IAN LIU

AI Research Intern, Wake Forest University Center for Artificial Intelligence Research
MS Student in Biostatistics, Brown University School of Public Health
ian_y_liu [at] brown.edu
linkedin.com/in/ian-liu-223010208/

EDUCATION

Brown University - M.S., Biostatistics, Expected May 2026
- Thesis Advisor: Prof. Ying Ma
- Academic Advisor: Prof. Stavroula Chrysanthopoulou
- Courses: Statistical Learning & Big Data, Statistical & AI-Powered Methods for High-Dimensional Genomics Data Analysis, Applied Longitudinal Data Analysis & Multilevel Modeling, Applied Generalized Linear Models, Fundamentals of Probability & Statistical Inference, Statistical Programming

Fei Tian College Middletown - B.S., Data Science, May 2024
- Magna Cum Laude
- Minor: Data Science
- **CS/AI & ML Courses**: Machine Learning & Artificial Intelligence, Data Mining, Data Structures & Algorithms, Database Management, Data Mining, Computational Analysis & Practical Programming, Front End & Back End Web Development, Intro to SAS, Statistical Computing (R)
- **Math/Stats Courses**: Linear Algebra, Intro to Probability, Statistical Theory & Methods, Applied Regression Analysis, Tools for Statistical Theory (Matrix Analysis), Survival Analysis, Calculus (I, II)

RESEARCH & TEACHING

Center for Artificial Intelligence Research, Wake Forest University,

AI Research Intern, Current

- Used graph neural networks to perform classification on Alzheimer's Disease progression
- Performed cortical thickness and curvature registration for 4000+ MRI images using freeSurfer

Northern Research Institute of Health, Visiting Part-Time Research Intern, 2023-Current

- Trained an optimal classifier (selected from more than 15 candidate models) using Pycaret AutoML (automated machine learning) with group K-fold cross-validation on a small dataset of 90 rows, achieving ROC-AUC of 0.875, accuracy of 0.9167, and F1 score of 0.9412 on a holdout dataset
- Reduced data dimensionality by 98% via feature selection with Variance Inflation Factor (VIF), Least Absolute Shrinkage and Selection Operator (LASSO), and Intraclass Correlation Coefficient (ICC)
- Extracted 800+ radiomics features from 136 PET images using Pyradiomics and performed reproducibility analysis via data visualizations, ICC, and Wilcoxon signed-rank tests with Benjamin-Hochberg correction

Brown University, Teaching Assistant, 2024

- Graded assignments, hosted labs & office hours for 21 students in the PHP2506 *Biostatistics for Public Health Research* class
- Covered topics such as hypothesis testing, regression, ANOVA, Stata, etc.

Institute of Statistical Science, Academia Sinica, Statistical & ML Research Intern, 2024

- Worked with [Prof. Tso-Jung Yen](#) to develop an extension of Locally Interpretable Model-Agnostic Explanations (LIME) to explain individual node influence in graph neural networks (GNN) via ElasticNet logistic regression and a novel adaptive class-balanced graph sampling algorithm
- GitHub: github.com/Ianyliu/fatty-liver-explainable-ai
- Poster: drive.google.com/file/d/1ygEee-VJjJPw7h4AEldsCci5rEsx2QCr/
- Evaluated 11 manifold learning and 7 clustering algorithms on DenseNet image embeddings for 135 subjects consisting of more than 2700 images, forming 200,000+ visualizations
- Quantified marginal & conditional image influence of 135 subject classifications of fatty liver with proposed method
- Assessed zero-shot automated segmentation of Segment Anything Model (SAM) for 732 liver ultrasound images

Institute for Informatics, Washington University in St. Louis,

Biomedical Informatics & Data Science (BIDS) Research Intern, 2023

- Worked with [Prof. Ruijin Lu](#) on *Longitudinal Investigation of Cognitive Reserve via Latent Variable Modeling* (paper in progress)
  - Poster: drive.google.com/file/d/19nbGk50l9uUzqIQwbEfdtYdu7Z4anohd
  - Report: docs.google.com/document/d/1uU1VO5fEyUX6yFhaXk4DuDrTJp97ODzk37hpXewO77E
  - Conducted confirmatory factor analysis in *lavaan* using 14 predictors and 3 latent factors to quantify cognitive reserve in dominantly inherited Alzheimer's Disease patients
  - Performed linear mixed effects modeling using *lmer4*, leading to the observation that cross-sectional cognitive reserve indeed declines over time
  - Identified cognitive reserve as a statistically significant interaction for Clinical Dementia Rating with a zero-adjusted (zero-altered) log normal hurdle mixed effects model using *GLMMadaptive*
- Worked with [Prof. Jing Li](#) on *Qualitative Causal Inference for Project ACHIEVE's Transitional Care Strategies* (paper in progress)
  - Report: docs.google.com/document/d/1ZcHExTnMnalTzhDx8P655Oh8IvUuKSiJ4zUlm8HONEQ
  - Conducted qualitative comparative analysis (QCA) and coincidence analysis (CNA) on combinations of 11 transitional care strategies in 29 hospitals to derive 29 causal paths for optimizing patient satisfaction and patient-reported outcomes
  - Performed 4 series of robustness tests (including fit-oriented robustness, case-oriented robustness, etc.) to evaluate holistic robustness of QCA solutions

Northern Horizon/Trillium Learning, Research Associate Intern, 2021

- Database manager & full-stack programmer in JavaScript, jQuery, HTML, CSS, MySQL, & NodeJS working on COVID Toolkit, a visualizer for COVID & vaccination data
- Improved website load speed 3x & reduced memory usage 4x via code optimization, lazy loading, re-modeling MySQL database, & optimizing SQL queries via A/B testing– resulting in 1.2+ million rows of data accessible with lower latency

- Enhanced user experience by designing and implementing an accordion menu that automatically collapses/expands

RESEARCH INTERESTS & SKILLS

Research Interests: spatial transcriptomics, genomics, neuroimaging, explainable AI (XAI), graph neural networks (GNNs), causal inference, computer vision, AutoML

Programming Languages: Python, R, SQL, JavaScript, HTML, jQuery, CSS, Stata

Bioinformatics & Neuroimaging: Scanpy, freeSurfer, PETSurfer, Pyradiomics, MONAI, Nilearn

Other: Pandas, Numpy, PyTorch, Tableau, Git, AWS, Docker, Bash, ETL, web scraping, Scikit-Learn, Matplotlib, ggplot2, GitHub Actions, Catboost, XGBoost

Exposure to: SAS, Terraform, TensorFlow, Keras, Kubernetes, MATLAB, Google Cloud, React JS, Neo4j, PySpark, Java, Golang

INDUSTRY

Hosta.AI, Data Science Intern, 2022-2023

- Saved 5 seconds on average for each computer vision image QA by making Python script more efficient
- Created QA training for annotation/labeling team of 40+ people and evaluated annotation quality of 1000+ samples

Cisco Systems, Technical Intern, 2022

- Full stack programmer working on [Blast Radius Fork](#), a Terraform infrastructure visualization tool with 40+ GitHub stars
- Enhanced app accessibility by building 8 multi-CPU architecture [Docker images](#) (1000+ Docker image pulls)
- Designed and implemented 3 new full-stack features, then hosted app on both AWS EC2 and Kubernetes clusters

Sound of Hope Media, Data Engineer Intern, 2021-2022

- Accelerated data analysis workflow via ETL (Extract, Transform, Load) data pipelines integrating YouTube APIs to extract & transform 1000+ rows of data monthly from 4 YouTube channels
- Automated uploading of 349 podcast episodes to 11 platforms (Apple Podcasts, Spotify, Patreon, etc.) via APIs & browser automation, saving 8+ hours of manual work
- Created a standalone executable app that retrieves & loads YouTube playlist info to Google Sheets, saving data analysts' 2 minutes per video

RELEVANT PROJECTS

Statistical Simulation: Importance of Luck, Talent, Risk Seeking, & Initial Capital in Life

GitHub: [github.com/Ianyliu/whats-important-in-life](#)
R Shiny Web App: [iantheliu.shinyapps.io/RFinalProj/](#)
- Modeled capital of individuals using hierarchical distributions adjustable by user-selected parameters, then measured influence of external and internal factors using stepwise nonlinear regression
- Created 29 animated dynamic visualizations displaying parameter distributions and relationships among variables

Computer Vision Classification of Breast Cancer Tumors via Deep Learning

Notebook:
colab.research.google.com/github/Ianyliu/feitian-courses/blob/main/CIS335%20Machine%20Learning%20%26%20AI/CIS421_Final_Project2_Breast_Cancer_Classification_Ian_Liu.ipynb

- Applied binary transfer learning of ResNet152 using PyTorch to achieve 99.3% accuracy & 0.02 loss on test dataset
- Achieved 90.30% accuracy on multi-class classification of 6 tumors via transfer learning of ResNet152

World Olympics Animated Tableau Dashboard

Tableau Dashboard: public.tableau.com/app/profile/ian.liu2530/viz/IanLiu-Olympics/Olympics

- Wrangled, integrated, and cleaned total of 270k+ rows, 80 columns of Olympics & GDP data via Microsoft PowerQuery and Excel
- Enhanced user experience by dynamic visualizations through 7 different charts and 5 different selectors

FELLOWSHIPS & AWARDS

- AI Business Fellowship, Perplexity AI, 2025
- Todd '87 and Christine Fisher P'17 Graduate Fellowship in Public Health, 2024-2025
- Distinguished College Achievement Award, Fei Tian College Middletown, 2024
- 2x Academic Excellence Award, Fei Tian College Middletown, 2023
- Provost's Outstanding Student's Award, Fei Tian College Middletown, 2023
- (7 semesters) Dean's List, Fei Tian College Middletown, 2020-2023
- Special Mention, Fei Tian College Middletown Hackathon, 2021
- 2x Semi-Finalist of SMART Scholarship, Department of Defense, 2021-2022
- 5th Place Overall (Diamond Rating), Beyond the Stars Dance Competition (Somerset, NJ Regionals), 2020
- Outstanding Performance Award, Northern Academy of the Arts Inaugural School-Wide Dance Competition, 2018
- 2nd Place, In-School American Mathematics Competitions 10, 2018
- Award from the Council of Yunlin County (Taiwan), 2017

Memberships in Associations and Institutional Affiliations

- Institute of Mathematical Statistics