

ISSN 2782-3806
ISSN 2782-3814 (Online)
УДК 004.8:61

ПРИМЕНЕНИЕ BIG DATA В ЛАБОРАТОРНОЙ МЕДИЦИНЕ

Овчинникова М. А.¹, Жиленкова Ю. И.^{1,2}, Черныш Н. Ю.^{1,3}

¹ Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр имени В. А. Алмазова» Министерства здравоохранения Российской Федерации, Санкт-Петербург, Россия

² Федеральное государственное бюджетное образовательное учреждение высшего образования «Северо-Западный государственный медицинский университет имени И. И. Мечникова» Министерства здравоохранения Российской Федерации, Санкт-Петербург, Россия

³ Федеральное государственное бюджетное учреждение «Российский научно-исследовательский институт гематологии и трансфузиологии Федерального медико-биологического агентства», Санкт-Петербург, Россия

Контактная информация:

Овчинникова Марина Александровна,
ФГБУ «НМИЦ им. В. А. Алмазова»
Минздрава России,
ул. Акkuratова, д. 2, Санкт-Петербург,
Россия, 197341.
E-mail: m.a.ovchinnikova@inbox.ru

Статья поступила в редакцию 10.05.2023
и принята к печати 31.05.2023.

РЕЗЮМЕ

Термин «большие данные» (Big Data) относится к массивам данных, объем которых превышает возможности традиционных баз данных в хранении, управлении и анализе информации. Возникновению алгоритмов больших данных способствовало внедрение мощных серверов, которые стали обладать достаточными ресурсами для обработки информации и компьютерных вычислений с целью применения больших данных для статистической обработки, анализа, прогнозирования и принятия решений. В лабораторной практике при наличии большого количества цифровой информации практическое использование Big Data в настоящее время не имеет большого распространения. Цель данной работы — провести ретроспективный обзор литературных источников по применению Big Data в сфере лабораторной медицины в период 2018–2023 гг. и проанализировать имеющиеся практические разработки, преимущества и достижения, связанные с аналитикой больших данных в сфере лабораторной медицины; проблемы, возможности и перспективы применения.

Ключевые слова: анализ больших данных, информационные системы, лабораторная медицина.

Для цитирования: Овчинникова М.А., Жиленкова Ю.И., Черныш Н.Ю. Применение Big Data в лабораторной медицине. Российский журнал персонализированной медицины. 2023; 3(4):77-87. DOI: 10.18705/2782-3806-2023-3-4-77-87.

APPLICATION OF BIG DATA IN LABORATORY MEDICINE

Ovchinnikova M. A.¹, Zhilenkova Yu. I.^{1,2}, Chernysh N. Yu.^{1,3}

¹ Almazov National Medical Research Centre, Saint Petersburg, Russia

² North-Western State Medical University named after I. I. Mechnikov, Saint Petersburg, Russia

³ Russian Research Institute of Hematology and Transfusiology of the Federal Medical and Biological Agency, Saint Petersburg, Russia

Corresponding author:

Ovchinnikova Marina A.,
Almazov National Medical Research Centre,
Akkuratova str., 2, Saint Petersburg, Russia,
197341.
E-mail: m.a.ovchinnikova@inbox.ru

Received 10 May 2023; accepted 31 May
2023.

ABSTRACT

The term “big data” (Big Data) refers to data sets, covering the excessive difference in differences between databases in the storage, management and analysis of information. The emergence of big data application algorithms has become the consumption of resources that use resource resources for information processing and computer calculations for the purpose of big data for statistical processing, analysis, forecasting and decision making. In laboratory practice, with a large amount of practical digital information, the use of big data is not currently widespread. The purpose of this work is to conduct a retrospective review of the literature on the use of big data in the field of laboratory medicine in the period 2018–2023. and evaluating the results of practical developments, benefits and achievements associated with big data analytics in the field of laboratory.

Key words: big data analysis, information systems, laboratory medicine.

For citation: Ovchinnikova MA, Zhilenkova YuI, Chernysh NYu. Application of Big Data in laboratory medicine. Russian Journal for Personalized Medicine. 2023; 3(4):77-87. (In Russ.) DOI: 10.18705/2782-3806-2023-3-4-77-87.

Список сокращений: АБД — аналитика больших данных, БД — базы данных, ИИ — искусственный интеллект, ЛИС — лабораторная информационная система, МИС — медицинская информационная система, МО — машинное обучение, СППВР — система поддержки принятия врачебных решений, СППКР — система поддержки принятия клинических решений, ЭМК — электронная медицинская карта, ФМ — фибрин-мономер.

ВВЕДЕНИЕ

Лабораторная медицина является медицинской дисциплиной с наивысшей степенью цифровизации. Все более распространенными становятся автоматизация, электронная передача результатов и электронная отчетность [1]. Медицинские лаборатории формируют цифровые базы данных результатов лабораторных исследований. Таким образом лабораторная медицина вносит большой вклад в концепцию цифровизации медицины. Однако разработки в области науки о больших данных, машинного обучения (МО), искусственного интеллекта (ИИ) все еще не достигли глобального распространения в сфере лабораторной медицины в настоящее время. Для расширения возможностей лабораторной медицины необходимо развивать и обеспечивать три ключевых направления: алгоритмы машинного обучения, необходимую вычислительную мощность для запуска указанных алгоритмов и большие объемы данных [2]. Определяющим вектором являются высококачественные надежные данные, на основании которых алгоритмы смогут сделать надежные выводы.

Понятие Big Data впервые было озвучено в 2008 году в выпуске журнала Nature, в статье главного редактора Клиффорда Линча [3]. В этом номере издания центральной темой был взрывной рост мировых объемов данных и их значимость в науке. Отмечается, что корректность сбора информации, ее доступность, алгоритмы машинной обработки и методы анализа, наряду с практическим применением возможностей науки о данных (data science) и технологий искусственного интеллекта (ИИ), представляются ведущими вызовами современности. Требуется модернизация всех подходов к использованию, аналитике и хранению этих данных в связи с объемами генерируемых и потенциально применимых данных в здравоохранении [4]. В настоящее время количество медицинских данных лавинообразно увеличилось в связи с переходом здравоохранения в цифровой формат — большая часть мировых медицинских учреждений использует различные цифровые сервисы, в первую оче-

редь — медицинские информационные системы (МИС) и хранимые в них электронные медицинские карты (ЭМК). ЭМК содержат в себе большой объем данных о пациенте: демографических, лабораторных, инструментальных, диагнозов и проводимой терапии. Персональные данные пациента накапливаются, следующим этапом осуществляется обезличивание и использование их для улучшения качества оказываемой помощи путем применения современных вычислительных инструментов. Стоит также отметить, что наиболее информативными и точными поставщиками медицинских данных являются результаты лабораторных тестов [5].

Задействование методов аналитики больших данных (Big Data), алгоритмов машинного обучения и искусственного интеллекта актуально в современной лабораторной диагностике также и в экономическом аспекте: проводится поиск решений, которые позволили бы эффективнее использовать ресурсы клиничко-диагностических лабораторий с целью повышения качества оказываемой медицинской помощи [6].

ЦЕЛЬ РАБОТЫ

Провести ретроспективный обзор отечественных и зарубежных литературных источников по применению Big Data в сфере лабораторной медицины в период 2018–2023 гг. Проанализировать имеющиеся практические разработки, преимущества и достижения, связанные с аналитикой больших данных в сфере лабораторной медицины; проблемы, возможности и перспективы применения big data для задач в рамках расширенной аналитики, включая искусственный интеллект, в рутинной клинической практике.

МАТЕРИАЛЫ И МЕТОДЫ

Поиск актуальной литературы осуществлялся в базах данных PubMed/Medline и научной электронной библиотеке E-library в период с 2018 по 2023 годы посредством формирования запросов по группам ключевых сочетаний в комбинации с логическими операторами: «big data in healthcare», «big data AND laboratory medicine», «AND machine learning», «AND clinical chemistry», «большие данные в здравоохранении» соответственно. По указанным запросам найдено суммарно 563 обзорных статьи. После изучения их аннотаций было отобрано 30 публикаций, соответствующих заданным целям обзора. Далее выполнялся качественный анализ текста каждой статьи и включение в обзор. Также было исключено 9 полнотекстовых статей

в связи с дублированием темы. В обзор была включена 21 полнотекстовая статья.

Общие понятия аналитики больших данных.

Наука о больших данных работает с информацией, трудно анализируемой традиционным программным обеспечением. Аналитика больших данных (АБД) включает в себя интеграцию разнообразных данных, контроль их качества и анализ, моделирование и интерпретацию, валидацию. Традиционно выделяют четыре основных типа аналитики: описательную, диагностическую, прогнозную и предписывающую [7].

1. Описательная аналитика. Включает в себя изучение прошлых данных, чтобы понять, что произошло. Обобщает исторические данные и дает представление о шаблонах, тенденциях и ключевых показателях.

2. Диагностическая аналитика. Выходит за рамки описательной аналитики, стремясь понять, почему что-то произошло. Это включает в себя углубление в данные, чтобы определить основные причины конкретного результата.

3. Предиктивная аналитика. Включает использование исторических данных и статистических методов для прогнозирования будущих событий или результатов. Использует закономерности и взаимосвязи, обнаруженные в данных, для оценки того, что может произойти в будущем.

4. Предписывающая аналитика. Расширяет возможности анализа данных, рекомендуя конкретные действия для оптимизации результатов на основе прогнозных моделей. Использует передовые алгоритмы, методы оптимизации и моделирование для получения практических идей.

Аналитика больших данных в медицине дает возможность анализировать большие массивы данных от пациентов, обнаруживать корреляции между массивами данных, вести разработку прогностических моделей с применением методов искусственного интеллекта. Посредством АБД в сфере медицины происходит объединение компетенции научных областей биоинформатики, медицинской визуализации, сенсорной информатики, медицинской информатики и информатизации системы здравоохранения.

АБД характеризуется рядом свойств, представленных в литературе в виде 3Vs [8]:

- Volume — объем — количество данных, генерируемое ежесекундно;
- Variety — разнообразие — различные типы генерируемых, накапливаемых и используемых данных;
- Velocity — скорость производства данных, увеличивающаяся с течением времени.

Позже были добавлены 3V свойства:

- Veracity — правдивость, неопределенность данных;
- Value — ценность — технологии аналитики больших данных превращают их в полезную информацию;
- Variability — вариабельность — различия данных по одной и той же теме, являющиеся большой сложностью для обработки, связанные форматом, методикой получения (единицы измерения, расчета).

С течением времени добавились еще 2V:

- Visualization — графическая и схематическая визуализация данных;
- Virality — жизнеспособность данных — наличие смысла и возможности анализа.

Относительно недавно было введено дополнительное свойство аналитики больших данных [9]:

- Complexity — сложность анализа данных — она тем выше, чем больше размер массива данных, их разнородность и неопределенность.

Медицинские данные зачастую представляются в разнообразных форматах. Также данные могут быть непрерывными или дискретными, неструктурированными, полуструктурированными или структурированными [10].

Актуальным вектором развития аналитики больших данных в медицинской сфере является производство и анализ омиксных данных — больших массивов данных, производимых на всех уровнях организации биологических процессов [11]. Термин исходит от английского Omics и представляет перечень биологических дисциплин, которым присущ общий подход к приемам изучения биологических процессов на разных уровнях организации (геномика, протеомика, транскриптомика, метаболомика). Омиксные данные представляют собой совокупность всех аналитов определенного уровня. В основе омиксных технологий лежит использование высокотехнологичных методов анализа, что приводит к накоплению больших массивов данных, которые анализируют специалисты data scientists. Это способствовало возникновению новой дисциплины — биоинформатики — на границе биологии, информатики и статистики. Применение алгоритмов машинного обучения с целью решения биологических задач стало следующим шагом и продолжает стремительно развиваться по сей день.

Возможные области применения ИИ в лабораторной медицине. Технологии ИИ могут быть внедрены в рутинную практику производственной деятельности лабораторий с целью оптимизации кодирования услуг и счетов, оценки степени загрузки

женности оборудования и персонала с предсказанием вероятных потребностей и штатного расписания, мониторинга обеспеченности реагентами и расходными материалами, автоматического считывания направлений и выдачи ответов, распознавания записей голосовых заключений.

Успешное развитие персонализированной медицины подразумевает применение технологий ИИ в научно-исследовательском направлении для анализа непрерывно увеличивающегося объема данных, производимых путем секвенирования, масс-спектрометрии и иных методов геномики, транскриптомики, протеомики, метаболомики. Продукт Watson for Genomics (WFG), разработка компании IBM, способен менее чем за 3 минуты проанализировать ряд мутаций пациента, выявленных ранее секвенированием экзонов. WFG предоставляет отчет с информацией об одобренных методах лечения, препаратах, доступных клинических испытаниях. В ходе тестирования продукта у 323 из 1 018 (31,7 %) пациентов он выявил дополнительные геномные события, имеющие потенциальное значение, со ссылками на существующие таргетные противоопухолевые препараты, протоколы и актуальные клинические исследования.

Во время пандемии COVID-19 компания WellAI разработала интеллектуальное диагностическое решение WellAI.COVID-19 для исследователей на базе нейронных сетей в виде инструмента поиска публикаций по теме COVID-19 с одновременным анализом. При формировании запроса COVID-19 и интересующего понятия отображается список концепций, ранжированных по вероятности, что дает возможность фокусироваться на исследовании взаимосвязи между запрашиваемыми понятиями [12].

Большие данные, генерируемые лабораторными тестами, методы машинного обучения и ИИ могут предсказывать новые тесты, помогая тем самым поиску новых диагнозов. Это позволяет обнаруживать скрытую информацию и предлагать дополнительные тесты, уменьшая количество ложноотрицательных результатов и помогая в раннем обнаружении ранее неизвестных заболеваний.

В своем исследовании Ma C. и соавторы представляют аспекты использования больших данных с упором на лабораторную медицину [13]. В нем было подчеркнута отсутствие стандартизации в клинических лабораториях и трудности с использованием данных в режиме реального времени, в основном из-за неструктурированных и ненадежных данных. Тем не менее, подчеркивается потенциал использования лабораторных данных вместе с такими аспектами, как установление ре-

ферентного диапазона, контроль качества, анализ факторов, влияющих на результаты значений анализов, создание диагностических и прогностических моделей, эпидемиологическое исследование, управление лабораторией и интеллектуальный анализ данных.

Клинические лаборатории отображают большинство результатов тестов в виде индивидуальных числовых значений. Однако результаты этих тестов, рассматриваемые изолированно, обычно имеют ограниченное значение для постановки диагноза. Чрезмерное использование лабораторий является обычным явлением. Необоснованное тестирование расходует ресурсы лабораторий и может нанести вред пациентам, способствуя развитию ятрогенной анемии. Ряд исследователей высказывают мысль о том, что путем применения алгоритмов ИИ в лабораторной медицине будет возможно свести к минимуму использование лабораторий при получении необходимой информации.

Так, Rachael C. применил методы машинного обучения для анализа базы данных госпитализированных пациентов Медицинского центра Стенфордского университета в период 2008–2014 гг. Модель включала основные клинические данные ЭМК, в том числе результаты 13 лабораторных тестов. Было установлено, что большая часть результатов повторных лабораторных тестов (проведенных в течение 14 дней после предыдущего назначения) значимо не отличались ($\pm 10\%$ или $\pm 0,1$ SD) от полученных ранее результатов, что дало возможность выделить ряд тестов, частоту выполнения которых можно сократить [14]. Приведенный подход актуален в тех ситуациях, когда проведение некоторых лабораторных тестов может быть ограничено финансовой составляющей, обеспеченностью реагентной и технической базой.

В своем исследовании ферритина Luo и коллеги обнаружили, что лабораторные тесты часто содержат избыточную информацию [15].

Так же Gunçar G. и соавторы выявили, что модели машинного обучения могут предсказывать гематологические заболевания, используя только клинический анализ крови. Был сформулирован вывод, что лабораторные тесты содержат больше информации, чем обычно считают медицинские работники [15].

На базе ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России В. С. Власов и коллеги посредством применения методов машинного обучения на данных скрининговых клоттинговых тестов (тромбиновое время), концентрации D-димера и срока гестации разработали прогностическую модель превышения верхней границы референтно-

го интервала для концентрации фибрин-мономера (ФМ) у беременных. При этом, несмотря на весьма ограниченную клиническую ценность применения тестов изолированного измерения ТВ и концентрации D-димера в период беременности, в составе модели выявили новые данные, а также обозначили целесообразность включения их в перечень лабораторных тестов, проводимых при беременности. При имеющейся малой доступности рутинного измерения концентрации ФМ иммунологическим методом, а также их патофизиологической значимости, разработанная модель может быть значима в контексте клинического заключения об активации системы свертывания крови в текущий момент времени и возникновения настороженности в отношении развития осложнений. Использование модели и представленного алгоритма может быть существенно упрощено при их интеграции в ЛИС. Следует отметить, что результаты разработки данной модели открывают перспективы создания подобных моделей и в иных областях лабораторной диагностики, что может способствовать реализации автоматизированного и персонализированного подхода к назначению исследований [6].

Пандемия SARS-CoV-2 и глобальная сложность ее разрешения создали стимул медицинскому сообществу для поиска инновационных подходов к разработке идентификационных тестов. Yang и соавторы оценили 3 356 подходящих пациентов [16]. Была разработана модель машинного обучения, включающая демографические характеристики пациентов (возраст, пол, раса) с 27 рутинными лабораторными тестами для прогнозирования инфекции SARS-CoV-2. Эта модель, использующая результаты рутинных лабораторных анализов, предлагает возможности для раннего и быстрого выявления пациентов, инфицированных SARS-CoV-2 с высоким риском, до того, как будут доступны результаты их ОТ-ПЦР. Это может сыграть важную роль в выявлении пациентов, инфицированных SARS-CoV-2, в районах, где тестирование ОТ-ПЦР недоступно из-за финансовых ограничений.

В медицинских лабораториях программное обеспечение на основе МО может выполнять анализ изображений. Так, анализаторы IRIS проводят автоматическую микроскопию мочи, идентифицируют объекты посредством анализа микрофотографий алгоритмом ИО, обученным на библиотеке изображений элементов мочи. Также уже в рутинной практике нашли применение программы, улучшающие информативность ручной микроскопии с помощью цифрового анализа морфологии клеток крови. Например, системы компаний Cella Vision, West Medica предварительно осуществляют клас-

сификацию клеточных элементов по подтипам в зависимости от характерных морфологических характеристик. Основной проблемой на пути повсеместного внедрения упомянутых цифровых систем является финансовый фактор. Данного недостатка лишено автоматическое устройство для микроскопии мазков крови российского разработчика Celly.AI [17]. Подсчет лейкоцитарной формулы с применением Celly.AI характеризуется высоким качеством, сопоставимым с результатами исследования лабораторными специалистами. Установлена высокая корреляция подсчета лейкоцитарной формулы с использованием автоматизированного подхода Celly с ручным методом для всех клеток, за исключением базофилов. Низкая стоимость системы Celly позволит внедрять автоматизацию в лаборатории с ограниченным бюджетом.

Лабораторная медицина как площадка для разработки моделей ИИ в РОСТ (Point of Care testing). При наличии огромного количества хорошо структурированных данных о пациентах в сочетании с возросшей вычислительной мощностью, лабораторная медицина является идеальной площадкой для разработки моделей ИИ. Однако в настоящее время только меньшинство систем, одобренных FDA, включают лабораторную диагностику [18]. Наиболее наглядно технологии ИИ реализуются в сфере РОСТ (Point of Care testing), что обусловлено разработкой биосенсоров, портативных устройств мониторинга, мобильными приложениями для пациентов. К примеру, система мониторинга глюкозы Guardian Connect (Medtronic) рассчитана на пациентов с инсулинозависимым сахарным диабетом. Устройство непрерывно измеряет уровень глюкозы в межклеточной жидкости под кожей с помощью небольшого сенсора. Пациентам доступны на своем смартфоне уровень глюкозы и динамика, предупреждения о текущей гипер/гипогликемии, включая прогнозы за 10–60 минут до события, возможность телемедицинской консультации с врачом для коррекции терапии. Или система ACR LAB Urine Analysis Test System Evaluation of Performance (Healthy.io 07.2019) — программа предназначена для качественного и полуколичественного определения в моче белка, глюкозы, нитритов и эритроцитов, а также альбумина, креатинина и отношения альбумина к креатинину с помощью колориметрического анализа мочевых полосок камерой смартфона посредством приложения, использующего алгоритм распознавания изображений. Или AI-система CLEW (CLEW 01.2021) — система поддержки принятия врачебных решений (СППВР) в интенсивной терапии на основе искусственного интеллекта. Система со-

бирает данные из электронной медицинской карты (ЭМК) и подключенных устройств интенсивной терапии, а также использует модели машинного обучения для расчета риска декомпенсации жизненно важных функций организма.

Большие данные как основа для создания моделей СППВР. Основной задачей любой медицинской лаборатории является измерение определенных маркеров в образце пациента, результаты которого в совокупности с диагностическими и клиническими данными возможно преобразовать в информацию, являющуюся основанием для принятия клинического решения. Стремительное развитие лабораторной медицины и возрастающий лавинообразно объем данных, нуждающийся в обработке, вынудили специалистов лаборатории фокусировать внимание только на пациентах с результатами, сильно отклоняющимися от границ установленного референтного интервала. В связи с этим, клиницисты сталкиваются со сложностью выбора и интерпретации тестов, что, в свою очередь, приводит к чрезмерному или недостаточному использованию лабораторных ресурсов [19].

Одним из способов решения представленной проблемы является разработка и применение лабораторных диагностических алгоритмов, представляющих собой системы экспертных правил, внедренные в ЛИС. Данная концепция известна с 1984 года и часто называется компьютеризированной системой поддержки принятия клинических решений (СППКР).

Стоит отметить, что смена вектора развития от хранения данных и применения минимальных аналитических функций к разработке предсказательных моделей и СППКР осуществляется крайне медленно.

Система IBM Watson представляет собой первый крупный мировой проект СППВР с использованием технологии ИИ. Однако проект был воспринят в научном мире как неудачный, несмотря на огромные инвестиции и широкую известность. В связи с чем в 2021 году компания IBM объявила о продаже проекта. Стремясь создать идеальную цифровую модель доктора, компания разработала специализированный продукт, в частности Watson for Oncology (WFO), который путем анализа информации от онкологического пациента в совокупности с генетическими исследованиями представлял варианты терапии и противопоказания. Данные метаанализа ряда исследований [20], включивших 2 463 пациентов, говорят о том, что между междисциплинарными командами врачей и WFO коэффициент соответствия решений составил 81,52 % и зависел от стадии рака, региона.

В России первой системой, прошедшей весной 2020 года независимые клинические и технические проверки (регистрационное удостоверение 2020/9958 Росздравнадзор) как программное медицинское изделие, стала платформа аналитики прогноза и управления рисками в сфере медицины на основе ИИ — Webiomed. Также в 2020 году регистрационных удостоверений были удостоены еще 2 продукта в сфере анализа медицинских изображений: платформа анализа и обработки медицинских изображений с применением технологий ИИ Botkin.Ai, платформа лучевой диагностики на базе технологий ИИ Care Mentor Ai [12].

СППВР Webiomed.DHRA представляет собой часть платформы Webiomed [5]. Цель развития Webiomed.DHRA состоит в улучшении возможностей по предсказанию подозрений на заболевания. Для создания модели была использована информация базы данных (БД) платформы прогнозной аналитики Webiomed, содержащая обезличенные данные ЭМК 2,9 млн пациентов, которые проходили обследование в медицинских учреждениях некоторых регионов Российской Федерации. Результат проведенного исследования был представлен моделью выявления подозрений на заболевания на основе данных результатов лабораторных исследований, представленных в ЭМК. Критериями отбора информации о пациентах были возраст старше 18 лет и наличие исследования клинического или биохимического анализа крови. В тестовой выборке демографические и лабораторные данные пациентов применялись как входные параметры для алгоритмов МО. Тестируемая выборка данных была случайно разбита на две — обучающую и тестовую выборки. При создании модели выявления подозрений на заболевания использовались инструменты машинного обучения — алгоритмы классификации: логистическая регрессия, «наивный» байесовский классификатор, дерево принятия решений, метод случайного леса, градиентный бустинг, адаптивный бустинг. В качестве показателя качества работы модели определена точность. Данный параметр рассчитывался как отношение количества записей, по которым классификатор принял правильное решение, к размеру всей выборки. Однако в результате проведенных экспериментов был обнаружен существенный недостаток — невысокая точность, не позволяющая на данный момент применять модель на практике [5].

Таким образом, по-прежнему актуальной задачей остается поиск чувствительных биомаркеров и обнаружение корреляционных связей между ними, наличие которых позволит проводить до-

стоверную и точную оценку вероятности наличия заболевания у пациента.

Новая концепция клинлабдомики (Клинлабдомика) — объединение клинической лабораторной медицины и ИИ. В эру больших данных технологии ИИ могут использовать большие наборы клинических данных для принятия клинических решений, выявления предикторов скрытых заболеваний, а также для создания новых проверяемых гипотез.

В настоящее время уже существует новая концепция радиомики для данных визуализации в сочетании с ИИ, но отсутствует новое определение клинических лабораторных данных в сочетании с ИИ, в связи с чем многие исследования в этой области не могут быть точно классифицированы. Поэтому Wen и коллектив авторов предлагают новую концепцию клинлабдомики (Клинлабдомики), объединяющую клиническую лабораторную медицину и ИИ [21]. Клинлабдомика может использовать высокопроизводительные методы машинного обучения для извлечения больших объемов данных из данных клинических лабораторных исследований. Затем, используя статистическую обработку, машин-

ное обучение, можно извлечь ранее неизвестную информацию.

Примечание: Первый этап — сбор образцов крови или биологических жидкостей и их тестирование. Далее из полученных клинических лабораторных данных с помощью методов глубокого машинного обучения извлекаются признаки, к примеру, основанные на данных клинических испытаний здоровых индивидуумов или пациентов с различными заболеваниями. Эти признаки используются для анализа, например, оценки на предмет их диагностической прогностической силы или связи со стадией заболевания. В совокупности это прокладывает путь к прецизионной и персонализированной медицине.

Из данных рутинных анализов биологических жидкостей Клинлабдомика извлекает, анализирует, проверяет и идентифицирует определенные воспроизводимые и важные показатели клинических лабораторных тестов для пациентов с клинически значимыми заболеваниями. Затем анализируется взаимосвязь между выбранными маркерами и результатами диагностики и лечения. Цель состоит в том, чтобы обеспечить точную диагностику заболевания, стратификацию риска и прогноз с помо-



Рис. 1. Визуализация рабочего процесса Клинлабдомики. Адаптировано из Wen X., Leng P., Wang J., et al. Clinlabomics: leveraging clinical laboratory data by data mining strategies [21]

щью глубокого машинного обучения на большом количестве данных и разработки прогностических моделей для связанных заболеваний.

Прогресс в клинической лабораторной медицине был обусловлен развитием лабораторных технологий, что создало условия для создания Клинлабоники. За последнее десятилетие клиническая лабораторная медицина развивалась в четырех различных областях [21]. Помимо этого, изменилось качество лабораторных результатов. Для клинических лабораторий качество данных так же важно, как и их количество. Ранее не существовало глобального руководства по обеспечению качества. Пока не была принята международная гарантия качества стандарта ISO15189, регламентирующего процесс управления качеством во всех областях лабораторной медицины, Стандартизация методов клиничко-лабораторных исследований и единые стандарты контроля качества сделали результаты клиничко-лабораторных исследований разных клинических лабораторий более сопоставимыми.

Клинлабоника и клиническое прогнозирование вызывают широкий интерес. К примеру, предсказание биологического старения не теряет своей актуальности. Это подтверждает проведенное исследование о влиянии курения на хронологический возраст, так как в настоящее время не существует информативных тестов для оценки его вклада в ускорение биологического старения [22]. Была разработана математическая модель, включающая 149 000 полностью анонимных индивидуальных записей, с целью предсказать статус курения пациентов, она включала только их пол и данные клинического и биохимического анализов крови. Модель продемонстрировала, что курение ускоряет старение человека и что статус курения можно также предсказать по результатам биохимического и клинического анализов крови. И, хотя эта модель оказалась менее точной, чем предикторы, основанные на метилировании ДНК, она дешевле и практичнее и включает только стандартные анализы крови.

Клинлабоника может предсказывать многие заболевания, в том числе риск возникновения сахарного диабета. Как всем известно, диабет представляет собой глобальную эпидемию, хроническую и неизлечимую, а длительное воздействие гипергликемии может вызвать хроническое повреждение тканей. Раннее прогнозирование способно резко снизить риск возникновения диабета. Результат исследования Yang показал, что применение Клинилабоники может помочь группам высокого риска своевременно и разумно принимать лекарства или менять образ жизни, чтобы они могли эффективно снизить риск развития диабета и даже предотвратить его [23].

Клинлабоника оказалась чрезвычайно важна в диагностике заболеваний, в частности для выявления COVID-19. Domínguez-Olmedo разработал модель для прогнозирования смертности пациентов с COVID-19, которая может оценивать смертность по лабораторным значениям с высокой степенью точности [24].

Клинлабоника также в некоторой степени корректно осуществляет лабораторное управление, клинический лабораторный контроль качества и автоматическую интерпретацию результатов лабораторных анализов. Референтные интервалы имеют решающее значение для интерпретации лабораторных результатов, и Клинилабоника также может помочь установить их.

Таким образом, Клинилабоника представляет собой новую концепцию, направленную на сбор ценной информации, полученной в результате рутинных лабораторных исследований.

ЗАКЛЮЧЕНИЕ

Лабораторная диагностика производит огромное множество цифровых данных, тем самым вносит весомый вклад в формирование цифрового портфолио пациента. Для лаборатории важно быть как генератором данных, так и иметь стремление осваивать все возможности инструментов компьютерной обработки информации. Технологии искусственного интеллекта открывают широкие перспективы в таком анализе, в том числе могут осуществить цифровую трансформацию лечебно-диагностического процесса с целью повышения результативности и эффективности.

Выполнив ретроспективный обзор отечественных и зарубежных литературных источников по применению технологий больших данных в сфере лабораторной медицины, можно говорить о явных преимуществах программ на основе МО при выполнении анализа разнообразных медицинских данных ввиду более высокой скорости и точности по сравнению с традиционными способами. Это открывает перспективу ускорению научных исследований, поиску новых биомаркеров, установлению новых взаимосвязей и патологических механизмов. Результаты исследований наглядно демонстрируют большой потенциал ИИ, который следует в перспективе реализовать в персонализированной медицине для улучшения и ускорения качества диагностики, принятия оптимальных клинических решений и рационального использования ограниченных ресурсов здравоохранения. Рассмотренные в обзоре исследования зачастую имеют ретроспективный характер и формат ком-

пьютерного симулирования эксперимента. В связи с этим предстоит большая работа по проведению масштабных проспективных исследований и клинической валидации моделей.

Изложенные выше обстоятельства накладывают ответственность на лаборатории в части повышения компетенции, развития аналитических навыков для сотрудничества с IT-специалистами. Качество любой программы на основе МО определяется качеством данных. Роль лаборатории заключается в стандартизации и экспертной оценке производимых данных, участии в разработке протоколов и процессе разметки данных для МО, инициативной постановке задач.

Аналитика больших данных в сфере лабораторной медицины может открыть новый путь для персонализированной медицины в будущем. Потенциал Клинической лабораторной медицины, который применяет машинное обучение к лабораторным данным для диагностических и прогностических целей, заслуживает большего внимания со стороны клиницистов. В будущем создание соответствующих баз данных с помощью стандартизированных и стандартных характеристик данных клинических исследований в различных медицинских учреждениях будет предоставлять высококачественную помощь в точной диагностике и лечении, тем самым будет сделан конкретный шаг к реализации прецизионной медицины.

Конфликт интересов / Conflict of interest

Авторы заявили об отсутствии потенциального конфликта интересов. / The authors declare no conflict of interest.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Cadamuro, J. Rise of the Machines: The Inevitable Evolution of Medicine and Medical Laboratories Intertwining with Artificial Intelligence — A Narrative Review. *Diagnostics* 2021, 11, 1399. doi.org/10.3390/diagnostics11081399
2. Damien Gruson, Thibault Helleputte, Patrick Rousseau, et al. Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation, *Clinical Biochemistry*, Volume 69, 2019, Pages 1–7, ISSN 0009-9120. doi.org/10.1016/j.clinbiochem.2019.04.013
3. Lynch C. Big data: how do your data grow? // *Nature*. 2008. Vol. 455. № 7209. P. 28–29.
4. Шляхто Е.В., Конради А.О., Курапеев Д.И. Информация как важнейший инструмент развития персонализированной медицины. Как научиться ей управлять на благо пациента. *Наука о «больших данных»*. Российский журнал персонализированной медицины. 2022;2(6):6–15. DOI: 10.18705/2782-3806-2022-2-6-6-15.
5. Гусев А.В., Новицкий Р.Э., Ившин А.А. и др. Машинное обучение на лабораторных данных для прогнозирования заболеваний. *Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология*. 2021;14(4):581–592.
6. Власов В.С., Спельников Д.М., Вавилова Т.В. и др. Прогнозирование повышения концентрации фибрин-мономера при беременности с использованием машинного обучения. *Лабораторная служба*. 2022;11(4):31–38. doi.org/10.17116/labs20221104131
7. Карнаухов Н.С., Ильяхин Р.Г. Возможности технологий «Big Data» в медицине. *Врач и информационные технологии*. 2019, № 1. С. 59–63.
8. Hassan M, Awan F.M, Naz A, et al. Innovations in Genomics and Big Data Analytics for Personalized Medicine and Health Care: A Review. *Int. J. Mol. Sci.* 2022, 23, 4645. doi.org/10.3390/ijms23094645
9. Маркина Н.В., Касюк С.Т., Шамаева Т.Н. Анализ данных в медицинских информационных системах с использованием технологии Data Mining. *Информатика, вычислительная техника и управление. Серия «Естественные и технические науки»*. 2019. № 6. С. 111–116.
10. Магеррамов З.Т., Рагимова Н.А., Абдуллаев В.Г. и др. Технология big data: потенциал, проблемы и применение в медицине и здравоохранении. *RI*. 2020. С. 54–68.
11. Hassan M, Awan FM, Naz A, et al. Innovations in Genomics and Big Data Analytics for Personalized Medicine and Health Care: A Review. *Int. J. Mol. Sci.* 2022, 23, 4645. doi.org/10.3390/ijms23094645
12. Гусев А.В., Евгина С.А., Годков М.А. Искусственный интеллект в здравоохранении России. Роль лаборатории. *Лабораторная служба*. 2022;11(2):5–8. doi.org/10.17116/labs2022110215
13. Ma C, Wang X, Wu J, et al. Real-world big-data studies in laboratory medicine: current status, application, and future considerations. *Clin Biochem*. 2020 Oct; 84:21–30. DOI: 10.1016/j.clinbiochem.2020.06.014.
14. Aikens RC, Balasubramanian S, Chen JH. A Machine Learning Approach to Predicting the Stability of Inpatient Lab Test Results. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci*. 2019:515–523.
15. Cardozo G, Tirloni SF, Pereira Moro AR, et al. Use of Artificial Intelligence in the Search for New Information Through Routine Laboratory Tests: Systematic Review. *JMIR Bioinform Biotech*. 2022 Dec 23;3(1):e40473. DOI: 10.2196/40473.
16. Yang HS, Hou Y, Vasovic LV, et al. Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection

Using Machine Learning. *Clinical Chemistry*. 2020 Nov;66(11):1396–1404. DOI: 10.1093/clinchem/hvaa200.

17. Глазкова А.И., Лукьянова Е.А., Артемьева О.А. и др. Подсчет лейкоцитарной формулы при помощи машинного обучения с использованием смартфона. *Лабораторная служба*. 2022;11(2):53–55. doi.org/10.17116/labs20221102153

18. Ronzio L, Cabitza F, Barbaro, et al. Has the Flood Entered the Basement? A Systematic Literature Review about Machine Learning in Laboratory Medicine. *Diagnostics* 2021, 11, 372. doi.org/10.3390/diagnostics11020372

19. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims. *Ann. Intern. Med.* 2006; 145:488–496. DOI: 10.7326/0003-4819-145-7-200610030-00006.

20. Jie Z, Zhiying Z, Li L. A meta-analysis of Watson for Oncology in clinical application. *Sci Rep*. 2021;11(1). doi.org/10.1038/s41598-021-84973-5

21. Wen X, Leng P, Wang J, et al. Clinlabomics: leveraging clinical laboratory data by data mining strategies. *BMC Bioinformatics*. 2022 Sep 24;23(1):387. DOI: 10.1186/s12859-022-04926-1.

22. Mamoshina P, Kochetov K, Cortese F, et al. Blood biochemistry analysis to detect smoking status and quantify accelerated aging in smokers. *Sci Rep*. 2019; 9:142. DOI: 10.1038/s41598-018-35704-w.

23. Yang H, Luo Y, Ren X, et al. Risk prediction of diabetes: big data mining with fusion of multifarious physical examination indicators. *Inf Fusion*. 2021; 75:140–149. DOI: 10.1016/j.inffus.2021.02.015.

24. Domínguez-Olmedo JL, Gragera-Martínez Á, et al. Machine learning applied to clinical laboratory data in Spain for COVID-19 outcome prediction: model development and validation. *J Med Internet Res*. 2021; 23: e26211. DOI: 10.2196/26211.

Информация об авторах:

Овчинникова Марина Александровна, ординатор кафедры лабораторной медицины и генетики ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России;

Жиленкова Юлия Исмаиловна, к.м.н., доцент кафедры лабораторной медицины и генетики, ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России, ФГБОУ ВО СЗГМУ им. И. И. Мечникова Минздрава России;

Черныш Наталия Юрьевна, к.м.н., доцент кафедры лабораторной медицины и генетики, ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России, ФГБУ РосНИИ ГиТ ФМБА России.

Authors information:

Ovchinnikova Marina A., resident of the Department of Laboratory Medicine and Genetics, Almazov National Medical Research Centre;

Zhilenkova Yuliya I., PhD., Associate Professor, Department of Laboratory Medicine and Genetics, Almazov National Medical Research Centre, Northwestern State Medical University named after I. I. Mechnikov;

Chernysh Natalia Yu., PhD., Associate Professor, Department of Laboratory Medicine and Genetics, Almazov National Medical Research Centre, Russian Research Institute of Hematology and Transfusiology of the Federal Medical and Biological Agency.