Yingjun Dong, Ph.D.

Postdoctoral Research Fellow | Phone: (607) 304-0550 | Email: Yingjun.Dong@uth.tmc.edu | Website: yingjundong.com

EDUCATION

Binghamton University, State University of New York

Ph.D. in System Science — January 2017 - August 2022

GPA: 3.90/4.00

Binghamton University, State University of New York

M.A. in Economics — August 2015 - December 2016

GPA: 3.22/4.00

Shandong Normal University

B.A. in Economics — September 2011 - June 2015

GPA: 3.00/4.00

PROFESSIONAL EXPERIENCE

University of Texas Health Science Center at Houston

Postdoctoral Research Fellow — September 2022 - Present

- Develop deep learning models for acute stroke imaging tasks using 3D CT angiography and related clinical data.
- Build self-supervised learning frameworks for paired 3D CTA images and radiology reports to support multiple downstream detection tasks.
- Apply multimodal image-text learning to improve representation learning and generalizability in stroke imaging analysis.
- Use LLaMA-based large language models to generate and summarize radiology report content.
- Explore vision-language and visual large language model approaches for medical imaging diagnosis and clinical decision support.
- Work with medical imaging preprocessing tools including FSLeyes, SimpleITK, and 3D Slicer.

Binghamton University, State University of New York

Graduate Research Assistant — August 2019 - August 2022

- Supported research activities in the Center for Collective Dynamics of Complex Systems.
- Assisted with CoCo seminar organization and invited-speaker communication.
- Independently organized the CoCo Student-Faculty Mixer event.
- Conducted research on machine learning, computational social systems, audio clustering, and narrative data analysis.

Binghamton University, State University of New York

Teaching Assistant — August 2017 - May 2018

- Teaching assistant for SSIE 505: Probability and Statistics.
- Teaching assistant for SSIE 520: Simulation and Modeling.

RESEARCH PROJECTS

Generalizable Self-Supervised Learning for Brain CTA in Acute Stroke — September 2022 – August 2025

- Developed self-supervised learning strategies using paired 3D CTA images and radiology reports for acute stroke-related prediction tasks.
- Fine-tuned pretrained representations on multiple clinically relevant labels to evaluate generalizability across tasks.
- Demonstrated the potential of image-text pretraining to improve medical imaging efficiency and real-world applicability.
- Tools: Python, PyTorch, Scikit-Learn, FSLeyes, SimpleITK.

Self-Supervised Learning on Image-Text Paired Medical Data September 2023 – August 2025

- Applied LLaMA to generate radiology report summaries.
- Built self-supervised learning workflows using paired imaging and report data for image classification tasks.
- Tools: Python, PyTorch, HuggingFace Transformers, LLMs.

Registration-Agnostic Imaging Models with 3D CTA Images September 2022 - May 2023

- Designed a self-supervised learning approach using paired registered and unregistered 3D CTA images.
- Reduced dependency on time-consuming image registration preprocessing for large vessel occlusion detection.
- Supported more practical deployment of CTA-based deep learning models in acute stroke workflows.
- Tools: Python, PyTorch, TensorFlow, Scikit-Learn, FSLeyes, SimpleITK.

Utterance Clustering on Real-World Audio Data August 2019 - August 2022

- Preprocessed stereo audio data with FFmpeg and shell scripting.
- Applied Gaussian mixture models to cluster utterances from real-world audio recordings.
- Improved preliminary clustering performance through signal preprocessing.
- Tools: Python, Scikit-Learn, FFmpeg, Shell Script.

Feature Selection on Facial Landmarks August 2018 - April 2019

- Processed images and facial landmarks using OpenCV.
- Applied a hybrid feature selection method combining information theory, hierarchical clustering, and genetic algorithms.
- Trained SVM classifiers on selected facial-landmark features to compare classification accuracy and computational efficiency.
- Used PyTorch and OpenCV to improve efficient landmark labeling.
- Tools: Python, Scikit-Learn, PyTorch, OpenCV.

Narrative Background Analysis with Doc2Vec August 2018 - April 2019

- Analyzed similarities among users' narrative background descriptions using Doc2Vec representations.
- Supported computational analysis of background diversity in collective ideation experiments.
- Tools: Python.

PUBLICATIONS

- Jeevarajan, Jerome A.; **Dong, Yingjun**; Ballekere, Anjan; Marioni, Sergio Salazar; Niktabe, Arash; Abdelkhaleq, Rania; Sheth, Sunil A.; Giancardo, Luca. **Can CTA-Based Machine Learning Identify Patients for Whom Successful Endovascular Stroke Therapy Is Insufficient?** *American Journal of Neuroradiology*, 46(12), 2500-2506, 2025.
- **Dong, Yingjun**; Pachade, Samiksha; Roberts, Kirk; Jiang, Xiaoqian; Sheth, Sunil A.; Giancardo, Luca. **Generalizable self-supervised learning for brain CTA in acute stroke.** *Computers in Biology and Medicine*, 184, 109337, 2025.
- Cao, Yiding; **Dong, Yingjun**; Kim, Minjun; MacLaren, Neil G.; Pandey, Sriniwas; Dionne, Shelley D.; Yammarino, Francis J.; Sayama, Hiroki. **Effects of network connectivity and functional diversity distribution on human collective ideation.** *npj Complexity*, 2(1), 2, 2025.
- **Dong, Yingjun**; Pachade, Samiksha; Liang, Xiaomin; Sheth, Sunil A.; Giancardo, Luca. **A self-supervised learning approach for registration agnostic imaging models with 3D brain CTA.** *iScience*, 27(3), 2024.
- Jeevarajan, Jerome; **Dong, Yingjun**; Azeem, Hussain; Salazar-Marioni, Sergio; Giancardo, Luca; Sheth, Sunil. **Machine Learning Enabled Outcome Predictions Using CT Angiography for Patients Undergoing Endovascular Stroke Therapy (P1-5.006).** *Neurology*, 102(7 supplement 1), 6321, 2024.
- Pachade, Samiksha; Datta, Surabhi; **Dong, Y.**; Salazar-Marioni, S.; Abdelkhaleq, Rania; Niktabe, A.; Roberts, Kirk; Sheth, Sunil A.; Giancardo, Luca. **Self-supervised learning with radiology reports, a comparative analysis of strategies for large vessel occlusion and brain CTA images.** *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 1-5, 2023.
- Pandey, Sriniwas; Cao, Yiding; **Dong, Yingjun**; Kim, Minjun; MacLaren, Neil G.; Dionne, Shelley D.; Yammarino, Francis J.; Sayama, Hiroki. **Generation and influence of eccentric ideas on social networks.** *Scientific Reports*, 13(1), 20433, 2023.

- Cao, Shun; MacLaren, Neil G.; Cao, Yiding; Marshall, Jason; **Dong, Yingjun**; Yammarino, Francis J.; Dionne, Shelley D.; Mumford, Michael D.; Connelly, Shane; Martin, Robert W. **Group Size and Group Performance in Small Collaborative Team Settings: An Agent-Based Simulation Model of Collaborative Decision-Making Dynamics.** *Complexity*, 2022(1), 8265296, 2022.
- Cao, Yiding; **Dong, Yingjun**; Kim, Minjun; MacLaren, Neil G.; Pandey, Sriniwas; Dionne, Shelley D.; Yammarino, Francis J.; Sayama, Hiroki. **Visualizing collective idea generation and innovation processes in social networks.** *IEEE Transactions on Computational Social Systems*, 10(5), 2234-2243, 2022.
- **Dong, Yingjun**; MacLaren, Neil G.; Cao, Yiding; Yammarino, Francis J.; Dionne, Shelley D.; Mumford, Michael D.; Connelly, Shane; Sayama, Hiroki; Ruark, Gregory A. **Utterance clustering using stereo audio channels.** *Computational Intelligence and Neuroscience*, 2021(1), 6151651, 2021.
- Cao, Shun; MacLaren, Neil G.; Cao, Yiding; **Dong, Yingjun**; Sayama, Hiroki; Yammarino, Francis J.; Dionne, Shelley D.; Mumford, Michael D.; Connelly, Shane; Martin, Robert. **An Agent-Based Model of Leader Emergence and Leadership Perception within a Collective.** *Complexity*, 2020(1), 6857891, 2020.
- Cao, Yiding; **Dong, Yingjun**; Kim, Minjun; MacLaren, Neil G.; Kulkarni, Ankita; Dionne, Shelley D.; Yammarino, Francis J.; Sayama, Hiroki. **Capturing the production of innovative ideas: An online social network experiment and "idea geography" visualization.** *Conference of the Computational Social Science Society of the Americas*, 341-354, 2020.
- **Dong, Yingjun**; Sayama, Hiroki. **Mutual-information-based feature selection for facial emotion recognition on light-weight devices.** *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2455-2461, 2019.
- Cao, Yiding; **Dong, Yingjun**; Kim, Minjun; MacLaren, Neil G.; Pandey, Sriniwas; Dionne, Shelley D.; Yammarino, Francis J.; Sayama, Hiroki. **Experimental data for: "Effects of network connectivity and functional diversity distribution on human collective ideation."** Dataset, 2024.
- Cao, Yiding; **Dong, Yingjun**; Kim, Minjun; MacLaren, Neil G.; Pandey, Sriniwas; Dionne, Shelley D.; Yammarino, Francis J.; Sayama, Hiroki; York, Binghamton. **Effects of Network Connectivity and Diversity Distribution on Human Collective Ideation.** *CoRR*, 2023.
- **Dong, Yingjun**. **Machine Learning Applications for Multimodal Human Behavior Analysis.** Doctoral dissertation, State University of New York at Binghamton, 2022.
- **Dong, Yingjun**; MacLaren, Neil G.; Cao, Yiding; Yammarino, Francis J.; Dionne, Shelley D.; Mumford, Michael D.; Connelly, Shane; Sayama, Hiroki; Ruark, Gregory A. **Speaker diarization using stereo audio channels: Preliminary study on utterance clustering.** *arXiv preprint arXiv:2009.05076*, 2020.

SELECTED PRESENTATIONS

- **Yingjun Dong** and Hiroki Sayama. **Optimizing Facial Feature Extraction for Emotion Detection on Mobile Devices.** NERCCS 2019: Second Northeast Regional Conference on Complex Systems, April 3-5, 2019, Binghamton, NY.
- Yiding Cao, **Yingjun Dong**, Minjun Kim, Neil MacLaren, Ankita Kulkarni, Shelley Dionne, Francis Yammarino, and Hiroki Sayama. **Examining the effects of expertise diversity on collective design and innovation using an online social network experiment and "idea geography" visualization: An initial report.** NERCCS 2019: Second Northeast Regional Conference on Complex Systems, April 3-5, 2019, Binghamton, NY.
- Yiding Cao, **Yingjun Dong**, Minjun Kim, Neil MacLaren, Ankita Kulkarni, Shelley Dionne, Francis Yammarino, and Hiroki Sayama. **Examining the effects of expertise diversity on collective design and innovation using an online social network experiment and "idea geography" visualization: A secondary report.** Conference on Complex Systems, September 30-October 4, 2019, Singapore.
- Yiding Cao, **Yingjun Dong**, Minjun Kim, Neil MacLaren, Shelley Dionne, Francis Yammarino, and Hiroki Sayama. **Background diversity and network structure in collective design and innovation: Perspectives with online social network experiments.** NetSci 2020: International School and Conference on Network Science, September 17-25, 2020, Rome, Italy.
- Sriniwas Pandey, Yiding Cao, **Yingjun Dong**, Minjun Kim, Neil G. MacLaren, Shelley D. Dionne, Francis J. Yammarino, and Hiroki Sayama. **Are we fascinated by eccentric ideas?** Poster presented at NetSci 2022: International School and Conference on Network Science, July 11-29, 2022, Shanghai, China / online.

ACADEMIC SERVICE

- Journal Reviewer: Journal of the American Heart Association (JAHA), Stroke: Vascular and Interventional Neurology (SVIN)
- Conference Reviewer: IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
- Poster Judge: The 32nd Keck Annual Research Conference.
- Editorial Staff: Northeast Journal of Complex Systems (NEJCS).

TECHNICAL SKILLS

Programming: Python, Shell Script, MATLAB, R, LaTeX, HTML, JavaScript, CSS

Machine Learning and Deep Learning: PyTorch, TensorFlow, Scikit-Learn, Transformers, HuggingFace, LangChain

Medical Imaging: 3D CTA analysis, FSLeyes, SimpleITK, 3D Slicer

Data and Media Processing: OpenCV, FFmpeg, MongoDB

Methods: Self-supervised learning, multimodal image-text learning, medical image analysis, natural language processing, large language models, data mining, speech processing, clustering, feature selection