Jiajia Li

jiajia.li1@anu.edu.au | +61477801339 | LinkedIn | GitHub | Based in Canberra, Australia

Education

Master of Data Science in AI and Computational Modelling Feb 2020 - Dec 2021
University of Canberra, Australia (2 academic years)
- GPA: 6.188/7.0 (Distinction)

Bachelor of Engineering in Bioengineering Sep 2015 - Jun 2019
Henan University, China (4 academic years)
- GPA: 3.42/4.0 (Distinction)

Research Interests

I am interested in applying machine learning, large language models, neural networks, bioinformatics, statistical methods, and other emerging technologies to address complex challenges in biology and medicine, with the goal of advancing research and improving outcomes in these fields.

Work Experience

Bioinformatics IT Support Officer Oct 2022 - present
Research School of Biology & Biological Data Science Institute, Australian National University (2 year 4 months)

- Delivered 10 short-term training programs on Linux, Python, Git, data visualisation, and other technologies, enhancing the department's research capabilities (details provided in the Teaching Experience section).
- Worked as a sessional tutor for 1 undergraduate course and 2 postgraduate courses, supporting students in understanding course content and assisting with assignment marking.
- Provided bioinformatics support to the Research School of Biology and the John Curtin School of Medical Research by advising students and researchers on bioinformatics tools, debugging coding errors, and identifying appropriate tools for specific research questions.
- Advised and supported students and researchers in utilising computing resources, including servers, the National Computational Infrastructure (NCI), virtual machines, and other common research computing tools.
- Collaborated with the department's IT user group and IT team to manage and administer computing infrastructure, and created/maintained comprehensive documentation, including course materials, teaching workflows, and infrastructure user guides.

Research Experience

Data standardisation and feature extraction with LLMs | Bioinformatician Nov 2024 - present
Supervisor: Dr. Emi Tanaka (2 months)
Biological Data Science Institute, Australian National University

- Co-developed SAI, an R package for data standardisation and feature extraction, utilising local and API-based large language models (LLMs).
- Designated and implemented functions `sai_clean_date` and `sai_clean_address`, integrating LLM-based prompts and R code to streamline tasks like date and address standardisation.
- Contributed to software documentation and co-authored a manuscript on the package development and applications, currently under preparation for publication.

Reduce noise in confocal microscopy images with CNNs | Master's Capstone Research Project Feb 2021 - May 2021
Supervisor: Prof. Girija Chetty (4 months)
School of Information Technology and Systems, University of Canberra

- Independently applied two pretrain CNNs (Noise2Noise and DnCNN) to denoise 12 confocal microscopy images of green fluorescence protein-stained cells, targeting Gaussian noise and improving image quality.
- Processed the images by resizing and cropping, then applied the CNNs to the green channel only. After denoising, recombined the denoised green channel with the original red and blue channels, followed by stitching the images together.
- Compared the performance of the two pretrained CNN models against a control group using MATLAB denoising tools, demonstrating a clear improvement in noise reduction.

Maize epidermis single-nucleus transcriptome analysis | Bachelor's Thesis Sep 2018 - May 2019
Supervisor: Prof. Guiling Sun (9 months)
State Key Laboratory of Crop Stress Adaptation and Improvement, Henan University

- Conducted single-nucleus RNA sequencing on maize epidermal cells, isolating nuclei using fluorescence-activated cell sorting (FACS) and obtaining 25,035 genes for analysis.

- Processed data with Cell Ranger and Seurat, identifying 13 distinct clusters and their differentially expressed genes, while removing batch effects and assessing organelle-encoded genes.
- Performed Gene Ontology (GO) enrichment analysis to assign clusters to specific cell types, including mesophyll, stomata, and epidermal subtypes.
- Identified a sub-population of pavement cells expressing previously unknown genes, and selected marker genes for further validation through RNA in situ hybridisation and transgenic plant generation.

Mutagenesis screening and selection of *Arabidopsis* mutants | Research Assistant — Sep 2017 - Jun 2018
Supervisor: Dr. Yanli Niu — (10 months)
State Key Laboratory of Crop Stress Adaptation and Improvement, Henan University

- Conducted mutagenesis screening for over 500 *Arabidopsis* samples, including planting mutated seeds, applying herbicide treatments to eliminate plants without marker genes, and monitoring growth.
- Collected plant leaves at designated growth stages for DNA analysis.
- Performed DNA extraction, PCR amplification, and gel electrophoresis to identify successfully mutated plants.
- Analysed electrophoresis results to select positive samples for further cultivation and seed harvesting.

Teaching Experience

Short-Term Training Programs | Instructor | Website
Biological Data Science Institute, Australian National University

- Oct 2024, Introduction to Git and GitHub (6 hours)
- Jul 2024, Introduction to Machine Learning (10 hours)
- May 2024, Data Visualisation with Python (10 hours)
- Mar 2024, Snakemake for Bioinformatics (10 hours)
- Mar 2024, Data Management, Reproducibility, and Integrity (2 hours)
- Oct 2023, Introduction to Python Programming (12 hours)
- Oct 2023, Introduction to Supercomputer Gadi (4 hours)
- Aug 2023, Introduction to Linux, HPC, and Variant Calling (12 hours)
- May 2023, Introduction to Python Programming (10 hours)
- Mar 2023, Introduction to Linux, HPC, and Variant Calling (10 hours)

Undergraduate Courses | Sessional Tutor
Research School of Biology, Australian National University

- Sep 2024, Theoretical and Applied Genetics (40 hours)

Postgraduate Courses | Sessional Tutor
Biological Data Science Institute, Australian National University

- Jul 2024, Scientific Reasoning and Consulting in Quantitative Biology (6 hours)
- Jul 2023, Scientific Reasoning and Consulting in Quantitative Biology (6 hours)

Skills

- **Programming Languages**: Python, R, Shell Scripting, MATLAB, SQL
- **Common Python and R Libraries**: Numpy, Pandas, Matplotlib, Seaborn, Plotly, Scikit-learn, XGBoost, TensorFlow, PyTorch, Keras, Dask, Jupyter Notebook, ggplot2, dplyr, tidyr, caret, Bioconductor, rmarkdown, Quarto
- **General Tools**: Git, Snakemake, Conda, Docker, Singularity, SLURM, PBS
- **Bioinformatics Tools**: FastQC, Fastp, Trimmomatic, Cutadapt, BWA, Bowtie2, SAMtools, VCFtools, bcftools, Guppy, Minimap2, Canu, Flye, QIIME2, DADA2, CHOPCHOP
- **Languages**: Proficient in English (PTE Academic 86, IELTS 7.5 - expired), Native Mandarin Speaker, Korean Beginner

Honours and Awards

- Nov 2020, Dean's Excellence Award Semester 2 2020 — University of Canberra
- Jun 2020, Dean's Excellence Award Semester 1 2020 — University of Canberra
- Feb 2020, UC High Achievers Scholarship (25% tuition fee) — University of Canberra
- May 2019, Merit Student Scholarship (1,000 CNY) — Henan University
- Sep 2018, National Encouragement Scholarship (5,000 CNY) — Ministry of Education of the People's Republic of China
- Sep 2018, Merit Student Scholarship (1,000 CNY) — Henan University

- Sep 2017, National Encouragement Scholarship (5,000 CNY) — Ministry of Education of the People's Republic of China
- Sep 2017, Merit Student Scholarship (1,000 CNY) — Henan University
- Sep 2016, Merit Student Scholarship (1,000 CNY) — Henan University

References

Available upon request.