Dr. Wan Xiang Shen

Personal website: www.shenwx.com   **Email:** wanxiang_shen@hms.harvard.edu
Address: Countway lib. room 310, 10 Shattuck St, Boston, MA 02115, USA

Education

National University of Singapore, Singapore.   Sept. 2019 - Aug. 2022
    Bioinformatics and Drug Discovery
    Department of Pharmacy
    Doctor of Philosophy (Ph.D.)

Tsinghua University, Beijing, China.   Sept. 2013 - Jul. 2016
    Cheminformatics
    Department of Chemistry
    Master of Science (M.S.)

Northwest University, Xi'an, China.   Sept. 2009 - Jul. 2013
    Major: Life science and technology (National Science Talent Base Class)
    Minor: Economics
    Bachelor of Science (B.S. Double Degree)

Work Experience

- Aug. 2023 - Present: QUAD Postdoctoral Fellow
  - Department of Biomedical Informatics (DBMI) at HMS, Boston, USA
  - Leader: Marinka Zitnik, PhD
- April 2022-Aug. 2023: Postdoc Fellow
  - Department of Chemistry, NUS, Singapore
  - Leader: Assoc. Prof. Wujie
- Aug.2018 - Sept. 2019: Research Fellow
  - Megvii Research (Face++), Beijing, China
  - Leader: Dr. Sun Jian
- Jun. 2017 - Aug.2018: Data Scientist
  - Tsinghua Data Innovation Base (D-Lab), Institute of Data Science, Beijing, China
  - Leader: Prof. Morgan C. Wang
- Aug. 2016 - Jun.2017: Data Mining Engineer (Intern)
  - Ping An Technology, Shanghai, China

Publications

1) **Shen W X,** Zeng X, Zhu F, et al. Out-of-the-box deep learning prediction of pharmaceutical properties by broadly learned knowledge-based molecular representations[J]. *Nature Machine Intelligence,* 2021, 3(4): 334-343.

2) **Shen W X,** Liu Y, Chen Y, et al. AggMapNet: Enhanced and Explainable Low-Sample Omics Deep Learning with Feature-Aggregated Multi-Channel Networks[J]. *Nucleic Acids Research.*, 2022, 50(8): e45-e45.

3) **Shen W X,** Tan Y, Chen Y, et al. Shen W X, Liang S R, Jiang Y Y, et al. Enhanced metagenomic deep learning for disease prediction and consistent signature recognition by restructured microbiome 2D representations[J]. *Patterns*, 2023, 4(1).

4) **Shen W X***, Cui C*, Shi X C, et al. Online triplet contrastive learning enables efficient cliff awareness in molecular activity prediction[J]. *Under Review in Nature Machine Intelligence*, 2023.

5) Cheng K P*, **Shen W X***, Jiang Y Y, et al. Deep learning of 2D-Restructured gene expression representations for improved low-sample therapeutic response prediction[J]. *Computers in Biology and Medicine,* 2023, 164: 107245.

6) Liu Y, Sundah N. R, Ho R. Y., **Shen. W. X.,** et al. Bidirectional growth of templated DNA nanobarcodes for multiplexed profiling of complex protein interactions[J]. *Under review in Nature Biomedical Engineering*, 2023.

7) **Shen W X,** Chen S. Y., Liu F., *et al.* Predicting Enzymatic Hydrolysis Half-lives of New Chemicals Using Support Vector Regression Models Based on Stepwise Feature Elimination [J]. *Molecular Informatics*, 2017, doi:10.1002/minf.201600153.

8) Jin, Y., Du, N., Huang, Y., **Shen, W. X.,** Tan, Y., Chen, Y. Z., ... & Tan, C. Fluorescence Analysis of Circulating Exosomes for Breast Cancer Diagnosis Using a Sensor Array and Deep Learning. *ACS sensors,* 2022, 7(5), 1524–1532.

9) Wong, L. R., Tan, E. A., Lim, M. E. J., **Shen, W. X.,** Le Lian, X., Wang, Y., ... & Ho, P. C. L. Functional effects of berberine in modulating mitochondrial dysfunction and inflammatory response in the respective amyloidogenic cells and activated microglial cells–In vitro models simulating Alzheimer's disease pathology. *Life Sciences,* 2021, 282, 119824.

10) Chia, W., Gomez-Lorenzo, M. G., Castellote, I., Tong, J. X., Chandramohanadas, R., Thu Chu, T. T., **Shen, W. X.,** ... & Tan, K. S. High-Content Phenotypic Screen of a Focused TCAMS Drug Library Identifies Novel Disruptors of the Malaria Parasite Calcium Dynamics. *ACS Chemical Biology*, 2021, 16(11), 2348-2372. (Cover)

11) Zeng, X., Yang, X., Fan, J., Tan, Y., Ju, L., **Shen, W. X.,** ... & Chen, Y. Z. MASI: microbiota—active substance interactions database. *Nucleic Acids Research*, 2021, 49(D1), D776-D782.

12) Feng T. T. *, Cao W.*, **Shen W X,** *et al.* Arctigenin inhibits STAT3 and exhibits anticancer potential in human triple-negative breast cancer therapy [J]. *Oncotarget*, 2017, 8(1):329-344.

13) Li L, **Shen W X,** Zheng J., *et al.* Mutation of the conserved GRG motif and decreasing activity of human RNase H2 [J]. *Open Life Science.*, 2015, 10(1).

Interviews & Talks

1) **Shen W X** and Chen Y Z. Towards ordered omics data science: Researchers in turning metagenomic chaos into image-like patterns [J], Cell Press *Patterns*, the People of Data special issue, Jan. 15, 2023, https://doi.org/10.1016/j.patter.2022.100673.

2) **Shen W X**. Invited talk on Knowledge-Based Small Molecule Drug Representation and Property Prediction, Talk, Yu Muhan Live open seminar, Zoom Meeting, Beijing, Berlin, Singapore, Mar. 20, 2022. https://www.bilibili.com/video/BV1tS4y1U7g6.

3) **Shen W X**. Invited talk on Data Representation and Convolutional Neural Networks in Omics Machine Learning. Talk, Syncedreview, Zoom Meeting, Beijing, Singapore, Feb. 24, 2021. https://www.bilibili.com/video/BV1eP4y1c7pN.

Academic and Project Experience

Computer Vision and Data Processing Frameworks   Feb.2018 - 2019
- ◆ Developing algorithms and toolbox for automatic facial data labelling and de-blurring by GANs, clustering the images based on machine learning and creating customer profile and consuming behavior database for retail companies.
- ◆ Developing model and pipeline for human facial detection, recognition, and tracking, de-blurring the human face images.
- ◆ Building, operation and maintenance of big data platform on clusters by Cloudera, including Hadoop, Hive, Spark and Hue, and responsible for processing and analyzing structured and unstructured data acquired or generated during video data processing.

Automating Model Toolbox in Data Science   Sept. 2016 - Feb.2018
Algorithm & Software Development：
- ◆ Developed a biomedical statistics package BioStat: https://github.com/shenwanxiang/BioStat
- ◆ Developed a post-prune method in decision tree model (based on cost complexity pruning (CCP) algorithm): https://github.com/shenwanxiang/sklearn-post-prune-tree
- ◆ Developed a time-series data analyzing and modelling tool for incidence of hand-foot-and-mouth and influenza: https://github.com/shenwanxiang/tsp
- ◆ Developed a stepwise feature elimination/addition method: https://github.com/shenwanxiang/SFE
Applications：
- ◆ Manipulated data with Apache Hive SQL (data integration, filling, cleaning and query for personal medical information).
- ◆ Data mining for medical institution information and patient visit behavior (clustering and association rule analysis).
- ◆ Constructed knowledge graph for medical insurance anti-fraud detection, relationship networks analyzing between medical institutions and patients.

Human Resource Data Analysis of Bank   Dec. 2017 - Feb.2018
- ◆ Building highly explanatory models (such as KPI, KBI, KCI models) for evaluation of the competitiveness index of the customer manager of the China Construction Bank(CCB) based on available data with statistical and machine learning technologies.
- ◆ Data quality report for human resource data by using Pandas, and to present key findings using data visualization techs and applications.

Machine Learning and Statistical Modeling in Agricultural   Jun. 2017 - Feb.2018
- ◆ Developed supply-demand balance model (production, consumption, import, export, and inventory for a variety of agricultural products) and prediction models of time-series (ARIMA, LSTM) for Agricultural data of AII, CAAS (Agricultural Information Institute (AII) of Chinese Academy of Agricultural Sciences).
- ◆ Agricultural data cleaning, filling, visualization and one-dimensional statistical analysis.

Machine Learning & Deep Learning in Drug Discovery   Nov. 2013 - Jun. 2016
- ◆ Mapped chemical space to reveal pharmacokinetic characteristics of approved and clinical drugs

(2-D hierarchical clustering using structure features and physicochemical properties).
- Developed algorithm of Stepwise Feature Selection in SVR model (Find best number of features to prevent model from overfitting or underfitting).
- Developed algorithm of generating framework and scaffold fingerprint(bite) features for chemical drugs (hunting the scaffolds and frameworks based on maximum similar substructure after hierarchical clustering).
- Developed pharmacokinetic patterns for drug combinations (comparison the pharmacokinetic (Half-life, Clearance, and Volume distribution for chemical drugs) differences between approved and clinical drugs).
- Developed Decision tree, Random Forest, Adaboost+, Bayes, SVM models for drug pharmacokinetics property prediction.
- Developed *In silico* prediction models of cytochrome P450 inhibitors for virtual screening (using 3D Convolution Neural Network model by the 3-D electron cloud density data of each compound).
- Constructed database for drugs pharmacokinetic parameters and inhibitors/substrates of drug metabolizing enzymes, practiced in web crawler technology for data collection.

Biological and Genetic Data Analysis   Sept. 2014 - Jun. 2015
- Developed feature extraction method for DNA sequence and amino acid sequence (k-mer extraction method).
- Developed RNN model to predict metastasis-related genes in Breast Cancer based on direct DNA sequence (gene expression in the cell pseudopods) and indirect feature extraction method.

Graph Based Drug Design and Compute Biology   Mar. 2013 - Aug. 2015
- Manipulated drug-design software (Discovery Studio, Sybyl) for molecular docking, protein structure modeling and pharmacological property calculation.
- Manipulated software of NAMD and Amber for molecular dynamics simulation (simulation of RNase H2 binds with DNA) to reveal the structural conformation of the enzymes.
- Teaching assistant for Computer-Aided Drug Design class.

Biotechnology Experiments   Mar. 2012 - Aug. 2013
- Transgenic experiment in model plant Arabidopsis (detection of transgenic plants using green fluorescent protein GFP as a reporter gene).
- Directed saturation mutagenesis experiment in RNase H2 Protein (Both biological experiment and computer dynamics simulation results reveal the GRG motif is vital for the enzyme activity).

Awards and Certificates

- "Mei Yiqi" Memorial Scholarship of Tsinghua University.
- "Shenzhen, Hong Kong and Macao" Innovation and Entrepreneurship Excellence Award.
- First Class Scholarship of Northwest University.
- National Inspirational Scholarship of Northwest University.
- Second Prize for School Mathematics Modeling Contest of Northwest University.
- Basic Authentication for Global Mathematical Modeling Ability.
- First prize for School "Challenge Cup" Entrepreneurship Competition of China (team leader).

- Challenge Cup" Bronze Award in Shaanxi Province of China (team leader).
- Gold Medal on Proteomics for the AI4Science.io hackathon 2022 (team leader).
- Gold Medal on Chemistry for the AI4Science.io hackathon 2022 (team member).
- Second Prize on Open Category for the AI4Science.io hackathon 2022 (team member).

Service and Leadership

- Vice chair of American Association of Pharmaceutical Scientists (AAPS) NUS student Chapter: Oct.2020 - Sept.2021