Department of Biostatistics and Medical Informatics Seminar

Yin Li, Ph.D.

Associate Professor
Biostatistics & Medical Informatics
University of Wisconsin–Madison
https://yinli-vision.github.io/

Friday, October 9, 2026
12:00-1:00 pm – Morgridge Hall Seminar Room 7560

Zoom Meeting: 966 3372 9112
https://uwmadison.zoom.us/j/96633729112?pwd=tHFc9i1dAAgmXe05uWtw8wXBIQZxGB.1 **Passcode: 621125**

From Medical Video Analysis to Multimodal Intelligence

Abstract: Video is becoming an increasingly important form of data in healthcare. Cameras are now deployed in operating rooms, intensive care units, exam rooms, and other clinical settings, generating a growing volume of medical video that captures rich information about disease signals, care delivery, and patients' responses. Yet much of this video is collected primarily for documentation or quality assurance, and its potential to support clinical care and medical research remains largely unexplored. At the same time, recent advances in multimodal large language models (MLLMs) offer new opportunities for medical video analytics. By jointly processing video, audio, text, and other modalities, these models provide a flexible interface for understanding complex clinical events and supporting a broad range of analysis tasks.

In this talk, I will first introduce our prior work on extracting clinically relevant signals from high-dimensional video data. We have developed specialized deep learning models to detect medical tools, recognize care activities, characterize patients' responses, and quantify diagnostic signals. Building on these efforts, I will then present our current and ongoing work toward generalist MLLMs for streaming video understanding, with the goal of supporting diverse medical video analysis tasks within a unified framework. I will highlight several key directions, including integrating multimodal signals beyond video and text, developing long-term memory for continuous video streams, and enabling richer interaction between AI models and human users. Collectively, these efforts aim to move medical video analysis beyond narrowly tailored perception models toward multimodal AI systems that can continuously observe, reason about, and interact within complex clinical environments.

Dr. Yin Li is a tenured Associate Professor in the Department of Biostatistics and Medical Informatics at the University of Wisconsin-Madison, with affiliate appointments in the Departments of Computer Sciences and Electrical and Computer Engineering. His research focuses on computer vision and multimodal machine learning, with an emphasis on developing AI systems that understand images, videos, and other multimodal data, and on translating these methods to applications in digital health. He is a recipient of the NSF CAREER Award and the Google Research Scholar Award. He has served as Area Chairs for major conferences in computer vision and machine learning, including CVPR, ICCV, ECCV, NeurIPS, ICLR, and IJCAI. His work has received Best Student Paper Awards at MobiHealth 2014 and IEEE Face and Gesture 2015, as well as a Best Demo nomination at ECCV 2020. His research has also been featured in outlets including MIT Technology Review, New Scientist, BBC, and Forbes.

School of Medicine and Public Health
UNIVERSITY OF WISCONSIN-MADISON