Editorial

DOI: 10.5582/irdr.2025.01030

Artificial intelligence applications in rare and intractable diseases: Advances, challenges, and future directions

Kenji Karako*

Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

SUMMARY: Rare and intractable diseases affect an estimated 3.5% to 5.9% of the global population but remain largely underserved in terms of diagnosis and treatment, with effective therapies available for only about 5% of conditions. This paper presents an overview of recent advances in artificial intelligence (AI) applications targeting these challenges. In diagnostic support, AI has been utilized to analyze genomic data and facial images, enhancing the accuracy and efficiency of identifying rare genetic syndromes. In therapeutic development, AI-driven analysis of biomedical knowledge graphs has enabled the prediction of potential treatment candidates for diseases lacking existing therapies. Additionally, generative models have accelerated drug discovery by identifying novel targets and designing candidate compounds, some of which have progressed to clinical evaluation. AI has also facilitated clinical trial support by automating patient eligibility screening using electronic health records, improving recruitment efficiency for trials that often struggle with small, geographically dispersed patient populations. Despite these advancements, challenges remain in ensuring data quality, interpretability of AI outputs, and the standardization of infrastructure across institutions. Moving forward, international data-sharing platforms integrating diverse modalities — clinical, genomic and image — are expected to play a pivotal role in enabling reliable, scalable, and ethically responsible AI applications. These developments hold the potential to transform the landscape of rare disease diagnosis, treatment, and research.

Keywords: rare diseases, intractable diseases, artificial intelligence

1. Introduction

Although rare and intractable diseases are individually uncommon, they are estimated to affect approximately 3.5% to 5.9% of the global population, corresponding to 263 to 446 million people worldwide (*1*). Most of these diseases are chronic and progressive, severely impairing patients' quality of life and requiring long-term support from healthcare and welfare systems. To date, effective treatments have been established for only about 5% of rare diseases (*2*), leaving the majority of patients to live with uncertainty in diagnosis and treatment. One of the most pressing challenges in this field is the frequent occurrence of delayed and inaccurate diagnoses. A large-scale European survey reported that the average time from symptom onset to a confirmed diagnosis is approximately 4.7 years, with one in four patients requiring over five years (*3*). This so-called "diagnostic odyssey" often involves consultations with more than five physicians and multiple instances of misdiagnosis before an accurate diagnosis is reached.

The causes of diagnostic delay are multifaceted, including insufficient knowledge of rare diseases among clinicians, nonspecific symptoms, symptoms spanning multiple medical specialties, and limited access to specialized medical centers. Misdiagnosis not only delays appropriate treatment but may also result in unnecessary or harmful interventions, deterioration of trust between patients and healthcare providers, and wasteful use of medical resources. Furthermore, the economic burden associated with treatment is a critical concern. Orphan drugs developed for rare diseases tend to be extremely expensive, with an average annual treatment cost in the United States reaching approximately USD 219,000 (*4*).

Thus, the field of rare and intractable diseases faces substantial unresolved challenges in diagnosis, treatment, and financial support. In recent years, artificial intelligence (AI) technologies have attracted increasing attention as promising tools to address these challenges. Applications of AI in this domain have demonstrated utility in diagnostic support, patient screening, drug discovery, and literature analysis using natural language processing. This article highlights notable AI-based studies on rare and intractable diseases, reflects on the challenges of implementing AI models and the critical role of data platforms, and offers insights into future

developments in the field.

2. Applications of AI in rare and intractable diseases

2.1. Diagnostic support

Various AI-driven approaches have been explored to support the diagnosis of rare and intractable diseases. In recent years, notable advancements have been reported in fields such as genomic data analysis and AI-based medical imaging.

In the domain of genomic diagnosis, De La Vega *et al*. developed an AI-powered whole genome interpretation system aimed at improving the efficiency of genetic diagnoses in patients with rare diseases (*5*). This system integrates genomic variant data with clinical phenotypes to automatically rank candidate disease-causing genes. Validation using data from 119 patients demonstrated that in 92% of cases, the correct gene was ranked within the top two candidates. Moreover, for cases involving structural variants, the system successfully identified the causative gene as the top-ranked candidate in 17 out of 20 cases. However, the method has primarily been validated on known diagnostic cases, and its applicability to novel gene discovery and unresolved cases requires further investigation. Additionally, as the AI-generated outputs necessitate expert validation, the role of AI in diagnosis is best situated within a human-in-the-loop framework, emphasizing collaboration with clinicians.

In the field of medical imaging, AI has also been employed to support the diagnosis of rare diseases. A prominent example is DeepGestalt, a deep learning model developed by Gurovich *et al*. that analyzes facial photographs to suggest potential genetic syndromes. Trained on more than 17,000 facial images, the model can recognize over 200 syndromes (*6*). Clinical evaluations have reported a top-10 accuracy of 91%, indicating strong potential as a diagnostic support tool. In a study involving 25 patients with KBG syndrome, the correct diagnosis was included within the top five suggestions in 80% of cases (*7*). These findings highlight the utility of image-based AI in objectively capturing subtle morphological features characteristic of rare syndromes.

2.2. Prediction of therapeutic options

Currently, only approximately 5% of rare and intractable diseases have established and effective treatment options. To bridge this gap, increasing attention has been paid to drug repurposing — extending the indications of existing drugs — and the application of AI technologies to novel drug design.

One notable example is the development of TxGNN, a graph neural network-based AI model (*8*). This model successfully identified treatment candidates for over 17,000 diseases, including rare and ultra-rare conditions. TxGNN is the first AI model capable of predicting therapeutic indications for untreated diseases through zero-shot learning. Compared to conventional approaches, it demonstrated approximately 50% higher accuracy in candidate identification. By learning from comprehensive biomedical knowledge graphs, TxGNN systematizes what was previously a process reliant on serendipity or clinical experience, thereby facilitating the rapid identification of potential treatments for rare diseases. However, the therapeutic effectiveness of candidate drugs proposed by TxGNN requires experimental validation, and concerns remain regarding the completeness of the underlying knowledge graphs and potential data biases. Despite these limitations, this study represents a landmark example of AI's potential in the rare disease domain.

Additional studies have also explored AI-driven discovery of therapeutic candidates using large-scale biological data and knowledge bases. For instance, Cong *et al*. proposed a two-step machine learning approach that first clusters diseases based on gene expression patterns and then evaluates drugs that can reverse abnormal expression profiles (*9*). Using this method, 22 drug candidates were identified, including the HDAC inhibitor vorinostat. These agents showed potential therapeutic effects for rare inflammatory myopathies such as inclusion body myositis, polymyositis, and dermatomyositis.

These studies demonstrate how deep learning and knowledge graph analysis can uncover latent associations between diseases and drugs, offering opportunities for drug repurposing that may be overlooked by traditional methods. Nonetheless, challenges persist, including the high cost of experimental validation and the inherent data scarcity associated with rare diseases, which may limit model accuracy. Moving forward, a key focus will be on developing efficient strategies to prioritize and validate AI-generated therapeutic predictions.

2.3. Drug discovery

AI technologies are also increasingly being applied to drug discovery for rare and intractable diseases. Traditionally, drug development has been a time- and cost-intensive process, spanning from target identification to lead compound design and clinical evaluation. However, recent advances in generative AI models have begun to significantly accelerate these steps. A prominent example is the AI-driven drug discovery platform PandaOmics, which successfully identified a novel therapeutic target for idiopathic pulmonary fibrosis and designed a small-molecule compound targeting it within just a few years. This candidate has already progressed to a Phase IIa clinical trial, marking the first case in which both the disease target and the compound were discovered and designed entirely by AI and

subsequently advanced to human trials.

These developments suggest that generative AI can make identification of previously unknown targets feasible and enable rapid design of novel therapeutic candidates, even for diseases previously considered untreatable. Nevertheless, compounds proposed by AI must still undergo rigorous validation to confirm their efficacy and safety through conventional clinical trials. Therefore, while AI has potential to streamline the drug discovery pipeline, cautious and thorough evaluation remains essential. Despite these limitations, such pioneering examples highlight the transformative potential of AI in reshaping drug discovery strategies for rare and intractable diseases. Integration of AI into various stages of drug development is expected to become increasingly impactful in the years ahead.

2.4. AI applications in clinical trial support and patient recruitment

In the context of rare and intractable diseases, recruiting eligible participants for clinical trials poses a significant challenge due to the limited number of potential candidates and their wide geographical distribution. Traditionally, patient enrollment has been a time-consuming and labor-intensive process, often leading to delays or even discontinuation of trials. To address this issue, recent advances have focused on leveraging AI to improve the efficiency of patient identification and eligibility screening. In addition to structured data from electronic health records (EHRs), AI techniques such as natural language processing and machine learning have shown promise in extracting relevant information from unstructured clinical narratives and assessing trial eligibility.

For example, a study involving breast cancer patients demonstrated that an AI system could evaluate trial eligibility using EHR data with an accuracy of approximately 87.6% (*10*). Such AI-supported systems can automate the otherwise complex and manual process of matching patients to trial criteria, thereby reducing human errors and oversight. Moreover, these technologies can accelerate patient recruitment and reduce the time required to initiate trials. With continued advancements in EHRs standardization and AI model refinement, efficient execution of clinical trials for rare diseases is becoming increasingly feasible and practical.

3. Challenges and future perspectives in application of AI

Application of AI to rare and intractable diseases has produced promising advances in a variety of areas, including diagnosis, treatment, drug discovery, and clinical trial support. AI, with its capabilities in computational processing and information integration, has shown great potential in domains where conventional medical technologies and research approaches face limitations. However, practical implementation in clinical settings and broader societal adoption are hindered by a range of technical, ethical, and regulatory challenges. To ensure effective and sustainable AI deployment, it is essential to systematically address these issues and establish a clear vision for future development. This section outlines the major obstacles currently facing AI applications and discusses strategies for overcoming them, along with prospects for future advancement.

3.1. Overcoming data scarcity and learning bias

The performance of AI models is highly dependent on the quality and quantity of training data. In the field of rare diseases, the intrinsic "rarity" of each condition presents a major obstacle — available case numbers are extremely limited. This scarcity makes it difficult to compile large, high-quality datasets required for effective machine learning, increasing the risk of reproducibility bias, where model performance is skewed due to imbalanced training data in terms of disease type, ethnicity, or age group.

Additionally, structural inconsistencies across institutions — such as variations in EHR formats and discrepancies in diagnostic terminology — further complicate data integration. Under such constraints, models are prone to overfitting and may lack sufficient generalizability, compromising their reliability in real-world clinical applications. To mitigate these issues, strategies such as transfer learning, data augmentation, and the use of simulated data are increasingly employed. Synthesized datasets that emulate characteristics of rare diseases, as well as transfer learning from similar conditions using large existing databases, may help build robust models even in data-constrained environments.

3.2. Model transparency and clinical accountability

Clarifying the rationale behind AI-generated predictions and diagnoses is essential for clinical applications. Particularly when AI is used to assist in medical decision-making, a lack of transparency in the model's reasoning process can undermine clinicians' trust and reliability of their judgments. This challenge is especially prominent in deep learning-based systems, often referred to as "black-box" models.

To address this, the field has seen increasing interest in explainable AI. Techniques such as SHapley Additive exPlanations (*11*) and Local Interpretable Model-Agnostic Explanations (*12*) can provide visual explanations of feature contributions and justify predictions by comparing similar cases. Moreover, attention mechanisms and graph-based architectures with inherent interpretability have been proposed, which can further support integration into clinical workflows. These advances are expected to facilitate human-in-the-loop

collaboration by enabling physicians to better understand and verify AI-driven insights.

3.3. Future directions and perspectives

To enable AI to function effectively in the realm of rare and intractable diseases, a multifaceted approach is necessary — not only technological innovation but also supportive policy frameworks and social infrastructure. Among the most critical priorities is establishment of international data-sharing platforms. Given the inherently limited case numbers for rare diseases, data collection at a single institution or within a single country is insufficient. Instead, large-scale data integration through multi-institutional and multinational collaborations, using standardized data formats, is essential. These platforms are not merely repositories; they serve as information hubs connecting researchers, clinicians, and patients, and play a vital role in enhancing both diversity and statistical validity of data used for AI model development.

Furthermore, platforms capable of integrating multidimensional data — such as clinical records, genomic profiles, imaging data, and lifestyle information — are essential for building high-accuracy, generalizable AI models. In the future, such platforms are expected to serve as implementation infrastructures for AI-based diagnostic and treatment support systems in actual clinical practice. In addition, platforms designed to continuously collect real-world data will enable ongoing model refinement and feedback learning, contributing to sustainable performance improvement and overall healthcare quality. Thus, development and operationalization of such platforms will be a cornerstone in long-term evolution of AI-enabled precision medicine.

In conclusion, rare disease research inherently faces critical limitations — such as disease heterogeneity, extremely limited patient populations, and diagnostic complexity — that conventional medical technologies and research approaches often struggle to address. In this context, AI holds great promise due to its strengths in information integration and knowledge analysis. Application of AI to diverse data types — such as genomic, imaging, and natural language data — has led to tangible progress in improving diagnostic accuracy, identifying new therapeutic candidates, and enhancing patient recruitment in clinical trials. However, successful implementation of AI in real-world clinical practice still faces significant hurdles. These include data scarcity and bias, the need for model interpretability, and establishment of appropriate regulatory and ethical frameworks. Addressing these issues will require not only technological innovation but also structural efforts such as international collaboration and development of inclusive, patient-centered data infrastructures. In particular, construction and operation of international data-sharing platforms will be key to ensuring the reliability and generalizability of AI models in the rare disease domain. Such platforms, by enabling standardized collection and analysis of multidimensional data and supporting cyclical feedback between model development and real-world deployment, are expected to significantly accelerate practical adoption and long-term advancement of AI-driven medical support systems.

Funding: None.

Conflict of Interest: The author has no conflicts of interest to disclose.

References

1. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A. Estimating cumulative point prevalence of rare diseases: Analysis of the Orphanet database. Eur J Hum Genet. 2020; 28:165-173.
2. Rare Disease Diversity Coalition, National Organization for Rare Disorders. Inequities in the rare disease community: The voices of diverse patients and caregivers. *https://rarediseases.org/wp-content/uploads/2024/08/RDDC-Survey-Report.pdf* (accessed May 15, 2025).
3. Faye F, Crocione C, Anido de Peña R, *et al*. Time to diagnosis and determinants of diagnostic delays of people living with a rare disease: Results of a Rare Barometer retrospective patient survey. Eur J Hum Genet. 2024; 32:1116-1126.
4. Althobaiti H, Seoane-Vazquez E, Brown LM, Fleming ML, Rodriguez-Monguio R. Disentangling the Cost of Orphan Drugs Marketed in the United States. Healthcare (Basel). 2023; 11:558.
5. De La Vega FM, Chowdhury S, Moore B, *et al*. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. Genome Med. 2021; 13:153.
6. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, Basel-Salmon L, Krawitz PM, Kamphausen SB, Zenker M, Bird LM, Gripp KW. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med. 2019; 25:60-64.
7. Nishat SMH, Shahid Tanweer A, Alshamsi B, Shaheen MH, Shahid Tanveer A, Nishat A, Alharbat Y, Alaboud A, Almazrouei M, Ali-Mohamed RA. Artificial intelligence: A new frontier in rare disease early diagnosis. Cureus. 2025; 17:e79487.
8. Huang K, Chandak P, Wang Q, Havaldar S, Vaid A, Leskovec J, Nadkarni GN, Glicksberg BS, Gehlenborg N, Zitnik M. A foundation model for clinician-centered drug repurposing. Nat Med. 2024; 30:3601-3613.
9. Cong Y, Shintani M, Imanari F, Osada N, Endo T. A new approach to drug repurposing with two-stage prediction, machine learning, and unsupervised clustering of gene expression. OMICS. 2022; 26:339-347.
10. Haddad T, Helgeson JM, Pomerleau KE, Preininger AM, Roebuck MC, Dankwa-Mullan I, Jackson GP, Goetz MP. Accuracy of an artificial intelligence system for cancer clinical trial eligibility screening: Retrospective pilot study. JMIR Med Inform. 2021; 9:e27767.
11. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-*

Paper.pdf (accessed May 15, 2025).
12. Ribeiro MT, Singh S, Guestrin C. "why should i trust you?" explaining the predictions of any classifier. In: NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session. *https://aclanthology.org/N16-3.pdf* (accessed May 15, 2025).

Received April 3, 2025; Revised May 16, 2025; Accepted May 23, 2025.

**Address correspondence to:*
Kenji Karako, Department of Surgery, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8655, Japan.
E-mail: tri.leafs@gmail.com

Released online in J-STAGE as advance publication May 28, 2025.