



Машинное обучение в прогнозировании неблагоприятных сердечно-сосудистых событий у пациентов с сахарным диабетом 2 типа. Мнение по проблеме

Наговицин А. К.^{1,2}, Бакутина Ю. Ю.¹, Кочкина И. А.¹, Чернуцкий В. В.¹

В работе анализируются перспективы использования машинного обучения для оценки сердечно-сосудистого риска у пациентов с сахарным диабетом 2 типа. Показано, что современные алгоритмы, в частности нейронные сети, демонстрируют высокую эффективность в прогнозировании риска. Однако, несмотря на высокую точность моделей, их внедрение ограничено рядом проблем, таких как дефицитом внешней проверки на независимых выборках и сложностью интеграции в системы поддержки врачебных решений. Отмечена необходимость в стандартизации дизайна исследований и обеспечении прозрачности работы алгоритмов для внедрения машинного обучения в широкую клиническую практику.

Ключевые слова: сахарный диабет 2 типа, сердечно-сосудистый риск, машинное обучение, искусственный интеллект, прогнозирование осложнений, стратификация риска.

Отношения и деятельность: нет.

¹ФГБОУ ВО Воронежский государственный медицинский университет им. Н. Н. Бурденко Минздрава России, Воронеж, Россия; ²ООО "Медицинский центр Широких Сердец", Воронеж, Россия.

Наговицин А. К.* — ассистент кафедры клинической лабораторной диагностики; врач-кардиолог, ORCID: 0000-0002-2452-1105, SPIN: 3024-7503, Бакутина Ю. Ю. — ассистент кафедры клинической лабораторной диагнос-

тики, ORCID: 0000-0003-4829-3511, SPIN: 7733-9395, Кочкина И. А. — студент, ORCID: 0009-0005-1533-8770, Чернуцкий В. В. — студент, ORCID: 0009-0001-5979-383X, SPIN: 1783-8850.

*Автор, ответственный за переписку (Corresponding author): nagovizina@gmail.com

СД — сахарный диабет, ССЗ — сердечно-сосудистые заболевания, ССР — сердечно-сосудистый риск, ЭМК — электронная медицинская карта, ML — машинное обучение (machine learning), SHAP — SHapley Additive exPlanations.

Рукопись получена 09.03.2026

Рецензия получена 20.03.2026

Принята к публикации 24.03.2026



Для цитирования: Наговицин А. К., Бакутина Ю. Ю., Кочкина И. А., Чернуцкий В. В. Машинное обучение в прогнозировании неблагоприятных сердечно-сосудистых событий у пациентов с сахарным диабетом 2 типа. Мнение по проблеме. *Российский кардиологический журнал*. 2026;31(2S):6899. doi: 10.15829/1560-4071-2026-6899. EDN: INOCHX

Machine learning in predicting adverse cardiovascular events in patients with type 2 diabetes. Opinion on the problem

Nagovitsyn A. K.^{1,2}, Bakutina Yu. Yu.¹, Kochkina I. A.¹, Chernutskiy V. V.¹

This paper analyzes the prospects for using machine learning to assess cardiovascular risk in patients with type 2 diabetes. It is shown that modern algorithms, particularly neural networks, demonstrate high efficiency in risk prediction. However, despite the high accuracy of the models, their implementation is limited by several issues, such as the lack of external validation on independent samples and the difficulty of integrating them into clinical decision support systems. The authors emphasize the need for standardized study design and transparency in algorithm performance to implement machine learning in widespread clinical practice.

Keywords: type 2 diabetes, cardiovascular risk, machine learning, artificial intelligence, complication prediction, risk stratification.

Relationships and Activities: none.

¹Burdenko Voronezh State Medical University, Voronezh, Russia; ²ООО "Shirokih Serdets" Medical Center, Voronezh, Russia.

Nagovitsyn A. K.* ORCID: 0000-0002-2452-1105, Bakutina Yu. Yu. ORCID: 0000-0003-4829-3511, Kochkina I. A. ORCID: 0009-0005-1533-8770, Chernutskiy V. V. ORCID: 0009-0001-5979-383X.

*Corresponding author: nagovizina@gmail.com

Received: 09.03.2026 **Revision Received:** 20.03.2026 **Accepted:** 24.03.2026

For citation: Nagovitsyn A. K., Bakutina Yu. Yu., Kochkina I. A., Chernutskiy V. V. Machine learning in predicting adverse cardiovascular events in patients with type 2 diabetes. Opinion on the problem. *Russian Journal of Cardiology*. 2026;31(2S):6899. doi: 10.15829/1560-4071-2026-6899. EDN: INOCHX

Сердечно-сосудистые заболевания (ССЗ) являются ведущей причиной смертности в мире [1–3]. Зачастую пациенты с ССЗ также страдают заболеваниями прочих органов и систем, особое место среди которых занимает сахарный диабет (СД) 2 типа [1]. Его наличие существенно увеличивает вероятность развития ССЗ и их прогрессирования [1]. По данным Joseph JJ, et al. (2022) у пациентов с СД 2 типа чаще регистрируются

ишемическая болезнь сердца, инсульт, сердечная недостаточность и поражение периферических артерий, что связано с сочетанным влиянием гипергликемии, инсулинорезистентности, дислипидемии, артериальной гипертензии, хронического воспаления и эндотелиальной дисфункции [1]. В связи с этим точная оценка сердечно-сосудистого риска (ССР) у больных СД 2 типа имеет не только прогностическое, но

Ключевые моменты

- Традиционная стратификация риска имеет существенные ограничения у пациентов с диагнозом сахарный диабет 2 типа.
- Машинное обучение и искусственный интеллект способны выявлять скрытые закономерности для индивидуализированного прогнозирования осложнений.
- Интерпретируемые модели позволяют точнее оценивать сердечно-сосудистый риск, но их внедрение требует обязательной внешней проверки на независимых выборках.

Key messages

- Traditional risk stratification has significant limitations in patients diagnosed with type 2 diabetes.
- Machine learning and artificial intelligence can identify hidden patterns for personalized complication prediction.
- Interpretable models allow for more accurate cardiovascular risk assessment, but their implementation requires mandatory external validation on independent samples.

и практическое значение, поскольку позволяет своевременно выделять пациентов с неблагоприятным прогнозом, интенсифицировать профилактические мероприятия и более обоснованно выбирать терапевтическую тактику [1, 4].

Традиционно для стратификации риска сердечно-сосудистых событий применяются шкалы SCORE2, ASCVD, QRISK3, PREVENT и другие модели, основанные на классических факторах риска [5]. Их преимущества — простота и возможность использования в повседневной клинической практике [4, 5]. Однако у больных СД 2 типа потенциал подобных инструментов нередко оказывается ограниченным [4, 5]. Как отмечают Dziopa K, et al. (2022), значительная часть шкал была разработана на общепопуляционных выборках и не всегда учитывает особенности диабетической когорты [4]. Кроме того, такие модели плохо отражают сложные взаимосвязи между множественными клиническими, лабораторными и поведенческими переменными, на что указывают Бакулин Г.Г. и др. (2025) [5]. Ещё один недостаток — статичность традиционных калькуляторов и шкал: они оценивают риск на основании ограниченного набора параметров и не адаптируются к накоплению новых данных [5]. Поскольку исходный риск сердечно-сосудистых осложнений у пациентов с СД 2 типа выше, чем в общей популяции, а его распределение внутри самой диабетической когорты выражено неоднородно, проблема оценки рисков сердечно-сосудистых событий является особенно актуальной [1, 3]. Как показало исследование Kwiendacz H, et al. (2023) в рамках проекта Silesia Diabetes-Heart, применение методов кластерного анализа позволяет выделять группы пациентов с качественно различными профилями ССР, обусловленными разными сочетаниями клинических характеристик и коморбидной патологии [6]. В такой ситуации для врача важно не только отнести пациента к конкретной категории риска, но и получить более точную, индивидуализированную оценку вероятности конкрет-

ных неблагоприятных исходов [3, 5]. Именно поэтому, как подчёркивают Бакулин Г.Г. и др. (2025), а также Oikonomou EK, Khera R (2023), персонализированные модели, основанные на технологиях машинного обучения (machine learning, ML), в последние годы рассматриваются как одно из наиболее перспективных направлений кардиометаболической медицины [3, 5].

По сравнению с классическими регрессионными моделями методы ML обладают рядом принципиальных преимуществ [5, 7]. Они способны анализировать большие массивы информации, одновременно учитывать большое количество признаков, выявлять скрытые закономерности и формировать более динамические прогнозы [5, 8]. Это особенно актуально для больных СД 2 типа, поскольку прогноз течения заболевания определяется сложными взаимодействиями между возрастом, полом, длительностью заболевания, уровнем гликированного гемоглобина, артериальным давлением, характеристиками липидного обмена, функцией почек, массой тела, сопутствующей патологией, проводимой терапией, данными методов визуализации и особенностями анамнеза [8–11].

Современные исследования показывают, что при прогнозировании сердечно-сосудистых событий на основе структурированных клинических данных наиболее часто используются ансамблевые алгоритмы, включая метод случайного леса (random forest), градиентного бустинга (gradient boosting) и аналогичные им методы [8, 12]. Их востребованность во многом обусловлена способностью эффективно работать с неоднородными признаками, относительной устойчивостью к выбросам и погрешностям измерений и высокой точностью прогнозов [8]. По данным систематического обзора Kee OT, et al. (2023), наилучшую прогностическую эффективность при предсказании ССЗ у пациентов с СД 2 типа показала модель на основе нейронной сети (AUC 0,91), за которой следовал градиентный бустинг, тогда как классические подходы, включая логистическую регрессию, демонстрировали худшие результаты [12]. Для неструктурированных данных, таких как данные элек-

трокардиографии, медицинские изображения и текстовые фрагменты медицинской документации, всё более активно применяются подходы глубокого обучения (Deep learning) [3, 13]. Однако, как отмечают Oikonomou EK, Khera R (2023), нейронные сети и модели глубокого обучения являются слишком сложными для понимания человеком, а обеспечение интерпретируемости нейронных сетей и моделей глубокого обучения превращается в одно из ключевых методологических требований, несоблюдение которого существенно ограничивает возможность их широкого внедрения в клиническую практику [3].

Особый интерес представляют работы, в которых прогностические модели создавались на основе электронных медицинских карт (ЭМК) и крупных баз данных [6, 9, 14]. Подобный подход позволяет приблизить алгоритмы к условиям реальной практики, поскольку в их основе лежат данные, отражающие повседневное ведение пациентов. В исследовании Sang H, et al. (2024) были разработаны прогностические модели для оценки риска ССЗ у пациентов с СД, прошедшие валидацию на двух независимых корейских когортах и продемонстрировавшие высокую прогностическую точность [14]. Xu C, et al. (2025) разработали и произвели валидацию на независимой выборке модель оценки риска ССЗ у пациентов с СД 2 типа, подтвердившую применимость методов ML в реальной клинической практике [9]. В рамках проекта Silesia Diabetes-Heart предпринималась попытка выделения различных профилей ССР у пациентов с диабетом, что имеет принципиальное значение для перехода от усредненной стратификации к более точному фенотипированию больных [6]. Практическое значение имеют и узкоспециализированные модели: Liu Y, et al. (2025) была разработана и валидирована модель для прогнозирования риска развития сердечной недостаточности у больных СД 2 типа [10], а Tang D, et al. (2025) применили объяснимую модель с SHAP-анализом для оценки факторов риска коронарной болезни сердца, продемонстрировав возможность сочетания высокой прогностической силы с клинической интерпретируемостью [15]. Tang Y, et al. (2025) показали, что объединение клинических факторов риска с характеристиками перикоронарной жировой ткани способно дополнительно улучшать прогнозирование ССЗ при СД 2 типа [11].

Ключевое преимущество таких моделей связано не только с возможным повышением точности прогноза. Модели ML позволяют произвести оценку индивидуального уровня риска и установить вклад конкретных клинических характеристик пациента в формирование повышенного риска [3, 6, 13, 15]. Для врача-клинициста это означает переход к пониманию вклада отдельных факторов: так, в исследовании Tang D, et al. (2025) применение SHAP-анализа позволило установить, что длительность диабетического анамнеза, уровень глюкозы крови и значение гликированного ге-

моглобина вносят наибольший вклад в прогноз коронарной болезни сердца у больных СД 2 типа [15].

Следует отметить, что при разработке моделей прогнозирования с использованием ML существует ряд проблем и ограничений. Так, качество модели зависит от качества исходных данных, на которых обучаются современные прогностические модели [8, 12]. В большинстве случаев они используют информацию из ЭМК, лабораторных информационных систем, регистров и популяционных когорт, что, с одной стороны, расширяет возможности прогнозирования, а с другой — повышает гетерогенность данных и их зависимость от условий сбора [8, 9, 14]. В реальной клинической практике данные собираются в процессе повседневного ведения пациента, поэтому они могут быть неполными, неоднородными и зависимыми от особенностей конкретного учреждения [8, 9]. По данным Nur A, et al. (2025), модели, в которых явно описана обработка пропущенных данных и применены методы их замещения, демонстрируют более высокую прогностическую эффективность [8]. Xu C, et al. (2025) также указывают, что неравномерное распределение исходов в выборке, неоднородность данных и пропущенные значения представляют собой серьезные трудности, требующие специальных методов предобработки [9]. При отсутствии этих этапов воспроизводимость исследования заметно снижается [12].

Ещё одним фактором, влияющим на прогностические характеристики модели, является выбор конечной точки [8, 12]. В исследованиях, посвящённых диабету и ССР, нередко используются комбинированные конечные точки, включающие инфаркт миокарда, инсульт, сердечно-сосудистую смерть и реваскуляризацию [8]. Такой подход удобен статистически, однако не всегда однороден с клинической точки зрения, поскольку факторы, определяющие риск сердечной недостаточности, могут отличаться от предикторов коронарных событий или инсульта [8, 10]. В связи с этим перспективным направлением представляется разработка моделей, ориентированных как на оценку интегрального риска, так и на развитие отдельных осложнений, что позволит точнее использовать прогнозы в повседневной практике [8, 10, 11, 15].

Само по себе повышение точности прогноза не может рассматриваться как достаточное основание для широкого внедрения алгоритмов в медицинскую практику [3, 5]. Существенно не только то, насколько хорошо модель предсказывает исход, но и то, насколько понятна логика её работы, возможно ли воспроизвести полученный результат на другой выборке и приводит ли применение такой модели к реальному улучшению клинических решений [3, 8]. Именно поэтому современная литература всё чаще концентрируется не столько на эффективности ML, сколько на качестве методологии, внешней валидации, воспроизводимости моделей в новых условиях и риске систематических ошибок [8].

Одним из наиболее значимых ограничений существующих исследований остаётся недостаточная внешняя валидация [3, 8]. Многие модели демонстрируют высокие значения площади под ROC-кривой в пределах одной базы данных, однако их эффективность нередко снижается при работе с другими популяциями [3, 8, 14]. По данным систематического обзора Nur A, et al. (2025), внешняя валидация неизменно сопровождается снижением прогностических характеристик, а зависимость моделей от лабораторных данных существенно ограничивает их применимость в условиях с различным ресурсным обеспечением [8]. Oikonomou EK, Khera R (2023) дополнительно указывают, что модели, обученные на рутинно собираемых клинических данных, воспроизводят систематические ошибки, накопленные в процессе реальной клинической практики, что указывает на риск переобучения и ограниченную универсальность алгоритмов [3, 8].

Помимо вышеуказанного, следует отметить проблему интерпретируемости результатов [3, 13]. Для клинициста принципиально важно понимание того, какие именно признаки определяют прогноз и подаются ли они клинической коррекции [13, 15]. Если модель остаётся "непрозрачным алгоритмом", доверие к ней снижается [3, 13]. В ответ на это активно развивается направление "интерпретируемого искусственного интеллекта": методы SHAP, LIME и иные подходы позволяют оценивать вклад отдельных признаков в итоговый прогноз и тем самым делать результат работы модели более доступным для понимания [3, 13, 15]. Это особенно важно в кардиологии и диабетологии, где оценка риска напрямую влияет на выбор характера лечения [1, 13]. Вместе с тем ML способно выявлять и менее очевидные комбинации факторов, которые трудно оценить в рамках традиционных статистических методов [5, 7]. Как указывают Бакулин Г.Г. и др. (2025) и Oikonomou EK, Khera R (2023), модели ML наиболее перспективны в задачах, где необходимо учитывать нелинейные взаимосвязи и интегрировать разнородные данные, тогда как их практическая реализуемость определяется возможностью служить инструментом поддержки клинических решений, а не их автоматизации [3, 5].

Во многих публикациях сохраняются проблемы несбалансированных выборок, пропущенных значений, различий в определении конечных точек и неполного описания этапов предобработки [8, 9]. При недостаточно корректном выполнении процедур предобработки высокие статистические показатели могут отражать не реальное качество алгоритма, а статистические артефакты, что обуславливает необходимость стандартизации разработки, отчётности и валидации прогностических моделей [3, 8]. Не менее существенным является организационный аспект: даже валидированная модель окажется малополезной, если её невозможно интегрировать в систему поддержки клини-

ческих решений, сопряжённую с ЭМК. Необходимо также учитывать медико-правовые и этические вопросы, включая ответственность за ошибочный прогноз, защиту персональных данных и возможность объяснить пациенту решение, принятое с учётом алгоритмической поддержки. Эти обстоятельства не уменьшают значимости технологии, однако показывают, что переход от экспериментальной разработки к полноценному клиническому инструменту представляет собой сложный, многоэтапный процесс [3, 8].

Качественная прогностическая система предполагает сочетание корректного дизайна исследования, обоснованного выбора признаков, клинической значимости конечных точек, строгой валидации и понятной интерпретации результатов [3, 7, 8]. ML при этом должно не противопоставляться клинической эпидемиологии, а дополнять её [5, 7]. При соблюдении этих условий модели ML действительно способны улучшать стратификацию риска у больных СД 2 типа и выявлять пациентов, нуждающихся в более активной профилактике [3, 5].

Заключение

Таким образом, ML открывает новые возможности для прогнозирования неблагоприятных сердечно-сосудистых событий у пациентов с СД 2 типа. Наиболее перспективными представляются модели, разработанные на больших клинических массивах, проверенные на независимых популяциях и дополненные инструментами объяснимого искусственного интеллекта. Их практическая значимость определяется не только потенциальным повышением точности, но и возможностью перехода к более индивидуализированной медицине, в рамках которой риск оценивается с учётом индивидуальных характеристик пациента. Вместе с тем было бы преждевременно рассматривать ML как альтернативу клиническому мышлению: при отсутствии внешней валидации, понятной интерпретации результатов и возможности интеграции в реальную практику даже качественный алгоритм рискует остаться экспериментальной разработкой. Дальнейшее развитие данной области связано не с полной автоматизацией принятия решений, а с рациональным взаимодействием врача и алгоритма: необходимы более строгие проспективные исследования, стандартизированная внешняя валидация и доказательства того, что использование моделей действительно улучшает клинические исходы, а не только статистические показатели.

Отношения и деятельность: все авторы заявляют об отсутствии потенциального конфликта интересов, требующего раскрытия в данной статье.

Декларация ИИ. Технологии генеративного ИИ использовались исключительно как вспомогательный инструмент на финальных стадиях работы над текстом рукописи.

Литература/References

- Joseph JJ, Deedwania P, Acharya T, et al. Comprehensive Management of Cardiovascular Risk Factors for Adults With Type 2 Diabetes: A Scientific Statement From the American Heart Association. *Circulation*. 2022;145(9):e722-e759. doi:10.1161/CIR.0000000000001040.
- GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet*. 2023;402(10397):203-34. doi:10.1016/S0140-6736(23)01301-6.
- Oikonomou EK, Khera R. Machine learning in precision diabetes care and cardiovascular risk prediction. *Cardiovasc Diabetol*. 2023;22(1):259. doi:10.1186/s12933-023-01985-3.
- Dziopa K, Asselbergs FW, Gratton J, et al. Cardiovascular risk prediction in type 2 diabetes: a comparison of 22 risk scores in primary care settings. *Diabetologia*. 2022;65(4):644-56. doi:10.1007/s00125-021-05640-y.
- Bakulin GG, Serezhina EK, Obrezan AG. Analysis of current approaches to cardiovascular risk assessment: the contribution of traditional risk scales and machine learning. *Kardiologiya: novosti, mneniya, obucheniye* [Cardiology: News, Opinions, Training]. 2025;13(4):26-33. (In Russ.) Бакулин Г.Г., Серезина Е.К., Обрезан А.Г. Анализ актуальных подходов к оценке сердечно-сосудистого риска: вклад традиционных шкал риска и машинного обучения. *Кардиология: новости, мнения, обучение*. 2025;13(4):26-33. EDN: HRUBSQ.
- Kwiendacz H, Wijata AM, Nalepa J, et al. Machine learning profiles of cardiovascular risk in patients with diabetes mellitus: the Silesia Diabetes-Heart Project. *Cardiovasc Diabetol*. 2023;22(1):218. doi:10.1186/s12933-023-01938-w.
- Hunter DJ, Holmes C. Where Medical Statistics Meets Artificial Intelligence. *N Engl J Med*. 2023;389(13):1211-9. doi:10.1056/NEJMr2212850.
- Nur A, Tjandra S, Yumnanisha DA, et al. Predicting the risks of stroke, cardiovascular disease, and peripheral vascular disease among people with type 2 diabetes with artificial intelligence models: a systematic review and meta-analysis. *Narra J*. 2025;5(1):e2116. doi:10.52225/narra.v5i1.2116.
- Xu C, Shi F, Ding W, et al. Development and validation of a machine learning model for cardiovascular disease risk prediction in type 2 diabetes patients. *Sci Rep*. 2025;15(1):32818. doi:10.1038/s41598-025-18443-7.
- Liu Y, Wang P, Wang M, et al. Development and External Validation of a Machine Learning-Based Model for Predicting Heart Failure Risk in Type 2 Diabetes. *Diabetes Metab Syndr Obes*. 2025;18:4177-91. doi:10.2147/DMSO.S558687.
- Tang D, Liang F, Gu X, et al. Exploration and analysis of risk factors for coronary artery disease with type 2 diabetes based on SHAP explainable machine learning algorithm. *Sci Rep*. 2025;15(1):29521. doi:10.1038/s41598-025-11142-3.
- Kee OT, Harun H, Mustafa N, et al. Cardiovascular complications in a diabetes prediction model using machine learning: a systematic review. *Cardiovasc Diabetol*. 2023;22(1):13. doi:10.1186/s12933-023-01741-7.
- Salih AM, Galazzo IB, Gkontra P, et al. A review of evaluation approaches for explainable AI with applications in cardiology. *Artif Intell Rev*. 2024;57(9):240. doi:10.1007/s10462-024-10852-w.
- Sang H, Lee H, Lee M, et al. Prediction model for cardiovascular disease in patients with diabetes using machine learning derived and validated in two independent Korean cohorts. *Sci Rep*. 2024;14(1):14966. doi:10.1038/s41598-024-63798-y.
- Tang Y, Zheng X, Yang X, et al. Machine learning-based integration of pericoronary adipose tissue and clinical risk factors for cardiovascular risk prediction in type 2 diabetes: a retrospective cohort study. *Eur J Med Res*. 2025;30(1):961. doi:10.1186/s40001-025-03237-4.

Адреса организаций авторов: ФГБОУ ВО Воронежский государственный медицинский университет им. Н. Н. Бурденко Минздрава России, Студенческая ул., д. 10, Воронеж, 394036, Россия; ООО "Медицинский центр Широких Сердец", Пограничная ул., д. 2, Воронеж, 394000, Россия.

Addresses of the authors' institutions: Burdenko Voronezh State Medical University, Studencheskaya str., 10, Voronezh, 394036, Russia; ООО "Shirokih Serdets" Medical Center, Pogranichnaya str., 2, Voronezh, 394000, Russia.