BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Jiang, Xiaoqian

eRA COMMONS USER NAME (credential, e.g., agency login): X1JIANG

POSITION TITLE: Christopher Sarofim Professor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Start Date MM/YYYY	Completion Date MM/YYYY	FIELD OF STUDY
Shanghai Maritime University	B.S.	09/1999	07/2003	Computer Science
University of Iowa	M.S.	09/2003	05/2006	Computer Science
Carnegie Mellon University	Ph.D.	08/2006	12/2010	Computer Science
University of California, San Diego	Postdoc	01/2011	06/2013	Biomedical informatics

A. Personal Statement

Sharing and analyzing massive amounts of biomedical data requires tackling privacy and efficiency to preserve the utility. My data privacy and machine learning background give me a unique perspective on integrated solutions. I am the Associate Vice President of Medical AI at the University of Texas Health Science Center at Houston (UTHealth). I also hold the Department Chair of Data Science and Artificial Intelligence position and am honored to be the Christopher Sarofim family professor. Additionally, I serve as the center director of Secure Artificial Intelligence for Health Care (SAFE) at the McWilliams School of Biomedical Informatics (MSBMI). Before joining UTHealth, I was an associate professor with tenure in the Department of Biomedical Informatics (DBMI) at UCSD. I received my Ph.D. from the School of Computer Science at Carnegie Mellon University (advised by Dr. Latanya Sweeney) with visiting student experience at the Decentralized Information Group (DIG) of Computer Science and Artificial Intelligence Lab (CSAIL) at MIT in the last two years of my graduate study. I received biomedical informatics postdoctoral training on ethics, legal, and social implications (ELSI). I investigated privacy technology for healthcare at UCSD (under Dr. Ohno-Machado's supervision) through the NLM Pathway to Independence Award. These experiences gave me a profound way to apply machine learning in biomedical informatics to address data privacy and efficiency issues simultaneously.

My research interests include privacy-preserving distributed data analysis and secure genome outsourcing. In the last six years, I received R00, R13, R21, R01, U01, OT, CPRIT Rising Stars, and UT STARs awards as principal investigator, as well as distinguished and best paper awards from AMIA Summit on Clinical Research Informatics (CRI), Translational Bioinformatics (TBI), and Annual Symposium in 2012, 2013, 2016, 2020, and 2024. I serve as the associate editor for *BMC Medical Informatics and Decision Making,* a guest lead editor for *Cancer Informatics*, and an editorial board member of the *Journal of the American Medical Informatics Association* and *Harvard Papers on Technology Science (H-POTS)*. I co-chaired the *2nd IEEE Conference on Health Informatics, Imaging, and System Biology.* I am the steering committee chair for *the International Workshop on Genome Privacy and Security (GenoPri)*. I co-organized the *iDASH genome privacy workshop series* (2014-now) reported by *GenomeWeb* and *Nature News*. My unique experiences of constantly integrating these worlds give me unusual insights to contribute to the interdisciplinary research of data privacy and machine learning in medicine.

Publications:

Pena D, Barman A, Suescun J, **Jiang X**, Schiess MC, Giancardo L, Alzheimer's Disease Neuroimaging Initiative. Quantifying Neurodegenerative Progression With DeepSymNet, an End-to-End Data-Driven Approach. Front Neurosci 2019 Oct 4;13:1053. PMID:31636533

Kim Y, **Jiang X**, Giancardo L, Pena D, Bukhbinder AS, Amran AY, Schulz PE, Alzheimer's Disease Neuroimaging Initiative. Multimodal Phenotyping of Alzheimer's Disease with Longitudinal Magnetic Resonance Imaging and Cognitive Function Data. Sci Rep 2020 Mar 26;10(1):5527. PMID:32218482

Dai Y, Yu H, Yan Q, Li B, Liu A, Liu W, **Jiang X**, Kim Y, Guo Y, **Zhao Z**. Drug-Target Network Study Reveals the Core Target-Protein Interactions of Various COVID-19 Treatments. Genes. 2022 Jul 6;13(7):1210–1210. PMCID: PMC9316565

Dai Y, Hsu YC, Fernandes BS, Zhang K, Li X, Enduru N, Liu A, Manuel AM, **Jiang X**, **Zhao Z**, Alzheimer's Disease Neuroimaging Initiative. Disentangling Accelerated Cognitive Decline from the Normal Aging Process and Unraveling Its Genetic Components: A Neuroimaging-Based Deep Learning Approach. J Alzheimers Dis. IOS Press; 2024;97(4):1807–1827. PMID: 38306043

Ongoing Research Support

R01AG066749 (MPIs Jiang, Zheng)
Finding combinatorial drug repositioning therapy for ADRD   04/01/20 - 03/31/25

R01AG083039 (MPIs Jiang, Tao, Bian)
An end-to-end informatics framework to study Multiple Chronic Conditions (MCC)'s impact on Alzheimer's disease using harmonized electronic health records   09/01/23 - 05/31/26

U01CA274576-01A1 (MPIs Long, Jiang)
Robust privacy-preserving distributed analysis platform for cancer research: addressing data bias and disparities   06/01/23 - 05/31/26

U01AG079847 (MPIs Zhao, Jiang)   09/15/23 - 08/31/28
AIM-AI: an Actionable, Integrated and Multiscale genetic map of Alzheimer's disease via deep learning

R01LM014520 (MPIs Ayday, Jiang, Vaidya)
Accelerating Genomic Data Sharing and Collaborative Research with Privacy Protection 09/01/23 - 08/31/28

R01AG082721 (MPIs Kim, Jiang)
Harmonizing multiple clinical trials for Alzheimer's disease to investigate differential responses to treatment via federated counterfactual learning   09/01/23 - 08/31/28

U24LM013755 (MPIs Ohno-Machado/Aronoff-Spencer/Xu)   Role: site-PI   12/21/20 - 11/30/24
RADx-rad Discoveries & Data: Consortium Coordination Center Program Organization

B.   Positions, Scientific Appointments, and Honors
Positions and Employment
2018 - Present	Director of Center of Secure Artificial Intelligence for hEalthcare (SAFE), School of Biomedical Informatics (SBMI), UTHealth
2020 - Present	Christopher Sarofim Professor, School of Biomedical Informatics, UTHealth
2023 - Present	Associate Vice President for Medical AI, UTHealth
2023 - Present	Chair, Department of Data Science and Artificial Intelligence

Other Experience and Professional Memberships
2010 - Present	Member, American Medical Informatics Association (AMIA)
2009 - Present	Member, Institute of Electrical and Electronics Engineers (IEEE)

2003 - Present	Member, Association of Computing Machinery (ACM)
2023 - Present	Fellow, International Academy of Health Sciences Informatics (IAHSI)
2023 - Present	Fellow of the American College of Medical Informatics (FACMI)

Honors

2012	Distinguished Paper Award, AMIA Summit on Clinical Research Informatics
2013	Distinguished Paper Award, AMIA Summit on Clinical Research Informatics
2016	Best Poster Award Health Data Exploration Network's 2nd Annual Network Meeting
2016	Best Paper Award in the AMIA 2016 Joint Summits on Translational Science
2016	Outstanding achievement award in secure genomic data analysis with SGX by Intel
2018	University of Texas System Stars Award
2018	Cancer Preventative and Research Institute of Texas (CPRIT) Rising Star
2019	Dean's Excellence Award for Innovation at UTHealth
2020	Distinguished Paper Award, AMIA Annual Symposium
2022	Dean's Excellence Award for Outstanding Faculty at UTHealth
2024	Best Student Paper Award in the American Medical Informatics Association (AMIA) 2024 Joint Summits on Translational Science

C.   Contributions to Science

1. **Model calibration** is the topic of my early publications. Calibration is essential in predictive models to generate individualized estimates for medically relevant clinical care outcomes and translational research. In predictive models based on binary results, the outputs constitute probability estimates that the event of interest will occur (e.g., a given patient has an 8% chance of having a myocardial infarction, given her risk factors). In this context, we measure the individualized prediction's calibration by checking how close this prediction is to the 'true' underlying probability of the event for that particular patient. Calibration is essential for these personalized medicine tools since estimates (i.e., predictions) are often used to determine a patient's risk (e.g., ICU mortality calculators and MELD score for liver transplantation). High risk can guide critical clinical decisions, such as initializing anti-lipid pharmacotherapy for an individual at high risk for cardiovascular disease or referral for chemoprevention trials for a woman with high chances of developing breast cancer. I developed novel models to ensure accurate probability estimation for individualized estimates for risk assessment. My research has improved the state-of-the-art calibration methodology, resulting in measures that are more suitable for individualized predictions than estimates that were calibrated using previous methods.

   a. **Jiang X**, Kim J, Wu Y, Ohno-Machado L. Selecting Cases for Whom Additional Tests Can Improve Prognostication, *AMIA Annual Symposium*, 1260-1268, 2012. PMID: 23304404

   b. Menon A, **Jiang X**, Vambu S, Elkan C, and Ohno-Machado L. Prediction of accurate probabilities with a ranking loss, *International Conference on Machine Learning (ICML),* 2012: CoRR abs/1206.4661.

   c. **Jiang X**, Osl M, Kim J, Ohno-Machado L. Calibrating predictive model estimates to support personalized medicine. *Journal of the American Medical Informatics Association*, 2012; 19(2): 263-274. PMID: 21984587

   d. **Jiang X**, Boxwala A, El-Kareh R, Kim J, Ohno-Machado L. A patient-Driven Adaptive Prediction Technique (ADAPT) to improve personalized risk estimation for clinical decision support. *Journal of the American Medical Informatics Association*, 19:e137-e144, 2012. PMID: 22493049

2. **Deep learning models for healthcare.** I have been studying context-aware deep learning models for healthcare applications and developed customized solutions to synthesize clinical knowledge and data to predict healthcare outcomes better. Working closely with clinical collaborators, we built novel models (e.g., supervised, semi-supervised, weakly supervised, reinforcement) to tackle the specific challenges in different tasks, including drug repurposing, lab test reduction, integrative modeling for electronic healthcare records including clinical notes, and

   a. Lee D, **Jiang X**, Yu H. Harmonized representation learning on dynamic EHR graphs. *J Biomed Inform* Academic Press; 2020 Jun;106:103426. PMID:32339747

   b. Yu L, Li L, Bernstam E, **Jiang X**. A deep learning solution to recommend laboratory reduction strategies in ICU. *Int J Med Inform* Elsevier; 2020 Dec;144:104282. PMID:33010730

    c. Hsieh K-L, Wang Y, Chen L, Zhao Z, Savitz S, **Jiang X**, Tang J, Kim Y. Drug Repurposing for COVID-19 using Graph Neural Network with Genetic, Mechanistic, and Epidemiological Validation. *Res Sq [Internet]* 2020 Dec 11; PMID:33330858

    d. Kim Y, Suescun J, Schiess MC, **Jiang X**. Computational medication regimen for Parkinson's disease using reinforcement learning. *Sci Rep* 2021 Apr 29;11(1):9313. PMID:33927277

3. **Distributed data analysis** for building models without sharing data is an active area of my research. Due to privacy concerns, it is hard for data custodians to collaborate efficiently because naïve procedures involving patient-level data transfer increase privacy risk and may violate institutional policies for sharing. For example, the general policy is that patient data must remain on U.S. Department of Veterans Affairs (V.A.) servers. It is non-trivial to enable cross-institutional collaboration that respects patient privacy and institutional policies. This motivated the development of federated data analysis algorithms that transmit only aggregated statistics as a practical alternative. I have developed many federated data analysis models for distributed data that do not exchange patient-level information. I have produced as accurate results as if data were centralized, which is not the case when meta-analysis methods are used.

    a. Kim Y, Sun J, Yu H, **Jiang X**. Federated Tensor Factorization for Computational Phenotyping. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17),* New York, NY, USA: ACM; 2017. p. 887–895.

    b. Chang C, Deng Y, **Jiang X**, Long Q. Multiple imputations for analysis of incomplete data in distributed health data networks. Nat Commun 2020 Oct 29;11(1):5467. PMID:33122624

    c. Deng Y, **Jiang X**, Long Q. Privacy-preserving Methods for Vertically Partitioned Incomplete Data. *AMIA Annual Symposium*, 2020 (**Distinguished paper award**)

    d. Dai W, **Jiang X**, Bonomi L, Li Y, Xiong H, Ohno-Machado L. VERTICOX: Vertically distributed cox proportional hazards model using the alternating direction method of multipliers. *IEEE Trans Knowl Data Eng Institute of Electrical and Electronics Engineers* (IEEE); 2020;1–1.

4. **Healthcare privacy** is the central area of my research. Data sharing and information exchange are critical in big data science to improve quality, speed up discovery, and promote meaningful use. Privacy protection is a big challenge in biomedical data sharing, as an inappropriate information process can put sensitive patient and family member's personal information at risk. Recent studies have shown the vulnerability of "anonymized" data, especially when attackers have background knowledge. My research focuses on *differential privacy*, which emerged as an ad-omnia solution. It roughly ensures that (almost and quantifiably) no additional privacy risk to individual patients by contributing their information to a statistical database. Based on biomedical data characteristics, I developed novel methods to de-identify data to support exploratory and decentralized data analysis in a privacy-preserving manner.

    a. **Jiang X**, Kim M, Lauter K, Song Y. Secure Outsourced Matrix Computation and Application to Neural Networks. *25th ACM Conference on Computer and Communications Security (CCS)* 2018.

    b. Zhang Q, Ma J, Lou J, Xiong L, **Jiang X**. Towards training robust private aggregation of teacher ensembles under noisy labels. *IEEE International Conference on Big Data (Big Data)* IEEE; 2020. p. 1103–1110.

    c. Aziz MMA, Kamali S, Mohammed N, **Jiang X**. Online algorithm for Differentially Private genome-wide Association Studies. ACM Trans Comput Healthcare Association for Computing Machinery (ACM); 2021 Mar;2(2):1–27.

    d. Lee D, Yu H, **Jiang X**, Rogith D, Gudala M, Tejani M, Zhang Q, Xiong L. Generating sequential electronic health records using dual adversarial autoencoder. *J Am Med Inform Assoc* Oxford University Press; 2020 Jul 1;27(9):1411–1419. PMID:32989459

5. **Multi-Scale, Integrated, and Contextualized Approaches for Aging and Disease in Public Health** is to understand and address the complex interplay between biological, environmental, and social factors. My research focuses on leveraging advanced computational methods and collaborative frameworks to enhance the understanding and treatment of age-related diseases, particularly those affecting the brain. This includes developing computational models for phenotyping aging-related diseases using multi-modal data, creating deep learning models to predict Alzheimer's disease progression, investigating the impact of social and behavioral factors on Alzheimer's disease, utilizing graph representation learning to predict health outcomes, and exploring methods to enhance fairness in disease prediction models. By integrating these multi-scale and contextualized approaches, my research aims to advance our understanding of aging

and disease, ultimately improving public health through more accurate predictions, targeted interventions, and equitable healthcare solutions.

a. Ren Y, Lou J, Xiong L, Ho JC, **Jiang X**, Bhavani SV. MULTIPAR: Supervised Irregular Tensor Factorization with Multi-task Learning for Computational Phenotyping. *Machine Learning for Health (ML4H)*, 498-511, 2023.

b. Ma X, Shyer M, Harris K, Wang D, Hsu YC, Farrell C, Goodwin N, Anjum S, **Jiang X**, Kim Y. Deep learning to predict rapid progression of Alzheimer's disease from pooled clinical trials: A retrospective study. *PLOS Digital Health*, 3(4), e0000479, 2024.

c. Bhattarai K, Rajaganapathy S, Das T, Kim Y, Chen Y, Dai Q, Li X, Jiang X, Zong N. Learning Physician's Treatment for Alzheimer's Disease based on Electronic Health Records and Reinforcement Learning. *IEEE 11th International Conference on Healthcare Informatics (ICHI).* IEEE; 2023. p. 525–526.

d. Ciciora D, Vásquez E, Valachovic E, Hou L, Zheng Y, Xu H, **Jiang X**, Huang K, Gabriel KP, Deng HW, Gallant MP, Zhang K. Social and Behavior Factors of Alzheimer's Disease and Related Dementias: A National Study in the U.S. *Am J Prev Med*. Elsevier; 2024 Apr;66(4):573–581. PMID: 37995949

Complete List of Published Work in My NCBI Bibliography:
https://www.ncbi.nlm.nih.gov/myncbi/xiaoqian.jiang.1/bibliography/public/