

Feng Xie

300 N Pasteur Dr, S238, Stanford, CA 94305 | 510-714-7947 | fengx@stanford.edu

Summary

- 6 years of experience in medical informatics and clinical data science using large-scale electronic health records (EHR)
- Abundant collaboration experience with clinicians, engineers, and healthcare researchers
- Extensive experience in machine learning software package development/release and large EHR data analysis
- Strong publication track record with 8 first-author research papers in well-known medical or informatics journals

Education

Duke-National University of Singapore (NUS) Graduate Medical School, **Singapore**
Joint Ph.D. by Duke and NUS, Health Service and Systems Research (Data Science) *August 2017-May 2022*
Thesis: Interpretable Automated Machine Learning Scoring Systems for Emergency Care

Tsinghua University, Bachelor of Science, School of Life Sciences **Beijing, China**
GPA 91/100, Outstanding Graduate Honor (Top 3%), Outstanding Thesis Award (Top 2%) *August 2013-July 2017*
Second Degree, Bachelor of Management, School of Economics and Management

Experience

Stanford University School of Medicine **Stanford, CA**
Postdoctoral Scholar *September 2022-Present*

- Multimodal prediction of neonatal complications with artificial intelligence and unstructured EHRs
- Developing domain-specific large language model to support clinical decision making in neonatal care.

Duke-National University of Singapore (NUS) Graduate Medical School **Singapore**
PhD Candidate & Fellow *August 2017- September 2022*

- Conceptualized new research ideas and conducted original research, including AutoScore for automating clinical score generation with interpretable machine learning, emergency triaging system, and readmission prediction tools
- 8 first-author publications in internationally high-impact journals, with 8 additional co-author papers
- Utilized and processed large-scale EHR database, such as Medical Information Mart for Intensive Care (MIMIC)
- Established and managed a 10-year relational database with over 200,000 patients for emergency care in Singapore
- Led a team of engineers, clinicians, and other researchers to benchmark risk models using large EHR database
- Presented original research in international conferences (e.g., American Medical Informatics Association [AMIA])
- Instructed 7 undergraduates, MD, and PhD students with co-author papers in the pipeline

Tsinghua University, Center for Applied Mathematics **Beijing, China**
Research Assistant (Undergraduate Thesis Project) *December 2016-July 2017*

- Built a mathematical model through ordinary differential equations on cell death and performed numerical simulation

University of Toronto, SickKids Hospital **Toronto, Canada**
Research Assistant (Overseas Research Exchange Program) *July 2016-September 2016*

- Used super-enhancers to identify Gli2 target genes in the 3T3-L1 cells through bioinformatical analysis

Skills

Programming: R, Python, SQL

Big Data & Machine Learning Tools: R (caret, ggplot2, glmnet, Tidyverse, survival, Knitr), Python (scikit-learn, NumPy, pandas, Matplotlib, TensorFlow, Keras, Pytorch), Linux environment

Data Science Projects: Pipeline for processing large-scale EHR data (cleansing, wrangling, visualization, modeling, interpretation), Statistical and machine learning analysis, software development, Variable selection

Research: Clinical study design, Systematic review, Cohort study (e.g., in emergency medicine and critical care), Research paper publication, Peer review

Software Development

- **AutoScore:** a novel framework to automate the development of a clinical scoring model with machine learning for predefined outcomes under R environment. Accepted and published by the Comprehensive R Archive Network (CRAN): Published at: <https://cran.r-project.org/web/packages/AutoScore/index.html> and <https://nliulab.github.io/AutoScore/>
Software manual: <https://cran.r-project.org/web/packages/AutoScore/AutoScore.pdf>
Overall downloads of over 10,000 since its publication at CRAN from June 2021
- **MIMIC-IV-ED Benchmark:** Python suite to benchmark predictive risk models for emergency departments from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database (Xie et al., *Scientific Data*. 2022)
Published at: <https://github.com/nliulab/mimic4ed-benchmark>

Selected Publications

- **Xie F**, Ning Y, Liu M, et.al. A universal AutoScore framework to develop interpretable scoring systems for predicting common types of clinical outcomes. *STAR Protocols*. **2023** <https://doi.org/10.1016/j.xpro.2023.102302>
- **Xie F**, Zhou J, Lee JW, et.al. Benchmarking Risk Triage Models for Emergency Department with Large Public Electronic Health Records. *Scientific Data*. **2022** <https://www.nature.com/articles/s41597-022-01782-9>
- **Xie F**, Liu N, Yan L, et al. Development and Validation of an Interpretable Machine Learning Scoring Tool for Estimating Time to Emergency Readmissions. *EClinicalMedicine*. **2022**; 45:101315
[https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370\(22\)00045-1/fulltext#%20](https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(22)00045-1/fulltext#%20)
- **Xie F***, Yuan H*, Ning Y, et al. Deep Learning for Temporal Data Representation in Electronic Health Records: A Systematic Review of Challenges and Methodologies. *Journal of Biomedical Informatics*. **2022**.
<https://doi.org/10.1016/j.jbi.2021.103980>
- **Xie F**, Ning Y, Yuan H, et al. AutoScore-Survival: Developing interpretable machine learning-based time-to-event scores with right-censored survival data. *Journal of Biomedical Informatics*. **2021**.
<https://doi.org/10.1016/j.jbi.2021.103959>
- **Xie F**, Ong MEH, Liew JNMH, et al. Development and Assessment of an Interpretable Machine Learning Triage Tool for Estimating Mortality After Emergency Admissions. *JAMA Network Open*. **2021**.
<https://doi.org/10.1001/jamanetworkopen.2021.18467>
- **Xie F**, Chakraborty B, Ong MEH, et al. AutoScore: A Machine Learning–Based Automatic Clinical Score Generator and Its Application to Mortality Prediction Using Electronic Health Records. *JMIR Med Inform*. **2020**.
<https://doi.org/10.2196/21798>
- **Xie F**, Liu N, Wu SX, et al. Novel Model for Predicting Inpatient Mortality After Emergency Admission to Hospital in Singapore: Retrospective Observational Study. *BMJ Open*. **2019**.
<http://dx.doi.org/10.1136/bmjopen-2019-031382>
- Yu JY, **Xie F**, Liu N et al. An external validation study of the Score for Emergency Risk Prediction (SERP), an interpretable machine learning-based triage score for the emergency department. *Scientific Report*. **2022**
- Yuan H, **Xie F**, Ong MEH, et al. AutoScore-Imbalance: An Automated Machine Learning Tool to Handle Data Imbalance in Interpretable Clinical Score Development. *Journal of Biomedical Informatics*. **2022**.
- Yu JY, Heo S, **Xie F** et al. Development and Asian-wide validation of the Grade for Interpretable Field Triage (GIFT) for predicting mortality in pre-hospital patients using the Pan-Asian Trauma Outcomes Study (PATOS). *The Lancet Regional Health - Western Pacific*. **2023**
- Saffari SE, Ning Y, **Xie F** et al. AutoScore-Ordinal: An Interpretable Machine Learning Framework for Generating Scoring Models for Ordinal Outcomes. *BMC Medical Research Methodology*. **2022**