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 - Present
Supervisor: Dr. Quaid Morris & Dr. Justin Jee
Research Areas: Machine Learning, LLMs, Precision Oncology, Time Series

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, Precision Medicine, Time Series, Causality

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

T. Behrouzi *, **J. Yu** *, R. Yang, A. Boch, A. Goldenberg, S. Goodday and S. Friend. *Maternal health Aggregated Trends Can Mislead: The power of N-of-1 Level Wearable Data Analysis for Personalized Pregnancy Monitoring.*, [Journal Article], JMIR Formative Research, 2026.

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.*, [Proceedings Paper], Machine Learning for Health (ML4H) Symposium, Vancouver, Canada, 2025.

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, 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, 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, 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, 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], 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] IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science, Denver, CO, USA, 2022.

Under Review

J. Yu *, M. Darmofal, M. Waters, J. Choy, T. N. Tran, C. Fu, L. Morales, K. U, R. Levine, N. Schultz, M. Berger, Q. Morris, J. Jee, *Generalist large language models complement tailor-made predictors for tumor genomics interpretation.*, [Preprint], 2026. under review at *Nature Communications*

C. Fu, **J. Yu**, R. Shi, M. Waters, , A. Arfe, N. Schultz, J. Jee, Q. Morris, *Routine Clinical Time Series Predict Cancer Adverse Events.*, 2026. under review at *Cancer Cell*

P. Chuang, T. Behrouzi, **J. Yu**, R. Yang, K. Stemmler, S. M. Goodday, S. H. Friend, A. Goldenberg, *CodeWearBook: Retrieval-Grounded Natural-Language Querying of Wearable Health Datasets Using Large Language Models.*, 2026. under review at *JAMIA Open*

*First author

Abstract

C. Fu, **J. Yu**, R. Shi, M. Waters, A. Arfe, N. Schultz, J. Jee, Q. Morris, *Routine Clinical Time Series Predict Cancer Adverse Events.*, Annual International Conference on Intelligent Systems for Molecular Biology 2026 (ISMB'2026, Best Oral Presentation)

J. Yu *, M. Darmofal, M. Waters, J. Choy, T. N. Tran, C. Fu, L. Morales, K. U, R. Levine, N. Schultz, M. Berger, Q. Morris, J. Jee, *Large language models for tumor genomic interpretation.*, American Association for Cancer Research Annual Meeting 2026 & Annual International Conference on Intelligent Systems for Molecular Biology 2026 (AACR'2026, ISMB'2026)

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).

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. Quaid Morris, Dr. Justin Jee

- Contributing to a transformer-based foundation model pretrained on 99.4M longitudinal lab/vital measurements from 224,047 cancer patients to predict treatment toxicities and mortality (NeurIPS TS4H 2025).
- Building an LLM-based pipeline mapping unstructured pathology reports to standardized OncoTree codes, generating 158,000+ cancer diagnoses across 112,000 patients for cohort-scale model training.
- Benchmarking zero-shot LLMs against supervised baselines on cancer-type, somatic-vs-CHIP, and variant-oncogenicity classification (100,000+ samples, 19 institutions); LLM–specialist ensembles raised cross-institution accuracy up to 55%.

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

Graduate researcher:

- Designed the agentic workflows to be adaptable to new data sources and contexts while ensuring data privacy and security by restricting LLM access to raw data through controlled SQL and Python execution layers
- Conducted a study on exploring the risks of using wearable data to inform pregnant individuals within the context of ongoing digital health research
- Built a graph neural network to forecast pregnancy symptoms such as edema and fatigue with wearables feature interpretability.
- Proposed the concept of individualized delivery readiness and developed ML forecasting models.
- 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.
- 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: Undergraduate 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: Undergraduate 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 preprocessing, 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

Grants

2025: Society of Memorial Sloan Kettering (MSK) Grant – $100,000
This grant is to support MSK researchers who are expanding access to pioneering clinical trials and accelerating the development of lifesaving cancer therapies.

2023, 2024: Vector Research Grant – $4,000 × 2
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
2021: BioTalent Canada Work Placement Program Funding – $7,500

Awards & Achievements

2025: American Association for Cancer Research (AACR) Assoiate Member
2025: American Society of Clinical Oncology (ASCO) Member
2023: Google CS Research Mentorship Program Recipient
2022: Vector Scholarship in Artificial Intelligence – $17,500
2022: Ontario Graduate Scholarship – $15,000
2019, 2020: President's Research Award – $1,500 × 2
2018: President's Scholarship of Distinction – $2,000

Academic Service

PAN-Asian Symposium on Data and AI Collaborations in Biomedicine, *Student organizer*	2026
Knowledge Discovery in Databases (KDD) AI for Genomics, *Reviewer*	2026
Discovery Oncology, *Reviewer*	2026
Nature Cancer, *Reviewer*	2025
JCO Precision Oncology, *Reviewer*	2025, 2026
BMC Bioinformatics, *Reviewer*	2025
Heliyon, *Reviewer*	2025, 2026
PeerJ Computer Science, *Reviewer*	2025, 2026
NeurIPS 2025 Workshop on Learning from Time Series For Health, *Reviewer*	2025.09
Machine Learning in Computational Biology, *Reviewer*	2024, 2025
6th & 7th Annual Conference on Health, Inference, and Learning, *Reviewer*	2025, 2026
Machine Learning for Health, *Reviewer*	2024
Machine Learning in Computational Biology, *Reviewer*	2024, 2025
International Conference on AI & Human-Computer Interaction, *Reviewer*	2023, 2024
ICLR 2024 Workshop on Learning from Time Series For Health, *Reviewer*	2023.02
BioTEC Conference, *Event Organizer Lead*	2022-2023

Student Supervising and Mentorship

Student supervision during Master's and PhD: Emily Valera, Phoebe Chuang

UofT Women in STEM, *Advisor*	2023
HerCode, *Mentor, Instructor*	2023
UWaterloo Engineering Student Society Resume Critique, *Mentor*	2021.01-03
UWaterloo Women in Engineering, *Ambassador*	2017-2022