HelixRare - A Novel Approach to Match Patients with Clinical Trials

- **Team DeepDx**

Executive Summary:

This project addresses the critical challenge of efficient and accurate clinical trial participant matching for rare diseases. Our solution leverages Artificial Intelligence (AI), specifically a combination of rule-based clinical matching, variant novelty assessment, and a Large Language Model (LLM) for nuanced eligibility criteria evaluation. We have integrated publicly available genomic datasets (gnomAD and ClinVar) to enhance the model's accuracy. This memo details our approach, highlighting the novel aspects of our model compared to existing solutions.

Problem Statement and Context:

The rarity of rare diseases poses significant obstacles to clinical trial recruitment. Traditional methods often fail to capture the complexity of patient profiles. Existing matching solutions may lack the capability to handle nuanced eligibility criteria, **variant novelty**, and geographic considerations. Our project aims to develop an AI algorithm that matches potential clinical trial participants based on their medical history, demographics, and geographic location from electronic health records (EHRs) along with their **genomic data**.

Huntington's Disease (HD) Description:

It is primarily caused by a mutation in the *HTT* gene, which provides instructions for making the huntingtin protein. This mutation involves an abnormal expansion of a repetitive DNA segment known as a CAG trinucleotide repeat. In healthy individuals, this CAG repeat occurs a limited number of times, but in HD, it's repeated excessively (more than 35 times). This expansion leads to the production of an elongated, unstable huntingtin protein that accumulates and damages neurons, particularly in brain regions controlling movement, cognition, and emotion.

Methodology:

1) **Clinical Match Assessment:** We developed a rule-based system to evaluate basic clinical eligibility criteria, such as CAG repeat counts and active symptoms and their severity for Huntington's Disease (HD). This provides a foundational assessment of patient suitability.
2) **Variant Novelty Assessment:** This is a core innovation of our model. We integrated gnomAD and ClinVar datasets to assess the novelty of patient genetic variants. Specifically,

    i. Variants not found in our combined database are assigned a high novelty score (1.0) because identifying these variants is crucial for understanding disease mechanisms.

    ii. Variants found are further evaluated based on ClinVar classification ("Uncertain significance") and gnomAD allele frequency (extremely low).

    iii. This allows us to prioritize patients with rare or potentially impactful genetic variations, a critical aspect of rare disease trials.

3) **LLM-Driven Eligibility Evaluation:** We utilized the "*facebook/bart-large-mnli*" LLM via the Hugging Face Transformers library. The LLM is provided with a comprehensive prompt containing Patient medical history, demographics, trial eligibility criteria, location, and travel requirements.

4) **Final Scoring and Ranking:** A final score is calculated by combining the clinical match score, LLM output, and variant novelty score, with adjustable weights. Geographic considerations (patient location vs. trial location and travel preference) are also factored into the final score. Patients are ranked based on their final scores, providing a prioritized list for trial recruitment.

Existing Model:

TrialGPT is an AI model designed to match patients to clinical trials by analyzing patient medical records and trial eligibility criteria. It uses natural language processing (NLP) to understand both patient data and trial requirements, then calculates a matching score to rank potential candidates.

Novelty of Our Model:

1. **Comprehensive Variant Novelty Assessment:** Integrating gnomAD and ClinVar databases to quantify variant novelty is a unique contribution. This addresses the critical need to identify patients with rare or potentially impactful genetic variations. Assigning a full score for variants not in the databases allows the model to find very rare variants which have not been discovered yet.

2. **Integrated Geographic Considerations:** Our model explicitly incorporates patient location and travel preferences, ensuring that matches are not only clinically relevant but also practically feasible. For example, the model reduces the score of a patient, if they are not willing to travel to a trial that requires travel.

3. **Modular and Extensible Design:** The modular nature of our algorithm allows for easy adaptation to specific rare diseases and specific clinical trials. The scoring weights and eligibility criteria can be adjusted to reflect the specific requirements of each trial.

Evaluation and Results:

We simulated patient data for Huntington's Disease and evaluated our algorithm's performance on a sample clinical trial. The results demonstrate the model's ability to accurately identify potential trial participants based on clinical criteria, prioritize patients with novel genetic variants and provide a ranked list of patients.

Future Directions:

- Integrate EHR data for more comprehensive patient profiles.
- Expand the variant database to include other relevant databases.
- Develop a user interface for easy trial configuration and patient matching.
- Fine tune the LLM prompt or incorporate a better LLM, to improve accuracy.
- Seek expert opinions to refine and adjust the scoring weights

Conclusion:

Our AI-powered clinical trial matching algorithm offers a novel and effective solution for addressing the challenges of rare disease research. By combining rule-based clinical matching, variant novelty assessment, and LLM-driven eligibility evaluation, we have created a powerful tool that can accelerate clinical trial recruitment and ultimately improve patient outcomes.