Zifan Yu

University of Tennessee – Knoxville, TN

+1 2406443948 • zyu20@uthsc.edu • learningMalanya.github.io

Education

PhD Student in Data Science and Engineering — 4.0/4.0
University of Tennessee - Knoxville (UTK) — *2023-Present*

Master of Science in Biostatistics — 3.7/4.0
University of Washington - Seattle (UW) — *2020-2022*

Bachelor of Science in Mathematics — 3.8/4.0 (Major GPA: 3.9/4.0)
University of Maryland - College Park (UMD) — *2016-2020*

Publications

(* *indicates first-authorship*)

BulkLMM: Real-time genome scans for multiple quantitative traits using linear mixed models [doi]
Zifan Yu*, Gregory Farage, Robert W. Williams, Karl W. Broman, Śaunak Sen. *bioRxiv 2023.12.20.572698*

Software (Packages) Development

- **BulkLMM.jl** [github] - Fast computational methods of linear mixed model association testing for a large number of quantitative traits.
- **MatrixLMnet.jl** [github] - Core functions for performing penalized regression of high-throughput data modeled in matrix linear models.

Research Experiences

Fern spores images characterization using novel computer vision techniques
Work under Dr. Hector Santos-Villalobos, Electrical Engineering and Computer Science, UTK — *Nov. 2024 - Present*

- Researching on fern spore classification and self-supervised segmentation using deep neural networks.
- Developed a front-end web application (Flask app) to facilitate efficient and user-friendly human annotation.

Investigating large vision-language models' capabilities in text-guided segmentation
COSC650 class project, Electrical Engineering and Computer Science, UTK — *Sep. 2024 - Dec. 2024*

- Developed a novel strategy for deploying off-the-shelf large vision-language models to automate medical image segmentation.
- Implemented key components of the model and conducted model evaluation on standard object detection datasets (e.g., MS-COCO).

Fast applications of linear mixed model (LMM) for multiple-trait association scans
Work under Dr. Śaunak Sen, University of Tennessee Health Science Center (UTHSC) — *June 2022 - June 2023*

- Developed computationally efficient algorithms of linear mixed model association genome-scans, utilizing **statistical techniques to allow vectorized computation** and **parallel programming** to achieve further speed-ups.
- Proposed solution achieved **significantly faster (e.g., 20k×) computational speeds** than previous state-of-the-art, GEMMA, for the task to perform genome-scans over large numbers (e.g., 30k) of traits.
- Implemented and compiled solutions into the software package **BulkLMM.jl**.

Applications of nowcasting methods to notifiable disease surveillance
UW Biostatistics capstone project, Washington Department of Health (WADOH) — *Sep. 2021 - Mar. 2022*

- Deployed Bayesian-based methods to predict unobserved hospitalizations (due to missing) of COVID-19 from 2020-2021, given partial data observed by the WADOH - a task referred to as "nowcasting".
- Solution achieved high precision in nowcasting for at least eight days, demonstrating the potential to reduce DOH's exclusion period from 16 days to 8 days, thereby accelerating real-time decision-making.

Sparse matrix linear models (MatrixLM) with the Elastic-net regularization
Internship project at Senresearch, UTHSC — *June 2021 - Sep. 2021*

- Derived proximal gradient descent solutions to evaluating MatrixLMs with the Elastic-net regularization.
- Implemented and compiled solutions into the software package **MatrixLMnet.jl**.

Teaching

COSC 325: Introduction to Machine Learning
Graduate Teaching Assistant, Department of EECS, UTK *Aug. 2024 - Dec. 2024*

- Preparing class materials: created 3/6 programming assignments with solutions and 12/12 weekly quizzes.
- Assisting students: Holding weekly office hours, monitoring and assisting students in class Discord server.
- Grading: Summarizing peer-review scores and grading final project grades.

Oral Presentations

Real-time linear mixed models for association studies on multiple quantitative traits
Complex Trait Community Annual Meeting 2023 (Oral), Virtual *October 11, 2023*
UTHSC Biostatistics Seminar Series (Invited guest speaker), Virtual *September 18, 2023*
ENAR 2023 (Oral), Nashville *March 21, 2023*

Applications of nowcasting methods to notifiable disease surveillance
Washington Department of Health Seminar Series (Student researcher), Virtual *March 23, 2022*

Survival analysis reading group session on the Kaplan-Meier estimator
University of Washington Biostatistics Survival Analysis Reading Group (Session lead), Virtual *July 21, 2021*

Honors

Graduate School Fellowship
University of Tennessee - Knoxville

Meritorious Winner Award
Mathematical Contest in Modeling (MCM) 2020

Skills

- **Programming:** Julia, Python, R, Matlab, Java, high-performance computing, distributed version control (Git), package development, LaTeX
- **Highlighted coursework:** Statistical inference, (classical) ML/DL/RL, mathematical modeling, stochastic processes, data structures and algorithms, object-oriented programming
- **Others:** Effective communication, team collaboration, oral presentation, academic/technical writing