

# Возможности машинного обучения для диагностики орфанных заболеваний

Дмитриева Н. Ю. 

АО «Астон Консалтинг», Москва, Российская Федерация

## Аннотация

Редкие или орфанные заболевания относятся к одним из наиболее тяжёлых групп заболеваний. При этом ранняя и точная диагностика таких заболеваний является серьёзной проблемой для врачей общей практики, педиатров и терапевтов. В статье рассмотрены возможности применения методов машинного обучения, в том числе искусственного интеллекта, для улучшения диагностики редких болезней. Приводится информация о различных моделях, разработанных, как международными специалистами, так и российскими исследователями.

**Ключевые слова:** орфанные заболевания; машинное обучение; редкие болезни; клинические регистры; наблюдательные программы; данные реальной клинической практики; RWD; RWE

### Для цитирования:

Дмитриева Н. Ю. Возможности машинного обучения для диагностики орфанных заболеваний. *Реальная клиническая практика: данные и доказательства*. 2023;3(3):36–39. <https://doi.org/10.37489/2782-3784-myrd-40>

**Поступила:** 20 августа 2023 г. **Принята:** 30 августа 2023 г. **Опубликована:** 12 октября 2023 г.

# Machine learning capabilities for the diagnosis of orphan diseases

Nataliya Yu. Dmitrieva 

AO "Aston Consulting", Moscow, Russian Federation

## Abstract

Rare or orphan diseases belong to one of the most severe groups of diseases. At the same time, early and accurate diagnosis of such diseases is a serious problem for general practitioners, pediatricians and therapists. The article discusses the possibilities of using machine learning methods, including artificial intelligence, to improve the diagnosis of rare diseases. Information is provided on various models developed by both international experts and Russian researchers.

**Keywords:** orphan diseases; machine learning; rare diseases; clinical registries; observation programs; real-world data; RWD; RWE

### For citation:

Dmitrieva NYu. Machine learning capabilities for the diagnosis of orphan diseases. *Real-World Data & Evidence*. 2023;3(3):36–39. <https://doi.org/10.37489/2782-3784-myrd-40>

**Received:** August 20, 2023. **Accepted:** August 30, 2023. **Published:** October 12, 2023.

## Введение

К одним из наиболее тяжёлых групп заболеваний относятся редкие или орфанные заболевания. Хотя они затрагивают небольшую часть популяции: в России считаются заболеваниями, которые имеют распространённость не более 10 случаев на 100 тысяч населения, без вовремя начатого лечения в большинстве своём они приводят к ранней инвалидизации пациента. Что значительно ухудшает жизнь пациента, сказывается на физическом и эмоциональном состоянии как его самого, так и его близких, а также накладывает большую экономическую нагрузку на государство.

Орфанные заболевания в основном являются генетическими, и, сопровождают пациента в течение всей жизни, при этом симптомы могут проявляться не сразу. Для врачей общей практики, педиатров и терапевтов, которые не имеют достаточных знаний для их выявления, ранняя и точная диагностика таких заболеваний является серьёзной проблемой [1]. Кроме того, редкие заболевания обычно показывают самые разнообразные проявления, что может сделать диагностику ещё более сложной. Доктор медицинских наук, директор ФГБНУ «Медико-генетический научный центр имени Н. П. Боч-

кова» Сергей Куцев отмечает: «Классический пример — болезнь Фабри, где от первых симптомов до корректного диагноза проходит 7–10 критических для жизни пациента лет» [2].

В связи с описанными выше проблемами показало интересным рассмотреть вопрос о возможностях применения методов машинного обучения, в том числе искусственного интеллекта, для улучшения диагностики редких болезней.

С одной стороны, небольшое число таких пациентов ставит под сомнение такую возможность, но изучение литературы показало, что такие задачи решаются и весьма успешно.

### Международный опыт

Анализ проведён на основе запроса «Orphan diseases machine learning Rare diseases» к базе PUBMED. В анализ были включены публикации 2022–2023 гг. на английском языке с возможностью доступа к полному тексту статьи.

Отобранные работы посвящены различным орфанным нозологиям.

Так в статье [4] показано, что искусственный интеллект продемонстрировал высокую эффективность в выявлении пациентов с болезнью Фабри. Набор данных содержал 4978 пациентов с подтверждённой болезнью Фабри и 1 000 000 пациентов без указанного диагноза. При испытаниях он достиг площади под кривой рабочих характеристик приемника (AUROC) 0,82. Высокие показатели были сохранены при тестировании на когортах только для мужчин и женщин, с AUROC 0,83 и 0,82 соответственно. Как пишут авторы, алгоритм использовал сотни фенотипических сигналов для прогнозирования и включал как знакомые симптомы, связанные с болезнью Фабри, такие как почечные проявления, так и менее изученные характеристики.

В работе [5] сообщается о создании модели машинного обучения на 205571 полной электронной медицинской карте из одного медицинского центра для диагностики острой печёночной порфирии на основе 30 известных случаев. В результате были идентифицированы 22 пациента с классическими симптомами заболевания. Из них у 7 человек был взят анализ мочи на порфобилиногенин, но ни один не был положительным. Как указывают авторы, это заставляет их работать над алгоритмом дальше, улучшая его.

В [6] указывается на возможности вычислительного извлечения смысла из естественного человеческого языка. Эти алгоритмы принимают в качестве входных данных документ и выводят полезное преобразование документа, например, позволяя выделять признаки более высокого уровня из электронных медицинских карт и проводить на основе их диагностику.

Учёные из Китая [7], использовали методы машинного обучения для диагностики фенилкетонурии. В большинстве случаев скрининг заболевания проводится путём оценки уровней фенилаланина (PHE) и тирозина (TYR) в неонатальных сухих кровяных пятнах, но он приносит высокий уровень ложноположительных результатов. А это в свою очередь, как пишут авторы, приводит к потере большого количества медицинских ресурсов и может вызвать панику в вовлечённых семьях. Таким образом, существует большая клиническая ценность для повышения точности скрининга фенилкетонурии. В данном исследовании учёные применили «алгоритм случайного леса» для улучшения модели скрининга, что позволило уменьшить уровень ложноположительных результатов.

В работе [8] приводятся факторы которые важны, для того, чтобы врачи и пациенты смогли доверять результатам полученным с помощью искусственного интеллекта. На основе опроса специалистов делаются следующие выводы: инструменты, используемые для диагностики редких заболеваний, могут восприниматься как более надёжные, если пользователь может поручиться за надёжность и точность технологии, а человек, использующий и/или разрабатывающий их, заслуживает доверия.

### Опыт Российской Федерации

В России данное направление тоже развивается, так в Научной электронной библиотеке eLIBRARY.RU представлена информация о нескольких грантах, посвящённых данной теме, но пока доступных результатов работ мало.

В статье [3] сообщается, о возможностях перепрофилирования лекарственных средств для помощи пациентам с орфанными заболеваниями. В данном случае с помощью методов глубокого обучения (глубинных нейронных сетей) проанализированы данные профиля экспрессии генов для предсказания терапевтических категорий препаратов.

По информации из Internet одной из компаний, занимающейся разработками в области выявления редких (генетических) заболеваний с помощью искусственного интеллекта, является Платформа прогнозной аналитики Webiomed компании ООО «К-СКАЙ» [9]. Как указано на сайте компании (<https://webiomed.ru/>), созданы алгоритмы для выявления болезни Фабри, синдрома Хантера и болезни Гоше. Для обучения системы взяты данные из электронных медицинских карт (ЭМК) медицинских учреждений.

Компания АО «Астон Консалтинг» с 2016 года является техническим оператором проекта «Аудит орфанных заболеваний» и ведёт проекты по управлению регистрами пациентов по 20 орфанным заболеваниям: онкологическим, онкогематологи-

ческим, аутоиммунным, лизосомным болезням накопления [10]. За время ведения проекта накоплена информация более чем по 15 000 орфанных пациентов. Важной характеристикой собранных данных является их структурированность, т. е. информация собирается не как текстовая выписка из ЭМК, а сразу разносится по необходимым полям. При этом структура карты близкая между различными нозологиями, что позволяет собирать в единый датасет информацию по различным орфанным нозологиям и выявлять совокупность фенотипических и клинических факторов, позволяющих проводить дифференцирование между близкими по симптоматике заболеваниями и верификацию поставленных диагнозов.

Динамическое наблюдение за пациентами, а также информация о терапии и оперативных вмешательствах позволяет отследить возникновение различных событий, таких как прогрессирование заболевания в течение определённого периода наблюдения, осложнения, нежелательные явления и другие, а методы машинного обучения позволяют выявить факторы, влияющие на их развитие у таких пациентов.

## ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

### Конфликт интересов.

Автор декларирует отсутствие конфликта интересов.

## СВЕДЕНИЯ ОБ АВТОРАХ

**Дмитриева Наталия Юрьевна** — к. б. н., начальник отдела информационных систем АО «Астон Консалтинг», Москва, Российская Федерация

**e-mail:** n.dmitrieva@aston-health.com

 <https://orcid.org/0000-0003-1072-243X>

## Литература/References

- Segura-Bedmar I, Camino-Perdones D, Guerrero-Aspizua S. Exploring deep learning methods for recognizing rare diseases and their clinical manifestations from texts. *BMC Bioinformatics*. 2022 Jul 6;23(1):263. doi: 10.1186/s12859-022-04810-y. PMID: 35794528; PMCID: PMC9258216.
- «Такеда» и «Фармимэкс» будут развивать диагностику редких заболеваний в России. *PHARM-PROM*. Интернет-ресурс. <https://pharmprom.ru/take-da-i-farmimeks-budut-razvivat-diagnostiku-red-kix-zabolevanij-v-rossii/?ysclid=lf5p34sow569226820>. (Режим доступа: 02.09.2023)
- Мирошниченко И.И., Вальдман Е.А., Кузьмин И.И. Новое предназначение старых лекарств (обзор). *Разработка и регистрация лекарственных средств*. 2023;12(1):182-190. <https://doi.org/10.33380/2305-2066-2023-12-1-182-190> [Miroshnichenko I.I., Valdman E.A., Kuz'min I.I. Old Drugs, New Indica-

tions (Review). *Drug development & registration*. 2023;12(1):182-190. (In Russ.)].

Помимо ЭМК важным источником информации могут быть и данные программ лабораторной диагностики. Например, для болезни Фабри на территории Российской Федерации такие программы есть у компаний Sanofi [11] и Takeda [12]. В таких программах собирается информация обо всех пациентах, направленных на диагностику, а также о результатах тестирования. Возможно, связь этих данных с информацией из ЭМК позволило бы создать новые модели машинного обучения, улучшающие качество диагностики орфанных заболеваний в России.

## Заключение

Таким образом, мы видим, что методы машинного обучения, в том числе и искусственный интеллект, активно используются для диагностики орфанных заболеваний во всём мире. Но, при этом существующие модели не решают всех проблем в данной области и остаётся много направлений для применения возможностей машинного обучения, как дополнительного средства для обработки данных реальной клинической практики (RWD) и доказательств из реальной клинической практики (RWE), для помощи пациентам с орфанными заболеваниями.

## ADDITIONAL INFORMATION

### Conflict of interests.

The author declares no conflict of interests.

## ABOUT THE AUTHORS

**Natalia Y. Dmitrieva** — Ph.D. Sc., Head of Information Systems Department, Aston Consulting JSC, Moscow, Russian Federation

**Corresponding author**

**e-mail:** n.dmitrieva@aston-health.com

 <https://orcid.org/0000-0003-1072-243X>

4. Jefferies JL, Spencer AK, Lau HA, Nelson MW, Giuliano JD, Zabinski JW, Boussios C, Curhan G, Gliklich RE, Warnock DG. A new approach to identifying patients with elevated risk for Fabry disease using a machine learning algorithm. *Orphanet J Rare Dis*. 2021 Dec 20;16(1):518. doi: 10.1186/s13023-021-02150-3. PMID: 34930374; PMCID: PMC8686369.
5. Hersh WR, Cohen AM, Nguyen MM, Benschung KL, Deloughery TG. Clinical study applying machine learning to detect a rare disease: results and lessons learned. *JAMIA Open*. 2022 Jun 30;5(2):ooac053. doi: 10.1093/jamiaopen/ooac053. PMID: 35783073; PMCID: PMC9243401.
6. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med*. 2019 Nov 19;11(1):70. doi: 10.1186/s13073-019-0689-8. PMID: 31744524; PMCID: PMC6865045.

7. Song Y, Yin Z, Zhang C, Hao S, Li H, Wang S, Yang X, Li Q, Zhuang D, Zhang X, Cao Z, Ma X. Random forest classifier improving phenylketonuria screening performance in two Chinese populations. *Front Mol Biosci.* 2022 Oct 11;9:986556. doi: 10.3389/fmolb.2022.986556. PMID: 36304929; PMCID: PMC9592754
8. Hallowell N, Badger S, Sauerbrei A, Nellåker C, Kerasidou A. "I don't think people are ready to trust these algorithms at face value": trust and the use of machine learning algorithms in the diagnosis of rare disease. *BMC Med Ethics.* 2022 Nov 16;23(1):112. doi: 10.1186/s12910-022-00842-4. PMID: 36384545; PMCID: PMC9670402
9. Выявление редких (генетических) заболеваний с помощью искусственного интеллекта. Интернет-ресурс. <https://webiomed.ru/novosti/zapuskaiut-proekt-elektronnaia-dispanserizatsiia/?ysclid=1f5tjhu-fyd561951251> (Режим доступа: 02.09.2023).
10. Орфанные заболевания. Интернет-ресурс. <https://aston-health.com/terapevticheskie-oblasti/orfannyye-zabolevaniya.html> (Режим доступа: 02.09.2023).
11. Диагностика болезней: МПС I, Фабри, Помпе, Ниманна-Пика типа А/В. <https://docsfera.ru/dbcontent/> (Режим доступа: 02.09.2023).
12. Программа диагностики НАО и ЛБН. Интернет-ресурс. <https://gendisease.aston-health.com/Home/About> (Режим доступа: 02.09.2023). 