CONTACT

Phone
+1(431)3376843

Email
ysun2443@uwo.ca

GitHub
https://github.com/syan1992

SKILLS

- **Programming&Frameworks**
Python, PyTorch, TensorFlow

- **Deep Learning:**
Graph Neural Networks, Generative Model (Transformers, VAEs), Contrastive Learning

- **Bioinformatics:**
RDKit, molecular docking (AutoDock Vina), PLIP

- **Frontier Tools & Models:**
AlphaFold2, ESMFold, Boltz-2

Yan Sun

PhD in Computational Biology & Computer Science with 4+ years of experience developing state-of-the-art ML methods for drug discovery, including molecular property prediction and drug–target interaction. Strong track record of publications (Bioinformatics, Commun. Chem., Nat. Commun.) and industrial collaborations (GE Healthcare, Servier).

Research Highlights

1. **Representation Learning**
   - Improved the ability of GNNs to capture long-range molecular dependencies by augmenting GNNs with xLSTM-based message passing, enhancing representation learning and predictive performance. (Commun. Chem., 2025).
   - Extended contrastive learning to regression tasks to structure embeddings by target values, improving molecular property prediction (BIBM, 2024).
2. **Imbalanced Regression**
   - Proposed a distribution-aware Mixture-of-Experts model to address data imbalance in molecular property regression, improving performance on rare but important regions without sacrificing generalization (Under Submission).

Work Experience

CT Image Quality Engineer
GE Healthcare 08/2017 - 12/2020

Data Science Intern
Servier Canada Inc., Quebec, Canada 07/2021 - 08/2022

Education

Visiting Student in Computer Science
Western University, Canada 06/2023 - 09/2025

PhD in Computer Science
University of Manitoba, Canada 09/2020 - 09/2025

Master in Computer Technology
Xiamen University, China 09/2014 - 06/2017

Bachelor in Cognitive Science and Technology
Xiamen University, China 09/2010 - 06/2014

Publication

Published articles

- **Sun, Y.**, Li, Y. Y., Leung, C. K., & Hu, P. (2024). iNGNN-DTI: prediction of drug-target interaction with interpretable nested graph neural network and pretrained molecule models. **Bioinformatics** (Oxford, England), 40(3), btae135. https://doi.org/10.1093/bioinformatics/btae135
- **Sun, Y.**, Islam, M., Zahedi, E., Kuenemann, M., Chouaib, H., & Hu, P. (2022, December). Molecular Property Prediction based on Bimodal Supervised Contrastive Learning. In 2022 IEEE International Conference on Bioinformatics and Biomedicine (**BIBM**) (pp. 394-397). IEEE.
- Lu, Y., Li, Y. Y., **Sun, Y.**, & Hu, P. (2025). FusionCLM: enhanced molecular property prediction via knowledge fusion of chemical language models. **Journal of Cheminformatics**, 17(1), 1-12.
- Zhang, C., **Sun, Y.**, & Hu, P. (2025). An interpretable deep geometric learning model to predict the effects of mutations on protein–protein interactions using large-scale protein language model. **Journal of Cheminformatics**, 17(1), 35.
- Hadipour, H., Li, Y. Y., **Sun, Y.**, Deng, C., Lac, L., Davis, R., ... & Hu, P. (2025). GraphBAN: An inductive graph-based approach for enhanced prediction of compound-protein interactions. **Nature Communications**, 16(1), 2541.
- Chen, L., Huang, Z. H., **Sun, Y.**, Domaratzki, M., Liu, Q., & Hu, P. (2024). Conditional probabilistic diffusion model driven synthetic radiogenomic applications in breast cancer. **PLOS Computational Biology**, 20(10), e1012490.
- Huang, Z.H., Chen, L., **Sun, Y.**, Liu, Q., Hu, P., Conditional generative adversarial network driven radiomic prediction of mutation status based on magnetic resonance imaging of breast cancer. **J Transl Med** 22, 226 (2024). https://doi.org/10.1186/s12967-024-05018-9
- Liu, C., **Sun, Y.**, Davis, R. et al. ABT-MPNN: an atom-bond transformer-based message-passing neural network for molecular property prediction. **J Cheminform** 15, 29 (2023). https://doi.org/10.1186/s13321-023-00698-9

Manuscript accepted

- **Sun, Y.**, Lu, Y., Li, Y., Jing, Z., Leung, C., Hu, P., Mol-xLSTM: A Graph-Based xLSTM Framework for Enhanced Molecular Property Prediction, accepted by **Communications Chemistry**, 2025
- Jing, Z., **Sun, Y.**, Li, Y. Y., Janarthanan, S., Deng, A., & Hu, P. (2025). Structure-Aware Fusion with Progressive Injection for Multimodal Molecular Representation Learning. Proceedings of the Thirty-Ninth Annual Conference on Neural Information Processing Systems (**NeurIPS 2025**), San Diego, USA.

Manuscripts under submission

- **Sun, Y.**, Shi, Y., Deng, A, Jing, Z., Leung, C. & Hu, P., DistRouting: Distribution-Aware Expert Routing Guided by Physicochemical Descriptors for Imbalanced Molecular Property Regression

Patents

- **Yan Sun**, Xueli Wang, Mingyang Yang, Bingjie Zhao, Tissue boundary determination apparatus and method. Patent Publication No. 20220383491.
- Bingjie Zhao, Xueli Wang, **Yan Sun**, A system and a method for predicting truncation image, a method for preparing data and medium thereof.
- Dejun Wang, Yaan Ge, **Yan Sun**, Buer Qi, Medical imaging method and system. Patent No. 11398012.