CLIMB: Data Foundations for Large Scale Multimodal Clinical Foundation Models

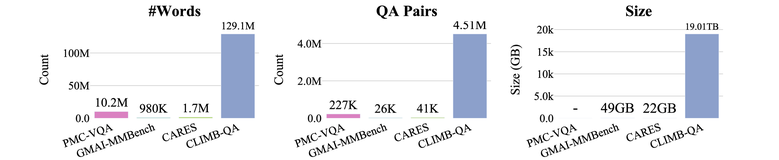
Wei Dai [1] Peilin Chen [1] Malinda Lu [1] Daniel Li [1] Haowen Wei [1,2] Hejie Cui [3] Paul Pu Liang [1]

[1]Massachusetts Institute of Technology [2]Harvard Medical School [3]Stanford University

Dataset Paper Model Paper

We collected *CLIMB*, one of the largest multimodal clinical dataset across 15 modalities. Models trained on CLIMB achieve up to 32.54% AUC improvements over previous SoTA.

We introduce CLIMB (Clinical Large-scale Integrative Multi-modal Benchmark), a comprehensive dataset unifying **4.51M samples across 44 datasets, totaling 19.01 TB**. Multitask pretraining on CLIMB significantly improves performance by up to 32.54% especially for understudied modalities, enables strong few-shot transfer to novel tasks, and enhances multimodal fusion.



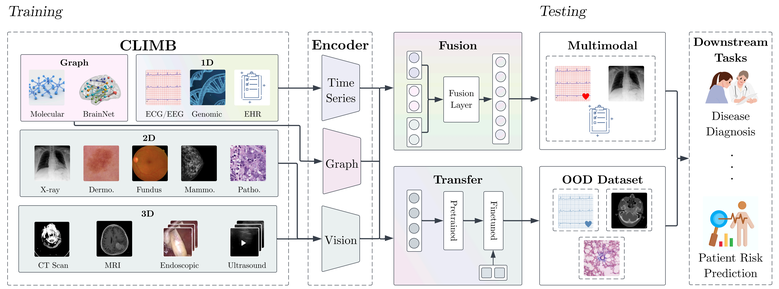
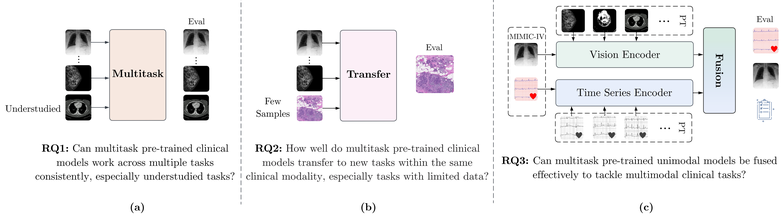


Figure 1. CLIMB unifies diverse clinical data across 15 modalities including imaging (2D/3D), time series (EEG/ECG), graphs, and multimodal combinations.

Experimental Setups

* We start from current pretrained SoTA, then train the model on CLIMB with the following tasks:



RQ1: Can multitask pre-trained clinical models work across multiple tasks consistently, especially understudied tasks?

(a)

RQ2: How well do multitask pre-trained clinical models transfer to new tasks within the same clinical modality, especially tasks with limited data?

(b)

RQ3: Can multitask pre-trained unimodal models be fused effectively to tackle multimodal clinical tasks?

(c)

CLIMB is the current largest and most diverse public clinical diagnosis dataset. You can download the data easily via our downloader on Github.

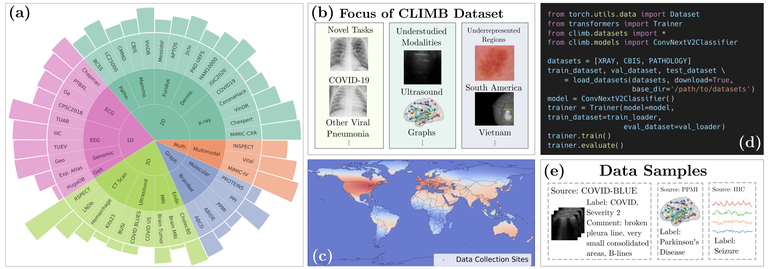


Figure 2. (a) Visualization of CLIMB dataset composition. (b) Focus of dataset collection. (c) Distribution of data collection sites in CLIMB. (d) Example code usage on CLIMB framework. (e) Sample data from CLIMB.

Experimental Results

Multitask pretraining significantly improves performance across tasks, achieving up to **32.54%** AUC improvement in understudied areas.

Few-shot performance on models trained on CLIMB demonstrate significant improvements, achieving up to **29% improvement**.

Multimodal fusion: Single-modality pretraining on CLIMB enhances multimodal learning performance, leading to successful transfer to MIMIC-IV.

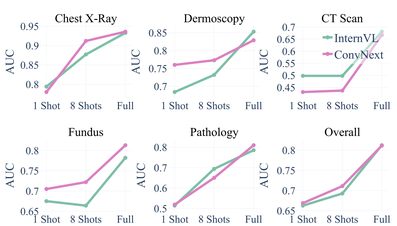


Figure 4. Few-shot performance of models across different pretraining (PT) datasets.

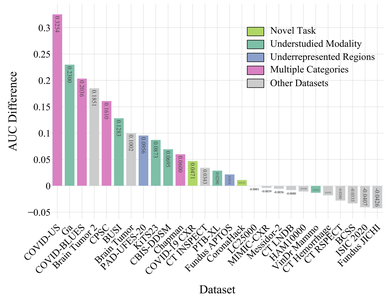
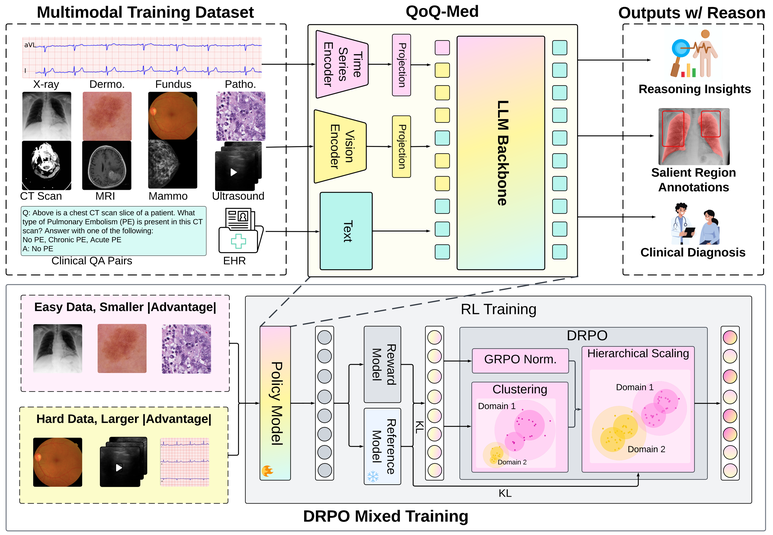


Figure 3. Difference in AUC achieved by the multitask model compared to single-task training.

Below describes our preprint

QoQ-Med: Building Multimodal Clinical Foundation Models with Domain-Aware GRPO Training



Challenge: In multimodal training, frequent modalities and easy samples dominates training.
Idea: Hierarchically clusters samples during training, upscaling rewards from rare and hard domains.

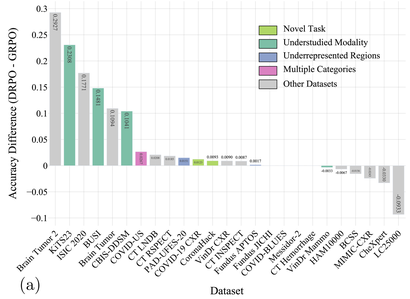
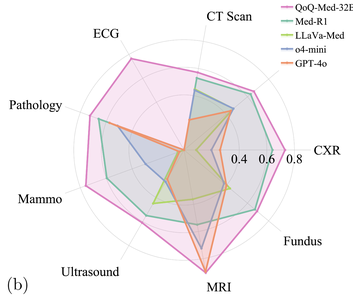


Figure 5. (a) DRPO performance vs GRPO.



(b) QoQ-Med vs other open/closed source models