

YanJun Li

1345 CENTER DR MSB P6-33 Gainesville, FL, 32610

+1 (352) 273-9957

✉ yanjun.li@ufl.edu

🌐 www.yanjun-li.com

RESEARCH INTERESTS

AI-driven Drug Discovery, Machine Learning, Deep Learning and their applications.

Molecule Design and Optimization: Deep Generative Modeling for Small Molecule, RNA, and Protein Design.

Molecular Recognition: Geometric Deep Learning for Binding Affinity and Structure Prediction.

Drug Discovery & Computational Biology: Drug Repurposing and Synergistic Combinations, Genetic Variant Effect Prediction, Single-Cell RNA-sequencing, Spatial Transcriptomics, Image-based Profiling Analysis, etc.

EDUCATION

Doctor of Philosophy in Computer Science

Aug. 2016 – May. 2021

University of Florida, Gainesville, FL, USA

GPA: 4.0 / 4.0

Advisors: Dr. Dapeng Oliver Wu & Dr. Xiaolin Andy Li

Master of Engineering in Communication and Information System

Sept. 2012 – Jun. 2015

University of Electronic Science and Technology of China, Chengdu, China

Master of Engineering in Information, Production and Systems Engineering

Sept. 2013 – Mar. 2015

Waseda University, Kitakyushu, Japan

Bachelor of Engineering in Information Engineering

Sept. 2008 – Jul. 2012

University of Electronic Science and Technology of China, Chengdu, China

Ranking top 5%

PROFESSIONAL EXPERIENCE

Assistant Professor

Mar. 2023 – Present.

Department of Medicinal Chemistry, College of Pharmacy, University of Florida, Gainesville, FL

- Machine Learning, Deep Learning, and Drug Discovery
- AI-driven Drug Discovery and Optimization
- Geometric Deep Learning for Molecular Recognition
- Computational Biology

Machine Learning Scientist

Sept. 2022 – Mar. 2023

Calico Life Sciences, Alphabet Inc., South San Francisco, CA

- Develop novel deep learning algorithms for general peptide, antibody, and protein *de novo* design.
- Design peptide binder for disordered protein regions with geometric deep learning and diffusion models.

Senior Research Scientist

Jun. 2021 – Jul. 2022

Institute of Deep Learning, Baidu Research, Sunnyvale, CA

- Developed a structure complementary contrastive learning approach to predict protein-protein interaction, achieving state-of-the-art performance on the challenging across-species prediction with superior generalizability.
- Incorporating sequence-based cross attention mechanism, masked language modeling, and multi-task learning with protein complex structure-informed design, our model can also recognize the contact residues within the interacting proteins.

AI/ML Ph.D. Intern

Jun. 2020 – Aug. 2020

GE Healthcare, Remote at Gainesville, FL

- Developed deep learning based approaches to accurately register the X-ray image pairs with large sizes and limited overlaps. The proposed two-stages (from coarse-grained to fine-grained) registration and synthetic patch data augmented training significantly improved model performance compared with the benchmark.

SELECTED PUBLICATIONS

* indicates equal contribution; † indicates corresponding author ([Google Scholar](#))

- X. Pan, Y. Li, P. Huang, H. Staecker, M. He, Extracellular Vesicles for Developing Targeted Hearing Loss Therapy. *Journal of Controlled Release*. 2024.
- Y. Li[†], C. Sun, D. Romanova, D. Wu, L. Moroz, Analysis and Visualization of Single-Cell Sequencing Data with Scanpy and MetaCell: A Tutorial. *Methods in Molecular Biology, Springer*, 2024.
- Q. Tang, R. Ratnayake, G. Seabra, Z. Jiang, R. Fang, L. Cui, Y. Ding, T. Kahveci, J. Bian, C. Li, H. Luesch, Y. Li[†], Depicting Drug Discovery: A Survey on Deep Learning for Morphological Profiling, *arXiv*, 2024.
- Z. Wang, Z. Feng, Y. Li, B. Li, Y. Wang, M. He, X. Li, BatmanNet: Bi-branch Masked Graph Transformer Autoencoder for Molecular Representation. *Briefings in Bioinformatics*, 2023.
- X. Pan, P. Huang, S. Ali, T.E. Hutchinson, N. Erwin, Z.F. Greenberg, Z. Ding, Y. Li, N.E. Fernandez, H. Staecker, and M. He. CRISPR-Cas9 Extracellular Vesicles for Treating Hearing Loss. *bioRxiv*, 2023.
- C. Sun*, Y. Li*, S. Marini, D. Wu, M. Salemi, B. Magalis, Phylogenetic-informed graph deep learning to classify dynamic transmission clusters in infectious disease epidemics. *bioRxiv*, 2023.
- Y. Ren*, Y. Li*, T. Loftus*, J. Balch, K. Abbott, S. Datta, M. Ruppert, Z. Guan, B. Shickel, P. Rashidi, T. Baslanti, Identifying acute illness phenotypes via deep temporal interpolation and clustering network on physiologic signatures. *Scientific Reports*, 2023. (Under review)
- Y. Li, D. Zhou, G. Zheng, D. Wu, X. Li, Y. Yuan, DyScore: A Boosting Scoring Method with Dynamic Properties for Identifying True Binders and Non-binders in Structure-based Drug Discovery. *Journal of Chemical Information and Modeling*, 2022.
- Y. Ren, T. Loftus, Y. Li, Z. Guan, M. Ruppert, S. Datta, G. Upchurch, P. Tighe, P. Rashidi, B. Shickel, T. Baslanti, A. Bihorac, Physiologic signatures within six hours of hospitalization identify acute illness phenotypes. *Plos Digital Health*, 2022.
- B. Liu*, Y. Li*, L. Zhang, Analysis and visualization of spatial transcriptomic data. *Frontiers in Genetics*, 2021.
- M. Rezaei, Y. Li, D. Wu, X. Li, C. Li, Deep Learning in Drug Design: Protein-Ligand Binding Affinity Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020.
- S. Datta*, Y. Li*, M. Ruppert*, Y. Ren, B. Shickel, T. Baslanti, P. Rashidi, A. Bihorac, Reinforcement learning in surgery. *Surgery*, 2020.
- Y. Li, S. Yu, J. Principe, X. Li, D. Wu, PRI-VAE: Principle of Relevant Information Variational Autoencoders. *arXiv*, 2020.
- T. Loftus, A. Filiberto, Y. Li, J. Balch, P. Tighe, G. Upchurch, P. Rashidi, X. Li, A. Bihorac, Decision Analysis and Reinforcement Learning in Surgical Decision-Making. *Surgery*, 2020.
- Y. Li, M. Rezaei, C. Li, X. Li, DeepAtom: A Framework for Protein-Ligand Binding Affinity Prediction. *Intl. Conf. on Bioinformatics and Biomedicine (BIBM)*, 2019. (acceptance rate 18%)
- Y. Li, K. Heng, K. Ye, X. Li, FoldingZero: Protein Folding from Scratch in Hydrophobic-Polar Model. *Deep Reinforcement Learning Workshop (Oral) of NeurIPS*, 2018.(acceptance rate < 10%)
- L. Sundaram, H. Gao, S. Padigepati, J. McRae, Y. Li, et al. Predicting the clinical impact of human mutation with deep neural networks. *Nature Genetics*, 2018. (impact factor: 41.38)
- Q. Zhu, Y. Li, X. Li. Character Sequence-to-Sequence Model with Global Attention for Universal Morphological Reinflection. *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, 2017.

- X. Li, Y. Li, O. Yoshie. Story Generation Based on Reinforcement Learning with Action Selection Policy. *Intl. Conf. on Electrical Engineering*, 2014.
- Y. Li, X. Li, O. Yoshie. Traffic Engineering Framework with Machine Learning based Meta-layer in Software-defined Networks. *Intl. Conf. on Network Infrastructure and Digital Content (IC-NIDC)*, 2014.

RESEARCH EXPERIENCE

Research Assistant

Aug. 2016 – May. 2021

NSF Center for Big Learning, University of Florida, Gainesville, FL

DeepDynaTree: Geometric Deep Learning for Infectious Disease Forecasting

Dec. 2020 – Dec. 2021

- Worked as a deep learning research scientist on an **NIH-R01** project and supervised a Ph.D. student.
- Proposed a phylogenetic-informed graph deep learning system -DeepDynaTree- to jointly learn from phylogenetic tree statistics and structure to accurately classify dynamic transmission clusters in infectious disease epidemics.
- Proposed a novel graph neural network variant Primal-Dual Graph Long Short-Term Memory (PDGLSTM) model to learn the evolutionary information in a memory-driven fashion with state-of-the-art performance.

DeepDrug: Data-driven Structure-based Drug Discovery via Deep Learning

Jan. 2019 – May. 2021

- Lead a team of two graduate students to collaborate with researchers from UF College of Pharmacy.
- Propose a data-driven approach **DeepAtom** to predict the protein-ligand binding affinity. Taking three-dimensional structural data of the protein and ligand as input, DeepAtom achieves state-of-the-art scoring performance.
- Propose a general framework **DyScore** to identify the native binding pose among computer-generated decoys. As the first work to incorporate the static structure and dynamic properties of the compound, DyScore achieves state-of-the-art ranking performance.
- Collaborate with Mayo Clinic, utilized the DeepDrug approach to significantly increase the virtual screening absolute success rate by **7 times** in a practical *de novo* drug design task for a cancer treatment research.
- Design and develop a public web server to deploy the proposed machine learning model. Realized job submission, job tracking functions, etc.

DC-EHR: Deep Learning Clustering Approach for EHR Data

Jul. 2019 – Dec. 2020

- Worked as a deep learning research scientist in an **NIH-R01** project with clinicians of UF Health Shands Hospital.
- Developed deep sequence models to learn from the irregular and sparse time series EHR data. Proposed a multi-task learning framework to cluster patients using vital sign data measured within the early stages of hospital admissions. Discovered multiple phenotypes with unique pathophysiological signatures and clinical outcomes.

PRI-VAE: Principle of Relevant Information Variational Autoencoders

Aug. 2019 – Jan. 2020

- Investigated the learning dynamics of the existing VAE models from an information-theoretic perspective.
- Proposed a novel model, termed principle of relevant information VAE (PRI-VAE), to learn disentangled representations by explicitly regularizing the degrees of compactness and disentanglement of the latent representation.

FoldingZero: Protein Folding from Scratch via Deep Reinforcement Learning

Feb. 2018 – Dec. 2018

- Initiated and led a team of three graduate students to tackle the centennial open challenges of protein folding.
- Designed a novel framework -FoldingZero- to self-fold *de novo* protein two-dimensional structure using a reinforcement learning approach based on a deep neural network and the Monte Carlo tree search.

- Training from scratch, FoldingZero successfully learns the latent folding knowledge to stabilize the protein structure without any supervision and domain knowledge.

PrimateAI: Predicting the Clinical Impact of Human Mutation via Deep Learning *Aug. 2017 – Jul. 2018*

- Developed an end-to-end system PrimateAI, based on DeepFolding, to identify pathogenic gene mutations.
- PrimateAI achieves the state-of-the-art accuracy to identify the pathogenic mutations in rare disease patients and enables the discovery of 14 new candidate genes in intellectual disability at genome-wide significance.

DeepFolding: De Novo Protein Structure Prediction via Deep Learning *Aug. 2016 – Dec. 2017*

- Developed deep learning based approaches to accurately predict protein secondary structure (sequence-to-sequence prediction task) and contact map (sequence-to-two-dimensional prediction task) from the primary sequence and conservation data.

GRANTS

ExoTarget Platform as a programmable delivery system *Dec. 2023 – Dec. 2024*

NATL INST OF HLTH, Co-Investigator

TEACHING EXPERIENCE

Graduate Course, PHA6467C Drug Design II, University of Florida *2024 Spring*

Pharm.D. Course, PHA 6935: Introduction to Artificial Intelligence in Pharmacy, University of Florida 2023 Fall

Organizer, Knowledge Sharing Group, Baidu Research USA *Dec. 2021 – Jul. 2022*

Teaching Assistant, Summer Deep Learning Workshop, University of Florida *Jun. 2018 – Aug. 2018*

Graduate Course, Teaching Assistant, EEL 6935 Big Data Ecosystems, University of Florida Jan. 2018 – May. 2018

PROFESSIONAL SERVICES

Organizer or member of executive committee for conferences

- Session Chair for 2024 UF Drug Discovery Symposium, AI-driven Drug Discovery Session, Gainesville, FL, April 15-16, 2024.

Program Committee

- GenBio workshop at Conference on Neural Information Processing Systems (NeurIPS) 2024
- Association for the Advancement of Artificial Intelligence (AAAI) 2022, 2023, 2024
- SIAM International Conference on Data Mining (SDM) 2024
- ACM SIGKDD conference on Knowledge Discovery and Data Mining (KDD) 2023, 2024
- American Medical Informatics Association (AMIA) 2022 Informatics Submit
- World Congress on Computational Intelligence (WCCI) 2022

Editorship

- Review Editor, Frontiers in Artificial Intelligence
- Review Editor, Frontiers in Neuroscience
- Review Editor, Frontiers in Pharmacology

Paper Reviewer

- Briefings in Bioinformatics
- IEEE Transactions on Network Science & Engineering
- IEEE Transactions on Vehicular Technology
- PLOS Digital Health
- PLOS One

- Digital Discovery
- ChemistrySelect
- SN computer science
- Journal of Imaging Science and Technology

Research Advisor

Ph.D. Student Committee Member

- Zhuobiao Qiao, *Department of Electrical and Computer Engineering, UF* Mar. 2023 – Present.
- Samantha Ali, *Department of Pharmaceutics, UF* Mar. 2023 – Present.
- Meng-Lun Hsieh, *Department of Medicinal Chemistry, UF* Aug. 2023 – Present.

Research Assistant Advisor

- Riddhish Thakare, *Department of Computer and Information Science and Engineering, UF* Jul. 2023 – Present.
- Junge Gu, *Department of Computer and Information Science and Engineering, UF* Jul. 2023 – Dec. 2023
- Shiyu Jiang, *Department of Computer Science, JHU* Jun. 2023 – Present.
- Weihang You, *Department of Computer and Information Science and Engineering, UF* May. 2023 – Present.

SELECTED AWARDS & HONORS

NIH Targeted Genome Editor Delivery Challenge Phase I Winner Award, NIH	Dec. 2023
Student Travel Grant, IEEE Intl. Conf. on Bioinformatics and Biomedicine	Nov. 2019
Mentorship Excellence Award, Summer Deep Learning Workshop, University of Florida	Jun. 2018 – Aug. 2018
Honors Scholarship for International Students, Waseda University	Sept. 2013 – Sept. 2014
Scholarship for Advancement of Industry, Science and Technology, Kitakyushu, Japan	Sept. 2013 – Sept. 2014
Outstanding Graduates Awards, UESTC	Jun. 2012
Top-Class Scholarship with Honor, UESTC	Sept. 2008 – Jun. 2015

INVITED TALKS

- UF PHA5930 Seminar in Pharmacy Research on Nov. 30, 2023. Host: Dr. Lari Cavallari (University of Florida) & Dr. Mei He (University of Florida). Title: "Artificial Intelligence in drug discovery".
- BioFlorida Annual Conference on Oct. 12. Host: Dr. Patrick Gardella. Title: Acceleration of Generative AI/Machine Learning in the Life Sciences".
- UF SCRIPPS chemical biology & 5th UF drug discovery symposium on Apr. 20, 2023. Host: Dr. Thomas Kodadek (UF Scripps Biomedical Research) & Dr. Hendrik Luesch (University of Florida). Title: "Deep Learning for Drug Target Identification and Protein-protein Interaction".
- The virtual science seminar on "Advanced Topics in Data Mining" on Nov. 14, 2022. Host: Dr. Dongsheng Luo (Florida International University). Title: "Data Mining in Drug Discovery".
- The virtual conference on "Data Science for Genomics and Medicine" on Sept. 19, 2022. Host: Dr. Boxiang Liu (National University of Singapore). Title: "Artificial Intelligence driven Drug Discovery".
- IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM 2019), San Diego, CA, USA, Nov. 2019. Host: Dr. Xiong Liu (Eli Lilly and Company). Title: "DeepAtom: A Framework for Protein-Ligand Binding Affinity Prediction".
- Conference on Neural Information Processing Systems (NeurIPS) 2018, Deep Reinforcement Learning Workshop, Montréal Canada. Host: Dr. David Silver (DeepMind). Title: "FoldingZero: Protein Folding from Scratch in Hydrophobic-Polar Model."
- NSF Center for Big Learning Kickoff Meeting, Gainesville, FL, USA, May. 31 2018. Host: Dr. Xiaolin Andy Li (University of Florida). Title: "DeepFolding: End-to-End Prediction of Human Protein De Novo Structures".