Qingyang Yin

323-347-1532 yinq@usc.edu linkedin.com/in/qyyin0516 github.com/qyyin0516 qyyin0516.github.io

Education

University of Southern California 08/2021 – 05/2026 (Expected)
Doctor of Philosophy in Computational Biology and Bioinformatics (GPA: 3.97/4.0) Los Angeles, CA, USA
Courses: Algorithms, Machine learning, Artificial intelligence, Database systems, Computational molecular biology.

Xiamen University 09/2017 – 06/2021
Bachelor of Engineering in Automation (GPA: 3.81/4.0, Ranking: 1/81) Xiamen, China
Courses: Software fundamentals (Data structure), Principles of computer, Computer networks, Case analysis for big data.

Skills

Programming languages: Python, C, C++, Java, HTML/CSS, JavaScript, SQL, bash, R, MATLAB
Tools: TensorFlow, Keras, PyTorch, Node.js, MapReduce, Hadoop, Linux, Git, LaTeX
Others: deep learning (DNN, CNN, RNN, autoencoder, XAI, transformer, GPT, etc.), bioinformatics (next-generation sequencing toolsets, molecular biology, etc.), high-performance computing

Experience & Selected Projects

University of Southern California, Los Angeles, CA, USA | **Research Assistant** 08/2021 – Present

Project 1: Cancer driver gene and mutation detection with XAI | *Python* 06/2023 – Present
- Utilized **explainable AI** techniques to detect cancer driver genes and mutations with **TensorFlow**.
- Built a novel **explainable DNN** model to predict the mutation status of genes based on the DepMap cancer dependency data. The AUROC can reach 0.96 for some candidate cancer driver genes.
- Designed and implemented a knowledge-primed **autoencoder** to generate driver-like artificial mutations.
- Used **layer-wise relevance propagation** to find important pathways of cancer.

Project 2: Cell type identification with XAI | *Python, R* 08/2021 – 06/2023
- Developed CellTICS, an **explainable DNN** model for cell-type identification and pathway finding in scRNA-seq data.
- Compared the classification performance of CellTICS with other state-of-the-art methods using 6 datasets, demonstrating superior performance with macro F1 scores ranging between 0.9406 and 0.9993.
- Used **activation values** of neural network neurons to find important pathways of each cell type.

Project 3: Dimensionality reduction of images | *MATLAB, Python* 05/2022 – 08/2022
- Implemented the **sparsity and geometry preserving graph embedding** (SGPGE) algorithm for dimensionality reduction of images.
- Resolved the issue of the non-positive definiteness in the eigenvalue problem.
- Compared SGPGE with traditional methods (**PCA, LDA, LPP, SPP**) and its variants using **kNN** classification on the ORL face dataset. Achieved an accuracy of 0.94 with SGPGE.

Project 4: Development and application of Burrows–Wheeler Transform | *C++* 01/2022 – 05/2022
- Used the **DC3 algorithm** to construct the suffix array of input strings.
- Implemented the **Burrows–Wheeler transform** and conducted a query using the compression to determine the exact match count of the pattern in the input string.
- Validated the results by comparing with those obtained from the **KMP algorithm** and checked for linear complexity.

Xiamen University, Xiamen, China | **Undergraduate Researcher** 01/2020 – 06/2021

Project 1: Feature extraction and classification for scRNA-seq data | *R, Python* 06/2020 – 06/2021
- Employed **stacked denoising sparse autoencoders** with **Keras** for feature extraction and built an **ensemble learning** model to predict cell types and ASD status. Developed a framework called scIAE.
- Evaluated scIAE against 11 other feature extraction methods through **t-SNE** visualization, **clustering**, and **SVM** classification. scIAE ranked first in SVM classification performance.
- Compared scIAE with 15 other classification tools using accuracy, mean F1 score, and median F1 score metrics. Paired t-tests indicated that scIAE had greater classification power than most of the other methods.

Project 2: Cell-type-specific predictive models for ASD | *R* 01/2020 – 06/2020
- Performed **recursive feature elimination with cross-validation** to select ASD-related genes.
- Used **partial least squares** to predict ASD status. The average AUROC can reach 0.85 for all cell types.

Publications

[1] **Yin, Q.**, & Chen, L. (2024). CellTICS: an explainable neural network for cell-type identification and interpretation based on single-cell RNA-seq data. *Briefings in Bioinformatics*, 25(1), bbad449.
[2] **Yin, Q.**, Wang, Y., Guan, J., & Ji, G. (2022). scIAE: an integrative autoencoder-based ensemble classification framework for single-cell RNA-seq data. *Briefings in Bioinformatics*, 23(1), bbab508.
[3] Guan, J., Wang, Y., Lin, Y., **Yin, Q.**, Zhuang, Y., & Ji, G. (2021). Cell type-specific predictive models perform prioritization of genes and gene sets associated with autism. *Frontiers in Genetics*, 11, 628539.