Leveraging Large Language Models for Enhanced GWAS Analysis and Rare Disease Diagnosis

Project Description:

Rare diseases present a significant diagnostic challenge due to their diverse and often subtle symptoms, coupled with their low prevalence. Genome Wide Association Studies (GWAS) have identified associations between genetic variants and diseases, but pinpointing the causal genes, particularly in monogenic rare diseases, remains complex. This project leverages the power of Large Language Models to enhance GWAS analysis and improve rare disease diagnosis. Our approach explores two key aims: (1) Given a set of top pathogenic gene candidates identified through GWAS, predict the likely conditions associated with these genes, considering their complex interactions and potential pleiotropic effects, and (2) Develop an LLM-driven diagnostic framework that integrates diverse patient characteristics, including age, sex, detailed lab results, and clinical notes, to predict the most probable rare disease diagnosis. This approach aims to improve diagnostic accuracy and reduce the time to diagnosis for rare disease patients. We will validate our models using published case studies of rare diseases, comparing our performance to existing diagnostic methods.

Responsibilities:

1. Conduct literature review on LLMs for rare disease diagnosis, focusing on genomic data and clinical notes.
2. Develop and implement an LLM-driven diagnostic framework that provides potential diagnoses based on patient characteristics.
3. Validate the framework's performance against published rare disease case studies.

Required Skills:

1. Strong proficiency in Python and R, with experience in machine learning techniques and deep learning.
2. Knowledge of genomic data analysis and bioinformatics tools such as GWAS is a plus.
3. Strong research skills, with the ability to critically analyze scientific literature and contribute to innovative solutions.

Learning Opportunities:

Interns will gain exposure to the application of large language models (LLMs) in healthcare, particularly in rare disease diagnosis and genomics. They will work with real-world data, learning to develop AI models for disease prediction and diagnostic frameworks. Moreover, interns will be mentored by experiences scientists at QCRI, offering valuable professional development in both AI and biomedical research.

Team Structure: The team will consist of 2-3 interns working closely with their assigned research mentors.

Mentors: Dr. Mohamad Saad {msaad@hbku.edu.qa}