Xinyue Zeng

(+1)9845282873 | xyzeng@vt.edu

EDUCATION

University of North Carolina at Chapel Hill (UNC-Chapel Hill) **Chapel Hill, United States**
Gillings School of Public Health
*Master of Science in **Biostatistics*** **May 2024**

Zhejiang University **Hangzhou, China**
School of Mathematical Sciences & School of International Studies
*Bachelor of Science in **Statistics** & Bachelor of Arts in **English Language and Literature*** **June 2022**

Main Courses.
Introduction to Statistical Computing and Data Management, Introduction to Data Science, Intermediate Statistical Methods, Analysis of Categorical Data, Probability and Statistical Inference, Advanced Probability and Statistical Inference, Intermediate Linear Models, Applied Longitudinal Data Analysis, Optimization for Machine Learning and Data Science, Field Observations in Biostatistics, Principles of Epidemiology for Public Health, Introduction to Survival Analysis, Statistical Computing, Principal of Statistics Collaboration, Mathematical Practice, Quantum Information and Quantum Computer, Statistical Learning, Multivariate Statistical Analysis, Linear Regression Analysis, Computer Simulation, Mathematical Software, Stochastic Processes, Partial Differential Equations, Mathematical Statistics, Scientific Computing, Real Variable Functions, Probability Theory, Ordinary Differential Equations.

Technical Skill:
Proficiency in programming R, SAS, Python, JAVA, Matlab.

RELATED EXPERIENCE

AI in Healthcare; Research Assistant **July 2024 - Present**
Computer Science department at Virginia Tech **Blacksburg, USA**
- Conducted research on deep learning algorithms for protein structure prediction with a focus on AlphaFold implementation and optimization.

Cell-type Specific RNA-sequence Genetic Deconvolution; Graduate Research Assistant **Aug 2022 – May 2024**
UNC-Chapel Hill **Chapel Hill, USA**
- Developed as leader a probabilistic non-parametric modeling approach with a nonlinear dynamic covariate matrix system integrating single-cell RNA-seq reference and bulk RNA-seq data from target samples for accurate CTS inference specific to each target sample.
- Did extension simulations which demonstrated that our method achieves 7.99%~123.1% higher Pearson Correlation Coefficient than state-of-art methods in almost every scenario assessed, demonstrating robustness across datasets.

Mixed Effects Modeling via Deep Learning; Student Assistant **Oct 2022 - May 2023**
UNC-Chapel Hill **Chapel Hill, USA**
- Utilized a deep-learning-based semiparametric regression model to identify important biomarkers in R, adjusting for complex confounding structures.
- Did comparative effectiveness analysis of simulations and observational electronic health record data through Linux shell script.

Deep-Learning-based Wildlife Image Recognition; Undergraduate Assistant **Sep 2021 - May 2022**
Zhejiang University **Hangzhou, China**
- Constructed a segmentation model based on UNet to eliminate the interference of the background information and create a segmented data set.
- Compared the performance of classification models including VGG, ResNet, MobileNet and InceptionsV3 on the new dataset and selected the one with the best recognition.
- Introduced Z-score standardization to the previous models and proposed a target recognition model with better accuracy and less loss

Thyroid Medical Image Recognition; Student Assistant **Sep 2021 - Jan 2022**
Zhejiang University **Hangzhou, China**
- Communicate with hospital staff for data collection and pre-processing.
- Compared CNN models in recognizing and classifying the thyroid medical images, and migrated the best method to video data for recognition of potential thyroid.

Deep Learning Model for Coffee Bean Image Recognition; Research Assistant **July 2021 - Nov 2021**
McGill University **Remote**
- Compared CNN models for characterizing images in large-scale network data by using their loss functions and accuracies.
- Introduced the Convolutional Block Attention Module (CBAM) to the architecture of EfficientNet to enhance characteristics of the channel and space, and improved accuracy from less than 50% to 84.24%.

Stock Investment System Based on Big Data Analysis and Statistical Optimization **Mar 2021 - May 2022**
Zhejiang University **Hangzhou, China**
- Developed a mid-frequency Resistance Support Relative Strength (RSRS) model of Decision Tree and Hidden Markov Model using industry factors to predict trends of individual stocks.
- Building an automated transactions platform with an interactive interface using MATLAB and Java.

Personalized Fashion Recommendation Software Based on Big Data **Mar 2020 - Mar 2021**
Zhejiang University **Hangzhou, China**
- Used web crawler to obtain photo collections and personal preference questionnaires from open source.
- Developed face recognition model combining support vector machine (SVM) and principal component analysis (PCA), and created a Local Adversarial Disentangling Network for Facial Makeup and De-Makeup (LADN) model with a complete makeup migration to show its effects.

Study of Biostatistics; Research Assistant **Oct 2020 - Feb 2021**
Zhejiang University **Hangzhou, China**

- Implemented algorithms for penalized regression-based clustering, sequence repetitiveness quantification and de novo repeat detection by weighted k-mer coverage, based on Python and R.
- Used support vector machine (SVM) and Bayesian information criterion to implement the whole-genome sequencing feature selection which improved the accuracy of copy number variation detection.

Study of Statistics; Research Assistant **Apr 2020 - Aug 2020**
Zhejiang University **Hangzhou, China**

- Used R for data analysis to explore the association between sociodemographic variables.
- Cowrote a report with graduate students.

INTERN EXPERIENCE

Software Design Engineer in Test Intern **July 2021 - Sep 2021**
Alibaba **Hangzhou, China**

- Replaced previous static routing algorithm with a Java-based Dynamic Source Routing (DSR) algorithm and introduced a node energy state *E* to judge the state of routes to smooth the process and meet the business requirements.

Quantitative Research Intern **Jan 2021 - Feb 2021**
Tenbagger Capital Management **Hangzhou, China**

- Used web crawlers to build and continuously update a financial database, including transaction rate, gross profit margin and leverage ratio; built a long-short system trading strategy on Python based on it

Assistant Researcher Intern **Apr 2020 - Oct 2020**
Academy of Internet Finance at Zhejiang University **Hangzhou, China**

- Researched mechanisms to evaluate future cities and published a paper detailing the statistical analysis techniques employed including Decision Tree and Cross-Validation Analysis to optimize parameters.

Trainee Quantitative Engineer Intern **Oct 2019 - Feb 2020**
Man Xueqiu Investment Management **Hangzhou, China**

- Developed a RSRS model, which was predicated on Z-score on rolling regression betas between the intra-day highs and lows, achieving an annual return of 19.10%.

PUBLICATIONS

- "Future Cities Report 2021", 2021, *Academy of Internet Finance at Zhejiang University*.
- "Analysis and Research on Macroeconomic Regulation and Control under Market Fluctuations", 2020, *China's Strategic Emerging*, Volume 202009, Page 3.
- "Personalized Fashion Recommendation Software Based on Big Data", 2020, *Student Research Training Program at Zhejiang University*.

PRESENTATIONS

- "Gaussian-Process-Based Cell Type Specific Unmixing of Bulk Expression Profiles", 2024, *ENAR 2024 Spring Meeting*.
- "Wildlife recognition based on deep learning", 2022, *Dissertation Defense for Undergraduate Thesis at Zhejiang University*.
- "Development of stock investment system based on big data analysis and statistical optimization", 2021, *National Innovation and Entrepreneurship Program at Zhejiang University.*

Ongoing Manuscript

- Xinyue Zeng, Will Tang, Quan Sun, Jia Wen, Yun Li. *Gaussian-Process-Based Cell Type Specific Unmixing of Bulk Expression Profiles*. (In progression for submission in 2024).
- Will Tang, Xinyue Zeng, Quan Sun, Jia Wen, Yun Li. *Deep Learning-Based Cell Type Specific Unmixing of Bulk Expression Profiles*. (In progression for submission in 2024).
- Will Tang, Quan Sun, Xinyue Zeng, Jia Wen, Yun Li. *Deconvolution of bulk expression profiles elucidates cell-type specific mechanism*. (Submitted in 2024).

AWARDS

- Biostatistics Travel Award for 2023-2024, *UNC-Chapel Hill*.
- Undergraduate Innovation and Entrepreneurship Award for 2021-2022, *Zhejiang University*.
- Third Prize of the Fifth "LSCAT" Cup Zhejiang Translation Contest for 2020, *China Translators Association*.
- Public Service Model for 2019-2020, *Zhejiang University.*

OTHERS

- **Software**: Statistical Analysis Platform (software copyright in China).
- **Languages:** Mandarin (native), English (fluent).