

УДК 004.032.26

doi 10.54708/22259309_2025_13260

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И БОЛЬШИЕ ДАННЫЕ В ДИАГНОСТИКЕ РЕДКИХ ЗАБОЛЕВАНИЙ

К. С. ЛАБУТКИН¹

¹kostya222lb@mail.ru

¹ФГБОУ ВО «Уфимский университет науки и технологий» (УУНИТ)

Аннотация. Редкие заболевания, представляющие значительную проблему из-за их низкой распространенности и сложных диагностических путей, теперь могут быть диагностированы точнее и быстрее благодаря технологиям обработки больших данных и искусственного интеллекта. Эти инновационные подходы позволяют анализировать массивы информации и выявлять закономерности, которые ранее были незаметны, сокращая время до постановки диагноза и открывая путь к более эффективному лечению. В работе приведен анализ существующих решений и научных статей в сфере применения искусственного интеллекта и методов анализа больших данных в сфере диагностики и лечения редких заболеваний. Также рассматриваются перспективы развития применения такого подхода в различных направлениях исследований.

Ключевые слова: редкие заболевания; искусственный интеллект; большие данные; анализ генома; диагностика.

ВВЕДЕНИЕ

Редкие заболевания, поражающие сравнительно небольшое количество людей, представляют собой серьезный вызов для современной медицины. Их диагностика зачастую сопряжена с долгими годами поисков, скитанием по врачам и, к сожалению, нередкими ошибками. Причина тому – разнообразие симптомов, неспецифичность проявлений и, главное, банальная редкость самих болезней, которая затрудняет накопление необходимого опыта у врачей. Однако стремительное развитие технологий обработки больших данных и искусственного интеллекта (ИИ) открывает новую эру в диагностике редких заболеваний. Способность ИИ анализировать огромные массивы разнородной информации – от генетических данных и медицинских изображений до текстовых записей и результатов анализов – позволяет выявлять тонкие закономерности и сложные взаимосвязи, которые ускользают от человеческого внимания. Эти инновационные подходы не только сокращают время, необходимое для постановки диагноза, но и повышают его точность, открывая путь к более эффективному лечению и, в конечном итоге, улучшению качества жизни людей, страдающих от этих невидимых недугов. В данной статье мы рассмотрим, как именно ИИ и методы анализа больших данных трансформируют диагностику редких заболеваний и какие перспективы это открывает для будущего медицины.

РЕДКИЕ ЗАБОЛЕВАНИЯ

Редкие (орфанные) заболевания – это группа патологий, характеризующихся низкой распространенностью каждой конкретной болезни в общей популяции. Общее количество таких заболеваний достигает 7000 [1], при этом они проявляются множеством разных симптомов, варьирующихся даже между пациентами с одним и тем же диагнозом. Несмотря на индивидуальную редкость каждого заболевания, в сумме они затрагивают сотни тысяч людей по всему миру. Критерии отнесения заболеваний к редким различаются: в ЕС распространение считается редким, если оно не превышает 1 случая на 2000 человек, а в России – не более 10 случаев на 100 000 населения. Большая часть этих заболеваний являются генетическими.

Проблема редких заболеваний носит поистине глобальный характер, затрагивая миллионы жизней по всему миру. Около 30 миллионов человек в США и Европе живут с редкими заболеваниями, а общее количество страдающих от них по всему миру достигает примерно 350 миллионов [1]. Генетическая природа лежит в основе большинства (около 80 %) этих недугов, делая их спутниками человека на протяжении всей жизни. К сожалению, особо уязвимы дети: половина пациентов – это дети, и 30 % из них не доживают до пятилетнего возраста. Более того, редкие заболевания становятся причиной 35 % смертей в первый год жизни. Примечательно, что значительная часть пациентов (около 80 %) страдает от сравнительно небольшого числа этих болезней – около 350. Медицина все еще бессильна против большинства из них, ведь для 95 % редких заболеваний нет одобренного лекарственного лечения. За первые 25 лет действия закона о лекарствах для орфанных заболеваний было одобрено всего 326 новых препаратов. Около 6 % обращений в специализированный информационный центр связаны с недиагностированными случаями, а примерно половина редких болезней не имеет своей специфической поддержки или исследований. Эти факты подчеркивают серьезность проблемы редких заболеваний и необходимость дальнейших исследований и разработок для улучшения диагностики и лечения этих недугов.

Для пациентов с редкими заболеваниями первым и самым серьезным вызовом является даже осознание наличия болезни. От появления первых симптомов до постановки диагноза (так называемая диагностическая задержка) часто проходит от пяти до семи лет [2]. В течение этого времени пациенты проходят сложный путь, обращаясь к многочисленным врачам и сталкиваясь с серией диагностических процедур, прежде чем будет установлен точный диагноз. В половине случаев они посещают не менее пяти специалистов и получают хотя бы один ошибочный диагноз. В среднем пациентам с редкими заболеваниями ставят три неверных диагноза, что может привести к неправильному лечению и серьезным последствиям.

ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Лишь недавно учеными было объявлено о полном секвенировании генома человека. Революция в технологиях секвенирования генома привела к существенному прорыву в нашем понимании биологии и диагностики, особенно редких заболеваний. Благодаря появлению высокопроизводительных секвенаторов, таких как Illumina и PacBio, процесс получения геномных данных стал гораздо быстрее и доступнее. Теперь секвенирование целого генома можно выполнить за относительно короткое время и с меньшими затратами, что открывает новые перспективы для научных исследований и клинического применения. Эти технологические достижения позволили увеличить точность и длину прочтения ДНК, проводить секвенирование отдельных клеток и изучать эпигеном, что ранее было затруднено. Однако эти прорывы привели к появлению огромных массивов данных, требующих новых подходов к анализу. В этих условиях искусственный интеллект становится ключевым инструментом, предоставляя возможности для эффективной обработки и интерпретации геномных данных. ИИ, используя методы машинного и глубокого обучения, может выявлять патогенные мутации, предсказывать развитие болезней на основе генотипа, анализировать

геномные данные в сочетании с другими типами информации, открывать новые гены и пути, связанные с заболеваниями, и автоматизировать процессы анализа, освобождая ресурсы ученых.

Искусственный интеллект становится мощным инструментом, предлагающим ряд значительных преимуществ в диагностике.

Анализ сложных генетических данных. Секвенирование генома, используемое для выявления мутаций, создает огромные объемы данных. ИИ, в частности методы машинного обучения, способны быстро и эффективно обрабатывать эти массивы, выявляя даже самые незначительные изменения, которые могут быть связаны с редким заболеванием. Человек, анализируя такие данные, может упустить важные детали. Генетические заболевания, также называемые моногенными, возникают из-за нарушений в структуре генома. Примером может служить галактоземия, когда ферменты, преобразующие молочный сахар в глюкозу, работают недостаточно эффективно. Установлено, что дефектный ген, вызывающий заболевание, может передаваться от одного или обоих родителей, влияя на степень активности ферментов. Существуют решения, использующие нейронные сети для анализа ДНК для диагностики таких заболеваний, как болезнь Гентингтона [3] и гемофилия [4]. Также существуют исследования, связанные с применением алгоритмов машинного обучения на основе случайного леса для диагностики приоритетных генов моногенных заболеваний [5, 6].

Фенотипирование. Диагностика редких заболеваний зачастую представляет собой сложную задачу, поскольку симптомы могут быть разнообразными, неспецифичными и варьироваться от пациента к пациенту. В данном контексте фенотипирование, то есть подробное описание клинических проявлений заболевания и диагностика заболеваний на основе фенотипического описания, играет ключевую роль. Традиционный подход к фенотипированию требует значительных усилий со стороны врачей и может быть субъективным. Однако искусственный интеллект предлагает новые возможности для автоматизированного и объективного фенотипирования, что значительно ускоряет и улучшает диагностику редких заболеваний. Для диагностики заболеваний на основе фенотипа используются методы обработки изображений посредством сверточных нейронных сетей. Существуют решения, анализирующие изображения лица [7, 8] а также кожных покровов [9] для диагностики редких генетических заболеваний. Наиболее традиционный подход здесь – распознавание различных патологических признаков на снимках МРТ [10].

Искусственный интеллект становится мощным союзником в борьбе с редкими заболеваниями, особенно в области разработки новых лекарственных препаратов. Традиционный процесс создания лекарств – это долгий, дорогостоящий и трудоемкий путь, но ИИ обещает существенно ускорить его. ИИ-алгоритмы могут анализировать огромные объемы биологических данных, включая геномные, протеомные и метаболомные данные, для выявления потенциальных мишеней для лекарств и предсказания их эффективности и токсичности. ИИ может автоматизировать процессы скрининга молекул, моделировать взаимодействие лекарств с биологическими мишенями и даже разрабатывать новые молекулы с заданными свойствами [11–13]. Это позволяет существенно сократить время и затраты на доклинических и клинических этапах исследований, делая разработку лекарств для редких заболеваний более осуществимой и приближая момент, когда эффективная терапия станет реальностью для многих пациентов, страдающих от этих невидимых недугов.

Всё больше медицинских учреждений оснащается единой базой данных. Общая база данных позволяет экономить время медицинских специалистов и пациентов в постановке диагнозов [14]. Подобная тенденция способствует развитию внутренних систем медицинских учреждений и поиску эффективных подходов к анализу данных, позволяющих помочь докторам в постановке диагноза и подборе терапии. Это позволяет применять подход, основанный на анализе больших данных [15].

Интеграция генетических данных с клинической информацией. ИИ способен объединять генетические данные пациента с его клиническими проявлениями (симптомами, результатами

анализов, изображениями). Это позволяет более точно определить, является ли генетическая мутация причиной конкретного заболевания, и исключить другие возможные диагнозы. Исследователи активно изучают возможность объединения различных типов данных (генетических, клинических, изображений и т.д.) для создания более полных и точных фенотипических профилей с помощью ИИ. Рассматриваются подходы к использованию нейронных сетей для фенотипирования на основе объединения генетических данных и клинических проявлений.

NLP-алгоритмы могут извлекать информацию о симптомах и истории болезни из неструктурированных медицинских записей, что позволяет создавать комплексные фенотипические профили пациентов. Исследования демонстрируют, как использование глубокого обучения на текстовых данных из медицинских записей может помочь в диагностике редких заболеваний, ориентируясь на фенотипические проявления [16].

Клинические исследования играют важнейшую роль в разработке новых методов лечения редких заболеваний, однако одной из главных проблем является набор достаточного количества пациентов. Редкость самих заболеваний, их разнообразие и часто неспецифичность симптомов усложняют поиск и идентификацию потенциальных участников. Искусственный интеллект предлагает инновационные решения для преодоления этих трудностей [17]. ИИ-алгоритмы способны анализировать огромные массивы данных, включая электронные медицинские карты, генетические данные, результаты анализов и даже информацию из социальных сетей, для выявления пациентов с подозрением на редкие заболевания, которые ранее могли оставаться незамеченными. ИИ может автоматически сопоставлять клинические проявления с известными фенотипами редких заболеваний, а также выявлять новые, неописанные ранее синдромы. Это позволяет значительно ускорить процесс идентификации и рекрутинга пациентов для клинических исследований, что, в свою очередь, приближает момент, когда новые, эффективные методы лечения станут доступны для людей, страдающих от этих невидимых недугов.

Анализ генома является мощным инструментом в диагностике редких заболеваний, многие из которых имеют генетическую природу. Однако интерпретация геномных данных, особенно при поиске редких мутаций, является сложной и трудоемкой задачей. Именно здесь ИИ открывает новые горизонты, предлагая беспрецедентные возможности для ускорения и повышения точности диагностики.

ПЕРСПЕКТИВЫ РАЗВИТИЯ И ПРОБЛЕМЫ ПРИМЕНЕНИЯ ИИ В ПОЛНОГЕНОМНОМ СЕКВЕНИРОВАНИИ ДЛЯ ДИАГНОСТИКИ РЕДКИХ ЗАБОЛЕВАНИЙ

Полногеномное секвенирование (WGS) предоставляет беспрецедентный объем генетической информации, но его интерпретация для диагностики редких заболеваний остается сложной задачей. Внедрение искусственного интеллекта в этот процесс обещает революционизировать диагностику, предлагая новые возможности, но и ставит ряд серьезных проблем.

Персонализированная интерпретация генетических вариантов. ИИ может учитывать индивидуальные особенности каждого пациента, такие как его клиническая картина, семейная история и этническая принадлежность, чтобы более точно интерпретировать значение генетических вариантов. Это позволит избежать неправильной интерпретации генетических данных и повысит точность диагностики.

Интеграция мультиомиксных данных. В будущем ИИ сможет интегрировать данные WGS с другими омиксными данными (транскриптомика, протеомика, метаболомика) для создания более полной картины биологических процессов, связанных с редкими заболеваниями. Это поможет лучше понять патогенез заболеваний и разработать более эффективные методы лечения.

Ускорение разработки лекарств. ИИ может помочь в разработке новых методов лечения, основанных на данных WGS и позволяющих создавать персонализированные лекарственные препараты, нацеленные на конкретные генетические нарушения. Методы проектирования

белков с конкретными свойствами с помощью методов искусственного интеллекта были использованы в период пандемии COVID, что дает надежду на дальнейшее развитие направления.

ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ ИИ В ГЕНОМНОМ АНАЛИЗЕ ДЛЯ РЕДКИХ ЗАБОЛЕВАНИЙ

Идентификация новых генов, связанных с редкими заболеваниями. ИИ может помочь исследователям выявлять новые гены и генетические варианты, связанные с редкими заболеваниями. Анализируя геномные данные больших когорт пациентов, ИИ может обнаруживать ранее неизвестные связи между генами и заболеваниями.

ИИ может интегрировать геномные данные пациента с другой медицинской информацией (клинические записи, данные визуализации и т. д.) для создания более персонализированного диагностического заключения.

Анализ геномных данных с помощью ИИ может помочь исследователям лучше понять механизмы развития редких заболеваний. Это открывает перспективы для разработки новых методов лечения, направленных на устранение причин, а не только симптомов болезни.

Основным препятствием во внедрении подходов обработки больших данных в это направление медицины является, несомненно, малый набор оцифрованных данных, большая часть из которых либо не хранится в цифровом виде, либо не обработана для использования в алгоритмах.

ЗАКЛЮЧЕНИЕ

Использование ИИ в анализе генома для диагностики редких заболеваний является многообещающим направлением. ИИ не только ускоряет процесс анализа, но и позволяет делать более точные и полные заключения. Это открывает новые возможности для ранней и точной диагностики редких заболеваний, что является ключевым фактором для улучшения прогноза и качества жизни пациентов. Продолжающиеся исследования и разработки в этой области позволят раскрыть полный потенциал ИИ в геномике и продвинуть медицину на новый уровень.

СПИСОК ЛИТЕРАТУРЫ

1. **Nguengang Wakap S., Lambert D. M., Olry A. et al.** Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database // *Eur J Hum Genet.* 2020. 28. P. 165–173.
2. **James K. Stoller.** The Challenge of Rare Diseases// *CHEST.* 2018. Vol. 153. Issue 6. P. 1309–1314.
3. **Alfonso Perez G., Caballero Villarraso J.** Neural Network Aided Detection of Huntington Disease // *J Clin Med.* 2022. Apr 10; 11(8):2110.
4. **Lopes T. J. S., Rios R., Nogueira T. et al.** Prediction of hemophilia A severity using a small-input machine-learning framework // *npj Syst Biol Appl.* 2021. 7. 22.
5. **Mao D., Liu C., Wang L., Ai-Ouran R., Deisseroth C., Pasupuleti S., Kim S. Y., Li L., Rosenfeld J. A., Meng L., Burrage L. C., Wangler M. F., Yamamoto S.; Undiagnosed Diseases Network; Santana M., Perez V., Shukla P., Eng C. M., Lee B., Yuan B., Xia F., Bellen H. J., Liu P., Liu Z.** AI-MARRVEL – A Knowledge-Driven AI System for Diagnosing Mendelian Disorders // *NEJM AI.* 2024. May; 1(5):10.1056/aioa2300009.
6. **De La Vega F. M., Chowdhury S., Moore B., Frise E., McCarthy J., Hernandez E. J., Wong T., James K., Guidugli L., Agrawal P. B., Genetti C. A., Brownstein C. A., Beggs A. H., Löscher B. S., Franke A., Boone B., Levy S. E., Öunap K., Pajusalu S., Huentelman M., Ramsey K., Naymik M., Narayanan V., Veeraraghavan N., Billings P., Reese M. G., Yandell M., Kingsmore S. F.** Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases // *Genome Med.* 2021. Oct 14;13(1):153.
7. **Qiang J., Wu D., Du H., Zhu H., Chen S., Pan H.** Review on Facial-Recognition-Based Applications in Disease Diagnosis // *Bioengineering (Basel).* 2022. Jun 23;9(7):273.
8. **Hong D., Zheng Y. Y., Xin Y., Sun L., Yang H., Lin M. Y., Liu C., Li B. N., Zhang Z. W., Zhuang J., Qian M. Y., Wang S. S.** Genetic syndromes screening by facial recognition technology: VGG-16 screening model construction and evaluation // *Orphanet J Rare Dis.* 2021. Aug 3;16(1):344.
9. **Duong D., Waikel R. L., Hu P., Tekendo-Ngongang C., Solomon B. D.** Neural network classifiers for images of genetic conditions with cutaneous manifestations.

10. Goel A., Shih G., Riyahi S., Jeph S., Dev H., Hu R., Romano D., Teichman K., Blumenfeld J. D., Barash I., Chicos I., Rennert H., Prince M. R. Deployed Deep Learning Kidney Segmentation for Polycystic Kidney Disease MRI // *Radiol Artif Intell.* 2022. Feb 16;4(2):e210205.
11. Réda Clémence et al. Machine learning applications in drug development // *Computational and Structural Biotechnology Journal.* 2020. Vol. 18. 241–252.
12. Alves V. M., Korn D., Pervitsky V., Thieme A., Capuzzi S. J., Baker N., Chirkova R., Ekins S., Muratov E. N., Hickey A., Tropsha A. Knowledge-based approaches to drug discovery for rare diseases // *Drug Discov Today.* 2022. Feb; 27(2):490–502.
13. Álvarez-Machancoses Ó., Fernández-Martínez J. L. Using artificial intelligence methods to speed up drug discovery // *Expert Opinion on Drug Discovery.* 2019. 14(8). 769–777.
14. Choon Y. W., Choon Y. F., Nasarudin N. A., Al Jasmi F., Remli M. A., Alkayali M. H., Mohamad M. S. Artificial intelligence and database for NGS-based diagnosis in rare disease // *Front. Genet.* 2024. 14:1258083.
15. Svenstrup D., Jørgensen H. L., Winther O. Rare disease diagnosis: A review of web search, social media and large-scale data-mining approaches // *Rare Diseases.* 2015. 3(1).
16. Chloe M. M., Pauline Pao S. W., Felicite E. S., Felix Chi Hang Ch., Grace Pui Ying Ch., Tony Long Fung P., Brian Siu Chun A., Toby Chun Hei Ch., Yeow Kuan Ch., Eric Chun Yiu L., Ching Wan L. Computer-assisted patient identification tool in inborn errors of metabolism – potential for rare disease patient registry and big data analysis // *Clinica Chimica Acta.* 2024. Vol. 561. 119811. ISSN 0009-8981.
17. Huml R. A., Dawson J., Lipworth K. et al. Use of Big Data to Aid Patient Recruitment for Clinical Trials Involving Biosimilars and Rare Diseases // *Ther Innov Regul Sci.* 2020. 54. 870–877.

ОБ АВТОРАХ

Лабуткин Константин Сергеевич, студ. каф. ТК.

METADATA

Title: Artificial intelligence and big data in the diagnosis of rare diseases

Author: K.S. Labutkin¹

Affiliation:

¹ Ufa University of Science and Technology (UUST), Russia.

Email: kostya222lb@mail.ru

Language: Russian.

Source: *Molodezhnyj Vestnik UGATU* (scientific journal of Ufa University of Science and Technology), no. 1 (32), pp. 60-65, 2025. ISSN 2225-9309 (Print).

Abstract: Rare diseases, which pose a significant problem due to their low prevalence and complex diagnostic pathways, can now be diagnosed more accurately and faster thanks to big data processing technologies and artificial intelligence. These innovative approaches make it possible to analyze arrays of information and identify patterns that were previously unnoticeable, reducing the time to diagnosis and paving the way for more effective treatment. The paper provides an analysis of existing solutions and scientific articles in the field of artificial intelligence and big data analysis methods in the field of diagnosis and treatment of rare diseases. The prospects for the development of the application of this approach in various research areas are also being considered.

Keywords: rare diseases, artificial intelligence, big data, genome analysis, diagnostics.

About authors:

Labutkin Konstantin Sergeevich, student, Dept. of Technical Cybernetics (UUST).