Jie (Jennifer) Yu

Education	**Tri-Institutional PhD Program at Cornell University/Weill Cornell Medicine, Memorial Sloan Kettering Cancer Center, Rockefeller University**, New York, NY, USA

PhD in Computational Biology and Medicine — August 2024 - May 2029

University of Toronto, Toronto, ON, Canada — September 2022 - June 2024
Master of Science, Computer Science - *fully-funded research Master's Degree*
Supervisor: Dr. Anna Goldenberg
Research Areas: Machine Learning, Digital Health, Digital Pathology, Precision Medicine, Time Series, Computer Vision, Causality, Spatial Biology

University of Waterloo, Waterloo, ON, Canada — September 2017 - June 2022
Bachelor of Applied Science, *Honours Biomedical Engineering - Co-op, Graduate with Distinction*
Minor: AI and Computing

Publications

J. Yu *, Z. Wu, A. T. Mayer, A. Trevino and J. Zou. *A Multi-Granularity Approach to Similarity Search in Multiplexed Immunofluorescence Images.* [Proceedings Paper] Proceedings of the 18th Machine Learning in Computational Biology meeting, [10.1101/2023.11.26.568745v1], Seattle, WA, USA, 2023 (MLCB '2023).

J. Yu *✉, T. Behrouzi *, K. Garg *, S. Tonekaboni and A. Goldenberg. *Dynamic Interpretable Change Point Detection for Physiological Data Analysis.* [Proceedings Paper] Proceedings of the 3rd Machine Learning for Health Symposium, 10.48550/arXiv.2211.03991, New Orleans, LA, USA, 2023 (ML4H '2023).

J. Yu *, M. Ali, R. Zhu, P. Edke. and A. Goldenberg. *Childhood Adversity's Impact on Dynamic Mental Health During and Post Pregnancy: A Causal Approach* [Proceedings Paper], IEEE SDS2023 Workshop: Data science techniques on data for neurodegenerative diseases and mental disorders, ceur-ws.org/Vol-3521/paper1.pdf, Zurich, Switzerland, 2023 (IEEE SDS '2023).

A. Hussain, Z. Zhang, **J. Yu**, R. Wei, H. Arshad, J. Lew, C. Jagan, Y. Wang, JH. Chen and J. Huizinga. *Haustral Rhythmic Motor Patterns of the Human Large Bowel Revealed by Ultrasound.* [Journal Article]. American Journal of Physiology-Gastrointestinal and Liver Physiology, 10.1152ajpgi.00068.2023, 2023.

H. Abedi, M. Ma, **J. Yu**, J. He, A. Ansariyan and G. Shaker. *Deep Learning-Based In-Cabin Monitoring and Vehicle Safety System Using a 4D Imaging Radar Sensor* [Journal Article], 10.1109/JSEN.2023.3270043, IEEE Sensor Journal, 2023.

H. Abedi, M. Ma, **J. Yu**, J. He, A. Ansariyan and G. Shaker. *On the Use of Machine Learning and Deep Learning for Radar- Based Passenger Monitoring* [Proceedings Paper], 10.1109/AP-S/USNC-URSI47032.2022.9887034 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science, Denver, CO, USA, 2022.

Under Review

R. Zhu, **J. Yu**, S. Friend, S. M. Goodday, B. Wang, A. Goldenberg. *Towards a personalized pregnancy experience: Forecasting symptoms using graph neural networks and digital health technologies.* in review for Machine Learning for Health (ML4H) Symposium, Vancouver, Canada, 2024.

S. M. Goodday, R. Yang, E. Karlin, J. Tempero, C. Harry, A. Brooks, T. Behrouzi, **J. Yu**, A. Goldenberg, M. Francis, D. Karlin, C. Centen, S. Smith and S. Friend. *Does anyone fit the average? Describing the heterogeneity of pregnancy symptoms using wearables and mobile apps.*, [10.1101/2024.04.26.24306455], 2024.

Abstract

JH. Chen, L. Liu, **J. Yu**, and J. D. Huizinga. *Evaluating stimulus-based heart rate variability assessment in chronic constipation using a machine learning approach.*, Machine Learning in Healthcare, Toronto, 2024 (MLHC'2024)

J. Yu †, D. Foster, and S. Pai.*Extending Similarity Network-Based Classifiers to Non-Coding Genome and Deep Learning*, Machine Learning in Computational Biology, Online, 2021 (MLCB'2021).

---

*First author

Under Preparation	**J. Yu** *, S. Goodday, A. Goldenberg, and S. Friend. *Time Series Analysis of Wearable Data for Understanding Delivery Readiness in Pregnancy.*
	J. Yu *, T. Behrouzi *, S. Goodday, A. Goldenberg, and S. Friend. *Exploring the risks of using wearable data to inform pregnant individuals.*

Invited Talks	*Childhood Adversity's Impact on Dynamic Mental Health During and Post Pregnancy: A Causal Approach* [Lighting Talk], IEEE SDS2023 Workshop: Data science techniques on data for neurodegenerative diseases and mental disorders, Zurich, Switzerland, 2023 (IEEE SDS '2023).
	Time for Baby: Leveraging Wearables Data to Enhance Personalized Pregnancy Outcomes - Delivery Readiness [Research Talk], Vector Endless Summer School: Health Roundup Seminar, Toronto, Canada, 2023.

Research
Experience

Cornell University/Weill Cornell Medicine (in affiliation with Memorial Sloan Kettering Cancer Center and the Rockefeller University) 2024.08-Present
Position: Graduate Researcher | Supervisor: Dr. Christina Leslie
- Designing and implementing deep learning models to unravel complex biological processes.
- Developing and implementing algorithms to integrate multi-omic data from diverse sources to uncover the mechanisms of gene regulation that play a critical role in cancer development.

University of Toronto (in affiliation with the Vector Institute & the Hospital for Sick Children) 2022.05-2024.08
Position: Graduate Researcher | Supervisor: Dr. Anna Goldenberg
Collaborators: Dr. Stephen Friend, Dr. Sarah Goodday
Worked on *Dynamic Interpretable Change Point Detection for Physiological Data Analysis* as the first author, and working on *Time Series Analysis of Wearable Data for Understanding Delivery Readiness in Pregnancy*
- Conducted a study on exploring the risks of using wearable data to inform pregnant individuals within the context of ongoing digital health research
- Proposed the concept of individualized delivery readiness and developed ML forecasting models.
- Defined research problems and objectives for the *Better Understanding the Metamorphosis of Pregnancy (BUMP)* study by identifying research gaps in existing work.
- Designed and implemented a novel dynamic change point detection method to detect pregnancy-related events with better performance and interpretability for physiological data than SOTA methodologies.
- Developed interpretable time-series forecasting models for pregnancy due date and delivery readiness prediction.

Worked on *Modelling Childhood Adversity's Impact on Dynamic Mental Health During and Post Pregnancy* as the first author
- Investigated and modeled the dynamic relationships between mental health variables during and post-pregnancy using causal frameworks such as the PC algorithm and DAGs with NO TEARS method.

Project manager:
- Facilitated knowledge-sharing by developing an onboarding Wiki page and organizing journal review sessions, enhancing team productivity.
- Managing a team of 6 to work on a wearables data missingness imputation software and discussed with key collaborators, including 4YouandMe, Evidation Health, Oxford University, and MIT.
- Participated in the recruitment process for data engineering roles by conducting interviews and assessing candidates to secure top-tier talent for advancing data management and software infrastructure.

Enable Medicine (in collaboration with Stanford University) 2023.05-09
Position: Research Scholar | Supervisors: Dr. Alexandro Trevino, Dr. James Zou, Zhenqin Wu
Worked on *A Multi-Granularity Approach to Similarity Search in Multiplexed Immunofluorescence Images* as the first author.
- Developed a similarity search pipeline on CODEX multiplexed tissue images, leveraging advanced self-supervised learning and multimodal models for feature extraction.

- Developed a patient similarity search algorithm that employs an entropy-based aggregation method to enable searching at higher, multi-granular levels.
- Benchmarked various feature generation approaches to handle high-dimensional images and tested them on various vision foundation models such as DinoV2, ResNet and PLIP image encoders.
- Conducted evaluations using datasets from different tissues on both patch-level and patient-level to demonstrate the framework's effectiveness and generalizability.

University of Waterloo (in collaboration with McMaster University)   2021.09-2022.04
Position: Researcher | Supervisor: Dr. Alexander Wong
Collaborators: Dr. Jan Huizinga & Dr.Jihong Chen
Worked on *Haustral Rhythmic Motor Patterns of the Human Large Bowel Revealed by Ultrasound*.
- Designed and implemented an active contour-based segmentation method that lowered the mean absolute error of colon wall detection by 90% in abdominal ultrasound images.
- Developed a novel algorithm to generate spatiotemporal maps of colon diameter change over time to reveal colon motility patterns.

Ontario Institute for Cancer Research   2021.05-08
Position: Computational Biology Research Intern | Supervisor: Dr. Shraddha Pai
Collaborators: Dr. Gary Bader, Dr. Duncan Forster
Worked on *Extending Similarity Network-Based Classifiers to Non-Coding Genome and Deep Learning* as the first author.
- Developed a deep-learning patient classifier with an accuracy of 88% using graph attention networks to predict clinical outcomes of patients with cancer.
- Conducted research on model explainability and used SHAP values to explain feature importance for patient outcome prediction.

University of Waterloo   2020.07-12
Position: Researcher | Supervisor: Dr. George Shaker
Worked on *Deep Learning-Based In-Cabin Monitoring and Vehicle Safety System Using a 4D Imaging Radar Sensor* .
- Built an ML model with high precision (0.90) and recall (0.95) for in-car occupant detection using 4D MIMO radar.
- Designed an ML data pipeline for multi-label classification with big data, including data pre-processing, model training & testing and performance evaluation.
- Analyzed and benchmarked different neural network architectures, including 3D-CNN, Long Short-term Memory (LSTM) and Temporal Convolution Network

Nuralogix Corporation (in collaboration with University of Toronto)   2021.01-07
Position: Machine Learning Research Intern | Supervisor: Dr. Kang Lee
Worked on *Deep Learning-Based Facial Blood Pressure Estimation*.
- Designed and developed a 3D-CNN computer vision model with an ensemble learning approach using facial videos; improving blood pressure estimation accuracy by 7%
- Proposed and implemented multiple model explainability methods, including permutation feature importance and saliency map

University of Guelph - Centre for Biodiversity Genomics   2019.01-04, 08-12
Position: Bioinformatics Research Intern | Worked on *ML model for taxonomic classification*.
- Developed a multi-class ML model for taxonomic classification on nucleotide & amino acid sequences that achieved 96% accuracy
- Performed error analysis and visualized key findings using data visualization tools

Grad Course Projects

Automated Clinical Note Generation from Doctor-Patient Conversations using Large Language Models   Winter 2022
Our team developed a top-performing large-language-model (LLM) solution for automating clinical note generation from physician-patient dialogues, thereby demonstrating the potential of LLMs in improving healthcare documentation and communication.

Exploring Model Compression Techniques for Deep Learning based Image Compression Models   Winter 2022
Our team implemented model compression techniques, such as pruning and quantization, for image compression algorithms to mitigate high data transfer and computation costs, and compared the performance of compressed models to their original models.

Analyzing Participant Engagement in Mental Health Studies Through Data Collection Processes   Fall 2022
Our team analyzed data collected in a mental health study, utilizing Exploratory Data Analysis, visualizations and statistical testing. Based on the insights gained from this analysis, we gave recommendations aimed at enhancing user engagement.

Grants	2023, 2024: Vector Research Grant – $4,000 This grant is provided for top-performing graduate students who conduct research in AI at Vector Institute.
Fellowships & Fundings	2022-2024: University Of Toronto Tuition Fellowship $10,000 2022-2024: University Of Toronto Top-up Fellowship $4,000 2022-2023: University of Toronto Fellowship (Arts and Science) – $10,000 2020: BioTalent Canada Work Placement Program Funding – $7,500
Awards & Achievements	2023: Google CS Research Mentorship Program Recipient 2022: Vector Scholarship in Artificial Intelligence – $17,500 2022: Ontario Graduate Scholarship – $15,000 2020, 2021, 2022: Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award (USRA) – $6,500 x 3 2019, 2020: President's Research Award – $1,500 x 2 2018: President's Scholarship of Distinction – $2,000

Community Involvement

Machine Learning for Health 2024, *Reviewer (4 papers)*	2024.09
Machine Learning in Computational Biology 2024, *Reviewer (5 papers)*	2024.07
International Conference on AI & Human-Computer Interaction 2024, *Reviewer (1 paper)*	2024.07
ICLR 2024 Workshop on Learning from Time Series For Health, *Reviewer (2 papers)*	2023.02
International Conference on AI & Human-Computer Interaction 2023, *Reviewer (1 paper)*	2023.09
UofT Women in STEM, *Advisor*	2023
HerCode, *Mentor, Instructor*	2023
BioTEC Conference, *Event Organizer Lead*	2022-2023
UWaterloo Engineering Student Society Resume Critique, *Mentor*	2021.01-03
UWaterloo Women in Engineering, *Ambassador*	2017-2022

Skills

Programming Languages: Python, MATLAB, R, C, C++, C#, SAS, SQL, JavaScript, Java, HTML/CSS, Excel & VBA
Software Development Tools: Git, VS Code, AWS, Jupyter Notebook, Domino MLOps, Confluence, Jira, Docker, Bash
Software Libraries: PyTorch, Tensorflow, Keras, TFLearn, OpenCV, NumPy, Pandas, scikit-learn, scikit-image, Matplotlib, Scipy, seaborn
Languages: English, Mandarin