Star Liu

U.S. Citizen | Personal Website | github.com/StarLiu1 | Latest CV | Google Scholar | sliu197@jhmi.edu

Education

2023-Present **Johns Hopkins School of Medicine (JHSOM)**
Degree: PhD in Biomedical Informatics and Data Science | Mentor : Harold P. Lehmann
Dissertation: *Aligning Predictive Models With Clinical Decision Making via Elicitation, Calibration, and Threshold Selection*

2021-2023 **JHSOM**
Degree: Master of Science in Biomedical Informatics and Data Science Research | Mentor : Harold P. Lehmann
Thesis: *Model Usefulness : Aligning Utility and Prevalence in Use and in Training of Clinical Machine Learning Models*

2017-2021 **Emory University**
Degree: Bachelor of Science in Quantitative Sciences, Biology Concentration | Mentor : Seunghwa Rho
Capstone: *Building a Predictive Model for Fire Incident Risk in Atlanta*

Competencies

Programming & ML: Python (PyTorch, pandas, numpy, scikit-learn) • R • SQL • Deep Learning • LangChain • Azure
Health Economics: Markov Models, Decision Trees, TreeAge
Databases: Microsoft SQL Server, CosmosDB, MongoDB
Vocabularies: ICD-9, ICD-10, RxNorm, LOINC, SNOMED-CT, CPT
Data Sources: EHR, Claims, Healthcare Cost and Utilization Project, Medical Expenditure Panel Survey
Softwares: GitHub, Overleaf, Microsoft Excel, Slurm, FileZilla, Tableau, QGIS, Google Analytics

Publications and Preprints

Published - Journal and Conference

1. Sood PD, **Liu S**, Kalyani RR, Kharrazi H. Comparing computable type 2 diabetes phenotype definitions in identifying populations of interest for clinical research. ***BMJ Health Care Inform***. 2025. PMID : 41125310
2. Barrett R, Lawler B, Liu S, Park WY, Davoodi M, Martin B, Kalyanam SM, Makker K, Kuiper JR, Aziz KB. Transforming neonatal care through informatics : A review of artificial intelligence, data, and implementation considerations. ***Semin Perinatol***. 2025. PMID : 40973564
3. Sood PD, **Liu S**, Pandya C, Kharrazi H. Assessing the Impact of Computable Type 2 Diabetes Phenotypes on Predicting Healthcare Utilization Using Electronic Health Records and Administrative Claims. ***Healthcare (Basel)***. 2025 PMID : 41008424
4. Sood PD, **Liu S**, et al. Measuring the Impact of Data Quality and Computable Phenotypes on Potential Racial Disparities in Predicting Healthcare Utilization Among Type 2 Diabetes Populations. ***Racial and Ethnic Health Disparities***. 2025 PMID : 40425977
5. Al Gharaibeh FN, **Liu S**, Wynn JL, Aziz KB. The utility of neonatal sequential organ failure assessment in mortality risk in all neonates with suspected late-onset infection. ***J Perinatol***. 2025. PMID : 40251303
6. Sood PD, **Liu S**, Lehmann HP, Kharrazi H. Assessing the Effect of Electronic Health Record Data Quality on Identifying Patients with Type 2 Diabetes. ***JMIR Medical Informatics***. 2024. PMID : 39189917
7. **Liu S**, Golozar A, Buesgens N, McLeggon JA, Black A, Nagy P. A framework for understanding an open scientific community using automated harvesting of public artifacts. ***JAMIA Open***. 2024. PMID : 38425704
8. Hassan S, **Liu S**, Patel SA, Emmert-Fees K, Suvada K, Tandon N, Sosale A, Anjana RM, Mohan V, Chwastiak L, and Ali MK. Association of collaborative care intervention features with depression and metabolic outcomes in the INDEPENDENT study : A mixed methods study. ***Prim Care Diabetes***. 2024. PMID : 38360505
9. **Liu S**, Wei S, Lehmann HP. Applicability Area : A novel utility-based approach for evaluating predictive models, beyond discrimination. ***AMIA Annu Symp Proc***. 2024. PMID : 38222359
10. Benameur K, Gross PJ, **Liu S**, Koneru S. The Mediterranean Soul Food Diet Intervention After Stroke, a Feasibility Study in the Stroke Belt. ***American Journal of Lifestyle Medicine***. 2023. Sage Journal Article : 10.1177/15598276231208383
11. **Liu S**, Ding X, Belouali A, Bai H, Raja K, Kharrazi H. Assessing the Racial and Socioeconomic Disparities in Postpartum Depression Using Population-Level Hospital Discharge Data : Longitudinal Retrospective Study. ***JMIR Pediatrics Parenting***. 2022. PMID : 36103575
12. Belouali A, Bai H, Raja K, **Liu S**, Ding X, Kharrazi H. Impact of Social Determinants of Health on Improving the LACE Index for 30-Day Unplanned Readmission Prediction. ***JAMIA Open***. 2022. PMID : 35702627

Under Review

13. Li ZY, **Liu S**, Ho JC, Narayan KMV, Ali MK, Varghese JS. Translating Subphenotypes of Newly Diagnosed Type 2 Diabetes from Cohort Studies to Electronic Health Records in the United States ***Nature Medicine***. Preprint

In Preparation

14. **Liu S**, Park WY, Al Gharaibeh FN, Wynn JL, Aziz KB. Antibiotics Stewardship - Characterizing antibiotics administration patterns among neonates < 33 weeks and 1500g with negative blood cultures relative to organ dysfunction. ***JAMA Pediatrics***.
15. **Liu S**, Ding X, Lehmann HP. Bootstrapping Bezier Curves to Quantify the Uncertainty Associated with the Decision-Analytic Optimal Point on the Receiver Operating Characteristic Curve. ***Medical Decision Making***.
16. **Liu S** and Martin B, Park WY, Northington F, Valdez RC, Makker K, Kuiper J, Nagy PG, Minty E and Aziz KB. Phenotype Development and Evaluation for Neonatal Hypoxic Ischemic Encephalopathy Using Electronic Health Record. ***Frontiers in Neurology***.

**See latest CV for updates (found in the header).

Presentations, Posters, and Case Studies

1. **Liu S**, Park WY, Al Gharaibeh FN, Wynn JL, Aziz KB. Antibiotics Stewardship - Characterizing antibiotics administration patterns among neonates < 33 weeks and 1500g with negative blood cultures relative to organ dysfunction. ***Neonatal Research Conference - UT San Antonio***. 2025. [Oral Presentation] Link to Slide
2. **Liu S**, Barrett RB, Zollo-Venecek K, Riesser B, Martin B. Advancing Electronic Clinical Quality Measure (eCQM) Interoperability : Model Context Protocol (MCP)-Orchestrated CQL-to-OMOP Translation. ***Observational Health Data Sciences and Informatics (OHDSI) Symposium***. 2025. [Software Demonstration] Link to Demo
3. **Liu S**, Lehmann HP. Clinical Utility Profiling (CUP) : A practical method for choosing predictive model cutoff ranges based on target deployment. ***AMIA Annual Symposium***. 2024. [Podium Presentation] Link to Slide and Abstract
4. **Liu S**, Lehmann HP. Identifying and Visualizing the Optimal Cutoff on the Receiver Operating Characteristic (ROC) Curve. ***Society for Medical Decision Making (SMDM)***. 2024. [Poster] Link to Poster
5. **Liu S**, Wei SX, Lehmann HP. Applicability Area : A Novel Utility-based Approach for Evaluating Predictive Models Beyond Discrimination. ***AMIA Annual Symposium***. 2023. [Podium Presentation] Link to Podium Presentation
6. **Liu S**, Lehmann HP. Assessing the Dimensions of a Dataset Most Sensitive to Cost-Sensitive Loss Functions : A Simulation Study. ***SMDM***. 2023. [Poster] Link to Poster
7. Sood P, **Liu S**, Kharrazi H. Evaluating the Variations in Healthcare Utilization Predictions Across Common Phenotypes of Type 2 Diabetes Using EHR and Claims Data. ***AMIA Annual Symposium***. 2023. [Poster] Link to Poster
8. Sood P, **Liu S**, Kharrazi H. Measuring the Effect of EHR Data Quality in Identifying Type-2 Diabetes Population Across Common Phenotype Definitions of Diabetes. ***AMIA Annual Symposium***. 2022. [Poster] Link to Poster
9. **Liu S**, Golozar A, McLeggon JA, Black A, Nagy P. The OHDSI Community Dashboard : Tracking the Health and Impact of the Open Science Observational Health Data Sciences and Informatics Community. ***Observational Health Data Sciences and Informatics (OHDSI) Symposium***. 2022. [Software Demonstration] Link to Abstract Link to Software
10. **Liu S**, Wall E, Patel SA, Park Y. COVID-19 Health Equity Dashboard - Addressing Vulnerable Populations. ***IEEE Visualization for Communication Workshop***. 2020. doi:10.31219/osf.io/2frha. [Case Study] Link to Case Study

Research Experience

Johns Hopkins School of Medicine — Baltimore, MD

Present - Jan. 2025

Independent Research - Self-initiated with PhD colleagues

Clinical quality measures automation : LLM-powered mapping from CQL to OMOP SQL

- Co-engineered Model Context Protocol (MCP) integration to provide LLMs with structured access to OMOP vocabularies and VSAC value sets, improving mapping accuracy through enhanced context retrieval
- Helped reduce mapping time from days to minutes via LLM-powered pipeline using LangChain to automate translation of clinical quality measures from CQL to executable SQL on OMOP databases
- Evaluated system performance against gold-standard clinician reviews; preparing manuscript for JAMIA

Python | LangChain | LLMs | MCP | OMOP CDM | SQL | Healthcare Standards

Present - Jan. 2024

Graduate Researcher | PI : Jim C. Fackler | Anesthesia and Critical Care Medicine

Estimating the risk of adverse drug events (ADEs) associated with broad-spectrum antibiotics among septic patients

- Conduct cohort discovery and characterize the risk of ADE across partner OMOP databases
- Define the risk profiles of ADEs and support broad-spectrum antibiotics administration
- Predict the risk of ADEs occurrence among pediatric patients with sepsis using statistical and NLP methods

ADEs | Antibiotics Stewardship | Pediatrics | Sepsis | OMOP | Network Study | Phenotyping

Present - Jun. 2023

Graduate Researcher | PI : Khyzer B. Aziz | NICU Precision Medicine Center of Excellence

Neonatal sepsis and antibiotics stewardship : Analyzing treatment patterns and outcomes

- Queried 90K+ neonates across 9 years of multi-million row NICU EHR data using SQL to analyze neonatal outcomes and antibiotic treatment patterns relative to organ dysfunction
- Engineered 20+ clinical features from temporal data to train ML algorithms (Random Forest, XGBoost, Clustering, LSTM), revealing that poor predictive performance signals inconsistent stewardship practices
- Leading manuscript for JAMA Pediatrics analyzing antibiotic use patterns; co-authoring multi-center studies to inform sepsis treatment guidelines across multi-center study sites

Python | SQL | Large-scale EHR data | Feature Engineering | Statistical Methods | Predictive Modeling | LSTM

Present - Nov. 2021

Graduate Researcher | Mentor : Harold P. Lehmann | Decision Sciences Lab

Clinical predictive model evaluation in the context of use : Aligning ML models with clinical decision-making

- Developed novel ML evaluation framework (Applicability Area) published at AMIA 2023, improving clinical utility assessment by incorporating harm-benefit tradeoffs, enabling better model selection for healthcare deployment
- Deployed interactive Azure dashboard for real-time sensitivity analysis across clinical scenarios to optimize ML model prediction thresholds based on context-specific cost-benefit ratios Link to iCUDA dashboard
- Built 1000-stratum Markov microsimulation evaluating suicide risk prediction strategies across 100K patients with probabilistic sensitivity analysis
- Leading manuscript establishing method for identifying optimal ML model operating points using decision analysis; conducting 25 physician interviews to quantify clinical harm-benefit tradeoffs

Python | R | Statistical Modeling | Markov Models | Decision Analysis | Azure | Interactive Dashboards | Qualitative Methods

Oct. 2025 - Nov. 2021

Research Assistant | PI : Hadi Kharrazi | Center for Population Health IT

Understanding the effect of data quality on using phenotype definitions to identify diabetes cohorts

- Processed 200K+ patient EHR and claims data; simulated 3 data quality scenarios across medical code systems (ICD, RxNorm, LOINC) to evaluate 4 diabetes phenotypes
- Built 20+ logistic regression models in R, demonstrating how data quality and phenotype choice impact downstream racial disparities in healthcare utilization predictions
- Co-authored 4 manuscripts informing equitable ML model development for population health

SQL | R | Large-scale EHR and Claims data | Logistic Regression | Statistical Modeling | Health Equity

May 2023 Oct. 2021

Graduate Researcher | Advisor : Paul G. Nagy | JHSOM

OHDSI Community Dashboard : Assessing the health and impact of the OHDSI ecosystem (`http://ohdsi.azurewebsites.net/`)

- Co-developed an automated ETL pipeline from PubMed, Google Scholar, YouTube, and EHDEN and mapped texts to standard terminologies using standard APIs
- Implemented/debugged the first version of the interactive tool with 3 front-end dashboards each with 3 features using Python, Flask, Dash, and Azure back-end
- Established an open-source framework for evaluating open-source scientific communities
- Led the writing of a manuscript accepted by JAMIA Open

OHDSI | APIs | Flask | Dash | HTML | Azure CosmosDB | GitHub | Manuscript

Sep. 2022 Oct. 2021

Graduate Researcher | Advisor : Hadi Kharrazi | JHSOM

Assessing the racial and socioeconomic disparities in postpartum depression using population-level hospital discharge data : A longitudinal retrospective study [Published]

- Conducted literature reviews on postpartum depression, diagnosis, and racial disparities
- Constructed logistic regression, multinomial regression, survival analysis, and sensitivity analysis using R and Healthcare Cost and Utilization Project datasets for Maryland (2016-2019)
- Published a manuscript and led its writing, journal submission, revisions, and correspondence

Manuscript | Maternal Health Disparity | HCUP | R | Regression Analysis

Emory University and Rollins School of Public Health — Atlanta, GA

May 2023 Jan. 2021

Research Assistant | PI : Mohammed K. Ali | Rollins School of Public Health

The associations of collaborative care intervention components with depressive symptoms and metabolic outcomes in the INDEPENDENT study : A mixed methods study

- ETL and analyzed 196 patients' longitudinal data and conducted multiple logistic regressions using R
- Wrote the statistical methods section and produced the appropriate figures for the manuscript
- Assisted the first author in writing a manuscript accepted by Primary Care Diabetes (second author)

Longitudinal Clinical Trials Data | Mixed Methods | R | Regression Analysis | Manuscript

Jun. 2021 – May 2020 | **Research Assistant | PI : Shivani Patel | Rollins School of Public Health**
COVID-19 Health Equity Dashboard (`https://covid19.emory.edu/`)
- Managed daily ETL process via Python scripts, integrating incident, demographic, vaccine, and SDoH data
- Implemented and managed back-end MongoDB database solutions using Python
- Implemented/debugged 20+ front-end features using NPM, React JS, React Simple Maps, and HTML
- Led the writing as the first author of a case study accepted by the IEEE VisComm 2020

COVID-19 Interactive Dashboard | Python | MongoDB | JavaScript | Case Study Presentation

May 2020 – Jan. 2020 | **Undergraduate Researcher | Mentor : Seunghwa Rho | Emory University**
Atlanta Fire and Rescue Department : Predicting fire incident risk Link to Report Link to PPT
- Pre-processed and aggregated +8 million observations of geospatial data on historical fires
- Optimized a random forest model to predict fires with 40% sensitivity and 90% specificity
- Delivered an R Shiny interactive dashboard to visualize predictions in use
- Received positive feedback from the Atlanta Mayor's Office, City Planning, and the Fire Department

Geospatial and Longitudinal Data | Predictive Modeling | Random Forest | Interactive Shiny App

Jun. 2020 – Sep. 2019 | **Volunteer Research Assistant | Rollins School of Public Health**
Diabetes care utilization and health equity research
- Wrangled Medical Expenditure Panel Survey and Healthcare Cost and Utilization Project data using R
- Conducted literature reviews on diabetes, preventable hospitalization, and care utilization
- Conducted consistent 30, 90, and 120-day phone interviews with a cohort of 24 patients with stroke

Diabetes | Survey Data | Panel Data | R | Literature Review | Phone Interviews

Professional Experience

RightPatient Dunwoody, GA

Aug. 2019 – May 2019 | **Analyst | Supervisor : Michael Trader**
- Liaised and delivered 2 million images from public safety agencies for the R&D project on opioid addiction
- Developed 4 client leads for the ACO patient-centered care software and relayed to the demo team
- Was entrusted with managing 358 client accounts on Zoho CRM for our patient wellness initiative

Healthcare Technology Startup | R&D | Customer Relations | Sales

Teaching & Mentorship

2025 – 2023 | **Graduate Teaching Assistant | Johns Hopkins School of Medicine**
Clinical Decision Analysis
- Assisted the faculties in teaching decision sciences and utility theory and hosted weekly office hours

TA | Decision Analysis | Treeage | Office Hours

2025 – 2023 | **Graduate Teaching Assistant | Johns Hopkins School of Medicine**
Implementing Fast Healthcare Interoperability Resources (FHIR)
- Assisted the faculties in teaching FHIR and hosted weekly office hours

TA | FHIR | Office Hours

2023 – 2022 | **Graduate Teaching Assistant | Johns Hopkins School of Medicine**
Clinical Data Analysis with Python
- Assisted the faculties in teaching Python and hosted weekly office hours

TA | Python | Office Hours

Jul. 2022 – Jul. 2022 | **Bootcamp Consultant | Instructor : Hadi Kharrazi | Bowie State University**
Public Health Informatics and Technology 4-Week Bootcamp
- Assisted the program faculties in teaching R and Tableau to a class of 15 students
- Hosted daily office hours and coached 4 capstone project groups

Bootcamp | R | Visualization | Capstone projects

May 2021 – May 2018 | **Academic Fellow Captain, Academic Fellow | Emory University**
Annual International Student Welcome and Mentoring Program
- Coordinated two orientations with 14 colleagues, oversaw 26 fellows, and mentored 24 mentees

Peer Mentorship | Program coordination

Peer Reviews

2021 - Present Reviewer | JAMIA Open

Certificates

Apr. 2021 - No Expiration	Information Privacy Security | JHMI CITI Program | Credential ID : 45798326
Apr. 2021 - No Expiration	Health Privacy Issues for Researchers | JSPH CITI Program | Credential ID : 45562106
Oct. 2021 - Oct. 2026	Basic Human Subjects Research | JSPH CITI Program | Credential ID : 45562104
Oct. 2021 - Oct. 2024	Social and Behavioral Research Best Practices for Clinical Research | JSPH CITI Program | Credential ID : 45562105
Oct. 2021 - Oct. 2023	Human Subjects Research - Biomedical Research | JHMI CITI Program | Credential ID : 45733041

Languages

English	Native
Chinese	Native
Spanish	Beginner