

Министерство образования и науки Российской Федерации
федеральное государственное бюджетное образовательное учреждение высшего
образования

НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ им. Р.Е. АЛЕКСЕЕВА (НГТУ)

УДК 528.854

№ госрегистрации АААА-А17-117092860072-5 от 28.09.2017 г.

Инв. № 16/2377

УТВЕРЖДАЮ

Проректор по научной работе,
докт. техн. наук, доц.

Н.Ю. Бабанов

«9» октября 2017 г.



ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

Изучение спектра кардиографических отклонений у больных с нарушением
регуляции глюкозы

Заключительный

Начальник УНИР и ИР

В. В. Беляков

Начальник ПФУ

С. Ю. Обыденнова

San Директор ИРИТ

В. Г. Баранов

Заведующий кафедрой ВСТ

В. В. Кондратьев


Руководитель темы

San В. Е. Гай


Нижний Новгород 2017


Stef


СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель темы,  9.10.2017 г. В. Е. Гай
к. т. н. подпись, дата (введение, раздел(ы) 2, заключение)

Исполнители темы:

Техник  9.10.2017 г. Н. А. Домнина
подпись, дата (введение, раздел(ы) 1)

Инженер  9.10.2017 г. И. В. Поляков
подпись, дата (раздел(ы) 2)

Техник  9.10.2017 г. М. О. Дербасов
подпись, дата (раздел(ы) 1)

СОДЕРЖАНИЕ

СПИСОК ИСПОЛНИТЕЛЕЙ.....	2
ОПРЕДЕЛЕНИЯ	4
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	5
ВВЕДЕНИЕ	6
1 Аналитический обзор современной научно-технической, нормативной, методической литературы, затрагивающей научно-техническую проблему, исследуемую в рамках НИР, в том числе обзор научных информационных источников: статьи в ведущих зарубежных и (или) российских научных журналах, монографии и (или) патенты) – не менее 15 научно-информационных источников за период 2006–2016 гг.	7
1.1 Инвазивные методы оценки уровня глюкозы в крови	7
1.2 Неинвазивные методы оценки уровня глюкозы в крови.....	12
1.3 Заключение по разделу	15
2 Выбор и обоснование методов и средств, направления исследований и способы решения поставленных задач	17
2.1 Обоснование метода неинвазивной оценки уровня глюкозы в крови с позиций теории активного восприятия.....	17
2.2 Вычислительный эксперимент.....	24
2.2.1 Описание набора данных	24
2.2.2 Описание методики тестирования	25
2.2.3 Результаты вычислительного эксперимента	30
2.3 Заключение по разделу	41
ЗАКЛЮЧЕНИЕ.....	43
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	44

ОПРЕДЕЛЕНИЯ

Теория	активного	–	теория, описывающая методы формирования признакового описания и предварительной обработки изображения
Классификатор		–	алгоритм отнесения некоторого объекта в соответствии с заданным решающим правилом к одному из классов
Машинное обучение		–	обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться
Группа	(полная, замкнутая)	–	Подмножество множества операторов

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ТАВ	–	теория активного восприятия
ЭКГ	–	электрокардиограмма

ВВЕДЕНИЕ

Ключевые слова: ЭЛЕКТРОКАРДИОГРАФИЧЕСКИЙ СИГНАЛ; УРОВЕНЬ ГЛЮКОЗЫ; МАШИННОЕ ОБУЧЕНИЕ; ТЕОРИЯ АКТИВНОГО ВОСПРИЯТИЯ.

Известно существует большое количество задач, связанных с анализом данных различной физической природы. Все труднее становится осуществлять их решение вручную, поэтому набирает популярность машинное обучение, которое позволяет автоматически находить закономерности в массивах данных с помощью алгоритмов.

Машинное обучение применяется во многих сферах, в том числе и в медицине. Было проведено множество исследований, в которых говорится, что каждое заболевание по-особому изменяет ритм сердца. Следовательно, анализируя электрокардиографический сигнал и находя признаки, специфичные для конкретной болезни, можно строить системы автоматической диагностики, что открывает большой простор для машинного обучения.

Одна из возможных проблем организма, которую можно диагностировать с помощью анализа ЭКГ-сигнала – это высокое содержание сахара в крови, или глюкозы, которое возникает по причине недостаточного воздействия инсулина. Речь идет о таком заболевании, как сахарный диабет.

В течение последних нескольких десятилетий распространение диабета неуклонно растет. По данным Всемирной организацией здравоохранения число людей с этим заболеванием по всему миру возросло с 108 миллионов в 1980 году до 422 миллионов в 2014 году. В Российской Федерации в 2015 году число больных составляет 4,4 миллиона человек. По сравнению с 2014-м годом количество больных сахарным диабетом увеличилось на 5,6%, а за 3 года с 2013 по 2015 годы – на 23%.

Болезнь опасна как ранними осложнениями - диабетической и гипогликемической комой, так и поздними, возникающими при длительном повышенном уровне глюкозы в крови. К таким осложнениям относятся инсульт, болезни сердца, почек, потеря зрения и слуха.

Результаты лечения диабета показали, что более частый контроль глюкозы и инсулина в крови может предотвратить многие из долгосрочных осложнений сахарного диабета. Пациентам рекомендуется проводить самоконтроль сахара в крови несколько раз в день, чтобы корректировать свою диету и применение сахароснижающих препаратов.

Действие самых популярных на сегодняшний день приборов для мониторинга содержания глюкозы в крови основано на измерении электрического тока, возникающего при окислении глюкозы капиллярной крови на поверхности тестовой полоски. Очевидным недостатком такого метода является его инвазивность, т.е. необходимость нарушения кожных покровов, прокола кожи.

Цель данной НИР – разработка метода неинвазивной оценки уровня глюкозы в крови.

1 Аналитический обзор современной научно-технической, нормативной, методической литературы, затрагивающей научно-техническую проблему, исследуемую в рамках НИР, в том числе обзор научных информационных источников: статьи в ведущих зарубежных и (или) российских научных журналах, монографии и (или) патенты) – не менее 15 научно-информационных источников за период 2006–2016 гг.

Данный раздел отчета описывает результаты работ, проведенных в соответствии с п. 1 Календарного плана работы и соответствует требованиям п. 7.1.1.1 Технического задания на выполнение НИР.

Начальный этап работ по созданию нового метода неинвазивной оценки уровня глюкозы в крови включает изучение текущего состояния науки и отрасли в сопряженных с затрагиваемых в работе направлениях.

1.1 Инвазивные методы оценки уровня глюкозы в крови

Метод, описанный в [1], служит для прогнозирования значений глюкозы в течение 15, 30 и 45 минут. Предварительная обработка данных включает в себя уменьшение влияния шума путем предварительной фильтрации с использованием каузального метода фильтрации Калмана. Для предсказания значений глюкозы используется 2 метода – искусственная нейронная сеть и модель авторегрессии.

В методе, основанном на искусственной нейронной сети, входная информация представляет собой текущее измерение уровня глюкозы и его метки времени вместе с ограниченным количеством предыдущих образцов глюкозы в системе. Нейронная сеть учитывает измерения глюкозы за 20 минут до текущего времени. Поскольку частота дискретизации варьируется от одной системы к другой, количество входов сети различно для каждого набора данных. Выходной сигнал сети представляет собой предсказание глюкозы в заданный момент времени. Параметры сети (веса и смещения) случайным образом инициализируются и, впоследствии, обновляются в соответствии с алгоритмом оптимизации Левенберга-Марквардта с обратной проработкой. Алгоритм обучения применяется в пакетном режиме, в котором веса и смещения обновляются только тогда, когда предоставляются все входы и цели. Для каждого набора данных для обучения нейронной сети использовалось 3 человека (каждый с двухдневными профилями). Пропущенные образцы были рассчитаны с использованием сплайновых методов для восстановления отсутствующих образцов перед использованием профилей глюкозы для обучения.

В методе, основанном на модели авторегрессии, берется текущее значение глюкозы и значения, собранные за некоторое время до начала предсказания. Все эти данные участвуют в предсказании с различными относительными экспоненциальными весами. Для каждого конкретного человека берется свой так называемый коэффициент забывания для вычисления этих

весов. Это необходимо для минимизации среднеквадратической ошибки между исходными данными обучающего набора и прогнозируемыми значениями.

В каждой из двух моделей вычисляется точность и задержка. Точность оценивается как среднеквадратичная ошибка прогнозируемого и реального значения глюкозы. Задержка модели оценивалась путем расчета задержек между исходным и прогнозируемым профилями при пересечении трех разных пороговых значений, определяемых следующим образом. Во-первых, пики идентифицируются в исходном профиле датчика. Затем пороговые значения размещаются на 25%, 50% и 75% от расстояния от пика до пика для положительных и отрицательных трендов. Окончательная задержка модели рассчитывается как среднее значение пороговых задержек для каждой положительной и отрицательной тенденции. Чтобы легче идентифицировать пороговые пересечения, учитывая ретроспективный характер оценки, исходный и предсказанный профили были отфильтрованы фильтром нижних частот Баттерворта первого порядка.

Результаты работы моделей оценивались на двух наборах данных:

1. Данные девяти пациентов с диабетом первого типа. Пациенты использовали систему с перерывами в течение 72 часов в неделю в течение четырех недель. Набор включает 12 ежедневных профилей для каждого пациента: 6 дневных и 6 записей за половину дня.
2. Данные шести пациентов с диабетом первого типа. Пациенты использовали систему около 72 часов. Набор включает 2 полных ежедневных профиля для каждого пациента.

Среднеквадратическая ошибка для первого набора данных (mg/dl) при использовании нейронной сети равна 9.74 ± 2.71 , при использовании авторегрессии 9.26 ± 2.97 . Среднеквадратическая ошибка для второго набора данных (mg/dl) при использовании нейронной сети равна 10.38 ± 3.15 , при использовании авторегрессии 10.46 ± 3.55 .

Задержка для первого набора данных при подъеме с использованием нейронной сети равна 3.92 ± 1.21 минут, с использованием авторегрессии 0.75 ± 0.28 минут. При спаде с использованием нейронной сети 5.10 ± 1.65 минут, с использованием авторегрессии 1.30 ± 0.38 минут.

Задержка для второго набора данных при подъеме с использованием нейронной сети равна 4.58 ± 1.42 минут, с использованием авторегрессии 1.745 ± 0.50 минут. При спаде с использованием нейронной сети 5.37 ± 2.16 минут, с использованием авторегрессии 3.18 ± 1.58 минут.

В [2] рассматривается задача прогнозирования уровня глюкозы в крови с помощью непрерывного её мониторинга. Предсказывается уровень и скорость, с которой значение глюкозы будет изменяться в течение 30 минут после последнего измерения.

При предварительной обработке применяется фильтр Баттерворта первого порядка для удаления выбросов, вызванных, например, движением пациента. Метод представляет собой

полуконтролируемое обучение. Он основан на аппроксимации функций с использованием диффузионных полиномов.

Для оценки точности используется метод PRED-EGA, анализирующий сеть ошибочных прогнозов. Метод сообщает процент точных предсказаний, доброкачественных и ошибочных предсказаний отдельно для трех диапазонов: гипогликемическом, эугликемическом и гипергликемическом.

Эксперименты проводились на 25 пациентах. Обучение проводилось при использовании 30% данных для обучающей выборки и 50% данных. Лучшие результаты были получены при 50%. Проценты верных прогнозов:

- 93,21% в гипогликемической диапозоне
- 97.68% в эугликемическом диапозоне
- 86,78% в диапозоне гипергликемии.

В статьях [3] и [4] используется физиологическая модель, которая оценивает уровень глюкозы по изменению различных систем организма (почки, печень и т.д.). Её рассматривают как непрерывную динамическую модель, которая описывается состоянием ее переменных, входными переменными и функцией перехода состояния, которая вычисляет следующее состояние с учетом текущего и входных переменных. Вектор состояния переменных организован в соответствии с тремя блоками следующим образом:

1. Динамика поглощения пищи
2. Динамика инсулина
3. Динамика уровня глюкозы

Для предсказания используется метод опорных векторов с отдельной настройкой параметров для каждого конкретного пациента. Для оценки точности используется квадрат среднеквадратической ошибки.

Эксперимент был проведен на 5 пациентах, всего набор данных содержал 200 ситуаций для оценки качества работы модели. Рассматриваемая модель может предсказать 23% гипогликемических событий с 1% ложных срабатываний.

Статья [5] описывает метод диагностики сахарного диабета и его стадии. Перед началом классификации применяется метод линейного дискриминантного анализа. Он используется для отделения людей, страдающих сахарным диабетом от здоровых. Для получения стадии диабета используется многокомпонентный метод опорных векторов Вейвлета.

Оценивается чувствительность, специфичность и точность классификации, используется матрица ошибок. Эксперименты проводились на 768 образцах. Полученная чувствительность – 83.33%, специфичность – 93.75%, точность – 89.74%.

В статье [6] изучается взаимосвязь между гематологическими параметрами и гликемическим статусом для выявления людей, имеющих сахарный диабет, и людей, у которых он появится в скором времени.

В качестве входных данных используются: лейкоциты, эритроциты, гемоглобин и параметры глюкозы. Для предсказания используется 2 метода: метод опорных векторов и искусственная нейронная сеть. В методе опорных векторов входные данные были преобразованы в пространство высокой размерности с помощью ядра радиальной базисной функции.

Эксперименты проводились на 190 людях. Для оценки точности строится матрица ошибок. В обоих методах достигнуто 98% точности.

В [7] метод служит для предсказания гипогликемического эпизода на следующей неделе. На вход метода подается значение глюкозы и метка времени, из которой потом получаются такие переменные как время дня, день недели, месяц и т.д.

Используется 2 набора входных переменных. Первый набор состоит из значений глюкозы за последнюю неделю, собранных в час дня, а также из разницы между текущим значением и предыдущим. Второй набор включает помимо данных из первого набора, информацию о принятых лекарствах и их дозах за последние 7 дней.

Для предсказания применяется четыре алгоритма: случайный лес, метод опорных векторов, метод К ближайших соседей и наивный Байесовский метод. Для оценки точности строится матрица ошибок, по которой считается точность, чувствительность и специфичность.

Эксперименты проводились на 10000 собранных данных. Лучше результаты показали модель случайного леса и модель, основанная на методе опорных векторов. Точность этих моделей на первом наборе данных 95%. На втором наборе данных точность этих моделей составляет 97%. Чувствительность модели на первом наборе данных для прогнозирования гипогликемии составила 92%, а специфичность 70%. В модели, построенной на втором наборе данных, специфичность увеличилась до 90%.

Метод, описанный в статье [8], осуществляет непрерывный мониторинг уровня глюкозы с использованием микроконтроллера смешанного сигнала MSP430. Алгоритм основан на комбинации метода сохранения энергии с модулем интеграции датчиков, который собирает такие физиологические параметры как насыщение кислородом крови, скорость кровотока и частоту сердечных сокращений. Предложены новые методы для определения базальной скорости обмена в сочетании с тепловым балансом человеческого тела и характерными сигналами фотолейзографии, а также теорией двойных упругих камер.

Алгоритм, применяемый в исследуемой модели, объединяет в себе дерево принятия решений и нейронной сети с обратным распространением ошибки, которая классифицирует

физиологические и экологические параметры на три категории, а затем устанавливает соответствующую модель прогнозирования для трех категорий.

Эксперимент проводился на пациентах с диабетом, на пожилых и на здоровых людях. Алгоритм обеспечивает точность 88,53%. Минус метода состоит в необходимости постоянно прикрепленного на тело модуля датчиков.

Статья [9] описывает метод предсказания глюкозы на 15, 30, 60 и 120 минут. Каждые 5 минут записывается среднее значение глюкозы. Каждую минуту считаются энергетические затраты организма. Также человек самостоятельно заносит информацию о приеме пищи и дозе инсулина.

Для предсказания используется метод опорных векторов. Для оценки точности применяется корень среднеквадратической ошибки при проведении перекрестной проверки 10 раз, а так же считается коэффициент корреляции.

Эксперимент был проведен на 27 пациентах, которые были поделены на 3 группы:

- А. 15 пациентов, предоставляющих полную информацию
- В. 5 пациентов, без информации о деятельности
- С. 7 пациентов, предоставляющих только значения уровня глюкозы

Рассматривается 6 моделей:

- 1) Для предсказания используются только значения глюкозы за последние 30 минут.
- 2) К значениям уровня глюкозы за последние 30 минут добавляется переменные, отражающие реакцию организма на инсулин и на прием пищи.
- 3) За последние 90 минут берется суммарное количество экзогенной глюкозы в плазме крови, а также значения уровня глюкозы за это же время.
- 4) Суточный контроль метаболизма глюкозы. К модели 3 добавляется переменная, идентифицирующая 24-часовые интервалы для фиксации ежедневных ритмов концентрации глюкозы.
- 5) К модели 3 добавляется информация о физической активности за последние 3 часа.
- 6) к модели 4 добавляется информация о физической активности за последние 3 часа.

Наилучший результат показывает группа А с краткосрочными прогнозами (15 и 30 минут). С увеличением предоставляемой пациентами информацией качество прогноза увеличивается, лучший прогноз – в модели 6.

В критических ситуациях процент успешных прогнозов постепенно улучшается от случая 1 до модели 3. В частности, для случая 3 - 89%, 85%, 76% и 70% значений группы А успешно прогнозируются как гипогликемические, тогда как 96%, 95%, 91% и 89% успешно прогнозируются как гипергликемический для каждого временного интервала соответственно (15, 30, 60, 120 минут). Для моделей 4-6 производительность улучшается, хотя между ними нет

систематических различий. Метод 6 дает 91%, 87%, 83% и 85% успешных гипогликемических прогнозов и 96%, 96%, 94% и 92% гипергликемических.

Для моделей 4 и 6 группы А также рассматривается количество точных, доброкачественных и ошибочных прогнозов для гипогликемии, эугликемии и гипергликемии.

В обоих случаях для всех временных интервалов более 90% предсказаний являются точными или с доброкачественными ошибками. Краткосрочные прогнозы являются более точными в гипогликемическом диапазоне по сравнению с гипергликемическим. Точность уменьшается для гипогликемии при увеличении временного интервала. В целом, модели 4, 5, и 6 работают одинаково в отношении этого типа анализа, в то время как модели 1-3 связаны с более ошибочными предсказаниями.

В статье [10] описывается предсказание уровня глюкозы на 30 и 60 минут. Признаки включают в себя следующее переменные:

- 1) Уровень глюкозы в настоящее время.
- 2) Скользящее среднее из 4 предыдущих значений глюкозы, включая текущее.
- 3) Экспоненциально сглаженная скорость изменения 4 предыдущих значений глюкозы, включая текущее.
- 4) Дозы Болюса, начинающиеся за 30 минут до времени прогнозирования.
- 5) Средняя базальная скорость, с интервалом в 5 или 15 минут.
- 6) Количество принятых углеводов за предшествующие 30 минут.
- 7) Среднее значение интенсивности упражнений за прошедшие 5, 30 или 60 минут.

Для предсказания используется метод опорных векторов. Оценка точности производится с помощью построения диаграммы Кларка.

Результаты работа метода:

- Для 30-минутного предсказания: зона А – 87.8%, В – 11.8%, D – 0.4%.
- Для 60-минутного предсказания: зона А – 74.5%, В – 21.5%, С – 2.2%, D – 1.8%.

1.2 Неинвазивные методы оценки уровня глюкозы в крови

В статье [11] исследуется возможность анализа сигналов дыхания как способ контроля уровня глюкозы в крови. Система химических датчиков собирает образцы выдыхаемого человеком воздуха для дальнейшего анализа уровня ацетона. Вместе с этим собираются значения уровня глюкозы в крови, полученные анализом крови, для оценки результатов прогнозирования.

Предварительная обработка включает в себя удаление шумов, базовую манипуляцию и нормализацию. Базовая манипуляция осуществляется для компенсации дрейфа, усиления контраста и масштабирования. Нормализация используется для компенсации разброса выборки, вызванной концентрацией аналита и давлением кислорода.

Для предсказания используется метод опорной ординальной регрессии для классификации образцов в четыре группы. Экспериментальные результаты показывают, что точность классификации составляет до 68.66%. Минус данного метода состоит в том, что для анализа уровня глюкозы необходимо выдохнуть в одноразовый мешок, что может быть не всегда удобно.

Для измерения уровня глюкозы в статье [12] используется гальваническая реакция на кожу и измерение температуры тела совместно с сигналом фотоплетизмографа. Для предсказания используется два метода машинного обучения - множественная линейная регрессии и искусственная нейронная сеть.

Исследование проводили на 50 людях, не страдающих диабетом, с индексом массы тела $27,3 \pm 3 \text{ кг/м}^2$. Для оценки точности используется диаграмма ошибок Кларка и подсчет коэффициента корреляции. Результаты показали 86.01% значений в зоне А и 13.99% - в зоне В, коэффициент корреляции равен 0.94

Работа [13] представляет собой систему для неинвазивной оценки уровня глюкозы в крови с использованием фотоплетизмографа и методов машинного обучения. Из формы сигнала, полученного с прибора, модуль обработки извлекает функцию, которая в дальнейшем обрабатывается для получения значения глюкозы в крови. Для предсказания используются 4 метода машинного обучения: линейная регрессия, многослойный персептрон, метод опорных векторов и случайные леса. Наилучший результат показал метод случайных лесов.

Работа методов была проверена на 410 людях без какой-либо персональной калибровки. Для оценки точности используется диаграмма ошибок Кларка и среднеквадратическая ошибка.

Результаты зон Кларка представлены для метода случайного леса, так как среднеквадратическая ошибка этого метода была наименьшей. Результаты были вычислены после перекрестной проверки: зона А - 87,7%, зона В - 10,3%, зона D - 1,9%.

В исследовании [14] уровень глюкозы в крови оценивается с помощью оценки электрохимических параметров слюны человека. Перед обработкой данных проводится масштабирование существенных признаков для всех параметров, а именно кислотность, окислительно-восстановительный потенциал, проводимость и концентрации электролитов, возраст добровольца, для получения нормализованных данных в диапазоне от -1 до 1. Это было сделано для того, чтобы избежать любых смещений, вызванных различиями в единицах измерения параметров.

Для предсказания уровня глюкозы используется 3 метода: логистической регрессии, искусственная нейронная сеть с обратной связью и метод опорных векторов.

В методе логистической регрессии высокий уровень содержания глюкозы в крови определяется по возрастным и слюновым электрохимическим параметрам. Модель генерирует выходные значения с точки зрения вероятностей, в качестве порога выбрано значение 0.5,

эквивалентное 120 mg/dl . Для тренировки модели применяется алгоритм градиентного спуска. С его помощью вычисляются значения коэффициентов с использованием метода средней квадратичной ошибки.

В искусственной нейронной сети используется метод градиентного спуска с обратным распространением ошибки. Архитектура классификатора состоит из входного слоя с 7 нейронами (по одному для каждого параметра), 33 нейронов скрытого слоя и двух узлов на выходном уровне, каждый из которых имеет один нейрон. Нейронная сеть была обучена путем сокращения среднеквадратической ошибки учебного набора данных. После её минимизации значения полученных констант были сохранены для проверки модели на половине оставшихся данных.

В методе опорных векторов используются линейная и Гауссова (радиальная базисная) функции ядра. Для разработки оптимальной модели были выбраны два ключевых параметра: C и γ для ядер. C обычно известен как параметр штрафа, который контролирует перенастройку модели. В случае радиальной базисной функции классификация обычно лучше из-за более высокого значения C , что делает классификацию более корректной. Параметр γ контролирует степень нелинейности модели. C обычно используется для реализации как линейных, так и радиальных базисных функций, тогда как γ используется специально для ядра радиально базисной функции.

Для оценки точности все предсказанные значения глюкозы делятся на:

- истинно-положительное (и реальное, и предсказанное значение ≥ 120 mg/dl)
- истинно-отрицательное (и реальное, и предсказанное значение < 120 mg/dl)
- ложно-положительное (реальное значение < 120 mg/dl, предсказанное ≥ 120 mg/dl)
- ложно-отрицательное (реальное значение ≥ 120 mg/dl, предсказанное < 120 mg/dl)

Чтобы получить разумную производительность бинарного классификатора, отношение истинно-положительных предсказаний к ложно-положительным должно быть высоким.

После распределения получившихся значений на данные группы вычисляются:

- Ассигасу (доля правильных ответов) – отношение количества правильно предсказанных случаев и общего количества предсказаний
- Precision (точность) - доля фактических обнаруженных случаев превышения уровня глюкозы
- Recall (полнота) - оценка действительно выявленных случаев превышения глюкозы среди фактических
- F_1 score (F-мера) – среднее гармоническое между точностью и полнотой

Лучший результат был получен при использовании модели, основанной на методе опорных векторов с Гауссовым ядром. Он показывает долю правильных ответов $(84.09 \pm 2.8)\%$, точность $(83.75 \pm 3.3)\%$, полноту $(84.92 \pm 4.5)\%$ и F-меру $(84.06 \pm 2.9)\%$.

В исследовании [15] был создан электронный дневник, в котором люди документируют все свои действия, приемы пищи, эмоциональные состояния и т.д. На основании этого нейронная сеть предсказывает уровень глюкозы в ближайшее время.

В данной работе применяются запаздывающие нейронные сети с обратной связью. Эти нейронные сети классифицируются как многослойные персептроны, которые имеют компоненты памяти для хранения предыдущих значений данных в сети. Нейронные сети были обучены с использованием метода обратного распространения ошибок. Оптимальные веса для минимизации ошибки в предсказательной модели получены с помощью алгоритма градиентного спуска.

Модели, разработанные в этом исследовании, включали оптимизацию с использованием генетических алгоритмов. Оптимизация с использованием генетического алгоритма была использована для минимизации числа нейронов и входов в нейронную сеть. Генетический алгоритм эффективно определяет, какие входы влияют на предсказание, и минимизирует различные взаимосвязи между нейронами в нейронной сети. Генетический алгоритм также определяет наилучшее значение для размера и импульса шага для нейронной сети.

Для оценки точности результатов вычисляется средний процент абсолютной разницы (MAD %) по всему набору данных, а также гипогликемические (≤ 70 mg/dl) и гипергликемические экстремумы (≥ 80 mg/dl).

Результаты вычисляются при 11-17 пациентах, используемых для обучения нейронной сети. Для оценки всегда используются данные 18-го пациента. Лучшие результаты показала модель с 16 пациентами в обучающей выборке:

- Средний процент абсолютной разницы - 22.1%
- Средний процент абсолютной разницы по гипергликемическим экстремумам - 11.2%
- Средний процент абсолютной разницы по гипогликемическим экстремумам - 54.1%
- Процент гипергликемии, предсказанный системой - 70.3%
- Процент гипогликемии, предсказанный системой - 1.1%

1.3 Заключение по разделу

Данный раздел отчета описывает результаты работ, проведенных в соответствии с п. 1 Календарного плана работы и соответствует требованиям п. 7.1.1.1 Технического задания на выполнение НИР.

В результате проведения аналитического обзора современной научно-технической, были достигнуты следующие результаты:

- Изучена и подтверждена актуальность задачи неинвазивных способов оценки уровня глюкозы в крови;

- Собрана информация о известных подходах неинвазивной и инвазивной оценке уровня глюкозы в крови, определены их основные недостатки.

2 Выбор и обоснование методов и средств, направления исследований и способы решения поставленных задач

Данный раздел отчета описывает результаты работ, проведенных в соответствии с п. 1 Календарного плана работы и соответствует требованиям п. 7.1.1.2 Технического задания на выполнение НИР.

2.1 Обоснование метода неинвазивной оценки уровня глюкозы в крови с позиций теории активного восприятия

Метод неинвазивной оценки уровня глюкозы в крови можно представить в виде совокупности трёх этапов (см. рис. 2.1): предварительная обработка, вычисление признаков и принятие решения.

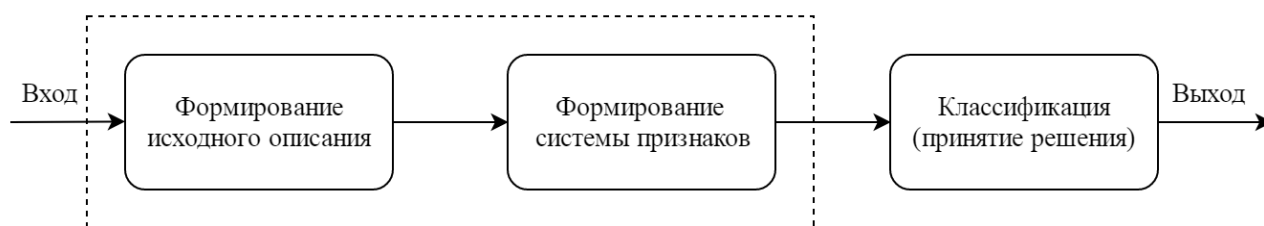


Рисунок 2.1 – Структура системы распознавания

Рассмотрим решение задачи распознавания уровня глюкозы в крови по ЭКГ сигналу с позиций теории активного восприятия [16].

Этап предварительной обработки заключается в выполнении Q -преобразования, которое заключается в применении к сегментам исходного сигнала операции сложения:

$$g(t) = \sum_{k=(t-1)*L+1}^{t*L} f(k), t = \overline{1, N} \quad (2.1)$$

где $g(t)$ – t -ый отсчёт сигнала g ;

g – результат применения Q -преобразования к сигналу f ;

L – число отсчётов, входящих в сегмент;

$f(k)$ – k -ый отсчёт сигнала f ;

N – число сегментов сигнала.

Формирование признакового описания исходного сигнала заключается в применении к сигналу g множества фильтров Уолша системы Хармута:

$$\mu(k, c(t)) = \sum_{i=0}^{M-1} F_k(i) * g\left(\frac{(t-1)*M+1}{t*M}\right) \quad (2.2)$$

где $\mu(k, c(t))$ – результат применения множества фильтров Уолша системы Хармута к сигналу g ;

$k = \overline{0, M-1}$, $t = \overline{0, |c|-1}$, $c = \{1, P, 2 \cdot P, 3 \cdot P, \dots, N - T \cdot P\}$ – множество значений смещений по сигналу g ;

$|c|$ – мощность множества c ;

P – величина смещения по сигналу g ($1 \leq P \leq M$);

M – число используемых фильтров.

Таким образом, признаковое описание сигнала представляет собой матрицу размером $M \times |c|$, причём каждая строка признакового описания представляет собой результат U -преобразования сегмента сигнала.

Последовательное применение к сигналу Q -преобразования и системы фильтров реализуют U -преобразование, являющееся базовым в теории активного восприятия.

Используемые фильтры представлены на рис. 2.2.

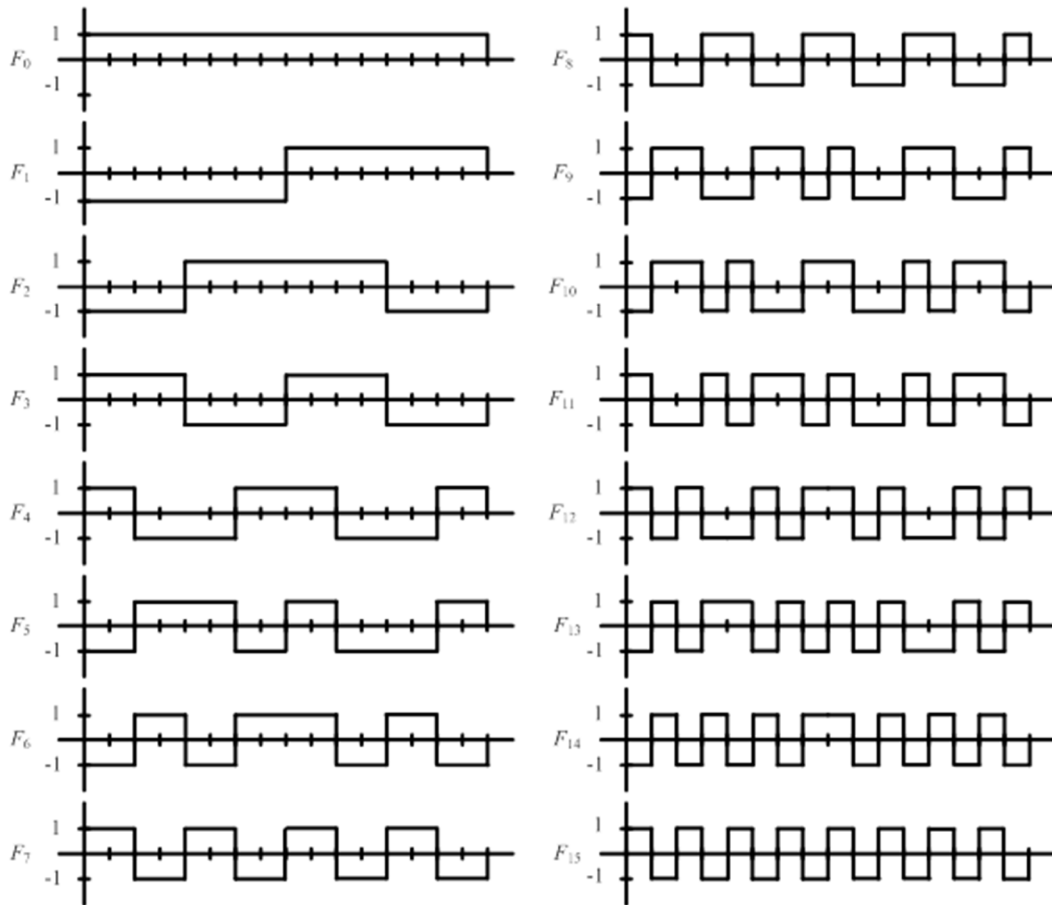


Рисунок 2.2 – Система фильтров

U -преобразование имеет минимально возможную вычислительную сложность, поскольку при его реализации используются простейшие операции – сложение и вычитание. Стандартные преобразования, требуют реализации свертки, а на уровне весовых коэффициентов – операции арифметического умножения.

Этап формирования системы признаков также включает в себя алгебру групп. Обнаруженные зависимости допускают своё использование на этапах принятия решения и понимания анализируемого сигнала.

Пусть каждому фильтру $F_i \in \{F_i\} \equiv \mathbf{F}$ соответствует бинарный оператор $V_i \in \{V_i\} \equiv \mathbf{V}$; тогда компоненте $\mu_i \neq 0$ вектора μ допустимо поставить в соответствие оператор V_i либо \bar{V}_i в зависимости от знака компоненты. В результате вектору μ ставится в соответствие подмножество операторов из $\{V_i\}$, имеющих аналогичную фильтрам конструкцию, но разное значение элементов матрицы ($+1 \leftrightarrow 1$; $-1 \leftrightarrow 0$). Задавая на множестве $\{V_i\}$ операции теоретико-множественного умножения и сложения, имеем алгебру описания сигнала в одномерных булевых функциях. С учётом инверсий всего существует 15 операторов, которые могут использоваться при формировании признакового описания, так как оператор V_0 принимает только прямое значение.

На множестве операторов формируется алгебра групп (этап синтеза) анализируемого сигнала:

1. семейство алгебраических структур (названных полными группами) $\{P_{ni}\}$ вида $P_{ni} = \{V_i, V_j, V_k\}$ мощности 35;
2. семейство алгебраических структур (названных замкнутыми группами) $\{P_{si}\}$ вида $P_{si} = \{V_i, V_j, V_k, V_r\}$ мощности 105, где каждая группа образована из пары определенным образом связанных полных групп.

Схематическое представление алгебры групп представлено на рис. 2.3.

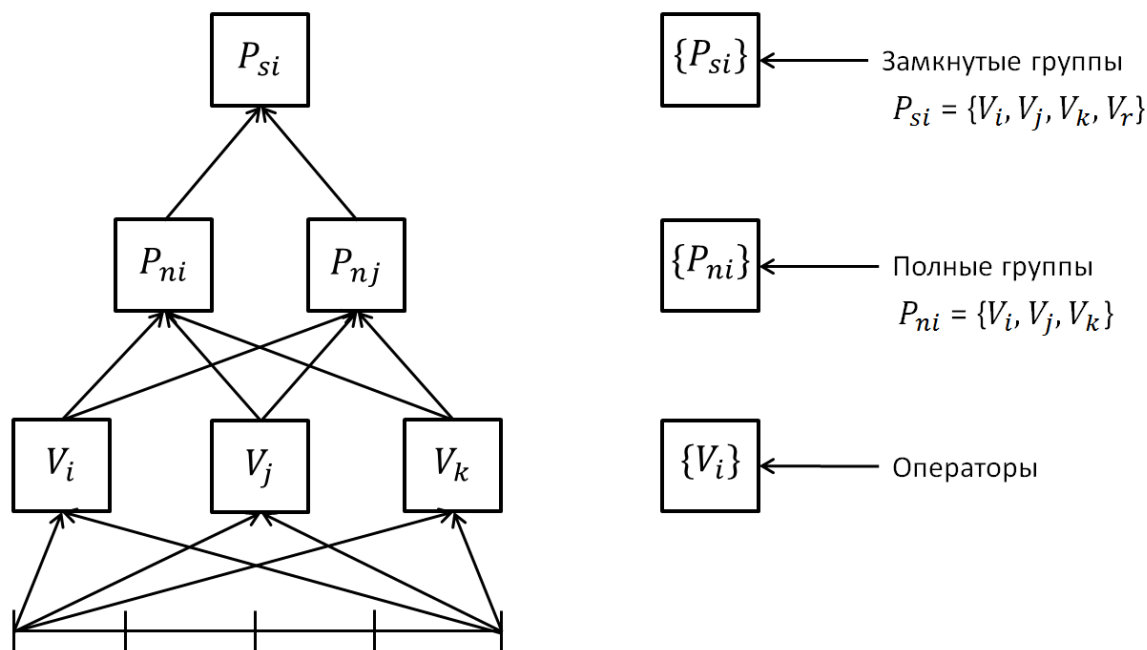


Рисунок 2.3 – Алгебра групп

Среди полных групп выделяют полные группы на операции сложения и на операции умножения, среди замкнутых групп – замкнутые группы и замкнутые множества.

Две группы (полные или замкнутые) называются несовместными, если в их состав входят операторы с одинаковыми номерами, но с разными знаками.

С помощью замкнутых и полных групп выполняется спектрально-корреляционный анализ. Полные группы позволяют выявить корреляционные связи между операторами, замкнутые группы - между полными группами. Если множество операторов – алфавит, то множества групп – более сложные грамматические описания наблюдаемого сигнала: полная группа – слово, замкнутая группа – словосочетание.

Используя спектральное представление сигнала μ , формируется множество операторов описывающий данный сигнал, а затем множества полных и замкнутых групп:

$$\begin{aligned} V &= GV[\mu], P_{na} = GP_{na}[\mu, V], P_{nm} = GP_{nm}[\mu, V], \\ P_s &= GP_s[\mu, V, P_{na}, P_{nm}], P_c = GP_c[\mu, V, P_{na}, P_{nm}], \end{aligned} \quad (2.3)$$

где GV – оператор вычисления по спектральному представлению сигнала признакового описания V на основе операторов;

GP_{na} (GP_{nm}) – на основе полных групп на операции сложения P_{na} (умножения, P_{nm});

GP_c (GP_s) – на основе замкнутых групп P_s (замкнутых множеств, P_c).

Матрицы вероятностей переходов - это модель признакового описания, в которой учитываются связи между соседними сегментами сигнала. Метод вычисления признакового описания в таком случае состоит в формировании матрицы вероятностей переходов между описаниями соседних сегментов.

Предлагаются следующие системы признаков, основанные на матрицах вероятностей переходов:

1. Система признаков PVI, описывающая вероятности переходов между значениями операторов (оператор может принимать три возможных значения: прямое, инверсное и равное нулю), вычисленными по соседним сегментам сигнала, без учёта связей между различными операторами, размерность пространства признаков – $3 \times 3 \times 15$ (рис. 2.4).
2. Система признаков PVD, описывающая вероятности переходов между значениями операторов, вычисленными по соседним сегментам сигнала, с учётом связи между операторами, размерность пространства признаков – 45×45 (рис. 2.5).
3. Система признаков PVS, описывающая вероятности переходов между описаниями сегментов, представленных в виде полных групп, размерность пространства признаков – 140×140 ; при использовании полных групп допустимо использовать только несколько максимальных по сумме отсчётов, находящихся под их образами,

групп (граф переходов для полных групп подобен графу переходов для операторов, рис. 2.5).

4. Система признаков, описывающая вероятности переходов между описаниями сегментов, представленных в виде замкнутых групп, размерность пространства признаков – 840×840 ; при использовании замкнутых групп допустимо использовать группы, сумма отсчётов, под образом которых максимальна.

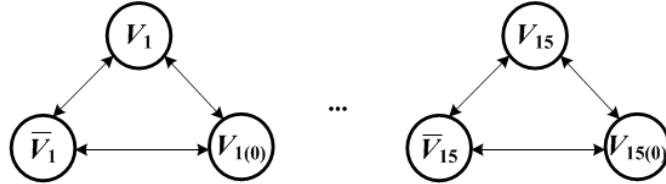


Рисунок 2.4 - Графы переходов между операторами (без учёта связей между операторами)

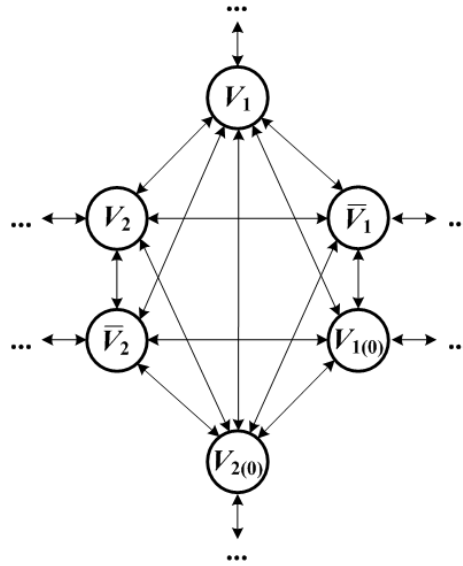


Рисунок 2.5 - Графы переходов между операторами (с учётом связей между операторами)

Алгоритм формирования системы признаков на основе полных (замкнутых) групп, D_i – описание i -го сегмента сигнала в виде полных (замкнутых) групп:

$$\begin{aligned}
 \forall i = \overline{1, (N-1)} \\
 \forall k = \overline{1, |D_i|} \\
 \forall l = \overline{1, |D_{i+1}|} \\
 MPG_{D_i}[D_i[k], D_{i+1}[l]] = MPG_{D_i}[D_i[k], D_{i+1}[l]] + 1.
 \end{aligned} \tag{2.4}$$

Далее будут использоваться следующие обозначения PP_{na} (PP_{nm}) – матрица вероятностей переходов между полными группами на операции сложения (умножения), PP_s – на основе замкнутых групп, PP_{ca} – на основе замкнутых множеств на операции сложения, PP_{cm} – на основе замкнутых множеств на операции умножения.

При использовании в качестве систем признаков матриц вероятностей переходов, между описаниями сегментов можно учитывать вероятности переходов не только между i и $(i + 1)$ сегментом сигнала, но и учитывать связи между большим числом сегментов. Для i -го сегмента возможен учёт не только $(i + 1)$, $(i + 2)$ и дальнейших сегментов, но и $(i - 1)$, $(i - 2)$ сегментов, т.е. не только «будущего», но и «прошлого».

Пример формирования признакового описания сигнала в виде матрицы вероятностей переходов размером 140×140 элементов для полных групп показан на рис. 2.6. При вычислении значений матрицы рассматривались связи между описаниями только пары соседних сегментов.

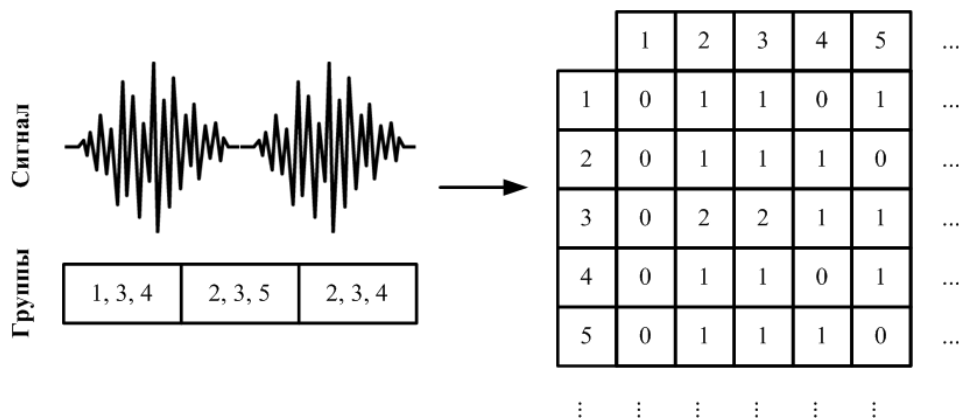


Рисунок 2.6 - Матрица вероятностей переходов между группами

Пример формирования признакового описания сигнала в виде 15 независимых матриц вероятностей переходов для операторов на показан рис. 2.7. При вычислении значений матрицы рассматривались связи между описаниями только пары соседних сегментов и учитываются вероятности переходов только для одного оператора.

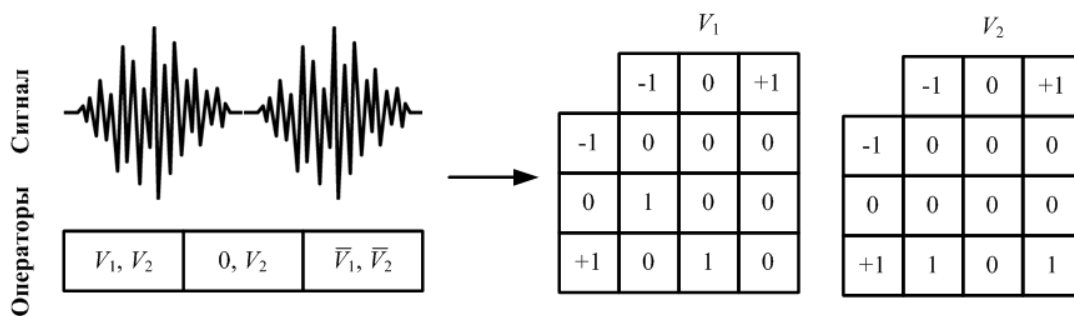


Рисунок 2.7 - Матрицы вероятностей переходов между операторами

Пример формирования признакового описания сигнала в виде матрицы вероятностей переходов для операторов размером 45×45 показан на рис. 2.8. При вычислении значений матрицы рассматривались связи между описаниями только пары соседних сегментов. В описаниях сегментов приведены значения только операторов V_1 и V_2 .

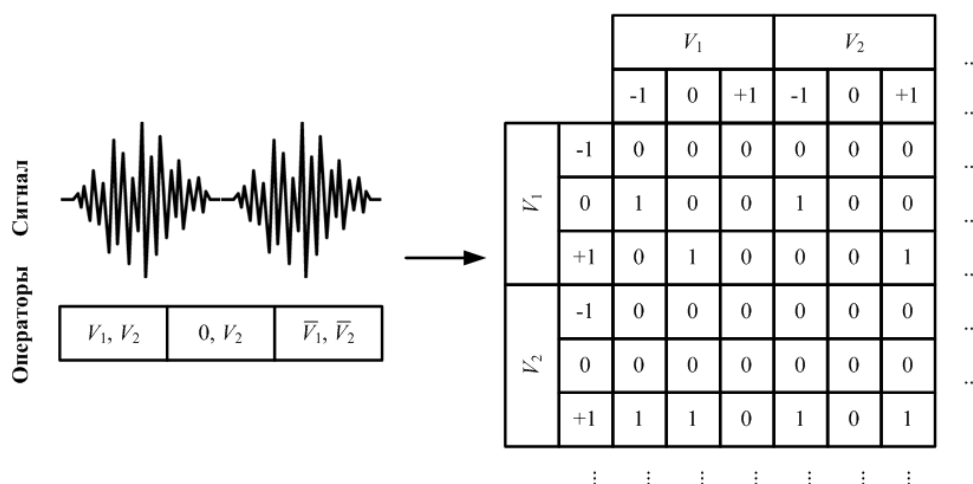


Рисунок 2.8 - Матрица вероятностей переходов между операторами

Сравнивая ТАВ с известными подходами к формированию признакового описания сигнала можно отметить следующее:

- 1) по сравнению с вейвлет-преобразованием и преобразованием Фурье ТАВ позволяет вычислять, по отношению к спектральным коэффициентам, признаки более высокого уровня (за счёт использования алгебры групп);
- 2) по сравнению с моделями глубокого обучения в ТАВ признаковое описание вычисляется без использования обучения, а по заранее заданным шаблонам;
- 3) при вычислении U -преобразования используются только операции сложения и вычитания.

Этап принятия решения основан на алгоритме композиции случайных деревьев принятия решения. Композиция - это объединение нескольких алгоритмов в один. Идея заключается в том, чтобы обучить алгоритмы, а затем усреднить полученные от них ответы. Чтобы построить композицию, нужно сначала обучить N базовых алгоритмов, причем их нельзя обучать на всей обучающей выборке, так как в этом случае они получаются одинаковыми.

Один из способов сделать базовые алгоритмы различными - использовать рандомизацию - обучать базовые алгоритмы на разных подвыборках обучающей выборки. В нашем случае рандомизация достигается двумя способами: путем генерации случайного подмножества обучающей выборки и случайного подмножества признаков из всех существующих.

При этом производится выборка случайных двух третей наблюдений для обучения, а оставшаяся треть используется для оценки результата. Эти операции проделываем сотни раз. Результирующая модель будет получена результатом «голосования» набора полученных при моделировании деревьев.

Исходя из вышеописанного, составим алгоритм принятия решения:

1. Получаем N случайных подвыборок.

2. Каждая получившаяся подвыборка используется как обучающая выборка для построения соответствующего решающего дерева. Причем:

2.1. Дерево строится, пока в каждом листе окажется не более определенного числа объектов. Чем меньше объектов в каждом листе, тем получаются более сложные и переобученные решающие деревья с низким смещением.

2.2. Процесс построения дерева рандомизирован: на этапе выбора оптимального признака, по которому будет происходить разбиение, он ищется не среди всего множества признаков, а среди случайного подмножества.

2.3. Случайное подмножество выбирается заново каждый раз, когда необходимо разбить очередную вершину.

4. Построенные деревья объединяются в композицию: для нашей задачи регрессии берется усредненное значение результатов работы всех алгоритмов.

Главные преимущества композиции случайных деревьев заключаются в следующем:

1. Они не переобучаются при росте числа базовых алгоритмов.
2. При отсутствии корреляции между отдельно взятыми деревьями алгоритм может дать идеальный результат, так как разброс ответов всей композиции равен разбросу отдельного дерева деленного на число деревьев.

2.2 Вычислительный эксперимент

2.2.1 Описание набора данных

Все исходные данные разбиваются на 3 части:

1. Обучающая выборка.
2. Выборка для тестирования полученной модели.
3. Итоговая тестовая выборка.

Обучающая выборка строится, исходя из следующего принципа: в первых двух частях исходных данных мы вычисляем количество одинаковых объектов каждого вида и для обучения берем пропорциональное этому число. Приведем поясняющую таблицу.

Таблица 2.1 – Принцип построения обучающей выборки

Общее число объектов	Число объектов для обучения
От 1 до 4	1
От 5 до 10	2
От 11 до 14	3
От 15 до 20	4
От 21 до 25	5

Проиллюстрируем на графике (см. рис. 2.9).

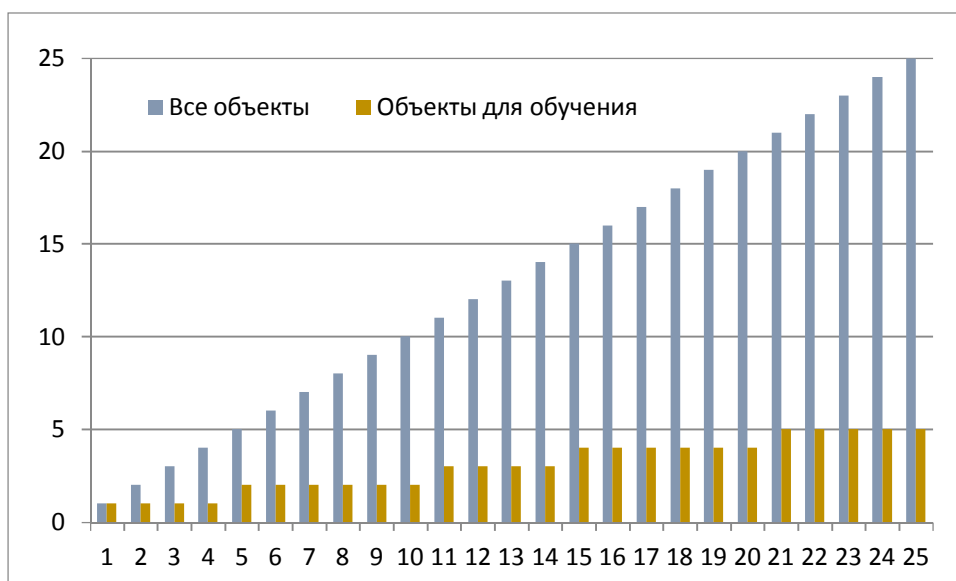


Рисунок 2.9 – Принцип построения обучающей выборки

Таким образом, чем чаще встречается объект определенного вида, тем больше он принимает участия в обучении. Другими словами, чем чаще у человека встречается определенный уровень содержания в крови глюкозы, тем чаще он встречается обучающей выборке.

Итоговое тестирование полученной модели проводится по третьей части данных, по независимой выборке. В ней могут встречаться значения глюкозы, на которых не проводилось обучение.

2.2.2 Описание методики тестирования

Для проверки полученных результатов используется несколько методов: диаграмма Кларка, коэффициент корреляции Спирмена и среднеквадратичное отклонение. Рассмотрим каждый из них.

2.2.2.1 Диаграмма Кларка

Диаграмма Кларка используется для оценки клинической значимости значений разности между значениями глюкозы, полученными точным лабораторным методом и значениями, полученными с помощью экспериментального прибора. Диаграмма разделена на зоны, каждая из которых соответствует степени тяжести ошибок измерения.

Вид диаграммы представлен на рис. 2.10.

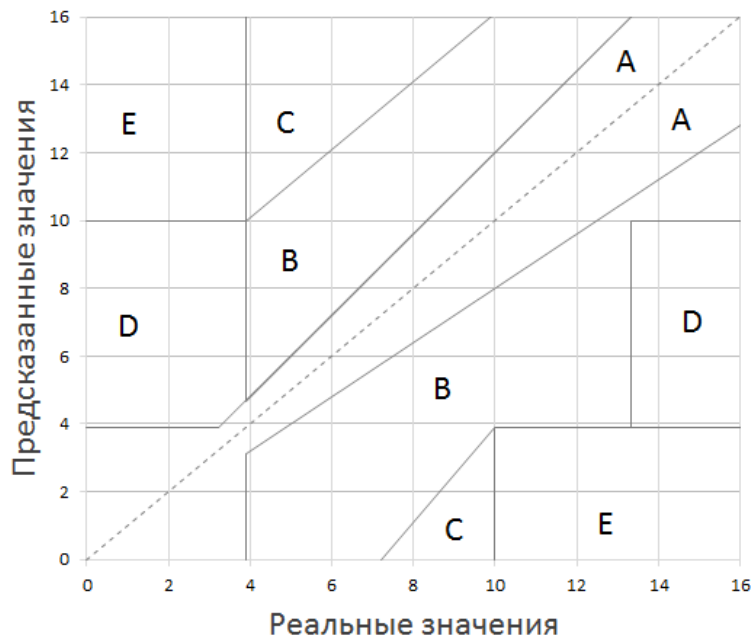


Рисунок 2.10 – Диаграмма Кларка

Рассмотрим каждую из зон:

- А - зона отсутствия ошибок. Она представляет собой значения глюкозы, которые отклоняются от эталонного не больше, чем на 20%. Значения, попадающие в этот диапазон, приведут к клинически правильным решениям относительно лечения.
- В - зона несущественных ошибок. Она представляет значения, которые отклоняются от эталонных больше, чем на 20%, но приведут к безвредному лечению или отсутствию его на основе наших предположений.
- С - зона существенных ошибок. Значения приведут к переходу границы приемлемых уровней глюкозы в крови; такое лечение может привести к фактическому падению глюкозы в крови до уровня ниже 70 мг/дл или повышению его выше 180 мг/дл, следовательно, это нанесет вред больному.
- D - зона опасных ошибок. Она представляет собой опасный отказ для выявления и лечения ошибок. Фактические значения глюкозы находятся вне целевого диапазона, но значения для пациента генерируются в пределах целевого диапазона.
- Е - зона жизненно опасных ошибок. Она является зоной ошибочного лечения. Генерируемые значения в пределах этой зоны противоположны эталонным значениям, и, следовательно, соответствующие решения в отношении лечения будут противоположным необходимым.

Таким образом, значения, содержащиеся в зонах А и В являются клинически приемлемыми, в то время как значения в зонах С, D, и Е являются потенциально опасными и, следовательно, являются клинически значимыми ошибками.

Пример диаграммы Кларка в работе представлен на рис. 2.11.

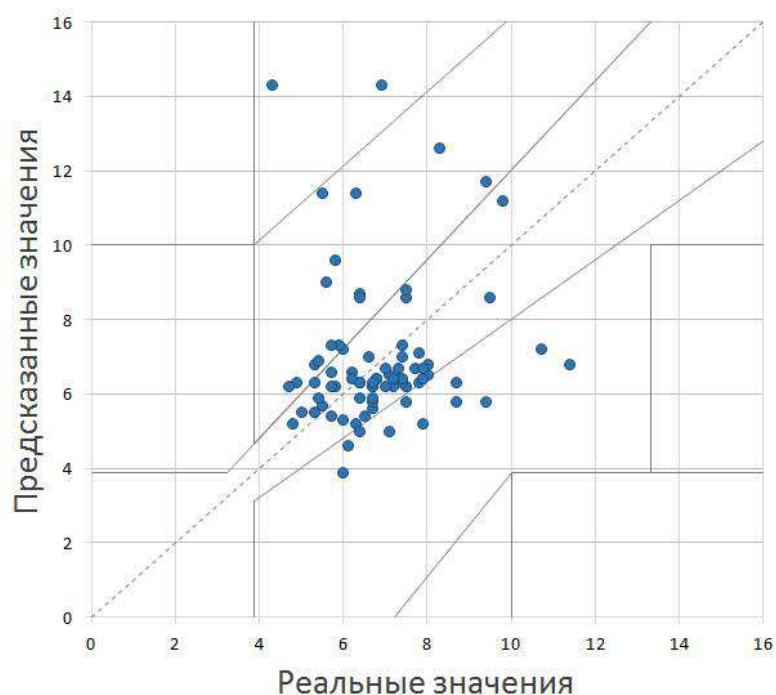


Рисунок 2.11 – Пример работы диаграммы Кларка

Здесь, как мы видим, большая часть предсказанных значений попадает в зоны А и В, то есть в зону отсутствия ошибок и зону несущественных ошибок, но при этом два значения находятся в зоне С, а значит говорят о существенной ошибке в предсказании, которая может нанести вред больному.

Если посмотреть на значения этих двух точек (табл. 2.2), то мы видим, что значения отклоняются от правильных более, чем в два раза.

Таблица 2.2 – Значения глюкозы, попавшие в зону С

Значение глюкометра	Значение алгоритма	Разность значений	Зона
6,9	14,3	7,4	С
4,3	14,3	10,0	С

2.2.2.2 Коэффициент корреляции Спирмена

Корреляция – это статистическая взаимосвязь нескольких случайных величин, либо величин, которые можно с некоторой допустимой степенью точности считать таковыми. Математической мерой корреляции двух случайных величин служит коэффициент корреляции. Это величина в диапазоне от -1 до +1, которая характеризует степень связи величин. Значение +1 говорит о том, что при увеличении одной переменной увеличивается значение другой переменной, при -1 наоборот.

Значения коэффициента корреляции можно интерпретировать по шкале Чеддока, которая характеризует показатели тесноты связи между двумя величинами. Шкала показана в табл. 2.3.

Таблица 2.3 – Шкала Чеддока

Величина коэффициента корреляции	0.1 – 0.3	0.3 – 0.5	0.5 – 0.7	0.7 – 0.9	0.9 – 1.0
Характеристика силы связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

Средняя

Сильная

Есть несколько видов коэффициентов корреляции. Рассмотрим их и выберем наиболее подходящий для нашей задачи.

1. Коэффициент Фехнера.

Зависит только от знаков отклонений величин от своих средних значений и не учитывает их величину, он характеризует не столько тесноту связи, сколько ее наличие и направление.

2. Метод наименьших квадратов.

Метод чувствителен к выбросам. Требуется нормальное распределение. При попытке описать изучаемое явление с помощью математического уравнения, прогноз будет точен для небольшого периода времени и уравнение регрессии следует пересчитывать по мере поступления новой информации.

3. Регрессионный анализ.

Выбросы увеличивают стандартную ошибку коэффициента, снижают значение R-квадрат.

4. Коэффициент корреляции Пирсона.

Подходит для оценки взаимосвязи между нормальными распределениями переменных. Не очень устойчив к выбросам - при их наличии можно ошибочно сделать вывод о наличии корреляции между переменными.

5. Коэффициент ранговой корреляции Спирмена.

Непараметрический аналог коэффициента корреляции Пирсона. Подходит, если распределение исследуемых переменных отличается от нормального, или возможны выбросы. Позволяет выявлять не только линейные связи, а любые, которые могут быть описаны монотонной функцией.

6. Коэффициент Кендалла.

Он дает несколько более строгую оценку связи, нежели коэффициент Спирмена. Считается более информативным. Обычно коэффициент Кендалла меньше коэффициента Спирмена.

7. Критерий Ширахатэ.

Является аналогом критерия значимости ранговой корреляции Спирмена, но более эффективен для малых выборок.

Для того чтобы выбрать коэффициент корреляции, нужно определиться с некоторыми моментами:

1. Имеются ли выбросы.
2. Нормальное ли распределение величин.

Поскольку результат работы алгоритма существенно зависит от того, как был снят ЭКГ-сигнал (не шевелился ли пациент, плотно ли прижал палец к чехлу и т.д.), будем считать, что в работе алгоритма возможны выбросы.

Теперь проверим распределение на нормальность. Для этого воспользуемся показателями асимметрии и эксцесса.

Асимметрия (А) - это мера несимметричности графика плотности реального распределения в сравнении с нормальным распределением. Эксцесс (Е) - мера вытянутости графика плотности реального распределения в сравнении с нормальным распределением.

По численным значениям асимметрии и эксцесса можно приближенно оценить нормальность распределения результатов испытаний.

А и Е рассчитываем по формулам (2.5) и (2.6)

$$A = \frac{1}{ns^3} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (2.5)$$

$$E = \frac{1}{ns^4} \sum_{i=1}^n (x_i - \bar{x})^4 - 3 \quad (2.6)$$

Или при помощи статистических функций в Microsoft Excel СКОС (для А) и ЭКСЦЕСС (для Е). Далее рассчитаем дисперсию А и Е с помощью формул (2.7) и (2.8).

$$D(A) = \frac{6(n-1)}{(n+1)(n+3)} \quad (2.7)$$

$$D(E) = \frac{24(n-2)(n-3)n}{(n-1)^2(n+3)(n+5)} \quad (2.8)$$

Где n – число испытаний, в нашем случае число измерений глюкозы.

Если формулы (2.9) и (2.10) верны, то распределение считается нормальным.

$$|A| \leq 3\sqrt{D(A)} \quad (2.9)$$

$$|E| \leq 5\sqrt{D(E)} \quad (2.10)$$

Теперь произведем расчеты для всех пациентов. Результаты занесем в табл. 2.4.

Таблица 2.4 – Результаты проверки типа распределения

	Пациенты		
	1003	1430	1696
n	104	370	211
A	1.187	0.306	0.242
E	2.559	-0.264	0.531
D(A)	0.055	0.016	0.028
D(E)	0.208	0.063	0.108
A	0.704	0.379	0.499
E	2.279	1.255	1.643
Распределение	Ненормальное	Нормальное	Нормальное

Видим, что у значений глюкозы пациента 1003 распределение отлично от нормального, у 1430 нормальное, но близко к границе, у 1696 нормальное. Для того чтобы оценка результатов была одинаковой для всех пациентов мы выберем коэффициент корреляции, в котором возможно ненормальное распределение значений, а также возможны выбросы. Этим условиям удовлетворяет коэффициент корреляции Спирмена.

2.2.3 Результаты вычислительного эксперимента

Рассмотрим результаты работы алгоритма по трем пациентам. По каждому представим проценты зон А-Е диаграммы Кларка, графическое представление диаграммы, значения коэффициента корреляции Спирмена, разницу среднеквадратического отклонения значений, полученных от инвазивного глюкометра и значений, полученных с помощью алгоритма, график соответствия реальных и предсказанных значений глюкозы. Для каждого пациента рассмотрим результат, полученный на основе трех систем признаков, PVI, PVD и PVS, представленных в пункте 3.3, и произведем сравнение результатов. В скобках около каждой системы признаков указаны параметра классификатора, которые дали лучший результат.

2.2.3.1 Пациент 1003

Система признаков PVI (Nodesize = 2, Mtry = 75, NumRandomCuts = 45).

Обобщенные результаты представлены в табл. 2.5.

Таблица 2.5 – Результаты работы алгоритма пациента 1003, PVI

Зона А	Зона В	Зона С	Зона D	Зона Е	Спирмен	SD _{гл} -SD _{ал}
97.33%	2.67%	0.00%	0.00%	0.00%	0.807	0.132

График соответствия реальных и предсказанных значений глюкозы показан на рис. 2.12, диаграмма ошибок Кларка - на рис. 2.13.

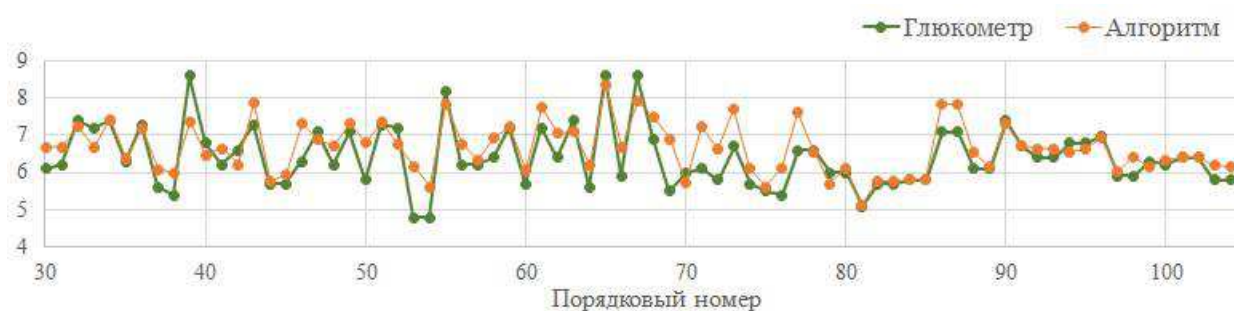


Рисунок 2.12 – Реальные и предсказанные значения глюкозы пациента 1003, PVI

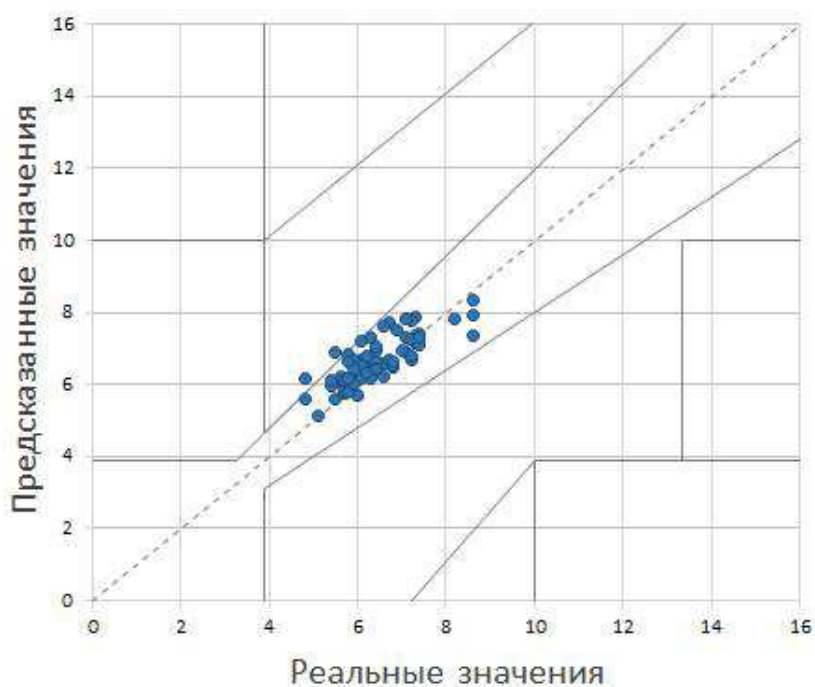


Рисунок 2.13 – Диаграмма ошибок Кларка пациента 1003, PVI

Система признаков PVD (Nodesize = 1, Mtry = 55, NumRandomCuts = 15).

Обобщенные результаты представлены в табл. 2.6.

Таблица 2.6 – Результаты работы алгоритма пациента 1003, PVD

Зона А	Зона В	Зона С	Зона D	Зона Е	Спирмен	$SD_{\text{гл}}-SD_{\text{ал}}$
96.00%	4.00%	0.00%	0.00%	0.00%	0.716	0.296

График соответствия реальных и предсказанных значений глюкозы показан на рис. 2.14, диаграмма ошибок Кларка - на рис. 2.15.

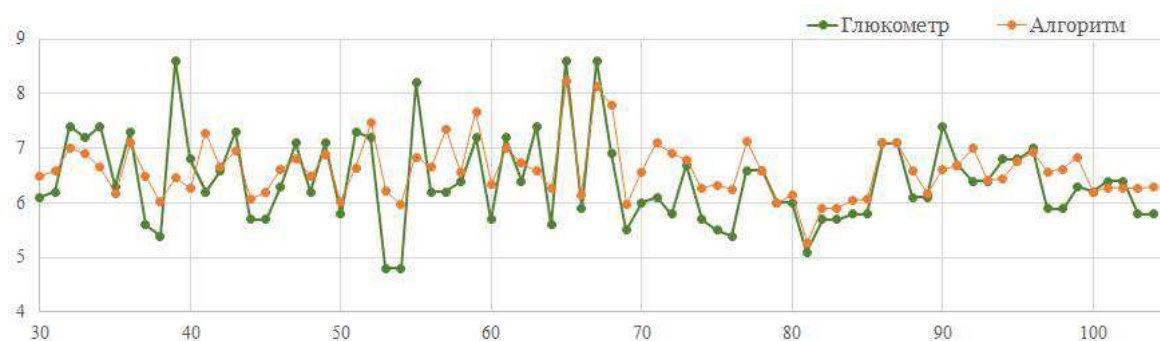


Рисунок 2.14 – Реальные и предсказанные значения глюкозы пациента 1003, PVD

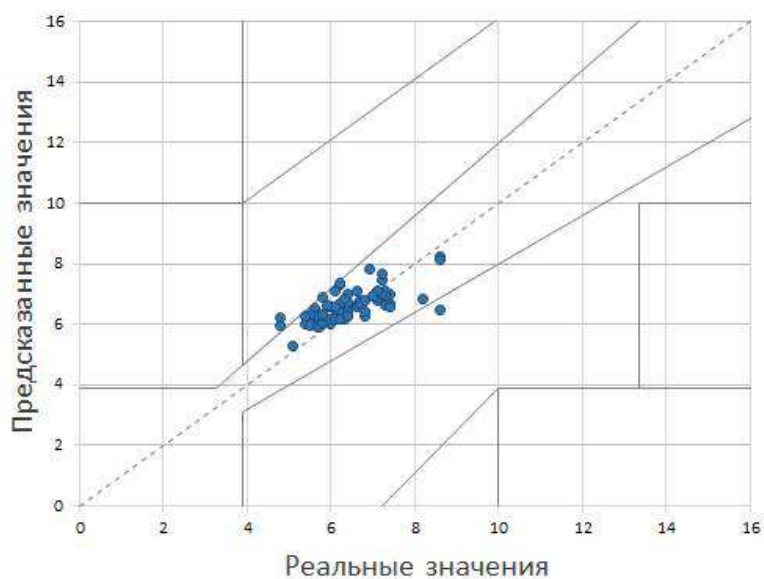


Рисунок 2.15 – Диаграмма ошибок Кларка пациента 1003, PVD

Система признаков PVS (Nodesize = 1, Mtry = 65, NumRandomCuts = 30).

Обобщенные результаты представлены в табл. 2.7.

Таблица 2.7 – Результаты работы алгоритма пациента 1003, PVS

Зона А	Зона В	Зона С	Зона D	Зона Е	Спирмен	$SD_{\text{гл}}-SD_{\text{ал}}$
97.33%	2.67%	0.00%	0.00%	0.00%	0.741	0.329

График соответствия реальных и предсказанных значений глюкозы показан на рис. 2.16, диаграмма ошибок Кларка - на рис. 2.17.

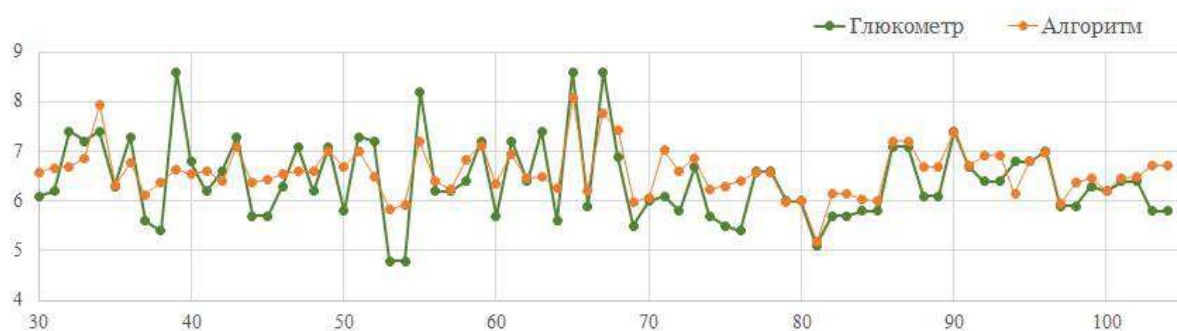


Рисунок 2.16 – Реальные и предсказанные значения глюкозы пациента 1003, PVS

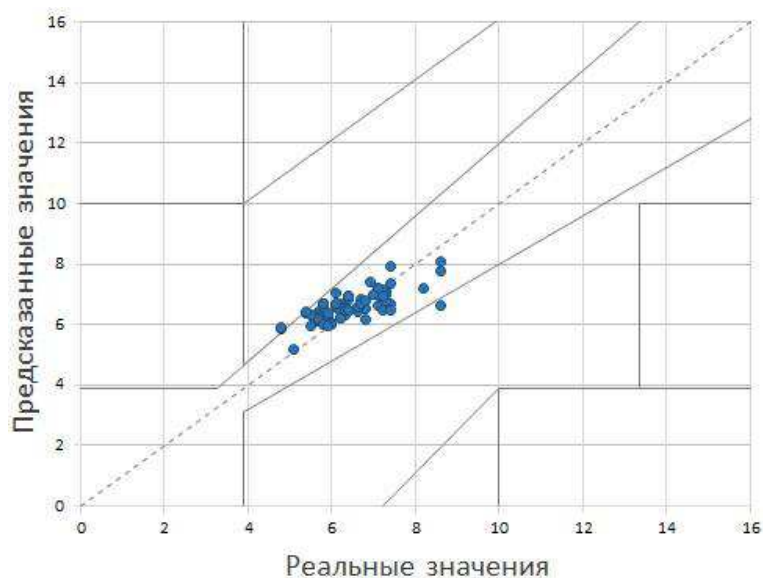


Рисунок 2.17 – Диаграмма ошибок Кларка пациента 1003, PVS

Сравнение результатов по трем системам признаков пациента 1003 приведены в табл. 2.8.

Таблица 2.8 – Сравнение систем признаков пациента 1003

Спирмен	$SD_{гл}-SD_{ал}$	Зона А	Зона В	Зона С	Зона D
PVI					
0.807	0.132	97.33%	2.67%	0.00%	0.00%
PVD					
0.716	0.296	96.00%	4.00%	0.00%	0.00%
PVS					
0.741	0.329	97.33%	2.67%	0.00%	0.00%

2.2.3.2 Пациент 1430

Система признаков PVI (Nodesize = 1, Mtry = 15, NumRandomCuts = 35).

Обобщенные результаты представлены в табл. 2.9.

Таблица 2.9 – Результаты работы алгоритма пациента 1430, PVI

Зона А	Зона В	Зона С	Зона D	Зона Е	Спирмен	SD _{гл} -SD _{ал}
64.98%	33.18%	1.84%	0.00%	0.00%	0.64	1.456

Графики соответствия реальных и предсказанных значений глюкозы показаны на рис. 2.18, диаграмма ошибок Кларка - на рис. 2.19.

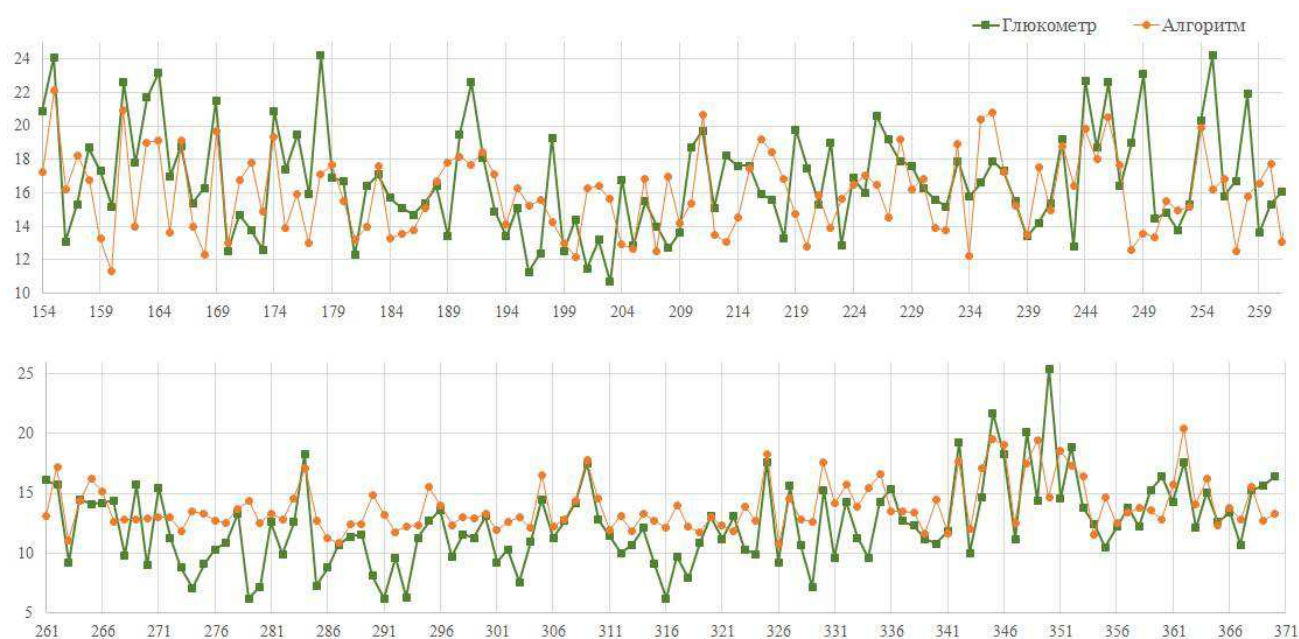


Рисунок 2.18 – Реальные и предсказанные значения глюкозы пациента 1430, PVI

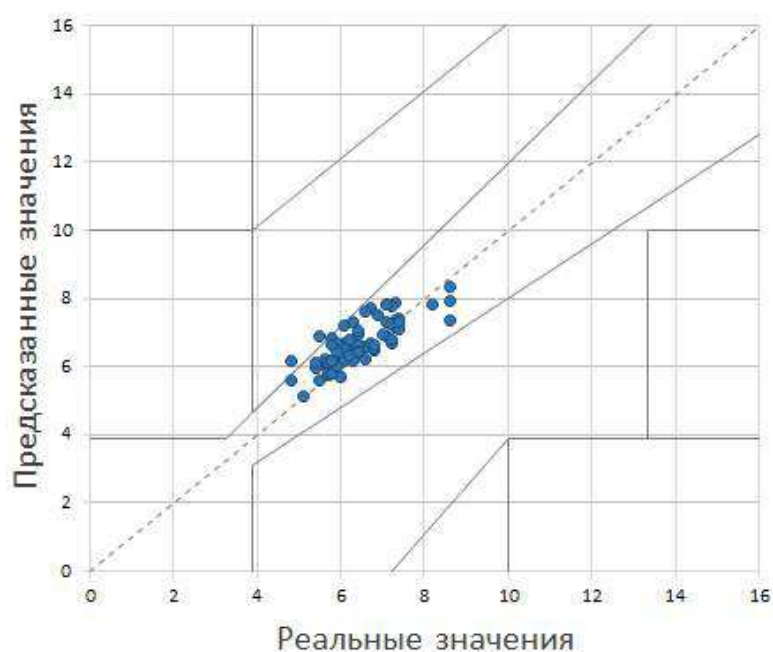


Рисунок 2.19 – Диаграмма ошибок Кларка пациента 1430, PVI

Система признаков PVD (Nodesize = 2, Mtry = 35, NumRandomCuts = 15).

Обобщенные результаты представлены в табл. 2.10.

Таблица 2.10 – Результаты работы алгоритма пациента 1430, PVD

Зона А	Зона В	Зона С	Зона D	Зона Е	Спирмен	SD _{гл} -SD _{ал}
65.44%	31.80%	2.76%	0.00%	0.00%	0.60	1.409

Графики соответствия реальных и предсказанных значений глюкозы показаны на рис. 2.20, диаграмма ошибок Кларка - на рис. 2.21.

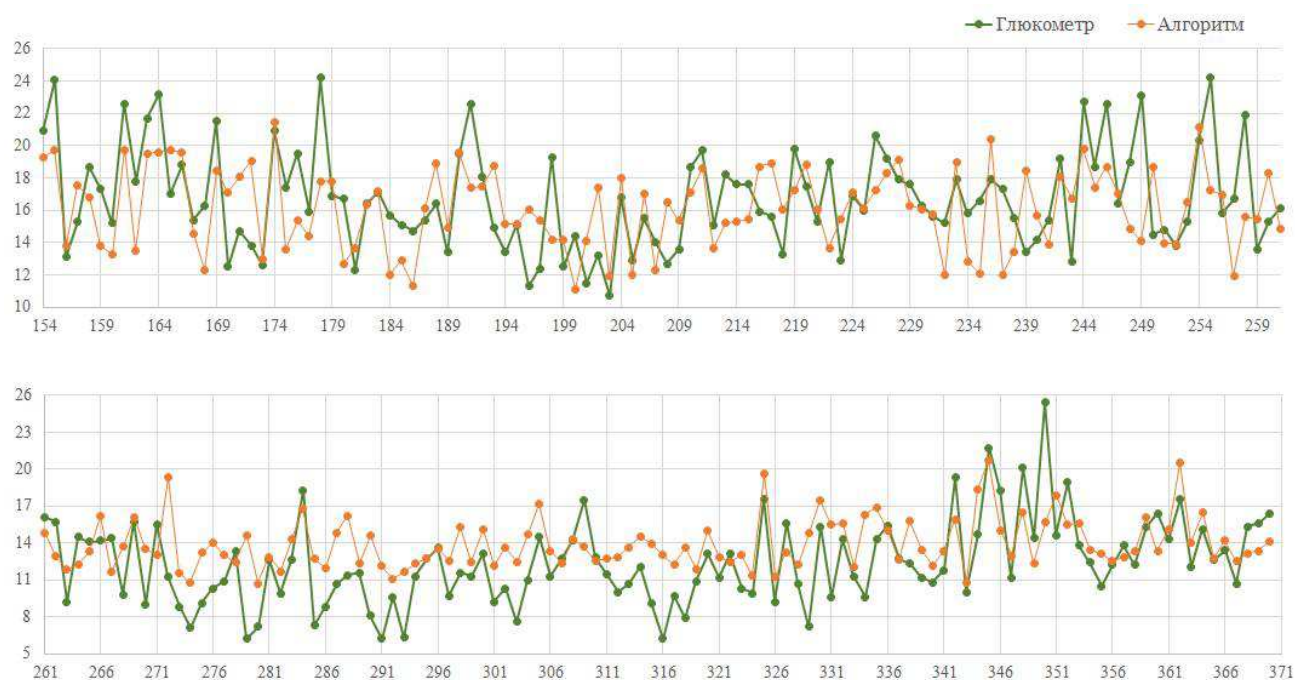


Рисунок 2.20 – Реальные и предсказанные значения глюкозы пациента 1430, PVD

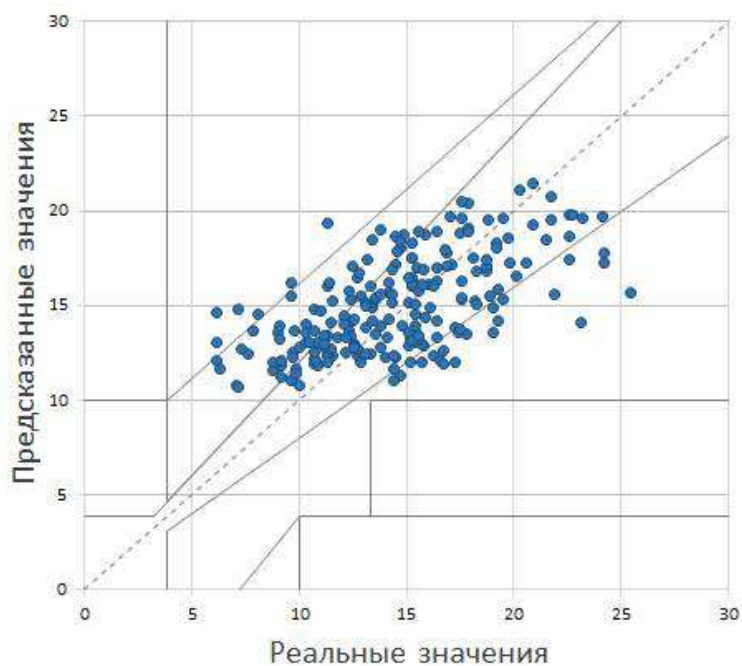


Рисунок 2.21 – Диаграмма ошибок Кларка пациента 1430, PVD

Система признаков PVS (Nodesize = 1, Mtry = 20, NumRandomCuts = 25).

Обобщенные результаты представлены в табл. 2.11.

Таблица 2.11 – Результаты работы алгоритма пациента 1430, PVS

Зона А	Зона В	Зона С	Зона D	Зона E	Спирмен	SD _{гл} -SD _{ал}
65.44%	32.72%	1.84%	0.00%	0.00%	0.62	1.328

Графики соответствия реальных и предсказанных значений глюкозы показаны на рис. 2.22, диаграмма ошибок Кларка - на рис. 2.23.

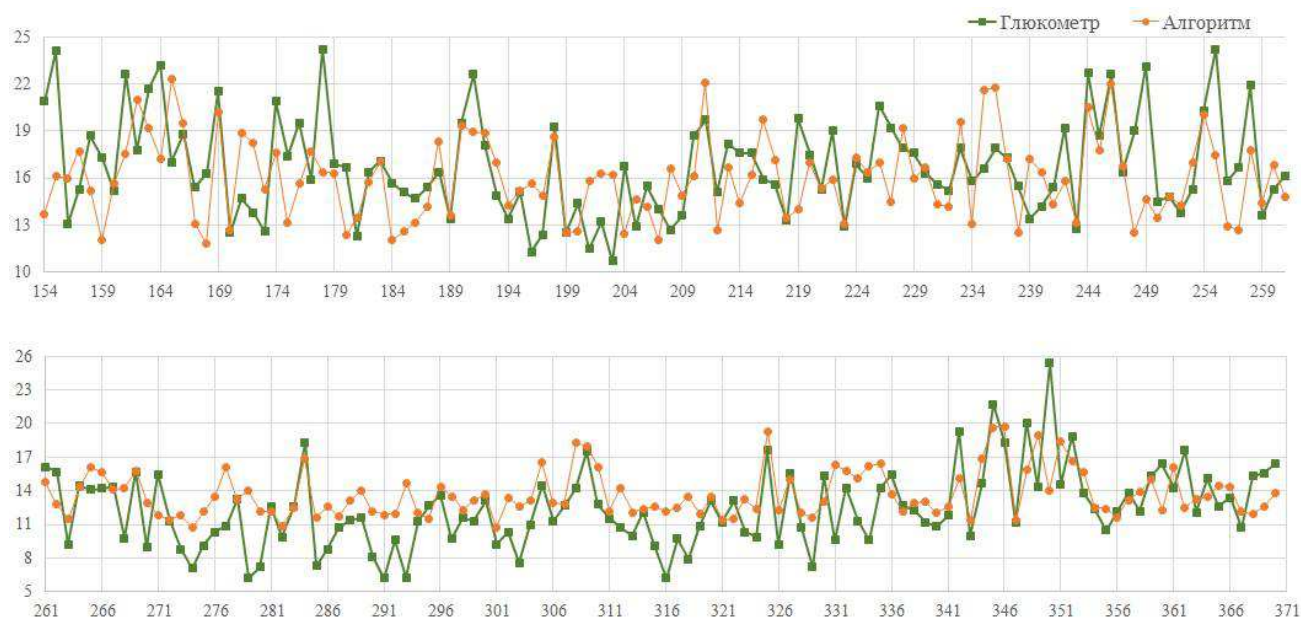


Рисунок 2.22 – Реальные и предсказанные значения глюкозы пациента 1430, PVS

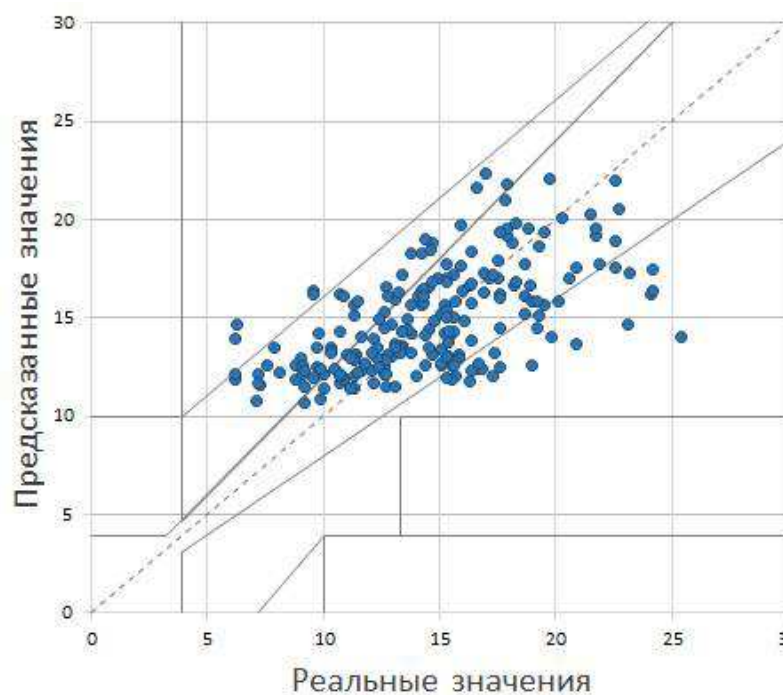


Рисунок 2.23 – Диаграмма ошибок Кларка пациента 1430, PVS

Сравнение результатов по трем системам признаков пациента 1430 приведены в табл. 2.12.

Таблица 2.12 – Сравнение систем признаков пациента 1430

Спирмен	$SD_{гл}-SD_{ал}$	Зона А	Зона В	Зона С	Зона D
PVI					
0.64	1.456	64.98%	33.18%	1.84%	0.00%
PVD					
0.60	1.409	65.44%	31.80%	2.76%	0.00%
PVS					
0.62	1.328	65.44%	32.72%	1.84%	0.00%

2.2.3.3 Пациент 1696

Система признаков PVI (Nodesize = 2, Mtry = 35, NumRandomCuts = 10).

Обобщенные результаты представлены в табл. 2.13.

Таблица 2.13 – Результаты работы алгоритма пациента 1696, PVI

Зона А	Зона В	Зона С	Зона D	Зона Е	Спирмен	$SD_{гл}-SD_{ал}$
91.67%	8.33%	0.00%	0.00%	0.00%	0.50	1.159

График соответствия реальных и предсказанных значений глюкозы показан на рис. 2.24, диаграмма ошибок Кларка - на рис. 2.25.

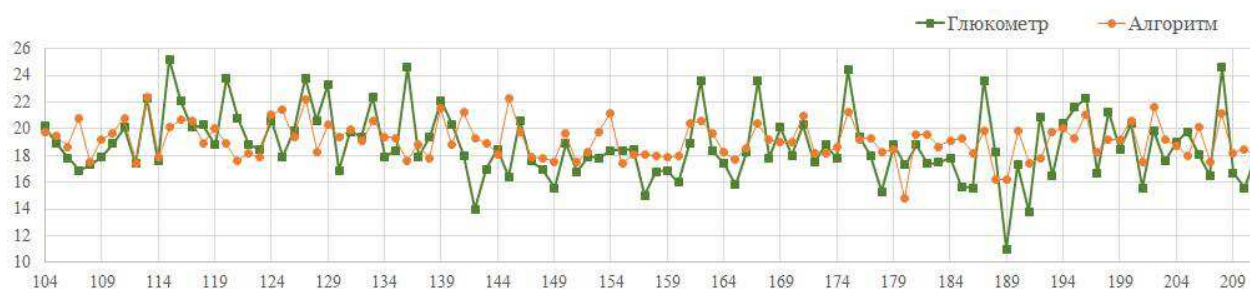


Рисунок 2.24 – Реальные и предсказанные значения глюкозы пациента 1696, PVI

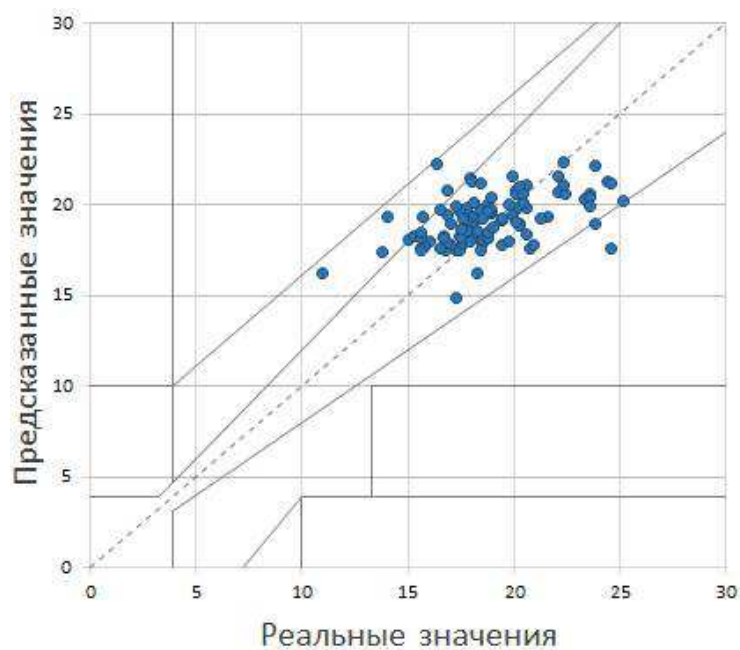


Рисунок 2.25 – Диаграмма ошибок Кларка пациента 1696, PVI Система признаков PVD (Nodesize = 2, Mtry = 65, NumRandomCuts = 20).
Обобщенные результаты представлены в табл. 2.14.

Таблица 2.14 – Результаты работы алгоритма пациента 1696, PVD

Зона А	Зона В	Зона С	Зона D	Зона Е	Спирмен	SD _{гл} -SD _{ал}
90.74%	7.41%	1.85%	0.00%	0.00%	0.47	0.972

График соответствия реальных и предсказанных значений глюкозы показан на рис. 2.26, диаграмма ошибок Кларка - на рис. 2.27.

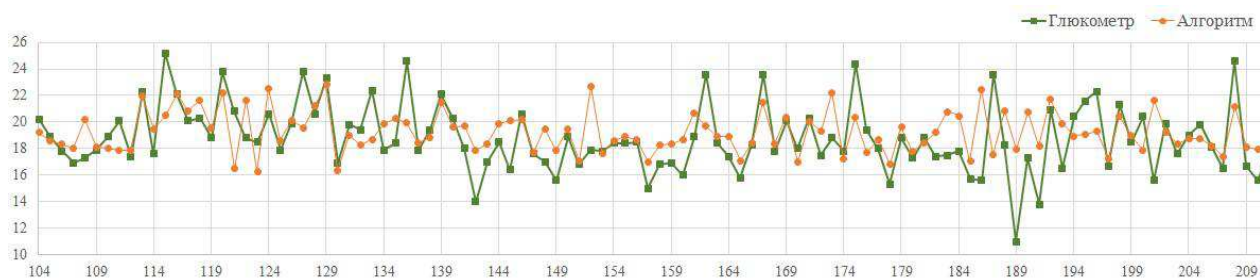


Рисунок 2.26 – Реальные и предсказанные значения глюкозы пациента 1696, PVD

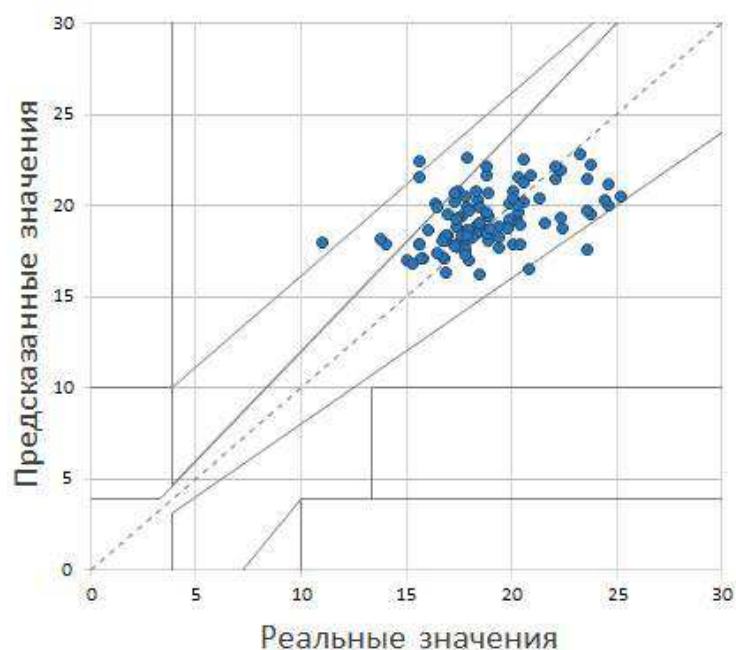


Рисунок 2.27 – Диаграмма ошибок Кларка пациента 1696, PVD

Система признаков PVS (Nodesize = 2, Mtry = 50, NumRandomCuts = 30)

Обобщенные результаты представлены в табл. 2.15.

Таблица 2.15 – Результаты работы алгоритма пациента 1696, PVS

Зона А	Зона В	Зона С	Зона D	Зона Е	Спирмен	SD _{гл} -SD _{ал}
90.74%	7.41%	1.85%	0.00%	0.00%	0.46	1.789

График соответствия реальных и предсказанных значений глюкозы показан на рис. 2.28, диаграмма ошибок Кларка - на рис. 2.29.

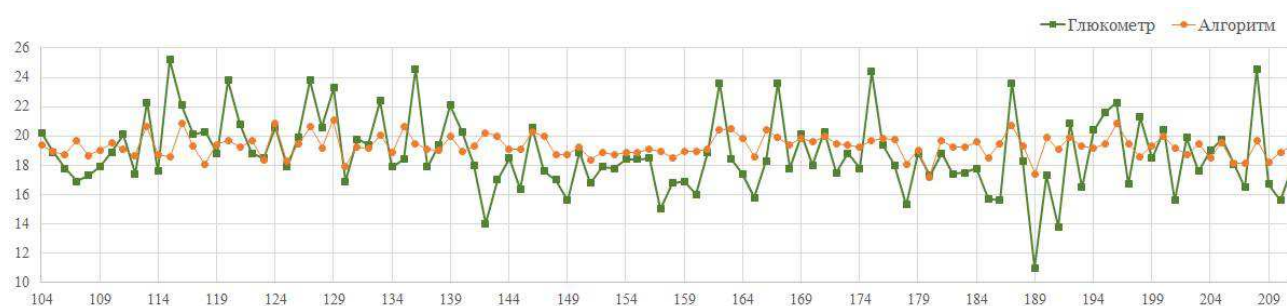


Рисунок 2.28 – Реальные и предсказанные значения глюкозы пациента 1696, PVS

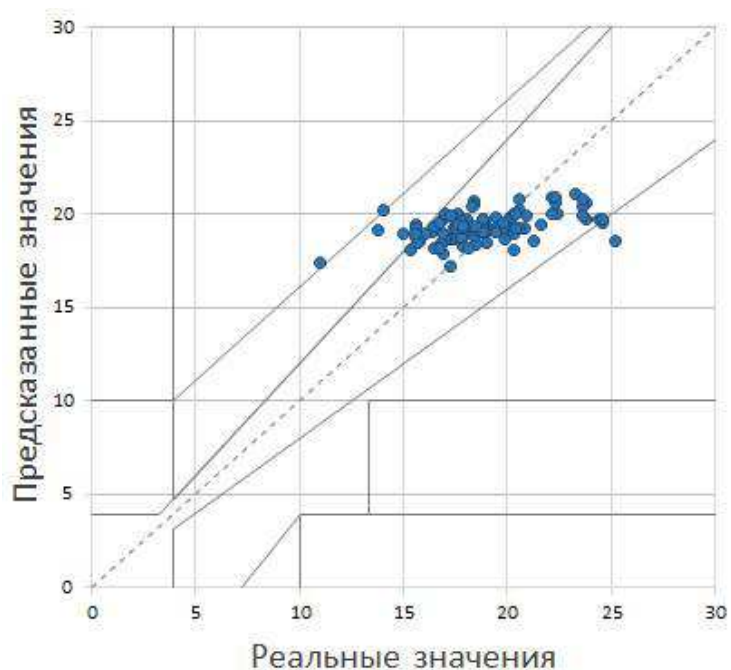


Рисунок 2.29 – Диаграмма ошибок Кларка пациента 1696, PVS

Сравнение результатов по трем системам признаков приведены в табл. 2.16.

Таблица 2.16 – Сравнение систем признаков пациента 1430

Спирмен	$SD_{гл}-SD_{ал}$	Зона А	Зона В	Зона С	Зона D
PVI					
0.50	1.159	91.67%	8.33%	0.00%	0.00%
PVD					
0.47	0.972	90.74%	7.41%	1.85%	0.00%
PVS					
0.46	1.789	90.74%	7.41%	1.85%	0.00%

Сведем лучшие из получившихся результатов в табл. 2.17.

Таблица 2.17 – Обобщенные результаты

Спирмен	$SD_{гл}-SD_{ал}$	Зона А	Зона В	Зона С	Зона D
Пациент 1003					
0.81	0.132	97.33%	2.67%	0.00%	0.00%
Пациент 1430					

0.62	1.328	65.44%	32.72%	1.84%	0.00%
Пациент 1696					
0.50	1.159	91.67%	8.33%	0.00%	0.00%

Сравним результаты, полученные с помощью классификатора extraTrees (в основе которого лежит алгоритм композиции случайных деревьев) с результатами, полученными ранее с помощью классификаторов xgboost (градиентный бустинг) и svm (метод опорных векторов).

Лучшие результаты, полученные ранее, представлены в табл. 2.18.

Таблица 2.18 – Результаты, полученные ранее

Спирмен	SD _{гл} -SD _{ал}	Зона А	Зона В	Зона С	Зона D
Пациент 1003					
0.72	-0.104	92.47%	7.53%	0.00%	0.00%
Пациент 1430					
0.46	0.156	59.35%	29.67%	7.72%	3.25%
Пациент 1696					
0.52	-0.131	83.49%	16.51%	0.00%	0.00%

Результаты, полученные в ходе данной работы, лучше. Рассмотрим их отдельно по каждому пациенту.

1003: Коэффициент корреляции Спирмена увеличился на 9%, результаты предсказаний в зоне В уменьшились на 5%, соответственно в зоне А они увеличились на 5%.

1430: Коэффициент корреляции Спирмена увеличился на 18%, зона D (опасных ошибок предсказания) полностью исчезла, зона С сократилась на 5.8%, зона А увеличилась на 5.3%.

1696: Коэффициент корреляции Спирмена остался почти без изменений (+ 0,02%), зона В сократилась на 8%, зона А соответственно увеличилась на 8%.

2.3 Заключение по разделу

Данный раздел отчета описывает результаты работ, проведенных в соответствии с п. 1 Календарного плана работы и соответствует требованиям п. 7.1.1.2 Технического задания на выполнение НИР.

В результате выполнения выбора и обоснования методов и средств решения поставленных задач, были достигнуты следующие результаты:

- в качестве методики выполнения предварительной обработки сигнала и формирования признаков используется теория активного восприятия;

- для решения задачи регрессии будет использован метод решающих деревьев, как обеспечивающий наивысшую точность решения задачи предсказания уровня глюкозы в крови по ЭКГ сигналу.

ЗАКЛЮЧЕНИЕ

В ходе выполнения НИР «Изучение спектра кардиографических отклонений у больных с нарушением регуляции глюкозы» были получены следующие основные научно-технические результаты в соответствии с Техническим заданием и Планом-графиком:

- 1) Аналитический обзор современной научно-технической, нормативной, методической литературы, затрагивающей научно-техническую проблему, исследуемую в рамках НИР. В результате обзора были выявлены и систематизированы основные методы инвазивной и неинвазивной оценки уровня глюкозы в крови.
- 2) Предложен подход к неинвазивной оценке уровня глюкозы в крови с позиций теории активного восприятия. Выполнен вычислительный эксперимент. Полученные результаты подтверждают возможность использования предложенного подхода. Необходимы дополнительные исследования на больших выборках данных.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Perez-Gandia C., Facchinetti A., Sparacino G., Cobelli C., Gomez E.J., Rigla M., de Leiva A., and Hernando M.E. Artificial neural network algorithm for on-line glucose prediction from continuous glucose monitoring // *Diabetes Technol Ther.* – 2010. – № 12(1). - С. 81-88
2. Mhaskar H.N., Pereverzyev S.V., van der Walt M.D. A deep learning approach to diabetic blood glucose prediction // *J. arXiv preprint arXiv:1707.05828* – 2017
3. Plis K., Bunescu R., Marling C., Shubrook J., Schwartz F. A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management // *AAAI Publications, Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence.* – 2014
4. Bunescu R., Struble N., Marling C., Shubrook J., Schwartz F. Blood Glucose Level Prediction using Physiological Models and Support Vector Regression // *Machine Learning and Applications International Conference on. IEEE.* – 2013. – С. 135-140
5. Çalışır D., Doğanterkin E. An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier // *Expert Systems with Applications.* – 2011. – № 38(7). - С. 8311-8315
6. Worachartcheewan A., Nantasenamat C., Prasertsrithong P et al. Machine learning approaches for discerning intercorrelation of hematological parameters and glucose level for identification of diabetes mellitus // *EXCLI Journal.* – 2013. – №12. – С. 885-893
7. B.Sudharsan, M.Peeples, M.Shomali. Hypoglycemia prediction using machine learning models for patients with type 2 diabetes // *J Diabetes Sci Technol.* – 2015. – №9(1). – С. 86-90.
8. Yang Zhang, Jian-ming Zhu, Yong-bo Liang, Hong-bo Chen, Shi-min Yin and Zhen-cheng Chen. Non-invasive blood glucose detection system based on conservation of energy method // *Physiological Measurement.* – 2017. – № 38(2). - 325 с.
9. E.I.Georga, V.C. Protopappas, D. Ardigo et al. Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression // *IEEE J Biomed Health Inform.* – 2013. – №17(1). – С. 71-81.
10. Marling C., Wiley M., Bunescu R., Shubrook J., Schwartz F. Emerging Applications for Intelligent Diabetes Management // *AI Magazine.* – 2017. – №38(3). – 67 с.
11. Guo D., Zhang D., Zhang L., Lu G. Non-invasive blood glucose monitoring for diabetics by means of breath signal analysis // *Sensors and Actuators B: Chemical.* – 2012. – № 173. - С. 106-113
12. Jyoti Yadav, Asha Rani, Vijander Singh, Bhaskar Mohan Murari. Investigations on Multisensor-Based Noninvasive Blood Glucose Measurement System // *J. Med. Devices.* – 2017. – № 11(3)
13. Monte-Moreno E. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques // *Artificial Intelligence in Medicine.* – 2011. – № 53(2). –С. 127-138

14. Malik S., Khadgawat R., Anand S., Gupta S. Non-invasive detection of fasting blood glucose level via electrochemical measurement of saliva // SpringerPlus
15. Pappada, S.M., Brent, B.S., Cameron, D., Rosman, P.M. Development of a Neural Network for Prediction of Glucose Concentration in Type 1 Diabetes Patients // J Diabetes Sci Technol. – 2008. – № 2(5). - C. 792–801
16. Utrobin, V. A., Physical interpretation of the elements of image algebra // J. Advances in Physical Sciences. – 2004. – №47. - C. 1017-1032