

Few-Shot Learning: The AI Breakthrough for Rare Disease Diagnosis

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Published: November 28, 2023 | AI Diagnostics

DOI: [10.5281/zenodo.17997294](https://doi.org/10.5281/zenodo.17997294)

Abstract

Few-Shot Learning: The AI Breakthrough for Rare Disease Diagnosis The field of artificial intelligence (AI) has revolutionized healthcare, offering unpr...

Few-Shot Learning: The AI Breakthrough for Rare Disease Diagnosis

The field of artificial intelligence (AI) has revolutionized healthcare, offering unprecedented capabilities in diagnostics, drug discovery, and personalized medicine. However, a significant challenge persists in the realm of **rare diseases**. With over 7,000 distinct rare diseases identified, many of which are genetic, the low prevalence of each condition translates directly into a profound **data scarcity** problem for machine learning models. Traditional deep learning approaches, which require thousands of labeled data points for effective training, often fail when confronted with the limited patient data available for a single rare disorder. This scarcity contributes to the diagnostic odyssey, where approximately 70% of individuals seeking a diagnosis remain undiagnosed [1].

The Data Scarcity Challenge in Rare Disease AI

Rare diseases, defined in the United States as conditions affecting fewer than 200,000 people, present a unique hurdle for AI in healthcare. The heterogeneity of clinical presentations, coupled with clinicians' limited experience with such diseases, makes diagnosis a complex and time-consuming process. While deep learning excels in common diseases with large, well-annotated datasets, its reliance on big data renders it ineffective for the "long tail" of rare conditions. The core issue is that for a model to learn the subtle patterns of a rare disease, it needs examples, and those examples are inherently few.

What is Few-Shot Learning (FSL)?

Few-Shot Learning (FSL) emerges as a critical AI paradigm designed

specifically to overcome this data limitation. FSL is a machine learning approach that enables a model to make accurate predictions or classifications after being exposed to only a handful of labeled examples—sometimes even zero—of a new class. Instead of learning a disease from scratch with limited data, FSL models are trained to learn *how to learn* from prior knowledge, a concept known as **meta-learning**.

In the context of rare diseases, FSL models leverage information from common diseases or related biological knowledge to quickly adapt and generalize to a new, rare condition with minimal patient data. This shift from data-intensive learning to knowledge-intensive learning is the key to unlocking AI's potential in this underserved area of medicine.

SHEPHERD: A Knowledge-Guided FSL Approach

A prominent example of FSL's application in this domain is the **SHEPHERD** model, a few-shot learning approach developed for multi-faceted rare disease diagnosis [1]. SHEPHERD directly addresses data scarcity by integrating two powerful techniques: **knowledge graphs** and **simulated data**.

1. **Knowledge Graph Integration:** SHEPHERD performs deep learning over a **biomedical knowledge graph** that is richly populated with relationships between genes, phenotypes (observable characteristics), and diseases. This graph provides a structured, comprehensive map of biological and clinical knowledge. 2. **Graph Neural Networks (GNNs):** The model uses a **Graph Neural Network (GNN)** to represent each patient as a set of phenotype terms embedded within the knowledge graph's structure. The GNN jointly embeds the patient's phenotype subgraph and candidate genes/diseases into a latent representation space. This ensures that the patient's data is interpreted not in isolation, but in the context of established biomedical relationships, effectively placing patients near their causal genes or diseases in the embedding space. 3. **Simulated Training:** To provide the necessary volume for deep learning, SHEPHERD is trained on a large dataset of **simulated rare disease patients**. This adaptive simulation approach generates realistic patient profiles, allowing the model to learn robust diagnostic patterns even before encountering real-world labeled examples.

By leveraging the structural information of the knowledge graph and the scale of simulated data, SHEPHERD demonstrates that FSL can move beyond the limitations of traditional deep learning, offering a path to earlier and more accurate diagnosis for patients with rare genetic conditions.

The Future of Digital Health and Rare Disease Diagnosis

The success of few-shot learning in rare disease diagnosis marks a significant milestone in **digital health**. It proves that AI is not limited to areas with abundant data but can be strategically deployed to solve the most challenging problems in medicine, particularly those characterized by extreme data sparsity. As FSL techniques continue to evolve, they will likely be integrated into clinical decision support systems, empowering clinicians to make more informed and timely diagnoses. This advancement promises to shorten the diagnostic journey for millions of patients globally, transforming the landscape

of rare disease management.

For more in-depth analysis on the intersection of few-shot learning, AI, and digital health, the resources at [www.rasitdinc.com] (<https://www.rasitdinc.com>) provide expert commentary and cutting-edge research insights.

**

References

[1] Alsentzer, E., Li, M. M., Kobren, S. N., Noori, A., *Undiagnosed Diseases Network*, Kohane, I. S., & Zitnik, M. (2025). Few shot learning for phenotype-driven diagnosis of patients with rare genetic diseases. *npj Digital Medicine**, 8(1), 380. [URL: <https://www.nature.com/articles/s41746-025-01749-1>]

Rasit Dinc Digital Health & AI Research

<https://rasitdinc.com>

© 2023 Rasit Dinc