



DISSERTATION DEFENSE



Meera Krishnamoorthy

Towards Scalable and Stable Machine Learning in Clinical Contexts: Addressing Computational Efficiency and Dataset Shift

Thursday, May 15, 2025

1:00pm – 3:00pm

3725 Beyster / Hybrid - [Zoom](#)

ABSTRACT: Clinical machine learning (ML) models hold promise for improving patient outcomes, but realizing this potential requires models that are both computationally efficient and robust to dataset shift. Efficiency enables deployment on bedside devices, enhancing privacy and reliability. Robustness to dataset shift is equally critical, as clinical ML models must perform predictably over time. In this thesis defense, I will address both challenges, proposing novel approaches that improve model efficiency and performance under dataset shift.

With respect to computational efficiency, I will examine two domains. First, in genomic sequence classification, I will demonstrate standard approaches are inefficient because they rely on large reference databases. To address this, I will introduce an ML approach that does not depend on a reference database at inference time, enabling accurate and memory-efficient genomic classification. Next, I will address computational efficiency in multiple instance learning (MIL) for large medical image analysis. I will show that while transformer-based approaches offer high accuracy, they are computationally costly. I will propose a lightweight positional encoding wrapper that improves standard MIL accuracy without increasing computational overhead.

With respect to dataset shift, I will demonstrate how conventional model selection methods often favor models that rely on unstable correlations, leading to poor generalization. I will present a novel model selection strategy that selects models with stronger generalization performance over time. Finally, I will explore dataset shift in survival analysis, in which the probability of censoring can change over time. I will present a novel approach to leveraging censored data during training, designed to improve time-to-event prediction for individuals similar to those censored in the training data but uncensored in the test data.

Overall, I will present methods to enhance the efficiency and stability of clinical ML models. These contributions aim to support the development of models that are easier to integrate into clinical workflows and more reliable in real-world healthcare settings.

CHAIR: Prof. Jenna Wiens