(Frank) Shuyu Lu

Ph.D. in Biomedical Informatics | (346)503-5197 | shuyu.lu@uth.tmc.edu | Houston, TX
Publications | LinkedIn

Summary

Biomedical informatics Ph.D. student working on machine learning, large language model reasoning, multimodal learning, and longitudinal representation learning. My research focuses on developing reliable, interpretable AI methods for complex real-world and biomedical data, including structured EHR, clinical narratives, behavioral signals, and longitudinal patient trajectories.

Education

Ph.D. in Biomedical Informatics | McWilliams School of Biomedical Informatics, UT Health Houston — 08/2024 - Present
GPA 3.75/4.0 | Advisor: Hongfang Liu | Anticipated Graduation: 11/2028 — Houston, TX
- Research Focus: Real-world clinical data modeling and representation learning.

Master of Science in Biomedical Informatics | University of Pittsburgh — 08/2019 - 05/2021
GPA 3.75/4.0 | Advisor: Xinghua Lu — Pittsburgh, PA

Bachelor of Engineering in Computer Science | Central China Normal University — 09/2015 - 06/2019
GPA 3.45/4.0 | Merit Student (2016-2017, 2017-2018) — Wuhan, Hubei, China
- RoboCup China Open 2016, Humanoid Dancing Robot category - First Prize (No. 3)

Bachelor of Art in French (Minor) — 03/2017 - 06/2019
Wuhan University | GPA 3.8/4.0 — Wuhan, Hubei, China

Research Experience

Real-World EHR Modeling of Obesity and Anti-Obesity Medications — 09/2024 - Present, Houston, TX
- Conducted real-world EHR studies on obesity phenotyping, anti-obesity medication effectiveness, adverse events, and longitudinal treatment outcomes across large clinical populations.
- Applied GRU-D autoencoders and clustering methods to pretreatment longitudinal EHR data to identify latent obesity phenotypes and characterize clinically distinct patient subgroups.
- Performed cross-institutional longitudinal analyses of GLP-1/GIP receptor agonists, evaluating weight-loss trajectories, treatment discontinuation, and temporal adverse-event patterns to identify heterogeneous treatment response and tolerability profiles.
- Exploring LLM-based clinical narrative modeling and multi-agent frameworks to integrate unstructured notes with structured EHR data for patient phenotyping and personalized treatment analysis.

Temporal and Multimodal Perioperative Trajectory Modeling for Postoperative Outcomes Prediction — 11/2024 - Present, Houston, TX
- Constructed longitudinal perioperative cohorts integrating preoperative, intraoperative, and postoperative EHR data to support dynamic modeling of postoperative complications.
- Investigated postoperative organ injuries (POIs) across surgical populations and applied multivariable logistic regression and Kaplan–Meier survival analysis to characterize risk factors and their associations with clinical outcomes and hospital stay.
- Developing temporal and multimodal modeling strategies for individualized postoperative risk prediction, with the long-term goal of enabling continuous rather than static perioperative risk assessment.

Multimodal Learning for Cognitive Phenotyping — 05/2025 - 08/2025
Algorithm Research Scientist | iFLYTEK Health — Beijing, China
- Developed a multimodal learning pipeline integrating speech, language, memory, and handwriting signals from 1,166 participants for community-based cognitive risk modeling (89.3% test accuracy; 77.5% F1).
- Applied SHAP and Gaussian mixture modeling for interpretable cognitive phenotyping and examined associations with rs-fMRI features in 83 participants; contributed to an attention-based multimodal fusion patent and research proposals.

Heart Failure EHR Mining and Predictive Modeling — 10/2019 - 02/2021, Pittsburgh, PA
- Developed interpretable machine learning models using structured and unstructured EHR data to identify heart-failure readmission risk and clinically relevant predictors.
- Applied NLP, tree-based machine learning, and SHAP-based interpretation to characterize important clinical features and uncover latent patient-level patterns.
- Investigated data-driven approaches for early risk stratification and personalized management, contributing to subsequent peer-reviewed work on interpretable heart-failure prediction.

Selected Publications

Tracking the Dynamic Trajectories: A Global-to-Local Pharmacovigilance Analysis of GLP-1 Receptor Agonists
- *First Author. Accepted by AMIA Annu Symp Proc. 2026.*

A Dynamic Risk Prediction Framework for Alzheimer's Disease and Related Dementias with Interpretability
- *Ruan, X., L**u, S.**, Fu, S. et al. A dynamic risk prediction framework for Alzheimer's disease and related dementias with interpretability. npj Digit. Med. **9**, 446 (2026). https://doi.org/10.1038/s41746-026-02732-0*

Deep Phenotyping of Obesity: Electronic Health Record–Based Temporal Modeling Study
- *Ruan X, **Lu S**, Wang L, Wen A, Murali S, Liu H. Deep Phenotyping of Obesity: Electronic Health Record–Based Temporal Modeling Study. J Med Internet Res 2025;27:e70140. URL: https://www.jmir.org/2025/1/e70140. DOI: 10.2196/70140.*

Understanding Heart Failure Patients EHR Clinical Features via SHAP Interpretation of Tree-Based Machine Learning Model Predictions
- ***Lu S**, Chen R, Wei W, Belovsky M, Lu X. Understanding Heart Failure Patients EHR Clinical Features via SHAP Interpretation of Tree-Based Machine Learning Model Predictions. AMIA Annu Symp Proc. 2022 Feb 21;2021:813- 822. PMID: 35308970; PMCID: PMC8861751.*

Current Status of Anti-obesity Medications and Performance, an EHR based Survey

- *Ruan, X., Li, R., Wang, L. et al. Current status of anti-obesity medications and performance, an EHR based survey. BMC Endocr Disord (2026). https://doi.org/10.1186/s12902-026-02423-7*

Evaluating FFT-based Algorithms for Strided Convolutions on ARMv8 Architectures

- *HUANG X D, WANG Q L, **LU S Y**, et al.Evaluating FFT-based algorithms for strided convolutions on ARMv8 architectures[J].Performance Evaluation, 2021,152:102248*

NUMA-aware FFT-based Convolution on ARMv8 Many-core CPUs

- *X. Huang, Q. Wang, **S. Lu,** R. Hao, S. Mei and J. Liu, "NUMA-aware FFT-based Convolution on ARMv8 Many-core CPUs," 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), New York City, NY, USA, 2021, pp. 1019-1026, doi: 10.1109/ISPA-BDCloudSocialCom-SustainCom52081.2021.00142*

Technical Skills

- **Programming:** Python, SQL, JavaScript/TypeScript, R, Django, Vue, React, Node.js, Tailwind CSS
- **Generative AI & LLM Systems:** Prompt Engineering, agentic workflows, multi-agent system
- **Database & Analytics:** PostgreSQL, MySQL, SQL Server, MongoDB, Spark, Tableau
- **Machine Learning & AI:** PyTorch, Scikit-Learn, multimodal learning, representation learning, clustering
- **Clinical Database / Data Structure:** Epic Cosmos, OMOP CDM, IQVIA

Additional Experience

Research Assistant 03/2023 - 06/2024

School of Public Health -- Zhejiang University, Healthy Zhejiang One Million People Cohort Hangzhou, Zhejiang, China

- Contributed to the Healthy Zhejiang One Million People Cohort, a large-scale digital public health initiative integrating population, environmental, clinical, and health-related data to support disease prevention and precision health research.
- Developed a full-stack lung-cancer risk assessment application using Vue, Tailwind CSS, and Django, translating an established statistical model into a user-facing healthcare tool.
- Performed clinical data preprocessing, quality control, and database management to support cohort-based epidemiological analyses and downstream modeling.

Software Development Engineer 09/2021 - 03/2023

China Electronics Technology Group Corporation- 38th Research Institute Hefei, Anhui, China

- Developed front-end components for two enterprise web applications, including a data-center management system and a distributed data management platform deployed across the institute.
- Standardized workflows for extracting, cleaning, aggregating, and transforming operational database records into analysis-ready outputs for internal reporting and workforce analytics.

Languages

Chinese -- Native | English -- Professional | French -- Basic