

Jay (Zhuosheng) Liu

zslu@ucdavis.edu | jlk666.github.io | linkedin.com/in/zhuosheng-liu |  Google Scholar

2025 graduate seeking internship/full-time opportunities in SDE/ML*(i485 Pending)*

Experience

Metagenomi Inc: Data Scientist Internship

June 2024 - Sept 2024

- Conducted **data preprocessing and feature engineering** using **highly noisy and high-dimensional data set from inhouse PostgreSQL database** to improve the accuracy and efficiency of predictive models.
- Assisted in the development and implementation of machine learning models (**KNN, Random Forest Regressor achieving 0.88 cosine similarity**) to predict the indel profiles of inhouse Crispr nuclease
- Assisted in prototyping AI solutions by implementing and optimizing deep learning frameworks(**CNN, LSTM achieving 0.995 accuracy of prediction the most common gene editing outcomes**) to enhance the understanding of key features in CRISPR nuclease activity, ultimately guiding in-house CRISPR screening.

Information Fusion and Mining Lab, Graduate Research Assistant co-advised by Dr. Jiawei Zhang

May 2023 - Now

- Developed and optimized **classic machine learning and deep learning** models for microbial pangenome, achieving rapid and accurate pathogenicity assessment of pathogenic bacteria.
- Drove advanced solutions and compared different **parameter efficient-fine-tuning** methods for **vision language model**.
- Systematical evaluation of different **GenAI text-to-video and image-to-video models**.

Data-intensive Biology Lab: Graduate Research Assistant co-advised by Dr. Titus Brown

March 2021 - Now

- Conducted and analyzed large-scale transcriptomics data of microbial-system by building **high-throughput and fully automatic bioinformatics pipelines**
- Organized and maintained large scientific projects on **High Performance Computer** over the cloud and kept **git version control** up-to-date

Education

3.97/4.0 **PhD. in Food Microbiology**, UC Davis | California, USA 2020-25

3.94/4.0 **Master's degree in Computer Science**, UC Davis | California, USA 2023-25

3.84/4.0 **Master's degree in Biotechnology**, Columbia University | New York, USA 2019-20

Research Focus: Applied Bioinformatics | Applied ML/AI | AI4Science

Courses: ECS 271 **Advanced Machine Learning (A+)** | ECS 174 **Computer Vision (A)** | ECS 122A/222A **Algorithm Analysis and Design (A)** | ECS 124 **Bioinformatics Theory and Practice (A)** | STA 141 **Advanced Statistical Computing A-B-C (A)** | ECS 201A **Computer Architecture (A)** | ECS 289 G **Advanced Deep Learning (A+)** | ECS 289 L **Advanced Artificial Intelligence (A+)**

Skills

Programming Python (NumPy, SciPy, pandas, Matplotlib, Jupyter), C/C++, Java, Perl, R, SQL, Scala, Git, Bash Scripting, LaTeX
Software Linux, AWS, Tensorflow, Pytorch, Keras, OpenCV, Snakemake, scikit-learn, DBeaver, Tableau

Projects

Parameter-Efficient Fine-Tuning for Vision-Language Models

June 2023 - Now

Preliminary report available at [Here](#)

- Conducted an extensive study on the application of **Parameter-Efficient Fine-Tuning (PEFT)** methods to vision-language models, particularly focusing on the **CLIP**.
- Conducted comprehensive empirical analysis across various datasets, such as VTAB-1K, to understand the adaptability and effectiveness of PEFT methods in different scenarios.
- Investigated the effectiveness of **prompt engineering** and **adapter technique such as CoOP, VPT and UPT** in improving model performance
- Implemented and optimized various hyperparameters through **systematic grid search** methods to ascertain the most effective settings for the models.

Utilizing the Efficient Segment-Anything Model (ESAM) for Enhanced Biomedical Image Segmentation

June 2023 - June 2024

Preliminary report available at [Here](#)

- Implemented **ESAM** for complex biomedical image segmentation using box and point prompts, significantly improving model performance through fine-tuning.
- Developed a Python-based framework for extracting and analyzing segmentation results with **GPT-4** for insightful medical recommendations.
- Conducted **statistical evaluation** using Intersection over Union (IoU) and Dice Similarity Coefficient (DSC) metrics on over 1500 medical images, enhancing diagnostic accuracy.

Virulence prediction of pathogenic foodborne pathogens using ML/DL based on Pangenome

June 2023 - Now

Manuscript in preparation

- Constructed pangenome (100,000 unique gene feature) using **fully integrated and scalable bioinformatics pipeline** (automation achieved using Snakemake)
- Visualized high-dimensional pangenomic gene feature using **dimensional reduction PCA**
- Trained, optimized, validated and statistically compared fundamental **unsupervised and supervised machine learning model** performance, including **Kmean, GMM, KNN, SVM, RF, Naive Baye**
- Trained and validated **deep neural network (MLP and CNN)** and achieved 93% virulence prediction accuracy
- Obtained valuable biological understandings by applying an **interpretable CNN with Grad-CAM**