

Jie Hou

Assistant Professor
 Department of Computer Science
 Program of Bioinformatics and Computational Biology
 School of Science & Engineering
 Saint Louis University
 St. Louis, MO 63103-2007
 Phone: 573-864-2105
 Email: jhou4@slu.edu
 Website: <http://cs.slu.edu/~hou>

I. Basic Information

Education

Ph.D. in Computer Science, University of Missouri-Columbia, Columbia, MO, USA July 2019
 Dissertation title: *Improving protein structure prediction by deep learning and computational optimization*
 Ph.D. advisor: Prof. Jianlin Cheng
M.A. in Statistics, University of Missouri-Columbia, Columbia, MO, USA May 2014
B.S. in Mathematics, Shanghai Maritime University, Shanghai, China July 2012

Experience

Assistant Professor, Department of Computer Science August 2019–Present
 Affiliated Faculty, Program of Bioinformatics and Computational Biology
 Saint Louis University, St. Louis, MO, USA
Assistant Professor, Department of Health and Clinical Outcomes Research (HCOR) May 2024–Present
 School of Medicine (Secondary Appointment)
 Saint Louis University, St. Louis, MO, USA
Research Assistant, Department of Computer Science August 2014–July 2019
 University of Missouri, Columbia, MO, USA

Honors and Awards

- 2024, Received *School of Science & Engineering Faculty and Staff Excellence in Teaching Award*, Saint Louis University
- 2019, Received *Outstanding PhD Student Award in the College of Engineering*, University of Missouri

II. Research

Research Interests

General Areas: Machine learning, data mining, deep learning, bioinformatics

Specific Areas: Protein/RNA structure prediction, protein quality assessment, next-generation sequencing data analysis

Funding

Expected Grants

- **NIH/NIGMS, R15 (PI, \$467,334)**, “*Improving Artificial Intelligence Readiness of RNA Motif Data for Structure Analysis and Modeling*,” 08/2024 – 07/2027. **Impact score: 24, Recommended for**

Funding by NIH/NIGMS Program on 05/16/2024

Status: Administrative review pending

Awarded Grants

- **President's Research Funds (PRF), Saint Louis University (co-PI, \$25,000)**, "*Privacy Analysis of Distributed Learning with Defensive Measures*", 05/2022–04/2023, PI - Dr. Reza Tourani
- **WUSTL Institute of Clinical and Translational Sciences (co-PI, \$15,000)**, "*Pharmacogenetic Refinement of the Warfarin Dose Using Machine Learning*", 03/2022–02/2023, PI - Dr. Brian Gage
- **President's Research Funds (PRF), Saint Louis University (Sole PI, \$15,000)**, "*Advancing Protein Structure Prediction using Small-angle X-ray Scattering data*", 05/2021–11/2022.

Pending Grants

- **NIH/NHLBI, R01 (co-PI, \$1,554,645)**, "*Pharmacogenetic Refinement of the Warfarin Dose Using Machine Learning*," 04/2024 – 03/2028, PI - Gage Brian. **A0 Percentile: 13%; A1 Percentile: 9%** (within the NHLBI grant payline in 2023 at 14%)
Status: Council review pending
- **NSF (co-PI, \$1,200,000)**, "*Collaborative Research: SaTC: Medium: Fostering Explainability of Privacy Attacks Towards Robust Privacy-Preserving Collaborative Learning*," 08/2024–07/2028, PI – Reza Tourani
Status: Scientific review pending
- **NIH/SBIR (co-PI, \$94,207)**, "*Developing an AI Powered Harness for a Medical Device to Improve Independence in Patients in Rehabilitation Hospitals*," 11/2024 – 10/2025, PI - Sheila Buswell (Buswell Biomedical)
Status: Scientific review pending
- **NSF (co-PI, \$700,000)**, "*CC* Compute-Campus: Modernizing Campus Cyberinfrastructure for AI-Enhanced Research and Education (ModernCARE)*," 01/2025 – 12/2026, PI - Dr. Ted Ahn
Status: Scientific review pending

Publications

Statistics of Publication Impact

- **50 Journal Articles Published in 22 International Journals**
- **Total Journal Impact Factors: +175**
- **+2760 Citations** (estimation according to Google Scholar as of June, 2024: https://scholar.google.com/citations?user=Rsc9b_wAAAAJ&hl=en&oi=ao)
- **H-Index: 24** (24 publications each having a citation number ≥ 24)
- **H-Index (since 2019): 23** (23 publications each having a citation number ≥ 23 since 2019)
- **i10-Index: 42** (42 publications each having ≥ 10 citations)
- **i10-Index (since 2019): 41** (41 publications each having ≥ 10 citations since 2019)
- **5 Conference Papers, 13 Conference Abstracts/Posters**
- **2 Book Chapters**

(Underscore - Hou as first-author/co-author, * - Hou as corresponding author)

Refereed Journals

1. Chen, J., Zia, A., Luo, A., Meng, H., Wang, F., Hou, J., Cao, R., and Si, D. (2024). Enhancing cryo-em structure prediction with deeptreader and alphafold2 integration. *Briefings in Bioinformatics*, 25(3):bbae118. - Impact factor (2024): 9.5
2. Si, T., Hopkins, Z., Yanev, J., Hou, J., and Gong, H. (2023). A novel f-divergence based generative adversarial imputation method for scRNA-seq data analysis. *Plos one*, 18(11):e0292792. - Impact factor (2023): 3.7
3. Zhang, L., Wang, S., Hou, J., Si, D., Zhu, J., and Cao, R. (2023). Complexqa: a deep graph learning approach for protein complex structure assessment. *Briefings in Bioinformatics*, 24(6):bbad287. - Impact factor (2023): 9.5; Citations (in Google Scholar): 2
4. Yang, H., Shi, X., Chen, C., Hou, J., Ji, T., Cheng, J., and Birchler, J. A. (2023). Genomic imbalance modulates transposable element expression in maize. *Plant Communications*, 4(2). - Impact factor (2023): 10.5; Citations (in Google Scholar): 2
5. Nakamura, A., Meng, H., Zhao, M., Wang, F., Hou, J., Cao, R., and Si, D. (2023). Fast and automated protein-dna/rna macromolecular complex modeling from cryo-em maps. *Briefings in Bioinformatics*, 24(2):bbac632. - Impact factor (2023): 9.5; Citations (in Google Scholar): 5
6. Shi, X., Yang, H., Chen, C., Hou, J., Ji, T., Cheng, J., and Birchler, J. A. (2022a). Dosage-sensitive miRNAs trigger modulation of gene expression during genomic imbalance in maize. *Nature Communications*, 13(1):3014. - Impact factor (2022): 16.6; Citations (in Google Scholar): 2
7. Hippe, K., Lilley, C., William Berkenpas, J., Chandana Pocha, C., Kishaba, K., Ding, H., Hou, J., Si, D., and Cao, R. (2022). Zoomqa: residue-level protein model accuracy estimation with machine learning on sequential and 3d structural features. *Briefings in bioinformatics*, 23(1):bbab384. - Impact factor (2022): 9.5; Citations (in Google Scholar): 7
8. Shi, X., Yang, H., Chen, C., Hou, J., Ji, T., Cheng, J., and Birchler, J. A. (2022b). Effect of aneuploidy of a non-essential chromosome on gene expression in maize. *The Plant Journal*, 110(1):193–211. - Impact factor (2021): 7.09; Citations (in Google Scholar): 10
9. Liu, J., Wu, T., Guo, Z., Hou, J., and Cheng, J. (2022). Improving protein tertiary structure prediction by deep learning and distance prediction in casp14. *Proteins: Structure, Function, and Bioinformatics*, 90(1):58–72. - Impact factor (2021): 4.08; Citations (in Google Scholar): 21
10. Guo, Z., Wu, T., Liu, J., Hou, J., and Cheng, J. (2021b). Improving deep learning-based protein distance prediction in casp14. *Bioinformatics*, 37(19):3190–3196. - Impact factor (2021): 6.93; Citations (in Google Scholar): 9
11. Chen, C., Hou, J., Shi, X., Yang, H., Birchler, J. A., and Cheng, J. (2021a). Deepgrn: prediction of transcription factor binding site across cell-types using attention-based deep neural networks. *BMC bioinformatics*, 22(1):1–18. - Impact factor (2021): 3.30; Citations (in Google Scholar): 57
12. Blavet, N., Yang, H., Su, H., Solanský, P., Douglas, R. N., Karafiátová, M., Šimková, L., Zhang, J., Liu, Y., Hou, J., et al. (2021). Sequence of the supernumerary B chromosome of maize provides insight into its drive mechanism and evolution. *Proceedings of the National Academy of Sciences*, 118(23). - Impact factor (2021): 12.77; Citations (in Google Scholar): 36

13. Si, D., Nakamura, A., Tang, R., Guan, H., Hou, J., Firozi, A., Cao, R., Hippe, K., and Zhao, M. (2021). Artificial intelligence advances for de novo molecular structure modeling in cryo-electron microscopy. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, page e1542. - Impact factor (2021): 11.50; Citations (in Google Scholar): 15
14. Chen, X., Liu, J., Guo, Z., Wu, T., Hou, J., and Cheng, J. (2021b). Protein model accuracy estimation empowered by deep learning and inter-residue distance prediction in casp14. *Scientific Reports*, 11(1):1–12. - Impact factor (2021): 4.99; Citations (in Google Scholar): 18
15. Guo, Z., Hou, J., and Cheng, J. (2021a). Dnss2: Improved ab initio protein secondary structure prediction using advanced deep learning architectures. *Proteins: Structure, Function, and Bioinformatics*, 89(2):207–217. - Impact factor (2021): 4.08; Citations (in Google Scholar): 44
16. Shi, X., Yang, H., Chen, C., Hou, J., Hanson, K. M., Albert, P. S., Ji, T., Cheng, J., and Birchler, J. A. (2021). Genomic imbalance determines positive and negative modulation of gene expression in diploid maize. *The Plant Cell*, 33(4):917–939. - Impact factor (2021): 12.08; Citations (in Google Scholar): 24
17. Yang, H., Shi, X., Chen, C., Hou, J., Ji, T., Cheng, J., and Birchler, J. A. (2021). Predominantly inverse modulation of gene expression in genomically unbalanced disomic haploid maize. *The Plant Cell*, 33(4):901–916. - Impact factor (2021): 12.08; Citations (in Google Scholar): 22
18. Adhikari, B., Shrestha, B., Bernardini, M., Hou, J., and Lea, J. (2021). Disteval: A web server for evaluating predicted protein distances. *BMC bioinformatics*, 22(1):1–9. - Impact factor (2021): 3.30; Citations (in Google Scholar): 9
19. Wu, T., Guo, Z., Hou, J., and Cheng, J. (2021). Deepdist: real-value inter-residue distance prediction with deep residual convolutional network. *BMC bioinformatics*, 22:1–17. - Impact factor (2021): 3.307; Citations (in Google Scholar): 63
20. Lawson, C. L., Kryshchuk, A., Adams, P. D., Afonine, P. V., Baker, M. L., Barad, B. A., Bond, P., Burnley, T., Cao, R., Cheng, J., et al. (2021). Cryo-em model validation recommendations based on outcomes of the 2019 emdataresource challenge. *Nature methods*, 18(2):156–164. - Impact factor (2020): 28.547; Citations (in Google Scholar): 74
21. Grunz-Borgmann, E. A., Nicholas, L. A., Spagnoli, S., Trzeciakowski, J. P., Valliyodan, B., Hou, J., Jilong, L., Kerley, M., Fritsche, K., Parrish, A. R., et al. (2020). The renoprotective effects of soy protein in the aging kidney. *Medical Research Archives*, 8(3). - Impact factor (2020): 2.235; Citations (in Google Scholar): 5
22. Chen, C., Hou, J., Shi, X., Yang, H., Birchler, J. A., and Cheng, J. (2020a). Gnet2: an r package for constructing gene regulatory networks from transcriptomic data. *Bioinformatics*. - Impact factor (2020): 6.937; Citations (in Google Scholar): 5
23. Chen, C., Hou, J., Tanner, J. J., and Cheng, J. (2020b). Bioinformatics methods for mass spectrometry-based proteomics data analysis. *International Journal of Molecular Sciences*, 21(8):2873. - Impact factor (2020): 5.923; Citations (in Google Scholar): 202
24. Jasmer, K. J., Hou, J., Mannino, P., Cheng, J., and Hannink, M. (2020). Heme oxygenase promotes b-raf-dependent melanosome formation. *Pigment Cell & Melanoma Research*, 33(6):850–868. - Impact factor (2020): 4.693; Citations (in Google Scholar): 10

25. Si, D., Moritz, S. A., Pfab, J., Hou, J., Cao, R., Wang, L., Wu, T., and Cheng, J. (2020). Deep learning to predict protein backbone structure from high-resolution cryo-em density maps. *Scientific Reports (Nature Publisher Group)*, 10(1). - Impact factor (2020): 4.379; Citations (in Google Scholar): 91
26. Johnson, A. F., Hou, J., Yang, H., Shi, X., Chen, C., Islam, M. S., Ji, T., Cheng, J., and Birchler, J. A. (2020). Magnitude of modulation of gene expression in aneuploid maize depends on the extent of genomic imbalance. *Journal of Genetics and Genomics*. - Impact factor (2020): 4.275; Citations (in Google Scholar): 18
27. Wu, T., Hou, J., Adhikari, B., and Cheng, J. (2020). Analysis of several key factors influencing deep learning-based inter-residue contact prediction. *Bioinformatics*, 36(4):1091–1098. - Impact factor (2020): 6.937; Citations (in Google Scholar): 32
28. Hou, J., Adhikari, B., Tanner, J. J., and Cheng, J. (2020a). Saxsdom: Modeling multidomain protein structures using small-angle x-ray scattering data. *Proteins: Structure, Function, and Bioinformatics*, 88(6):775–787. - Impact factor (2020): 3.756; Citations (in Google Scholar): 10
29. Lensink, M. F., Brysbaert, G., Nadzirin, N., Velankar, S., Chaleil, R. A., Gerguri, T., Bates, P. A., Laine, E., Carbone, A., Grudin, S., et al. (2019). Blind prediction of homo-and hetero-protein complexes: The casp13-capri experiment. *Proteins: Structure, Function, and Bioinformatics*, 87(12):1200–1221 (conducted prior to joining SLU). - Impact factor (2019): 2.828; Citations (in Google Scholar): 117
30. Song, H., Zhou, H., Qu, Z., Hou, J., Chen, W., Cai, W., Cheng, Q., Chuang, D. Y., Chen, S., Li, S., et al. (2019). From analysis of ischemic mouse brain proteome to identification of human serum clusterin as a potential biomarker for severity of acute ischemic stroke. *Translational Stroke Research*, 10(5):546–556 (conducted prior to joining SLU). - Impact factor (2019): 5.780; Citations (in Google Scholar): 25
31. Cheng, J., Choe, M.-H., Elofsson, A., Han, K.-S., Hou, J., Maghrabi, A. H., McGuffin, L. J., Menéndez-Hurtado, D., Olechnovič, K., Schwede, T., et al. (2019). Estimation of model accuracy in casp13. *Proteins: Structure, Function, and Bioinformatics*, 87(12):1361–1377 (conducted prior to joining SLU). - Impact factor (2019): 2.828; Citations (in Google Scholar): 91
32. Zhou, N., Jiang, Y., Bergquist, T. R., Lee, A. J., Kacsoh, B. Z., Crocker, A. W., Lewis, K. A., Georghiou, G., Nguyen, H. N., Hamid, M. N., et al. (2019). The cafa challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. *Genome biology*, 20(1):1–23 (conducted prior to joining SLU). - Impact factor (2019): 10.806; Citations (in Google Scholar): 352
33. Hou, J., Wu, T., Cao, R., and Cheng, J. (2019). Protein tertiary structure modeling driven by deep learning and contact distance prediction in casp13. *Proteins: Structure, Function, and Bioinformatics*, 87(12):1165–1178 (conducted prior to joining SLU). - Impact factor (2019): 2.828; Citations (in Google Scholar): 159
34. Hou, J., Shi, X., Chen, C., Islam, M. S., Johnson, A. F., Kanno, T., Huettel, B., Yen, M.-R., Hsu, F.-M., Ji, T., et al. (2018b). Global impacts of chromosomal imbalance on gene expression in arabidopsis and other taxa. *Proceedings of the National Academy of Sciences*, 115(48):E11321–E11330 (conducted prior to joining SLU). - Impact factor (2019): 9.412; Citations (in Google Scholar): 58
35. Lei, W., Lu, Y., Hou, J., Chen, C., Browning, J. D., Lubahn, D. B., Cheng, J., Folk, W. R., Sun, G. Y., and Fritsche, K. L. (2018). Rna sequence analysis reveals expected and novel immuno-modulatory activities by sutherlandia frutescens (conducted prior to joining SLU)

36. Keasar, C., McGuffin, L. J., Wallner, B., Chopra, G., Adhikari, B., Bhattacharya, D., Blake, L., Bortot, L. O., Cao, R., Dhanasekaran, B., et al. (2018). An analysis and evaluation of the wefold collaborative for protein structure prediction and its pipelines in casp11 and casp12. *Scientific reports*, 8(1):1–18. - Impact factor (2018): 4.011; Citations (in Google Scholar): 23 (conducted prior to joining SLU)
37. Adhikari, B., Hou, J., and Cheng, J. (2018a). Dncon2: improved protein contact prediction using two-level deep convolutional neural networks. *Bioinformatics*, 34(9):1466–1472 (conducted prior to joining SLU). - Impact factor (2018): 4.51; Citations (in Google Scholar): 179
38. Adhikari, B., Hou, J., and Cheng, J. (2018b). Protein contact prediction by integrating deep multiple sequence alignments, coevolution and machine learning. *Proteins: Structure, Function, and Bioinformatics*, 86:84–96 (conducted prior to joining SLU). - Impact factor (2018): 2.501; Citations (in Google Scholar): 21
39. Jiang, L., Wan, Y., Anderson, J. C., Hou, J., Islam, S. M., Cheng, J., and Peck, S. C. (2017). Genetic dissection of arabidopsis map kinase phosphatase 1-dependent pamp-induced transcriptional responses. *Journal of experimental botany*, 68(18):5207–5220 (conducted prior to joining SLU). - Impact factor (2017): 5.354; Citations (in Google Scholar): 10
40. Li, H., Hou, J., Adhikari, B., Lyu, Q., and Cheng, J. (2017). Deep learning methods for protein torsion angle prediction. *BMC bioinformatics*, 18(1):417 (conducted prior to joining SLU). - Impact factor (2017): 2.213; Citations (in Google Scholar): 63
41. Hou, J., Adhikari, B., and Cheng, J. (2018a). Deepsf: deep convolutional neural network for mapping protein sequences to folds. *Bioinformatics*, 34(8):1295–1303 (conducted prior to joining SLU). - Impact factor (2017): 5.481; Citations (in Google Scholar): 207
42. Cao, R., Adhikari, B., Bhattacharya, D., Sun, M., Hou, J., and Cheng, J. (2017). Qacon: single model quality assessment using protein structural and contact information with machine learning techniques. *Bioinformatics*, 33(4):586–588 (conducted prior to joining SLU). - Impact factor (2017): 5.481; Citations (in Google Scholar): 94
43. Adhikari, B., Nowotny, J., Bhattacharya, D., Hou, J., and Cheng, J. (2016). Coneva: a toolbox for comprehensive assessment of protein contacts. *BMC bioinformatics*, 17(1):1–12 (conducted prior to joining SLU). - Impact factor (2016): 2.448; Citations (in Google Scholar): 27
44. Song, H., Lu, Y., Qu, Z., Mossine, V. V., Martin, M. B., Hou, J., Cui, J., Peculis, B. A., Mawhinney, T. P., Cheng, J., et al. (2016). Effects of aged garlic extract and fruarg on gene expression and signaling pathways in lipopolysaccharide-activated microglial cells. *Scientific reports*, 6:35323 (conducted prior to joining SLU). - Impact factor (2016): 4.569; Citations (in Google Scholar): 23
45. Hou, J., Acharya, L., Zhu, D., and Cheng, J. (2016a). An overview of bioinformatics methods for modeling biological pathways in yeast. *Briefings in functional genomics*, 15(2):95–108 (conducted prior to joining SLU). - Impact factor (2016): 4.098; Citations (in Google Scholar): 28
46. Cao, R., Bhattacharya, D., Hou, J., and Cheng, J. (2016). Deepqa: improving the estimation of single protein model quality with deep belief networks. *BMC bioinformatics*, 17(1):495 (conducted prior to joining SLU). - Impact factor (2016): 2.448; Citations (in Google Scholar): 170
47. Lei, W., Lu, Y., Hou, J., Li, J., Browning, J., Eichen, P., Cheng, J., Lubahn, D., Folk, W., Sun, G., et al. (2015). Immunomodulation of innate immune cells by sutherlandia frutescens: A transcriptomic analyses. *The FASEB Journal*, 29(1_supplement):593–3 (conducted prior to joining SLU). - Impact factor (2016): 5.259

48. Hou, J., Stacey, G., and Cheng, J. (2015). Exploring soybean metabolic pathways based on probabilistic graphical model and knowledge-based methods. *EURASIP Journal on Bioinformatics and Systems Biology*, 2015(1):5 (conducted prior to joining SLU). - Impact factor (2015): NA
49. Jo, T., Hou, J., Eickholt, J., and Cheng, J. (2015b). Improving protein fold recognition by deep learning networks. *Scientific reports*, 5:17573 (conducted prior to joining SLU). - Impact factor (2015): 5.228; Citations (in Google Scholar): 136
50. Li, J., Hou, J., Sun, L., Wilkins, J. M., Lu, Y., Niederhuth, C. E., Merideth, B. R., Mawhinney, T. P., Mossine, V. V., Greenlief, C. M., et al. (2015b). From gigabyte to kilobyte: a bioinformatics protocol for mining large rna-seq transcriptomics data. *PLoS one*, 10(4):e0125000 (conducted prior to joining SLU). - Impact factor (2015): 3.057; Citations (in Google Scholar): 13

Book Chapters

1. Hou, J., Wu, T., Guo, Z., Quadir, F., and Cheng, J. (2020c). The multicom protein structure prediction server empowered by deep learning and contact distance prediction. In *Protein Structure Prediction*, pages 13–26. Springer
2. Shi, X., Chen, C., Yang, H., Hou, J., Ji, T., Cheng, J., Veitia, R. A., and Birchler, J. A. (2020). The gene balance hypothesis: Epigenetics and dosage effects in plants. In *Plant Epigenetics and Epigenomics*, pages 161–171. Springer

Refereed Conference Papers

1. Yan, D., Azad, A., Hou, J., Chen, J. Y., and Zaki, M. J. (2023). 22nd international workshop on data mining in bioinformatics (biokdd 2023). In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5897–5898. - Acceptance rate: NA
2. Dougherty, S., Kumar, A., Hou, J., Tourani, R., and Tabakhi, A. M. (2023). A stealthy inference attack on split learning with a split-fuse defensive measure. In *2023 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9. IEEE. - Acceptance rate: NA
3. Rahbar, M., Chauhan, R. K., Shah, P. N., Cao, R., Si, D., and Hou, J* (2022). Deep graph learning to estimate protein model quality using structural constraints from multiple sequence alignments. In *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–10. - Acceptance rate: 29%; Citations (in Google Scholar): 1
4. Kosfeld, T., McMillan, J., DiPaolo, R. J., Hou, J., and Ahn, T.-H. (2020). Performance evaluation of viral infection diagnosis using t-cell receptor sequence and artificial intelligence. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–10. - Acceptance rate: 29%; Citations (in Google Scholar): 1
5. Chen, X., Akhter, N., Guo, Z., Wu, T., Hou, J., Shehu, A., and Cheng, J. (2020c). Deep ranking in template-free protein structure prediction. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–10. - Acceptance rate: 29%; Citations (in Google Scholar): 5

Theses (Hou and his graduated students)

1. Omar Al Akkad (2023). *Evaluation of Privacy Inference Attack on Deep Recurrent Neural Networks*. Master Thesis, Saint Louis University
2. Chauhan, Rahul Kumar (2021). *A Deep Learning Approach to Estimate Protein Model Quality Using Structural Constraints from Multiple Sequence Alignments*. Master Thesis, Saint Louis University

3. Hou, J (2019). Improving protein structure prediction by deep learning and computational optimization. *PhD Dissertation*, University of Missouri - Columbia

Conference Abstracts/ Posters

1. Hou, J, Mahdi Rahbar, Sheng Wang, R. C., and Hou, J. (2022). A deep graph learning approach to rank protein-protein interaction structures. *The 14th Critical Intelligent Systems for Molecular Biology (ISMB)*, 2022
2. Hou, J, Guo, Z., Liu, J., Wu, T., and Cheng, J. (2020b). Improving protein single-model and consensus quality assessment using inter-residue distance prediction and deep learning. *The 14th Critical Assessment of Techniques for Protein Structure Prediction (CASP14)*, 2020
3. Hou, J and Cheng, J. (2018a). Deep convolutional neural networks for predicting the quality of single protein structural model. *The 13th Critical Assessment of Techniques for Protein Structure Prediction (CASP13)*, Riviera Maya, Mexico, 2018
4. Hou, J and Cheng, J. (2018b). Large-scale integration of protein model quality assessment using deep learning and contact predictions. *The 13th Critical Assessment of Techniques for Protein Structure Prediction (CASP13)*, Riviera Maya, Mexico, 2018
5. Hou, J and Cheng, J. (2018c). Tertiary structure prediction assisted by saxs data, probabilistic modeling, deep learning, and contact predictions. *The 13th Critical Assessment of Techniques for Protein Structure Prediction (CASP13)*, Riviera Maya, Mexico, 2018
6. Hou, J, Wu, T., Cao, R., and Cheng, J. (2018c). Casp13 tertiary structure prediction by the multicom human group. *The 13th Critical Assessment of Techniques for Protein Structure Prediction (CASP13)*, Riviera Maya, Mexico, 2018
7. Hou, J, Wu, T., and Cheng, J. (2018d). Improving protein tertiary structure prediction by deep learning, contact prediction and domain recognition. *The 13th Critical Assessment of Techniques for Protein Structure Prediction (CASP13)*, Riviera Maya, Mexico, 2018
8. Hou, J, Adhikari, B., Cao, R., and Cheng, J. (2016b). Saxs-assisted tertiary structure prediction by multicom. *The 12th Critical Assessment of Techniques for Protein Structure Prediction (CASP12)*, Gaeta, Italy, 2016
9. Hou, J, Adhikari, B., Cao, R., Crivelli, S. N., and Cheng, J. (2016c). Tertiary structure prediction by wfall-cheng. *The 12th Critical Assessment of Techniques for Protein Structure Prediction (CASP12)*, Gaeta, Italy, 2016
10. Hou, J, Li, J., Adhikari, B., and Cheng, J. (2016d). Improved integration of template-based and template-free model sampling methods for protein structure prediction. *The 12th Critical Assessment of Techniques for Protein Structure Prediction (CASP12)*, Gaeta, Italy, 2016
11. Hou, J, Li, J., and Cheng, J. (2016e). Protein tertiary structure prediction by multicom-cluster server. *The 12th Critical Assessment of Techniques for Protein Structure Prediction (CASP12)*, Gaeta, Italy, 2016
12. Li, J., Hou, J, Sun, L., Wilkins, J. M., Lu, Y., Niederhuth, C. E., Merideth, B. R., Mawhinney, T. P., Mossine, V. V., Greenlief, C. M., et al. (2015a). From gigabyte to kilobyte: A bioinformatics protocol for mining large rna-seq. *ACM-BCB*, Atlanta, GA, USA, 2015

13. Jo, T., Hou, J., Eickholt, J., and Cheng, J. (2015a). Deep learning network for protein fold recognition. *The Mid-South Computational Biology and Bioinformatics Conference (MCBIOS)*, Little Rock, AR, USA, 2015

Selected Talks

1. **Invited talk** at *Barnes-Jewish Hospital, Saint Louis*, 11/2023
Talk title: "Machine Learning applications in healthcare"
2. **Short Research talk** at *SLU, 2023 AI for Medicine conference*, 10/2023
Talk title: "AI and big data analytics in healthcare data"
3. **Department Colloquium talk** at *SLU Computer Science Department Colloquium*, 09/2023
Talk title: "The applications of deep learning techniques in protein structure prediction"
4. **Paper presentation at 2023 SIGKDD conference, BioKDD workshop**, Long Beach, 08/2023
Talk title: "AnglesRefine: refinement of 3D protein structures using Transformer based on torsion angles"
5. **Department Colloquium talk** at *SLU Computer Science Department Colloquium*, 10/2022
Talk title: "The applications of deep learning techniques in protein structure prediction"
6. **Poster presentation at 2022 ISMB conference**, Madison, 07/2022
Poster title: "A deep graph learning approach to rank protein-protein interaction structures."
7. **Invited talk** in *Data Analysis & Intelligent Systems (DAIS) group*, University of Washington, Bothell, WA, 02/2021,
Talk title: "Integration of cryo-EM and inter-residue distance prediction from protein folding",
8. **Department Colloquium talk** in *Parks College Colloquium*, Saint Louis University, St. Louis, MO, US, 09/2020.
Talk title: "DeepRank2: Utilizing computer vision techniques to improve protein model quality assessment using inter-residue distance prediction and deep learning"
9. **Short Research talk** in *Department of Biological Sciences Research Retreat*, Saint Louis University, St. Louis, MO, US, 09/2019
Talk title: "Protein Structure Prediction using Deep Learning and Computational Optimization"
10. **Poster presentation at The 13th Critical Assessment of Techniques for Protein Structure Prediction (CASP13)**, Riviera Maya, Mexico, Dec 2018
Poster title: "CASP13 Tertiary Structure Prediction by the MULTICOM Human Group".
11. **Invited talk at The 13th Critical Assessment of Techniques for Protein Structure Prediction (CASP13)**, Riviera Maya, Mexico. Dec 2018
Talk title: "Large-scale integration of protein model quality assessment using deep learning and contact predictions"
12. **Paper presentation at 2015 MidSouth Computational Biology and Bioinformatics Society (MCBIOS)**, Little Rock, Arkansas, US, March 2015
Talk title: "Improving protein fold recognition by deep learning networks"

III. Teaching and Advising

Teaching

Statistics of Teaching Impact

- **Taught 3 Different Courses** in Machine Learning, Algorithms in Computational Biology, Introduction to Computer Science, and **2 General Research Courses**
- **Co-taught 1 course** in Bioinformatics Colloquium
- **Average Teaching Evaluation Score from 2019 to 2024:**
 - **3.93/4.0** - CSCI 1020: Introduction to Computer Science: Bioinformatics
 - **3.90/4.0 (3.97/4.0 since 2022)** - CSCI 4750/5750: Introduction to Machine Learning
 - **3.98/4.0** - BCB 5300: Algorithms in Computational Biology

Courses Designed and Taught (5)

1. CSCI 1020: Introduction to Computer Science: Bioinformatics (Fall 2019, Spring 2020, Fall 2020, Spring 2021, Spring 2023)
2. CSCI 4750/5750: Introduction to Machine Learning (Spring 2020, Fall 2020, Fall 2021, Spring 2022, Fall 2022, Spring 2023, Fall 2023, Spring 2024)
3. BCB 5300: Algorithms in Computational Biology (Fall 2021, Fall 2022, Fall 2023)
4. CSCI 5970: Research Topics (Summer 2021, Spring 2022, Summer 2022, Fall 2022, Spring 2023)
5. CSCI 5990: Thesis Research (Spring 2021, Fall 2022, Spring 2023)

Courses Co-Taught (1)

1. BCB 5810: Bioinformatics Colloquium (Fall 2020)

Advising

Statistics of Advising Impact

- **Advised 2 graduate students as chairperson of master's written thesis**
- **Advised 8 master's students as supervisor in research projects**
- **Co-supervised 2 PhD students**
- **Co-advised 5 master's students in University of Washington-Bothell**
- **Advised 3 undergraduate students, 2 high school students**
- **Mentored 33 CS undergraduates and BCB graduate students in course selection**

Current Lab Members

- Vinay Chaudhari, MS-AI SLU (09/2023 - Present); Project: Development of AI-driven RNA motif data analysis framework

Prior SLU Graduate Students Advising as Thesis Chair

1. Omar AI Akkad; MS-AI SLU (01/2022 - 05/2023) Current: PhD program of GIS at Saint Louis University

2. Rahul Chauhan; SLU (09/2019-05/2021); Publication: [Rahbar et al. (2022)]
Master thesis: A deep learning approach to estimate protein model quality using structural constraints from multiple sequence alignments.
Current: Software Development Engineer II at GoDaddy

Prior SLU Graduate Students Advising as Project Supervisor

1. Charlz Nithin Jerold; MS-BCB SLU (09/2022 - 05/2024). Current: full-time Bioinformatics Research Analyst at Washington University at St. Louis
2. Nexhi Sula (Female), MS-AI SLU (01/2023 - 05/2024). Next step: will join PhD program of Computer Science at SLU in Fall 2024
3. Mahdi Rahbar; MS-AI SLU (09/2021 - 05/2023). Publication: [Rahbar et al. (2022)]
4. Chandana Enugala (Female), MS-AI SLU (06/2022 - 07/2023)
5. Varun Julakanti; MS-CS SLU (Spring 2023)
6. Pankil Shah; MS-CS SLU (01/2021-05/2022)
7. Yu Zhang (Female), MS-BCB SLU (01/2021- 05/2022). Current: full-time Bioinformatician at Washington University at St. Louis
8. Aasritha Nallapu (Female), MS-BCB SLU (Spring 2022). Current: full-time Bioinformatician at Washington University at St. Louis

Graduate Student Co-Advising as Thesis / Projects / Dissertation Chair (including co-supervision of 2 PhD students and and 5 MS students in total at University of Washington - Bothell)

1. Tong Si (Female); PhD student in the Department of Mathematics at SLU. PhD research co-advisor (09/2022 - Present).
Publication: [Si et al. (2023)]
2. Noor Al-Hammadi (Female); PhD student in the Department of Health & Clinical Outcomes Research at SLU. PhD research co-advisor (10/2020 - 08/2022). Current: Assistant Professor at SLU
3. Esha Rajesh Gavali; Graduate student in the Department of Computer Science at University of Washington Bothell. Committee Chair: Dong Si, Ph.D (02/2023 - Present).
4. Chloe Ma; Graduate student in the Department of Computer Science at University of Washington. Committee Chair: Dong Si, Ph.D (05/2023 - 05/2024).
5. Jason Chen; Graduate student in the Department of Computer Science at University of Washington. Committee Chair: Dong Si, Ph.D (02/2021 - 12/2023).
Publication: [Chen et al. (2024)]
6. Andrew Hitoshi Nakamura; Graduate student in the Department of Computer Science at University of Washington. Committee Chair: Dong Si, Ph.D (07/2021 - 08/2022).
Publication: [Nakamura et al. (2023)]
7. Nathan Ranno; Graduate student in the Department of Computer Science at University of Washington. Committee Chair: Dong Si, Ph.D (02/2020 - 06/2021).

Prior SLU Undergraduate Students Advising as Project Supervisor

1. Bhanu Pujari; BS Neuroscience SLU (FL2021 - SP2024). Next step: Join SLU Medical School (SOM) to start graduate study in Fall 2024
2. Daniel Mao; BS Computer Science (SP2022 - SP2023). Current: Software Developer at Saint Louis University

3. Ciri Chandana Pocha (Female); BS Neuroscience SLU (FL2020 - FL2021); Publication: 1 published journal [Hippe et al. (2022)]

High School Intern / Student Advising (2 students)

1. 10/2023 - Present, Grace Fu; Parkway south high school, Manchester, Missouri, United States.
2. 06/2022-08/2022, Gabriel Schwarz; Highland High School, Missouri, United States.

Academic mentor (33 students in total)

- Provide active mentorship to 33 CS undergraduates and BCB graduate students since 2019

IV. Services

Statistics of Main Internal and External Services

- Served on the degree committee of 17 graduate students
- Served on +8 various committees of department, college, and university
- Collaborating/Consulting +10 research groups on bioinformatics, machine learning, artificial intelligence research at SLU
- Served as US Department of Energy (DOE) panelist and grant application reviewer 1 time
- Associate Editor of Protein Bioinformatics (specialty section of Frontiers in Bioinformatics)
- Served as handling editor for journals such as Protein Bioinformatics, and IEEE/ACM Transactions on Computational Biology and Bioinformatics
- Membership with AAAS and ACM
- Reviewed papers for 16 journals
- Co-chaired BioKDD workshop in KDD 2023 conference
- Served on program committees of conferences and workshops 2 times

Department Services

Graduate Thesis / Dissertation Committees at SLU (11 students): Jinhui Li (current), Cory Gardner (current), Tong Si (Female, current), Nexhi Sula (Female, SP2024), Wazma Ali (Female, SP2024), Muhammad Aqeel (FL2024), Andrew Beiler (FL2023), Matt Dreyer (SU2023), Mahdi Rahbar (SP2023), Noor Al-Hammadi (Female, SU2022), Princewill Okorie (SP2022)

Graduate Thesis / Capstone Committees at University of Washington-Bothell (6 students): Esha Rajesh Gavali (Female, current), Saurav Jayakumar (current), Chloe Ma (Female, SP2024), Jason Chen (FL2023), Andrew Hitoshi Nakamura (SU2022), Nathan Ranno (SU2021)

Research Collaboration Services in the CS Department: Collaborating with many faculty in the computer science department on various research projects, including published work Kosfeld et al. (2020), Dougherty et al. (2023).

Department Committee Services

1. 2024, PhD qualifying exam committee of Cory Gardner, Department of Computer Science, Saint Louis University
2. 2024, Serve on the Ad-hoc Bioinformatics and Computational Biology Academic Program Review Committee, Saint Louis University
3. 2023 - Present, Technology Committee, Department of Computer Science, Saint Louis University

4. 2022-2023, Supervisor for Computer Science Capstone projects, Department of Computer Science, Saint Louis University
5. 2022 - Present, Co-organizer of MS-BCB program orientation, Saint Louis University
6. 2022 - Present, Department representative in Curriculum Committee of School of Science and Engineering (SSE)
7. 2022, Faculty representative in Academic fair, Department of Computer Science, Saint Louis University
8. 2021-2022, Committee member in department undergraduate curriculum/affairs committee, Department of Computer Science, Saint Louis University
9. 2020-2021, Supervisor for Computer Science Capstone projects, Department of Computer Science, Saint Louis University
10. 2020 - Present, Serve as an academic advisor to multiple MS-BCB graduate students, Department of Computer Science, Saint Louis University.
11. 2020, Co-organizer of weekly BCB colloquium seminar series, Department of Computer Science, Saint Louis University
12. 2019 - Present, Serve as an academic mentor to multiple CS undergraduate students, Department of Computer Science, Saint Louis University.
13. 2019 - Present, Provided reference letters for BCB/CS/AI students in support of internships, job openings, and PhD applications, Department of Computer Science, Saint Louis University
14. 2019 - Present, Participating in almost all Computer Science (CS) faculty meetings

College Services

1. 2023 - Present, Chair of SSE Curriculum Committee, School of Science and Engineering (SSE), Saint Louis University
2. 2023 - Present, Participating in faculty meetings and activities, School of Science and Engineering (SSE), Saint Louis University

University Services

Affiliation Services

Affiliated with *SLU Master of Bioinformatics and Computational Biology (BCB) program* (since 2019), *SLU AHEAD Institute* (since 2024). Participating in meetings, symposiums and external / internal seminars organized by these two organizations.

Research Collaboration and Consulting Services at Saint Louis University (+5 groups / PIs)

1. Collaboration with Dr. Longping Victor Tse in Department of Molecular Microbiology & Immunology to develop computational methods for identification of protein epitope structure
2. Collaboration with Dr. Brent Znosko in Department of Chemistry to develop computational framework for RNA motif analysis
3. Collaboration with Dr. Dapeng Zhang in Department of Biology to develop bioinformatics methods for Genome-wide discovery and analysis of bacterial toxin systems
4. Collaboration with Dr. Haijun Gong in Department of Mathematics to develop deep learning algorithms for scRNA-seq data analysis.

5. Collaboration with Dr. Noor Al-Hammadi in Department of Health & Clinical Outcomes Research to develop machine learning algorithms for warfarin dosing prediction.
6. Collaboration with Drs. Chihou Lei (Aerospace and Mechanical Engineering), Yan Gai (Biomedical Engineering), Benjamin Hutz (Mathematics) on a variety of research topics and grant writing.

Other University Services

1. 2024, Served as a PhD committee member for Jinhui Li who is a PhD student in the Department of Biology, Saint Louis University
2. 2023 - Present, Served as a PhD committee member for Tong Si who is a PhD student in the Department of Mathematics, Saint Louis University
3. 2022, Moderator in SLU Sigma Xi Symposium, Saint Louis University
4. 2022, Judge in SLU GSA Research Symposium, Saint Louis University
5. 2020 - 2022, Serve as a PhD committee member for Noor Al-Hammadi who is a PhD student in the Department of Health Outcomes Research
6. 2020, Served as a PhD committee member for Sourav Bhadra who is a PhD student in the Department of Environmental Science and GIS

Regional, National and International Services

Membership

AAAS (since 2020), ACM (since 2019)

Journal Paper Reviewing (16 International Journals)

Nature Computational Science, Journal of Biotechnology, IEEE Transactions on Pattern Analysis and Machine Intelligence, Frontiers in Bioinformatics, Briefings in Bioinformatics, Bioinformatics, BMC Supplements, Computational and Structural Biotechnology, Nature Scientific Report, Plos One, Chinese Journal of Electronics, Journal of Digital Discovery, Journal of Neural Networks, Journal of Entropy, Journal of Applied Sciences, Database: the Journal of Biological Databases and Curation.

Workshop co-chair, Committee Member, Editorial Board, Grant Review, and Other Services

- 2024, Program Committee, IEEE International Conference on Bioinformatics and Biomedicine (BIBM)
- 2023-2024, Co-editor, IEEE/ACM Transactions on Computational Biology and Bioinformatics
- 2023, Handling editor, Protein Bioinformatics
- 2023, Workshop co-chair: The 22nd International Workshop on Data Mining in Bioinformatics (BIOKDD), Long Beach, CA, USA, Aug. 2023.
- 2023, Panel Reviewer: US Department of Energy (DOE) Biological & Environmental Research (BER).
- 2022–Present, Associate Editor: Journal of Frontiers in Bioinformatics.
- 2022, Judge in DataFest competition at Pacific Lutheran University