

Yuying Lu

✉ yl5839@cumc.columbia.edu ☎ +1 6462286733 Ⓛ github.com/YUYING-LU

EDUCATION

Mailman School of Public Health, Columbia University, New York, USA
Ph.D. Candidate of Biostatistics

Sept. 2024 - Now
Overall GPA: 4.11/4.3

Shanghai Jiao Tong University (SJTU), Shanghai, China
Bachelor of Mathematics and Applied Mathematics (Wu Wenjun Program)

Sept. 2020 - Jun. 2024
Rank: 1/8 Overall GPA: 3.88/4.3

RESEARCH INTERESTS

Transfer Learning; Semi-supervised Learning; Online Statistical Inference; Synthetic Data; Diffusion Model; Electronic Health Records (EHRs); Precision Medicine.

SELECTED RESEARCH EXPERIENCE

ReFIT: Federated Transfer Learning for Sequential Prediction

May. 2025 - Dec. 2025

- Designed a streaming (online) optimization algorithm for federated, semi-supervised transfer learning, enabling parameter updates as source data arrive sequentially.
- Implemented the method in R and ran numerical experiments on simulated data, using conformal inference for calibrated prediction sets.
- Built a model-ready cohort from breast cancer EHR data in All of Us Workbench by querying, cleaning, and standardizing records for 40,000+ participants, with validation for missing data and multicollinearity.
- First author; submitted a manuscript to the journal *Statistics in Biosciences (SIBS)*.

DRIFT: Multi-Domain Robust Estimation of Individualized Treatment Effect

Jan. 2025 - Jan. 2026

- Proved consistency of a latent-factorized outcome model and a maximin individualized treatment effect (ITE) estimator.
- Modified and implemented an R optimization algorithm for maximin ITE and ran numerical experiments on simulated data, achieving a 110% average gain in worst-case predictive accuracy over leading baselines.
- Co-first author; prepared a manuscript for submission to the Journal of the American Statistical Association (JASA).

CoxTL: Transfer Learning for Time-to-Event Outcomes

Sept. 2024 - Dec. 2024

- Developed a two-step transfer learning method to enhance time-to-event outcome prediction for the target population with limited sample sizes and published an R package ‘CoxTL’ on GitHub.
- CoxTL outperformed the best baseline by +4.5% C-index for time-to-event prediction under covariate shift on simulated data and achieved the top AUC on real data among 5 methods (4 alternatives).
- First author; paper ‘Adaptive Transfer Learning for Time-to-Event Modeling with Applications in Disease Risk Assessment’ is accepted by the journal *Biostatistics*.

SC2CD: Co-optimizes Spatial Clustering and Cell-type Decomposition

Jul. 2023 - May. 2025

- Derived an iterative framework that combines a graph-based neural network and a matrix-factorization-based decomposition model to jointly improve spatial domain identification and cell-type decomposition.
- Built a Python pipeline for spatial cell-type decomposition and applied it to 3 Visium datasets spanning 14 tissues; increased median PC1 regression R² by 25.4% (*p*-value = 0.0041).
- Co-first author; submitted the paper to the conference RECOMB and published a Python package on GitHub.

FEST: Federated Semi-supervised Transfer Learning for Risk Prediction.

Apr. 2023 - Sept. 2023

- Built an R implementation of FEST to fuse external labeled data with unlabeled internal data and train a global model across multi-site datasets without individual data sharing.
- Our method outperformed benchmarks on simulated data (AUC +4%, MSE -30% vs. best baseline) and ranked first in AUC among all 5 methods (4 alternatives) on real data application.
- First author; paper ‘Enhancing genetic risk prediction through federated semi-supervised transfer learning with inaccurate electronic health record data’ is published in the journal *Statistics in Biosciences (SIBS)*.

HONORS & AWARDS

Outstanding Graduates of Shanghai, China	2024 Summer Research Internship Scholarship from Mitacs	2023
Excellent Monitor of Shanghai Jiao Tong University		2023
Excellent Undergraduate Scholarship, Shanghai Jiao Tong University		2021 & 2022 & 2023
Wu Wenjun Scholarship		2021
Second Prize in National College Students Mathematics Competition		2021
Merit Student of Shanghai Jiao Tong University		2021

TEACHING & EXPERIENCE

Teaching Assistant, Statistical Inference I, Columbia University
Teaching Assistant, Biostatistical Methods II, Columbia University
Member of Rongchang Talent Students Program, Shanghai Jiao Tong University
Member of Student Union, School of Mathematical Sciences, Shanghai Jiao Tong University

Sept. 2025 - Dec. 2025
Jan. 2025 - May 2025
May 2021 - May 2024
Feb. 2021 - Jul. 2022

POSTERS

University Biostatistics Annual Symposium 2025 Poster Competition
Data Science Day 2025 Poster, Columbia University Data Science Institute

2025
2025

PUBLICATIONS & PREPRINTS

Yuying Lu, Tian Gu & Rui Duan. *Enhancing Genetic Risk Prediction through Federated Semi-Supervised Transfer Learning with Inaccurate Electronic Health Record Data*. Statistics in Biosciences (2023).

Yuying Lu, Tian Gu & Rui Duan. *Adaptive Transfer Learning for Time-to-Event Modeling with Applications in Disease Risk Assessment*. Biostatistics (2025). accepted.

Yuying Lu, Lan Luo, & Tian Gu. *ReFIT: Federated Transfer Learning for Sequential Prediction and Uncertainty Quantification Using Streaming EHR Data*. Statistics in Biosciences (2025). submitted.

Yuying Lu, Camille Morencé, Yong Jin Kweon, Archer Y. Yang & Jun Ding. *SC2CD: Co-optimizes Spatial Clustering and Cell Decomposition through Bidirectional Information Flow in Single-Cell Spatial Transcriptomics*. preprint.

Wenbo Fei, **Yuying Lu** (co-first), Yuanjia Wang, & Molei Liu. *Maximin Learning of Heterogeneous Treatment Effect on Multi-Domain Outcomes*. In prep.