

МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ГЛИКЕМИИ
НА ОСНОВЕ СМЕШАННЫХ ВРЕМЕННЫХ РЯДОВ

Мингалиев З.З., Новикова С.В.

Казанский национальный исследовательский технический университет имени
А.Н. Туполева – КАИ, г. Казань

Поступила в редакцию 11.12.2024, после переработки 15.01.2025.

В статье рассматривается задача прогнозирования уровня глюкозы в крови с использованием данных, содержащих слабо выраженные зависимости между параметрами, включая временные ряды и физиологические показатели. Предложен подход, основанный на применении нейронных сетей с долгой краткосрочной памятью (LSTM), способный производить прогнозирование будущих значений уровня глюкозы (SGV), а также выявлять аномалии в данных. Для улучшения качества модели использован метод кластеризации DBSCAN, позволяющий выделить группы данных с похожими характеристиками. Также разработан алгоритм заполнения пропущенных данных на основе среднего значения в кластере, что позволяет повысить точность прогнозирования. Проведены численные эксперименты на данных, собранных с помощью мониторинга уровня глюкозы, которые продемонстрировали эффективность предложенного подхода для прогнозирования SGV с учётом временных зависимостей и влияния сопутствующих факторов.

Ключевые слова: временные ряды, DBSCAN, аномалии в данных, пропущенные данные, мониторинг уровня глюкозы, SGV (Sensor Glucose Value), моделирование временных зависимостей.

Вестник ТвГУ. Серия: Прикладная математика. 2025. № 1. С. 67–87.
<https://doi.org/10.26456/vtpmk731>

Введение

Прогнозирование уровня глюкозы в крови (гликемии) является критически важной задачей для автоматизации инсулинотерапии пациентов с сахарным диабетом. Автоматизированная система опирается на прогнозное значение уровня сахара крови для расчетов, и должна поддерживать гликемию в целевом диапазоне не менее 70% времени [1]. Ошибки в прогнозах могут привести к гипогликемии (снижению уровня глюкозы ниже нормы) или гипергликемии (повышению уровня глюкозы выше нормы), что несет серьезные риски для здоровья [2]. Поэтому точные методы прогнозирования глюкозы крови чрезвычайно важны.

Классические методы прогнозирования временных рядов, такие как линейная регрессия, ARIMA, использование многослойного персептрона, фильтр Калмана и экспоненциальное сглаживание [3], оказываются малоэффективными из-за невозможности учета индивидуальных особенностей пациента: метаболических особенностей, реакции на инсулин, образа жизни, питания, физической активности, уровня стресса и др. Эти параметры варьируются от человека к человеку, что делает универсальные модели малоприменимыми.

С другой стороны, современные технические средства индивидуального мониторинга (носимые сенсоры и инсулиновые дозаторы) позволяют формировать для каждого пациента гетерогенные временные ряды измерений, включающие информацию о глюкозе крови, потреблении углеводов, дозировке инсулина и физической активности, собираемую с разной частотой и точностью [4]. Стандартные методы прогноза требуют однородных и стационарных данных, в то время как временные ряды, получаемые с сенсоров, имеют выраженную нелинейность и сложную зависимость от множества внешних факторов. Помимо этого, данные, поступающие с сенсоров, могут содержать шумы или пропуски, с которыми классические подходы плохо справляются.

Таким образом, задача прогнозирования уровня гликемии выходит за рамки возможностей традиционных моделей, которые не учитывают индивидуальные особенности пациентов, не способны эффективно обрабатывать смешанные временные ряды и не адаптированы к работе с данными, содержащими шумы и пропуски. Это требует применения более сложных методов, таких как глубокие нейронные сети, которые способны извлекать скрытые нелинейные зависимости и учитывать влияние множества факторов, персонализируя прогнозы для каждого пациента.

1. Постановка задачи

Задачей настоящего исследования является разработка метода, способного:

- учитывать и корректно обрабатывать аномалии и пропущенные значения в данных, минимизируя их влияние на точность прогнозирования;
- применять кластеризацию методом DBSCAN для выявления групп сходных данных и повышения качества прогнозирования;
- прогнозировать уровень глюкозы в крови на основе временных рядов с использованием нейронных сетей с краткой долгосрочной памятью (LSTM);

В качестве исследуемого набора данных используется датасет, содержащий временные ряды показателей уровня глюкозы в крови (SGV), дополненные информацией о колебаниях (Delta), средней динамике изменений (Avg Delta), а также другими параметрами, такими как IOB (активный инсулин) и COB (углеводы в крови). Данные включают примерно 5% пропусков, вызванных техническими сбоями в мониторинговых устройствах.

2. Существующие подходы к прогнозированию гликемии

Существующие методы прогнозирования уровня глюкозы в крови, предложенные различными авторами, опираются на подходы машинного обучения, матема-

тической статистики и анализа временных рядов. Однако каждый из них имеет свои ограничения, особенно в контексте персонализированных моделей.

Методы экстраполяции и прогнозирования трендов (например, ARIMA, Holt-Winters) [5] часто используются для прогноза данных временных рядов. Они эффективны в предсказании гладких трендов, но плохо работают с зашумленными данными, характерными для гликемических уровней. Такие подходы игнорируют дополнительные факторы, влияющие на колебания глюкозы, например физическую активность или стресс [6].

Симуляторы используют математическое моделирование для прогнозирования уровня глюкозы на основе усреднённых данных [7]. Проблема их применения заключается в недостаточной точности в реальных условиях, поскольку симуляторы редко адаптируются к индивидуальным особенностям организма пациента и часто ограничены параметрами, заданными разработчиками [8].

Ансамблевые методы [9], показали высокую точность при сравнении различных подходов. Они используют такие параметры, как уровень глюкозы, прием пищи и физическая активность. Однако ограничениями являются сложность настройки моделей и необходимость в большом объеме обучающих данных. Эти подходы также склонны к переобучению на небольших выборках данных, что ограничивает их применение для индивидуализированных предсказаний, особенно для пациентов с редкими паттернами изменений уровня глюкозы [10].

Таким образом, современные методы сталкиваются с проблемами персонализации, устойчивости к шуму в данных, и сложностями в практическом использовании, что требует дальнейших исследований для создания относительно простых, но точных моделей.

3. Методология

Предлагаемая методология прогнозирования уровня глюкозы строится на использовании трехэтапного комбинированного подхода, который в качестве математического аппарата обработки данных включает:

- формирование смешанных временных рядов ключевых параметров пациента, влияющих на прогнозируемый уровень сахара крови;
- анализ данных, выявление тенденций, очистка от шумов;
- кластеризацию полученных данных;
- непосредственно прогнозирование при помощи нейросетевых рекуррентных моделей типа LSTM на выделенных кластерах.

Такой подход позволяет учитывать как индивидуальные физиологические особенности пациента, так и скрытые закономерности в данных. Опишем предлагаемый метод подробно.

Этап 1. Формирование рядов измерений. На данном этапе формируется смешанный временной ряд, включающий три ключевых параметра:

1. Текущий уровень глюкозы в крови (SGV).

2. Количество активного инсулина (IOB Sum).
3. Активные углеводы (COB).

Как показано в [11], выделенные параметры предоставляют исчерпывающее описание текущего состояния пациента, и их динамика имеет прямую связь с изменением уровня глюкозы в будущем. Так как прогнозируемый уровень глюкозы в ретроспекции напрямую связан со всеми тремя вышеназванными параметрами, их совместное использование позволяет построить более адекватную модель, чем методы, основанные только на одном из параметров.

В исследовании использован открытый набор логгированных данных добровольческого интернет-сообщества «Open Humans», куда люди с диабетом загружают данные со своих систем мониторинга и терапии для широких научных исследований. Все данные полностью анонимны, все участники дают добровольное информированное согласие на использование данных в некоммерческих исследовательских целях (<https://www.openhumans.org/>). Файлы логгированных данных представляют собой записи в формате json [12], фиксирующие данные о текущем уровне сахара в крови, а также параметры, связанные с активностью инсулина и потреблением углеводов в определенные моменты времени. Кроме того, файлы логгирования содержат временные метки, записи о техническом состоянии сенсоров и иных устройств, другие важные параметры, что делает их ценным источником информации для создания моделей.

Данные лог-файлов были трансформированы с использованием алгоритмов, основанных на регулярных выражениях, что позволило точно извлечь ключевые показатели. Затем данные были структурированы в табличный формат, обеспечивая удобство для дальнейшего анализа и использования в машинном обучении. Фрагмент сформированного смешанного временного ряда, используемого в исследовании, приведен в Таблице 1.

ТАБЛИЦА 1: Фрагмент сформированного временного ряда

SGV	IOB Sum	COB
73,51058	2,473206	9,901214
73,62766	2,449787	9,793387
73,82663	2,403566	9,5725
74,07911	2,342227	9,270861
74,40508	2,284576	8,984037
74,85807	2,251429	8,819815
75,45117	2,246876	8,799519
76,10171	2,250468	8,822541
76,67069	2,232998	8,751288

Этап 2. Предварительный анализ и обработка. После формирования смешанного ряда его необходимо подвергнуть анализу с целью определения степени его гладкости, однородности, а также вида присутствующего в нем шума. В ходе предварительных экспериментов было установлено, что данные временных рядов трех ключевых показателей (глюкоза-инсулин-углеводы) обладают существенной

неоднородностью, однако им присущи общие тенденции, позволяющие выделять относительно однородные подмножества рядов данных. Подобное поведение на качественном уровне может быть интерпретировано как, например, состояния, связанные с нормальным уровнем глюкозы, гипогликемией или гипергликемией, либо состояния, связанные с различными режимами и уровнем зашумления работы носимых датчиков, и т.п. Также на данном этапе необходимо провести очистку данных от шумов с использованием сглаживающих фильтров. Тип фильтра определяется по результатам анализа характеристик, присутствующих в данных шумов.

Этап 3. Кластеризация данных. Согласно проведенному на предыдущем этапе анализу, полученные ряды данных подвергаются кластеризации. Метод кластеризации выбирается на основе результатов анализа неоднородности на предыдущем этапе. Полученный в результате применения алгоритма кластеризации код кластера становится дополнительным параметром для прогноза. Это добавляет модели способность учитывать структурные особенности данных, которые не всегда очевидны при обработке временных рядов.

Этап 4. Прогноз гликемии. Для непосредственного прогнозирования уровня глюкозы применяется сеть с долгой краткосрочной памятью LSTM, разработанная специально для работы с временными рядами [13]. Она способна учитывать как долгосрочные, так и краткосрочные зависимости между данными, что особенно важно для задачи прогнозирования гликемии.

Входные данные:

- смешанный временной ряд, состоящий из трех связанных рядов измерений (SGV, IOB Sum, COB), синхронизированных во времени;
- кластерный код конкретного измерения;
- горизонт прогноза – количество временных интервалов в будущем, на которое необходимо сделать прогноз;
- промежуток ретроспекции – количество временных интервалов в прошлом, на основании которых строится прогноз в будущем.

Выходные данные:

- уровень глюкозы (SGV) через заданный промежуток времени.

Структурную схему разработанного метода представляет Рис. 1.

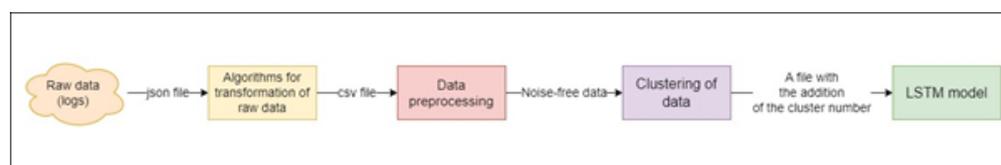


Рис. 1: Взаимодействие этапов комбинированного метода прогнозирования уровня гликемии

4. Вычислительные эксперименты

Для проверки эффективности предлагаемого метода были использованы данные от трех различных пациентов с различными физиологическими характеристиками, что позволило проверить универсальность предложенной модели.

Пациент №1:

- Пол: мужчина,
- Возраст: 35 лет,
- Диагноз: сахарный диабет 1 типа,
- Степень чувствительности к инсулину: средняя,
- Уровень физической активности: низкая.

Пациент №2:

- Пол: женщина,
- Возраст: 50 лет,
- Диагноз: сахарный диабет 2 типа инсулинопотребный,
- Степень чувствительности к инсулину: ниже среднего,
- Уровень физической активности: умеренная.

Пациент №3:

- Пол: мужчина,
- Возраст: 28 лет,
- Диагноз: сахарный диабет 1 типа,
- Степень чувствительности к инсулину: лабильная,
- Уровень физической активности: высокая, аperiodическая.

Для кластеризации использован метод DBSCAN (Density-Based Spatial Clustering of Applications with Noise) – алгоритм для обнаружения кластеров в зашумленных данных на основе плотности. DBSCAN объединяет в кластеры точки с большим количеством «соседей», тогда как отдельные единичные выбросы интерпретирует как шум [14]. Такой подход соответствует особенностям рассматриваемых в нашем исследовании данных, и позволяет дополнительно очищать данные от шумов, которые не удалось нивелировать на этапе фильтрации. В качестве алгоритма очистки данных от шума использован фильтр Гаусса. Пропуски в данных были заполнены методом линейной интерполяции.

Параметры алгоритма кластеризации выбраны на основе метода анализа расстояний между ближайшими соседями. Рис. 2 демонстрирует решающее правило выбора ключевого параметра алгоритма – порога расстояния между соседями ϵ .

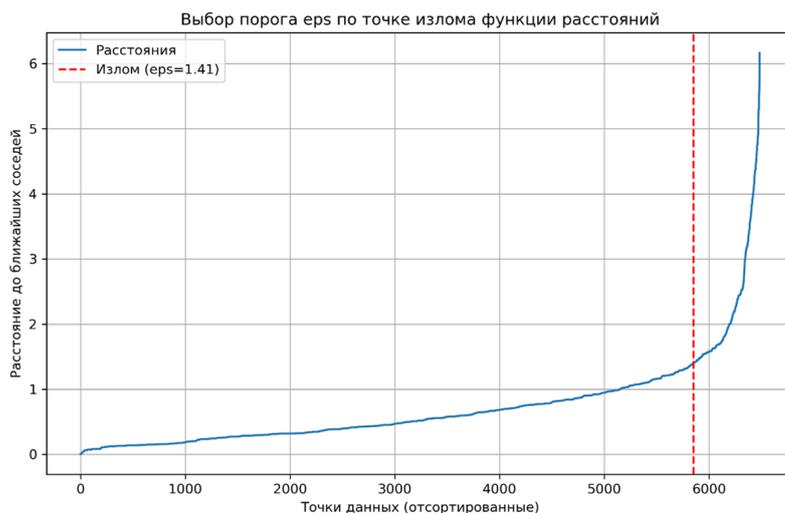


Рис. 2: Выбор порога ϵ по точке излома функции расстояний

В результате кластеризации данные были сгруппированы в три плотных кластера, которые можно интерпретировать как «показания, характерные для нормального уровня сахара», «показания, соответствующие повышенному уровню сахара» и «показания, соответствующие пониженному уровню сахара».

Для оценки точности прогнозов использована средняя абсолютная процентная ошибка, широко применяемая в машинном обучении (MAPE, Mean Absolute Percentage Error) [15], определяемая формулой:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100, \quad (1)$$

где y_i — фактическое значение на i -м шаге, \hat{y}_i — предсказанное значение на i -м шаге, n — количество данных для анализа (количество временных шагов после обработки данных).

Точность определялась как величина, обратная ошибке:

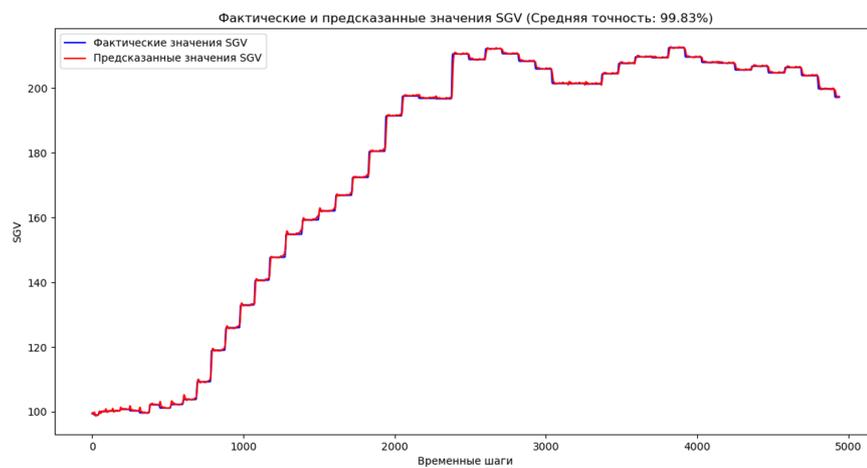
$$\epsilon = 100 - MAPE. \quad (2)$$

Временной шаг данных составил 5 минут. Данные были разделены на обучающую и тестовую выборки в пропорции 70 к 30.

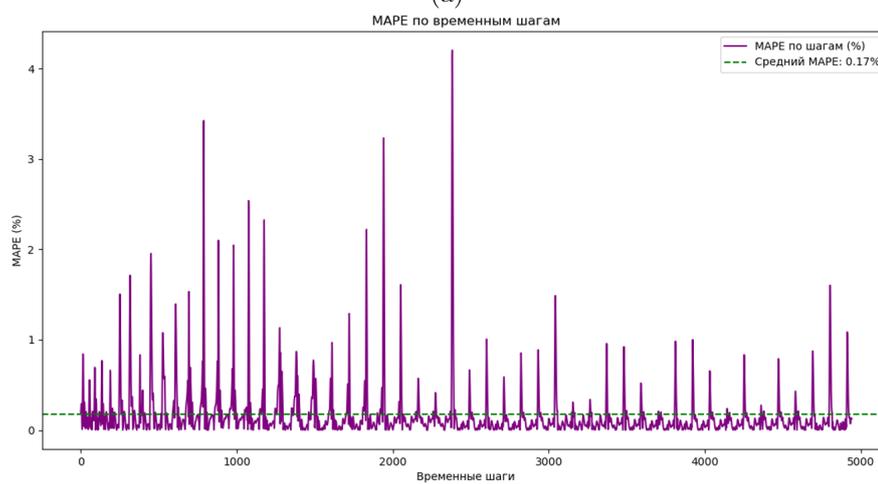
4.1. Результаты экспериментов

Пациент №1.

Прогнозирование по обученной модели на тестовых данных для первого пациента показало точность 99,83% при использовании окна ретроспекции в 60 временных интервалов (300 минут), и 96,41% при ретроспекции в 10 интервалов (50 минут). Графическое представление результатов прогнозирования отражено на Рис. 3 и Рис. 4 соответственно.

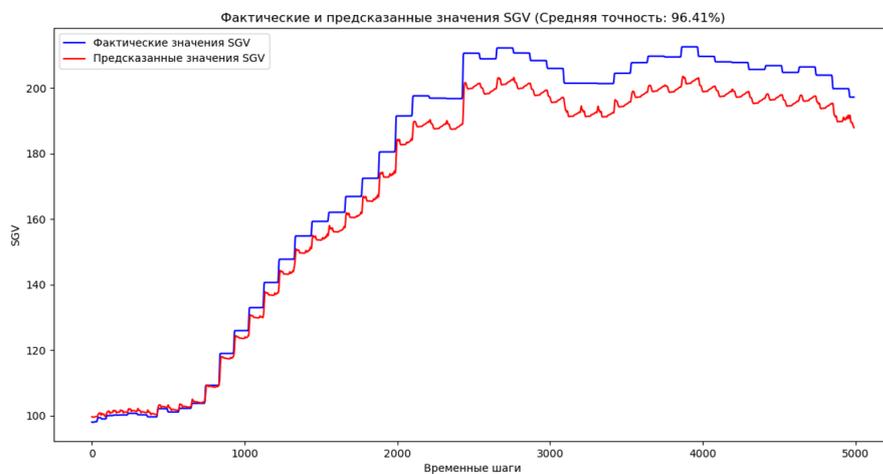


(а)

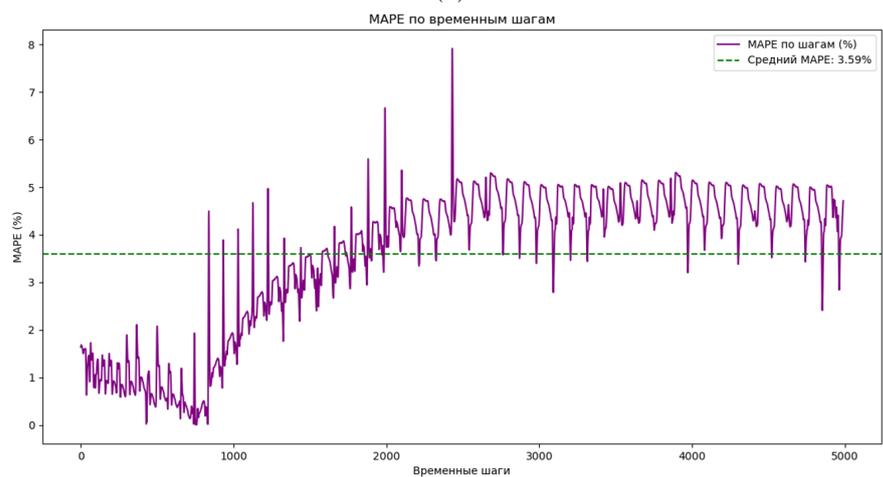


(б)

Рис. 3: Результаты прогнозирования для Пациента №1, ретроспекция 300 минут: (а) отклонение прогнозных значений от реальных; (б) поминутный уровень ошибки прогноза



(а)

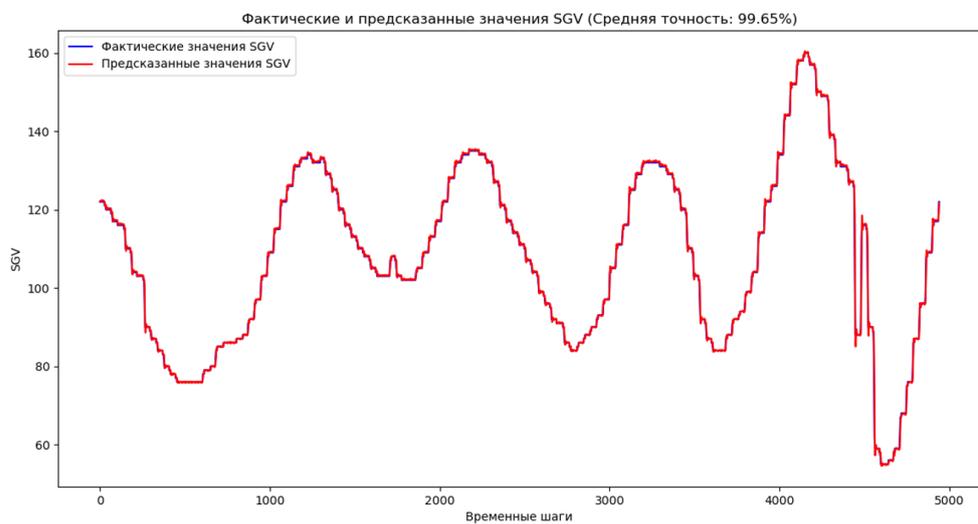


(б)

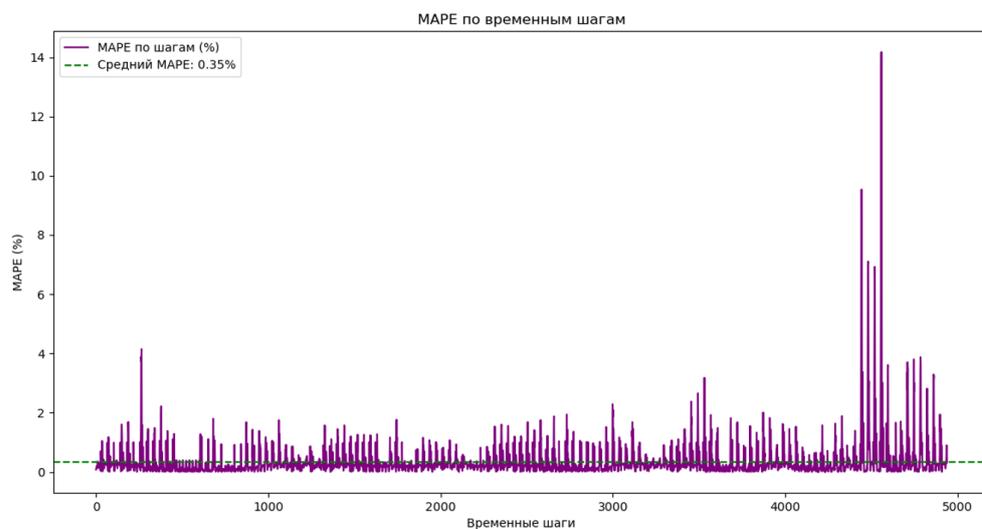
Рис. 4: Результаты прогнозирования для Пациента №1, ретроспекция 50 минут: (а) отклонение прогнозных значений от реальных; (б) поминутный уровень ошибки прогноза

Пациент №2.

Прогнозирование по обученной модели на тестовых данных для второго пациента показало точность 99,65% при использовании окна ретроспекции в 60 временных интервалов (300 минут), и 96,69% при ретроспекции в 10 интервалов (50 минут). Графическое представление результатов прогнозирования отражено на Рис.5 и Рис. 6 соответственно.



(а)



(б)

Рис. 5: Результаты прогнозирования для Пациента №2, ретроспекция 300 минут: (а) отклонение прогнозных значений от реальных; (б) поминутный уровень ошибки прогноза

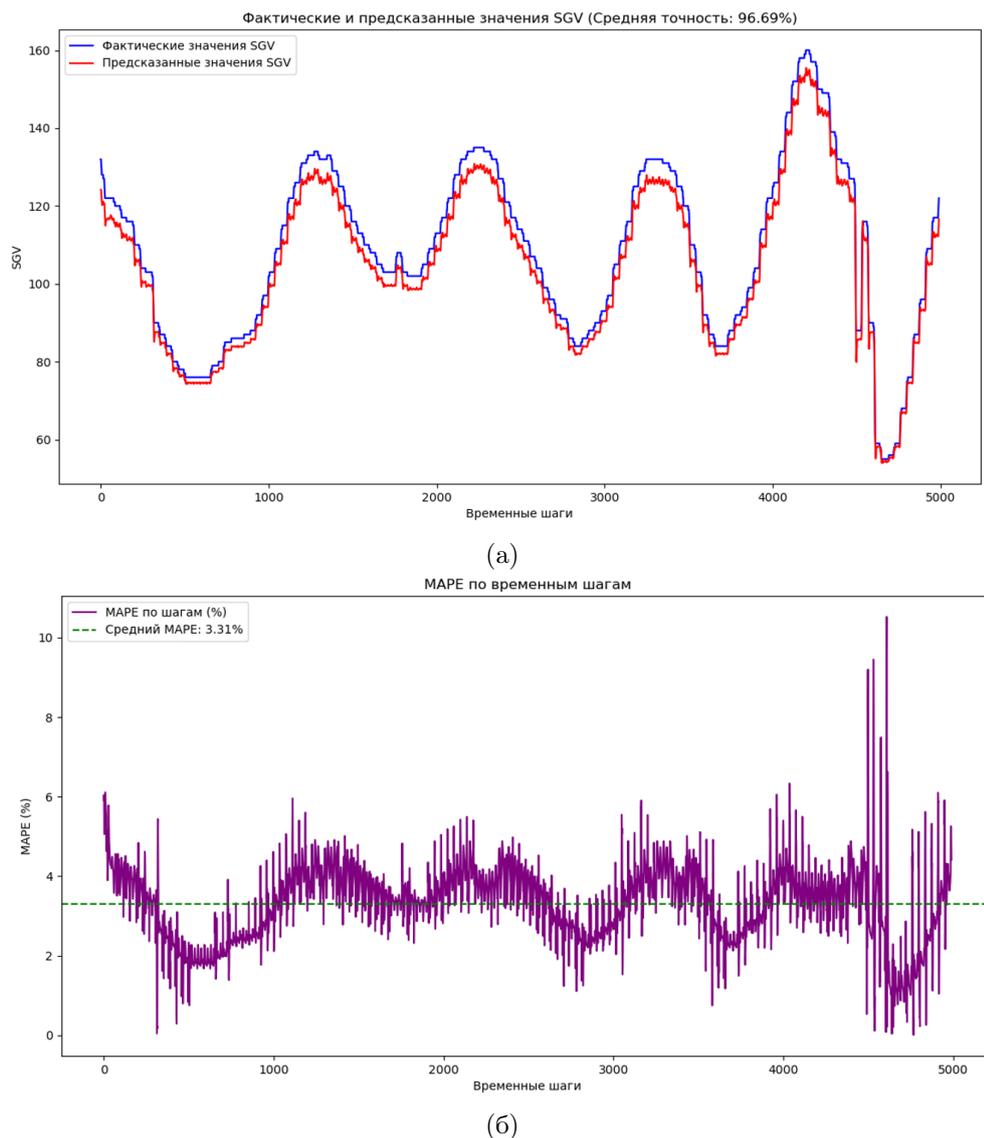


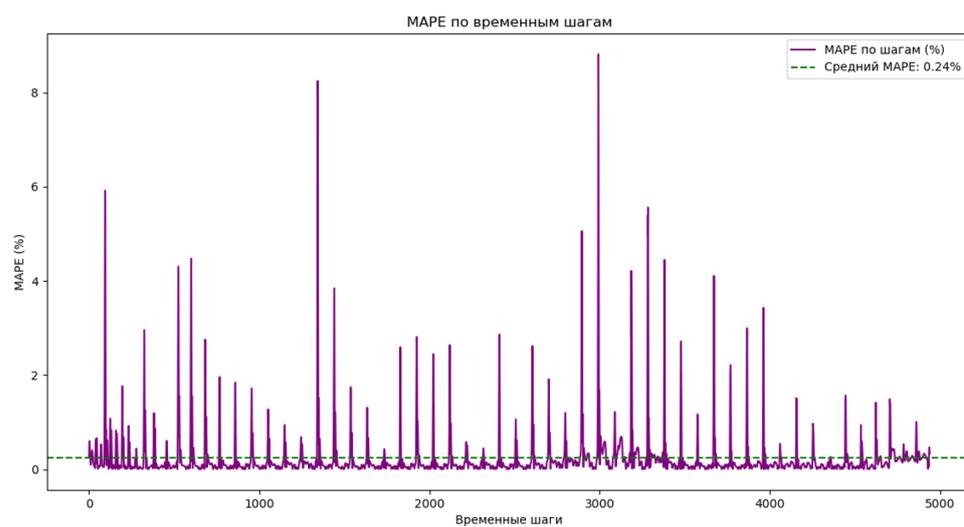
Рис. 6: Результаты прогнозирования для Пациента №2, ретроспекция 50 минут: (а) отклонение прогнозных значений от реальных; (б) поминутный уровень ошибки прогноза

Пациент №3.

Прогнозирование по обученной модели на тестовых данных для третьего пациента показало точность 99,76% при использовании окна ретроспекции в 60 временных интервалов (300 минут), и 99,16% при ретроспекции в 10 интервалов (50 минут). Графическое представление результатов прогнозирования отражено на Рис.7 и Рис. 8 соответственно.



(а)

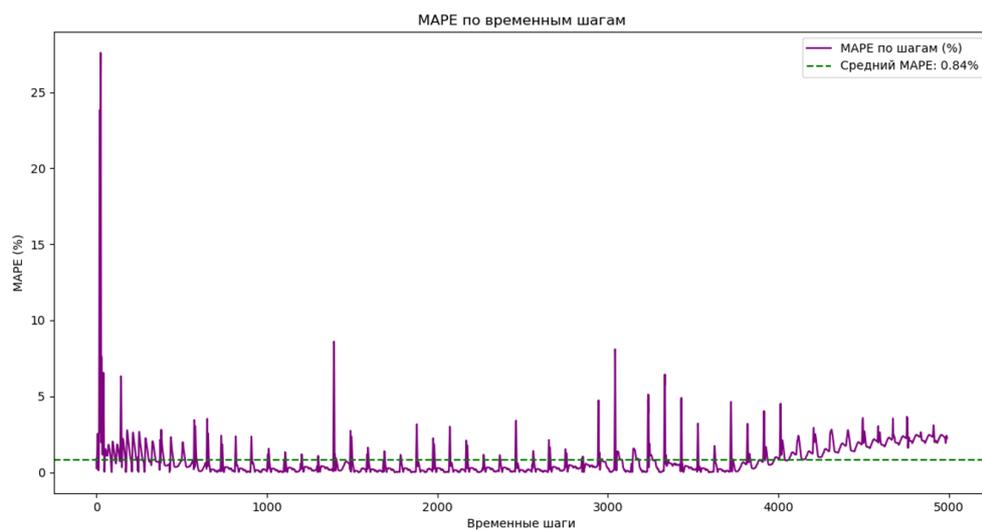


(б)

Рис. 7: Результаты прогнозирования для Пациента №3, ретроспекция 300 минут: (а) отклонение прогнозных значений от реальных; (б) поминутный уровень ошибки прогноза



(а)



(б)

Рис. 8: Результаты прогнозирования для Пациента №3, ретроспекция 50 минут: (а) отклонение прогнозных значений от реальных; (б) поминутный уровень ошибки прогноза

Таким образом, эксперименты для всех трех пациентов продемонстрировали высокую точность предложенного метода.

5. Сравнение с существующими аналогами

Для доказательства преимущества разработанного метода в рамках исследования было произведено сравнение эффективности прогнозирования с другими методами, пригодными для обработки временных рядов.

Для сравнения были использованы ряды данных Пациента № 1. Сравнение предложенного метода проводилось по показателю Точность в смысле формулы (2) со следующими методами:

- ARIMA;
- Многослойный персептрон со структурой 6:2:1 (6 входов, 1 скрытый слой с 2 нейронами, 1 выход);
- LSTM по смешанному временному ряду без предварительной кластеризации;
- LSTM по временному ряду гликемии с предварительной кластеризацией.

Результаты тестирования четырех методов, а также предложенного комбинированного метода представлены в Таблице 2.

ТАБЛИЦА 2: Результаты тестирования методов

Модель	Точность	Время обучения	Время отклика для получения результата прогнозирования
ARIMA	72,49%	0.93 секунд	0,74 секунд
Многослойный персептрон	93,41%	6.1 секунд	1.03 секунд
LSTM (без кластеризации)	86,61%	72.63 секунд	1.06 секунд
LSTM (с кластеризацией, только глюкоза)	92,86%	75.63 секунд	1.07 секунд
LSTM (с кластеризацией)	99,83%	83.51 секунд	1.07 секунд

Рис. 9 демонстрирует графическое сравнение эффективности рассматриваемых методов.

Эксперименты, проведенные на данных Пациентов №2 и №3, продемонстрировали аналогичные результаты.

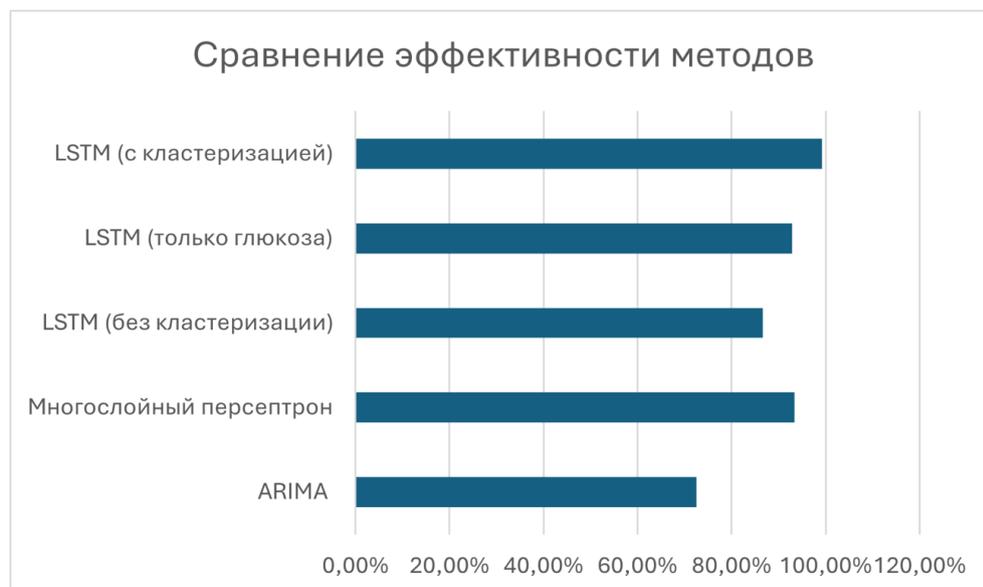


Рис. 9: Сравнение эффективности прогнозирования

6. Обсуждение

В ходе анализа результатов экспериментов можно сделать несколько ключевых наблюдений. Разработанные варианты комплексной модели продемонстрировали стабильно высокую точность в 95% и выше для данных трех пациентов с различными физиологическими и другими особенностями. Это подтверждает возможность применения разработанной модели прогноза уровня глюкозы в персонализированной медицине с высокой точностью и надежностью.

Наиболее высокая точность предсказания уровня глюкозы в крови была достигнута при использовании шага ретроспекции в диапазоне 60 шагов (300 минут). Это обусловлено тем, что меньшие значения окна ретроспекции не дают достаточного объема данных для выявления временных закономерностей, а большие значения могут включать устаревшие данные, снижая актуальность прогноза.

Метод группировки данных DBSCAN продемонстрировал высокую эффективность в составе комплексной модели, и может быть рекомендован как основной метод этапа кластеризации.

Сравнение результатов работы разработанного комплексного метода с аналогичными подходами продемонстрировало его преимущества как в точности прогнозирования, так и в скорости вычислений. В частности, по сравнению с самым точным из анализируемых методов (многослойным перцептроном), разработанный подход обеспечивает увеличение точности прогнозирования с 93% до 99%. В среднем, точность разработанного метода превышает показатели аналогов на 6%. Кроме того, анализ времени отклика показывает, что использование комплексного метода обеспечивает скорость вычислений на уровне лучших из сравниваемых решений, не уступая им по эффективности.

Заключение

В данной статье приведены результаты разработки и исследования эффективности комплексного метода прогнозирования уровня глюкозы в крови с использованием машинного обучения. Основной особенностью метода является предварительная кластеризация данных смешанных временных рядов и включение меток классов как дополнительных характеристик временного ряда. Для прогнозирования таких дополненных смешанных рядов использованы сети долгой краткосрочной памяти LSTM.

Эксперименты проводились на реальных данных, включающих широкий диапазон показателей, таких как уровень глюкозы, углеводная нагрузка и активный инсулин. Результаты исследования показали, что предварительная кластеризация данных с использованием алгоритма DBSCAN значительно улучшает качество прогнозов, позволяя эффективно обрабатывать аномалии, такие как резкие скачки уровня глюкозы. Достигнутая точность прогноза в 99% позволяет сделать вывод об эффективности применения разработанного метода на практике.

Сравнение предложенного подхода с традиционными моделями и другими архитектурами нейронных сетей показало его превосходство в точности и скорости вычислений.

Разработанная модель адаптирована для работы в реальном времени и может быть интегрирована в системы автоматизированного управления диабетом.

В качестве направлений будущих исследований рассматривается возможность включения в модель данных о физической активности и физиологическом состоянии пациента с носимых устройств, таких как смарт-часы, фитнес-браслеты и т.п.

Список литературы

- [1] Battelino T., Danne T., Bergenstal R.M. et al. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range // *Diabetes Care*. 2019. Vol. 42, № 8. Pp. 1593–1603. <https://doi.org/10.2337/dci19-0028>
- [2] Pikulin S., Yehezkel I., Moskovitch R. Enhanced blood glucose levels prediction with a smartwatch // *PLoS ONE*. 2024. Vol. 19, № 7. ID e0307136. <https://doi.org/10.1371/journal.pone.0307136>
- [3] Rancati S., Bosoni P., Schiaffini R., Deodati A., Mongini P.A., Sacchi L., Toffanin C., Bellazzi R. Exploration of Foundational Models for Blood Glucose Forecasting in Type-1 Diabetes Pediatric Patients // *Diabetology*. 2024. Vol. 5. Pp. 584–599. <https://doi.org/10.3390/diabetology5060042>
- [4] Vettoretti M., Cappon G., Facchinetti A., Sparacino G. Advanced Diabetes Management Using Artificial Intelligence and Continuous Glucose Monitoring Sensors // *Sensors*. 2020. Vol. 20. ID 3870. <https://doi.org/10.3390/s20143870>
- [5] Hamilton J. *Time Series Analysis*. Princeton University Press, 1994.

- [6] Кисиль С.И., Залетова Т.С. Существующие подходы к математической модели «инсулин-глюкоза» // Национальная Ассоциация Ученых. 2020. Т. 54, № 1. С. 4–7. <http://doi.org/10.31618/nas.2413-5291.2020.1.54.192>
- [7] Albers D.J., Levine M., Gluckman B., Ginsberg H., Hripesak G., et al. Personalized glucose forecasting for type 2 diabetes using data assimilation // . 2017. Vol. 13, № 4. ID e1005232. <https://doi.org/10.1371/journal.pcbi.1005232>
- [8] Гоменюк С.М., Емельянов А.О., Карпенко А.П., Чернецов С.А. Методы прогнозирования оптимальных доз инсулина для больных сахарным диабетом I типа. Обзор // Машиностроение и компьютерные технологии. 2009. № 4.
- [9] Khadem H., Nemat H., Elliott J., Benaissa M. Blood Glucose Level Time Series Forecasting: Nested Deep Ensemble Learning Lag Fusion // Bioengineering. 2023. Vol. 10. ID 487. <https://doi.org/10.3390/bioengineering10040487>
- [10] Tena F., Garnica O., Lanchares J., Hidalgo J.I. Ensemble Models of Cutting-Edge Deep Neural Networks for Blood Glucose Prediction in Patients with Diabetes // Sensors. 2021. Vol. 21. ID 7090. <https://doi.org/10.3390/s21217090>
- [11] Ghimire S., Celik T., Gerdes M., Omlin C.W. Deep learning for blood glucose level prediction: How well do models generalize across different data sets? // PLoS ONE. 2024. Vol. 19, № 9. ID e0310801. <https://doi.org/10.1371/journal.pone.0310801>
- [12] Lee J.M., Hirschfeld E., Wedding J. A Patient-Designed Do-It-Yourself Mobile Technology System for Diabetes: Promise and Challenges for a New Era in Medicine // JAMA. 2016. Vol. 315, № 14. Pp. 1447–1448. <https://doi.org/10.1001/jama.2016.1903>
- [13] El Idrissi T., Idri A., Abnane I., Bakkoury Z. Predicting blood glucose using an LSTM Neural Network // Proceedings of the 2019 Federated Conference on Computer Science and Information Systems. Eds. by M. Ganzha, L. Maciaszek, M. Paprzycki. 2019. Pp. 35–41. <http://doi.org/10.15439/2019F159>
- [14] Chien C.-F., Suwattananuruk B. Density-Based Spatial Clustering of Applications With Noise (DBSCAN) for Probe Card Production for Advanced Quality Control of Wafer Probing Test // IEEE Transactions on Semiconductor Manufacturing. 2024. Vol. 37, № 4. Pp. 567–575. <https://doi.org/10.1109/TSM.2024.3468000>
- [15] Mean Absolute Percentage Error (MAPE) [Electronic resource] // Statistics How To. URL: <https://www.statisticshowto.com/mean-absolute-percentage-error-mape/>.

Образец цитирования

Мингалиев З.З., Новикова С.В. Модель прогнозирования гликемии на основе смешанных временных рядов // Вестник ТвГУ. Серия: Прикладная математика. 2025. № 1. С. 67–87. <https://doi.org/10.26456/vtppmk731>

Сведения об авторах**1. Мингалиев Заид Зульфатович**

аспирант кафедры Прикладной математики и информатики Казанского национального исследовательского технического университета им А.Н. Туполева – КАИ.

Россия, 420015, г. Казань, ул. Большая Красная, д. 55, КНИТУ-КАИ.

E-mail: zaidmingaliev@yandex.ru

2. Новикова Светлана Владимировна

профессор кафедры Прикладной математики и информатики Казанского национального исследовательского технического университета им А.Н. Туполева – КАИ.

Россия, 420015, г. Казань, ул. Большая Красная, д. 55, КНИТУ-КАИ.

E-mail: SVNovikova@kai.ru

A MIXED TIME SERIES GLYCEMIC PREDICTION MODEL

Mingaliyev Z.Z., Novikova S.V.

Kazan National Research Technical University – KAI, Kazan

Received 11.12.2024, revised 15.01.2025.

The article considers the problem of predicting blood glucose levels using data containing weakly expressed dependencies between parameters, including time series and physiological parameters. An approach based on the use of neural networks with long short-term memory (LSTM) is proposed, which is capable of predicting future glucose values (SGV), as well as identifying anomalies in the data. To improve the quality of the model, the DBSCAN clustering method is used, which allows you to identify groups of data with similar characteristics. An algorithm for filling in missing data based on the average value in the cluster is also developed, which improves the accuracy of forecasting. Numerical experiments were carried out on data collected by monitoring glucose levels, which demonstrated the effectiveness of the proposed approach for predicting SGV, taking into account time dependencies and the influence of associated factors.

Keywords: time series, DBSCAN, data anomalies, missing data, glucose monitoring, SGV (Sensor Glucose Value), time history modeling.

Citation

Mingaliyev Z.Z., Novikova S.V., “A mixed time series glyceemic prediction model”, *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2025, № 1, 67–87 (in Russian). <https://doi.org/10.26456/vtpmk731>

References

- [1] Battelino T., Danne T., Bergenstal R.M. et al., “Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range”, *Diabetes Care*, **42**:8 (2019), 1593–1603, <https://doi.org/10.2337/dci19-0028>.
- [2] Pikulin S., Yehezkel I., Moskovitch R., “Enhanced blood glucose levels prediction with a smartwatch”, *PLoS ONE*, **19**:7 (2024), e0307136, <https://doi.org/10.1371/journal.pone.0307136>.
- [3] Rancati S., Bosoni P., Schiaffini R., Deodati A., Mongini P.A., Sacchi L., Toffanin C., Bellazzi R., “Exploration of Foundational Models for Blood Glucose Forecasting in Type-1 Diabetes Pediatric Patients”, *Diabetology*, **5** (2024), 584–599, <https://doi.org/10.3390/diabetology5060042>.

- [4] Vettoretti M., Cappon G., Facchinetti A., Sparacino G., “Advanced Diabetes Management Using Artificial Intelligence and Continuous Glucose Monitoring Sensors”, *Sensors*, **20** (2020), 3870, <https://doi.org/10.3390/s20143870>.
- [5] Hamilton J., *Time Series Analysis*, Princeton University Press, 1994.
- [6] Kasil’ S.I., Zaletova T.S., “Existing approaches to the mathematical model of ”insulin-glucose””, *Natsional’naya Assotsiatsiya Ucheny’kh [National Association of Scientists]*, **54**:1 (2020), 4–7 (in Russian), <http://doi.org/10.31618/nas.2413-5291.2020.1.54.192>.
- [7] Albers D.J., Levine M., Gluckman B., Ginsberg H., Hripcsak G., et al., “Personalized glucose forecasting for type 2 diabetes using data assimilation”, **13**:4 (2017), e1005232, <https://doi.org/10.1371/journal.pcbi.1005232>.
- [8] Gomenyuk S.M., Emel’yanov A.O., Karpenko A.P., Chernetsov S.A., “Methods for predicting optimal insulin doses for patients with type I diabetes mellitus. Review”, *Mashinostroenie i komp’yuternyye tekhnologii [Mechanical engineering and computer technology]*, 2009, №4 (in Russian).
- [9] Khadem H., Nemat H., Elliott J., Benaissa M., “Blood Glucose Level Time Series Forecasting: Nested Deep Ensemble Learning Lag Fusion”, *Bioengineering*, **10** (2023), 487, <https://doi.org/10.3390/bioengineering10040487>.
- [10] Tena F., Garnica O., Lanchares J., Hidalgo J.I., “Ensemble Models of Cutting-Edge Deep Neural Networks for Blood Glucose Prediction in Patients with Diabetes”, *Sensors*, **21** (2021), 7090, <https://doi.org/10.3390/s21217090>.
- [11] Ghimire S., Celik T., Gerdes M., Omlin C.W., “Deep learning for blood glucose level prediction: How well do models generalize across different data sets?”, *PLoS ONE*, **19**:9 (2024), e0310801, <https://doi.org/10.1371/journal.pone.0310801>.
- [12] Lee J.M., Hirschfeld E., Wedding J., “A Patient-Designed Do-It-Yourself Mobile Technology System for Diabetes: Promise and Challenges for a New Era in Medicine”, *JAMA*, **315**:14 (2016), 1447–1448, <https://doi.org/10.1001/jama.2016.1903>.
- [13] El Idrissi T., Idri A., Abnane I., Bakkoury Z., “Predicting blood glucose using an LSTM Neural Network”, *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems*, eds. M. Ganzha, L. Maciaszek, M. Paprzycki, 2019, 35–41, <http://doi.org/10.15439/2019F159>.
- [14] Chien C.-F., Suwattananuruk B., “Density-Based Spatial Clustering of Applications With Noise (DBSCAN) for Probe Card Production for Advanced Quality Control of Wafer Probing Test”, *IEEE Transactions on Semiconductor Manufacturing*, **37**:4 (2024), 567–575, <https://doi.org/10.1109/TSM.2024.3468000>.
- [15] *Mean Absolute Percentage Error (MAPE)*, Statistics How To, <https://www.statisticshowto.com/mean-absolute-percentage-error-mape/>.

Author Info**1. Mingaliyev Zaid Zulfatovich**

PhD student at the Department of Applied Mathematics and Computer Science,
Kazan National Research Technical University named after A.N. Tupolev – KAI.

Russia, 420015, Kazan, 55 Bolshaya Krasnaya str., KNRTU-KAI.

E-mail: zaidmingaliyev@yandex.ru

2. Novikova Svetlana Vladimirovna

Professor at the Department of Applied Mathematics and Computer Science, Kazan
National Research Technical University named after A.N. Tupolev – KAI.

Russia, 420015, Kazan, 55 Bolshaya Krasnaya str., KNRTU-KAI.

E-mail: SVNovikova@kai.ru