Yu Li
E-mail: liyu@cse.cuhk.edu.hk
Homepage: liyu95.com
Phone: +86 18655062219

January 26, 2021

Assistant Professor
Department of Computer Science and Engineering
The Chinese University of Hong Kong

Employment

Dec 2020 – present, Assistant Professor, CSE, The Chinese University of Hong Kong (CUHK)

Education

King Abdullah University of Science & Technology (KAUST) — Jeddah, Saudi Arabia
Ph.D. in Computer Science — 2017 – 2020
–Advisor: Xin Gao
–Thesis: Towards Structured Prediction in Bioinformatics with Deep Learning

King Abdullah University of Science & Technology (KAUST) — Jeddah, Saudi Arabia
M.S. in Computer Science — 2015 – 2016

University of Science & Technology of China (USTC) — Hefei, China
B.S. (Honored) in Biosciences, BEI Shizhang Elite Class, GPA Top 2% — 2011 – 2015

Research Interests

Deep learning, Healthcare, Bioinformatics, Machine learning

- Machine learning for health care
- Biomolecular function prediction with deep learning
- Biological sequence analysis and Nanopore sequencing
- Biomolecular structure reconstruction and prediction
- Structured learning

Publications

Journal:(*Equal contribution, †Co-corresponding)

32. **Y Li***, Z Xu*, W Han*, H Cao*, R Umarov, A Yan, M Fan, H Chen, L Li, P Ho, X Gao. "HMD-ARG: hierarchical multi-task deep learning for annotating antibiotic resistance genes". *Microbiome*. (2020). *accepted*.

31. **Y Li***, S Wang*, C Bi, Z Qiu, M Li, X Gao. "DeepSimulator1.5: a more powerful, quicker and lighter simulator for Nanopore sequencing". *Bioinformatics*. (2020). 10.1093/bioinformatics/btz963.

30. T Zhang, **Y Li**, Y Li, S Sun, and X Gao. "A Self-adaptive Deep Learning Algorithm for Accelerating Multi-component Flash Calculation". *Computer Methods in Applied Mechanics and Engineering*. (2020). 10.1016/j.cma.2020.113207.

29. H Li*, S Tian*, **Y Li***, R Tan, Y Pan, C Huang, Y Xu, and X Gao. "Modern Deep Learning in Bioinformatics". *Journal of Molecular Cell Biology.* (2020). 10.1093/jmcb/mjaa030.

28. C Bi, L Wang, B Yuan, X Zhou, **Y Li**, S Wang, Y Pang, X Gao, Y Huang, M Li. "Long-read Individual-molecule Sequencing Reveals CRISPR-induced Genetic Heterogeneity in Human ESCs". *Genome Biology.* (2020). 10.1186/s13059-020-02143-8.

27. Z Li, Y Li, B Zhang, **Y Li**, Y Long, X Zou, M Zhang, Y Hu, W Chen†, X Gao†. "DeeReCT-APA: prediction of alternative polyadenylation site usage through deep learning". *Genomics, Proteomics & Bioinformatics (GPB).* (2020). accepted.

26. J Lam*, **Y Li***, L Zhu†, R Umarov, H Jiang, A Heliou, F Sheong, T Liu, Y Long, Y Li, L Fang, R Altman, W Chen†, X Huang†, X Gao†. "A deep learning framework to predict binding preference of RNA constituents on protein surface". *Nature Communications.* (2019). s41467-019-12920-0.

25. G Jia, **Y Li**, H Zhang, I Chattopadhyay, A Jensen, D Blair, L Davis, P Robinson, T Dahln, S Brunak, M Benson, G Edgren, N Cox, X Gao, A Rzhetsky. "Estimating heritability and genetic correlations from large health datasets in the absence of genetic data". *Nature Communications.* (2019). s41467-019-13455-0.

24. J Lei, G Sheng, P Cheung, S Wang, **Y Li**, X Gao, Y Zhang, Y Wang, X Huang. "Two symmetric Arginine residues play distinct roles in Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage". *Proceedings of the National Academy of Sciences of the United States of America (PNAS).* (2019). 10.1073/pnas.1817041116.

23. **Y Li**, T Zhang, S Sun†, X Gao†. "Accelerating Flash Calculation through Deep Learning Methods". *Journal of Computational Physics (JCP).* (2019). 10.1016/j.jcp.2019.05.028.

22. **Y Li**, C Huang, L Ding, Z Li, Y Pan, X Gao. "Deep learning in bioinformatics: introduction, application, and perspective in big data era". *Methods.* (2019). 10.1016/j.ymeth.2019.04.008. **Cover article of the Methods issue: Deep Learning in Bioinformatics**.

21. Z Zou, S Tian, X Gao, **Y Li**. "mlDEEPre: Multi-functional enzyme function prediction with hierarchical multi-label deep learning". *Frontiers in Genetics.* (2019) 10.3389/fgene.2018.00714.

20. R Umarov, H Kuwahara, **Y Li**, X Gao†, V Solovyev†. "Promoter analysis and prediction in the human genome using sequence-based deep learning models". *Bioinformatics.* (2019). 10.1093/bioinformatics/bty1068.

19. U Hameed, C Liao, A Radhakrishnan, F Huser, S Aljedani, X Zhao, A Momin, F Melo, X Guo, C Brooks, **Y Li**, X Cui, X Gao, J Ladury, L Jaremko, M Jaremko, J Li, S, Arold. "H-NS uses an autoinhibitory conformational switch to achieve environment-controlled gene silencing". *Nucleic Acids Research (NAR).* (2018). 10.1093/nar/gky1299.

18. Zhihao Xia, **Y Li**, B Zhang, Z Li, Y Hu, W Chen†, X Gao†. "DeeReCT-PolyA: a robust and generic deep learning method for PAS identification". *Bioinformatics.* (2018). 10.1093/bioinformatics/bty991.

17. **Y Li**, R Han, C Bi, M Li, S Wang†, X Gao†. "DeepSimulator: a deep simulator for nanopore sequencing". *Bioinformatics.* (2018). 10.1093/bioinformatics/bty223.

16. V Kordopati, A Salhi, R Razali, A Radovanovic, F Tifratene, M Uludag, **Y Li**, A Bokhari, A AlSaieedi, A Raies, C Neste, M Essack, V Bajic. "DES-Mutation: System for Exploring Links of Mutations and Diseases". *Scientific Reports.* (2018). 10.1038/s41598-018-31439-w.

15. **Y Li***, F Xu*, F Zhang, P Xu, M Fan, L Li, X Gao†, R Han†. "DLBI: Deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy". *Bioinformatics*. (2018). 10.1093/bioinformatics/bty241.

14. S Wang, S Fei, Z Wang, **Y Li**, J Xu, F Zhao, X Gao. "PredMP:a web server for de novo prediction of membrane protein". *Bioinformatics*. (2018). 10.1093/bioinformatics/bty684.

13. R Han, X Wan, L Li, A Lawrence, P Yang, **Y Li**, S Wang, F Sun, Z Liu, X Gao, F Zhang. "AuTom-dualx: a toolkit for fully automatic alignment of dual-axis tilt series with simultaneous reconstruction". *Bioinformatics*. (2018). 10.1093/bioinformatics/bty620.

12. R Han, **Y Li**, X Gao, S Wang. "An accurate and rapid continuous wavelet dynamic time warping algorithm for end-to-end mapping in ultra-long nanopore sequencing". *Bioinformatics*. (2018). 10.1093/bioinformatics/bty555.

11. **Y Li**, S Wang, R Umarov, B Xie, M Fan, L Li, X Gao. "DEEPre: sequence-based enzyme EC number prediction by deep learning". *Bioinformatics*. (2017). 10.1093/bioinformatics/btx680.

10. H Dai, R Umarov, H Kuwahara, **Y Li**, L Song†, X Gao†. "Sequence2Vec: a novel embedding approach for modeling transcription factor binding affinity landscape". *Bioinformatics*. (2017). 10.1093/bioinformatics/btx480.

9. S Wu, D Wang, J Liu, Y Feng, J Weng, **Y Li**, X Gao, J Liu, W Wang. "The dynamic multisite interactions between two intrinsically disordered proteins". *Angewandte Chemie*. (2017). 10.1002/anie.201701883.

8. X Li, Q Tao, Y Fang, C Cheng, Y Hao, J Qi, **Y Li**, W Zhang, Y Wang, X Zhang. "Reward sensitivity predicts ice cream-related attentional bias assessed by inattentional blindness". *Appetite*. (2015). 10.1016/j.appet.2015.02.010.

Conference:(*Equal contribution, †Co-corresponding)

7. X Chen*, **Y Li***, R Umarov, X Gao, L Song. "RNA Secondary Structure Prediction By Learning Unrolled Algorithms". *Eighth International Conference on Learning Representations (ICLR-20),* ***Oral***.

6. X Chen, H Dai, **Y Li**, X Gao, and L Song. "Learning to Stop While Learning to Predict." *Thirty-seventh International Conference on Machine Learning (ICML-20)*.

5. L Ding, M Yu, L Liu, F Zhu, Y Liu, **Y Li**, L Shao. "Two Generator Game: Learning to Sample via Linear Goodness-of-Fit Test." *Thirty-third Conference on Neural Information Processing Systems (NeurIPS-19)*.

4. L Ding, Z Liu, **Y Li**, S Liao, Y Liu, P Yang, G Yu, L Shao, X Gao. "Linear Kernel Tests via Empirical Likelihood for High Dimensional Data". *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*.

3. L Ding, S Liao, Y Liu, **Y Li**, P Yang, Y Pan, C Huang, L Shao, X Gao. "Approximate Kernel Selection with Strong Approximate Consistency". *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*.

2. **Y Li***, F Xu*, F Zhang, P Xu, M Fan, L Li, X Gao†, R Han†. "DLBI: Deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy". *The Twenty-Sixth Conference on Intelligent Systems for Molecular Biology (ISMB-18)*.

1. R Han, **Y Li**, X Gao, S Wang. "An accurate and rapid continuous wavelet dynamic time warping algorithm for end-to-end mapping in ultra-long nanopore sequencing". *The Seventeenth European Conference on Computational Biology (ECCB-18).*

Under review & Preprint(*Equal contribution, †Co-corresponding)

8. K Wang, R Stevens, H Alachram, **Y Li**, L Soldatova, R King, S Ananiadou, M Li, F Christopoulou, J Ambite, S Garg, U Hermjakob, D Marcu, T Beibbarth, E Wingender, E Sheng, X Gao, B Khomtchouk, A Galstyan, B Chambers, J Evans, A Rzhetsky. "NERO: A Biomedical Named-entity (Recognition) Ontology with a Large, Annotated Corpus Reveals Meaningful Associations Through Text Embedding". *PLOS Computational Biology, under review*

7. G Jia*, **Y Li***, X Zhong, K Wang, M Pividori, R Alomairy, A Esposito, H Ltaief, C Terao, M Akiyama , K Matsuda, D Keyes, H Im, T Gojobori, Y Kamatani, M Kubo, N Cox, X Gao, A Rzhetsky. "Genome-wide Association Mapping of Geometric Space of Human Disease". *Science Advances, under review.*

6. S Chen*, **Y Li***, T Zhang, X Zhu, S Sun, X Gao. "Lunar Features Detection for Energy Discovery via Deep Learning". *Applied Energy, under review.*

5. H Dai, X Chen, **Y Li**, X Gao, L Song. "A Framework For Differentiable Discovery Of Graph Algorithms". *ICML-21, under review*

4. **Y Li**, Z Li, C Huang, L Ding, Y Pan, X Gao. "SupportNet: solving catastrophic forgetting in class incremental learning with support data". *arxiv.org/abs/1806.02942.*

3. **Y Li**, L Ding, X Gao. "On the decision boundary of deep neural network". *arxiv.org/abs/1808.05385.*

2. R Umarov, **Y Li**, H Kuwahara, C Van Neste, X Gao. "NNfold: RNA secondary structure prediction by deep learning with an architecture imposing contextual constraining". *under preparation.*

1. **Y Li**, H Kuwahara, P Yang, L Song†, X Gao†. "PGCN: Disease gene prioritization by disease and gene embedding through graph convolutional neural networks". *bioRxiv 532226*, 2019.

Experience

Le Song's group, Georgia Tech	Atlanta, USA
Visiting student (Advisor: Le Song)	Fall, 2018
LAMDA, Nanjing University	Nanjing, China
Visiting student (Advisor: Zhi-Hua Zhou)	Summer, 2018
Huang Research Group, HKUST	Hong Kong, China
Visiting student (Advisor: Xuhui Huang)	Summer, 2017
Cognitive Psychology Lab, USTC	Hefei, China
Research assistant (Advisor: Xiaochu Zhang)	2014 - 2015
New Oriental School	Hefei, China
Intern	Winter, 2014
Peking University	Beijing, China
Summer school	Summer, 2014
SIBS & IBP, CAS	Shanghai & Beijing, China
Intern	Summer, 2012

Patents

7. HMD-ARG: Hierarchical Multi-task Deep learning for annotating Antibiotic Resistance Genes (under application)
6. PGCN: Disease gene prioritization by disease and gene embedding through graph convolutional neural networks (under application)
5. Continuous Wavelet-Based Dynamic Time Warping Method and System (under application)
4. Incremental learning method through deep learning and support data, WO2019193462A1
3. Deep-learning based structure reconstruction method and apparatus, WO2019145767A1
2. Deepsimulator method and system for mimicking nanopore sequencing, WO2019116119A1
1. System, apparatus, and method for sequence-based enzyme ec number prediction by deep learning, WO2019077494A1

Professional Activities

Memberships:

- Association for the Advancement of Artificial Intelligence(AAAI) Member
- International Society for Computational Biology (ISCB) Member

Talks:

12. December 2020, Towards Understanding Biomolecular Structure and Function with Deep Learning, CUHK(SZ), Shen Zhen, China
11. May 2020, Towards Understanding Biomolecular Structure and Function with Deep Learning, CityU Hong Kong, Hong Kong, China
10. May 2020, Towards Understanding Biomolecular Structure and Function with Deep Learning, CUHK, Hong Kong, China
9. April 2020, Towards Understanding Biomolecular Structure and Function with Deep Learning, Georgia Tech, Atlanta, USA
8. April 2020, Towards Understanding Biomolecular Structure and Function with Deep Learning, TTIC, Chicago, USA
7. March 2020, Towards Structured Prediction in Bioinformatics with Deep Learning, Indiana University Bloomington, Bloomington, USA
6. October 2019, Towards understanding protein functions through deep learning, KAUST Bioengineering Graduate Seminar, Thuwal, Saudi Arabia
5. October 2019, Towards understanding protein functions through deep learning, SUSTech, Shenzhen, China
4. March 2019, Tutorial on Machine Learning and Hand-on Coding Training, CTPL Workshop on Machine Learning, Thuwal, Saudi Arabia

3. July 2018, DLBI: Deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy, The Twenty-Sixth Conference on Intelligent Systems for Molecular Biology (ISMB-18), Chicago, USA

2. June 2018, Deep learning in bioinformatics, Nanjing University, Nanjing, China

1. December 2017, DEEPre: sequence-based enzyme EC number prediction by deep learning, KAUST Research Conference on Big Data Analyses in Evolutionary Biology, Thuwal, Saudi Arabia

Reviewer:

- Nucleic Acids Research
- Bioinformatics
- PLOS Computational Biology
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- International Conference on Research in Computational Molecular (RECOMB-20)
- International Conference on Learning Representations (ICLR-20)
- AAAI Conference on Artificial Intelligence (AAAI-18, 19, 20)
- International Joint Conferences on Artificial Intelligence (IJCAI-18)
- IEEE Transactions on Neural Networks and Learning Systems (IEEE TNNLS)
- European Conference on Computer Vision (ECCV-2020)
- Genomics, Proteomics & Bioinformatics (GPB)
- Scientific Reports
- BMC Genomics
- BMC Bioinformatics
- Frontiers in Genetics
- Fuel
- IEEE International Conference on Bioinformatics and Biomedicine (BIBM-17)
- IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)
- Journal of Bioinformatics and Computational Biology (JBCB)
- Journal of Biomedical Informatics (JBI)
- PLOS ONE
- Proteins: Structure, Function, and Bioinformatics

Supervised Students

In KAUST:
Master students:

- Siyuan Chen
- Wenkai Han
- Rawan Albakri

Visiting students:

- Siyuan Chen (2019, from UESTC)
- Noura AlRasheed (Summer, 2019, KGSP student from UCSD)
- Longxi Zhou (Summer, 2018, from USTC)
- Ammar Alqatari (Summer, 2018, KGSP student from Stanford)
- Zhongxiao Li (Spring, 2018, from SUSTech)
- Zhenzhen Zou (2018, from CAS)

Teaching

In KAUST:
TA for CS320 (Probabilistic Graphical Models) in Spring 2019, Spring 2020.
TA for CS229 (Machine Learning) in Spring 2018.
TA for CS220 (Data Analytics) in Fall 2016, Spring 2017, Fall 2017, Fall 2019.

Awards & Honours

9. Commencement speaker candidate, the first Chinese candidate in the KAUST history, 2020, KAUST
8. Full scholarship for MS/PhD study (acceptance rate: 3%), 2015, KAUST
7. Certificate of Honor Rank (5% out of all graduates in 2015 at USTC), 2015, USTC
6. TianYi FeiYoung Scholarship, 2014, USTC
5. Shizhang Bei's Fellowship, 2013-2015, USTC
4. National Scholarship (2 out of 73 students), 2013, USTC
3. Xingye Scholarship of Responsibility (2 out of 90 students), 2012, USTC
2. Outstanding Student Scholarship, Golden Prize (5% chance of receiving), 2012, USTC
1. Cyrus Tangs Moral Education Scholarship (2 out of 90 students), 2011-2015, USTC