

Young Sang Choi

young.sang.choi@columbia.edu | youngsangchoi.github.io

EDUCATION

Columbia University in the City of New York Graduate School of Arts and Sciences
Doctoral Program in Biomedical Informatics, Fall 2023 - Current.
Advisors: Dr. Shalmali Joshi and Dr. Pierre Elias.
Master of Arts in Biomedical Informatics, 2024.

New York, NY
GPA: 4.05

Cornell University Cornell Tech
Master of Science in Information Systems, Concentration in Health Tech, 2019.
Advisor: Dr. Deborah Estrin.
Merit Scholarship (Academic Years of 2017 - 2019).

New York, NY

New York University College of Arts and Science
Bachelor of Arts in Mathematics, Minors in Computer Science and Creative Writing, 2017.

New York, NY

PEER-REVIEWED PUBLICATIONS

* denotes equal contribution.

Young Sang Choi*, Vincent Jeanselme*, Pierre Elias, and Shalmali Joshi. "ICYM2I: The Illusion of Multimodal Informativeness under Missingness." *International Conference on Learning Representations (ICLR)*, 2026.

Young Sang Choi, Shalmali Joshi, Linyuan Jing, and Pierre Elias. "Does Multimodality Help in Deep Learning-Based Structural Heart Disease Detection?" *Medical Imaging with Deep Learning (MIDL), Short Paper Track*, 2024.

Yong-Yeon Jo, **Young Sang Choi**, Jong-Hwan Jang, and Joon-myoungh Kwon. "ECGT2T: Towards Synthesizing Twelve-Lead Electrocardiograms from Two Asynchronous Leads." *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023.

Young Sang Choi, Jieun Oh, Seonhui Ahn, Yul Hwangbo, and Jinho Choi. "Automated Pulmonary Function Measurements from Preoperative CT Scans with Deep Learning." *Proceedings of the 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, IEEE, 2022. **(Oral Presentation)**

Yong-Yeon Jo*, **Young Sang Choi***, Hyun Woo Park, Jae Hyeok Lee, Hyojung Jung, Hyo-Eun Kim, Kyounglan Ko, Chan Wha Lee, Hyo Soung Cha, and Yul Hwangbo. "Impact of image compression on deep learning-based mammogram classification." *Scientific Reports* 11, no. 1 (2021): 1-9.

PREPRINTS & TECHNICAL REPORTS

* denotes equal contribution.

Chao Pang*, Vincent Jeanselme*, **Young Sang Choi**, Xinzhuo Jiang, Zilin Jing, Aparajita Kashyap, Yuta Kobayashi, Yanwei Li, Florent Pollet, Karthik Natarajan, and Shalmali Joshi. 2025. "FoMoH: A Clinically Meaningful Foundation Model Evaluation for Structured Electronic Health Records." *arXiv*. (2025).

Harry Reyes Nieva, Aparajita Kashyap, Erica A. Voss, Anna Ostropolets, Adit Anand, Mert Ketenci, Frank J. DeFalco **Young Sang Choi**, Yanwei Li, Monica N. Allen, Stephanie A. Guang, Karthik Natarajan, Patrick Ryan, Noémie Elhadad, "The Impact of Evolving Endometriosis Guidelines on Diagnosis and Observational Health Studies." *medRxiv* (2024).

CLINICAL ABSTRACTS AND SOFTWARE DEMOS

Matthew B.A. McDermott, Aleksia Kolo, Chao Pang, Edward Choi, Ethan Steinberg, Hyewon Jeong, Jack Gallifant, Jason Alan Fries, Jeffrey N. Chiang, Jungwoo Oh, Justin Xu, Kamilė Stankevičiūtė, Kiril Vadimovic Klein, Mikkel Frøelund Odgaard, Nassim Oufattole, Patrick Rockenschaub, Pawel Renc, Robin van de Water, Shalmali Joshi, Simon Austin Lee, Teya Bergamaschi, Tom Pollard, Vincent Jeanselme, Nigam Shah, Michael Wornow, Aparajita Kashyap, Xinzhou Jiang, Yanwei Li, **Young Sang Choi**, Yuta Kobayashi, Ryan King, "MEDS Decentralized, Extensible Validation (MEDS-DEV) Benchmark: Establishing Reproducibility and Comparability in ML for Health," Presented at: ML4H Demo Track, December 16, 2024, Vancouver, Canada.

Young Sang Choi, Shalmali Joshi, and Pierre Elias. "Multimodal Machine Learning for Structural Heart Disease Detection." Presented at: AMIA 2024 Annual Symposium, November 9-13, 2024, San Francisco, CA, United States of America.

Bella Mehta, **Young Sang Choi**, J. Alex Gibbons, et al. “Ultrasonography May Improve Osteoarthritis Pain Prediction: An Artificial Intelligence Approach.” Presented at: EULAR 2023 Annual European Congress of Rheumatology, May 31-June 3, 2023, Milan, Italy.

Young Sang Choi, Jieun Oh, Seonhui Ahn, Yul Hwangbo, and Jinho Choi. “Measuring FEV₁ from Preoperative Chest Computed Tomography Scans using Deep Learning.” Presented at: Radiological Society of North America 2022 Scientific Assembly and Annual Meeting, November 27–December 1, 2022, Chicago, IL, United States of America.

Yong-Yeon Jo, **Young Sang Choi**, Hyun Woo Park, Jae Hyeok Lee, Hyojung Jung, Hyo-Eun Kim, Kyounglan Ko, Chan Wha Lee, Hyo Soung Cha, and Yul Hwangbo. “Impact of Image Compression on Deep Learning based Classification Performance: Mammogram.” Presented at: Radiological Society of North America 2020 Scientific Assembly and Annual Meeting, November 29–December 5, 2020, Virtual.

EXPERIENCE

Columbia University

Graduate Research Assistant, reAIM Lab and CRADLE Lab

New York, NY

Sep. 2023 - Current

- **Learning from Irregular Multimodal Healthcare Event Streams**

Currently developing framework for learning from irregular, multimodal healthcare event streams to address challenges in medical data such as missing data, unaligned input-label pairs, and asynchronous sampling. Proposed framework enables event type, timing, and value predictions from input and label streams.

- **Multimodal Models for Healthcare with Incomplete Data**

Investigating the gap between clinical data acquisition and downstream modeling tasks driven by factors such as diagnostic stewardship and cost-benefit tradeoffs. Proposing a taxonomy of clinical multimodal data generation and using Partial Information Decomposition to quantify task-dependent interactions across diagnostic and prognostic tasks and leveraging inverse probability weighting-based methods for adjusting for modality missingness. Benchmarking implications of modality availability on model performance and information gain on synthetic and real-world healthcare datasets in radiology, intensive care, and cardiology. Preprint in submission.

- **Evaluation of Electronic Health Record-Based Foundation Models**

Benchmarking learning electronic health record (EHR)-based foundation models (FMs) for patient phenotyping and outcome prediction. Working with the cross-institutional Medical Event Data Standard (MEDS) effort for standardization and harmonization strategies for clinical data, processing pipelines, model training, and evaluation of EHR foundation models. Work presented as a software demo at ML4H and full paper preprint in submission.

Student Research Worker, Division of Cardiology

Jun. 2023 - Aug. 2023

- **Multimodal Learning for Cardiology**

Built and evaluated a system for the detection of structural heart disease using common clinical modalities in ECGs, chest X-rays, and structured EHR data in place of underutilized transthoracic echocardiograms. Work presented as a first author short paper in Medical Imaging with Deep Learning (MIDL) 2024.

National Cancer Center, Korea

Technical Research Personnel, Healthcare AI Team

Goyang-si, Republic of Korea

Dec. 2019 - Dec. 2022

- **Measuring Preoperative Pulmonary Function from Chest CT Scans**

Collected pre- and post-operative CT pairs with pulmonary function test values. Utilized multi-label CNN-RNN models to measure pre-operative lung function, and improved performance with self-supervised feature extractors and transformer-based sequence regressors. Measured post-operative pulmonary decline with video similarity networks. Work presented as a first author publication and oral presentation at IEEE-EMBS BHI 2022.

- **Evaluating the Impact of Image Compression on Mammogram Classification**

Evaluated the impact of lossy JPEG 2000-based image compression on downstream deep learning-based classifiers. Work presented as a co-first author publication in *Scientific Reports*.

- **Estimating Pain Symptoms from Knee Radiographs and Ultrasound Features**

Collaboration with Dr. Bella Mehta at the Hospital for Special Surgery in New York. Measured KOOS pain and KL-scores from knee x-rays and ultrasound features in end-stage osteoarthritis subjects using multi-modal classification networks. Work presented as a clinical abstract in EULAR 2023.

- **Opportunistic Osteoporosis Screening with Abdominal CT Scans**

Collaboration with the BISPL lab at Korea Advanced Institute of Science and Technology (KAIST). Collected a dataset of 276 CT-DXA pairs. Designed system for automated opportunistic screening of osteoporosis using maximum intensity projections to detect lumbar vertebrae with DenseNet-based image regressors.

- **Thyroid Incidentaloma Clinical Decision Support System**
Built a multi-institutional thyroid incidentaloma dataset as part of the Dr. Answer 2.0 project. Extracted screening chest CTs and diagnosis information from 1,380 subjects with and without benign and malignant thyroid nodules from NCC. Assisted in the thyroid segmentation and nodule object detection annotation Collaborated with Chungbuk National University Hospital, Monitor Corporation, and Cheongju St. Mary's Hospital to train and validate a clinical decision support system for flagging potential thyroid cancer during lung cancer screening.
- **Multi-Task Learning for Mammography Applications**
Collected 7,500 screening mammograms and parsed radiology reports for multi-label classification and enrolling additional study subjects. Performed empirical study on which tasks enable positive transfer and robustness to distribution shift.
- **Positron Emission Tomography Classification Dataset Collection**
Collaboration between six South Korean hospitals to create a multi-modal dataset of positron emission tomography scans and electronic health record data for a three-class classification problem with normal, lymphoma, and lung cancer subjects. Generated frontal and sagittal maximum intensity projections for annotation. Cleaned free-form clinical notes to JSON files.
- **Pathology Classification Dataset Collection**
Collected data from 4,000 subjects. Built a patch generator script with multiprocessing. Trained patch-level image classifiers as baseline models. Hosted the Korea Health Datathon 2020 competition with Naver Cloud Platform and Konyang University Hospital.

Cornell University

Research Intern, Small Data Lab at Cornell Tech

New York, NY

Dec. 2018 - Jan. 2019

- **YouTube Viewing History Pipeline**

Built a pre-processing pipeline for Google Takeout data for use in a YouTube watch study using Python and a PostgreSQL database.

Memorial Sloan Kettering Cancer Center

Graduate Student Intern, Healthcare Informatics Innovation Lab

New York, NY

Jun. 2018 - Aug. 2018

- **RFID-Based Interaction Management**

Built a rule-based model to flag and visualize patient-provider interactions. This tool uses Python and real-time location system data for re-identification. Data were sourced from RFID tags.

- **Augmented Reality for Early Ambulation**

Built an augmented reality artwork tour iOS app to motivate surgery patients to walk during post-operative care. This app, built using Swift and ARKit, used an object detection algorithm to trigger pre-recorded audio files from artists and curators.

PROFESSIONAL SERVICE

Journal Reviewer:

Transactions on Machine Learning Research (TMLR).

Conference Reviewer:

Machine Learning for Healthcare (MLHC): 2024, 2025.

International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI): 2025.

Machine Learning for Health Symposium (ML4H): 2025.

American Medical Informatics Association Annual Symposium (AMIA): 2025.

MENTORSHIP

Praveena Ratnavel, UCLA Undergraduate Student → UCLA Bioinformatics Doctoral Program

Summer 2023

Columbia DBMI Summer Research Program. Project Title: "Automated Chest X-Ray Diagnosis with Deep Learning."

TEACHING

Columbia University in the City of New York

Teaching Assistant, Machine Learning for Health and Medicine (BINF 4002)

Spring 2025

Teaching Assistant, Advanced Machine Learning for Health and Medicine (BINF 4008 / COMS 4995)

Fall 2023

Cornell University

Teaching Assistant, The Product Studio at Cornell Tech (TECH 5900)

Fall 2018