Yiquan Wang
• +86-19537838515 • ethan@stu.xju.edu.cn • Personal Homepage
• GitHub • Google Scholar • ORCID • OpenReview • IEEE Biometrics Council Member

---

Education

National Base for Mathematical Research and Teaching Talents, Xinjiang University Urumqi, Xinjiang
B.S. in Mathematics and Applied Mathematics 2023.09 – 2027.06

Tsien Excellence in Engineering Program, Tsinghua University & X-Institute Shenzhen, Guangdong
Joint Program, X Scholar 2024.06 – 2027.06

Institute of Neurological and Psychiatric Disorders, Shenzhen Bay Laboratory Shenzhen, Guangdong
Visiting Student, Wen Yuan's Research Group 2025.07 – 2025.09

Research Projects

Tsinghua University Tsien Excellence in Engineering Program ESRT 2024.08-2025.10
- **Project:** From Signal to Symphony: Exploring 2D Sequence Representations for Protein Function Prediction.
- Predicting protein function from its primary sequence is a fundamental challenge in computational biology. While deep learning has excelled, the optimal representation of sequence data remains an open question. This study explores protein sonification—the conversion of amino acid sequences into 2D spectrograms—as a representation for this task. To facilitate this investigation, we developed a benchmark dataset of 18,000 sequences spanning 12 functionally diverse protein classes. Our systematic evaluation suggests that the structural transformation from a 1D sequence to a 2D spectrogram may be a key contributor to the model's predictive performance. This observation is supported by ablation studies where models using either purely visual or acoustic features from the spectrogram demonstrated effective standalone performance, suggesting that the representation itself is a key source of this capability. For instance, a model using a sonification map without explicit biophysical meaning achieved 81.08% accuracy, while our biophysically-informed model reached 84.00%, indicating that such domain knowledge may offer a modest performance benefit. When trained from scratch on our dataset, our fusion model achieved performance comparable to or slightly exceeding that of standard transformer architectures like ESM-2 and ProtBERT, suggesting its potential for data efficiency in this specific context. The model's potential for generalizability was further supported by its performance on the external CARE enzyme classification benchmark, where it achieved 90.44% accuracy. Finally, as a proof-of-concept, we explore the utility of our encoding to guide a diffusion model in generating novel GFP variants, which were assessed for structural viability using computational methods. Our work provides evidence suggesting that the utility of sonification in this context may stem largely from its representational structure, offering a perspective on feature engineering for biological sequences.
- Project Repository. Supervised by Prof. Kai Wei.

National Undergraduate Innovation Training Program 2025.04 – 2026.04
- **Project:** Copy Number Variation Conditional Diffusion Model: For Alzheimer's Disease Risk Assessment.
- This project aims to integrate copy number variation (CNV) features from whole-genome sequencing data with multi-dimensional clinical data such as metabolic indicators. Leveraging the success of diffusion models in processing high-dimensional simulation data and protein phenotype prediction, we will construct a comprehensive framework consisting of CNV feature encoding, genomic region attention, and conditional U-Net diffusion modules. This will simulate CNV distribution changes and evolutionary processes in the genome, analyze the specific role of CNV in regulating Alzheimer's disease pathways, and ultimately improve disease risk assessment and early intervention accuracy.
- Supervised by Prof. Kai Wei.

Chinese Academy of Sciences (CAS) Innovation Practice Training Program 2024.11 – 2025.09
- **Project:** A Knowledge Graph-based Q&A System for Heatwave Disaster Adaptation.
- To address the urgent global challenge of increasingly severe heatwaves, this project details an automated pipeline for constructing a domain-specific knowledge graph from academic literature. The system leverages a Large Language Model to extract relational triples from unstructured text. A novel workflow then standardizes entities by using SentenceTransformer

embeddings and a FAISS index for semantic clustering, followed by LLM-based canonical name generation. The resulting knowledge graph powers a hybrid KG-RAG query engine that retrieves factual subgraphs to provide context for an LLM, enabling the generation of precise, source-grounded answers for complex queries on heatwave adaptation.
- Supervised by Researcher Yong Ge.

Provincial Undergraduate Innovation Training Program   2024.03 – 2025.06
- **Project:** Research on the Generation Cyclability of Cartesian Product Graphs and Related Problems.
- The k-ary n-cube (Q_n^k) is a critical topology for large-scale computing systems powering modern AI and HPC workloads, where fault tolerance is paramount. Traditional fault models, which assume faults are independent and random, yield unrealistic resilience estimates because real-world failures are spatially correlated, manifesting as topological clusters. This paper introduces the Region-Based Fault (RBF) model, a new paradigm that addresses this gap by directly modeling this spatial correlation. Our primary contribution is a proof that for odd $k \geq 3$ and $n \geq 2$, the Q_n^k remains Hamiltonian-connected—a property vital for deadlock-free routing and efficient task scheduling—under a sufficient set of RBF conditions. We present a constructive algorithm that finds a Hamiltonian path by leveraging an adaptive decomposition strategy. Experimental analysis demonstrates that our approach significantly enhances fault tolerance and remains robust far beyond its conservative theoretical guarantees. This work provides a practical, high-performance solution for maintaining connectivity in systems where fault clustering is prevalent.
- Project Repository. Supervised by Prof. Eminjan Sabir.

Professional Experience

Beijing Frontier Research Center for Biological Structure, Tsinghua University   Beijing, China
Intern   2025.07 – Present
- Developed the Peptide Design Competition Website, the Polypeptide Structure Database and engaged in protein and polypeptide design research.

Institute of Software, CAS & Huawei Mindspore Community   Remote
Intern   2024.09 – 2025.03
- Implemented a VGG19-based model for Pollock-style art generation, focusing on fractal and turbulent feature extraction and the creation of NFT-based digital labels. Featured on Huawei's official Wechat.

Selected Publications

[1] **Wang, Y.**, Cai, M., Dong, Y., et al. (2025). From Signal to Symphony: Exploring 2D Sequence Representations for Protein Function Prediction. *Journal of Chemical Information and Modeling* (JCR Q1, CAS Q2 Top).

[2] **Wang, Y.***, Cai, M., & Huang, T. Y. (2025). AI for disease prediction: Performance insights and key limitations. *Journal of Clinical Neuroscience, 138*, 111360.

[3] **Wang, Y.***, Huang, T. Y., Gao, Q., & Zhang, J. (2025). HeDA: An Intelligent Agent System for Heatwave Risk Discovery through Automated Knowledge Graph Construction and Multi-layer Risk Propagation Analysis. *arXiv preprint arXiv:2509.25112*.

[4] Wang, X., **Wang, Y.**, & Huang, T. Y. (2025). Crypto-ncRNA: Non-coding RNA (ncRNA) Based Encryption Algorithm. *ICLR 2025 Workshop*. (Co-first author).

[5] Wang, X., **Wang, Y.***, & Pan, J. (2025). Digital Art Creation and Copyright Protection in Pollock Style Using GANs, Fractal Analysis, and NFT Generation. *ICLR 2025 Workshop*. (Co-first author).

[6] Wang, X., **Wang, Y.***, Xu, L., et al. (2024). Octopus Inspired Optimization (OIO): A Hierarchical Framework for Navigating Protein Fitness Landscapes. *arXiv preprint arXiv:2410.07968*. (Co-first author).

[7] Zhao. H., Shi. W., Liu Chang., **Wang, Y.**. (2025). Boundary Box-Guided Targeted Adversarial Attacks with Semantic Perturbation. *In 2025 IEEE international conference on systems, man, and cybernetics.* IEEE. (CCF-C).

[8] **Wang, Y.***, Zhang, J., & Chang, Y. (2024, November). A probability prediction model for flood disasters based on Multi-layer Perceptron. In *Journal of Physics: Conference Series* (Vol. 2905, No. 1, p. 012003). IOP Publishing.

[9] Wang, J., & **Wang, Y.*** (2024, September). Multi-stage Crop Planting Strategy optimization Model Based on PSO Algorithm. In *2024 3rd International Conference on Electronics and Information Technology (EIT)* (pp. 915-919). IEEE.

[10] **Wang, Y.**, Zai, J., Liu, Z., et al. (2025). Hamiltonian Connectivity in k-Ary n-Cubes under a Region-Based Fault Model. (under review).

[11] **Wang, Y.***, Zuo, J.*, Zheng, T. (2025). Boosting MOOC Engagement: A Social-Constructivist Approach using Graph Neural Networks. (under review).

Awards & Honors

- **SynBio Challenges**: Silver Award $\times$ 2 — 2025
- **2025 X-Fusion "Global Innovator Fusion Conference"**: Best Poster Award — 2025
- **The Mathematical Contest in Modeling (MCM)**: Honorable Mention — 2025
- **Alibaba Cloud Tianchi University Student Competition**: National Finals, 17th Place — 2024
- **APMCM Asia-Pacific Mathematical Modeling Competition**: National Third Prize — 2024
- **"Alpha Egg Cup" National Go Championship**: 15th Place — 2024
- **Xinjiang Youth Amateur Go Competition**: 53rd Place — 2024
- **Hunan Province Spring Cup Go Competition**: 7th Place — 2024
- **National Youth Intellectual Sports Meeting Go Competition**: 9th Place — 2024
- **Xinjiang University Vulnerability Reporting Honor** — 2023
- **"Tianshan Fixed Network Cup" Cybersecurity Skills Competition**: 7th Place, Xinjiang Region — 2023

Academic Activities

Peer Reviewer:

- NeurIPS 2025 AI for Science Workshop
- NeurIPS 2025 MATH-AI Workshop
- ICLR 2025 Workshop on AI for Nucleic Acids
- ICLR 2025 Workshop on GenAI Watermarking
- ICML 2025 Workshop on AI for Math
- Mini-Reviews in Medicinal Chemistry
- Current Science
- F1000 Research

Academic Engagement:

- Tsinghua University-Peking University Center for Life Sciences Summer Camp — 2025.07
- Shenzhen Bay Laboratory / Shenzhen Medical Academy of Research and Translation Summer Research — 2025.07 – 2025.09
- Tsinghua University Tsien Excellence in Engineering Program – Zero One Scholar — 2024.06 – 2027.06
- AI Winter School, Brown University Department of Physics — 2025.01
- CAAI Artificial Intelligence and Technology Ethics Training Course — 2024.09 – 2024.12
- Fudan University Summer School of Mathematical Logic — 2024.08
- Jinan University Guangdong Thousand Villages Survey Project — 2024.08
- Wuhan University National Tianyuan Mathematics Center Discussion Class — 2024.03 – 2024.06

Website Development

- **Peptide Design Competition, Tsinghua University**: https://www.fbs.frcbs.tsinghua.edu.cn/competition/2025Peptide/
- **Polypeptide Structure Database, Tsinghua University**: https://www.frcbs.tsinghua.edu.cn/cpdb/
- **Tong Wang (Tsinghua University) Research Group**: https://tongwang.vercel.app/

Skills & Interests

Technical Skills: Python, C/C++, MATLAB, LaTeX, Linux, HTML/CSS/JavaScript

Research Interests: Artificial Intelligence, Deep Learning, AI for Science, Bioinformatics, Computational Biology, Mathematical Modeling, Neuroscience

Languages: Chinese (Native), English (Professional Working Proficiency)