LI, WENTAO

: mentosli@outlook.com ◇ : Li-Wentao ◇ : Wentao Li ◇ : wentaoli.net

EDUCATION

University of Texas Health Science Center at Houston (UTHealth) *Feb 2021 - Dec 2025 (exp)*
4^{th}-year PhD student in the School of Biomedical Informatics
Honor: *Dean's Excellent Award 2021, 2021; Jingchun Sun Memorial Scholarship of UTHealth, 2023; D.Bradley McWilliams Scholars Endowed Scholarship Award, 2024.*

University of California, San Diego *Sep 2018 - Jun 2020*
Master of Science in Statistics

Shanghai Maritime University, Shanghai, China *Sep 2014 - Jun 2018*
Bachelor of Science in Mathematics
Honor: *Dean's List of SMU, 2016; First Class Scholarship of SMU, 2017.*

TECHNICAL HIGHLIGHTS

Languages: Python (Pytorch, Tensorflow), JavaScript (node), plink, R, Matlab

Skills: Machine Learning, Deep Learning, Multi-omics Studies, Medical Imaging
Federated Learning, Privacy-preserving AI

WORK EXPERIENCE

Graduate Research Assistant, UT MD Anderson Cancer Center *Oct 2023 - present*

- Developed a novel cross-modal attention fusion method integrating multi-modal medical data (e.g., CT/PET-CT imaging, genomic, and clinical data) to enhance predictive modeling in cancer studies, this developed method overcomes the challenges of information mixture in different modalities with domain separation;
- Design and deploy a foundation model for chest CT scans, utilizing a dataset of 10,000 patients to improve diagnostic accuracy and efficiency for downstream tasks;
- Performed survival analysis on lung cancer patients in EGFR and ALK mutation clinical trials, demonstrating that radiological measurements of sublethal damage (SLD) can predict progression-free survival (PFS);
- Investigated brain regional interactions and genetic variant expressions associated with psychological disorders such as Bipolar Disorder, Depression, and Anxiety. Mapped genetic findings to brain spatial structures to provide potential therapeutic insights, with a novel approach that incorporates individualized brain patterns in genomic association analysis.

Graduate Research Assistant, UTHealth *Feb 2021 - Sep 2023*

- Developed and published Federated Generalized Linear Mixed Models (FedGLMMs) for Genome-Wide Association Studies (GWAS), introducing mixed effect models in a federated setting for the first time to address both scalability and privacy challenges in genomic research [1, 2, 3];
- Led privacy-preserving genomic data analysis evaluations using the OpenSNP dataset, addressing the critical challenge of protecting genomic data while demonstrating secure data-sharing capabilities [4];
- Engineered privacy-preserving correlation estimation and genetic imputation algorithms for GWAS, enhancing data privacy without compromising analytical power [5, 6, 7, 8];
- Developed COLLAGENE, a secure genomic analysis tool published in Genome Biology, offering a practical solution for privacy-preserving GWAS on binary phenotypes and secure meta-analysis, with the novel use of homomorphic encryption to protect data privacy [9];
- Created a deep learning model to accurately predict blood pressure from photoplethysmogram (PPG) signals, advancing non-invasive health monitoring technologies [10].

Research Intern, UTHealth *Jul 2020 - Jan 2021*

- Developed and published a privacy-preserving federated learning method to approximate the intractable marginal log-likelihood function in Generalized Linear Mixed Models (GLMMs) for cohort studies [11];
- Successfully hosted federated training across Houston, San Diego, and Munich using VERTIcal Grid Logistic Regression with Confidence Intervals (VERTIGO-CI), enhancing collaboration across multiple research institutions [12];

Research Assistant, School of Medicine, UCSD *Jun 2019 - Jun 2020*

- Conducted mathematical proofs for calibration measurements and models in clinical prediction research [13];
- Deployed and tested Docker containers for the prediction models (VERTIGO with Confidence Intervals & GLORE), evaluating privacy-preserving capabilities with data from Oklahoma, Texas, and San Diego.

PRESENTATION

AMIA 2021 Virtual Informatics Summit *Mar 2021*
Principal Speaker

- Presentation on published conference paper 'VERTIcal Grid lOgistic regression with Confidence Interval'

PROJECTS

Personal Website: `https://wentaoli.net` (for additional project details)

Federated Learning Platform (FedPlatform) Development *May 2022 - present*
Principal Developer

- Developed a lightweight cross-silo federated learning platform accessible via a web browser;
- Embedded a Python distribution within the browser to streamline federated learning tasks, eliminating the need for federated trainers to install dependencies;
- Simulated multi-party data collaboration tests for linear regression using federated learning;
- The ongoing project aims to bridge isolated data silos, providing an intuitive platform for non-professional users to engage in federated learning tasks.

FedML MLOpsCloud-Web Development *Sep 2022 - Jun 2023*
Research Developer

- Contributed to an open-source project under FedML Inc (`https://fedml.ai`), a US-based startup focused on building scalable, collaborative AI solutions;
- Developed a web-based cross-silo federated learning feature, facilitating secure collaboration between distributed data sources;
- Designed and implemented a generalized framework to align model structures during communication between browsers (Tensorflow.js) and servers (Pytorch), enabling seamless cross-platform federated learning.

PUBLICATIONS

[1] **W. Li**, H. Chen, X. Jiang, and A. Harmanci, "Federated generalized linear mixed models for collaborative genome-wide association studies," *iScience*, vol. 26, no. 8, p. 107227.

[2] **W. Li**, H. Chen, X. Jiang, and A. Harmanci, "Fedgmmat: Federated generalized linear mixed model association tests," *PLoS computational biology*, vol. 20, no. 7, p. e1012142, 2024.

[3] M. M. Anjum, N. Mohammed, **W. Li**, and X. Jiang, "Privacy preserving collaborative learning of generalized linear mixed model," *Journal of Biomedical Informatics*, vol. 127, p. 104008. Publisher: Elsevier.

[4] L. Dervishi, X. Wang, **W. Li**, A. Halimi, J. Vaidya, X. Jiang, and E. Ayday, "Facilitating federated genomic data analysis by identifying record correlations while ensuring privacy," *AMIA Annual Symposium Proceedings*, vol. 2022, pp. 395–404.

[5] S. Wang, M. Kim, **W. Li**, X. Jiang, H. Chen, and A. O. Harmanci, "Privacy-aware kinship inference in admixed populations using projection on reference panels," *bioRxiv*, pp. 2022–05. Publisher: Cold Spring Harbor Laboratory.

[6] S. Wang, M. Kim, **W. Li**, X. Jiang, H. Chen, and A. Harmanci, "Privacy-aware estimation of relatedness in admixed populations," *Briefings in Bioinformatics*, vol. 23, no. 6. Publisher: Oxford Academic.

[7] X. Wang, L. Dervishi, **W. Li**, X. Jiang, E. Ayday, and J. Vaidya, "Efficient federated kinship relationship identification," *AMIA Summits on Translational Science Proceedings*, vol. 2023, pp. 534–543.

[8] A. O. Harmanci, M. Kim, S. Wang, **W. Li**, Y. Song, K. E. Lauter, and X. Jiang, "Open imputation server provides secure imputation services with provable genomic privacy," *bioRxiv*, pp. 2021–09. Publisher: Cold Spring Harbor Laboratory.

[9] **W. Li**, M. Kim, K. Zhang, H. Chen, X. Jiang, and A. Harmanci, "COLLAGENE enables privacy-aware federated and collaborative genomic data analysis," *Genome Biology*, vol. 24, no. 1, p. 204.

[10] Y. Chu, K. Tang, Y.-C. Hsu, T. Huang, D. Wang, **W. Li**, S. I. Savitz, X. Jiang, and S. Shams, "Non-invasive arterial blood pressure measurement and SpO2 estimation using PPG signal: a deep learning framework," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 131.

[11] **W. Li**, J. Tong, M. M. Anjum, N. Mohammed, Y. Chen, and X. Jiang, "Federated learning algorithms for generalized mixed-effects model (GLMM) on horizontally partitioned data from distributed sources," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, p. 269. Publisher: Springer.

[12] J. Kim, **W. Li**, T. Bath, X. Jiang, and L. Ohno-Machado, "VERTIcal grid lOgistic regression with confidence intervals (VERTIGO-CI)," *AMIA Summits on Translational Science Proceedings*, vol. 2021, p. 355. Publisher: American Medical Informatics Association.

[13] Y. Huang, **W. Li**, F. Macheret, R. A. Gabriel, and L. Ohno-Machado, "A tutorial on calibration measurements and calibration models for clinical prediction models," *Journal of the American Medical Informatics Association*, vol. 27, no. 4, pp. 621–633. Publisher: Oxford University Press.