Self-Service Data Science in the EHR with Multimodal Patient Embeddings

Jason A. Fries[1], Keith Morse[2]

Ethan Steinberg[1], Michael Wornow[1], Rahul Thapa[1], Zepeng Huo[1], Shih-Cheng Huang[1], Lawrence Guo[3], Lillian Sung[3], Nigam H. Shah[1,2]

[1] Stanford University, [2]Stanford Health Care, [3]The Hospital for Sick Children *jfries@stanford.edu, KMorse@stanfordchildrens.org*

Stanford MEDICINE School of Medicine

AIMI HAI

AIMI-HAI Partnership Grant

HAI FIVE

Celebrating 5 Years of Impact

MOTIVATION

Foundation models are common for text and imaging, but not **electronic health records (EHRs).**

Is this patient most likely to get gastrointestinal cancer in 30 years, 10 years or 3 years?

Patient classification tasks are also common in healthcare AI, but **EHR data is complex to parse and difficult to featurize.**

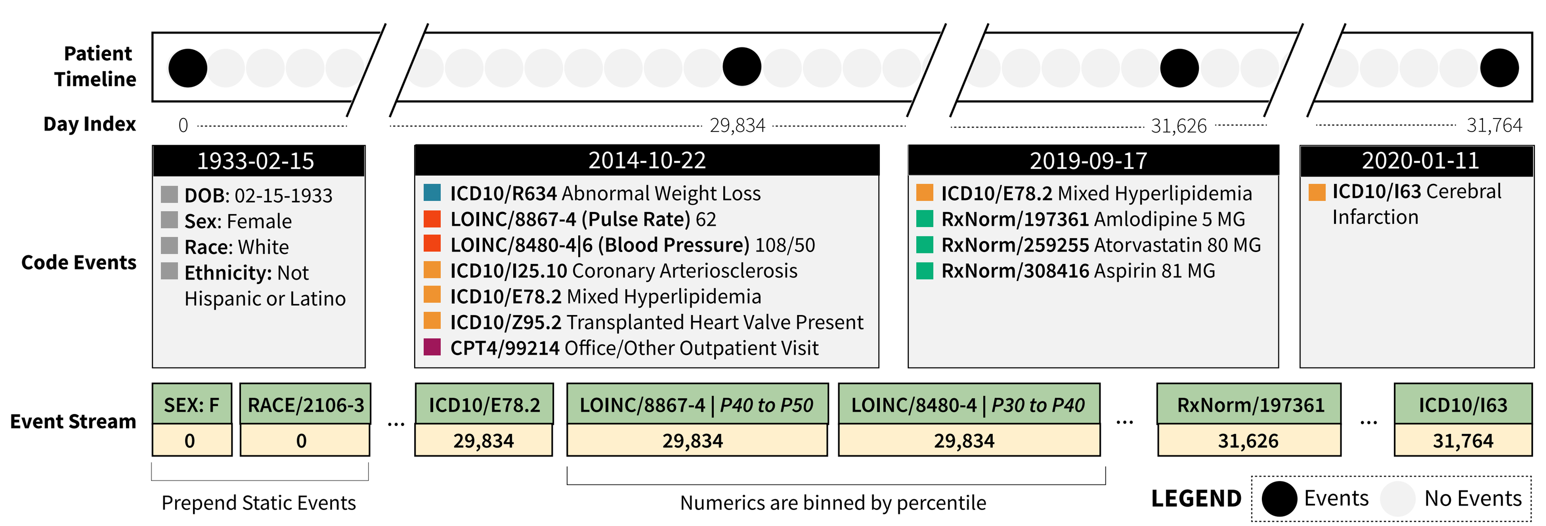
Identify all patients who have evidence of pulmonary hypertension.

We developed a publically accessible patient embedding model trained on millions of patients' longitudinal EHRs.

TRAINING DATA

Patient Timelines

Instead of viewing a patient's data as tabular, we represent as an ***event stream*** a sequence of discrete events ordered by time. Here multimodal events are transformed into tokens drawn from a finite symbol vocabulary.



Pretraining Data

We used self-supervised learning to pretrain on the entire patient population present in **Stanford Health Care** and **Stanford Children's Health.**

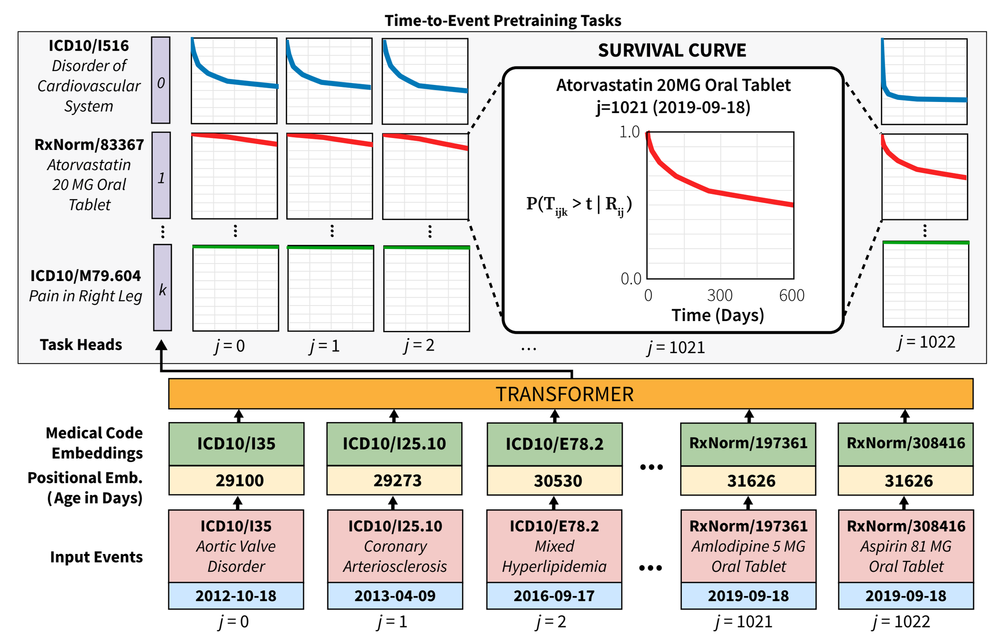
STANFORD STARR-OMOP (EHR)

2.7M Patients

4.5B Events

METHODS

MOTOR: Time-to-Event (TTE) Pretraining



Our Goal Estimate the probability distribution of event times $P(T_i = t)$ accounting for censoring

Pretrain on **8192 TTE tasks**

Max timeline length of **16,384 events**

EXPERIMENTS & RESULTS

Baseline Comparisons

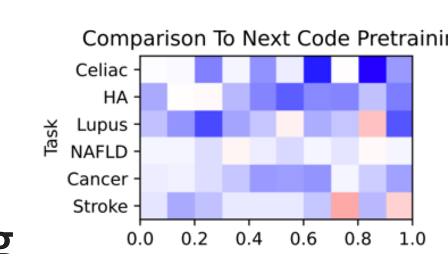
Method	Dataset	Celiac	HA	Lupus	NAFLD	Cancer	Stroke
Cox PH	EHR-OMOP	0.689	0.761	0.770	0.726	0.793	0.779
DeepSurv	-	0.704	0.823	0.790	0.800	0.811	0.830
DSM	-	0.707	0.828	0.784	0.805	0.809	0.835
DeepHit	-	0.695	0.826	0.807	0.805	0.809	0.833
RSF	-	0.729	0.836	0.787	0.802	0.824	0.840
MOTOR-Scratch	-	0.696	0.795	0.803	0.821	0.777	0.831
MOTOR-Probe	-	0.802	0.884	0.850	0.859	0.865	0.874
MOTOR-Finetune	-	**0.802**	**0.887**	**0.863**	**0.864**	**0.865**	**0.875**

Average 4.6% improvement in C-statistics over SOTA

Comparing Pretraining Objectives

Objective	Celiac	HA	Lupus	NAFLD	Cancer	Stroke
Next Code	0.774	0.862	0.842	0.860	0.860	0.857
Time-to-Event	**0.802**	**0.887**	**0.863**	**0.864**	**0.865**	**0.875**

Outperforms autoregressive pretraining



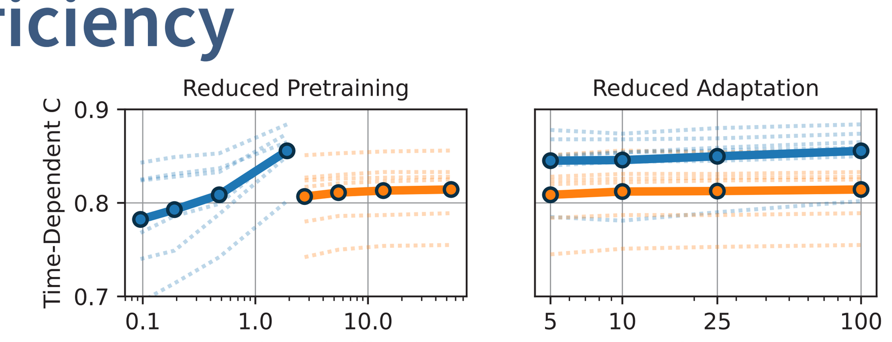
Better on Longer Time Horizons

Temporal Robustness & Cross-Site Adaptability

Method	Dataset	Celiac	HA	Lupus	NAFLD	Cancer	Stroke
Best Baseline	Out-of-time STARR-OMOP	0.682	0.776	0.810	0.750	0.786	0.768
Best MOTOR	-	**0.792**	**0.843**	**0.880**	**0.809**	**0.870**	**0.836**
Best Baseline	MIMIC-IV	0.625	0.820	0.807	0.736	0.748	0.780
Best MOTOR	-	**0.628**	**0.850**	**0.819**	**0.802**	**0.828**	**0.812**

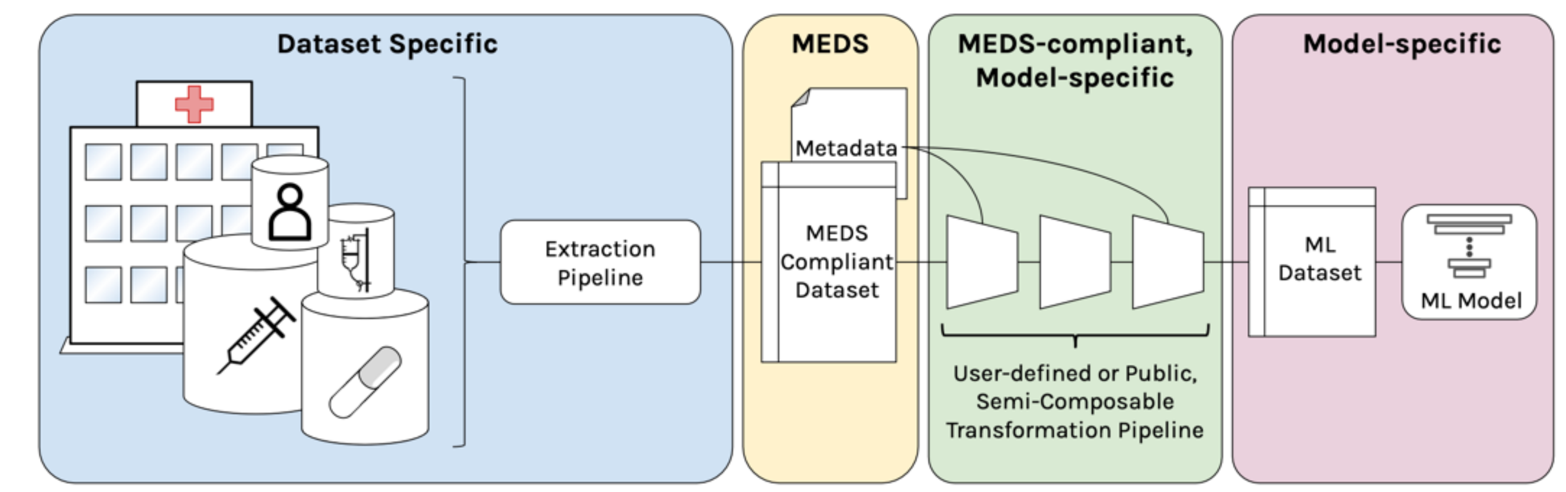
Sample & Label Efficiency

Requires 95% fewer labeled examples to match SOTA.



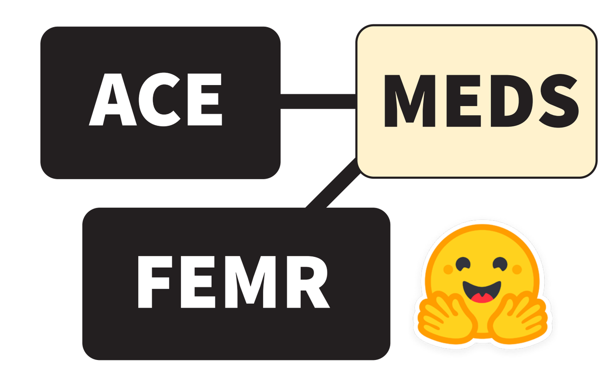
SHARED RESOURCES

Medical Event Data Standard (MEDS)



MEDS is a proposed data standard for representing EHR data and model inputs for sharing pretrained EHR models.

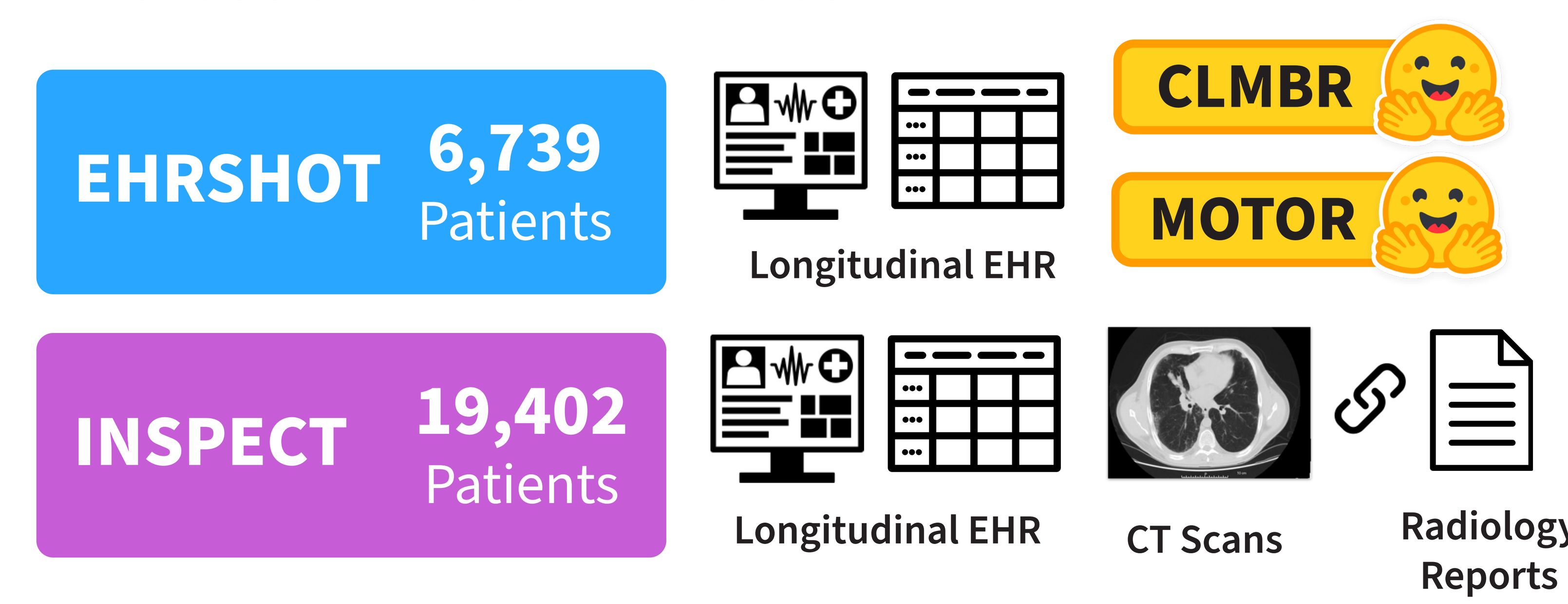
Data-Centric EHR Curation



Advanced Cohort Engine (ACE): Patient search engine with temporal query processing. Exports MEDS formatted data.

FEMR: Framework for curating EHR timelines for use in ML workflows with EHR foundation models.

Dataset & Model Releases



Papers

1. Guo et al. **A Multi-Center Study on the Adaptability of a Shared Foundation Model for Electronic Health Records.** *npj Digital Medicine 2024*
2. Steinberg et al. **MOTOR: A Time-To-Event Foundation Model For Structured Medical Records.** *ICLR 2024.*
3. MEDS Working Group. **Medical Event Data Standard (MEDS): Facilitating Machine Learning for Health.** *TS4H@ICLR2024*
4. Huang et al. **INSPECT: A Multimodal Dataset for Patient Outcome Prediction of Pulmonary Embolisms.** *NeurIPS D&B 2023.*
5. Wornow et al. **EHRSHOT: An EHR Benchmark for Few-Shot evaluation of Foundation Models.** *NeurIPS D&B 2023.*