STEVENS INSTITUTE OF TECHNOLOGY 1870

Ph.D. DISSERTATION DEFENSE

Candidate:	Adedolapo Aishat Toye
Degree:	Doctor of Philosophy
School/Department:	Charles V. Schaefer School of Engineering and Science (SES) / Computer Science
Date:	Friday, April 24th, 2026
Time:	3:30PM
Location:	Gateway South, Room 324 (Health and AI Lab)
Title:	Advancing Health Data Analysis: Methodological Approaches for Addressing Data Limitations and Informing Clinical Decision Making
Chairperson:	Dr. Samantha Kleinberg, Department of Computer Science
Committee Members:	Dr. Feng Liu, Department of Systems and Enterprises
	Dr. Yue Ning, Department of Computer Science
	Dr. Nikhil Muralidhar, Department of Computer Science
	Dr. Shaoyi Huang, Department of Computer Science

ABSTRACT

The increasing generation of health data presents significant opportunities for advancing healthcare research. However, leveraging this potential presents challenges for both researchers, who require access to quality data for model development, and clinicians, who rely on timely insights to guide interventions and improve patient care. A major challenge is limited access to health data due to privacy concerns, which restricts data sharing. While simulation provides a solution to this, existing simulation approaches either overestimate model performance or are black boxes, making them unsuitable for ablation studies. Another challenge is missing data, as imputation quality can significantly impact downstream machine learning (ML) tasks. However, many imputation methods assume data are missing completely at random (MCAR), which is often not the case in health data, where missingness may be missing at random (MAR) or not missing at random (NMAR).

On the clinical side, ground truth sparsity, such as the lack of frequent assessments of patient consciousness in intensive care units (ICUs), limits the ability to monitor patients continuously, which is important for timely interventions. In addition, the limited understanding of the underlying causal relationship between exposures and outcomes makes it challenging to develop effective clinical guidelines.

This thesis addresses these challenges through the following contributions. First, we develop methods for simulating realistic time series data with similar performance to real data on ML tasks, by augmenting simulated data with data properties (e.g., error and nonstationarity). Second, we develop an approach for robustly evaluating missing data imputation methods under real-world scenarios across multiple missing mechanisms (MCAR, MAR, and NMAR) and varying percentages. Third, we develop a framework for the hourly classification of consciousness in ICU patients using physiological signals, enabling more frequent monitoring. Finally, we investigate the

relationship between neonatal glucose exposure and early childhood neurodevelopmental outcomes and assess the potential causal effects of hyperglycemia to help inform clinical guidelines. Together, these contributions aim to advance health data analysis by addressing key data limitations and informing clinical decision-making.