

# Leshan (Louis) Zhao

(334) 559-2626 | leshan.zhao@emory.edu | linkedin.com/in/leshan-zhao/ | leshanzhao.github.io/

## SUMMARY

- M.S. candidate in **Biostatistics** at Emory University, with a prior master's degree in **Computer Sciences**.
- Statistical Research Assistant at Emory School of Medicine (methodology development). Practical experience in **spatial statistics** and extensive Python programming in analyzing **single-cell spatial transcriptomic** data.
- Proficient in **Python** (5 years), **R** (3 years), and **SAS** (SAS Certified Base Programming Specialist).
- Over 3 years of practical experience in **quantitative data analysis** and implementing **machine learning** algorithms.

## EDUCATION

**Emory University, GA** 08/2023 – 05/2025 (Expected)  
MSPH in Biostatistics & Bioinformatics | Cum GPA: 3.91/4.00

**Auburn University, AL** 09/2021 – 05/2023  
M.S. in Computer Science & Software Engineering | Cum GPA: 4.00/4.00

**The Chinese University of Hong Kong, Shenzhen (CUHK-SZ), China** 09/2017 – 07/2021  
B.S. in Bioinformatics

## RELATED SKILLS

**Programming Languages:** C/C++ | MATLAB | SQL | Linux Bash | **Python** | **R** | **SAS**

**Software/Platform:** ArcGIS | R Shiny | HPC

**Related Courses:** SAS | Relational Database | Machine Learning | Advanced Algorithms | Observational Study | Clinical Trials

**Packages:** Scikit-Learn | PyTorch | Scanpy | Transformer | DESeq2 | dplyr | ggplot2

## WORK EXPERIENCES

**Research Assistant** | AI in Genomics Lab, Emory School of Medicine 10/2023 – present

**Inferring Cell-Cell Communications in Tumor Ecosystems from Spatial Transcriptomics and Multi-omics Data**

- Developing a novel machine learning algorithm using Python to study cell-cell interactions.
  - **Quantified** the **cell-cell colocalization** and **gene-gene colocalization** patterns between B/T cell pairs and PD-1/PDCD1 gene pairs using modified **Bivariate Ripley's K** function in **brain tumor** and pancreatic tissues.
  - Planned to conduct graph convolution for **feature extraction** and test the correlation between B-T cell colocalization and specific gene pair colocalization in cancerous tissue regions compared to normal regions.

**Research Assistant** | Zhaohui “Steve” Qin Bioinfo-Lab, Emory School of Public Health 08/2023 – present

**De novo tissue-specific-Gene discovery in ALS**

- Investigated the abnormal repeat of the C9orf72 gene in Amyotrophic Lateral Sclerosis (ALS) disease.
  - Identified the **top 10 genes** differentially expressed in ocular motor neurons and spinal motor neurons in ALS patients.
  - Identified **repetitive elements** in the paired-end 100-bp-reads data using **HOMER** (a Unix Bioinformatics software).
  - **Formulated** the DE Analysis **pipeline** and the Repetitive Element Analysis **pipeline** for our lab by running SBATCH scripts on a High-Performance-Computing Cluster (Linux software and R). Conducted preprocessing, **batch-effect correction**, and **DE analysis**.

**Machine Learning and Data Science Project** | Auburn University 11/2021 – 12/2022

- Implemented from scratch a **Neural Network** model to **classify** the Kaggle Cardiovascular disease dataset based on 12 **clinical** and **observational** features, with a **1% higher accuracy** on average compared with other **benchmark** models in Scikit-learn.
  - Conducted **parallel** and matrix computing to speed up training by over **600% efficiency**.
- Developed from scratch a Logistic Regression **R package** (**R Shiny App** and **Vignette** guide available on leshanzhao.github.io/).

## COMPETITION

**Kaggle Competition: Feedback Prize - NLP Evaluating Student Writing** 01/2022 – 03/2022

- Won a **silver** medal, ranked **91<sup>st</sup>** out of 2,058 teams (**Top 5%**). Developed an automated feedback tool that identifies elements in student writings by automatically segmenting texts and classifying elements in essays.
- Implemented a **Named Entity Recognition (NER)** solution to categorize each word into one of 7 distinct labels. Optimized two **transformer** models and **improved accuracy by over 0.5%** (rose over **500 in ranking**).
  - Conducted Fast Gradient Sign Method (**FGSM**)-**Adversarial Training** in the word embedding layer (improved accuracy **by 0.2%**). Adopted **BIO tagging** to extend 7 labels to 15 in training and tuned the probability and length threshold for pruning during prediction (improved accuracy **by 0.1%**).
  - Balanced the learning cost between classes by applying **dice loss** to assign higher punishment weight to the classes that occurred less frequently (improved accuracy **by 0.2%**).
  - Conducted **random drop-out** and adopted **average cross-entropy loss** to reduce **30% memory usage** and enable lower batch size while guaranteeing the learning rate and the loss stability during training.

## PUBLICATIONS

1. Dong, X., Jin, Z., **Zhao, L.**, & Guo, Z. (2021, August). BioCanCrypto: An LDPC-coded bio-cryptosystem on fingerprint cancellable template. *2021 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 1-8). IEEE.
2. Huang, H. Y., Li, J., Tang, Y., Huang, Y. X., ... **Zhao, L.**, ... & Huang, H. D. (2021). MethHC 2.0: an information repository of DNA methylation and gene expression in human cancer. *Nucleic Acids Research*, 49(D1), D1268-D1275.