Leveraging Large Language Model (LLM) for Recommending Treatment Options in Cancer Trials

Xingye Tan(xvt5070@psu.edu), Le Bao(lub14@psu.edu), Department of Statistics, The Pennsylvania State University

ABSTRACT

- **Goal**: Rare cancers present unique challenges, including small sample sizes, limited research resources, diagnostic difficulties, and restricted treatment options. To address these issues, our overarching project leverages historical treatment effects and recommendation systems powered by large language models (LLMs) to assist in treatment choices for rare cancers. Within this framework, this subproject specifically focuses on improving data processing techniques to enhance the accuracy and reliability of the recommendation system.
- **Method**: We utilized an LLM (specifically, Hugging Chat's Nous model) to efficiently collect and summarize historical data from common cancers. These data encompasses information on study arms, cancer types, treatment protocols, trial phases, and outcomes from over **20,000** clinical trials involving cancer patients.
- **Results**: Comparing the original treatment data from ClinicalTrials.gov with the treatment protocols extracted via the LLM, we increased the treatment matching accuracy from **63%** to **95%**. We also successfully standardized cancer type classifications.

INTRODUCTION

- To tackle the challenges of limited sample sizes, scarce research resources, and few treatment options for rare cancers, this project integrates large language models (LLM) and transfer learning techniques. By utilizing data from common cancers, transfer learning allows the recommendation system to improve rare cancer treatments without starting from scratch, leveraging existing knowledge from common cancers.

- We initially downloaded all clinical trial data from **ClinicalTrials.gov**, but it lacked some essential features. Using the **NCT ID** numbers, we employed **Hugging Chat's Nous model** to extract additional key information, I managed the processing of **Phase 1** data, covering over **20,000 clinical trials**, achieving a **92%** data extraction success rate.

- The data required extensive standardization and verification. I focused on two key tasks:
  - **Standardization of Cancer types**
  - **Verification of treatments using NLP techniques.**

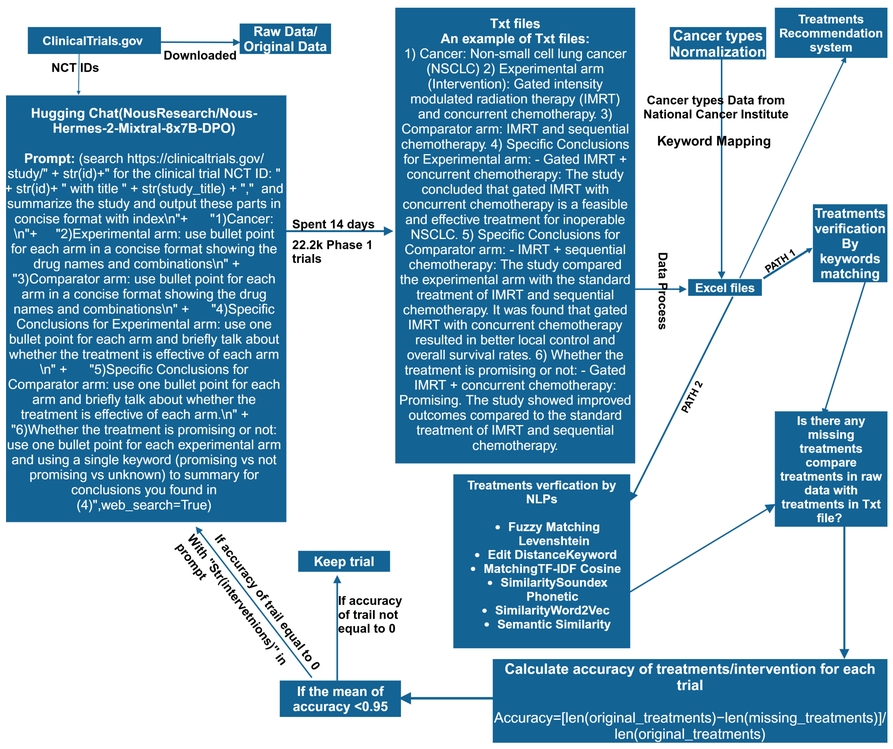
- By focusing on data standardization and verification, this subproject has significantly enhanced the system's reliability, providing more accurate and dependable support for treatment decisions in rare cancers.

- To clearly demonstrate the gaps in ClinicalTrials.gov's raw data and how they are addressed, we created a comparison between the raw data and the information provided by Hugging Chat's model.

	Columns	Hugging_Chat	Raw_Data
1	NCT.ID	Both	Both
2	Study.Title	Both	Both
3	Phases	Both	Both
4	Study.Status	Both	Both
5	Conditions	Both	Both
6	Study.Results	Both	Both
7	Experimental.arm	Hugging Chat	Both
8	Treatment/Interventions	Both	Both
9	Promising or Not	Hugging Chat	
10	Score	Hugging Chat	
11	Primary.Outcome.Measures		Raw Data
12	Study.Design		Raw Data
13	Locations		Raw Data

The table demonstrates how **Hugging Chat's model** provides additional columns that we need but are missing from the original **ClinicalTrials.gov** raw data. These include columns such as the experimental arm, whether the conclusion is promising, and the corresponding score for each trial. These columns further improve the reliability of rare cancer treatment recommendations.

METHODS



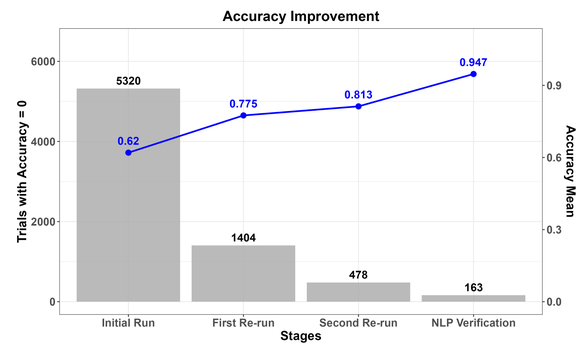
1. Initial Run (Path 1):
   - Processed data with Hugging Chat, achieving an accuracy mean of **0.62**.
   - **5,320** trials had accuracy = 0.

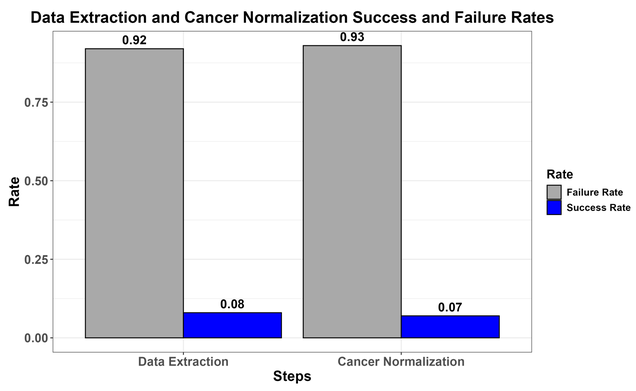
- First Re-run:
  - Reprocessed the 5,320 trials, improving accuracy mean to **0.775**.
  - **1,404** trials had accuracy = 0.

- Second Re-run:
  - Reprocessed the **1,404** trials, further improving accuracy mean to **0.813**.
  - **478** trials had accuracy = 0.

2. Switch to Path 2 (NLP Verification):
   - Since improving the **478** trials wouldn't raise overall accuracy to **0.95**, we applied NLP techniques.
   - This increased the final accuracy to **0.947**.
   - **163** trials had accuracy = 0.

RESULTS





- **Possible reasons for data extraction failures in the remaining trials**: The Hugging Chat model might have been unstable during execution, or the process may have taken too long, causing some trial requests to time out.
- **Possible reasons for NLP validation failures in the remaining trials**: Some treatments may have the same effects but use different names, making them difficult to match accurately.
- **Possible reasons for standardization failures in the remaining trials**: Some rare cancer types may not appear in our existing cancer list, making accurate standardization impossible.

SUMMARY & FUTURE WORK

- Successfully matched specific cancer types for 93% of the data using the cancer list from the National Cancer Institute website. **(Normalization)**
- Achieved a **92% success rate** in data extraction from over **20,000 clinical trials**.**(Extraction)**
- Improved treatment matching accuracy from an initial **62% to 94.7%.(Verification)**
- Re-examine **missed trials during data extraction** by conducting a thorough search using a flowchart approach.
- Continue validating **treatment accuracy** for **Phase 2, 3, and 4** clinical trials.
- Finalize and confirm the **model for the recommendation system** to optimize treatment suggestions.

REFERENCES

- https://www.cancer.gov/types/common-cancers
- https://www.medicalnewstoday.com/articles/rare-cancers
- https://www.cancer.gov/types