

LIYAN WANG

wliyan_2021@tju.edu.cn

+86 136-6208-5263

EDUCATION

Tianjin University
M.E. Chemical Machinery

Tianjin, China | NOW

Tianjin University
B.S. Processing Equipment and Control Engineering (GPA: 3.6/4.0)

Tianjin, China | Jun 2021

Major courses: Numerical Computational Methods and Matlab(94), Database Applied Technique(93), Linear Algebra(90), Processing Equipment and Control Technology and Application(Bilingual, 91.6)

RESEARCH INTEREST

Generally, I am interested in **multimodal and the application of various neural networks in medical scenarios and industry.**

Currently, I am exploring the following areas:

- **Contrastive Learning for Multimodal Fusion:** Applying the concept of contrastive learning to multimodal fusion in tabular data. Through the principle of optimizing representations based on similarity and difference, we can gain more comprehensive insights from complex data. For example, integrating multi-omics data in pan-cancer research.
- **Enhanced Cross-Modal Fusion of Molecular and Image:** With my model expertise, I'm delving into cross-modal fusion. I'm merging transcriptomic and radiomic data to extract richer features for improved subtype classification. This method uncovers hidden patterns, enhancing classification accuracy.

In addition to the current research experience, other research directions that I am interested in are as follows:

- **Deep learning of genome-wide features:** In genome sequencing projects, feature learning from high-dimensional sequence data using recurrent neural networks (RNN) and self-attention mechanisms such as Transformer can capture features of local and global mutation patterns. This helps to identify key mutations associated with the phenotype.
- **Integrating Multimodal Biological Data:** Integrating diverse biological data, including molecular structures and transcriptomics, with textual and functional information. Applying contrastive losses for unsupervised fusion to extract informative latent features. These fused features then guide subsequent tasks.

RESEARCH EXPERIENCE

Tianjin University | Pan-Cancer Multi-omics Data Fusion. | Jan 2022 - NOW

- To comprehensively understand the onset, progression, and treatment of cancer, we employed the concept of contrastive learning to effectively integrate multiple transcriptomic data sets related to pan-cancer.
- It was proved that our method can more effectively integrate multiple transcriptomic data in cancer research, producing superior feature representations for downstream tasks.
- I am going to submit the manuscript of this work as first author soon.

Tianjin University | Gene Functional Classification. | Jan 2022 - NOW

Gene editing technology allows for precise addition, deletion, or modification of yeast cell DNA sequences, thereby altering its genetic information and expression characteristics, which is of significant importance for further investigating gene functions, metabolic pathways, and signaling in yeast cells.

- To efficiently leverage large-scale genomic data and conduct functional reordering of genes, we have developed an automated functional sorting program using Java.
- **My part:** I was responsible for the development of a website for automatic editing of chromosome sequences, aimed at simplifying the sequencing process for non-programming researchers. Among them, the backend of the website is built using SpringBoot, and the frontend uses Bootstrap.

Tianjin University | Distributed parallel processing of methylation data. | Jun 2021 - Jan 2022

- To efficiently analyze the big pan-cancer methylation data and to overcome the co-methylation phenomenon, we proposed a MapReduce-based distributed and parallel-designed partial least squares approach.
- **My part:** I was mainly responsible for data cleaning and screening of about 485,000 probe sites from 1131 cancer samples in the TCGA database, using the empirical Bayesian method(ComBat) to eliminate batch effects on the data, and GO and KEGG enrichment analysis were performed on the 67 specific genes screened out.

- The method identified 22,000 potential methylation loci highly related to early-stage pan-cancer diagnosis, of which 67 loci were also considered pan-cancer signatures based on gene expression.
- This work has been accepted by *Frontiers in Genetics* (2022 Impact Factor = 4.772).

Tianjin University | Graduation Design

| Jan 2021 – May 2021

- To identify codons with feature-specific usage preferences, we utilized statistical learning methods to analyze the correlation between codon usage bias and tRNA copy numbers.
- Using a statistical learning approach based on multiple correlation coefficients (Pearson, Spearman, MIC) to investigate the relationship between codon usage bias and tRNA abundance, we identified three classes of codons with specific usage bias. This has important implications for guiding gene editing in *Saccharomyces cerevisiae*.
- This work was independently completed by me and was rated as an excellent graduation thesis of Tianjin University in 2021.

PUBLICATIONS

1. Qien He, Junxuan Zhu, **Liyan Wang**, En-ci Ding and Kai Song*, DNA methylation loci identification for pan-cancer early-stage diagnosis and prognosis using a new distributed parallel partial least squares method, *Frontiers in Genetics*, accepted, 2022.
2. Junxuan Zhu, Jinhan Zhang, **Liyan Wang**, Hao Huang, Zhibo Zhang, Kai Song* and Xiaofei Zhang*, Progressively Helical Multi-Omics Data Fusion GCN and Its Application in Lung Adenocarcinoma, *IEEE Access*, 2023.
3. Junxuan Zhu, Yifan Tong, Jinhan Zhang, **Liyan Wang**, Qien He and Kai Song*, Machine Learning Reveals Molecular Similarity and Fingerprints in Structural Aberrations of Somatic Cancer, *Symmetry*, 2023.
4. Junxuan Zhu, Yifan Tong, Jinhan Zhang, **Liyan Wang**, Qien He, Kai Song*, Machine Learning Reveals Molecular Similarity and Fingerprints in Structural Aberrations of Somatic Cancer, *Symmetry*, 2023.
5. **Liyan Wang**, Kai Song*, BioCON: A Contrastive Learning Framework for Integrating Multi-Omics Data in Pan-Cancer.

SKILLS

English: CET-6

Coding: Python, Html, Matlab

Machine Learning: Scikit-learn, Pytorch

Others: AutoCAD, ProE, Bootstrap, SW-6, Slicer, Video Editing, Samtools(Linux)

HONORS

Tianjin University First Prize Scholarship	2022
Outstanding Graduation Thesis, Tianjin University	2021
National Inspirational Scholarship awarded by the Ministry of Education	2019, 2020
Honorable Mention, Mathematical Contest In Modeling (MCM/ICM)	2018

OTHERS

- During my undergraduate years, I took **swimming** lessons for three semesters, an experience that greatly fueled my passion for the sport. Currently, I swim 2000 meters every week in my spare time.
- When I initially delved into the field of machine learning, apart from studying literature and books, I also gained substantial algorithmic knowledge through a Steam game called "**while True learn()**." This game covered various concepts, including decision trees, neural networks, and reinforcement learning. To this day, I continue to enjoy honing my skills by playing this game during my leisure hours, successfully progressing through its levels.
- I enjoy creating **vlogs** to document both my studies and daily life through video recordings.