Using Mixture-of-Experts for ClinicalT5 to Improve Classification of Rare ICD Codes

Yelena Yu

Introduction

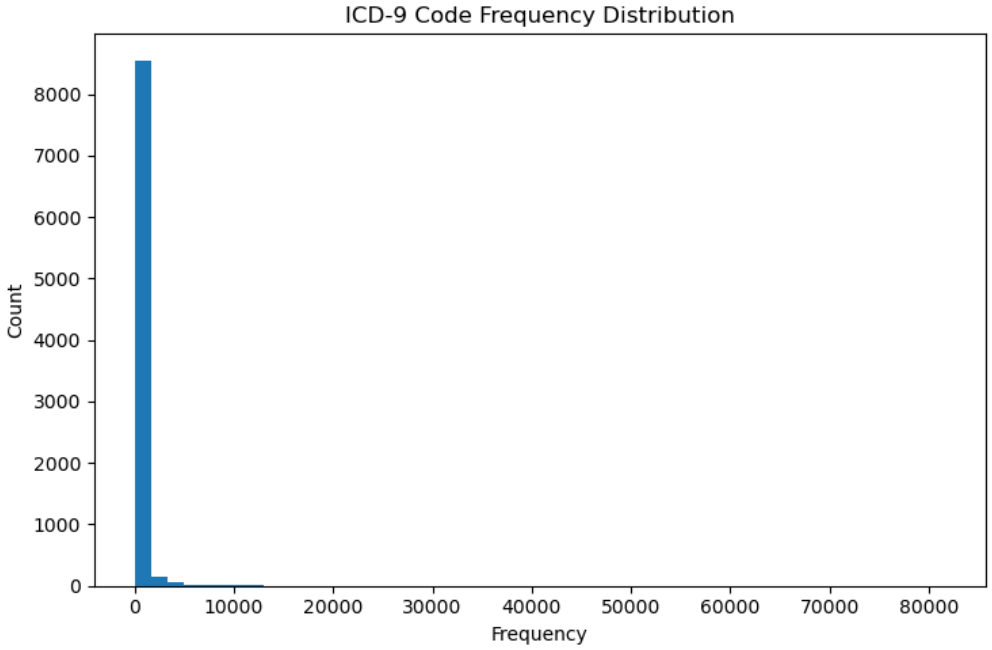
- ICD is a standard coding system for disease diagnosis
  - Used for record-keeping, plays an important role in patient care

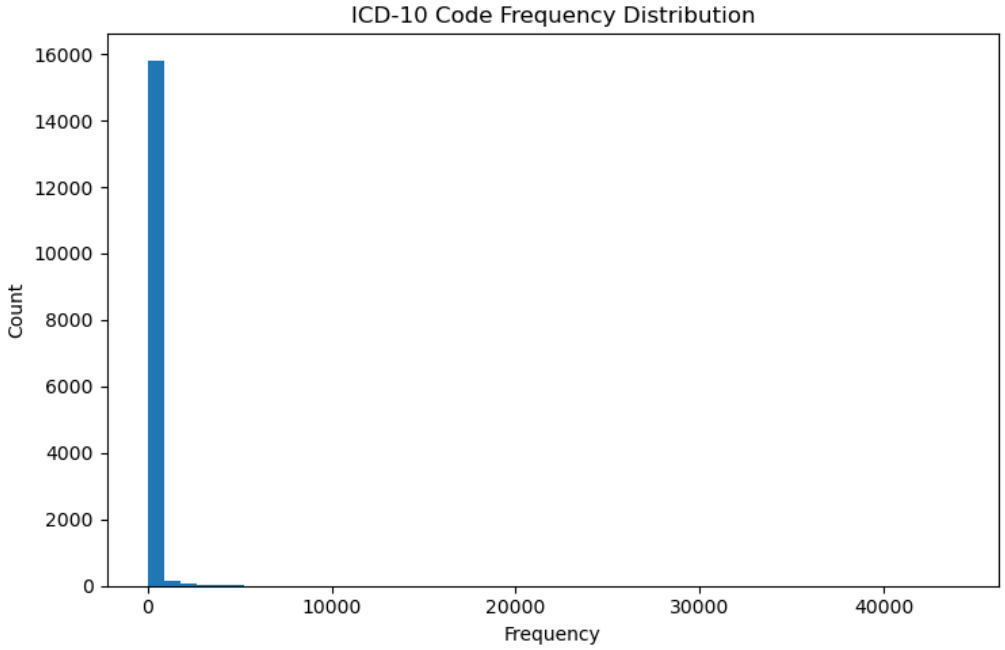
- A common procedure is assigning ICD codes to clinical notes to structure unstructured data
  - LLMs generate relevant ICD codes based on clinical text (ClinicalT5, MedBERT, etc)

- Intrinsically biased: In training datasets, common ICD codes appear more often than rare ones

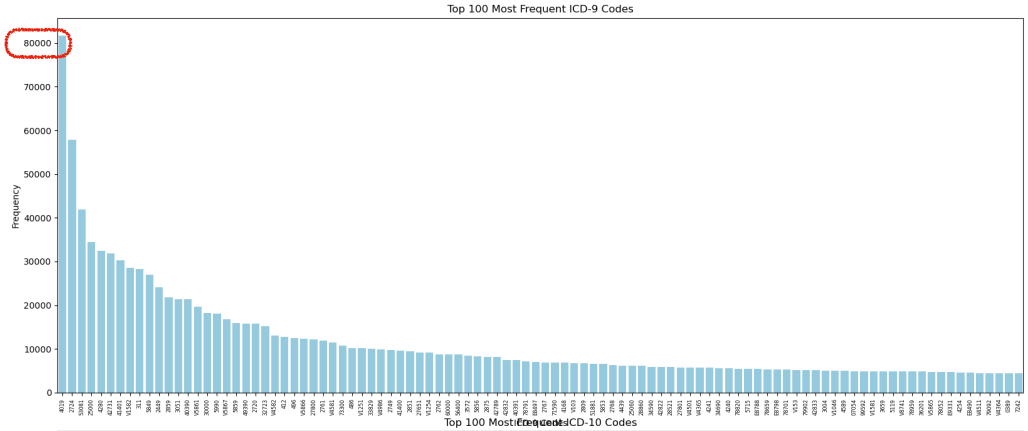
Rare ICD codes are less likely to be assigned by LLMs, leading to misdiagnosis.

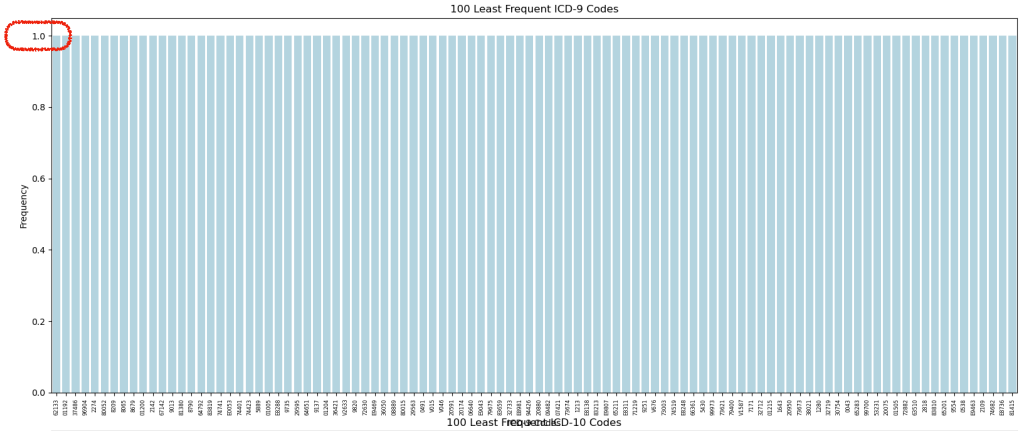
Problem Statement I Data Perspective

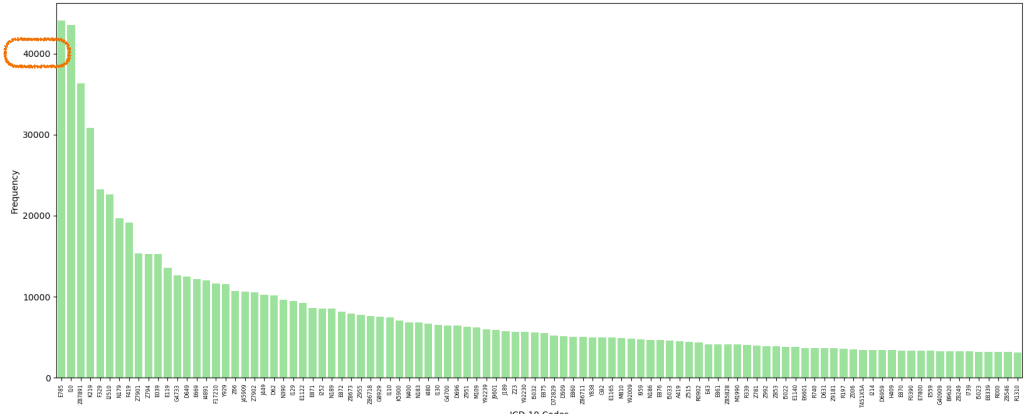


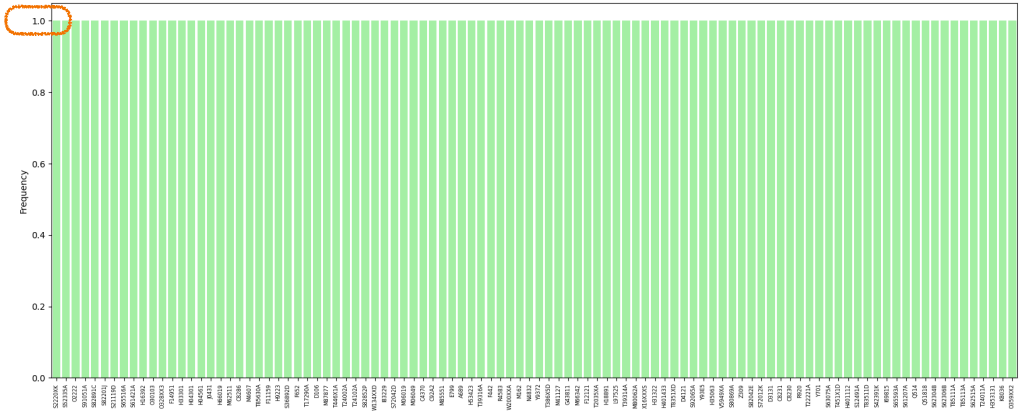


Problem Statement I Data Perspective









The disparity in frequency of common and rare ICD codes is very severe.
Over-representation of common codes lead to misdiagnosis of rare diseases.

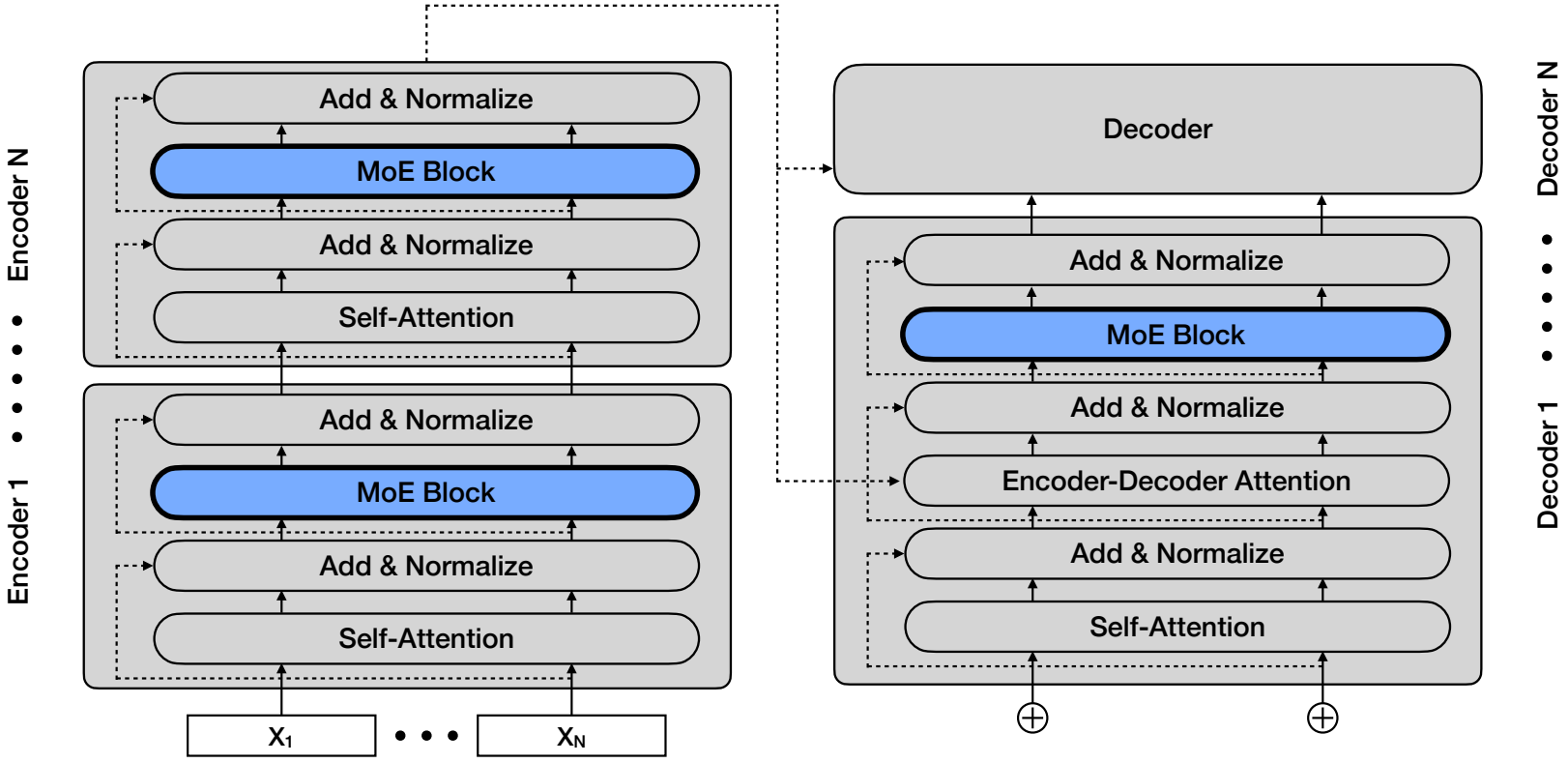
Problem Statement I Model Perspective

- A singular LLM is applied to an extremely diverse set of ICD codes
  - Patient diagnoses: cardiovascular, diabetes, cancer, sepsis, etc
  - Scarcity: Malformation of Coronary Vessels (rare), Essential Hypertension (common)

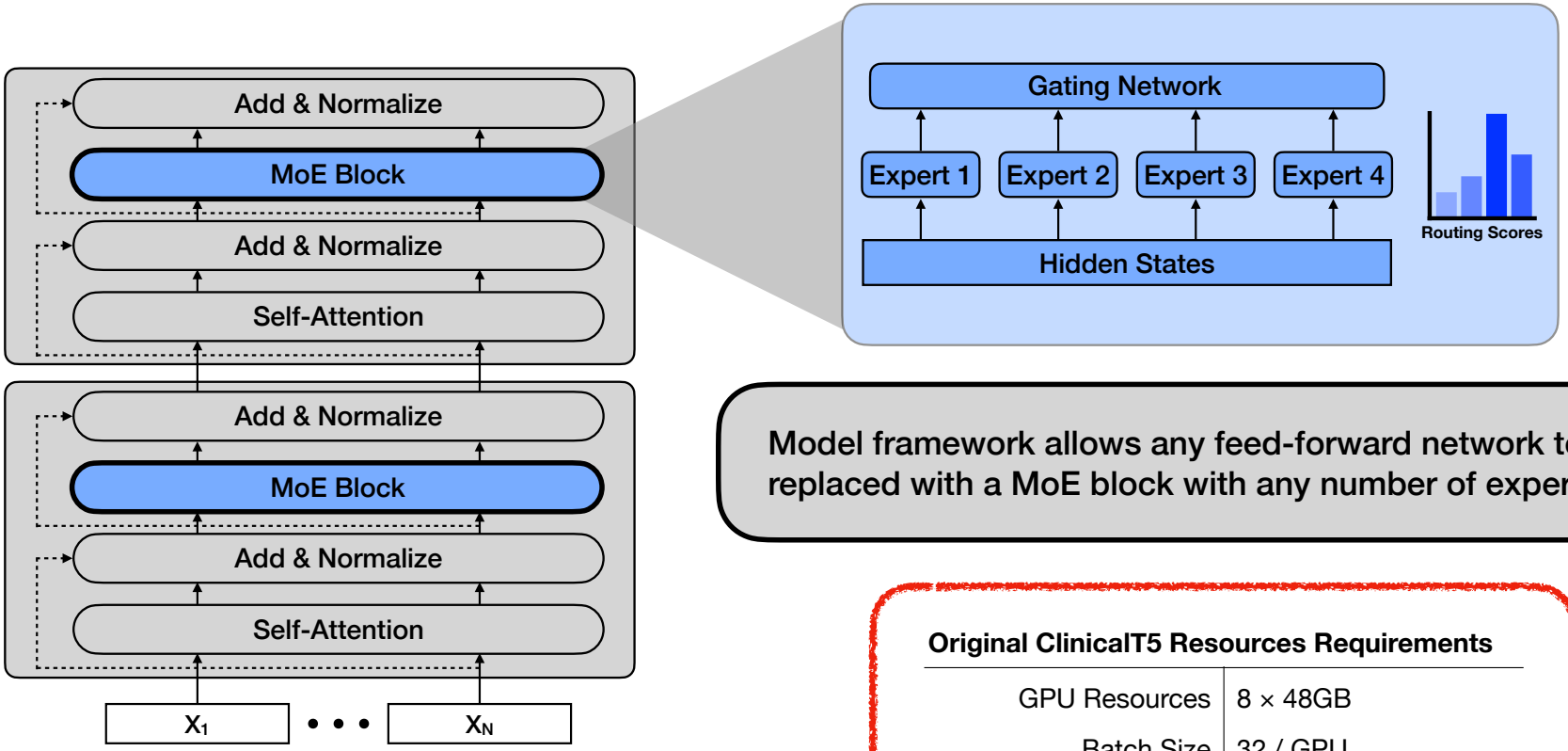
- Instead of large aggregated models, LLMs need some degree of expertise for specialized diagnosis
  - Idea: Partition specialized components tailored to different types of codes
  - Experts specialized in handling different codes, experts to handle rare codes

A mixture-of-expert architecture allows a LLM to achieve the required specialization.

Solution I LLM Architecture



Solution I MoE Architecture



Model framework allows any feed-forward network to be replaced with a MoE block with any number of experts.

Original ClinicalT5 Resources Requirements

GPU Resources	8 × 48GB
Batch Size	32 / GPU
Sequence Length	512

Summary

- Current LLM-based ICD code assignment is biased towards common ICD codes.
  - Leads to misdiagnosis and patient care for rare medical conditions.

- MoE allows the LLM to develop the required specialization to better handle ICD codes
  - Experts specialized in handling different codes, experts to handle rare codes).

- As part of this project:
  - MoE and sparse MoE blocks were developed based on research papers
  - A framework was developed to allow any feed-forward layer in the encoder and decoder to be replaced with an MoE block, with any number of experts.

Thank You