

Методический подход и рекомендации по научному описанию создания и валидации модели машинного обучения

© А.В. ГУСЕВ^{1,2}, А.В. ВЛАДЗИМИРСКИЙ^{3,4}, Г.Г. ГАВРИЛЕНКО¹

¹ООО «К-Скай», Петрозаводск, Россия;

²ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, Москва, Россия;

³ГБУЗ города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы», Москва, Россия;

⁴ФГАОУ ВО «Первый Московский государственный медицинский университет им. И.М. Сеченова» Минздрава России (Сеченовский Университет), Москва, Россия

РЕЗЮМЕ

Цель исследования. Стандартизировать требования к описанию результатов создания и валидации моделей машинного обучения для обеспечения прозрачности и доверия со стороны практического здравоохранения и надзорных органов.

Материал и методы. Выполнено аналитическое исследование данных на основе методологии [Cross-Industry Standard Process for Data Mining](#) (реализованы четыре первые фазы цикла исследования; фазы оценки и внедрения должны быть выполнены в формате внешней независимой валидации). Использованы аналитические методы научного познания: анализ, синтез, индукция.

Результаты. Разработаны практико-методические рекомендации по подготовке отчета (рукописи научной статьи) о разработке и валидации модели машинного обучения. Рекомендации оформлены в виде: а) методического документа; б) чек-листа для ускоренной проверки структуры, полноты и качества содержания рукописи. Принципиальные отличия авторской разработки от аналогов: наличие стандартной структуры медицинской научной статьи; баланс представления клинической и технологической информации; включение специфических аспектов машинного обучения, наиболее релевантных с медицинской точки зрения; методическая информация для поддержки авторов; обеспечение возможности применения чек-листа разными заинтересованными специалистами.

Заключение. Оригинальный методический подход и чек-лист могут повысить качество, прозрачность и воспроизводимость научных отчетов в сфере машинного обучения для здравоохранения. Разработка отличается балансом технологических и клинических аспектов, образовательной значимостью и комплексными возможностями. Предложенный инструментарий может быть валидирован независимыми исследователями.

Ключевые слова: цифровое здравоохранение, искусственный интеллект, машинное обучение, большие данные, системы поддержки принятия решений, телемедицина.

ИНФОРМАЦИЯ ОБ АВТОРАХ:

Гусев А.В. — <https://orcid.org/0000-0002-7380-8460>

Владзимирский А.В. — <https://orcid.org/0000-0002-2990-7736>

Гавриленко Г.Г. — <https://orcid.org/0000-0001-9024-2756>

Автор, ответственный за переписку: Гусев А.В. — e-mail: agusev@webiomed.ai

КАК ЦИТИРОВАТЬ:

Гусев А.В., Владзимирский А.В., Гавриленко Г.Г. Методический подход и рекомендации по научному описанию создания и валидации модели машинного обучения. *Медицинские технологии. Оценка и выбор.* 2022;44(3):12–30. <https://doi.org/10.17116/medtech20224403112>

Methodical approach and recommendations for scientific description of creation and validation of machine learning model

© A.V. GUSEV^{1,2}, A.V. VLADZIMIRSKIY^{3,4}, G.G. GAVRILENKO¹

¹K-SkAI LLC, Petrozavodsk, Russia;

²Central Research Institute of Organization and Informatization in Healthcare, Moscow, Russia;

³Scientific and Practical Clinical Center for Diagnosis and Telemedicine Technologies, Moscow, Russia;

⁴Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia

ABSTRACT

Objective. To standardize the requirements for describing creation and validation of machine learning models ensuring transparency and confidence in healthcare and regulatory authorities.

Material and methods. Analytical study of data was based on the Cross-Industry Standard Process for Data Mining methodology (the first four phases of the research cycle were implemented; evaluation and implementation phases should be performed as external independent validation). Analytical methods of scientific knowledge including analysis, synthesis and induction were used.

Results. Practical and methodical guidelines on description of development and validation of machine learning models have been developed. These guidelines are made in the form of methodical document and checklist for accelerated verification of structure, completeness and quality of content. The principal differences between the author's approach and analogues are standard structure of medical scientific article; balanced presentation of clinical and technological information; inclusion of specific aspects of machine learning that are most relevant from a medical point of view; methodical information to support the authors; ensuring the possibility of applying the checklist by various interested specialists.

Conclusion. An original methodical approach and checklist can improve the quality, transparency and reproducibility of scientific reports in machine learning for healthcare. The is distinguished by balanced technological and clinical aspects, educational significance and comprehensive capabilities. Independent researchers can validate the proposed approach.

Keywords: digital healthcare, artificial intelligence, machine learning, big data, decision making support systems, telemedicine.

INFORMATION ABOUT THE AUTHORS:

Gusev A.V. — <https://orcid.org/0000-0002-7380-8460>

Vladimirskiy A.V. — <https://orcid.org/0000-0002-2990-7736>

Gavrilenko G.G. — <https://orcid.org/0000-0001-9024-2756>

Corresponding author: Gusev A.V. — e-mail: agusev@webiomed.ai

TO CITE THIS ARTICLE:

Gusev AV, Vladimirskiy AV, Gavrilenko GG. Methodical approach and recommendations for scientific description of development and validation of machine learning model. *Medical Technologies. Assessment and Choice*. 2022;44(3):12–30. (In Russ.). <https://doi.org/10.17116/medtech20224403112>

Введение

В настоящее время ситуация в мировом здравоохранении характеризуется постоянно возрастающей нагрузкой и одновременно нехваткой ключевых ресурсов, включая финансовое и кадровое обеспечение [1]. Основной причиной этого является бремя хронических неинфекционных и социально-значимых инфекционных заболеваний, демографические тренды, экономические кризисы [2]. Еще более усугубила проблему пандемия COVID-19 в 2020—2022 гг. Необходимость выявления и сокращения неэффективных затрат, освобождение врача от нагрузки, не связанной с непосредственным оказанием медицинской помощи, в том числе от обработки большого объема медицинских данных, автоматический анализ информации с целью минимизации негативного влияния человеческого фактора относятся к самым актуальным задачам современного здравоохранения [3]. Их можно эффективно решать с помощью технологий искусственного интеллекта (ИИ), которые за последние годы совершили качественный переход от научных исследований в практическую плоскость. Множество компаний по всему миру разрабатывают самые разнообразные продукты на основе технологий ИИ [4]. Рост инвестиций и постепенное проникновение соответствующих продуктов в практическое здравоохранение обострили проблему доверия и ответственного отношения [5, 6].

В этой связи критически важной задачей, стоящей перед разработчиками и лицами, принимающими решения в сфере внедрения систем ИИ в практическое здравоохранение, является разработка и публикация подробных отчетов, описывающих процесс сбора и подготовки наборов данных, тестирования

и валидации моделей машинного обучения (ММО). Такие отчеты должны быть своеобразным мостом между разработчиками и медицинскими работниками, формирующим доверие к системам ИИ, а потому должны отвечать принципами доказательной медицины. Проблема качества указанных публикаций нашла свое отражение в появлении особого вида методических рекомендаций — чек-листов для быстрой проверки наличия в рукописи основных требуемых компонентов (MI-CLAIM, CONSORT-AI, SPIRIT-AI и т.д.) [7, 8]. Несовершенство подобных инструментов подтверждается достаточно редким их использованием. Принцип методики чек-листов, несомненно, является прогрессивным и эффективным; однако надежных и значимых инструментов в аспекте описания создания ММО нет.

Цель исследования — стандартизировать требования к описанию результатов создания и валидации ММО для обеспечения прозрачности и доверия со стороны практического здравоохранения и надзорных органов.

Материал и методы

Выполнено аналитическое исследование данных на основе методологии Cross-Industry Standard Process for Data Mining (CRISP-DM) [9], общий ход которого представлен на **рис. 1**.

В рамках данного исследования нами реализованы четыре первые фазы цикла исследования по методологии CRISP-DM: понимание целей; начальное изучение данных; подготовка данных; моделирование. Предполагается, что фазы оценки и внедрения будут выполнены независимыми коллективами исследователей для обеспечения объективности

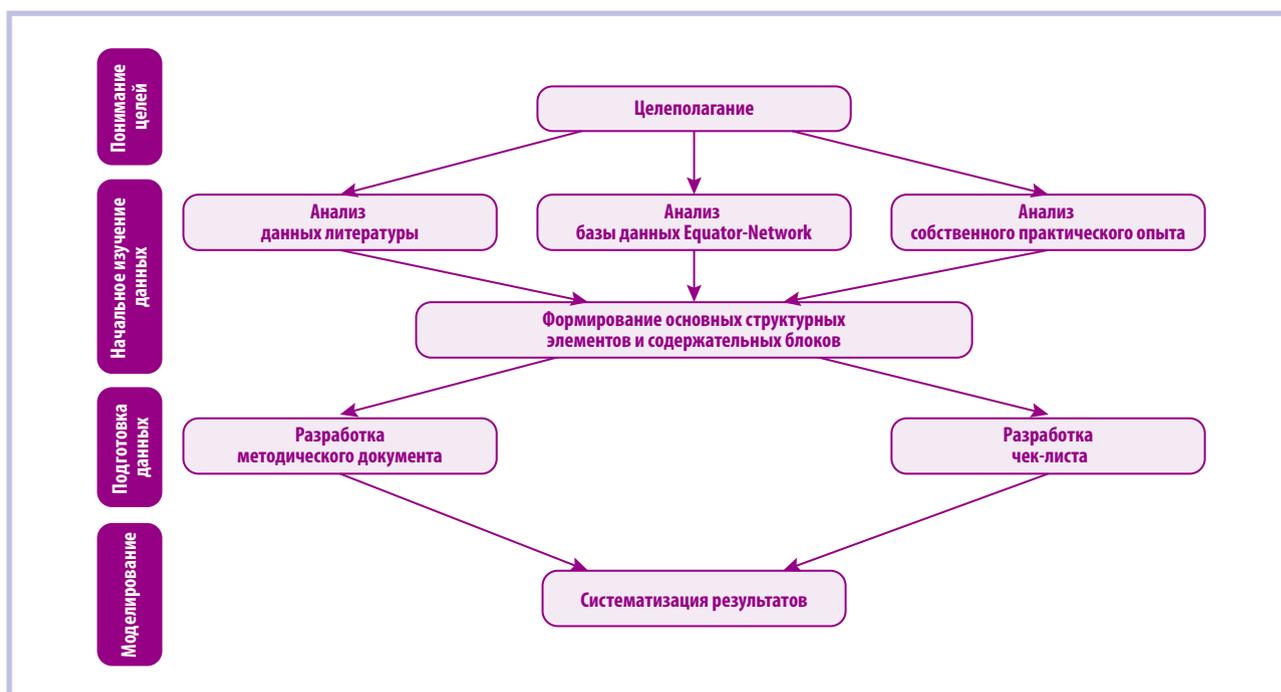


Рис. 1. Диаграмма общего хода исследования.

Fig. 1. Study flowchart.

оценок качества наших разработок. Для формирования стандартизированных требований проведена этапная обработка данных, полученных из трех источников:

- научной периодики (библиографические базы Российского индекса научного цитирования (РИНЦ, <https://www.elibrary.ru>), Национальной медицинской библиотеки США (PubMed, <https://pubmed.ncbi.nlm.nih.gov>);
- базы данных чек-листов и методологий Equator-Network (<https://www.equator-network.org>);
- собственного практического опыта работы в редакционных коллегиях рецензируемых научных журналов, а также в качестве ведущих исследователей и руководителей научных проектов в сфере технологий ИИ для здравоохранения.

Для иллюстрации отдельных положений использованы цитаты из оригинальных научных публикаций, ссылки на которые оформлены по общепринятым правилам и включены в пристатейный список литературы. В ряде случаев профессиональным переводчиком осуществлен перевод цитат на русский язык. Цитаты представлены в оригинальном виде, без каких-либо искажений или дополнений.

Использованы аналитические методы научного познания: анализ, синтез, индукция.

Результаты

В ходе исследования разработаны практико-методические рекомендации по подготовке отчета

(рукописи научной статьи) о разработке и валидации модели машинного обучения, предназначенной для исследований или применения в реальной клинической практике. Рекомендации оформлены в виде методического документа и чек-листа для ускоренной проверки структуры, полноты и качества содержания рукописи.

Методический документ

Общие требования

Отчет о создании модели машинного обучения должен состоять из нескольких частей (рис. 2). Далее рассмотрим каждую часть и рекомендации по ее оформлению подробнее. В заголовке подраздела в скобках будут даны наиболее распространенные в зарубежных рецензируемых изданиях обозначения структурных единиц публикации.

1. Информационная часть отчета

1.1. Название (Title)

Заглавие отчета должно полностью отражать содержание, тематику и кратко суммировать особенности разработки модели. Задача заглавия — раскрыть содержание работы. Желательно, чтобы заглавие содержало отдельные ключевые слова по тематике работы, это повышает его информативность и шансы включения в обзоры для последующего цитирования. Следует избегать как слишком общих, так и чрезмерно перегруженных детализацией заглавий. Объем заглавия — не более 10–12 слов.



Рис. 2. Общая структура отчета о создании и валидации прогнозной модели.

Fig. 2. Overall structure of report on development and validation of predictive model.

1.2. Аннотация (Abstract)

Выполняет функцию расширенного названия отчета и повествует о содержании работы. Цель аннотации — быстрое ознакомление читателя с содержанием основного текста. Аннотация должна быть написана в безличной форме, короткими предложениями, не содержать цитат, формул, аббревиатур, ссылок на литературные источники и т.д.

В аннотации должны быть **тезисно** изложены:

- проблематика работы (не более 2–3 предложений о решаемой с помощью ММО проблеме, существующих подходах и их ограничениях с акцентом на значимость данной работы);
- цель;
- источники данных для обучения и/или валидации модели;
- используемые методы;
- важнейшие результаты, включая метрики модели;
- выводы или заключение (с акцентом на актуальность и потенциальную практическую ценность).

Пример №1

История вопроса. Большая часть смертей от COVID-19 связана с тяжелым течением заболевания. Эффективное лечение тяжелых случаев остается проблемой из-за отсутствия ранней диагностики.

Цель. Исследование направлено на разработку эффективной модели прогнозирования тяжести COVID-19 путем объединения результатов лучевых исследований с клиническими биохимическими показателями.

Методы. Обследованы 46 пациентов с COVID-19 (10 — с тяжелым течением заболевания, 36 — с течением средней тяжести). Для построения прогностической модели сформирован набор данных из 27 медицинских документов пациентов с тяжелым течением болезни и 151 документа пациентов с течением средней тяжести. Собранные данные включают резуль-

таты лабораторных и лучевых (компьютерная томография — КТ) исследований. Нам удалось выделить характерные особенности изображений КТ пациентов с помощью сверточной нейронной сети. Обучение модели проводилось на комбинированных данных результатов КТ и лабораторных исследований.

Результаты. Мы представляем предиктивную модель, объединяющую результаты лучевых исследований пациентов с биохимическими показателями для выявления тяжелых случаев течения COVID-19. Метрики разработанной модели: AUROC 0,93, F1 0,89, что на 6% и 15% выше указанных показателей моделей, основанных только на данных лабораторных исследований. Кроме того, мы разработали статистическую модель для прогнозирования степени тяжести COVID-19 на основе результатов лабораторных исследований пациентов, полученных до того, как они были классифицированы как тяжелые случаи; эта модель достигла точности AUROC 0,81.

Выводы. Разработана модель, прогнозирующая клиническое прогрессирование COVID-19, а также оценивающая тяжесть заболевания на основе комбинированного анализа лабораторных тестов и результатов компьютерной томографии [10].

1.3. Ключевые слова (Keywords)

Основным критерием их выбора является ценность для трансляции содержания документа или для обеспечения его поиска в библиографических базах данных. Ключевые слова могут отражать основные понятия, термины, методы, концепции и т.д. Рекомендовано использовать базовую терминологию вместе с более сложной, допустимо использование синонимов. Могут быть представлены как отдельными словами, так и устойчивыми словосочетаниями. Рекомендованное количество — в пределах 10, при этом количество слов в составе ключевой фразы в пределах 3–4. При подборе ключевых слов на русском языке рекомендуется использовать список РИНЦ

(<https://www.elibrary.ru/keywords.asp>), на английском языке — справочник NLM MeSH (<https://www.nlm.nih.gov/mesh/meshhome.html>).

2. Основная (содержательная) часть

В основной части отчета авторы излагают цель, задачи, дизайн и ход, материалы и методы работы, приводят промежуточные и итоговые результаты, описывают процесс валидации (апробации) созданной ММО, также там в форме дискуссии обосновываются новизна и ценность полученных результатов. Материал данного раздела должен быть изложен в логической последовательности, повторяющей этапы работы, и содержать научную аргументацию со ссылками на статьи в рецензируемых научных журналах. С учетом современных тенденций допустимы ссылки на препринты; однако надо учитывать, что научные издания имеют различные варианты политики относительно препринтов, что может сказаться на решении о публикации отчета.

Согласно стандартным российским и международным требованиям, структура научной публикации должна включать следующие основные разделы: введение, материал и методы, результаты, обсуждение, выводы (заключение). Ниже приведены рекомендации по написанию и оформлению каждого основного раздела.

2.1. Введение (*Introduction*)

Объем раздела может составлять 6—20% общего объема основного текста документа (увеличение объема свыше 10% не рекомендовано для отчетов в формате научной статьи). Во «Введении» авторы представляют обоснование научной и практической ценности создания конкретной ММО. Вначале должна быть обоснована актуальность рассматриваемой задачи создания данной конкретной модели. Рекомендуется избегать общих рассуждений о полезности ИИ, сохранения здоровья и т.д. — это удел отдельных публикаций. Необходимо четко обозначить, почему разработка именно данной модели — на фоне иных исследований и моделей аналогичной тематики — является актуальной; привести известные способы решения задачи и их недостатки. Рассуждения об актуальности, решенных и еще не решенных проблемах, подкрепленные цитированием научных публикаций, должны обосновывать цель работы. При необходимости в этом разделе читателя знакомят со структурой отчета.

Раздел должен включать в себя следующие компоненты:

- разъяснение контекста работы (медицинского, научного и т.д.) и мотивации выбора темы;
- обоснование выбранного для данной работы подхода;
- краткий обзор существующих публикаций по теме создания аналогичных моделей;
- цель работы.

Перечисленные компоненты могут быть освещены как в отдельных подпунктах раздела, так и в не разделенном на структурные единицы тексте.

2.1.1. Разъяснение контекста работы (практического, научного и т.д.) и мотивации выбора темы

Контекст работы служит основой для объяснения цели. Выбор тематики базируется на актуальности проблемы, теоретической или практической значимости, новизне полученных результатов.

Пример №2

Пациенты с COVID-19 демонстрируют различную симптоматику, что затрудняет их сортировку. В то время как у некоторых инфицированных пациентов симптомы отсутствуют, у других развивается тяжелый острый респираторный синдром или полиорганная недостаточность. Выявление ключевых характеристик, которые определяют течение болезни в больших когортах пациентов, является актуальной задачей, особенно с учетом его потенциала для помощи врачам и больницам в прогнозировании течения заболевания, эффективном распределении основных ресурсов и улучшении результатов лечения пациентов. Перспективным подходом к решению данной задачи выглядит прогнозирование на основе методов машинного обучения [11].

2.1.2. Обоснование выбранного для данной работы подхода

Необходимо определить и теоретически обосновать концепцию или технологию, которая будет применяться для решения заявленной проблемы, и обозначить перспективность такого подхода; целесообразно кратко охарактеризовать подходы, которые применялись иными исследователями/разработчиками, и емко объяснить, почему полученные результаты неудовлетворительны.

Пример №3

Целенаправленные реабилитационные вмешательства, включая индуцированную ограничением двигательную терапию, двустороннюю тренировку рук, роботизированную терапию и зеркальную терапию, широко используются для устранения двигательной дисфункции у пациентов с хроническим инсультом. Систематические обзоры и метаанализы показали, что эти современные вмешательства эффективны в улучшении двигательной функции у пациентов с инсультом и должны широко применяться в практике. Машинное обучение может стать полезным инструментом для прогнозирования восстановления двигательных функций в процессе реабилитации путем выявления наиболее восприимчивых пациентов и облегчить тем самым практическое использование методик восстановительного лечения [12].

2.1.3. Краткий обзор существующих разработок (публикаций)

В этом разделе следует привести ссылки на публикации по схожей тематике либо обосновать дефицит таковых (фраза «в доступной литературе не обнаружено» и ее аналоги недопустимы). Раскрытие степени исследованности проблемы необходимо для понимания текущего состояния предметной области и обоснования актуальности. Если объем текста достаточно велик, то обзор выполненных исследований имеет смысл вынести в отдельный подпункт (Обзор литературы/*Related works/Theory etc.*), но в большинстве случаев его включают в главу «Введение». Если материал выносят в отдельную часть публикации, то он должен расширять, а не повторять изложенное во «Введении» и закладывать основу для дальнейшей работы. Если в обзоре явным образом содержится сравнение собственной разработки с аналогичными разработками/моделями/подходами, то этот текст целесообразно перенести в раздел отчета «Обсуждение»; во «Введении» тогда остается только обоснование актуальности.

Пример №4

Проблема пропущенных данных при прогнозировании выживаемости пациентов с раком молочной железы по-разному решалась исследователями. В работах *A.K. Rathore* и соавт. (2014), *H. Lotfnezhad Afshar* и соавт. (2015), *N. Shukla* и соавт. (2018) использован набор данных о заболеваемости раком, основанный базе данных *SEER*. В каждом исследовании выбраны различные прогностические переменные, такие как возраст, раса, локализация, семейное положение, первичная локализация, результаты гистологии, степень, размер опухоли, лимфатический узел, распространение, стадия *TNM* и т.д. *E. Acuna* и *C. Rodriguez* (2004) использовали средние значения вместо отсутствующих значений при предварительной обработке данных. В некоторых исследованиях (*D. Delen* и соавт., 2005; *K. Park* и соавт., 2013; *S. Boughorbel* и соавт., 2016) недостающие данные удалены. В исследовании *H. Lotfnezhad Afshar* и соавт. (2015) авторы заменили пропущенные значения, используя среднее значение каждого признака в полном наборе данных в качестве единого исходного значения. Многие алгоритмы могут управлять отсутствующими данными, но обработка выбросов и искаженных данных является сложной задачей и влияет на качество модели прогнозирования (*S. Fielding* и соавт., 2008) [13].

2.1.4. Цель исследования

В заключительной части введения авторы декларируют **цель разработки модели (цель работы)**. Необходимо однозначно указать, чему именно посвящено исследование — разработке модели, ее валидации, в том числе внешней, либо чему-то иному. Опционально цель работы может включать рабочую (нуле-

вую) гипотезу (например, в виде уровня диагностической точности, которого должна достичь ММО в определенных условиях). При формировании нулевой гипотезы могут быть использованы подходы *D.A. Korevaar* и соавт. [14].

2.2. Материал и методы (*Material and methods*)

Данный раздел — ключевой с точки зрения качества, доверия и воспроизводимости полученных результатов. Раздел начинают с указания дизайна исследования; при этом в большинстве случаев применим дизайн диагностического исследования. В качестве справочного ресурса рекомендуется использовать международную базу данных *Equator-Network* (<https://www.equator-network.org>).

Пример №5

Проведено ретроспективное диагностическое исследование. Описание дизайна и результатов выполнено в соответствии с методологией *STARD 2015* (*P.M. Bossuyt* и соавт., 2015). <...> Индекс-тест (исследуемый метод) — распознавание и анализ рентгенограмм посредством программного продукта на основе сверточных нейронных сетей типа *U-NET*, модифицированных и обученных специальным образом. <...> Референсный метод — двойной пересмотр ранее классифицированных рентгенограмм двумя квалифицированными врачами-рентгенологами со стажем работы 8–10 лет (*С. Ланге* и соавт., 2010). Оценку проводили путем бинарной классификации изображений: норма или патология. Клиническую информацию при интерпретации изображений не использовали. Результаты классификации накапливались в виде электронных таблиц, далее подвергались статистическому анализу с помощью ПО *MedCalc*. Значимость результатов определяли по 95%-му доверительному интервалу (ДИ). Проведен расчет следующих показателей: 1. Чувствительность — <...>. 2. Специфичность — <...>. 3. Отношение правдоподобия положительного/отрицательного результата — <...>. 4. Прогностическая ценность положительного результата — <...>. 5. Прогностическая ценность отрицательного результата — <...>. 6. Площадь под характеристической кривой (*AUC* — от англ. *area under curve*) — <...> [15].

Далее авторы должны описать ход работы, методику экспериментов и разработок, указать способы (прежде всего математические и статистические) оценки точности и воспроизводимости полученных результатов, указать критерии (метрики) оценки эффективности ММО.

Ход процесса создания и/или валидации модели целесообразно отобразить в виде диаграммы (*flow-chart*), на которой последовательно отображаются этапы работы, распределение данных и т.д. Для разных дизайнов возможно использование различных

вариантов диаграммы, соответствующие материалы можно почерпнуть из базы данных **Equator-Network** (<https://www.equator-network.org>).

Достаточно удачный пример схематического изображения общего дизайна исследования (**Study design**) приведен на **рис. 3** [16].

Раздел должен содержать информацию, достаточную для воспроизведения данной работы независимым исследователем.

В части данных раздел должен исчерпывающе раскрывать следующую информацию:

- источники данных;
- данные пациентов, включенных в набор;
- описание процесса подготовки данных;
- характеристику набора данных;
- описание процесса разработки и оценки модели;
- методы исследования.

2.2.1. Источники данных (Data sources)

Приводится описание источников информации, на основе которых формировали наборы данных для обучения и валидации. В частности, следует указать названия, типы и принадлежность информационных систем, регистров в сфере здравоохранения; общие характеристики содержащихся в них данных; условия и ограничения доступа к данным.

Пример №6

Мы провели ретроспективный анализ регистра Японской ассоциации неотложной медицины догоспитальной остановки сердца (**JAAM-ОНСА**). Этот регистр представляет собой проспективную общенациональную базу данных, созданную комитетом реестра **JAAM-ОНСА**, детально описанную **T. Kitamura** и соавт., 2018; **T. Irisawa** и соавт., 2018. Вкратце реестр включает догоспитальную, внутрибольничную информацию и сведения об исходах у пациентов с догоспитальной остановкой сердца, доставленных в отделение неотложной помощи. Данные по неотложной помощи собраны в отделениях 87 учреждений (66 университетских больниц и/или третичных центров интенсивной терапии) в Японии. Центры интенсивной терапии сертифицированы Министерством здравоохранения, труда и социального обеспечения Японии и проводят узкоспециализированное лечение, в частности чрескожное коронарное вмешательство. Из включенных в исследование 21 медицинское учреждение не сертифицировано как центр интенсивной терапии, но эти учреждения предоставляют неотложную медицинскую помощь населению, а некоторые могут обеспечить экстракорпоральную сердечно-легочную реанимацию и интенсивную терапию. Всего в реестре **JAAM-ОНСА** зарегистриро-



Рис. 3. Пример блок-схемы, описывающей общий дизайн исследования.

Fig. 3. Example of flowchart describing overall study design.

ваны 34 754 пациента с догоспитальной остановкой сердца с июня 2014 г. по декабрь 2017 г. [17].

2.2.2. Данные пациентов, включенных в набор (Study population/Participants/Patient populations etc.)

Приводят следующие данные:

1. Характер и объем общей выборки, то есть группы субъектов (пациентов), включаемых в набор данных для создания ММО; также указывают в том числе этническую, национальную, половозрастную, нозологическую, социальную и иные релевантные структуры общей выборки. Рекомендуется формировать наборы данных эквивалентно структуре популяции и заболеваемости (распространенности) целевой нозологией.

2. Критерии включения, невключения и исключения. Критерии включения определяются клинической или иной задачей, которая ставится перед разрабатываемой моделью, чаще они связаны с неким вмешательством — конкретным видом диагностического исследования, хирургического лечения или с конкретной нозологией (патологическим состоянием). Критерии исключения чаще определяют в ходе работы с первичными данными; в целом критерии могут носить медицинский (например, диапазон возраста, наличие сопутствующей патологии и т.д.) и/или технический (параметры и протоколы проведенных диагностических исследований) характер [18].

3. Условия, в которых формировались исходные данные, в том числе временной период, географический регион, виды, формы и условия оказания медицинской помощи и т.д.

4. Механизм обеспечения этики и законности использования данных; чаще всего это информация о соглашении с операторами информационных систем и биомедицинских данных, а также одобрение исследования комитетом по биоэтике. Опционально, в случае проспективного формирования набора данных, могут рассматриваться аспекты получения информированного добровольного согласия пациента.

Приводят характеристику включенных пациентов (базовые демографические признаки, клинические признаки, доступные предикторы), в том числе количество лиц (случаев) с отсутствующими, неполными данными и результатом. Практика представления этих результатов в табличном виде с развернутым комментарием наиболее оптимальна. В случае большого объема рекомендуется вынести наглядный материал в приложение.

Пример №7

В анализ включены 423 604 участника, для которых доступно достаточное количество информации. Средний возраст участников на исходном уровне составлял 56,4 (стандартное отклонение 8,1) года,

и 188 577 (44,5%) участников — мужчины. За период наблюдения (медиана 7 лет) зарегистрировано 6703 случая сердечно-сосудистых заболеваний (ССЗ). Средний возраст пациентов с ССЗ составил 60,5 года (60,2 года для мужчин и 61,1 года для женщин). Поскольку минимальный период наблюдения для всех участников составлял 5 лет, мы оценивали точность различных моделей в предварительном определении 5-летнего риска развития ССЗ. На 5-летнем горизонте общее количество случаев ССЗ составило 4801 [19].

2.2.3. Описание подготовки набора данных для создания модели

Следует описать все этапы предварительной обработки данных (препроцессинга), включая очистку и преобразование. Необходимо указать методы работы с пропущенными данными, критерии определения избыточных данных, очистки выбросов и шума (указывая критерии для удаления выбросов), нормализации, дискретизации, статистической обработки и т.д. По всем примененным методам предварительной обработки данных необходима подробная информация.

Если при подготовке набора данных использована разметка, то детально описывают методику такой разметки, включая использованные инструменты, квалификацию персонала и т.д. При этом рекомендуется придерживаться стандартной классификации качества разметки [12]:

- низкое (факт наличия находки) — формируется на основе медицинской документации;
- среднее (классификация находок) — формируется на основе экспертного мнения;
- высокое (подтвержденный диагноз) — формируется на основе результатов более чувствительного метода исследования либо динамического наблюдения (повторного проведения того же метода через определенный временной промежуток).

Пример №8

Для предварительной обработки данных категориальные переменные преобразованы в числовые с однократным кодированием, за исключением параметра «участок окклюзии», которому присваивался порядковый номер (общая сонная артерия — 0, внутренняя сонная артерия — 1, средняя мозговая артерия, сегмент M1 — 2, сегмент M2 — 3, сегмент M3 — 4). Параметры, имеющие размерность временной отметки, преобразованы в продолжительность между наблюдаемыми событиями. Непрерывные переменные дихотомизированы следующим образом: преморбидные по шкале mRS разделены на 0—2 (ранее не нуждавшиеся в уходе) и от 3 до 5 (ранее нуждавшиеся в уходе), поскольку информация о том, является ли пациент нуждающимся в уходе до начала инсульта, имеет решающее значение для прогноза развития событий

после инсульта. Время прибытия в больницу разделено на промежутки с 7:00 до 17:00 (прибыли в рабочее время) и с 17:00 до 7:00 (прибыли в нерабочее время), потому что качество лечения пациентов с окклюзией крупных сосудов в острой фазе может заметно различаться в эти периоды. Заполнение пропущенных значений не проводилось. Кроме того, мы не стандартизировали каждую переменную на этапе препроцессинга, потому что стандартизация со всем набором данных дает дополнительную информацию к каждой переменной. Каждая переменная стандартизирована только после того, как набор данных разделен на обучающие данные и тестовые данные [20].

2.2.4. Характеристика набора данных (Data collection/Study data)

Детально описывают объем, структуру и характеристику всех наборов данных, использованных в работе. Рекомендуется привести обоснование размера выборки.

В качестве характеристики набора данных обязательно должны быть указаны:

- общий объем;
- классы, баланс классов, количество записей в каждом классе;
- наличие и вид разметки данных;
- переменные-предикторы;
- стандарты предоставления медицинских данных (например, DICOM).

Раздел иллюстрируют таблицами с описанием переменных (тип данных, наличие пропущенных значений и т.д.) и образцом набора данных или статистическим резюме.

Пример №9

Набор данных для исследования сформирован в 2019 г. в НПКЦ ДиТ ДЗМ с использованием базы данных результатов программы скрининга рака легкого методом НДСТ. Отбор случаев из базы проведен случайным образом. В набор включены данные 100 лиц, прошедших скрининговое обследование и подписавших информированное добровольное согласие в городских поликлиниках Москвы. Всего включены данные 80 мужчин (средний возраст $64,8 \pm 9,4$ года) и 20 женщин (средний возраст $65,9 \pm 8,3$ года); это лица европеоидной расы, жители Москвы. Набор данных разделен на две подгруппы: «норма» ($n=32$) и «патология» ($n=68$). Общие критерии для подгрупп «норма» и «патология»:

- лица от 55 до 80 лет;
- индекс курения — более 30 пачка/лет;
- курение или отказ от курения в пределах 15 лет;
- отсутствие симптомов (кроме кашля, мокроты, одышки, связанных с курением);
- отсутствие КТ органов грудной клетки в течение последнего года;

- отсутствие в течение последнего месяца повышения температуры тела более $37,5^{\circ}\text{C}$, кровохарканья, розовой или «ржавой» мокроты, боли в грудной клетке, сиплого голоса, необъяснимого похудения;
- отсутствие рака легкого и метастазов в легких.

Критерии отнесения к подгруппе «норма»:

- выполнена первичная низкодозная КТ (НДКТ) из программы скрининга рака легкого;
- после первичного исследования имеется повторная скрининговая НДКТ, выполненная более чем через 600 дней после первичного исследования;
- после первичного и повторного скрининговых исследований у пациента не выявлен рак легкого.

Критерии отнесения к подгруппе «патология»:

- в течение 3 мес после первичной НДКТ для скрининга верифицировано злокачественное новообразование легкого.

Всем обследуемым выполнена НДКТ грудной клетки. Исследования проведены на 10 компьютерных томографах, имеющих 64 среза (Toshiba Aquilion 64 и Toshiba CLX). <...> Данные сканирования обрабатывали, анализировали и сохраняли с помощью программного обеспечения в составе Единого радиологического информационного сервиса. Классификация очагов проведена в соответствии с LungRADS (версия 2014 г.) [21].

При использовании наборов данных, созданных иными авторами и размещенных в открытом доступе, необходимо корректно процитировать такие наборы.

Пример №10

На сайте Московского эксперимента по компьютерному зрению в лучевой диагностике (<https://mosmed.ai>) размещен в открытом доступе набор данных с результатами КТ пациентов с COVID-19 [22]. В справочном файле представлена информация об условиях использования («Набор данных доступен под лицензией Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0) License») и строка для цитирования в научных и иных публикациях в случае использования этого набора данных сторонними разработчиками.

2.2.5. Описание процесса разработки и оценки/валидации модели (Model development/Model derivation and validation/Experimental evaluation)

В данном разделе приводят обоснование выбора алгоритмов, подробно описывают все манипуляции по построению модели, методику валидации и метрики оценки модели. Здесь же описывают, как разделен набор данных на обучающие и оценочные. Допустимо представление в схематическом виде с возвращенными комментариями.

Отметим, что значительную ценность работе придает независимая проверка модели: как минимум — на наборах данных, которые не использованы для об-

учения (не «оценочный» набор, а полностью новый, например взятый из сторонних открытых источников), как максимум — валидация независимой группой экспертов.

Метрики оценки модели рекомендуется выбирать с учетом правил биостатистики и принципов доказательной медицины.

Пример №11

Для разработки моделей прогнозирования использованы два алгоритма машинного обучения: **k-nearest neighbor (KNN)** и искусственная нейронная сеть (**ANN**). Алгоритм **KNN** — один из наиболее широко используемых инструментов интеллектуального анализа данных для классификации и прогнозирования закономерностей данных информатики здоровья (M. Zhu и соавт., 2007). Алгоритм **KNN** предсказывает, что похожие объекты будут существовать в непосредственной близости; в результате он маркирует класс цели на основе окружающих ее **k** соседей (T. Cover и соавт., 1967). Этот метод прогнозирования похож на процесс принятия клинического решения клиницистами/терапевтами. В большинстве случаев клиницист/терапевт может порекомендовать конкретное вмешательство новому пациенту, если профиль этого нового пациента совпадает с профилями тех пациентов, которые успешно вылечены этим конкретным вмешательством. Таким образом, алгоритм **KNN** можно рассматривать как искусственную экспертную систему, которая прогнозирует ответы участников на основе обширного опыта, полученного в ходе обучения. Алгоритм **ANN** основан на биологических нейронных сетях человеческого мозга (T. Manning и соавт., 2014). Подобно человеческой нейронной сети, вычислительная система **ANN** состоит из нескольких нейронов/узлов на разных уровнях, включая входной, скрытый и выходной уровни. Преимущество алгоритма **ANN** заключается в том, что он может фиксировать сложные нелинейные отношения между входными и выходными переменными посредством вычислений в скрытых слоях, что делает его одним из идеальных инструментов для прогнозирования результатов у пациентов с инсультом (N. Shahid и соавт., 2019). В этом исследовании использован алгоритм **ANN** с прямой связью. Мы приняли один скрытый слой и определили оптимальное количество скрытых нейронов в скрытом слое, используя метод перекрестной проверки в **k**-кратном размере [12].

2.2.6. Методы исследования

Раздел содержит методы исследования (методы научного познания, использованные для обоснования, создания и оценки модели); обязательно — метрики оценки качества, точности, надежности, эффективности ММО.

Ранее опубликованные в других работах методы, в том числе статистического анализа, перечисляют тезисно и представляют в виде ссылки. Любые модификации существующих методов должны быть подробно описаны. С учетом специфики сферы здравоохранения следует придерживаться правил биостатистики, в том числе в части определения объема выборки, выбора методов статистического анализа и конкретных метрик (например, диагностической точности, согласованности классификаций и т.д.), способа подтверждения значимости полученных результатов (*p*-значение вероятности, 95% ДИ). Настоятельно рекомендуется использовать метрики (статистические методы), стандартно применяемые в медицинской научной литературе, описанные в руководствах по медицинской и биологической статистике.

2.2.7. Содействие взаимодействию и воспроизведению (Promoting interoperability and replicability)

Раздел содержит информацию деонтологического и этического характера, в том числе транслирующую прозрачность работы, ее пригодность для проверки воспроизводимости результатов (например, «в представленной публикации мы опираемся на рекомендации **TRIPOD**», «исходный код выложен на **GitHub**», «набор данных представлен в открытом доступе» и т.д.).

2.3. Результаты

Важнейшим элементом отчета является представление результатов создания ММО (метрик, значений, классификаций). Рекомендуется формировать результаты работы модели с использованием принятой медицинской номенклатуры, на основе стандартов (например, Международной классификации болезней, **SNOMED**), справочников национального уровня (например, ФСИДИ), клинических рекомендаций.

Пример №12

Результатом нашего исследования стала 1-месячная выживаемость с благоприятным неврологическим исходом, определяемая по Категории церебральных функций (**CPC**) 1 или 2.14. Подробности приведены в Приложении 1 [17].

Желательно предоставлять результаты как можно большего количества проведенных экспериментов для последующего сравнения и анализа. Рекомендуется предоставлять разнообразную визуализацию полученных результатов для повышения наглядности — графики, диаграммы и т.д. Непосредственно в текст отчета выносят только важнейшие, принципиальные для дальнейшего обсуждения результаты, остальные материалы лучше представлять в виде приложений. В этом разделе также демонстрируют взаимосвязь отдельных предикторов с выходным результатом.

Пример №13

Для каждой модели **XGBoost** рассчитаны средние абсолютные значения **SHAP** (рис. 4) [11]. Важнейшими факторами для прогнозирования критических последствий в течение 1 нед оказались наличие острого повреждения почек, высокие и низкие уровни лактатдегидрогеназы (ЛДГ), частота дыхания и уровень глюкозы. К другим важным прогностическим факторам относятся уровни систолического и диастолического артериального давления, **pH**, общий уровень белка, содержание **C**-реактивного белка и **D**-димера. Что касается смертности, то как высокие, так и низкие значения возраста, анионного разрыва, уровня **C**-реактивного белка и ЛДГ оказались самыми сильными факторами, влияющими на прогноз смертельного исхода в течение 1 нед после госпитализации. Другие важные переменные для повышения точности прогноза смерти включают уровень сатурации при поступлении в больницу, уровень

азота мочевины крови, ферритина, ширину распределения эритроцитов (**RDW**), уровень диастолического артериального давления и содержание лактата крови. Используя оценки взаимодействия **SHAP**, мы обнаружили, что ковариантные взаимодействия между функциями вносят меньший вклад в точность предсказания моделей, чем значение каждой функции (рисунки **S11** и **S12** в мультимедийном приложении 2), за исключением случая **AKI**, где уровни ЛДГ, глюкозы и **C**-реактивного белка выступают сильными ковариатами. Для сравнения мы также оценили значимость функции модели **LASSO** для этих экспериментов (рисунок **S9** в мультимедийном приложении 2). Мы увидели совпадение ключевых характеристик, которые обе модели считают важными при прогнозировании критических событий и смерти на 7-й день. Мы обнаружили, что для критических событий наиболее важной особенностью обеих моделей является **AKI**. Учащенное поверхностное дыха-

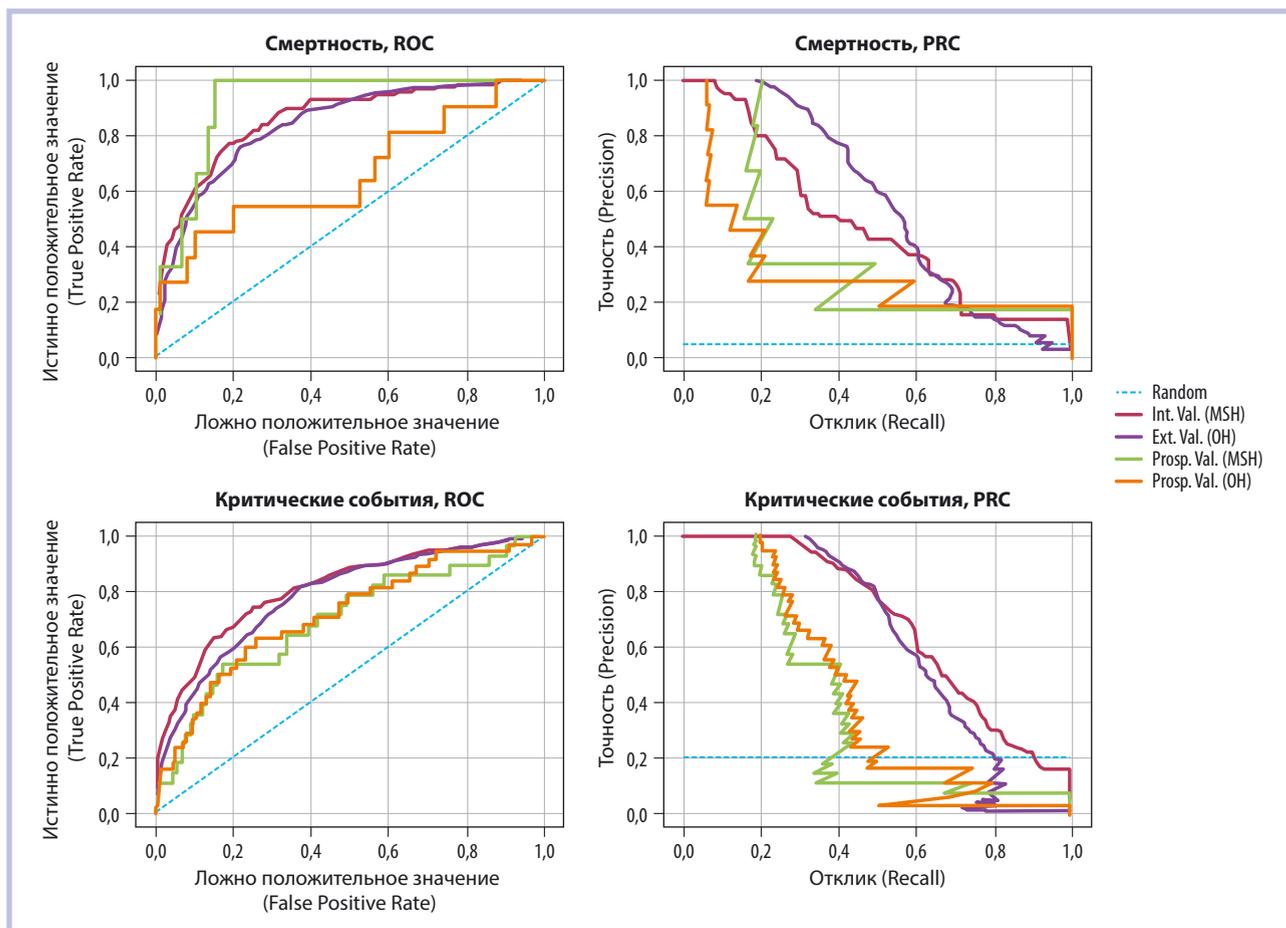


Рис. 4. Производительность классификатора **XGBoost**. ROC-кривые (левая панель) и кривые точности-отклика (правая панель) для смертности (вверху) и критических событий (внизу) в проверочных экспериментах. Для обобщения показаны результаты модели **XGBoost**, полученной на основе перекрестной проверки данных из **MSH** и примененной ко всем другим больницам.

Fig. 4. Performance of the **XGBoost** classifier. ROC curves (left panel) and precision-recall curves (right panel) for mortality (top) and critical events (bottom) in validation experiments. **XGBoost** model results derived from cross-validation of **MSH** data and applied to all other hospitals.

ние и повышенный уровень **D**-димера также связаны с более высокой частотой смертельных исходов, а более низкое диастолическое артериальное давление связано отрицательно. Что касается смертности, то мы также выявили соответствие ключевых характеристик обеих моделей. В частности, пожилой возраст и более высокий анионный разрыв играют роль сильных факторов в прогнозировании смертельного исхода в обеих моделях, а низкие значения диастолического артериального давления и сатурации демонстрируют отрицательную корреляцию. Обнадёживает тот факт, что многие функции, имеющие большое значение в основной модели **XGBoost**, также отнесены к приоритетным в классификаторе **LASSO**, что свидетельствует о прогностической надёжности этих функций. В мультимедийном приложении 7 перечислены 10 основных характеристик моделей «Критическое событие» и «Смертельный исход в течение 7 дней».

Большинство авторов избегают упоминать об экспериментах с отрицательным результатом. Между тем такие эксперименты могут оказаться полезнее экспериментов с положительным исходом. В этой связи рекомендуется описывать не только позитивные результаты работы, но и неудачные или результаты и показатели, которые не соответствуют ожидаемым метрикам качества и точности.

Необходимо предоставлять информацию о том, каким образом обеспечиваются стандартизация и объяснимость результатов работы ММО конечному пользователю. Концепция «объяснимого ИИ» критически важна для сферы здравоохранения: процесс, причины и результаты принятия решения ММО должны быть доступны (наглядно представлены) и понятны медицинскому работнику — пользователю ММО.

2.4. Обсуждение

Раздел принципиальной важности, во многом определяющий практическую и научную ценность созданной модели. Эта часть должна исследовать значение результатов работы, а не повторять их. Требуется изложение рассуждений о результатах, их значении, взаимосвязях, влиянии. Целесообразно не объединять разделы «Результаты» и «Обсуждение». Рекомендуется избегать многословных цитирований и подробного обсуждения литературных источников. Оптимальным является лаконичное сопоставление собственных результатов и заключений с представленными в литературе. Значение имеет не обзор, а критический самоанализ. Структура раздела должна предупредить возможные вопросы оппонентов. В данном разделе должен быть последовательно определен и описан ряд моментов, отраженных далее.

2.4.1. Интерпретация

Должна быть приведена общая интерпретация результатов с учетом заявленных целей, декларируе-

мых ограничений и допущений, дана авторская оценка эффективности разработанной модели в соответствии с поставленными задачами. Следует провести сравнение с результатами и выводами аналогичных исследований и другими релевантными доказательствами качества; указать значимость (научную, медицинскую, организационно-управленческую и т.д.) модели; перечислить все преимущества подхода и реализации. Обязательно анализируют воспроизводимость результатов работы ММО в разных клинических условиях, на разных наборах данных. Обсуждают аспекты неопределенности и объяснимости результатов работы ММО.

Пример №14

Наша новая модель ИИ, **EDRnet**, смогла предсказать смертность пациентов с **COVID-19**, используя 28 биомаркеров крови, полученных в течение 24 ч после поступления в больницу. В наборах данных независимого тестирования **EDRnet** показала отличную производительность прогнозирования с высокой чувствительностью (100%), специфичностью (91%) и точностью (92%). Мы смогли улучшить характеристики прогнозирования, приняв ансамблевой подход, сочетающий модели **DNN** и **RF**. Следует отметить, что **EDRnet** разработана путем обучения на данных китайских пациентов и тестирования на данных корейских пациентов. **EDRnet** имеет несколько преимуществ. Во-первых, **EDRnet** может предсказать, какие пациенты подвержены высокому риску смерти на ранней стадии госпитализации (то есть в течение 24 ч после госпитализации). Это существенное улучшение по сравнению с предыдущей моделью прогнозирования ИИ, которая прогнозировала смертельный исход за 10 дней до наступления выживания или смерти (**L. Yan** и соавт., 2020). Прогноз смерти на момент госпитализации может быть в значительной степени информативным для клиницистов, поскольку критическое время в отношении прогрессирования заболевания составляет 10–14 дней с момента появления симптомов, согласно предыдущим исследованиям (**Q. Ruan** и соавт., 2020). **EDRnet** может предоставить рекомендации по приоритетности лечения в отношении того, кого следует лечить интенсивно. Во-вторых, **EDRnet** использует только биомаркеры крови для прогнозирования смертельного исхода. Как правило, у пациентов с **COVID-19** исследование крови проводят при поступлении в больницу. Биомаркеры крови — это объективные показатели, которые используются для количественной оценки состояния пациентов, что может быть полезно для обеспечения надежности модели ИИ. Мы не включали ни субъективные биомаркеры, такие как симптомы, ни предрасполагающие факторы, такие как сопутствующие заболевания, потому что эти индексы трудно количественно

оценить и может возникать высокая вариабельность между пациентами. В-третьих, клиническое значение и значение биомаркеров крови, используемых в нашей модели **EDRnet**, хорошо изучены в ходе многих предыдущих клинических исследований. Таким образом, результаты прогнозируемой ИИ смерти объяснимы и легко понимаются врачами. Кроме того, в нашей модели **EDRnet** используется несколько основных биомаркеров крови.

В этом исследовании не наблюдалось значительных различий в использовании фармакологических средств между пациентами групп умерших и выживших, за исключением антибиотиков, и в применении противовирусных препаратов, таких как ремдесивир. Антибиотики или комбинированная терапия обычно используются при подозрении на бактериальную суперинфекцию. На сегодняшний день не существует безусловно эффективных фармакологических средств для лечения **COVID-19**. В этом исследовании фармакологическое лечение существенно не связано с выживаемостью.

EDRnet не требует всех 28 биомаркеров крови для прогнозирования смертельного исхода. **EDRnet** работала хорошо, если на момент госпитализации определялось не менее 19 биомаркеров крови. По сравнению с предыдущими моделями ИИ прогнозирования смерти от **COVID-19**, в которых использовались 3 биомаркера, могут возникнуть опасения, что **EDRnet** требует слишком много биомаркеров. Однако эти анализы крови обычно выполняются в нашей повседневной клинической практике для госпитализированных пациентов с **COVID-19**. Если будет накоплено больше данных, мы сможем уменьшить количество биомаркеров крови для прогнозирования смертельного исхода [23].

2.4.2. Ограничения

В этом разделе указывают все ограничения и допущения полученной модели. Обязательно обозначается отсутствие внешней валидации, если таковая не проводилась.

Пример №15

У нашего исследования имеется несколько ограничений. Во-первых, некоторые пациенты с очень большими гематомами или пребывающие в критическом состоянии не включены в исследование по решению семьи, что могло ограничить возможность обобщения результатов. Во-вторых, несмотря на хорошую характеристику алгоритмов, размер выборки относительно невелик. Исследования с более крупными выборками могут продемонстрировать большую прогностическую эффективность. В-третьих, мы не включили в эту модель информацию о раннем росте гематомы и распространении отека, которая важна для исхода. Наконец, отсутствует внешняя проверка для опреде-

ления обобщаемости и исключения институциональной предвзятости.

2.4.3. Перспективы применения

На основе метрик качества и точности модели следует оценить возможности ее применения в условиях реальной клинической практики, включая ограничения, особенности и другие важные данные, которые необходимо учесть во время развертывания и применения модели. Указывают возможности или опыт интеграции модели в действующие информационные системы в сфере здравоохранения; перспективы регистрации программного обеспечения на основе модели в качестве медицинского изделия.

2.4.4. Направление дальнейших улучшений модели

Если при создании модели предложены некоторые идеи в части дальнейшего улучшения метрик точности модели, ее эффективности и безопасности, их следует кратко отразить в данном разделе.

2.5. Выводы (Conclusions)

Основные итоги работы могут быть представлены в отдельном коротком разделе «Выводы», который допустимо включать как подраздел в главу «Обсуждение» или «Результаты и обсуждение». В этом разделе кратко подводятся итоги: каких результатов удалось добиться, какие проблемы остались открытыми с необходимостью проведения дальнейших исследований и разработок. Если ранее сформулирована не только цель, но и задачи работы, то целесообразно сформировать каждый вывод как ответ на конкретную задачу. Оптимально включить в выводы основные численные метрики использованных наборов данных и качества модели.

Пример №16

Модель машинного обучения с несколькими стандартными клиническими переменными перед лечением может улучшить прогноз исходов у пациентов с окклюзией крупных сосудов, которым выполнена механическая тромбэктомия. Даже худшая модель машинного обучения показала значительно более высокие оценки, чем лучшие предыдущие оценки прогнозов, включая оценку **PRE**, индекс **SPAN**, оценку **THRIVE**, оценку **NIAT** и оценку **NIAT2**. Модель машинного обучения также превосходила стандартную статистическую модель, основанную на логистической регрессии. Насколько нам известно, это первое исследование, показывающее, что машинное обучение может лучше предсказывать клинический результат у пациентов с окклюзией крупных сосудов, чем ранее разработанные методы оценки. Если модель будет подвергнута дальнейшей проверке и доработке, она может стать полезным инструментом для поддержки решения о выполнении машинного обучения.

3. Вспомогательная часть публикации

Включает список использованных источников, а также дополнительные схемы, таблицы, графики и т.д., оформленные в виде приложений к отчету. При формировании этой части необходимо учитывать требования политики научных рецензируемых журналов в части информирования об источниках финансирования, наличия конфликта интересов, степени и формы вовлеченности авторов в подготовку отчета.

3.1. Приложения

В этом разделе представляют графические и табличные сведения, не вошедшие в основную часть отчета. Ссылки на приложения приводят в соответствующих разделах основного текста. В случае публикации отчета онлайн допустимо и целесообразно представлять мультимедийное приложение. В приложениях могут размещаться ссылки на наборы данных, программный код, готовые программные продукты и прочее.

3.2. Список литературы (References)

Тщательный подбор и анализ литературы для отчета — важный компонент научной работы. Без изучения текущего состояния проблемы невозможно корректно и обоснованно сформулировать цель и задачи; без критичного сопоставления с ранее опубликованными материалами невозможно объективно проанализировать собственные результаты и наметить реальные пути дальнейшего развития.

Список литературы оформляют в соответствии с общими требованиями к описанию библиографических источников. Оптимально придерживаться «Единых требований к рукописям, представляемым в биомедицинские журналы», разработанных Международным комитетом редакторов медицинских журналов (так называемый Ванкуверский стиль). Однако необходимо учитывать требования рецензируемого научного журнала, если отчет планируется публиковать как статью.

Для обзора литературы используют электронные библиографические базы данных и индексы цитирования (РИНЦ, PubMed, Scopus и т.д.) и специальные инструменты. Аннотированный список источников целесообразно создавать с помощью библиографических менеджеров Zotero или Mendeley. В перечень работ рационально включать собственные предыдущие публикации (не более 3), в порядке самоцитирования использовать их во «Введении» и/или «Обсуждении».

Чек-лист

Должна существовать возможность быстрой, стандартизированной оценки качества рукописи научной статьи о разработке и валидации ММО. Для этого разработан оригинальный чек-лист, базирующийся на изложенном выше методическом подходе (таблица). Этот инструмент позволяет проводить ускоренную проверку структуры, полноты и качества содержания рукописи. Чек-лист отличается универсальностью: он может применяться авторами для самоконтроля, рецензентами, а также редакторами научных журналов для оценки качества и принятия решения о публикации. Оценка соответствия того или иного раздела рукописи соответствующему пункту чек-листа осуществляется путем выбора одного из трех вариантов: «да» — полное соответствие, «нет» — несоответствие, «частично» — неполное соответствие.

Обсуждение

В настоящее время существует несколько методических разработок близкой тематики.

F. Cabitza и A. Campagner предложили «развернутый» чек-лист для оценки и самооценки качества исследований на тему медицинского ИИ [24]. Этот инструмент, согласно позиции разработчиков, предназначен как для сотрудников редакционных коллегий научных журналов и рецензентов, так и для самих авторов. Углубленный анализ содержания вопросов ука-

Чек-лист для подготовки отчета о научном описании создания и валидации модели машинного обучения

Checklist for preparing a report on scientific description of development and validation of machine learning model

Раздел	Номер	Содержание	Особое внимание	Оценка соответствия
1. Информационная часть				
Название	1.1	Отражает содержание, тематику и особенности разработки модели	Объем 10–12 слов Ключевые слова по тематике работы	Да Нет Частично
Аннотация	1.2	Структурированное описание сути процесса и результатов работы для быстрого ознакомления читателя с содержанием основного текста	Структура идентична структуре основного текста	Да Нет Частично
Ключевые слова	1.3	Обеспечивают трансляцию содержания отчета и его поиск в библиографических базах данных	До 10 ключевых слов Использован справочник РИНЦ Использован справочник NLM MeSH	Да Нет Частично

Продолжение таблицы см. на след. странице

Чек-лист для подготовки отчета о научном описании создания и валидации модели машинного обучения (Продолжение)
 Checklist for preparing a report on scientific description of development and validation of machine learning model

Раздел	Номер	Содержание	Особое внимание	Оценка соответствия
2. Основная (содержательная) часть				
Введение	2.1	Обоснование научной и практической ценности создания модели машинного обучения	Лаконичный научный обзор	Да Нет Частично
	2.1.1	Контекст исследования (медицинского, научного и т.д.) и мотивация выбора темы	Клиническое целеполагание Роль и место модели в реальных процессах медицинских организаций Потенциальное влияние на показатели индивидуального и общественного здоровья, деятельности медицинских организаций	Да Нет Частично
	2.1.2	Обоснование выбранного подхода	Убедительное объяснение преимущества оригинального подхода	Да Нет Частично
	2.1.3	Краткий обзор существующих исследований/моделей	Критический анализ: — достигнутого уровня решения рассматриваемой проблемы — собственного предыдущего опыта	Да Нет Частично
	2.1.4	Цель исследования	Конкретная, однозначная формулировка	Да Нет Частично
Материал и методы	2.2	Дизайн и ход разработки модели	Дизайн работы, принятый в медицинской науке Блок-схема (flow-chart)	Да Нет Частично
	2.2.1	Источники данных	Типы и принадлежность информационных систем и регистров в сфере здравоохранения Условия и ограничения доступа к источникам данных	Да Нет Частично
	2.2.2	Данные пациентов, включенных в набор	Критерии включения, невключения, исключения Структура набора данных эквивалентна структуре популяции Математический расчет объема выборки	Да Нет Частично
	2.2.3	Характеристика набора данных	Общий объем Классы, баланс классов, количество записей в каждом классе Наличие и вид разметки данных Переменные-предикторы Стандарты предоставления медицинских данных Цитирование открытых наборов данных	Да Нет Частично
	2.2.4	Описание процесса подготовки данных	Тип разметки в соответствии с классификацией	Да Нет Частично
	2.2.5	Описание процесса разработки и оценки модели	Метрики качества модели соответствуют требованиями биостатистики Валидация на новых данных Внешняя независимая валидация	Да Нет Частично
	2.2.6	Методы исследования	Метрики оценки качества, точности, надежности, эффективности модели Подбор метрик и анализ на принципах медицинской и биологической статистики	Да Нет Частично
	2.2.7	Содействие взаимодействию и воспроизведению	Биоэтика, деонтология Компоненты в свободном доступе	Да Нет Частично
	2.3	Результаты, полученные в процессе разработки и валидации модели машинного обучения	Результаты работы модели соответствуют медицинской номенклатуре, стандартам, клиническим рекомендациям Графическое и табличное представление Информация о значимости Информация о стандартизации и объяснимости результатов работы модели	Да Нет Частично

Продолжение таблицы см. на след. странице

Чек-лист для подготовки отчета о научном описании создания и валидации модели машинного обучения (Окончание)
 Checklist for preparing a report on scientific description of development and validation of machine learning model

Раздел	Номер	Содержание	Особое внимание	Оценка соответствия
Обсуждение	2.4	Анализ полученных результатов	Баланс анализа положительных, отрицательных и нейтральных результатов Критический анализ, сравнение и сопоставление с данными литературы, собственными предыдущими результатами	Да Нет Частично
	2.4.1	Интерпретация	Принятие или отклонение нулевой гипотезы Проанализирована воспроизводимость результатов работы модели	Да Нет Частично
	2.4.2	Ограничения	Объективная характеристика недостатков, допущений и барьеров	Да Нет Частично
	2.4.3	Перспективы применения	Возможности или опыт интеграции модели в действующие информационные системы в сфере здравоохранения Перспективы регистрации программного обеспечения на основе модели в качестве медицинского изделия	Да Нет Частично
	2.4.4	Направление дальнейших улучшений модели	Конкретные действия по улучшению	Да Нет Частично
Выводы	2.5	Итоговая систематизация результатов относительно цели и задач исследования	Выводы соответствуют задачам Содержат численные метрики основных результатов	Да Нет Частично
3. Вспомогательная часть публикации				
	3.1	Приложения	Ссылки на набор данных, исходный код и т.д.	Да Нет Частично
	3.2	Список литературы	Баланс русскоязычных источников и источников на иностранных языках Самоцитирование	Да Нет Частично

зывает на «узкую» направленность этого чек-листа. В реальности он адресован рецензентам, позволяет последовательно оценить методическое содержание, логичность и последовательность изложения. Если же говорить об авторах, то данный чек-лист будет полезен только опытным исследователям (относительно немногочисленным). В целом этот инструмент помогает определиться в решении о принятии или отклонении рукописи с позиции редакционной коллегии журнала. Его возможности для самооценки качества самими авторами статей выглядят сомнительными.

Аналогичную оценку можно дать и одной из первых разработок в изучаемой сфере — чек-листу «минимальной информации для исследований клинических ММО» **MI-CLAIM** [8]. Этот инструмент обладает выраженной лаконичностью, однако не позволяет в должной мере осветить клиническую составляющую процесса создания ММО.

Предложена модификация чек-листа для формирования протоколов (планов) клинических исследований **SPIRIT-AI**. В основе этого инструмента лежит известный и широко валидированный чек-лист **SPIRIT** для описания универсального протокола научного исследования в сфере медицины [25]. Авторы **SPIRIT-AI** проделали большую методическую работу,

однако фактические дополнения в части «искусственного интеллекта» минимальны. Из 33 основных пунктов базового чек-листа только 7 содержат особенности, признанные авторами инструмента важными для описания ММО. Вместе с тем некоторые особенности выглядят «искусственными». Например, необходимость указывать в названии исследования факт применения технологий ИИ, область такого применения, суть вмешательства, производимого посредством ММО. Это достаточно общая рекомендация. С тем же успехом можно рекомендовать авторам статей по электрокардиографии использовать в названии статей слово «электрокардиография». Полезным методическим приемом можно считать рекомендацию указывать критерии включения и исключения не только для пациентов, но и отдельно для данных (наборов данных), а также описывать процедуру оценки и работы с некачественными или недоступными данными. В целом инструмент **SPIRIT-AI** может быть рекомендован для разработки плана проведения научного исследования, но не для описания его результатов.

Эту проблему тот же коллектив авторов попытался решить путем создания расширения ранее известного чек-листа **CONSORT** [26]. Здесь наблюда-

ется аналогичная ситуация. В основе — надежный и валидный методический инструмент для авторов научных статей в области медицины и здравоохранения. Для нескольких пунктов добавлены «расширения» в части ИИ, большая часть которых неинформативна, не содержит действительно специфических особенностей и требований. Например, о том, что в актуальности работы требуется описывать применение изучаемой ММО в конкретном клиническом контексте. Вместе с тем трудно предположить исследование нового лекарственного вещества без предоставления в статье информации о показаниях, противопоказаниях и режимах его применения. Соответственно, большинство дополнительных пунктов **CONSORT-AI** являются очевидными и совершенно неспецифичными для темы машинного обучения и технологий ИИ. Вместе с тем интересен пункт с требованием указывать уровень компетенций, обязательных для конечных пользователей изучаемой ММО. На наш взгляд, **CONSORT-AI** не имеет принципиальных отличий от базового чек-листа **CONSORT**, не содержит специфических или уникальных требований и особенностей.

К сожалению, несколько чек-листов, предложенных для описания ММО в отдельных клинических сферах (урологии, кардиологии, лучевой диагностике, стоматологии), можно расценить лишь как «дань моде» [27—29]. В них методически верно, очень подробно и системно описаны требования и структура научных медицинских исследований, но специфических особенностей, связанных с технологиями ИИ, фактически нет. Словосочетания «искусственный интеллект» и «машинное обучение» могут быть заменены любыми другими (например, «остеосинтез аппаратом внешней фиксации»), при этом на сути чек-листа это никак не скажется. Наблюдается и обратная ситуация, когда чек-лист соответствует дизайну научной статьи в области математики, при этом практически отсутствуют медицинская специфика и клинические составляющие.

Положительной оценки заслуживают отдельные компоненты, включенные в чек-лист **F. Schwendicke** и соавт. для описания научных исследований технологий ИИ в стоматологии [28]. В отличие от других исследователей авторы включили в свой инструмент такие пункты, как «Неопределенность» (описание неопределенностей в результатах работы ММО) и «Объяснимость» (оценка объяснимости, достоверности и прозрачности работы ММО). Помимо этого, в чек-лист включены отдельные вопросы подготовки наборов данных и достаточно детальные компоненты процесса создания ММО [28]. Такой подход мы считаем верным методологически, заслуживающим положительной оценки. Здесь наблюдается достаточно успешная интеграция специфических аспектов машинного обучения и стандартной медицинской научной статьи.

В своей работе мы подошли к проблеме создания чек-листа комплексно, определив:

- необходимость наличия стандартной структуры медицинской научной статьи;
- баланс представления клинической и технологической информации;
- включение специфических аспектов машинного обучения и технологий ИИ, причем наиболее релевантных с точки зрения медицинского сообщества;
- наличие общей методической информации, направленной на поддержку авторов;
- обеспечение возможности применения чек-листа разными заинтересованными специалистами (авторами, рецензентами, редакторами, критически настроенными читателями).

Такой подход позволил нам создать инструмент, принципиально отличающийся по качеству и возможностям от разработок других авторов. Наша разработка основана на фундаментальных требованиях к структуре и содержанию медицинской научной статьи — здесь есть взаимосвязь с принципами создания чек-листов **SPIRIT-AI** и **CONSORT-AI** [25, 26]. Однако наш инструмент дополнен информацией методического характера, которую часто упускают начинающие авторы научных статей. Это повышает образовательные и справочные возможности нашей разработки.

В изложении специфических аспектов машинного обучения и ИИ прослеживается некая преемственность с инструментами **MI-CLAIM** и **F. Schwendicke** и соавт. [28]. Однако в своей работе мы расширили перечень требований, дополнив чек-лист детальными характеристиками наборов данных, включая источники их формирования, особенности разметки (подготовки) и прочее.

В цитируемых чек-листах лишь констатируется потребность излагать применение ММО в клиническом контексте, но отсутствует детальная информация о том, как это необходимо делать. В своей разработке мы учли этот недостаток и соответствующим образом дополнили как содержательную часть методики, так и чек-лист.

В отличие от узкоспециализированных или слишком общих вариантов чек-листа [27—29] наша разработка может применяться авторами научных статей — для повышения компетенций и самоконтроля, рецензентами — для контроля качества структуры и содержания рукописей, редакторами — для принятия решений о публикации статьи. Крайне важно подчеркнуть методическую и образовательную составляющую. На фоне колоссального интереса к технологиям ИИ множество авторов впервые «пробуют перо», создавая рукописи для научных медицинских журналов. При этом сами авторы не принадлежат к сфере здравоохранения, не владеют навыками подготовки именно медицинских рукописей и не знают принци-

пов доказательной медицины. Из-за этого возникает значительное количество конфликтов и барьеров при публикации соответствующих работ. Наша разработка направлена на формирование у авторов, начинающих писать статьи для медицинских журналов, профильных, специфических компетенций и навыков.

Заключение

Для повышения качества научных публикаций и обеспечения их соответствия правилам доказательной медицины разработаны практико-методические рекомендации по подготовке рукописи о разработке и валидации моделей машинного обучения для здра-

воохранения. Для ускоренной проверки структуры, полноты и качества содержания рукописи разработан оригинальный чек-лист, который может применяться как авторами для самоконтроля, так и рецензентами, редакторами научных журналов для оценки качества и принятия решений о публикации. Практико-методические рекомендации и чек-лист отличаются балансом технологических и клинических аспектов, образовательной значимостью и комплексными возможностями. Предложенный инструментарий может быть валидирован независимыми исследователями.

Авторы заявляют об отсутствии конфликта интересов.
The authors declare no conflicts of interest.

ЛИТЕРАТУРА/REFERENCES

- Global Burden of Disease Health Financing Collaborator Network. Past, present, and future of global health financing: A review of development assistance, government, out-of-pocket, and other private spending on health for 195 countries, 1995-2050. *Lancet*. 2019;393(10187):2233-2260. [https://doi.org/10.1016/s0140-6736\(19\)30841-4](https://doi.org/10.1016/s0140-6736(19)30841-4)
- Cause of death, by non-communicable diseases (% of total). World Bank Open Data. Accessed June 16, 2022. <https://data.worldbank.org/indicator/SH.DTH.NCOM.ZS>
- Topol E, ed. *The Patient Will See You Now: The Future of Medicine Is in Your Hands*. New York: Basic Books; 2016.
- Topol E. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books; 2019.
- Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;(369):m1328. <https://doi.org/10.1136/bmj.m1328>
- Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*. 2021;3:199-217. <https://doi.org/10.1038/s42256-021-00307-0>
- Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nature Medicine*. 2020;26(9):1364-1374. <https://doi.org/10.1038/s41591-020-1034-x>
- Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nature Medicine*. 2020;26(9):1320-1324. <https://doi.org/10.1038/s41591-020-1041-y>
- Collins GS, Reitsma JB, Altman DG, Moons KG; TRIPOD Group. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. The TRIPOD Group. *Circulation*. 2015;131(2):211-219. <https://doi.org/10.1161/CIRCULATIONAHA.114.014508>
- Li D, Zhang Q, Tan Y, Feng X, Yue Y, Bai Y, et al. Prediction of COVID-19 Severity Using Chest Computed Tomography and Laboratory Measurements: Evaluation Using a Machine Learning Approach. *JMIR Medical Informatics*. 2020; 8(11):e21604. <https://doi.org/10.2196/21604>
- Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. *Journal of Medical Internet Research*. 2020;22(11):e24018. <https://doi.org/10.2196/24018>
- Thakkar HK, Liao WW, Wu C-Y, Hsieh Y-W, Lee T-H. Predicting Clinically Significant Motor Function Improvement after Contemporary Task-Oriented Interventions Using Machine Learning Approaches. *Journal of Neuroengineering and Rehabilitation*. 2020; 17(1):131. <https://doi.org/10.1186/s12984-020-00758-3>
- Kalafi EY, Nor NAM, Taib NA, Ganggayah MD, Town C, Dhillon SK. Machine Learning and Deep Learning Approaches in Breast Cancer Survival Prediction Using Clinical Data. *Folia Biologica*. 2019;65:212-220.
- Korevaar DA, Gopalakrishna G, Cohen JF, Bossuyt PM. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses. *Diagnostic and Prognostic Research*. 2019;3:22. <https://doi.org/10.1186/s41512-019-0069-2>
- Морозов С.П., Владимировский А.В., Ледихова Н.В., Соколина И.А., Кульберг Н.С., Гомболевский В.А. Оценка диагностической точности системы скрининга туберкулеза легких на основе искусственного интеллекта. *Туберкулез и болезни легких*. 2018;96(8):42-49. Morozov SP, Vladimirovskiy AV, Ledakova NV, Sokolina IA, Kulberg NS, Gombolevskiy VA. Evaluation of diagnostic accuracy of the system for pulmonary tuberculosis screening based on artificial neural networks. *Tuberkulez i bolezni legkikh*. 2018;96(8):42-49. (In Russ.). <https://doi.org/10.21292/2075-1230-2018-96-8-42-49>
- Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circulation Research*. 2017;121(9):1092-1101. <https://doi.org/10.1161/CIRCRESAHA.117.311312>
- Okada Y, Kiguchi T, Irisawa T, Yamada T, Yoshiya K, Park C, et al. Development and Validation of a Clinical Score to Predict Neurological Outcomes in Patients with Out-of-Hospital Cardiac Arrest Treated with Extracorporeal Cardiopulmonary Resuscitation. *JAMA Network Open*. 2020;3(11):e2022920. <https://doi.org/10.1001/jamanetworkopen.2020.22920>

18. Павлов Н.А., Андрейченко А.Е., Владимирский А.В., Ревязин А.А., Кирпичев Ю.С., Морозов С.П. Эталонные медицинские датасеты (MosMedData) для независимой внешней оценки алгоритмов на основе искусственного интеллекта в диагностике. *Digital Diagnostics*. 2021;2(1):49–66. Pavlov NA, Andreychenko AE, Vladzimirskyy AV, Revazyan AA, Kirpichev YuS, Morozov SP. Reference medical datasets (MosMedData) for independent external evaluation of algorithms based on artificial intelligence in diagnostics. *Digital Diagnostics*. 2021;2(1):49–66. (In Russ.). <https://doi.org/10.17816/DD60635>
19. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular Disease Risk Prediction Using Automated Machine Learning: A Prospective Study of 423,604 UK Biobank Participants. *PLoS One*. 2019;14(5):e0213653. <https://doi.org/10.1371/journal.pone.0213653>
20. Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, et al. Predicting Clinical Outcomes of Large Vessel Occlusion Before Mechanical Thrombectomy Using Machine Learning. *Stroke*. 2019;50(9):2379–2388. <https://doi.org/10.1161/STROKEAHA.119.025411>
21. Морозов С.П., Владимирский А.В., Гомболевский В.А., Кляшторный В.Г., Федулова И.А., Власенков Л.А. Искусственный интеллект в скрининге рака легкого: оценка диагностической точности алгоритма для анализа низкодозовых компьютерных томографий. *Туберкулез и болезни легких*. 2020;98(8):24–31. Morozov SP, Vladzimirskiy AV, Gomboleviskiy VA, Klyashtorny VG, Fedulova IA, Vlasenkov LA. Artificial intelligence in lung cancer screening: assessment of the diagnostic accuracy of the algorithm analyzing low-dose computed tomography. *Tuberkulez i bolezni legkikh*. 2020;98(8):24–31. (In Russ.). <https://doi.org/10.21292/2075-1230-2020-98-8-24-31>
22. Morozov SP, Andreychenko AE, Pavlov NA, Vladzimirskyy AV, Ledikhova NV, Gomboleviskiy VA, et al. 2020. MosMedData: Chest CT Scans with COVID-19 Related Findings Dataset. arXiv preprint arXiv:2005.06465. <https://doi.org/10.1101/2020.05.20.20100362>
23. Ko H, Chung H, Kang WS, Park C, Kim DW, Kim SE, et al. An Artificial Intelligence Model to Predict the Mortality of COVID-19 Patients at Hospital Admission Time Using Routine Blood Samples: Development and Validation of an Ensemble Model. *Journal of Medical Internet Research*. 2020;22(12): e25442. <https://doi.org/10.2196/25442>
24. Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies. *International Journal of Medical Informatics*. 2021;153:104510. <https://doi.org/10.1016/j.ijmedinf.2021.104510>
25. Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ; on behalf of the SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial reports for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ*. 2020;370:m3210. <https://doi.org/10.1136/bmj.m3210>
26. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; on behalf of the SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ*. 2020;370:m3164. <https://doi.org/10.1136/bmj.m3164>
27. Kwong JCC, McLoughlin LC, Haider M, Goldenberg MG, Erdman L, Rickard M, et al. Standardized Reporting of Machine Learning Applications in Urology: The STREAM-URO Framework. *European Urology Focus*. 2021;7(4):672–682. <https://doi.org/10.1016/j.euf.2021.07.004>
28. Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, et al.; IADR e-oral health network and the ITU WHO focus group AI for Health. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *Journal of Dentistry*. 2021;107:103610. <https://doi.org/10.1016/j.jdent.2021.103610>
29. Sengupta PP, Shrestha S, Berthon B, Messas E, Donal E, Tison GH, et al. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): A Checklist: Reviewed by the American College of Cardiology Healthcare Innovation Council. *JACC Cardiovascular Imaging*. 2020;13(9):2017–2035. <https://doi.org/10.1016/j.jcmg.2020.07.015>

Поступила 22.02.2022

Received 22.02.2022

Принята к печати 25.04.2022

Accepted 25.04.2022