Building a Global Rare Disease Information Hub to Advance Knowledge, Diagnosis, and Treatment

Naveen Khetarpal, Dandan Wang and Ebenezer Ajao

M.S. in Biotechnology Management & Entrepreneurship

Prashant Soni

M.S. in Artificial Intelligence

FACULTY MENTORS

Rana Khan, Ph.D. and Youshan Zhang Ph.D.

ACKNOWLEDGEMENTS

Supported by Katz School Faculty Research grant to Drs. Khan and Zhang.

Introduction

- Rare diseases (RD) are defined as conditions affecting fewer than 1 in 2000 people in any WHO region[1]. Despite their rarity, there are over 7000 known RD, collectively impacting approximately 300 million people globally and ~30 million in the U.S[2]. Around 80% of RD have a genetic cause, 70% of which manifest in childhood. Current data suggests that a staggering 95% of RD lack approved treatments and the average time to accurate diagnosis is 4.8 years, with 30% of affected children dying before age five[1].

- While considerable research and information are available, it's region-specific, scattered, and unstructured and can take significant time and effort to extricate needed information for diagnosis and treatment. A centralized repository is urgently needed to streamline access to this critical data.

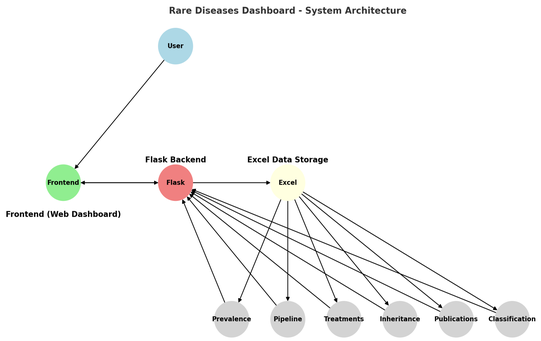
- **Aim:** Address an unmet need by developing an interactive and structured repository of clinical innovations in rare diseases. By consolidating RD-related information, the repository seeks to facilitate early diagnosis and treatment recommendations—two of the most significant challenges in this field[3].

Method

- This project focuses on approximately 4,000 of the most prevalent rare diseases (RDs). The methodology began with a comprehensive literature and internet search, gathering information on RD types, prevalence, geographic distribution, patient demographics, available treatments, therapies in development, and the companies involved in their development. This was followed by an in-depth scientific analysis of relevant databases, with Orphanet selected as a model due to the high quality and computability of its datasets available through the Orphadata platform.

- In addition, other reputable resources were leveraged, including the FDA's Orphan Drug Database, the FDA's Orange Book and PubMed. These sources contributed to the creation of a global RD information hub encompassing scientific classifications, symptoms, genetic variations, publications, available and emerging treatments.

Disease	Estimated prevalence (/100,000)
Down syndrome	95.0 BP
Cleft lip/palate	80.0 BP
Fetal and neonatal alloimmune thrombocytopenia	39.6307
Renal agenesis, unilateral	50.0 BP
Pneumonia caused by Pseudomonas aeruginosa infection	50.0 *
47,XYY syndrome	50.0 BP*
Iniencephaly	50.0 BP*
Congenital bilateral absence of vas deferens	50.0 *
Primary Sjögren syndrome	48.99 *
B-cell chronic lymphocytic leukemia	48.0



Results

- The collected data is organized into an Excel sheet, serving as the project's foundation.

- A dashboard has been successfully created through integrating the data sheet into a user-friendly web application accessible online.

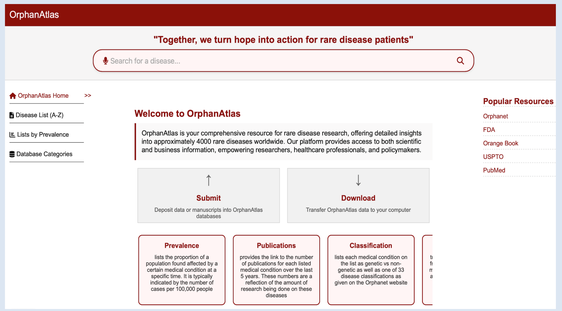
- This database will provide comprehensive insights, supporting:

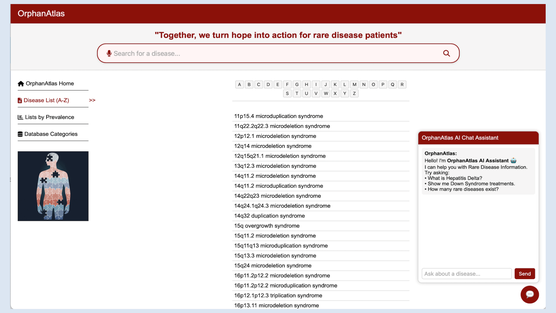
  - **Patients:** Early diagnosis and treatment options.

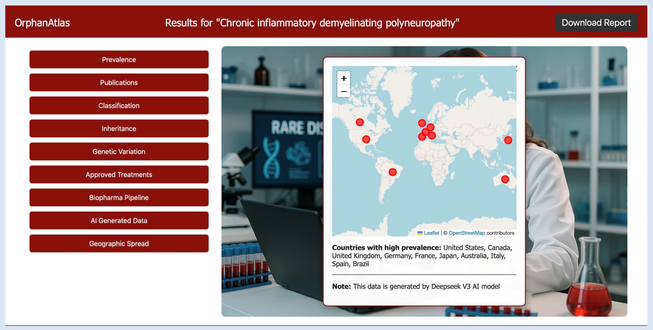
  - **Clinicians:** Ready access to detailed and validated data, aiding better clinical decisions.

  - **Researchers:** Accelerated discovery through consolidated data resources.

  - **Startups and Big Pharma:** Insights into competitive landscapes and strategic development opportunities.







Conclusions

- A standout feature of this project is that it brings together key information from across the RD landscape-both scientific and business-such as Biopharma pipelines, patents, and scientific publications to provide a clear and complete picture of how research and innovation are progressing in this critical area.

- The repository empowers stakeholders to have faster access to critical data, supporting early diagnosis and treatment planning.

- This initiative has the potential to transform care, improve patient outcomes, and accelerate advancements in research and treatment of RD.

Future Enhancements

- Implementing a new and better user authentication system for the personalized experiences and integrating AI/ML for predicting disease patterns and treatment suggestions.

- Expanding the database with more rare disease datasets and developing an API for third-party integrations.

References

1. The landscape for rare diseases in 2024. 2024. Lancet Glob Health. 12(3):e341. doi: 10.1016/S2214-109X(24)00056-1.

2. FDA. 2024. Rare Diseases at FDA. Accessed January 28, 2025. https://www.fda.gov/patients/rare-diseases-fda

3. IQVIA. 2023. Recent advancements in the rare disease landscape. Accessed January 28, 2025. https://www.iqvia.com/locations/united-kingdom/blogs/2023/02/recent-advancements-in-the-rare-disease-landscape

Katz School of Science and Health