LI, WENTAO

✉: mentosli@outlook.com ◇ ⌂: Li-Wentao ◇ 🎓: Wentao Li

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

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 (15%), 2016; First Class Scholarship of SMU (5%), 2017

TECHNICAL HIGHLIGHTS

Languages:	Python (Pytorch, Tensorflow), JavaScript (node), plink, R, Matlab
Skills:	Privacy-preserving AI, Machine Learning, Federated Learning, Differential Privacy, Secure Multi-party computation, Deep Learning, Homomorphic Encryption

WORK EXPERIENCE

Research Assistant, UTHealth *Feb 2021 - present*

· Developed and published a genetic algorithm for federated learning in the privacy-preserving genome-wide association studies (GWAS) using GLMMs [2];
· Conducted federated Genomic Data Analysis evaluation experiments [6] with OpenSNP dataset;
· Developed a privacy federated learning genetic algorithm based on R package 'Generalized linear Mixed Model Association Tests (GMMAT)', and publication is ongoing;
· Developed a privacy-preserving kinship relationship estimation algorithm using attention layer in transformer model for cross-silo GWAS, and publication is ongoing.

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 [1];
· Conducted experiments in adding differential privacy to federated GLMMs [5];
· Hosted successful federated training among Houston, San Diego, and Munich with previous published work VERTIcal Grid logistic regression with Confidence Interval (VERTIGO-CI) [4];

Research Assistant, School of Medicine, UCSD *June 2020 - June 2021*

· Developed two prediction models in R and Python (based on Logistic Regression) that can handle horizontally and vertically partitioned data, Grid Binary LOgistic REgression (GLORE) and VERTIcal Grid logistic regression (VERTIGO);
· Proved and developed an algorithm that can transmit Confidence Intervals based on VERTIGO and published the method as VERTIGO-CI [4];
· 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;
· Cleaned highly correlated genetic data with Quality control (QC) procedure in Plink;
· Developed a useful SNPs QC pipeline that can check for sex discrepancy, Minor allele frequency, Hardy-Weinberg equilibrium, Heterozygosity, relationships between individuals and association analyses;
· Conducted mathematical proofs in calibration measurements and models for clinical prediction [3].

SEMINARS & PROJECTS

AMIA 2021 Virtual Informatics Summit *March 2021*
Principal Speaker

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

FedML MLOpsCloud-Web development *Sep 2022 - present*
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).

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.

PUBLICATIONS

[1] 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, 2022. Publisher: Springer

[2] W. Li, H. Chen, X. Jiang, and A. Harmanci, "Federated Generalized Linear Mixed Models for Collaborative Genome-wide Association Studies," *arXiv preprint arXiv:2210.00395*, 2022

[3] 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, 2020. Publisher: Oxford University Press

[4] 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, 2021. Publisher: American Medical Informatics Association

[5] 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, 2022. Publisher: Elsevier

[6] 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," *arXiv preprint arXiv:2203.05664*, 2022

[7] 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, 2021. Publisher: Cold Spring Harbor Laboratory

[8] 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, 2022. Publisher: Oxford Academic

[9] 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, 2022. Publisher: Cold Spring Harbor Laboratory