для поддержки принятия врачебных решений с целью поиска подходящей траектории лечения пациента.

Апробация результатов работы. Основные положения и результаты диссертационного исследования докладывались на следующих научно-технических конференциях: Международный научно-технический конгресс "Интеллектуальные системы и информационные технологии – 2021" ("ИС & ИТ-2021", Таганрог, 2021 г.); Международная научно-практическая конференция студентов, аспирантов и молодых ученых "Молодёжь и современные информационные технологии" (Томск, 2018 г.); Международная научная конференция «Информационные технологии в науке, управлении, социальной сфере и медицине» (Томск, 2019 г.); Международная научная конференция "Информационные технологии в науке, управлении, социальной сфере и медицине" (Томск, 2018 г.); XXVI Международная научно-техническая конференция "Информационные системы и технологии – 2020" (Нижний Новгород, 2020 г.).

Достоверность полученных результатов обеспечивается использованием верифицированных данных клинических исследований, сравнением результатов, полученных предложенными методами с результатами известных методов.

Положения, выносимые на защиту:

1. Предложенная интерактивная интеллектуальная система поддержки принятия врачебных решений обеспечивает формирование наилучшей траектории лечения пациента в условиях стационара при снижении времени получения решения в среднем на 20-30 %.
2. Предложенный метод анализа аномальных значений предикторов на основе ансамбля, полученного агрегированием предсказаний индивидуальных методов, обеспечивает более точное выявление аномальных значений по сравнению с индивидуальными методами. В частности, выигрыш по внутригрупповой дисперсии очищенных данных составляет не менее 25 % на уровнях отсечения 0,05 и 0,1; а средний выигрыш по метрикам Жаккара / Дайса составляет от 7 до 30 % в зависимости от индивидуального метода.
3. Предложенный метод выделения кластеров значимых диагностических признаков позволяет определить группы эффективности комплекса лечения и наглядно продемонстрировать, насколько выбранный комплекс будет результативен для новых пациентов с учетом значений их диагностических предикторов.

Реализация и внедрение результатов работы. Результаты диссертационного исследования использованы при выполнении научно-исследовательских проектов: грант РФФИ № 19-37-90005 Аспиранты: "Система поддержки принятия врачебных решений в задачах персонализированной медицины (на примере лечения детей с ожирением)", 2019-2022 г., под руководством д.т.н. Берестневой О.Г.

Грант РФФИ №18-07-00543 А: "Интеллектуальная система поддержки принятия управленческих решений по инновационному развитию региональных научно-медицинских центров", 2018-2020 г., под руководством д.т.н. Берестневой О.Г.

Результаты диссертационной работы используются в Томском НИИКиФ ФФГБУ ФНУЦ МРиК ФМБА при первоначальной оценке диагностических предикторов пациентов; в учебном процессе ОИТ ИШИТР ТПУ.

Публикации: Основные результаты исследований отражены в 17 публикациях: 5 статей в научных журналах и изданиях, рекомендуемых ВАК, из которых 3 статьи проиндексированы в базе данных Scopus; 12 статей в сборниках трудов международных и российских конференций.

Структура и объем диссертации. Работа состоит из введения, четырех глав, заключения, списка литературы из 95 наименований и 5 приложений. Полный объем работы составляет 225 страниц, включая 115 рисунок и 69 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В введении обосновывается актуальность работы, формируется цель и задачи исследования, аргументируется научная новизна, обосновываются основные положения, выносимые на защиту.

В первой главе "Постановка задачи, состояние предметной области" исследованы виды, классификация и подходы к построению систем поддержки принятия врачебных решений. Выявлены проблемы построения подобного рода систем. Произведён анализ исследований в направлении подбора траекторий лечения пациентов, учитывающих персональные характеристики и анамнез заболевания. Проанализированы методы и алгоритмы, используемые для предсказания оптимальных траекторий лечения.

Анализ предметной области показал, что среди СПВР существует множество различных систем, однако, с учётом специфики заболеваний, остаётся необходимость в создании специализированных систем для решения конкретных задач. Кроме того, в большинстве случаев методы и алгоритмы, заложенные разработчиками в системы, не могут быть модифицированы.

В научной литературе, посвящённой медицинским информационным системам, выделяются два основных типа: электронная медицинская карта пациента (Electronic Medical Records, EMR) и системы поддержки принятия врачебных решений (Clinical Decision Support, CDS). В российских стандартах медицинские информационные системы классифицируются на пять типов, среди которых аналогом СПВР выступают "медико-технологические информационные системы", предназначенные для информационного обеспечения процессов диагностики, лечения, реабилитации и профилактики пациентов в лечебно-профилактических учреждениях. Функциональное назначение СПВР трактуется весьма широко, включая такие функции, как обеспечение своевременных напоминаний врачам, информирование о необходимых анализах и аллергических реакциях на лекарственные препараты.

Среди отечественных разработок можно отметить системы ЭКСНА, АССОД, ОТЭКС (г. Новосибирск). Внимания заслуживают инструментальный комплекс

ИМСЛОГ (Янковская А.Е., г. Томск), Универсальный Классификатор (Юдин В.Щ., г. Москва) и система Deep Data Diver (Дюк В.А. и Асеев М.Г., г. Санкт-Петербург).

Из зарубежных программ широко используются программы WIZWHY, Clementine, See5/C5.0 для поиска закономерностей в данных. Популярны статистические пакеты: SAS, SPSS, STATGRAPICS, STADIA, Statistica и другие, включающие элементы Data Mining. Активно используются нейронные сетевые системы, такие как OWL (HyperLogic), NeuroShell (Ward Systems Group). Можно выделить ряд экспертных системы в области медицины PathAI – анатомическая патология и диагностика рака; IBM Watson Health – анализ медицинских данных, диагностика заболеваний, выбор оптимальных вариантов лечения; DXplain – поддержка клинических решений; Germwatcher – диагностика инфекций.

На основе литературного обзора сделан вывод о том, что для решения поставленной задачи наиболее эффективной является реализация системы поддержки принятия врачебных решений выбора траектории лечения на основе методов интеллектуального анализа данных и машинного обучения.

Вторая глава "Алгоритмы подготовки данных для анализа и исследования траекторий лечения" описывает предлагаемый подход к определению оптимальной траектории лечения с использованием алгоритмов машинного обучения и получением системы объяснений, а также процедуры, которые позволяют привести данные к виду, необходимому для получения требуемого результата.

Траекторией лечения будем называть развёрнутую во времени последовательность *этапов лечения* E_i , представленную множеством $T = \{E_1, E_2, \dots, E_n\}$. На каждом i -м этапе лечения применяется конкретный j -й набор терапевтических мероприятий C_{ij} , называемый *комплексом*. Выбор комплекса на конкретном этапе лечения осуществляется с учётом значений диагностических предикторов и оценки возможных рисков для здоровья пациента. В лечебной практике такой выбор осуществляется врачом на основе протоколов лечения и опыта.

На выбранном этапе E_i лечения доступен комплекс C_{ij} , т.е. $E_i \leftarrow C_{ij}$, $C_{ij} \in C$, где C – это множество всех доступных в данной лечебной ситуации комплексов.

Введем обозначения: X – исходный набор диагностических предикторов; X_i^b – набор диагностических предикторов, которые известны перед i -м этапом лечения; $X_1^a, X_2^a, \dots, X_i^a$ – набор диагностических предикторов после каждого этапа лечения; Y_{ij} – преобразованное значение результата, интерпретированное как бинарный исход, где $Y_{ij} = 1$ если исход лечения в баллах выше среднего для всех пациентов, которые лечились на этом комплексе и чей результат в баллах свидетельствовал о положительной динамике (1 – высокая эффективность, 0 – низкая эффективность); M_{ij} – модель-классификатор, полученная для комплекса C_{ij} , оценивающая вероятность высокой эффективности.

В качестве одного из наборов данных, исследованных в работе, выступала база данных "Хронический пиелонефрит", содержащая информацию о диагностических предикторах пациентов до и после лечения на протяжении трёх этапов лечения, при этом для разных пациентов применялись разные наборы лечебных комплексов. Первый этап предполагал выбор комплексов из двух доступных C_{10}, C_{11} , где первый индекс 1 – номер этапа, второй индекс 0,1 – номер комплекса, второй

этап включал выбор из трех доступных комплексов C_{22} , C_{23} , C_{24} . Процесс выбора комплекса лечения C_{ij} начинался с первичного сбора диагностических предикторов пациентов X_i^b . По завершении этапа лечения производился повторный сбор предикторов X_i^a для оценки эффективности проведенной терапии и выбора комплекса для следующего этапа E_i .

Если после оценки результатов проведенного лечения у пациента отсутствовала положительная динамика, это указывало на то, что примененный комплекс лечения C_{ij} на данном этапе E_i был неэффективным. В случае наблюдения значительных улучшений комплекс лечения C_{ij} считался эффективным. На рисунке 1 показан пример формирования траектории лечения, путем выбора комплексов для набора данных "Хронический пиелонефрит".

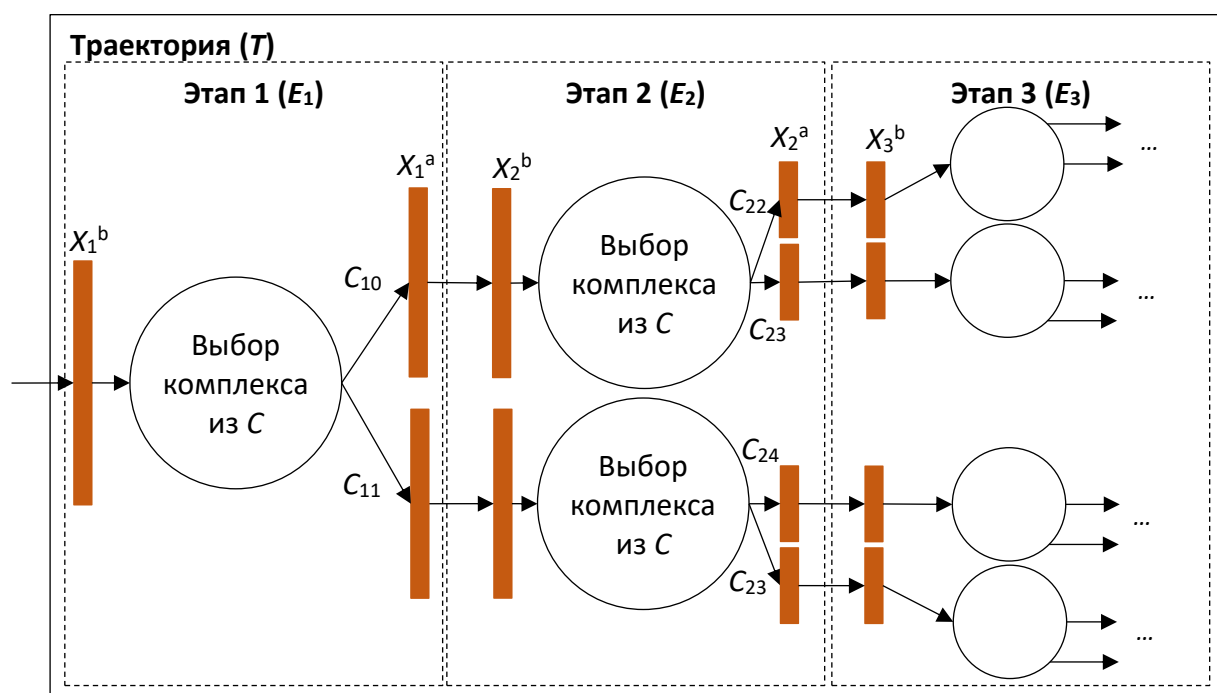


Рисунок 1 – Пример оцениваемых траекторий лечения пациентов с диагнозом хронический пиелонефрит в Томском НИИ курортологии и физиотерапии, где X_i^b , X_i^a набор диагностических предикторов до и после лечения; C – доступное множество комплексов C_{ij} на этапе лечения E_i

В диссертационной работе предложено использовать аддитивные объяснения Шепли для решения задачи адаптивного управления траекторией лечения T . Для этого предлагается выполнять оценку качества терапевтических мероприятий C_{ij} после каждого этапа лечения E_i путем активации моделей машинного обучения M_{ij} , настроенных для дифференциации эффективности лечения на значениях диагностических предикторов, полученных до и после проведенных этапов лечения E_i . Каждая обученная модель M_{ij} позволяет получить аддитивные объяснения Шепли $S_{ij} = \{\phi_f(x)_1, \phi_f(x)_2, \dots, \phi_f(x)_n\}$, оценивающие влияние каждого диагностического предиктора на результативность терапевтических процедур анализируемого этапа E_i .

В основе аддитивных объяснений Шепли (Shapley additive explanations) лежит агрегированная оценка разницы в выходе предсказываемой целевой перемен-

ной для случаев, когда значение анализируемого входного признака известно и когда на входе вместо реального значения используется его матожидание для разных наборов исследуемых зависимых переменных:

$$\varphi_f(x) = \sum_{F^S \subseteq N \setminus \{f\}} (1/n!) |F^S|! (n - |F^S| - 1)! (y^p(F^S \cup \{f\}) - y^p(F^S)), \quad (1)$$

где $\varphi_f(x)$ – значение Шепли для признака f пациента с диагностическими предикторами x ; n – количество признаков в выборке X , используемой для обучения модели M_{ij} ; F^S – подмножество признаков из X , не включающее f ; $|F^S|$ – количество признаков в подмножестве F^S ; $y(F^S)$ – выход модели при использовании только признаков из подмножества F^S . Положительные значения $\varphi_f(x)$ свидетельствуют о том, что реальные значения признака f выдают значение прогноза выше, чем для случая, когда на вход подаётся матожидание признака f . Более высокие значения Шепли по амплитуде оказывают большее влияние на целевую переменную. Отбор значений $\varphi_f(x)$ с наибольшей амплитудой позволяет выявить наиболее существенные предикторы, а знак $\varphi_f(x)$ отражает характер влияния (отрицательный или положительный) на прогнозное значение.

Таким образом, после расчета значений Шепли для модели M_{ij} выявляются диагностические предикторы в наибольшей степени повлиявшие на выход модели. При анализе признаков нового поступившего пациента с учётом значения $\varphi_f(x)$ можно выявить, насколько значения признаков, относящихся к существенным, близки к тем, которые имелись у пациентов после эффективного лечения для комплекса C_{ij} . Задача нахождения оптимальной траектории заключается в выборе такого комплекса C_{ij} на этапе лечения E_i , при котором агрегированное влияние всех существенных значимых признаков, указывающих на высокую вероятность эффективности, будет максимальным.

Метод анализа и прогнозирования траектории реализуется в два этапа.

Этап 1 "Анализ траекторий лечения", см. рисунок 2 содержит несколько шагов.

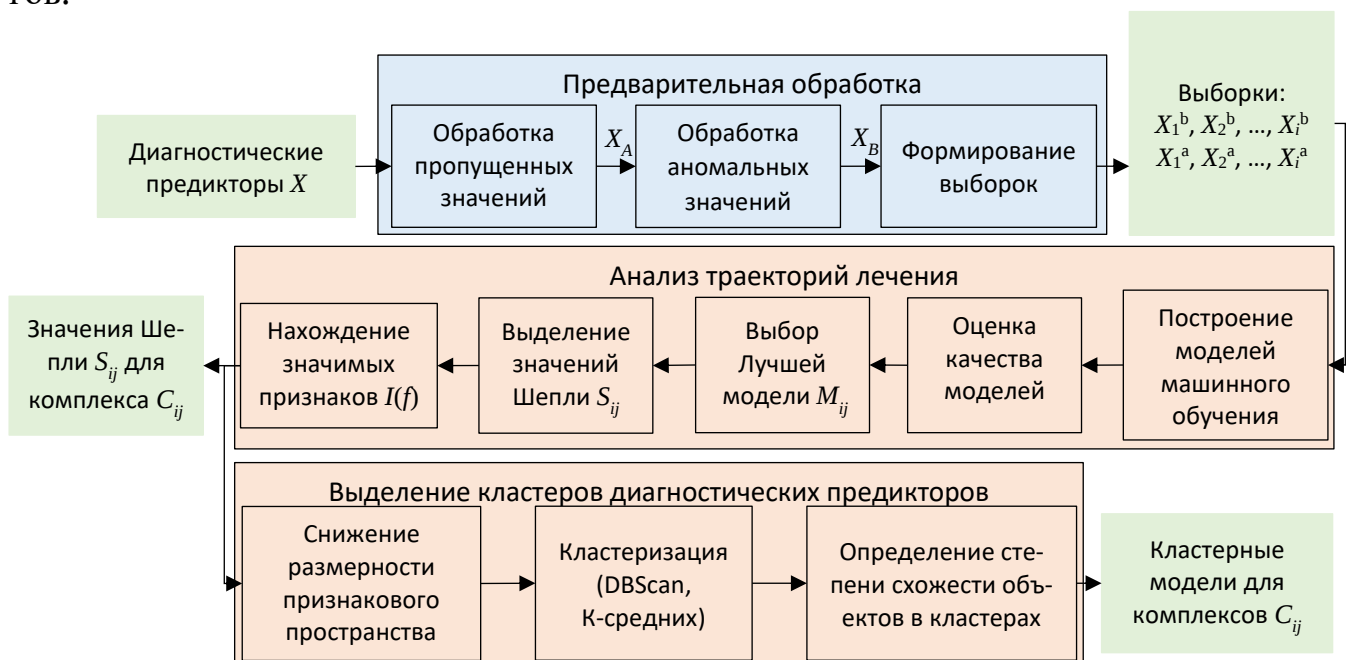


Рисунок 2 – Процесс анализа траекторий лечения

Шаг 1.1. Обработка пропущенных значений. Исходный набор, содержащий значения диагностических предикторов для пациентов X , оценивается на наличие пропущенных значений; если пропущено более 90% значений, то такой предиктор исключается из дальнейшего рассмотрения; X_A – набор данных, полученный после исключения предикторов с большим количеством пропущенных значений.

Шаг 1.2. Обработка аномальных значений. Из записей набора X_A удаляются векторы диагностических предикторов, классифицируемые как выбросы с использованием специально разработанного ансамбля четырех методов; X_B – набор данных после удаления аномальных предикторов.

Шаг 1.3. Формирование выборок. Формируются выборки $\{X_i^b\}$, $\{X_i^b, X_{i-1}^a, X_{i-1}^b\}$, $\{\dots\}$, отобранные из X_B для разных комплексов C_{ij} , включенных в этап лечения E_i .

Шаг 1.4. Построение "моделей машинного обучения. Для каждой из сформированных выборок строятся модели-классификаторы M_{ij} эффективности лечения для каждого комплекса C_{ij} . В качестве целевой переменной используются метки Y_{ij} . Для нахождения лучших моделей осуществляется подбор их гиперпараметров. В качестве моделей могут выступать модели бинарной классификации, полученные с помощью любого метода, позволяющего вычислять вероятность положительного класса – эффективность комплекса.

Шаг 1.5. Оценка качества модели. Построенные модели-классификаторы для каждого комплекса C_{ij} этапа лечения E_i оцениваются с использованием метрик Recall (чувствительность), Specificity (специфичность) и F1; проводится ROC-анализ, показывающий изменение чувствительности и ложноположительного уровня модели при разных порогах.

Шаг 1.6. Выбор лучшей модели. Для каждого комплекса C_{ij} модели M_{ij} с оценкой $F1 < 0,7$ отбраковываются, как обладающие низкой предиктивной способностью; среди оставшихся выделяется модель, обладающая максимальным значением метрики F1.

Шаг 1.7. Выделение значений Шепли. Найденная лучшая модель M_{ij} каждого комплекса C_{ij} используется для вычисления модели аддитивных объяснений Шепли S_{ij} , оценивающей вклад каждого диагностического предиктора в значение эффективности комплекса для конкретного пациента.

Шаг 1.8. Нахождение значимых признаков. На основе значений Шепли S_{ij} , полученных для лучшей модели M_{ij} каждого комплекса C_{ij} этапа лечения E_i , определяется влияние значимых признаков I . Влияние признака I определяется как среднее из значений $\varphi_f(x)_1$ по каждому признаку f :

$$I(f) = (1/n) \sum_x |\varphi_f(x)|. \quad (2)$$

Если для каждого признака выполняется $I > Z$, где Z – заданный экспертом порог, то признак I считается информативным.

Шаг 1.9. Выделение кластеров диагностических предикторов. К набору полученных информативных признаков применяется алгоритм кластеризации диагностических предикторов: значения признаков каждого комплекса C_{ij} проецируются на двумерное пространство с использованием алгоритма T-SNE и полученные

двумерные векторные вложения разбиваются на группы с помощью алгоритма кластеризации К-средних для получения кластеров пациентов с близкой эффективностью лечения.

Этап 2 "Определение оптимального комплекса в траектории", см. рисунок 3, включает следующие шаги.

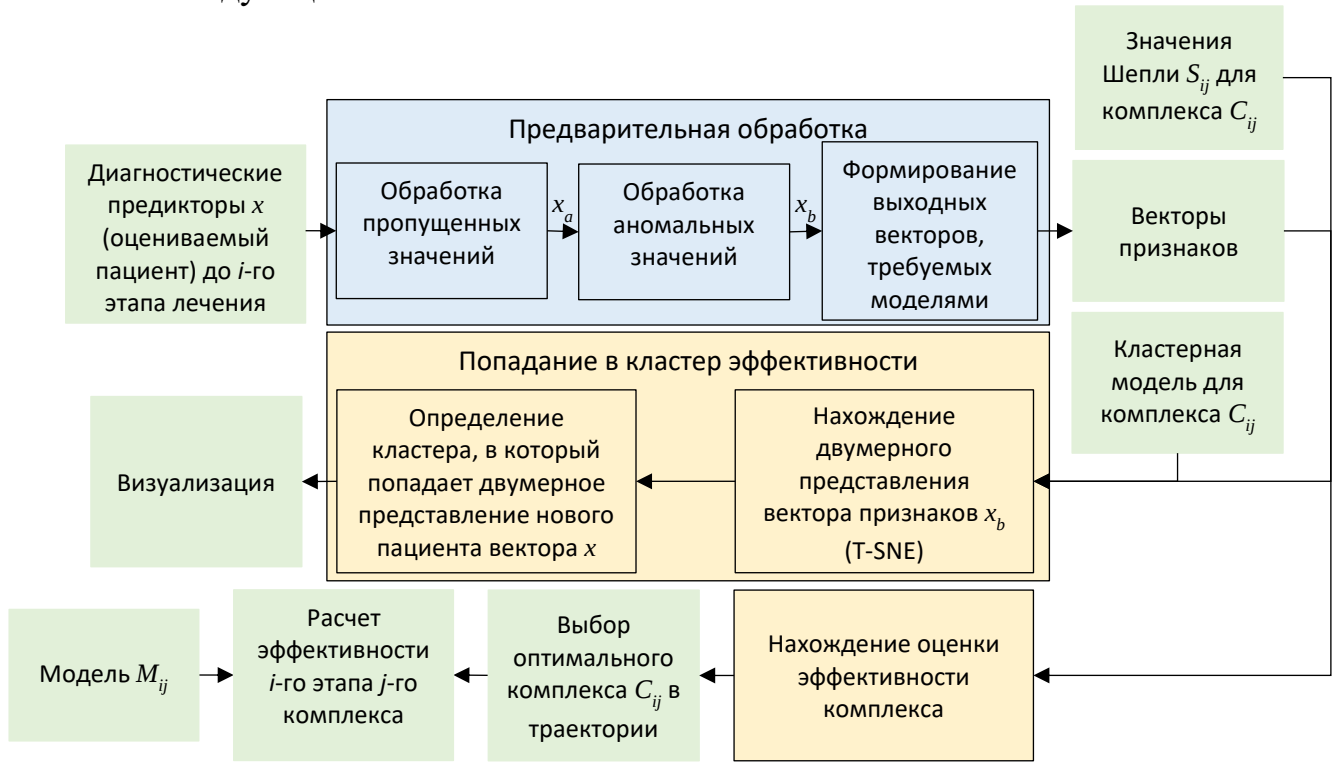


Рисунок 3 – Определение оптимального комплекса в траектории

Шаг 2.1. Обработка пропущенных значений. Для нового пациента обработка пропущенных значений осуществляется в соответствии с шагом 1.1. В случае отсутствия пропущенных значений $x_a \leftarrow x$ и переход к шагу 2.2, иначе выдается сообщение "Для принятия решения недостаточно данных".

Шаг 2.2. Обработка аномальных значений. Построенная на шаге 1.2 решающая граница ансамбля, определяющего аномальные значения, используется для отбора выбросов. Если вектор значений диагностических предикторов x_a попадает в область, ограниченную этой границей, то модели, оценивающие эффективность комплексов, считаются подходящими для анализа нового пациента; если нет, то на выходе блока $x_b = x_a$ выдается сообщение "Входящий вектор содержит аномальные значения".

Шаг 2.3. Формирование выходных векторов, требуемых моделью. Входные данные x_b приводятся к вектору x_c в соответствии с $I(f)$, определенными на шаге 1.8.

Шаг 2.4. Нахождение оценки эффективности комплекса. Рассчитывая сумму абсолютных значений Шепли с учетом значений диагностических предикторов, относящихся к информативным признакам, для конкретного комплекса лечения C_{ij} получаем величину $R(C_{ij})$, показывающую оценку эффективности комплекса C_{ij} для анализируемого пациента:

$$R(C_{ij}) = \sum_{f \in x_c} (V \cdot I(f)), \quad (3)$$

где f – информативный признак в наборе данных; x – вектор диагностических предикторов f ; V – знак влияния для признака f (+1 если с ростом значения признака увеличивается вероятность успеха лечения, –1 если с ростом значения признака вероятность успеха уменьшается). Индикаторная переменная W определяется по формуле:

$$W = w \frac{\sum \varphi(x)}{n}, \forall x: x_f > m(f), \quad (4)$$

где x_f – значение признака f для пациента с вектором диагностических предикторов x , $m(f)$ – ожидаемое значение признака.

Шаг 2.5. Выбор оптимального комплекса. Если $R(C_{ij}) > 0$, то комплекс C_{ij} является эффективным для пациента с вектором диагностических предикторов x . Если таких комплексов несколько, то предпочтение отдается комплексу, имеющему наибольшее значение $R(C_{ij})$ для конкретного пациента, т.е.

$$R^* = \max(R(C_{ij})), R(C_{ij}) > 0.$$

Шаг 2.6. Нахождение двумерного представления вектора признаков. К вектору диагностических предикторов нового пациента применяется алгоритм T-SNE для нахождения его двумерного представления.

Шаг 2.7. Определение кластера, в который попадает двумерное представление пациента. Двумерное представление нового пациента проецируется на кластерные решения, полученные на шаге 1.9.

Выбор ансамбля методов для применения на шагах 1.2. и 2.2 осуществлялся следующим образом. Поскольку медицинские базы данных зачастую содержат пропущенные и аномальные значения, что затрудняет настройку моделей машинного обучения, в работе предложен метод анализа аномальных значений предикторов на основе ансамбля методов. Для этого были проведены исследования различных методов обнаружения аномалий и для ансамбля осуществлен выбор четырех методов: изолирующий лес (Isolation Forest, IF), фактор локального выброса (LOF), эллиптическая оболочка (Elliptic Envelope, EE) и метод опорных векторов (OneClassSVM, OC).

Метод IF построен на предположении, что аномальные точки легче отделяются от нормальных данных. LOF считает выбросами точки, которые имеют существенно меньшую плотность, чем их соседи. EE построен на предположении, что нормальные точки данных генерируются из гауссовского распределения, и стремится построить эллипсоид минимального объема, который охватывает нормальные данные. OC предполагает нахождение разделяющей границы алгоритмом SVM с использованием нелинейного ядра (например, радиальной базисной функции), при этом вектора, попадающие внутрь ограниченной области, считаются нормальными, а остальные – аномальными.

Итоговый результат работы выбранных четырех алгоритмов определяется посредством мажоритарного голосования.

Была исследована эффективность работы предложенного ансамбля посредством вычисления суммы квадратов ошибок (SSE) для него и каждого из четырех методов в отдельности по формуле:

$$SSE = \sum_{l=1}^{NK} \sum_{\forall x_i \in K_l} \|x_i - \mu_l\|^2, \quad (5)$$

где NK – количество кластеров в данных, x_i – вектор диагностических предикторов пациента i , попавший в кластер K_l с центроидом μ_l .

Выделение кластеров диагностических предикторов на шаге 1.9. осуществлялось следующим образом. Поскольку исходные данные содержат большое количество признаков, возникают проблемы с их визуализацией, оценкой и построением связей. Для снятия этих ограничений в диссертационной работе предложено осуществлять переход от многомерного пространства признаков к двумерному пространству посредством использования алгоритма T-SNE (T-distributed Stochastic Neighbor Embedding). Этот алгоритм минимизирует расхождение Кульбака-Лейблера между условными вероятностями p_{ij} и q_{ij} :

$$KL(P \| Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}, \quad (9)$$

где p_{ij} – условная вероятность сходства между векторами диагностических предикторов x_i и x_j в исходном многомерном пространстве; q_{ij} – условная вероятность сходства между точками z_i и z_j в двумерном пространстве.

Из нового признакового пространства кластеры выделялись с использованием алгоритмов DBScan и K-средних. Определялись особенности кластеров и выделялись характеристики, объединяющие пациентов в выделенных кластерах, см. рисунок 2.

Для нахождения кластерного решения с учетом максимизации расстояния между кластерами и минимизации внутрикластерной дисперсии использовались относительные метрики: коэффициент силуэта, индекс Калински-Харабаша, индекс Дэвиса-Булдина. Для оценивания однородности кластеров относительно меток классов (объектов, которые входят в кластеры) использовались внутренние метрики: индекс Рэнда, индекс однородности, полноты и V-меры.

Рассмотренные индексы позволили оценить однородность сформированных кластеров с учётом групп пациентов, объединённых по результатам результативности терапии, т.е. определили соответствие кластеров и групп пациентов со схожей эффективностью лечения.

Третья глава "Программное обеспечение анализа траекторий лечения пациентов" содержит описание структуры системы поддержки принятия врачебных решений, включающую базу данных PatientDataBase и следующие блоки: предварительной обработки исходных данных; визуализации; определения траектории; построения ML моделей; формирования отчетности.

Для эффективного хранения и обработки результатов обследования пациентов была разработана база данных PatientDataBase. Концептуальная схема базы данных представлена на рисунке 6.

PatientDataBase представлена четырьмя основными таблицами для хранения результатов обследования пациентов Patients, Laboratory_research, Analyzes и Research_results. В качестве исходных данных были использованы три набора (1-3),

сформированных из базы данных PatientDataBase, и один (4) взят из открытого источника, см. таблицу 1.

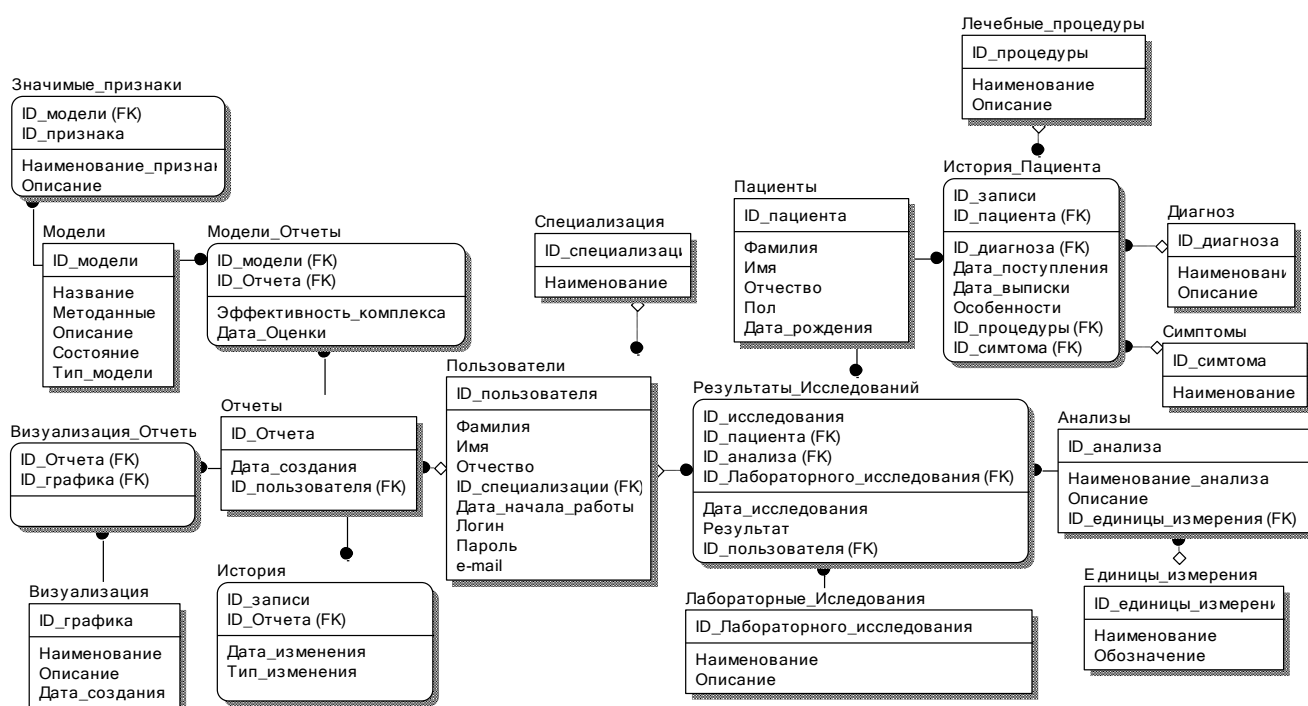


Рисунок 6 – Концептуальная схема разработанной базы данных PatientDataBase

Таблица 1 – Используемые наборы данных

№	Набор данных	Количество признаков	Количество экземпляров в наборе данных
1	Хронический пиелонефрит	480	230
2	Рожистые воспаления	44	114
3	Ожирение	162	345
4	Computing in Cardiology Challenge 2012	258	4010

Набор данных "Хронический пиелонефрит" содержит результаты клинко-лабораторных исследований детей от 5 до 17 лет с диагнозом "необструктивный хронический пиелонефрит", стадия частичной клинко-лабораторной ремиссии, проведенных на базе детского отделения ФГБУН "ТНИИКиФ ФМБА" с 2014 по 2019 год. Процесс лечения пациентов включал несколько этапов с фиксированием диагностических предикторов на каждом этапе до и после курса лечения. В таблице 2 представлены этапы лечения. Пациентам на этапах лечения, включенных в траекторию, назначались разные комплексы. Например, комплексы лечения для первого этапа подразделялись на лечение по стандарту с применением минеральной воды и лечение по стандарту без применения минеральной воды.

Таблица 2 – Динамика измерения показателя "Гемоглобин, гл"

Пациенты	C_{ij}	E_1		C_{ij}	E_2		C_{ij}	E_3	
		д/л	п/л		д/л	п/л		д/л	п/л
Пациент 1	C_{10}	137	130	C_{22}	137	129	C_{33}	144	132
Пациент 2	C_{10}	136	130	C_{23}	149	132	C_{32}	141	139
Пациент 3	C_{10}	137	140	C_{24}	144	138	C_{34}	150	142
Пациент 4	C_{11}	134	135	C_{24}	139	136	C_{34}	135	133
Пациент 4	C_{11}	148	149	C_{22}	152	156	C_{33}	157	146

Набор данных "Рожистые воспаления" представляет собой результаты исследований пациентов, проведенных СибГМУ в 2016-2017 г. Содержит данные температуры тела при поступлении, а также показатели общего и биохимического анализов крови.

Набор данных "Ожирение" получен в рамках сотрудничества с НИИ курортологии и физиотерапии г. Томска. Состояние каждого пациента оценивалось при поступлении на лечение и по окончании курса реабилитации. Пациенты – дети и подростки в возрасте от 6 до 17 лет, с проблемами веса. Группы характеризуют пациентов, проходящих лечение по разным траекториям.

Четвертая глава "Тестирование методов и программной реализации анализа траекторий лечения" содержит результаты апробации предложенных методов на экспериментальных данных. Для демонстрации работы предложенных подходов был использован набор данных "Хронический пиелонефрит".

Исследование эффективности подхода обнаружения аномалий в исходных данных с использованием методов ансамблирования по сравнению с индивидуальными алгоритмами осуществлялось путём выполнения серии тестов, оценивающих сумму квадратов ошибок SSE кластерного решения очищенных данных алгоритмом K-средних. Низкие значения SSE указывают на то, что кластеры получаются более компактными, а очищенные данные содержат меньше векторов, удаленных от центроидов и других объектов выборки. Для испытаний взяты признаки из наборов данных, выполнен поиск аномалий каждым из этих инструментов, а также выполнено сравнение полученных решений.

На рисунке 7а показаны результаты обнаружения аномальных значений индивидуальными алгоритмами (EE, OC, IF, LOF) для следующих признаков: лимфоциты до лечения, %; лимфоциты после лечения, %. На рисунке 7б представлен результат работы ансамбля поиска аномалий.

С помощью коэффициента Жаккара (intersection over union, IoU) J и коэффициента Дайса D произведено попарное сравнение результатов поиска аномалий всех исследуемых алгоритмов:

$$J = \frac{|A \cap B|}{|A \cup B|}, \quad D = \frac{2 \cdot |A \cap B|}{|A| + |B|}, \quad (8)$$

где: A и B множество аномальных значений, обнаруженных алгоритмом A и алгоритмом B соответственно; $A \cap B$ – количество элементов, отмеченных как аномалия алгоритмом A и алгоритмом B одновременно; $A \cup B$ – количество всех элементов, отмеченных как аномалия алгоритмом A и алгоритмом B . Коэффициент Дайса обладает большей чувствительностью к частичному совпадению в результатах обнаружения аномалий, что ценно при анализе небольших объемов данных. В таблице 3 представлено сравнение работы алгоритмов поиска аномалий путём расчета метрик J и D .

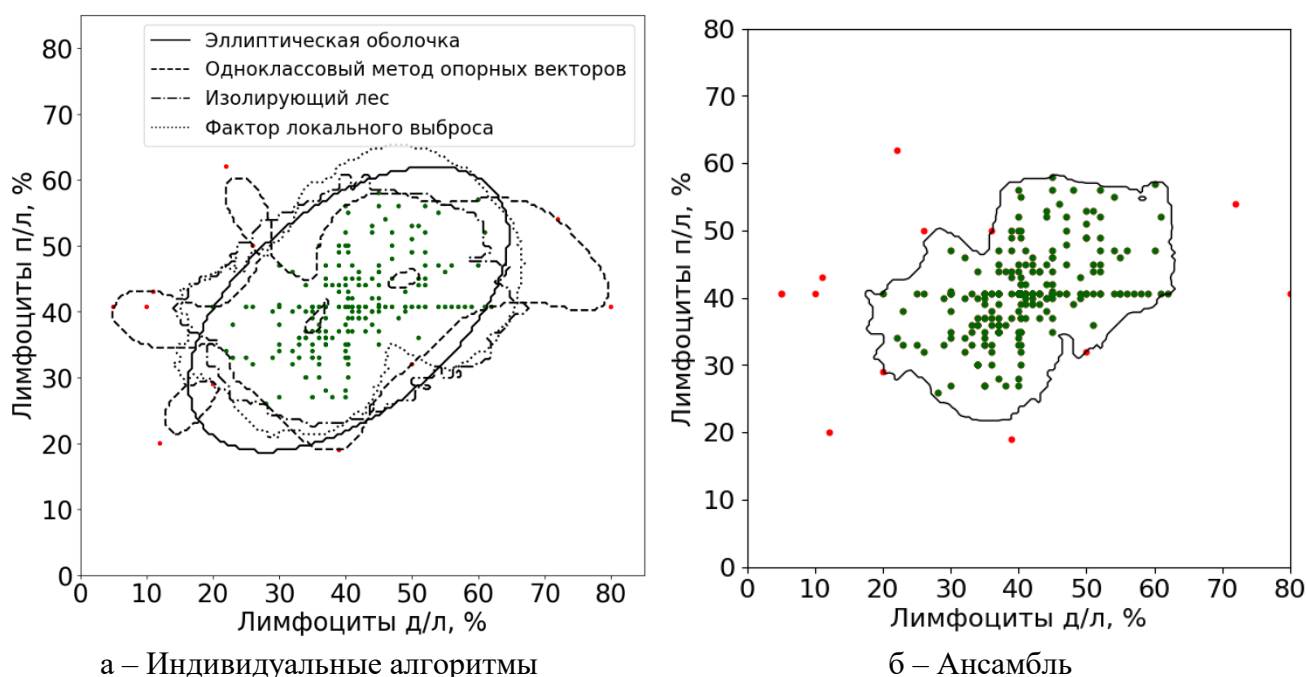


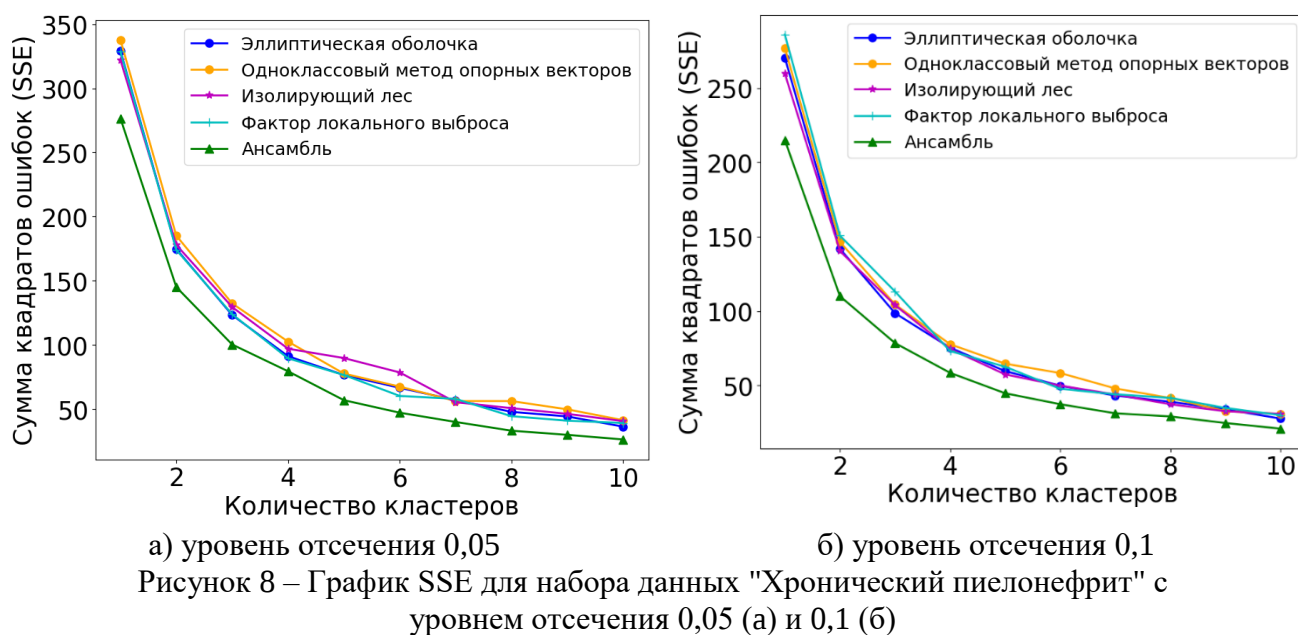
Рисунок 7 – Результат работы алгоритмов поиска аномалий (а – результат работы алгоритмов поиска аномалий; б – результат работы ансамбля поиска аномалий)

Таблица 3 – Значения мер Жаккара (J) и Дайса (D) алгоритмов обнаружения аномальных значений и их ансамбля в наборе данных "Хронический пиелонефрит"

	Уровень отсечения	OC		OC_SGD		IF		LOF		Ансамбль	
		J	D	J	D	J	D	J	D	J	D
EE	0,05	0,31	0,47	0,07	0,13	0,5	0,67	0,67	0,8	0,71	0,83
	0,1	0,62	0,77	0,23	0,38	0,58	0,78	0,51	0,68	0,65	0,87
OC	0,05			0,08	0,15	0,3	0,45	0,26	0,41	0,31	0,47
	0,1			0,22	0,36	0,56	0,77	0,55	0,71	0,63	0,81
OC_SGD	0,05					0,07	0,13	0,07	0,13	0,13	0,23
	0,1					0,34	0,47	0,24	0,39	0,37	0,54
IF	0,05							0,6	0,75	0,66	0,8
	0,1							0,57	0,73	0,75	0,86

Из данных таблицы видно, что разные алгоритмы в качестве аномальных данных отбирали разные объекты из выборок, видно также, что предложенный подход, основанный на ансамблировании, позволяет более точно оценивать аномалии по сравнению с индивидуальными методами.

Для того, чтобы показать эффективность алгоритма ансамблирования при разной доле аномалий, в работе исследовалось влияние уровня отсечения на величину SSE. В экспериментах по оценке компактности кластеров каждым из исследуемых алгоритмов были приняты уровни отсечения 0,05 и 0,1, соответствующие долям аномалий 5% и 10% соответственно. Графическая интерпретация результативности подходов для признаков "лимфоциты д/л, %" и "лимфоциты п/л, %" для уровней отсечения 0,05 и 0,1 представлена на рисунке 8.



Из рисунка 8 видно, что индивидуальные алгоритмы показывают схожие результаты и более высокие значения SSE при различных уровнях отсечения. Ансамбль алгоритмов демонстрирует наименьшее значение SSE по сравнению с каждым индивидуальным алгоритмом.

В качестве иллюстрации подхода определения оптимального комплекса в траектории представлен расчет показателя эффективности траектории для одного из пациентов. Значения Шепли оценки вклада признаков выбора схемы лечения, диагностические предикторы выбранного пациента и матожидание, используемые в формуле расчета для второго этапа лечения представлены в таблице 4.

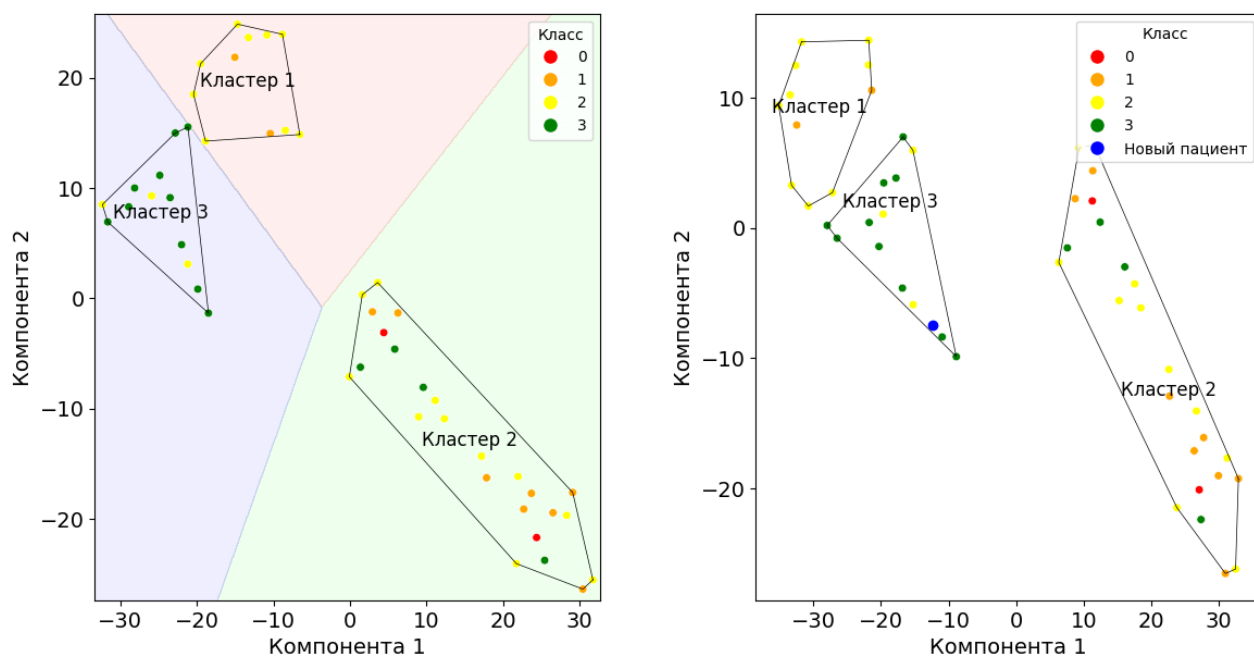
Таблица 4 – значения Шепли и диагностические предикторы

	Абсолютные значения важности для траектории вода	Значение предикторов для пациента	Матожидание признака
холестерин д/л, ммоль/л	0,017	5,2	4,31
общий диурез д/л, мл	0,01897	1430	1162,2
гемоглобин д/л, г/л	0,0192	117	132,56
креатинин д/л, мкмоль/л	0,0269	62,9	55,77
калий д/л, ммоль/л	0,0275	4,49	3,95
общий белок д/л, г/л	0,0295	70	70,26
СОЭ д/л, мм/ч	0,0365	18	11,72
лейкоциты в д/л, $10^9/л$	0,0547	4,9	9,63
мочевина д/л, ммоль/л	0,0562	6	4,66
эритроциты д/л, $10^{12}/л$	0,0864	3,93	4,67

Красным выделены предикторы, при увеличении которых эффективность комплекса возрастает. Согласно методу определения оптимального комплекса в траектории, оценка эффективности комплекса №1 второго этапа для конкретного пациента составила 0,214. Посчитанные аналогичным образом значения для двух других комплексов составили 0,192 для комплекса №2 и 0,150 для комплекса №3. Поскольку комплекс №1 демонстрирует наибольшее положительное значение, он считается оптимальным для применения на данном этапе лечения.

Значимые диагностические предикторы могут использоваться для визуализации групп пациентов на двумерной плоскости, для которых тот или иной комплекс лечения является эффективным. Для этого такие значения обрабатываются предложенным алгоритмом выделения кластеров значимых диагностических признаков, основанном на применении метода T-SNE.

Использование разных цветов, ассоциированных с определенной меткой, характеризующей эффективность лечения, позволяет обнаружить области пространства, где находятся векторы, связанные с пациентами, у которых этот комплекс лечения был эффективен, а также группы пациентов с низкой эффективностью выбранного комплекса. На рисунке 9 представлена визуализация векторных вложений. Области, выделенные различными цветами, отражают решения, полученные в процессе кластеризации алгоритмом K-средних, которые разделяют пациентов на группы с различной эффективностью лечения. Осям x и y соответствуют компоненты 1 и 2 векторов признаков, полученные с использованием алгоритма T-SNE.



а) кластерные решения для комплекса второго этапа с применением воды

б) отображение нового пациента на кластерные решения

Рисунок 9 – Векторные вложения и кластеризация алгоритмом K-средних для набора данных "Хронический пиелонефрит"; метки соответствуют исходу лечения пациента: 0 – отрицательный результат, 1 – незначительное улучшение, 2 – улучшение, 3 – значительное улучшение

На основании представленных результатов была получена лучшая кластерная структура для второго этапа комплекса с применением воды с использованием информативных признаков из набора данных "Хронический пиелонефрит". Значения относительных и внутренних метрик для полученной кластерной структуры представлены в таблице 5. Относительные метрики, показывают хорошую отделимость и компактность кластеров. Внутренние метрики показывают согласованность меток кластеров и исходов лечения.

Таблица 5 – Относительные и внутренние метрики кластерного решения при количестве кластеров = 3 для набора данных "Хронический пиелонефрит" второй этап лечения комплекс с применением воды

Относительные метрики	индекс Дэвиса-Булдина	0,92
	индекс Калински-Харабаша	218,35
	коэффициент силуэта	0,66
Внутренние метрики	индекс Рэнда	0,58
	индекс взаимной информации	0,055
	индекс однородности	0,11
	индекс полноты	0,12
	V-мера	0,12

Выделенные кластеры позволили сгруппировать пациентов по эффективности терапии для анализируемого комплекса. Кластер 2 включает пациентов с различными исходами, в том числе тех, для кого лечение оказалось неэффективным. Кластер 3 преимущественно объединяет пациентов с наилучшими результатами лечения, тогда как кластер 1 включает пациентов с умеренными улучшениями здоровья. Таким образом, попадание нового пациента в кластер 3 указывает на высокую вероятность положительного эффекта от предложенного лечения.

Векторы диагностических предикторов новых пациентов, составляющих контрольную группу, проецировались на полученное кластерное пространство. На основе этой проекции рассчитывался процент принадлежности новых пациентов к каждому из кластеров. Результат анализа распределения новых пациентов по кластерам представлен в таблице 6.

Таблица 6 – Распределение новых пациентов по кластерному решению

Метка класса	Кластер 1				Кластер 2				Кластер 3			
	0	1	2	3	0	1	2	3	0	1	2	3
Количество новых пациентов, попавших в кластер	1	2	7	1	1	1	2	6	0	0	3	6
Средний процент состава кластеров	0	34,14	64,41	1,45	9,78	17,55	38,28	33,39	0	10,42	43,06	46,33
Процент новых пациентов	9,09	18,18	63,64	9,09	10	10	20	60	0	0	33,33	66,67

Кластер 3 характеризуется пациентами с улучшениями (33 %) и существенными улучшениями (67 %) состояния здоровья. Что соответствует структуре сформированных кластеров. Таким образом, попадание новых пациентов в кластер 3 может свидетельствовать о высокой эффективности применяемого комплекса.

В заключении сформулированы основные результаты, полученные в ходе исследования.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Предложена система поддержки принятия врачебных решений, позволяющая формировать траекторию лечения пациента в условиях стационара с учетом динамики лечения и особенностей протекания заболевания; анализ и прогноз траектории лечения осуществляется на основе аддитивных объяснений Шепли, заключающихся в оценке вклада каждого диагностического предиктора в предсказание исхода лечения. Предложенная система поддержки принятия врачебных решений обеспечивает формирование наилучшей траектории лечения пациента в условиях стационара при снижении времени получения решения в среднем на 20-30 %.
2. Предложен и исследован метод анализа аномальных значений предикторов на основе ансамбля, полученного агрегированием предсказаний индивидуальных методов: одноклассового метода опорных векторов, изолирующего леса, фактора локального выброса и эллиптической оболочки, что обеспечивает синергетический эффект более точного выявления аномальных значений по сравнению с индивидуальными методами.
3. Предложенный метод анализа аномальных значений предикторов на основе ансамбля, полученного агрегированием предсказаний индивидуальных методов, обеспечивает более точное выявление аномальных значений по сравнению с индивидуальными методами. В частности, выигрыш по внутригрупповой дисперсии очищенных данных составляет не менее 25 % на уровнях отсечения 0,05 и 0,1; а средний выигрыш по метрикам Жаккара / Дайса составляет от 7 до 30 % в зависимости от индивидуального метода.
4. Предложен и исследован метод выделения кластеров значимых диагностических признаков для определения и двумерной визуализации групп пациентов с наиболее эффективным комплексом лечения на основе минимизации дивергенции Кульбака-Лейблера распределений оригинальных значений диагностических предикторов и их векторных вложений в двумерном пространстве. Метод позволяет определить группы эффективности комплекса лечения и показать, насколько выбранный комплекс будет результативен для новых пациентов по значениям их диагностических предикторов.
5. Результаты диссертационной работы используются в Томском НИИКиФ ФФГБУ ФНУЦ МРиК ФМБА при первоначальной оценке диагностических предикторов пациентов; в учебном процессе отделения информационных технологий Инженерной школы информационных технологий и робототехники ТПУ.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в изданиях, рекомендованных ВАК:

1. Лызин, И.А. Использование компьютерной визуализации в анализе кривых дыхания / О.Г. Берестнева, И.А. Осадчая, И.А. Лызин // Сибирский журнал клинической и экспериментальной медицины. – 2020. – Т. 35 – №. 4. – С. 57-64

(Scopus).

2. Лызин, И.А. Тематическое моделирование в контексте медицинских текстов / С.А. Землянский, С.В. Аксенов, И.А. Лызин, О.Г. Берестнева // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2021. – Т. 24. – № 4. – С. 58-64.
3. Лызин, И.А Опыт реабилитации детей с хроническим пиелонефритом в условиях промышленного города / Ю.В. Фаткина, Н.П. Степаненко, О.Г. Берестнева, И.А. Лызин // Сибирский журнал клинической и экспериментальной медицины. – 2021. – Т. 36. – № 2. – С. 98-106 (Scopus).
4. Лызин, И.А. Эффективность «Школы здоровья: социально-психологическая реабилитация детей с заболеванием почек» в рамках комплексной медицинской реабилитации / Ю.В. Фаткина, Н.П. Степаненко, Г.Г. Решетова, Т.Ю. Левицкая, О.Г. Берестнева, И.А. Лызин // Вопросы курортологии, физиотерапии и лечебной физической культуры. – 2023. – Т. 100. – № 3. – С. 28-38 (Scopus).
5. Лызин, И.А. Разработка системы поддержки принятия врачебных решений на примере решения задачи выбора наилучшей траектории лечения для детей с эндокринопатиями / О.Г. Берестнева, И.А. Лызин, С.В. Аксенов, О.В. Марухина, Н.П. Степаненко // Информационные и математические технологии в науке и управлении. – 2023. – № 4(32). – С. 127-138.

Статьи в сборниках трудов международных конференций

6. Lyzin, I.A. The use of computer visualization in the analysis of breathing curves / O.G. Berestneva, I.A. Lyzin, A.L. Yumasheva O.V. Marukhina, A. Trufanov // IOP Conference Series: Materials Science and Engineering. – 2021. – Vol. 1019. – 012007 (14th International Forum on Strategic Technology, IFOST 2019) (Scopus).
7. Lyzin, I.A. Application of Integral Health Index in Evaluation of Efficiency of Rehabilitation of Children with Obesity/ O.G. Berestneva, I.A. Lyzin, N.P. Stepanenko // Studies in Systems, Decision and Control. – 2022. – Vol. 416. – P. 87-100 (Scopus).
8. Лызин, И.А. Алгоритмическое, программное обеспечение системы поддержки принятия врачебных решений / И.А. Лызин, С.В. Аксенов, О.Г. Берестнева, Н.П. Степаненко, Ю.В. Фаткина // Информационные технологии в науке, управлении, социальной сфере и медицине: материалы VII Международной научной конференции (Томск, Иркутск, 17–25 октября 2022 г.), Томск: Томский государственный педагогический университет. – 2023. – С. 102-107.
9. Лызин, И.А. Анализ медицинских и психологических текстов средствами тематического моделирования / С.А. Землянский, С.В. Аксенов, И.А. Лызин // Информационные технологии в науке, управлении, социальной сфере и медицине: материалы VII Международной научной конференции (Томск, Иркутск, 17–25 октября 2022 г.), Томск: Томский государственный педагогический университет. – 2023. – С. 188-196.
10. Лызин, И.А. Технология формирования баз знаний для систем поддержки принятия врачебных решений / О.Г. Берестнева, И.А. Лызин, О.В. Марухина, Н.П. Степаненко // Труды Международного научно-технического конгресса «Интеллектуальные системы и информационные технологии – 2021 (ИС & ИТ-2021)», Таганрог, 2021. – С. 156-163.

11. Лызин, И.А. Системы поддержки принятия врачебных решений / С.В. Аксенов, И.А. Лызин, Е.В. Берестнева, О.В. Марухина // Фундаментальные и прикладные исследования: опыт, проблемы и перспективы. Сборник научных трудов по материалам Международной научно-практической конференции (Санкт-Петербург, 10 марта 2020 г.). – С. 143-147.
12. Лызин, И.А. Алгоритм оценки эффективности научно-медицинских центров / И.Г. Видяев, И.А. Лызин, О.Г. Берестнева, С.В. Аксёнов // Информационные системы и технологии – 2019: сборник материалов XXV Международной научно-технической конференции (Нижний Новгород, 16 апреля 2019 г.), Нижний Новгород: Изд-во НГТУ. – 2019. – С. 469-471.
13. Лызин, И.А. Информационная технология анализа результатов лечения пациентов с эндокринопатиями / И.А. Лызин, О.В. Марухина // Молодежь и современные информационные технологии: сборник трудов XVI Международной научно-практической конференции студентов, аспирантов и молодых ученых, Томск, 3-7 Декабря 2018. – Томск: ТПУ, 2019. – С. 92-93.
14. Лызин, И.А. Применение индивидуального интегрального показателя здоровья на примере оценки эффективности реабилитации детей с ожирением / Н.П. Степаненко, О.Г. Берестнева, А.Л. Юмашева, И.А. Лызин // Информационные системы и технологии – 2020: сборник материалов XXVI Международной научно-технической конференции, Нижний Новгород, 24-28 апреля 2020. – Нижний Новгород: Изд-во НГТУ, 2020. – С. 662-669.
15. Лызин, И.А. Система поддержки принятия решения для выбора траектории лечения детей с эндокринопатиями / Марухина О.В., Берестнева О.Г., Лызин И.А., Аксёнов С.В., Мокина Е.Е. // Международный научно-технический конгресс «Интеллектуальные системы и информационные технологии - 2020» «ИС & ИТ-2020», Таганрог, 2020. – Т. 1. – С. 217-222.

Статьи в других изданиях

16. Лызин, И.А. Индивидуальный интегральный показатель здоровья в оценке эффективности реабилитации детей с ожирением / И.А. Лызин, Н.П. Степаненко, О.Г. Берестнева, Е.В. Берестнева, О.В. Марухина // Математические методы в технологиях и технике. – №1. – 2021. – С. 158-163.
17. Лызин, И.А. Оценка эффективности реабилитационных мероприятий с помощью индивидуального интегрального показателя здоровья / Н.П. Степаненко, И.А. Лызин, А.Л. Юмашева // Международный журнал прикладных и фундаментальных исследований. – № 10. – 2019 – С. 179-183.