Qile (Keela) Dai

(203)-435-1155 | Atlanta, GA | qile.dai@emory.edu | LinkedIn | GitHub | Google Scholar

Summary

PhD candidate in Biostatistics with extensive experience in AI/ML and statistical modeling gained through 6+ years of research and 2 internships at GSK and Johnson & Johnson. Authorized to work in the U.S. without sponsorship.

Education

Emory University   Aug 2020 - May 2025
Ph.D. in Biostatistics; GPA: 3.99/4.00   Atlanta, GA

Yale University   Aug 2018 - May 2020
M.S. in Biostatistics; GPA: 4.00/4.00   New Haven, CT

University of International Business and Economics   Aug 2014 - Jun 2018
B.S. in Statistics; GPA: 3.70/4.00   Beijing, China

Skills & Interests

Programming Languages: Python, SQL, R, MATLAB, SAS, Bash
ML Frameworks: Scikit-Learn, PyTorch, TensorFlow, Keras, OpenCV, HuggingFace Transformers
Data Analysis: Pandas, NumPy, PySpark, Hadoop, Tableau, Power BI
Platform & Tools: AWS (SageMaker), Linux, Git, Docker

Professional Experience

Johnson & Johnson   May 2024 - Aug 2024
Data Science Intern   Spring House, PA
- Designed a **machine learning framework** leveraging **Random Forest** and **PageRank** to predict drug toxicity
- Implemented the framework using **Python (Scikit-learn, NetworkX)**, improving prediction precision by 5%
- Selected as the highlighted oral presenter (top 10%) at J&J Summer Intern Research Symposium

GlaxoSmithKline (GSK)   May 2023 - Aug 2023
Statistician Intern   Remote
- Designed a **hypothesis testing framework** leveraging **Bayesian regressions** to facilitate drug discovery
- Automated a **pipeline using SQL and R** to facilitate efficient data acquisition, model training, and visualization
- Applied the framework to a dataset with over **500,000+ samples** and **200,000+ features** across 17 diseases, improving drug target identification precision from 42% to 60%

Research Experience

Emory School of Public Health   Aug 2020 - Present
Graduate Researcher   Atlanta, GA
- Project 1: Regression using Summary-level Statistics [Nature Communication paper] [Python package]
  - Designed a **Bayesian regression framework** that uses summary-level statistics to avoid data privacy issues
  - Developed a **Python package** for efficient parallel data processing and model training
- Project 2: Multi-Dimensional Tensor Data Analysis [paper] [R package]
  - Designed an **R package** using **supervised tensor decomposition** for gene expression analysis
- Project 3: Alzheimer's Risk Prediction with Time-Series Data
  - Employed **recurrent neural networks (LSTM)** in **Python (Scikit-learn, PyTorch)** to predict brain pathologies using **longitudinal clinical features**, increasing prediction accuracy by 15%

Projects

Sentiment Analysis for Marketing via **Bidirectional Transformer Model** [code]
- Fine-tuned **BERT** with **PyTorch** and **Hugging Face Transformers** for sentiment assessment of Amazon Reviews, achieving 83% recall rate and 80% precision rate on predicting positive sentiment

Food Vision Classification via **Convolutional Neural Network** [code]
- Fine-tuned EfficientNet-B2 architecture with **TensorFlow** for food image classification, achieving 81% accuracy

Selected Publications

- Published:

  1. Omnibus proteome-wide association study identifies 43 risk genes for Alzheimer disease dementia.
     Tingyang Hu, Randy L Parrish, **Qile Dai**, Aron S Buchman, Shinya Tasaki, David A Bennett, Nicholas T Seyfried, Michael P Epstein, Jingjing Yang.
     The American Journal of Human Genetics, 2024. [tool] [paper]

  2. OTTERS: a powerful TWAS framework leveraging summary-level reference data.
     Qile Dai, Geyu Zhou, Hongyu Zhao, Urmo Võsa, Lude Franke, Alexis Battle, Alexander Teumer, Terho Lehtimäki, Olli T Raitakari, Tõnu Esko, Michael P Epstein, Jingjing Yang.
     Nature Communications, 2023. [tool] [paper]

  3. G2S3: A gene graph-based imputation method for single-cell RNA sequencing data
     Weimiao Wu, Yunqing Liu, **Qile Dai**, Xiting Yan, Zuoheng Wang.
     PLOS Computational Biology, 2021. [paper]

  4. Regulation and characterization of tumor-infiltrating immune cells in breast cancer.
     Qile Dai, Weimiao Wu, Amei Amei, Xiting Yan, Lingeng Lu, Zuoheng Wang.
     International immunopharmacology, 2021. [paper]

- Under Review:

  1. STACCato: Supervised Tensor Analysis tool for identifying condition-related Cell-cell Communications using scRNA-seq data across multiple conditions
     Qile Dai, Michael P Epstein, Jingjing Yang.
     Under review at *Briefs in Bioinformatics.* [R package] [preprint]

  2. CADET: Enhanced transcriptome-wide association analyses in admixed samples using eQTL summary data.
     Taylor Head, **Qile Dai**, Joellen Schildkraut, David J. Cutler, Jingjing Yang, Michael Epstein
     Under revision at *The American Journal of Human Genetics.* [preprint]

- In Preparation:

  1. Cell-Cell Communication Patterns in Alzheimer's Disease Dementia and Mild Cognitive Impairment Vary by Cortical Layers.
     Qile Dai, Shizhen Tang, Jian Hu, Philip L. De Jager, David A. Bennett, Aron S. Buchman, Jingjing Yang, Michael P. Epstein.

Honors and Awards

ASHG - Predoctoral semifinalist for the Trainee Research Excellence Awards (top 5%)	2024
Emory Biostatistics - 2nd place for Best Senior Student Presentation	2024
Emory Human Genetics - Best Student Presentation	2024
ASHG - Reviewer's Choice Abstract (top 10%)	2023
Emory Biostatistics - Best Teaching Assistant Award	2023
Emory Biostatistics - 1st place for Ph.D. method qualifying exam	2021
Yale Biostatistics - $5,000 Colin White Memorial Scholarship (awarded to the best master student)	2019
UIBE - Outstanding Student	2016
UIBE - First-class Scholarship	2016

Teaching Assistant Experience

Introduction to Bioinformatics (BIOS 540)	2024 Fall
High-Throughput Data Analysis using R and BioConductor (BIOS 555)	2024 Fall
Statistical Computing (BIOS 532)	2024 Spring
Modern Regression Analysis (BIOS 526)	2023 Fall
R Programming for BIOS students (BIOS 545)	2023 Spring
Biostatistical Methods (BIOS 508)	2022 Fall
Generalized Linear Models (BIOS 709)	2022 Spring
Survival Analysis Methods (BIOS 522)	2021 Fall
Statistical Methods I (BIOS 500)	2021 Fall

Journal Referee Services

Bioinformatics; Genome Medicine; Human Genetics and Genomics Advances; Nature Aging; PLOS Genetics