



DOENÇAS RARAS E INTELIGÊNCIA ARTIFICIAL: UMA ANÁLISE DOS DESAFIOS E OPORTUNIDADES EM BIOINFORMÁTICA E INFORMÁTICA MÉDICA COM DADOS DA BASE ORPHANET

Eloise dos Santos¹, Davi Augusto Bogdanovicz Antunes dos Santos¹, Michael Pereira Gomes, Murilo Cardoso¹, Wellynton Rafael Hass¹, Gabriel Passos de Jesus^{1,2,3}

¹Centro de Ensino Superior de Maringá, Ponta Grossa, Brasil (elosnt06@gmail.com)

²Pontifícia Universidade Católica do Paraná, Curitiba, Brasil

³Universidade Federal do Paraná, Curitiba, Brasil

Resumo: As doenças raras afetam cerca de 6% da população mundial, impondo desafios como atraso diagnóstico e fragmentação de dados. Esta revisão analisa o uso de IA em bioinformática, destacando aplicações em genômica, mineração de prontuários e suporte à decisão clínica. Apesar do potencial, persistem limitações como bases de dados pequenas, vieses e validação clínica. Conclui-se que a integração de dados clínicos, aliada à validação multicêntrica, é necessária para avanços robustos.

Palavras-chave: Doenças raras; Inteligência Artificial; Bioinformática; Medicina; Ciência da Computação

INTRODUÇÃO

As doenças raras são numerosas, heterogêneas e, coletivamente, representam carga sanitária e social expressiva. A análise epidemiológica da base Orphanet mostrou que essas doenças podem afetar entre 3,5% e 5,9% da população global, além de apresentarem forte predominância de condições genéticas e início pediátrico em grande parte dos casos (NGUYEN et al., 2019). Em paralelo, estudos de síntese recentes ressaltam que o campo permanece marcado por atraso diagnóstico, itinerários assistenciais fragmentados, escassez de evidências clínicas robustas e baixa disponibilidade de terapias específicas, o que amplia custos humanos, sociais e econômicos (FERREIRA, 2019; GROFT; POSADA; TARUSCIO, 2021).

Nesse cenário, a inteligência artificial desponta como eixo estratégico para a literatura em Ciência da Computação aplicada à saúde. Revisões de escopo indicam crescimento do uso de aprendizado de máquina em doenças raras, sobretudo em tarefas de diagnóstico, prognóstico e classificação, com emprego recorrente de dados de imagem, variáveis clínicas, dados demográficos e conjuntos ômicos (SCHAEFER et al., 2020). Contudo, a própria literatura computacional adverte que o domínio das doenças raras impõe um problema metodológico central: o reduzido número de amostras, combinado com alta dimensionalidade dos dados biomédicos, dificulta generalização, reprodutibilidade e avaliação externa dos modelos (BANERJEE et al., 2023).

No âmbito da bioinformática, a integração entre sequenciamento de exoma/genoma, bases de conhecimento biomédico e algoritmos de IA tem ampliado a capacidade de priorizar variantes e sugerir hipóteses diagnósticas mais rapidamente. Revisões recentes destacam que a IA aplicada ao diagnóstico genético pode reduzir o esforço interpretativo, apoiar a correlação genótipo-fenótipo e fortalecer estratégias de medicina de precisão em doenças raras (JAMES et al., 2023; HONG et al., 2024). De modo complementar, revisões sistemáticas sobre abordagens de aprendizado de máquina mostram utilidade crescente de modelos treinados com dados de sequenciamento para estratificação, identificação de mutações e apoio ao diagnóstico de doenças genéticas raras (ROMÁN-NARANJO; PARRA-PÉREZ; LÓPEZ-ESCÁMEZ, 2023).

Na informática médica, prontuários eletrônicos e técnicas de processamento de linguagem natural vêm sendo mobilizados para rastrear sinais clínicos dispersos e identificar pacientes potencialmente subdiagnosticados. Estudos e revisões demonstram que registros eletrônicos em saúde podem servir como base para sistemas de triagem e vigilância de doenças raras, ainda que desafios de padronização, qualidade de dado e interoperabilidade permaneçam substanciais (GARCELON; BURGUN; SALOMON; NEURAZ, 2020; COHEN et al., 2020). Além disso, a consolidação de registros de pacientes e infraestruturas de dados é apontada como requisito para treinamento, validação e monitoramento de sistemas inteligentes com impacto clínico real (HAGEMAN et al., 2023).



Diante desse panorama, este trabalho desenvolve uma revisão de literatura sobre desafios e oportunidades da IA aplicada às doenças raras, enfatizando duas frentes complementares: a bioinformática, voltada à análise genômica e multiômica, e a informática médica, centrada em registros clínicos, apoio à decisão e integração assistencial. O objetivo é sintetizar como a literatura recente descreve ganhos potenciais, limitações técnicas e implicações éticas do uso de IA nesse campo, oferecendo uma base conceitual para pesquisas futuras e para a adoção responsável dessas tecnologias no contexto biomédico.

MATERIAL E MÉTODOS

Este estudo caracteriza-se como uma revisão narrativa de literatura, de natureza qualitativa e exploratória, voltada à análise das contribuições da inteligência artificial para o campo das doenças raras, com foco em bioinformática e informática médica. A construção do trabalho foi orientada por publicações científicas recentes, prioritariamente entre 2016 e 2025, selecionadas por sua relevância temática e por seu potencial de representar o estado da arte nas interfaces entre computação e medicina.

O levantamento bibliográfico concentrou-se em artigos de revisão, revisões sistemáticas, estudos de escopo e trabalhos aplicados sobre aprendizado de máquina, processamento de linguagem natural, análise genômica, prontuários eletrônicos em saúde e medicina de precisão. Como critérios de inclusão, privilegiaram-se estudos que discutem aplicações de IA no diagnóstico, na priorização de variantes, na integração de dados clínicos e moleculares e na identificação de pacientes com doenças raras. Foram excluídos trabalhos com aderência temática limitada, duplicados ou sem contribuição direta para os eixos analíticos definidos.

Após a seleção, os estudos foram examinados por leitura analítica e organizados em duas categorias principais: (i) bioinformática, abrangendo métodos computacionais aplicados a dados genômicos e multiômicos; e (ii) informática médica, contemplando registros eletrônicos de saúde, apoio à decisão clínica e interoperabilidade de dados. A síntese dos achados buscou identificar oportunidades, limitações técnicas, desafios éticos e lacunas de pesquisa, permitindo discutir, de forma condensada, como a IA tem sido empregada no enfrentamento dos problemas diagnósticos e assistenciais associados às doenças raras.

RESULTADOS E DISCUSSÃO

Os estudos analisados convergem em um ponto central: a principal oportunidade da IA em doenças raras está na redução da chamada *diagnostic odyssey*, isto é, do percurso longo e fragmentado até a confirmação diagnóstica. Essa conclusão reforça tanto

as revisões já mobilizadas na introdução quanto levantamentos mais recentes, que mostram predominância de aplicações voltadas ao diagnóstico e à triagem clínica em detrimento de usos plenamente incorporados à rotina assistencial (SCHAEFER et al., 2020; BANERJEE et al., 2023; ACERO RUGE et al., 2025). Em bioinformática, esse potencial aparece de forma mais clara na priorização de variantes, no apoio à correlação genótipo-fenótipo e na integração de dados multiômicos. Trabalhos recentes indicam que pipelines combinando genômica, transcriptômica e apoio algorítmico conseguem ampliar o rendimento diagnóstico e acelerar a interpretação molecular, sobretudo em coortes pediátricas e em programas nacionais de doenças raras (JAMES et al., 2023; LUNKE et al., 2023; CHOON et al., 2024; ALI; LI; AGRAWAL, 2025). Assim, a oportunidade mais consistente em bioinformática não é apenas automatizar análises, mas articular IA com bases curadas, conhecimento biomédico e fluxos clínico-laboratoriais de medicina de precisão.

Na informática médica, os resultados sugerem que o ganho mais promissor reside na exploração de registros eletrônicos em saúde para identificação precoce de casos, construção de fenótipos computáveis e produção de evidência do mundo real. Estudos sobre mineração de prontuários e processamento de linguagem natural mostram que informações dispersas em texto livre podem antecipar suspeitas diagnósticas, como ilustrado em doenças neurológicas raras e em aplicações baseadas em prontuários hospitalares (GARCELON; BURGUN; SALOMON; NEURAZ, 2020; COHEN et al., 2020; LO BARCO et al., 2021). Revisões mais amplas sobre NLP em prontuários indicam, contudo, que o valor desses sistemas depende fortemente da qualidade documental, da padronização terminológica e da possibilidade de reutilização dos dados para decisão clínica (HOSSAIN et al., 2023). Nesse sentido, redes federadas e estratégias de governança de dados aparecem como oportunidade decisiva para ampliar reprodutibilidade, transparência e escala analítica sem exigir centralização irrestrita dos dados sensíveis, questão particularmente crítica em doenças raras devido ao baixo número de pacientes e à dispersão geográfica dos casos (VAN BAALEN et al., 2024; CHEN et al., 2025).

Ao mesmo tempo, a literatura revela lacunas importantes que respondem diretamente à pergunta do título deste trabalho. O principal desafio em bioinformática continua sendo a escassez de amostras representativas, agravada por alta dimensionalidade, desbalanceamento entre classes, anotações incompletas e dependência de bases de referência concentradas em poucos centros e populações. Já na informática médica, persistem barreiras de interoperabilidade, heterogeneidade semântica entre



sistemas, ausência de validação multicêntrica e baixa explicabilidade dos modelos quando aplicados ao apoio à decisão. Soma-se a isso o fato de que muitos estudos permanecem em nível de prova de conceito, com pouca demonstração de impacto prospectivo sobre desfechos clínicos ou fluxos de cuidado (BANERJEE et al., 2023; HAGEMAN et al., 2023; ACERO RUGE et al., 2025). Por outro lado, a literatura também amplia o horizonte de oportunidades: a IA pode acelerar descoberta e reposicionamento de fármacos, fortalecer estudos com dados do mundo real e apoiar ecossistemas colaborativos entre pesquisadores, serviços e associações de pacientes (NAPOLITANO et al., 2024; KLEIN et al., 2024). Desse modo, os achados da revisão indicam que o avanço mais relevante para doenças raras depende menos de algoritmos isolados e mais da integração entre bioinformática, informática médica, governança ética e infraestrutura compartilhada de dados, condição necessária para transformar promessas computacionais em benefício clínico mensurável.

CONCLUSÃO

A revisão de literatura permitiu concluir que a inteligência artificial tem potencial concreto para reduzir atrasos diagnósticos e qualificar o cuidado em doenças raras, especialmente quando aplicada à priorização de variantes, à integração de dados multiômicos e à análise de registros eletrônicos em saúde. Sob essa perspectiva, bioinformática e informática médica aparecem como campos complementares: a primeira fortalece a interpretação molecular e a medicina de precisão, enquanto a segunda amplia a capacidade de identificação precoce, monitoramento e apoio à decisão clínica.

Entretanto, o avanço do campo ainda depende do enfrentamento de lacunas estruturais, como escassez de dados, baixa interoperabilidade, limitada validação multicêntrica e desafios éticos relacionados à transparência e à governança da informação. Assim, responde-se à pergunta do título destacando que os principais desafios estão na qualidade, integração e validação dos dados, ao passo que as maiores oportunidades residem na construção de ecossistemas colaborativos, auditáveis e orientados ao benefício clínico, capazes de transformar soluções computacionais em impacto real para pacientes com doenças raras.

AGRADECIMENTOS

Os autores agradecem nominalmente à Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação da UniCesumar; ao Programa de Pós-Graduação em Informática da PUCPR; ao Programa de Pós-Graduação em Medicina Interna da UFPR; ao Programa de Pós-Graduação em Clínica Cirúrgica da UFPR; ao Programa de Pós-Graduação em Computação Aplicada da UEPG; ao Prof. Dr. Leonardo Camargo, do curso de Medicina da PUCPR - Londrina; à Dra. Maria de Fátima Domingos (FSP/USP); aos Prof. Dr. Felipe de Lara Janz, Prof. MSc. Josyel Olszewski, Prof. Dr. Wanderson Ribeiro e Prof. Dr. Fábio Brasil de Oliveira, à Dra. Danielly Augusta Ramos Felippi, à Dra. Camila Ceruti e ao Dr. Miguel Fabrin, e à acadêmica Ester Souza, do curso de Medicina da UFV, pelo apoio institucional, pelas contribuições acadêmicas e pelo incentivo ao desenvolvimento deste trabalho.

REFERÊNCIAS

- ABOZAIID, Ghada Mohammed; KERR, Katie; MCKNIGHT, Amy; AL-OMAR, Hussain A. Criteria to define rare diseases and orphan drugs: a systematic review protocol. *BMJ Open*, v. 12, n. 7, e062126, 2022.
- ACERO RUGE, L. M. et al. Artificial intelligence for the comprehensive approach to orphan/rare diseases: a scoping review. *Semergen*, v. 51, n. 5, art. 102434, 2025.
- ALI, Sara S.; LI, Qifei; AGRAWAL, Pankaj B. Implementation of multi-omics in diagnosis of pediatric rare diseases. *Pediatric Research*, v. 97, n. 4, p. 1337--1344, 2025.
- BANERJEE, Jineta. et al. Deepashree Venkatesh; GUINNEY, Justin; GREENE, Casey. Machine learning in rare disease. *Nature Methods*, v. 20, n. 6, p. 803--814, 2023.
- CHEN, Jie. et al.. Use of Real-World Data and Real-World Evidence in Rare Disease Drug Development: A Statistical Perspective. *Clinical Pharmacology & Therapeutics*, v. 117, n. 4, p. 946--960, 2025.
- CHOON, Yee Wen. et al. Artificial intelligence and database for NGS-based diagnosis in rare disease. *Frontiers in Genetics*, v. 14, art. 1258083, 2024.
- COHEN, Aaron M. et al. Detecting rare diseases in electronic health records using machine learning and knowledge engineering: case study of acute hepatic porphyria. *PLOS ONE*, v. 15, n. 7, e0235574, 2020.
- FERREIRA, Carlos R. The burden of rare diseases. *American Journal of Medical Genetics Part A*, v. 179, n. 6, p. 885--892, 2019.



- GARCELON, Nicolas; BURGUN, Anita; SALOMON, Rémi; NEURAZ, Antoine. Electronic health records for the diagnosis of rare diseases. *Kidney International*, v. 97, n. 4, p. 676-686, 2020.
- GROFT, Stephen C.; POSADA, Manuel; TARUSCIO, Domenica. Progress, challenges and global approaches to rare diseases. *Acta Paediatrica*, v. 110, n. 10, p. 2711--2716, 2021.
- HAGEMAN, Isabel C.; VAN ROOIJ, Iris A. L. M.; DE BLAAUW, Ivo; TRAJANOVSKA, Misel; KING, Sebastian K. A systematic overview of rare disease patient registries: challenges in design, quality management, and maintenance. *Orphanet Journal of Rare Diseases*, v. 18, art. 106, 2023.
- HONG, Juhyeon; LEE, Dajun; HWANG, Ayoung; KIM, Taekeun; RYU, Hong-Yeoul; CHOI, Jungmin. Rare disease genomics and precision medicine. *Genomics & Informatics*, v. 22, n. 1, art. 28, 2024.
- HOSSAIN, Elias. et al. Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review. *Computers in Biology and Medicine*, v. 155, art. 106649, 2023.
- JAMES, Kiely N.; PHADKE, Sujal; WONG, Terence C.; CHOWDHURY, Shimul. Artificial intelligence in the genetic diagnosis of rare disease. *Clinics in Laboratory Medicine*, v. 43, n. 1, p. 127--143, 2023.
- KLEIN, T. L. et al. A rare partnership: patient community and industry collaboration to shape the impact of real-world evidence on the rare disease ecosystem. *Orphanet Journal of Rare Diseases*, v. 19, n. 1, art. 262, 2024.
- LO BARCO, Tommaso. et al. Improving early diagnosis of rare diseases using Natural Language Processing in unstructured medical records: an illustration from Dravet syndrome. *Orphanet Journal of Rare Diseases*, v. 16, art. 309, 2021.
- LUNKE, Sebastian. et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nature Medicine*, v. 29, n. 7, p. 1681--1691, 2023.
- NAPOLITANO, Giulio; HAS, Canan; SCHWERK, Anne; YUAN, Jui-Hung; ULLRICH, Carsten. Potential of Artificial Intelligence to Accelerate Drug Development for Rare Diseases. *Pharmaceutical Medicine*, v. 38, n. 2, p. 79--86, 2024.
- NGUYEN, Alexandra. et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics*, v. 27, n. 2, p. 165--173, 2019.
- ROMÁN-NARANJO, Pedro; PARRA-PÉREZ, Alberto M.; LÓPEZ-ESCÁMEZ, José A. A systematic review on machine learning approaches in the diagnosis and prognosis of rare genetic diseases. *Journal of Biomedical Informatics*, v. 143, art. 104429, 2023.
- SCHAEFER, Julia; LEHNE, Moritz; SCHEPERS, Josef; PRASSER, Fabian; THUN, Sylvia. The use of machine learning in rare diseases: a scoping review. *Orphanet Journal of Rare Diseases*, v. 15, art. 145, 2020.
- VAN BAALEN, Valerie; DIDDEN, Eva-Maria; ROSENBERG, Daniel; BARDENHEUER, Kristina; VAN SPEYBROECK, Michel; BRAND, Monika. Increase transparency and reproducibility of real-world evidence in rare diseases through disease-specific Federated Data Networks. *Pharmacoepidemiology and Drug Safety*, v. 33, n. 4, e5778, 2024.