LI, WENTAO

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

EDUCATION

University of Texas Health Science Center at Houston (UTHealth) *Feb 2021 - present*
PhD student in the School of Biomedical Informatics
Honor: Dean's Excellent Award 2021, 2021; Jingchun Sun Memorial Scholarship of UTHealth, 2023.

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

Shanghai Maritime University, Shanghai, China *Sep 2014 - June 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, Genomic Studies, Medical Imaging Studies
Federated Learning, Privacy-preserving AI

WORK EXPERIENCE

Graduate Research Assistant, UTHealth *Feb 2021 - present*

- Developed and published a series of Federated Generalized Linear Mixed Models (FedGLMMs) for the Genome-Wide Association Studies (GWAS) [1, 2, 3];
- Developed deep learning model to predict blood pressure using PPG signal [4];
- Conducted privacy-preserving Genomic Data Analysis evaluation experiments [5] with OpenSNP dataset;
- Developed privacy-preserving correlations estimation and genetic imputation algorithms for GWAS [6, 7, 8, 9];
- Built a ready-to-run secure genomic analysis tool COLLAGENE [10], which provides practical privacy-preserving GWAS protocol for binary phenotypes and a secure meta-analysis protocol.

Research intern, UTHealth *July 2020 - Jan 2021*

- Developed and published a privacy federated learning method to approximate the intractable marginal log-likelihood function in the Generalized Linear Mixed Models (GLMMs) for cohort study [11];
- Hosted successful federated training among Houston, San Diego, and Munich with previous published work VERTIcal Grid logistic regression with Confidence Interval (VERTIGO-CI) [12];

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

- Conducted mathematical proofs in calibration measurements and models for clinical prediction [13];
- Developed two prediction models in R and Python that can handle horizontally and vertically partitioned data, Grid Binary LOgistic REgression (GLORE) and VERTIcal Grid logistic regression (VERTIGO);
- Set up Dockers for the prediction models (VERTIGO with Confidence Intervals & GLORE) and then tested the capability of privacy-preserving prediction with data from Oklahoma, Texas, and San Diego.

PRESENTATION

AMIA 2021 Virtual Informatics Summit *March 2021*
Principal Speaker

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

PROJECTS

Personal Website: `https://wentaoli.net` (for additional projects information)

Federated Learning Platform (FedPlatform) development *May 2022 - present*
Principal developer

- Developed a lightweight cross-silo federated learning platform based on the browser;
- Embed a Python distribution on the browser to accomplish federated learning tasks. This lightweight system can free federated trainers from installing any dependencies;
- Accomplished multi-party data collaboration simulation test on linear regression with federated learning;
- Ongoing project aims to bridge isolated data islands and provide an experience-friendly platform for non-professional users to collaborate on federated learning tasks.

FedML MLOpsCloud-Web development *Sep 2022 - Jun 2023*
Research developer

- Open source project under FedML Inc (`https://fedml.ai`), a US start-up company building open and collaborative AI anywhere at any scale.
- Developed a web-based cross-silo federated learning feature in FedML;
- Designed and deployed a generalised framework in web-based federated learning, which aligns model structures during communication between browsers (Tensorflow.js) and the server (Pytorch).

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." Pages: 2023.10.03.560753 Section: New Results.

[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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] **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.

[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.