HONGLIN WANG

honglin.wang@uconn.edu • (860) 771-8889 • harry-wang12.github.io • linkedin.com/in/honglin-harry-wang/

OVERVIEW

- 5+ years of experience in Data Mining, Machine Learning, Deep Learning, and Computer Vision
- Proficient in analyzing large-scale datasets using various Machine Leaning and Statistical models

EDUCATION

UNIVERSITY OF CONNECTICUT, **School of Engineering**   Storrs, CT
Ph. D. in Computer Science and Engineering   *March 2024*
Research interests: Developing and applying machine learning and deep learning in larger-scale bio-datasets and delivering potential biology pattern.
Awards: Predoctoral prize for research excellence; Doctoral Dissertation Fellowship.

UNIVERSITY OF CONNECTICUT, **School of Engineering**   Storrs, CT
M. Sc. in Computer Science and Engineering   *Jan 2019*
Master Thesis: A Method to Score Pathways Using Heuristic Rules.

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA   Chengdu, China
School of Information and Software Engineering   *May 2017*
B. Eng. in Software Engineering (Mainframe Computers)
Awards: First class of People Fellowships in 2014 and 2015.

WORK AND RESEARCH EXPERIENCE

UConn Computer Science and Engineering Department   *Storrs, CT*
Research Assistant   Jan 2019 − *Present*

- Apply Graph Neural Network and Variational Autoencoders to infer the cell communication with large spatial genomic data. (Transformer, GNN)
- Develop and deploy pipelines for analysis of large-scale biology dataset using multiple machine learning and statistical methods. (Dimensional reduction, Clustering, Stratification, Classification and Prediction)
- Design Machine Learning models to classify cell types with large scale multi-omics data.
- Develop and apply methods to identify cells from cell images. (Computer vision)

Biogen   Cambridge, MA
Co-op, Advanced Analytics, Digital Health   *Jun 2023 − Aug 2023*

- Implement pipeline to compare machine learning and statical models for generating risk score for patient using baseline clinical trials factors. (Feature selection, prediction and model performance)
- Develop and implement algorithm to optimally stratify the patients based on risk scores. (Stratification)
- Build and deploy a user-friendly website to visualize the stratification result.
- Discuss and find the clinical usefulness and superiority of this work with the clinicians.

UConn Information Technology Services   Storrs, CT
Web developer / Software Engineer   *May 2019 – Jun 2023*

- Collaborate with developers from variety departments to build interactive website to visualize dataset from scientific databases. (PHP, Python, JavaScript, MySQL, JSON)

TECHNICAL CAPABILITIES

- Machine Learning & Deep learning: Graph Neural Networks, Convolutional Neural Networks, SVM, Linear Regression, K-NN; K-means, Leiden, DBSCAN; PCA, t-SNE, UMAP; AUC, ROC, F1-score
- Data Visualization: PowerBI, Matplotlib, Plotly, Python Dash, Cytoscape, Tableau
- Programming Languages: Python, SQL, PHP, JavaScript, Java, R, C/C++, MATLAB
- Statistical skills: Probability, Distribution, Hypothesis testing
- DBMS: MySQL, MongoDB
- Miscellaneous Tools: Jupiter Lab, GitHub, Apache, Scikit-learn, Pandas, NumPy, PyTorch, OpenCV, Scikit-image.

RECENT PROJECT

A semi-supervised graph neural network to identify cell type and infer cell communication *Storrs, CT*
Research project Jan 2023 − *Present*

- Segment cell from tissue images using self-develop code with image thresholding, noise removing and watershed method.
- Map cell into node from a graph and design an algorithm to link the nodes.
- Build Graph Convolution Networks with trainable edge weight.
- Train the model with only 5% human-labeled sample and tune the model with validation dataset.
- Reach the 95.53% of accuracy in 4 classes of cell type classifying comparing to 89.47% accuracy in MLP model when applying with mouse brain dataset.
- Design code using heuristic rule to infer cell communication with trained edge weight from the trained model.
- Collaborate with experts from neurology department to validate the inference results.

RECENT PUBLICATIONS

Conferences

- **[23' BIBM]** [Submitted] **Honglin Wang**, Chenyu Zhang, Merissa Olmer, Hannah Swahn, Martin K. Lotz, Peter Maye, David Rowe, and Dong-Guk Shin, GeneNet3D: a 3D-CNN based framework to infer gene pathway crosstalk using both bulk and single cell human cartilage datasets, *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2023*.

- [**22' ACM-BCB**]. **Wang, Honglin**, Pujan Joshi, Chenyu Zhang, Peter F. Maye, David W. Rowe, and Dong-Guk Shin, rCom: a route-based framework inferring cell type communication and regulatory network using single cell data, 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2022.

- [**22' BIBM**]. C. Zhang, P. Joshi, **Wang, Honglin**, S. -H. Hong, R. Yan and D. -G. Shin, Pola Viz Reveals Microglia Polarization at Single Cell Level in Alzheimer's Disease, *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2022.*

- [**21' BIBM**]. **Wang, Honglin**, Pujan Joshi, Seung-Hyun Hong, Dong-Ju Shin, and Dong-Guk Shin, ctBuilder: A framework for building pathway crosstalks by combining single cell data with bulk cell data, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021.

- [**21' BIBM**]. Joshi, Pujan, **Wang, Honglin**, Salvatore Jaramillo, Seung-Hyun Hong, Charles Giardina, and Dong-Guk Shin, Identification of Crosstalk between Biological Pathway Routes in Cancer Cohorts, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021.

- [**20' BIBM**]. **Wang, Honglin**, Pujan Joshi, Seung-Hyun Hong, Peter F. Maye, David W. Rowe, and Dong-Guk Shin, cTAP: A Machine Learning Framework for Predicting Target Genes of a Transcription Factor using a Cohort of Gene Expression Data Sets, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020.

- [**20' BIBM**]. Joshi, Pujan, Brent Basso, **Wang, Honglin**, Seung-Hyun Hong, Charles Giardina, and Dong-Guk Shin, Identification of Key Biological Pathway Routes in Cancer Cohorts, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020.

Journals
- [**22' BMC Genomics**]. **Wang, Honglin**, Pujan Joshi, Seung-Hyun Hong, Peter F. Maye, David W. Rowe, and Dong-Guk Shin. Predicting the targets of IRF8 and NFATc1 during osteoclast differentiation using the machine learning method framework cTAP, BMC genomics 23, no. 1 (2022): 1-18.

 [**22' Methods**]. Joshi, Pujan, Brent Basso, **Wang, Honglin**, Seung-Hyun Hong, Charles Giardina, and Dong-Guk Shin. rPAC: Route based pathway analysis for cohorts of gene expression data sets. Methods 198 (2022): 76-87.