STEVENS
INSTITUTE OF TECHNOLOGY
1870

Ph.D. DISSERTATION DEFENSE

Candidate: Louis Adedapo Gomez
Degree: Doctor of Philosophy
School/Department: Charles V. Schaefer, Jr. School of Engineering and Science/ Department of Computer Science
Date: Monday, March 25th, 2024
Time/Location: 3 pm / Gateway North Room 303
Title: Learning Models from Health Time Series: Overcoming challenges of data availability and sufficiency

Chairperson: **Dr. Samantha Kleinberg, Department of Computer Science**

Committee Members: Dr. Yue Ning, Department of Computer Science
Dr. Ping Wang, Department of Computer Science
Dr. Nikhil Muralidhar, Department of Computer Science
Dr. Johannes Weickenmeier, Department of Mechanical Engineering

ABSTRACT

Large amounts of healthcare data are being generated through medical records and patient-generated data; however, the unique nature of these health datasets creates computational challenges that make machine learning and causal inference difficult. As data collected from individual patients solely depend on patient care, it creates a challenging scenario where different patients can have different variables measured, which for causal discovery leads to confounding and lack of generalization. Further, while simulation addresses a core challenge of evaluating algorithms as it has ground truth and can be shared without privacy concerns, current models either provide overly optimistic estimates of performance on machine learning tasks or do not allow for ablation studies into how data properties affect performance due to their black box nature, limiting their wide application in health.

In this thesis, I tackle these challenges by developing new methods for (i) data simulation that generates simulated data with similar performance to real data and allows for greater control over the kinds of data properties encoded in them [1] and (ii) learning causal models when we have multiple time series datasets with partially overlapping variable sets [2]. In addition, this thesis also focuses on addressing application-specific problems by helping clinicians better monitor patient health by introducing new techniques for classifying consciousness using only continuously recorded physiological signals [3] and helping individuals with type 2 diabetes better track their eating occasions by adapting a simulation-based approach to meal detection [4].

STEVENS
INSTITUTE OF TECHNOLOGY
1870

REFERENCE:

[1] **Gomez, L. A**., Toye, A. A., Hum, R. S., & Kleinberg, S. (2023). Simulating Realistic Continuous Glucose Monitor Time Series By Data Augmentation. *Journal of Diabetes Science and Technology*, 19322968231181138.

[2] **Gomez, L. A**., Claassen, J., & Kleinberg, S. Causal Inference for Time Series Datasets with Partially Overlapping Variables (In Review at IJCAI 2024)

[3] **Gomez, L. A**., Shen, Q., Doyle, K., Vrosgou, A., Velazquez, A., Megjhani, M., ... & Kleinberg, S. (2023). Classification of level of consciousness in a neurological ICU using physiological data. *Neurocritical care*, *38*(1), 118-128.

[4] Popp, C. J., Wang, C., Hoover, A., **Gomez, L. A**., Curran, M., St-Jules, D. E., ... & Kleinberg, S. (2023). Objective determination of eating occasion timing (OREO): Combining self-report, wrist motion, and continuous glucose monitoring to detect eating occasions in adults with pre-diabetes and obesity. *Journal of Diabetes Science and Technology*, 19322968231197205.