

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В ДИФФЕРЕНЦИАЛЬНОЙ ДИАГНОСТИКЕ СЛОЖНО КЛАССИФИЦИРУЕМЫХ ТИПОВ САХАРНОГО ДИАБЕТА

© Н.В. Русыева*, И.И. Голодников, И.В. Кононенко, Т.В. Никонова, М.В. Шестакова

ГНЦ РФ ФГБУ «Национальный медицинский исследовательский центр эндокринологии», Москва

Течение сложно классифицируемых типов сахарного диабета (СД) (медленно развивающийся иммуноопосредованный СД взрослых (LADA), моногенные формы СД (MODY)) имеет общие черты как с СД 1 типа (СД1), так и с СД 2 типа (СД2), поэтому зачастую остаются неверно диагностированными. Ошибки в определении типа диабета влекут за собой неверную тактику лечения, что приводит к плохому контролю гликемии, развитию осложнений, снижению качества жизни пациента, увеличению смертности.

Ключевым методом диагностики MODY служит секвенирование генов, ассоциированных с этим заболеванием, а LADA — иммунологический анализ крови в совокупности с особенностями клинической картины. Однако до сих пор не определены точные критерии для направления пациентов на данные исследования. Выполнение данных исследований всем без исключения пациентам с факторами риска может привести к неоправданным экономическим затратам, кроме того, доступ к ним зачастую затруднен. В связи с этим разработаны различные автоматизированные алгоритмы на основе статистических методов и машинного обучения (глубокие нейросети, «деревья решений» и др.) для выделения пациентов, которым наиболее оправданно проведение углубленного обследования. Среди них — алгоритмы дифференциальной диагностики СД1 и СД2, алгоритмы, специализирующиеся на диагностике только LADA или только MODY, лишь один алгоритм направлен на мультиклассовую классификацию пациентов с СД. Широко применяется один из алгоритмов, направленный на диагностику MODY у пациентов в возрасте до 35 лет. Однако существующие алгоритмы имеют ряд недостатков, как-то: малый размер выборки, исключение из исследования пациентов с MODY или пациентов более старшего возраста, отсутствие верификации диагноза с помощью соответствующих исследований, использование поздних осложнений СД в качестве параметров для диагностики. Зачастую в группу исследователей не входили практикующие врачи. Кроме того, ни один из алгоритмов не находится в открытом доступе и не протестирован для пациентов в России. В данной рукописи представлен анализ основных автоматизированных алгоритмов дифференциальной диагностики СД, разработанных в последние годы.

КЛЮЧЕВЫЕ СЛОВА: сахарный диабет; дифференциальная диагностика; машинное обучение; алгоритм; латентный аутоиммунный диабет взрослых; диабет зрелого типа у молодых

MACHINE LEARNING METHODS IN THE DIFFERENTIAL DIAGNOSIS OF DIFFICULT-TO-CLASSIFY TYPES OF DIABETES MELLITUS

© Nadezhda V. Rusyaeva*, Ivan I. Golodnikov, Irina V. Kononenko, Tatiana V. Nikonova, Marina V. Shestakova

Endocrinology research centre, Moscow

The course of difficult-to-classify types of diabetes mellitus (DM) (slowly developing immune-mediated DM of adults (LADA), monogenic forms of DM (MODY)) has common features with both type 1 DM (T1DM) and type 2 DM (T2DM), so often remain misdiagnosed. Errors in determining the type of diabetes lead to incorrect treatment tactics, which leads to poor glycemic control, the development of complications, a decrease in the patient's quality of life, and increased mortality.

The key method for diagnosing MODY is sequencing of genes associated with this disease, and LADA is an immunological blood test in combination with the features of the clinical picture. However, the exact criteria for referring patients to these studies have not yet been determined. Performing these studies on all patients without exception with risk factors can lead to unjustified economic costs, and access to them is often difficult. In this regard, various automated algorithms have been developed based on statistical methods and machine learning (deep neural networks, "decision trees", etc.) to identify patients for whom an in-depth examination is most justified. Among them are algorithms for the differential diagnosis of T1DM and T2DM, algorithms specializing in the diagnosis of only LADA or only MODY, only one algorithm is aimed at multiclass classification of patients with diabetes. One of the algorithms is widely used, aimed at diagnosing MODY in patients under the age of 35 years. However, existing algorithms have a number of disadvantages, such as: small sample size, exclusion of patients with MODY or older patients from the study, lack of verification of the diagnosis using appropriate studies, and the use of late complications of diabetes as parameters for diagnosis. Often the research team did not include practicing physicians. In addition, none of the algorithms are publicly available and have not been tested for patients in Russia. This manuscript presents an analysis of the main automated algorithms for the differential diagnosis of diabetes, developed in recent years.

KEYWORDS: diabetes mellitus; differential diagnosis; machine learning; algorithm; latent autoimmune diabetes of adults; maturity-onset diabetes of the young

ВВЕДЕНИЕ

По данным Росстата, общая численность пациентов с сахарным диабетом (СД) в России составляет 5,17 млн человек (3,23% всего населения), и это число неуклонно растет [1]. Под термином «сахарный диабет» объединены заболевания с различной этиологией, общим для которых является развитие стойкой гипергликемии, которая приводит к повреждению различных тканей и систем органов. В настоящее время Российской ассоциацией эндокринологов рекомендована классификация СД, принятая Всемирной организацией здравоохранения (ВОЗ) в 1999 г., согласно которой выделяют: СД 1 типа (СД1), СД 2 типа (СД2), гестационный СД (ГСД), другие специфические типы СД. При этом выделяют два подтипа СД1: иммуноопосредованный и идиопатический [2]. В 2019 г. ВОЗ была предложена новая классификация СД, где в отдельную группу выделены гибридные типы СД: медленно развивающийся иммуноопосредованный СД взрослых, или LADA (latent autoimmune diabetes in the adults — латентный аутоиммунный СД взрослых), а также СД2 со склонностью к кетозу [3].

Предпосылкой к выделению новых типов СД стала гетерогенность СД, обусловленная фенотипическим, генетическим, иммунологическим различиями. Некоторые исследователи предлагают альтернативные, более гибкие способы охарактеризовать пациента с СД, как, например, в многомерной модели «палитры», где «основные цвета» — это разные патофизиологические процессы, вносящие вклад в развитие СД (исходная масса β -клеток, их функция, аутоиммунитет, инкретиновая активность, ожирение, инсулинорезистентность и т.д.), и каждый пациент характеризуется различными «оттенками» этих «цветов» [4, 5].

Среди взрослых пациентов с аутоиммунным СД встречаются случаи медленного развития заболевания без ярких клинических проявлений в дебюте с отсутствием инсулинопотребности дольше 6 мес. Этот подтип СД был назван медленно прогрессирующим аутоиммунным СД взрослых (LADA). Таким образом, СД1 — это некий спектр, на одном полюсе которого располагается классический СД1 с яркими, быстро развивающимися проявлениями инсулиновой недостаточности, на другом — LADA с его мягким течением. Более высокая скорость снижения функции β -клеток ассоциирована с более молодым возрастом в дебюте, меньшим индексом массы тела (ИМТ), гаплотипами HLA DRB1 и DQB1 у европеоидов, особенностями гуморального и клеточного иммунитета [6–8]. Предлагаются различные подходы к более точной классификации СД1. По аналогии с эндотипами при бронхиальной астме некоторые исследователи предлагают выделение эндотипов СД1, в том числе с помощью кластеризации на основе вышеупомянутой модели «палитры». Концепция эндотипов основана на возможности развития одного и того же заболевания в исходе разных патофизиологических процессов на фоне разной генетической предрасположенности. Авторы предполагают, что можно выделить как минимум два эндотипа: «проинсулиновый аутоиммунный DR4» и «GAD аутоиммунный DR3» [9]. Такой подход в перспективе может стать «золотой серединой» между стандартным и индивидуализированным отношением к каждому пациенту.

Популяция пациентов с СД2 также неоднородна по антропометрическим показателям, выраженности инсулино-

дефицита и инсулинорезистентности, характеру течения заболевания, тяжести осложнений, ответу на сахароснижающие препараты. Эти различия становятся основанием для поиска новых подтипов СД2. Важным шагом стало исследование E. Ahlqvist и соавт., в котором проведен кластерный анализ данных 8980 пациентов из базы ANDIS (All New Diabetics in Scania) [10]. В анализ были включены такие параметры, как возраст в дебюте, ИМТ, уровень гликированного гемоглобина (HbA_{1c}), наличие антител к глутаматдекарбоксилазе (Ат к GAD), индексы HOMA2-IR и HOMA2- β (Homeostasis model assessment of insulin resistance/ β -cells — гомеостатическая модель оценки инсулинорезистентности/функции β -клеток). В результате выделено 5 подтипов СД. Все пациенты с аутоиммунным диабетом были отнесены в кластер тяжелого аутоиммунного диабета (severe autoimmune diabetes — SAID); в то время как СД2 и другие типы неиммунного диабета были разделены на 4 кластера: тяжелый инсулинодефицитный диабет (severe insulin-deficient diabetes — SIDD); тяжелый инсулинорезистентный диабет (severe insulinresistant diabetes — SIRD); мягкий, связанный с ожирением диабет (mild obesity-related diabetes — MOD); мягкий, связанный с возрастом диабет (mild age-related diabetes — MARD). Результаты были воспроизведены на других данных, в том числе данных пациентов со средней длительностью СД 10,2 года. Худший гликемический контроль и быстро развивающаяся потребность в инсулине наблюдались в группе SIDD. Выявлена ассоциация выделенных подтипов СД с определенными вариантами генов высокого риска СД, а также с темпом развития осложнений: у пациентов с SIRD чаще и быстрее прогрессировала нефропатия, хотя гликемический контроль не отличался от других групп, что может свидетельствовать о необходимости раннего назначения нефропротективной терапии в этой подгруппе. X.-F. Xiong и соавт. повторили исследование на китайской популяции, добавив к анализу уровень триглицеридов (ТГ) и мочевой кислоты, и выделили еще два подтипа: СД, связанный с отягощенной наследственностью (inheritance-related), и СД, связанный с уровнем мочевой кислоты (uric acid-related) и ассоциированный с повышенным риском ишемической болезни сердца, цереброваскулярной болезни и хронической болезни почек [11].

СЛОЖНОСТИ В ОПРЕДЕЛЕНИИ ТИПА САХАРНОГО ДИАБЕТА У МОЛОДЫХ ВЗРОСЛЫХ ПАЦИЕНТОВ

Проблема точного определения типа СД наиболее актуальна в группе взрослых пациентов молодого возраста (18–45 лет по ВОЗ), когда могут манифестировать не только классический СД1 и СД2, но и LADA, и MODY (Maturity-onset diabetes of the young — диабет взрослого типа у молодых). По разным данным, у 7,4–26,2% молодых пациентов тип СД определен неверно [12, 13]. Диагностика LADA и MODY затруднена ввиду неспецифической клинической картины, недостаточной осведомленности практикующих врачей, отсутствия четко прописанного алгоритма диагностики, низкой доступности иммунологического и молекулярно-генетического исследования (МГИ). Случаи «СД не-1 типа» (без типичных симптомов декомпенсации углеводного обмена) попадают в категорию СД2. Кроме того, к ошибкам может привести и неправильная интерпретация лабораторных данных.

Дифференциальная диагностика LADA

Распространенность LADA составляет 2–12% общего количества случаев СД [14]. По данным R. Cheheltani и соавт., у 10% пациентов с СД1 первично был диагностирован СД2 [15]. Вероятно, эти данные можно распространить и на популяцию пациентов с LADA, так как именно этот тип в дебюте протекает подобно СД2. По другим данным, почти 40% взрослых пациентов с СД1 первично поставлен неправильный диагноз и у 4–14% пациентов с СД2 выявлены антитела к β -клеткам [16, 17]. Ошибки в диагностике и тактике лечения LADA во многом обусловлены представлениями клиницистов о течении аутоиммунного СД. Латентное течение LADA в дебюте при отсутствии инсулинопотребности и признаков острой метаболической декомпенсации не укладывается в стереотипную клиническую картину СД1 [18]. Кроме того, СД1 традиционно считается заболеванием детей и подростков, хотя около половины пациентов заболевают в возрасте старше 18 лет [19].

Хотя LADA, согласно актуальной на текущий момент классификации 1999 г., относится к СД1, при переходе на новую классификацию этот тип СД будет выделен в отдельную категорию. Поэтому следует отметить основные отличия LADA от классического СД1: возраст дебюта старше 30 лет, длительно сохраняющаяся функция β -клеток (медленный темп снижения уровня С-пептида, отсутствие инсулинопотребности более 6 мес после дебюта). Чаще, чем при СД1, наблюдаются признаки инсулинорезистентности, метаболический синдром [20].

Возраст в дебюте менее 50 лет, аутоиммунные заболевания у пациента или его родственников, отсутствие метаболического синдрома и наличие антител (Ат) к β -клеткам (в 90% случаев — к GAD) позволяют заподозрить аутоиммунный генез СД у пациентов с фенотипом СД2 [21]. Ат к GAD считаются наиболее чувствительным маркером LADA при дифференциальной диагностике с СД2 [22, 23].

Диагностическая ценность определения базального С-пептида ограничена, так как его уровень может оставаться относительно сохранным длительное время. Однако при его повышенном уровне LADA можно исключить с высокой вероятностью [24]. Кроме того, на уровень С-пептида может влиять нарушение функции почек [25]. Уровень стимулированного С-пептида при LADA ниже, чем при СД2, но выше, чем при классическом СД1, однако четкие пороговые уровни его не установлены [26].

Дифференциальная диагностика MODY

Распространенность MODY относительно невысока (0,3–5% всех случаев СД) [27], но в масштабах российской популяции количество таких пациентов велико. Учитывая мягкое, бессимптомное течение основных типов MODY, отсутствие антител и склонности к кетоацидозу, сохранный уровень С-пептида, часто диагностируется СД2, однако из-за молодого возраста в дебюте и нормальной массы тела может быть диагностирован и СД1. До установления диагноза MODY проходит в среднем 10 лет, около 80% случаев остаются недиагностированными [28]. У 36% пациентов с MODY диагностируется СД1, у 51% — СД2, и лишь у 6% пациентов этот диагноз был верно установлен изначально [29]. Клиницист дол-

жен обращать внимание на такие признаки, как ранний дебют СД (чаще до 35 лет), умеренная гипергликемия натощак, высокая чувствительность к препаратам сульфонилмочевины (ПСМ), отягощенный семейный анамнез СД с аутосомно-доминантным характером наследования, сведения о неонатальном диабете/гипогликемии в семье, наличие пороков развития половых органов и кист почек. Отличия от СД1 — отсутствие аутоиммунитета к β -клеткам, сохранный секретин инсулина по истечении предполагаемого «медового месяца» (более 3 лет), низкая потребность в инсулине, отсутствие склонности к кетоацидозу; от СД2 — дебют в молодом возрасте, отсутствие ожирения, признаков инсулинорезистентности, дислипидемии [2, 30]. Однако все эти признаки недостаточно чувствительны и специфичны, и единственным достоверным методом диагностики MODY остается молекулярно-генетическое исследование.

Два наиболее распространенных типа MODY — MODY 2 (29,7–32%) и MODY 3 (31–62%) [28, 29, 31]. Еще 12 типов MODY встречаются существенно реже [32]. MODY 2 связан с мутацией гена глюкокиназы (GCK), при снижении активности которой развивается мягкая гипергликемия натощак. MODY 2 часто диагностируется в возрасте до 30 лет при случайном обследовании, например при прегравидарной подготовке или во время беременности. HbA_{1c} чаще не превышает 7,5%, компенсация достигается с помощью диеты, риск развития осложнений низкий. MODY 3 ассоциирован с мутацией в гене ядерного фактора гепатоцитов 1A (HNF1A) и манифестирует у 63% уже к 25 годам, еще у 16% — к 35 годам [33]. Дебют зачастую связан с повышением инсулинорезистентности (пубертат, ожирение) [34]. Характерны преимущественно постпрандиальная гипергликемия, глюкозурия при эугликемии. Чувствительность к ПСМ при MODY 3 в 4 раза выше, чем при СД2, и они являются препаратами выбора [35]. Однако прогрессирующее снижение секреции инсулина на 1–4% в год может привести к назначению инсулинотерапии [36]. Некоторые мутации HNF1A не приводят к развитию MODY 3, а лишь предрасполагают к СД2 [37]. В этом случае диабет выявляется в более старшем возрасте, нет повышенной чувствительности к ПСМ, не так очевиден аутосомно-доминантный тип наследования.

Пациенты с разными типами СД требуют разных терапевтических подходов. Сообщается, что даже пациенты из разных кластеров СД2 (по E. Ahlqvist) неодинаково реагируют на сахароснижающие препараты как в клинических исследованиях, так и в рутинной практике. Поэтому необходима точная дифференциальная диагностика типа СД у каждого пациента как при первичной диагностике, так и при длительном течении заболевания. Упрощение и автоматизация этого процесса возможны с помощью различных калькуляторов на основе математических методов. Наиболее передовой технологией в этой сфере являются методы искусственного интеллекта (ИИ), с помощью которых можно приблизиться к максимальной точности диагностики. Персонализация подхода к пациентам с СД с помощью ИИ отвечает современным тенденциям и стремлению эндокринологического сообщества к прецизионной диабетологии («precision diabetology»).

МАШИННОЕ ОБУЧЕНИЕ

Машинное обучение (МО) — это подвид ИИ, раздела информатики, который занимается созданием компьютерных систем, способных решать задачи, предназначенные для человеческого интеллекта (понимание языка, способность рассуждать). От обычных компьютерных программ МО отличается способностью обучаться на большом количестве однотипных задач, находя оптимальное решение и минимизируя количество ошибок. Основным преимуществом методов МО перед классическими статистическими методами является их гибкость, отсутствие заранее прописанных четких правил «если А, то Б», способность обрабатывать более сложные массивы данных и большее количество параметров, поэтому они способны обеспечить гораздо большую точность диагностики и прогнозирования заболеваний, чем классические статистические методы, помочь специалисту в выборе лечения, контроле заболевания [38, 39]. Алгоритмы МО, внедренные в электронные медицинские карты в виде систем поддержки принятия врачебных решений (СППВР), могут использоваться в качестве вспомогательного инструмента для скрининга заболеваний и выявления лиц из группы риска, планирования лечения и диспансерного наблюдения. Данные отечественной и зарубежной литературы отражают быстрое развитие методов МО в диабетологии [40, 41]. Особого внимания заслуживают СППВР, способные предлагать врачу или пациенту «подсказки» по образу жизни, предотвращению гипогликемии, коррекции сахароснижающей терапии, лекарственным взаимодействиям.

Виды алгоритмов машинного обучения

Основные разновидности МО — это классическое МО «с учителем» и «без учителя», МО «с подкреплени-

ем», ансамбли методов и глубокое обучение (ГО), которые различаются по структуре, функциональности, формату исходных данных, зависимости процесса обучения от разработчика, «прозрачности» работы, объему необходимых ресурсов. Архитектура алгоритмов классического МО должна соответствовать типу поставленной перед ним задачи, а входные данные для них должны быть представлены структурированно, «дискретно», то есть заранее обработаны и подготовлены исследователем. Алгоритмы ГО, основанные на искусственных нейронных сетях (НС), могут анализировать данные в их естественном, «сыром» виде, что типично для медицинской информации (текст в электронной медицинской карте, изображения, полученные с помощью визуализирующих методов). Однако НС ГО сложны в реализации, требуют намного больше материально-технических ресурсов и гораздо большего объема данных для обучения.

Отнесение пациента к одной из заранее известных категорий (типов СД) — одна из задач классификации, для решения которых чаще всего используются следующие алгоритмы: линейная и логистическая регрессия, метод k-ближайших соседей (k-Nearest Neighbors), метод опорных векторов (Support Vector Machine), «дерева решений» (Decision Tree) и его производные, например, метод «случайного леса» (Random Forest), наивный байесовский классификатор (Naive Bayes), НС.

Приведем краткое описание принципа работы наиболее часто используемых алгоритмов классификации. Модели на основе «дерева решений» наиболее привлекательны для построения дифференциально-диагностического алгоритма, так как формируют привычные для клинициста блок-схемы с дихотомическим делением в узлах, подобные тем, которые иллюстрируют алгоритм действий врача при том или ином заболевании в клинических рекомендациях (рис. 1).

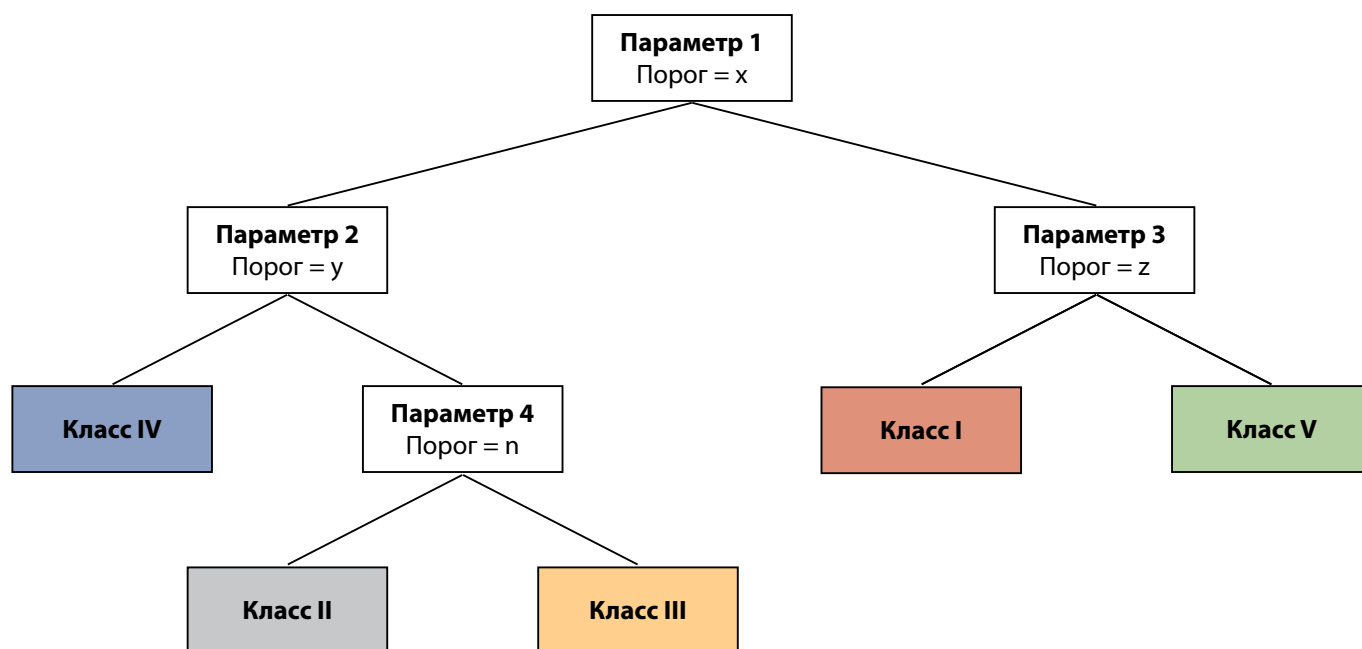


Рисунок 1. Схематическое изображение принципа работы алгоритма «дерева решений». Алгоритм выделяет значимые параметры и их иерархию, а также самостоятельно подбирает оптимальные пороговые уровни каждого параметра, которые дискриминируют два или несколько классов объектов.

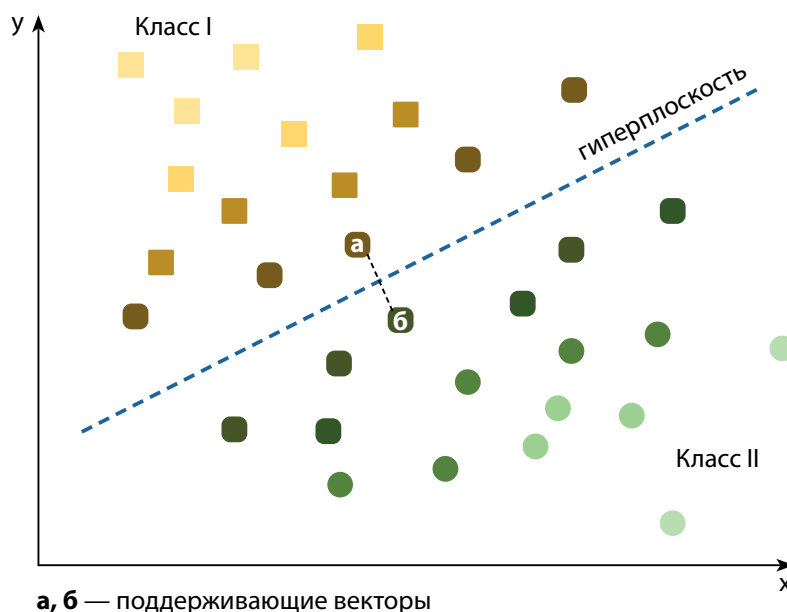


Рисунок 2. Схематическое изображение принципа работы алгоритма k-ближайших соседей. Гиперплоскость максимально эффективно разделяет два класса объектов. Объекты, лежащие близко к границе по обе стороны от нее и имеющие признаки обоих классов, называют поддерживающими векторами.

Метод опорных векторов основан на другом принципе — размещении объектов в пространстве признаков и вычислении оптимальной прямой (или плоскости, если речь идет о трехмерной модели), разделяющей группы объектов (рис. 2). Алгоритм k-ближайших соседей относит объект в определенную группу на основе сравнения с определенным количеством (k) его ближайших «соседей» в пространстве признаков.

НС — сложные системы связей между «нейронами», подобные естественным НС человеческого мозга. Они имеют входной, выходной слой «нейронов»

и скрытые слои (рис. 3). Если скрытых слоев больше одного, то НС считаются «глубокими». НС способны решать практически любые задачи, в том числе классификационные, но менее распространены, чем алгоритмы других типов, так как для их разработки необходимы значительные ресурсы и большой объем данных. Их обучение происходит методом обратного распространения ошибки и меньше зависит от активного вмешательства человека. НС вычленяет значимые факторы, находит взаимосвязи между ними и определяет их вес в принятии решения.

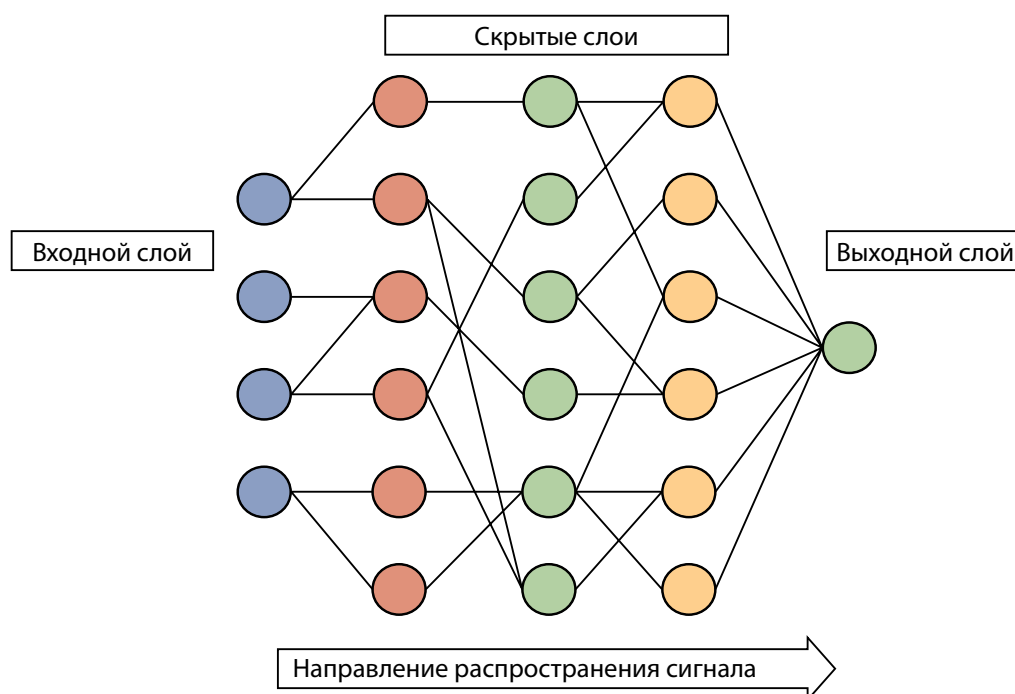


Рисунок 3. Схематическое изображение нейронной сети.

Круги соответствуют отдельным обрабатывающим блокам алгоритма («нейронам»), которые выстроены в слои и взаимосвязаны друг с другом прямыми и обратными связями.

В основе алгоритмов МО лежит общий принцип: учитывая свойства каждого объекта в базе данных (входные параметры), они обучаются находить оптимальное решение задачи путем поиска значимых для результата параметров, связей между ними и других закономерностей. Процесс обучения происходит на обучающей выборке, затем на проверочной выборке проводится валидация алгоритма. От выбранных входных параметров, качества базы данных, объема и состава выборок зависят показатели эффективности будущего алгоритма: точность (precision, accuracy), полнота (recall), F1-мера, площадь под ROC (AUROC), под кривой Precision-Recall (AUPRC). Все метрики должны оцениваться комплексно.

Применение компьютерных методов для дифференциальной диагностики типов сахарного диабета

Накопленные базы данных о пациентах с СД, насчитывающие десятки тысяч записей, обеспечивают достаточный объем данных для обучения алгоритма. Предложено немало моделей прогнозирования СД, бинарной классификации типов СД и прогнозирования диабетических осложнений. В качестве входных параметров для классификации СД принимаются самые разнообразные, в том числе неочевидные показатели. Большая часть исследований компьютерных технологий сосредоточена на дифференциальной диагностике лишь двух типов диабета — СД1 (без выделения подгруппы LADA) и СД2. Для решения таких задач используются в основном алгоритмы классического МО, как более простые в реализации.

Алгоритмы дифференциальной диагностики СД1 и СД2

А. Lynam и соавт. сравнивали эффективность метода логистической регрессии с 6 алгоритмами МО для дифференциальной диагностики СД1 и СД2 у пациентов 18–50 лет [42]. Объем обучающей выборки составил 960 человек, 14% из которых имели СД1. В качестве входных параметров использовались пол, возраст в дебюте СД, ИМТ, наличие Ат к GAD, уровень общего холестерина (ХО), липопротеидов высокой плотности (ЛПВП), ТГ. Общая дискриминационная способность модели оказалась наибольшей у метода логистической регрессии (AUROC 0,95 vs $\geq 0,92$ у алгоритмов). Клиническая ценность методов продемонстрирована с помощью анализа кривой принятия решений.

Группа Qureshi разработала метод классификации СД, используя базу данных из 2000 записей [43]. Первоначально отобрано 19 параметров: ИМТ, HbA_{1c}, инсулин натощак, уровни С-пептида и глюкозы в пероральном глюкозотолерантном тесте (ПГТТ), креатинин, ХО, ЛПВП, липопротеиды низкой плотности (ЛПНП), ТГ, тестостерон, Ат к GAD, к тирозинфосфатазе (IA-2), наличие микроальбуминурии, кетонурия, систолическое и диастолическое артериальное давление (АД). Из перечисленных параметров с помощью анализа главных компонент выбрано 12 наиболее значимых. В сравнении с методом k-средних, k-ближайших соседей, наивным байесовским классификатором и «деревом решений» метод опорных векторов показал наибольшую точность — 98%, а его чувствительность и специфичность составили 94 и 97% соответственно.

В финском исследовании P. Romantschuk собраны данные о 2224 пациентах с дебютом СД в возрасте от 16 лет из баз DIREVA и HUCH T1DM [44]. Проанализированы ИМТ, С-пептид, глюкоза плазмы натощак (ГПН), АД, ЛПНП, ЛПВП, ТГ, Ат к GAD в дебюте и С-пептид в течение 3–5 лет после установления диагноза. Показатель соотношения С-пептида с ГПН оказался наиболее значимым различием между группами. Неожиданным результатом анализа стало отсутствие различий по АД и липидному профилю, что, по мнению автора, обусловлено тем, что в анализ не вошел фактор наличия антигипертензивной и гиполипидемической терапии. Алгоритмы «дерева решений», в том числе «случайного леса», были обучены на 312 записях (153 пациента с СД1, 159 — с СД2), имеющих все необходимые данные (возраст дебюта, ИМТ, Ат к GAD, С-пептида в дебюте и на момент исследования). Сгенерирован алгоритм со следующими точками ветвления: соотношение С-пептид/глюкоза (cut-off 3,83 нмоль/ммоль), возраст (36,5 года), ИМТ (32,6 кг/м²). При тестировании на пациентах без сведений об антителах и уровне С-пептида алгоритм показал общую точность 91,83%, чувствительность 89,97%, специфичность 92,44%, AUROC 0,905.

Примечательны результаты китайского исследования, в котором использованы параметры кривых, полученных при непрерывном мониторинговании гликемии (288 измерений в сутки) [45]. Выделено 17 признаков из 1050 кривых: средний уровень глюкозы в течение суток, за 1 ч до еды, через 3 ч после еды, количество отклонений от порогового уровня вверх и вниз, длительность нахождения в целевом диапазоне, ниже и выше порогового уровня, площадь под кривой, стандартное отклонение уровня глюкозы, средняя и максимальная амплитуда перепадов уровня глюкозы, абсолютные средние различия между разными днями. Был обучен ансамбль методов «double-Class AdaBoost», затем оценивалась его эффективность при разных пороговых уровнях глюкозы. Максимальная эффективность классификации СД достигнута при пороговых значениях глюкозы 7 или 8 ммоль/л: процент ошибок составил 9,7%, точность — 90,3%.

R. Oram и соавт. разработали показатель генетического риска СД1 (T1D-GRS), который основан на анализе однонуклеотидных полиморфизмов (SNP), ассоциированных с СД [46]. Этот показатель с высокой точностью AUROC 0,88 (95% ДИ 0,87–0,89, $p < 0,0001$) выявлял СД1 среди сложно классифицируемых пациентов 20–40 лет европейской популяции. Определение SNP у пациентов с СД пока не применяется в рутинной практике, однако следует обратить внимание на другие результаты этого исследования: с помощью логистической регрессии и ROC-анализа выявлен независимый вклад в результат возраста в дебюте, ИМТ, специфических антител. AUROC этих трех показателей составила 0,94 (95% ДИ 0,89–0,98), а при добавлении к ним T1D-GRS увеличивалась до 0,96 (95% ДИ 0,94–0,99). На основе полученных данных создан инструмент для оценки вероятности СД1 и СД2, размещенный по URL-адресу: <https://www.diabetesgenes.org/t1dt2d-prediction-model/>. Из анализа исключены случаи с подтвержденным MODY, что является основным недостатком исследования.

На основе данных телефонного опроса и 129 684 записей из медицинского архива также были разработаны 2 алгоритма МО, однако они могут служить лишь для

административных целей и в рутинной практике не применимы, так как одним из входных параметров в этих алгоритмах является уже поставленный ранее диагноз [47, 48].

Необходимо отметить, что в перечисленных алгоритмах не учитываются гетерогенность СД1 и возможность наличия MODY у пациентов, поэтому они не отвечают задачам диагностики более редких типов СД. Также следует учесть, что в качестве входных параметров были использованы результаты дорогостоящих иммунологического и генетического исследований, а также метод непрерывного мониторирования глюкозы — все эти методы зачастую труднодоступны в рутинной практике. Кроме того, некоторые авторы концентрируются преимущественно на технических нюансах разработки алгоритмов и зачастую недостаточно глубоко затрагивают клинические аспекты СД.

Алгоритмы диагностики LADA

В исследовании S. Furlanos и соавт. проведен ретроспективный анализ пациентов с LADA (n=102) и СД2 (n=111) в возрасте 30–75 лет [49]. Информация была получена с помощью опроса. У пациентов с LADA чаще встречались: дебют в возрасте <50 лет, полидипсия, полиурия, полифагия, снижение массы тела, ИМТ < 25 кг/м², аутоиммунные заболевания у пациента или его родственников. На основании числа этих критериев у каждого пациента была сформирована шкала «Клинического риска LADA». При проспективном анализе 130 пациентов с дебютом СД наличие 2 из 5 критериев по этой шкале показало 90% чувствительность и 71% специфичность в отношении диагностики LADA, негативная прогностическая ценность при наличии ≤1 критерия составила 99%.

В ретроспективном исследовании Z. Wang сравнивались 880 пациентов с СД2 и 50 — с LADA, причем были использованы и те параметры, которые довольно редко оцениваются в рутинной практике [50]. Не выявлено статистических различий между группами LADA и СД2 по полу, этнической принадлежности (94–97% пациентов принадлежали к народу Хань), уровню образования, выраженности полидипсии, полиурии, полифагии, степени снижения массы тела, наличию «мушек» в поле зрения, онемению рук и ног, артериальной гипертензии, ишемической болезни сердца, дислипидемии, семейному анамнезу, употреблению алкоголя, ИМТ, соотношению обхвата талии к обхвату бедер, уровню инсулина натощак. При помощи простой логистической регрессии выявлена корреляция LADA с возрастом, кетозом, курением, HbA_{1c} показателями глюкозы в ПГТТ (натощак, в точках 60 и 120 мин), С-пептида в точках 60 и 120 мин, AUC показателей глюкозы и С-пептида, а также HOMA2-β (p<0,05). Затем были исследованы взаимосвязи между этими факторами и выделены 5 наиболее клинически значимых параметров: пол, кетоз, курение, AUC глюкозы плазмы в точке 60 мин и AUC С-пептида в точке 120 мин, на основе чего была построена модель множественной логистической регрессии, которая достигала чувствительности 78,57% и селективности 67,96% при Y=0,0472, AUROC 0,757 (p<0,05).

В 2022 г. R. Cheheltani и соавт. предложили модель, основанную на алгоритме градиентного бустинга (XGBoost) для выявления случаев СД1 среди пациентов с диагнозом

СД2 [15]. Проводился ретроспективный анализ данных более 737 000 пациентов о возрасте, демографических показателях, факторах риска, симптомах, лечении, результатах физикального осмотра, лабораторных и инструментальных исследований. Алгоритм выявил наиболее значимые предикторы ошибочного диагноза СД2: возраст, ИМТ, масса тела, особенности лечения, уровни HbA_{1c}, глюкозы крови. AUC модели была равна 0,81, при полноте 50% доля ложноположительных результатов составила 7%, а при полноте 10% точность составила 17%. Исследователи предполагают, что этот алгоритм способен примерно в 70 раз сократить группу пациентов, которым оправданно проведение иммунологического скрининга. Данный алгоритм предложен для внедрения в медицинскую информационную систему в США в качестве СППВР. Однако в этой работе из анализа были исключены пациенты с MODY и не учтена гетерогенность СД1.

A. Miller и соавт. использовали алгоритмы на основе метода опорных векторов, «случайного леса», метода k-ближайших соседей, НС (многослойного персептрона (НС с простейшей архитектурой) и NeuralNet) и наивного байесовского классификатора для анализа данных 1176 пациентов в возрасте 30–70 лет [51]. Исследователи ставили задачу оценить прогностическую ценность отдельных клинических признаков: возраста, пола, ИМТ, уровня холестерина, отягощенной наследственности, объема талии, причем наиболее значимым оказался последний показатель. Наибольшие показатели эффективности были достигнуты при использовании НС: точность (accuracy) 85,51%, чувствительность 84,09%, специфичность 86,93%, точность (precision) 86,93%, полнота 84,53%, F1-score 85,71%.

Алгоритмы диагностики MODY

С помощью метода логистической регрессии разработан калькулятор риска MODY, который размещен в открытом доступе на сайте www.diabetesgenes.org [52]. Он создан на основе данных, полученных при сравнении группы больных СД1 (N=278), СД2 (N=319) и пациентов с генетически подтвержденным MODY (N=594, 243 — с мутацией GCK, 296 — HNF1A и 55 — HNF4A). В исследовании участвовали все пациенты с дебютом СД в возрасте до 35 лет включительно, в том числе дети. Проанализированы данные, которые не составляет труда определить у любого пациента: пол, возраст в дебюте заболевания и на момент исследования, ИМТ, лечение в дебюте и в настоящий момент, время, прошедшее до начала инсулинотерапии, уровень HbA_{1c} и наличие СД у родителей. AUROC в отношении прогноза MODY составила более 0,94, что свидетельствует о хорошей дискриминационной способности калькулятора, чувствительность — 91% (vs. 72% у стандартных клинических критериев), а специфичность 94% (vs. 91%). Однако этот метод валидизирован только для лиц с дебютом СД в возрасте ≤35 лет. В настоящее время калькулятор дополнен полями для указания расовой принадлежности и наличия сопутствующих заболеваний (кисты почек, глухота, парциальная липодистрофия, тяжелая инсулинорезистентность в отсутствие ожирения, тяжелое синдромальное ожирение). До проведения МГИ создатели калькулятора рекомендуют измерение уровней С-пептида и специфических антител, так как с помощью

этих показателей можно с большой долей вероятности исключить или подтвердить MODY [53]. Прогностическая ценность калькулятора была подтверждена на португальской и мультиэтнической бразильской популяции, хотя при ретроспективном анализе пациентов с MODY, проведенном в Польше, не выявлено преимуществ применения калькулятора для отбора пациентов, подлежащих МГИ, по сравнению с рутинным подходом [54–56].

В исследовании R. Mulligan и соавт. на 330 пациентах (70 из них — с подтвержденным MODY) сравнивались несколько моделей: традиционная логистическая регрессия, алгоритмы ИИ (НС, дерево классификаций, кластеризация) [57]. При валидации алгоритмы показали максимальную точность выявления MODY, равную 75–80%, причем логистическая регрессия и НС показали одинаковую эффективность, а точность рутинного клинического подхода на той же выборке составила 42%.

МГИ остается решающим методом диагностики MODY. Традиционно патогенность генетических вариантов оценивается с помощью критериев American College of Medical Genetics (ACMG) [58], которые разработаны и унифицированы для любых генов, однако зачастую врач-генетик не может дать однозначное заключение о патогенности конкретной мутации, даже используя доступные современные методы компьютерного моделирования. Группа исследователей предложила дифференцированный подход к оценке *HNF1A*, *HNF4A*, *GCK* [59]. Ученые предположили, что критерии ACMG имеют разный вес в принятии решения о патогенности генетического варианта. Для оценки значимости критериев разработан алгоритм МО, использовавший в качестве входных данных 15 из 33 критериев ACMG. Данные были собраны из базы ClinVar и Human Gene Mutation Database: 899 уникальных генетических вариантов для *HNF1A*, 1037 — для *HNF4A*, 1664 — для *GCK*. Исследователи включили в анализ только независимые критерии, в то время как зависимые, фактор инфляции регрессии которых составил более 10, а коэффициент корреляции — более 0,8, были исключены. Предположение исследователей было подтверждено, и алгоритм показал высокую прогностическую ценность, точность составила 95%.

Алгоритмы мультиклассовой классификации СД

И наконец, в исследовании 2020 г. была предпринята попытка мультиклассовой классификации пациентов с СД на группы: классический СД1, СД2, LADA, MODY, ГСД [60]. Исследователи сравнивали эффективность алгоритмов «дерева решений», «случайного леса», *k*-ближайших соседей, логистической регрессии и многослойного персептрона (метод ГО). Для обучения использовались данные о 3378 пациентах из базы TablHealth Project. В качестве входных параметров были выбраны возраст, суточная калорийность рациона, глюкоза плазмы натощак, систолическое и диастолическое артериальное давление, ИМТ, HbA_{1c} , наличие нарушенной толерантности к глюкозе, метаболического синдрома, макросомии плода, микроальбуминурии, ишемической болезни сердца, артериальной гипертензии и церебральной васкулопатии. Наибольшую эффективность показала НС многослойный персептрон. Остальные алгоритмы тоже имели высокую точность, однако ввиду разной представленности классов в выборке их эффективность в выявлении

более редких типов диабета оказалась существенно ниже. К примеру, метод «случайного леса» имел показатель полноты 0,98–1,0 для всех типов СД, кроме LADA — 0,64. Для MODY наилучший показатель F1-score получен для методов «случайного леса» и НС — 0,98 при точности 0,96, полноте 1,0.

Исследователи не описывают критерии постановки того или иного диагноза и методы верификации LADA и MODY, что является существенным недостатком данного исследования. Кроме того, в группу исследователей не входили практикующие врачи — лишь математики и информатики. Диагнозы, установленные пациентам из базы данных, принимались как заведомо истинные, что могло привести к повторению алгоритмом тех же ошибок, которые совершаются клиницистами в рутинной практике. К тому же значительная часть использованных признаков представляет собой поздние диабетические осложнения, поэтому данный алгоритм невозможно применить для ранней диагностики.

ЗАКЛЮЧЕНИЕ

Гетерогенность сахарного диабета является причиной ошибок в диагностике и, следовательно, лечении СД. При нетипичном течении диабета клиницист должен оценить личный и семейный анамнез пациента, антропометрические показатели, характер течения заболевания, эффективность лечения, и на основании этих данных он эмпирически может заподозрить одну из сложно классифицируемых форм СД и назначить дополнительное обследование. Однако отдельные клинические признаки недостаточно чувствительны и специфичны. Многие исследования показывают неэффективность рутинного подхода, в связи с чем много случаев LADA и MODY остаются недиагностированными. Это влечет за собой тяжелые последствия в виде ухудшения гликемического контроля, быстрого развития осложнений. В случае MODY 2, напротив, проводятся ненужные лечебные мероприятия, что ухудшает качество жизни пациента. Для выявления пропущенных случаев возможно искусственное расширение круга пациентов, подлежащих дообследованию, но эта стратегия неоправдана, так как повлечет за собой избыточную нагрузку на систему здравоохранения.

В связи с этим перспективным является применение различных формализованных подходов к дифференциальной диагностике СД1 и СД2 и к оценке степени риска LADA и MODY лишь по анамнестическим и клиническим характеристикам пациента. Эти подходы эволюционировали от простых блок-схем до калькуляторов и шкал, разработанных на основе классических статистических методов (калькулятор риска MODY, шкала «Клинического риска LADA» и др.). Хотя в клинических исследованиях эти относительные методы показали высокую эффективность, остается большое пространство для дальнейшего улучшения инструментов путем оптимизации входных параметров, поиска неочевидных дифференциально-диагностических признаков и определения их «веса» в принятии решения. Методы МО являются наиболее подходящими для этой задачи. Хотя лучшую эффективность показывают НС ГО, методы классического МО не следует игнорировать, так как они просты в реализации и достаточно эффективны.

Опыт исследовательских групп показывает, насколько важно выделение статистически значимых различий между разными типами СД с целью выбора входных параметров для алгоритма и формирование объемного, полноценного набора данных, максимально соответствующего требованиям о полноте, сбалансированности, репрезентативности и др., поэтому работа по формированию диагностического алгоритма на основе методов ИИ должна выполняться в тесном сотрудничестве с IT-специалистами. Внедрение алгоритмов МО в широкую врачебную практику в качестве СППВР оправдано, эффективно и отвечает современным тенденциям. К сожалению, предложенные зарубежными специалистами модели в основном недоступны, так как не находятся в открытом доступе. В любом случае преимущества каждого алгоритма МО должны быть доказаны в клинических исследованиях.

Проведение дальнейших исследований, направленных на создание инструментов, максимально точных и удобных для использования практикующим врачом и, возможно, работающих в фоновом режиме, является перспективным направлением в медицине, так как позволит не только сократить расходы на излишние исследова-

ния, но и повысит эффективность лечения, снизит риск развития осложнений СД, улучшит качество жизни и прогноз пациентов.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Источники финансирования. Грант Министерства науки и высшего образования Российской Федерации, соглашение 075-15-2022-310 от 20.04.2022 г.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с содержанием настоящей статьи.

Участие авторов. Русяева Н.В. — анализ литературы, сбор материала, написание статьи; Голодников И.И. — анализ литературы, сбор материала, написание статьи; Никонова Т.В. — внесение правок в текст, вычитка рукописи; Кононенко И.В. — внесение правок в текст, вычитка рукописи; Шестакова М.В. — итоговое редактирование рукописи, внесение важных правок в текст статьи. Все авторы одобрили финальную версию статьи перед публикацией, выразили согласие нести ответственность за все аспекты работы, подразумевающую надлежащее изучение и решение вопросов, связанных с точностью или добросовестностью любой части работы.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

1. Котова Е.Г., Кобякова О.С., Стародубов В.И., и др. *Заболееваемость всего населения России в 2021 году с диагнозом, установленным впервые в жизни: статистические материалы*. М.: ЦНИИОИЗ Минздрава России; 2022. [Kotova EG, Kobiakova OS, Starodubov VI, et al. *Zabolevaemost' vsego naseleniia Rossii v 2021 godu s diagnozom, ustanovlennym v pervye v zhizni: statisticheskie materialy*. Moscow: TsNIOIz Minzdrava Rossii; 2022. (In Russ.)].
2. Дедов И.И., Шестакова М.В., Майоров А.Ю., и др. Алгоритмы специализированной медицинской помощи больным сахарным диабетом (10-й выпуск) // *Сахарный диабет*. — 2021. — Т. 24. — №51. — С. 1-235. [Dedov II, Shestakova MV, Mayorov AYU, et al. Standards of specialized diabetes care (10-th edition). *Diabetes Mellitus*. 2021;24(51):1-235 (In Russ.)]. doi: <https://doi.org/10.14341/DM12802>
3. World Health Organization. *Classification of diabetes mellitus*. 2019 [cited 13.02.2022]. Available from: <https://apps.who.int/iris/handle/10665/325182>
4. SpringerLink [Internet]. *Painting a new picture of personalised medicine for diabetes* [cited 13.02.2022]. Available from: <https://link.springer.com/article/10.1007/s00125-017-4210-x>
5. Schwartz SS, Epstein S, Corkey BE, et al. The time is right for a new classification system for diabetes: Rationale and implications of the β -cell-centric classification schema. *Diabetes Care*. 2016;39(2):179-186. doi: <https://doi.org/10.2337/dc15-1585>
6. Yin W, Luo S, Xiao Z, et al. Latent autoimmune diabetes in adults: a focus on β -cell protection and therapy. *Front Endocrinol (Lausanne)*. 2022;(13). doi: <https://doi.org/10.3389/fendo.2022.959011>
7. Zaharia OP, Bobrov P, Strassburger K, et al. Metabolic characteristics of recently diagnosed adult-onset autoimmune diabetes mellitus. *J Clin Endocrinol Metab*. 2018;103(2):429-437. doi: <https://doi.org/10.1210/je.2017-01706>
8. Horton V, Stratton I, Bottazzo GF, et al. Genetic heterogeneity of autoimmune diabetes: age of presentation in adults is influenced by HLA DRB1 and DQB1 genotypes (UKPDS 43). *Diabetologia*. 1999;42(5):608-616. doi: <https://doi.org/10.1007/s001250051202>
9. Battaglia M, Ahmed S, Anderson MS, et al. Introducing the endotype concept to address the challenge of disease heterogeneity in type 1 diabetes. *Diabetes Care*. 2020;43(1):5-12. doi: <https://doi.org/10.2337/dc19-0880>
10. Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*. 2018;6(5):361-369. doi: [https://doi.org/10.1016/S2213-8587\(18\)30051-2](https://doi.org/10.1016/S2213-8587(18)30051-2)
11. Xiong XF, Yang Y, Wei L, et al. Identification of two novel subgroups in patients with diabetes mellitus and their association with clinical outcomes: A two-step cluster analysis. *J Diabetes Invest*. 2021;12(8):1346-1358. doi: <https://doi.org/10.1111/jdi.13494>
12. Seidu S, Davies MJ, Mostafa S, et al. Prevalence and characteristics in coding, classification and diagnosis of diabetes in primary care. *Postgrad Med J*. 2014;90(1059):13-17. doi: <https://doi.org/10.1136/postgradmedj-2013-132068>
13. De Lusignan S, Khunti K, Belsey J, et al. A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. *Diabetic Medicine*. 2010;27(2):203-209. doi: <https://doi.org/10.1111/j.1464-5491.2009.02917.x>
14. Смирнова О.М., Кононенко И.В., Дедов И.И. Гетерогенность сахарного диабета. Аутоиммунный латентный сахарный диабет у взрослых (LADA): определение, распространенность, клинические особенности, диагностика, принципы лечения // *Сахарный диабет*. — 2008. — Т. 11. — №4. — С. 18-23. [Smirnova OM, Kononenko IV, Dedov II. Geterogennost' sakharnogo diabeta. Autoimmunnyy latentnyy sakharnyy diabetes u vzroslykh (LADA): opredelenie, rasprostranennost', klinicheskie osobennosti, diagnostika, printsipy lecheniya. *Diabetes mellitus*. 2008;11(4):18-23. (In Russ.)]. doi: <https://doi.org/10.14341/2072-0351-5583>
15. Cheheltani R, King N, Lee S, et al. Predicting misdiagnosed adult-onset type 1 diabetes using machine learning. *Diabetes Res Clin Pract*. 2022;(191):110029. doi: <https://doi.org/10.1016/j.diabres.2022.110029>
16. Muñoz C, Floreen A, Garey C, et al. Misdiagnosis and diabetic ketoacidosis at diagnosis of type 1 diabetes: Patient and caregiver perspectives. *Clin Diabetes*. 2019;37(3):276-281. doi: <https://doi.org/10.2337/cd18-0088>
17. Buzzetti R, Zampetti S, Maddaloni E. Adult-onset autoimmune diabetes: current knowledge and implications for management. *Nat Rev Endocrinol*. 2017;13(11):674-686. doi: <https://doi.org/10.1038/nrendo.2017.99>
18. Thomas NJ, Lynam AL, Hill AV, et al. Type 1 diabetes defined by severe insulin deficiency occurs after 30 years of age and is commonly treated as type 2 diabetes. *Diabetologia*. 2019;62(7):1167-1172. doi: <https://doi.org/10.1007/s00125-019-4863-8>
19. Leslie RD, Evans-Molina C, Freund-Brown J, et al. Adult-onset type 1 diabetes: Current understanding and challenges. *Diabetes Care*. 2021;44(11):2449-2456. doi: <https://doi.org/10.2337/dc21-0770>
20. Carlsson S, Midthjell K, Tesfamarian MY, Grill V. Age, overweight and physical inactivity increase the risk of latent autoimmune diabetes in adults: results from the Nord-Trøndelag health study. *Diabetologia*. 2007;50(1):55-58. doi: <https://doi.org/10.1007/s00125-006-0518-7>

21. Buzzetti R, Tuomi T, Mauricio D, et al. Management of latent autoimmune diabetes in adults: A consensus statement from an international expert panel. *Diabetes*. 2020;69(10):2037-2047. doi: <https://doi.org/10.2337/dbi20-0017>
22. Hawa MI, Kolb H, Schloot N, et al. Adult-onset autoimmune diabetes in Europe is prevalent with a broad clinical phenotype: Action LADA 7. *Diabetes Care*. 2013;36(4):908-913. doi: <https://doi.org/10.2337/dc12-0931>
23. Sørgerd EP, Skorpén F, Kvaløy K, et al. Time dynamics of autoantibodies are coupled to phenotypes and add to the heterogeneity of autoimmune diabetes in adults: the HUNT study, Norway. *Diabetologia*. 2012;55(5):1310-1318. doi: <https://doi.org/10.1007/s00125-012-2463-y>
24. Bell DSH, Ovalle F. The role of C-peptide levels in screening for latent autoimmune diabetes in adults. *Am J Ther*. 2004;11(4):308-311. doi: <https://doi.org/10.1097/01.mjt.0000102372.28723.2b>
25. Leighton E, Sainsbury CA, Jones GC. A practical review of C-peptide testing in diabetes. *Diabetes Ther*. 2017;8(3):475-487. doi: <https://doi.org/10.1007/s13300-017-0265-4>
26. Hernandez M, Mollo A, Marsal JR, et al. Insulin secretion in patients with latent autoimmune diabetes (LADA): half way between type 1 and type 2 diabetes: action LADA 9. *BMC Endocr Disord*. 2015;15(1):1. doi: <https://doi.org/10.1186/1472-6823-15-1>
27. Nkonge KM, Nkonge DK, Nkonge TN. The epidemiology, molecular pathogenesis, diagnosis, and treatment of maturity-onset diabetes of the young (MODY). *Clin Diabetes Endocrinol*. 2020;6(1):20. doi: <https://doi.org/10.1186/s40842-020-00112-5>
28. Shields BM, Hicks S, Shepherd MH, et al. Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia*. 2010;53(12):2504-2508. doi: <https://doi.org/10.1007/s00125-010-1799-4>
29. Pihoker C, Gilliam LK, Ellard S, et al. Prevalence, characteristics and clinical diagnosis of maturity onset diabetes of the young due to mutations in HNF1A, HNF4A, and glucokinase: Results from the SEARCH for diabetes in youth. *J Clin Endocrinol Metab*. 2013;98(10):4055-4062. doi: <https://doi.org/10.1210/jc.2013-1279>
30. Carroll RW, Murphy R. Monogenic diabetes: A diagnostic algorithm for clinicians. *Genes (Basel)*. 2013;4(4):522-535. doi: <https://doi.org/10.3390/genes4040522>
31. Schober E, Rami B, Grabert M, et al. Phenotypical aspects of maturity-onset diabetes of the young (MODY diabetes) in comparison with Type 2 diabetes mellitus (T2DM) in children and adolescents: experience from a large multicentre database. *Diabet Med*. 2009;26(5):466-473. doi: <https://doi.org/10.1111/j.1464-5491.2009.02720.x>
32. Urakami T. Maturity-onset diabetes of the young (MODY): current perspectives on diagnosis and treatment. *Diabetes, Metab Syndr Obes Targets Ther*. 2019;12(1):1047-1056. doi: <https://doi.org/10.2147/DMSO.S179793>
33. Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nat Clin Pract Endocrinol Metab*. 2008;4(4):200-213. doi: <https://doi.org/10.1038/ncpendmet0778>
34. Зубкова Н.А., Арбатская Н.Ю., Петрайкина Е.Е., и др. Сахарный диабет типа MODY3: клиническая и молекулярно-генетическая характеристика 9 случаев заболевания // *Проблемы Эндокринологии*. — 2014. — Т. 60. — №1. — С. 51-56. [Zubkova NA, Arbatskaya NYu, Petraykina EE, et al. Type 3 form of MODY: the clinical and molecular-genetic characteristic. Nine cases of the disease. *Problems of Endocrinology*. 2014;60(1):51-56. (In Russ.)]. doi: <https://doi.org/10.14341/probl201460151-56>
35. Oliveira SC, Neves JS, Pérez A, Carvalho D. Maturity-onset diabetes of the young: From a molecular basis perspective toward the clinical phenotype and proper management. *Endocrinol Diabetes Nutr*. 2020;67(2):137-147. doi: <https://doi.org/10.1016/j.endinu.2019.07.012>
36. Delvecchio M, Pastore C, Giordano P. Treatment options for MODY patients: A systematic review of literature. *Diabetes Ther*. 2020;11(8):1667-1685. doi: <https://doi.org/10.1007/s13300-020-00864-4>
37. Li L-M, Jiang B-G, Sun L-L. HNF1A: from monogenic diabetes to type 2 diabetes and gestational diabetes mellitus. *Front Endocrinol (Lausanne)*. 2022;13(1):1047-1056. doi: <https://doi.org/10.3389/fendo.2022.829565>
38. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317-1318. doi: <https://doi.org/10.1001/jama.2017.18391>
39. Kavakiotis I, Tsave O, Salifoglou A, et al. Machine learning and data mining methods in diabetes research. *Comput Struct Biotechnol J*. 2017;15:104-116. doi: <https://doi.org/10.1016/j.csbj.2016.12.005>
40. Makroum MA, Adda M, Bouzouane A, Ibrahim H. Machine learning and smart devices for diabetes management: Systematic review. *Sensors (Basel)*. 2022;22(5):1843. doi: <https://doi.org/10.3390/s22051843>
41. Климонтон В.В., Бериков В.Б., Сайк О.В. Искусственный интеллект в диабетологии // *Сахарный диабет*. — 2021. — Т. 24. — №2. — С. 156-166. [Klimontov V.V., Berikov V.B., Saik O.V. Artificial intelligence in diabetology. *Diabetes mellitus*. 2021;24(2):156-166. (In Russ.)]. doi: <https://doi.org/10.14341/DM12665>
42. Lynam AL, Dennis JM, Owen KR, et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagnostic Progn Res*. 2020;4(1):6. doi: <https://doi.org/10.1186/s41512-020-00075-2>
43. Ijcsis J of CS [Internet]. Diabetes type1 and type2 classification using machine learning technique [cited 13.02.2023]. Available from: https://www.academia.edu/33764602/Diabetes_Type1_and_Type2_Classification_Using_Machine_Learning_Technique
44. Romantschuk P. How to distinguish between type 1 and type 2 diabetes at diagnosis in ≥16-year-old patients. 2021 [cited 13.02.2023]. Available from: <https://helda.helsinki.fi/handle/10138/341686>
45. Wang Y, Liu S, Chen R, et al. A novel classification indicator of type 1 and type 2 diabetes in China. *Sci Rep*. 2017;7(1):17420. doi: <https://doi.org/10.1038/s41598-017-17433-8>
46. Oram RA, Patel K, Hill A, et al. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care*. 2016;39(3):337-344. doi: <https://doi.org/10.2337/dc15-1111>
47. Nooney JG, Kirkman MS, Bullard KM, et al. Identifying optimal survey-based algorithms to distinguish diabetes type among adults with diabetes. *J Clin Transl Endocrinol*. 2020;21:100231. doi: <https://doi.org/10.1016/j.jcte.2020.100231>
48. Lo-Ciganic W, Zgibor JC, Ruppert K, et al. Identifying type 1 and type 2 diabetic cases using administrative data: A tree-structured model. *J Diabetes Sci Technol*. 2011;5(3):486-493. doi: <https://doi.org/10.1177/193229681100500303>
49. Fournalos S, Perry C, Stein MS, et al. A clinical screening tool identifies autoimmune diabetes in adults. *Diabetes Care*. 2006;29(5):970-975. doi: <https://doi.org/10.2337/diacare.295970>
50. Wang Z, Zhang J, Xu H, et al. Development and validation of a prevalence model for latent autoimmune diabetes in adults (lada) among patients first diagnosed with Type 2 Diabetes Mellitus (T2DM). *Med Sci Monit*. 2021;27:e932725. doi: <https://doi.org/10.12659/MSM.932725>
51. Miller A, Panneerselvam J, Liu L. An empirical analysis of LADA diabetes case, control and variable importance. In: *Proceedings of the 14th IEEE/ACM International Conference on Utility and Cloud Computing Companion*. New York, NY, USA: ACM; 2021. Vol 5. 1-8. doi: <https://doi.org/10.1145/3492323.3495632>
52. Shields BM, McDonald TJ, Ellard S, et al. The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes. *Diabetologia*. 2012;55(5):1265-1272. doi: <https://doi.org/10.1007/s00125-011-2418-8>
53. McDonald TJ, Colclough K, Brown R, et al. Islet autoantibodies can discriminate maturity-onset diabetes of the young (MODY) from type 1 diabetes. *Diabetic Medicine*. 2011;28(9):1028-1033. doi: <https://doi.org/10.1111/j.1464-5491.2011.03287.x>
54. da Silva Santos T, Fonseca L, Santos Monteiro S, et al. MODY probability calculator utility in individuals' selection for genetic testing: Its accuracy and performance. *Endocrinol Diabetes Metab*. 2022;5(5):e00332. doi: <https://doi.org/10.1002/edm.2.332>
55. Tarantino RM, Abreu G de M, Fonseca ACP de, et al. MODY probability calculator for GCK and HNF1A screening in a multiethnic background population. *Arch Endocrinol Metab*. 2020;64(1):17-23. doi: <https://doi.org/10.20945/2359-399700000173>
56. Hohendorff J, Zapala B, Ludwig-Słomczynska AH, et al. The utility of MODY probability calculator in probands of families with early-onset autosomal dominant diabetes from Poland. *Minerva Med*. 2019;110(6):499-506. doi: <https://doi.org/10.23736/S0026-4806.19.06053-1>

57. Mulligan RP, Letourneau-Freiberg LR, Bowden TL, et al. 1131-P: Clinical prediction models for MODY genetic testing in the united states monogenic diabetes registry. *Diabetes*. 2021;70(S1):1-8. doi: <https://doi.org/10.2337/db21-1131-P>
58. Bahcall OG. ACMG guides on the interpretation of sequence variants. *Nat Rev Genet*. 2015;16(5):257-257. doi: <https://doi.org/10.1038/nrg3940>
59. Liu Y, Qu H-Q, Wenocur AS, et al. Interpretation of maturity-onset diabetes of the young genetic variants based on American College of Medical Genetics and Genomics Criteria: Machine-learning model development. *JMIR Biomed Eng*. 2020;5(1):e20506. doi: <https://doi.org/10.2196/20506>
60. Mainenti G, Campanile L, Marulli F, et al. Machine learning approaches for diabetes classification: Perspectives to artificial intelligence methods updating. In: *Proceedings of the 5th International Conference on Internet of Things, Big Data and Security*. New York, NY, USA: SCITEPRESS — Science and Technology Publications; 2020. Vol 5. P. 533-540. doi: <https://doi.org/10.5220/0009839405330540>

ИНФОРМАЦИЯ ОБ АВТОРАХ [AUTHORS INFO]

***Русяева Надежда Владимировна**, аспирант [**Nadezhda V. Rusyaeva**, MD, PhD student]; адрес: Россия, 117036 Москва, ул. Дм. Ульянова, д. 11 [address: 11 Dm. Ulyanova street, 117036 Moscow, Russian Federation]; ORCID: <https://orcid.org/0000-0003-3628-2102>; Researcher ID: AAY-6365-2021; Scopus Author ID: 57220024968; eLibrary SPIN: 7256-6090; e-mail: nadshul@gmail.com

Голодников Иван Иванович, аспирант [Ivan I. Golodnikov, MD, PhD student]; ORCID: <https://orcid.org/0000-0003-0935-9004>; Researcher ID: AAJ-8843-2021; Scopus Author ID: 57208628509; eLibrary SPIN: 3213-0916; e-mail: golodnikov.ivan@endocrincentr.ru

Никонова Татьяна Васильевна, д.м.н. [Tatiana V. Nikonova, MD, PhD]; ORCID: <https://orcid.org/0000-0001-5656-2596>; eLibrary SPIN: 8863-0201; e-mail: tatiana_nikonova@mail.ru

Конonenко Ирина Владимировна, к.м.н., доцент [Irina V. Kononenko, MD, PhD, associate professor]; ORCID: <https://orcid.org/0000-0003-4929-1526>; Researcher ID: H-5947-2016; Scopus Author ID: 35744972400; eLibrary SPIN: 6528-7986; e-mail: shakhtarina@bk.ru

Шестакова Марина Владимировна, д.м.н., профессор, академик РАН [Marina V. Shestakova, MD, PhD, Professor]; ORCID: <https://orcid.org/0000-0002-5057-127X>; Researcher ID: D-9123-2012; Scopus Author ID: 7004195530; eLibrary SPIN: 7584-7015; e-mail: shestakova.mv@gmail.com

ЦИТИРОВАТЬ:

Русяева Н.В., Голодников И.И., Конonenко И.В., Никонова Т.В., Шестакова М.В. Методы машинного обучения в дифференциальной диагностике сложно классифицируемых типов сахарного диабета // *Сахарный диабет*. — 2023. — Т. 26. — №5. — С. 473-483. doi: <https://doi.org/10.14341/DM13070>

TO CITE THIS ARTICLE:

Rusyaeva NV, Golodnikov II, Kononenko IV, Nikonova TV, Shestakova MV. Machine learning methods in the differential diagnosis of difficult-to-classify types of diabetes mellitus. *Diabetes Mellitus*. 2023;26(5):473-483. doi: <https://doi.org/10.14341/DM13070>