Stanford MEDICINE | Department of Biomedical Data Science

Beyond Prediction: Causal Validity in ML-Driven Drug Discovery and Health Monitoring

Emily Fox
Professor of Statistics and Computer Science at Stanford University and Chief Technical Advisor
February 5th, 2026
1:30PM-2:50PM
****R358****

Abstract:

While machine learning shows tremendous promise for extracting disease-relevant signals from rich, multimodal clinical and cellular datasets, predictive accuracy alone does not yield decision-ready science—establishing causality is critical. This talk traces a path from pattern-finding to causal insight in two parts.

First, I will present our approach at insitro to therapeutic target discovery, which builds target conviction by triangulating two complementary lenses. On the clinical side, we leverage human genetics as experiments of nature to identify causal genetic drivers, exploiting ML-derived precision phenotypes imputed at scale from rich clinical data modalities. On the cellular side, we run perturbation screens in disease-relevant models to measure phenotypic consequences across high-content modalities including microscopy and omics. Because the biases of these ClinML and CellML lenses differ, cross-lens concordance holds promise in raising confidence in causal targets. I'll highlight key sources of bias and how modeling choices and assay design chip away at them to refine target prioritization, including our work on multimodal embeddings and sparse autoencoders for aligning and interrogating ML-derived phenotypes.

Second, I'll shift to my Stanford work on leveraging mechanistic knowledge as causal inductive bias. When training blackbox models on observational data, high predictive accuracy can coexist with low causal validity—the counterfactual simulation of a glucose trajectory may respond nonsensically to a simulated insulin intervention. We address this by encoding domain knowledge about sets of treatment effect rankings into a causal loss that, combined with standard predictive loss, biases learning toward causally valid models. Using type 1 diabetes as a case study, I'll show how hybrid models—combining mechanistic ODE dynamics with flexible neural network components—trained with our proposed hybrid loss achieve both state-of-the-art predictive performance and causal validity for the challenging task of modeling glucose dynamics post-exercise.

Reading list:

- Zou, B.J., Levine, M.E., Zaharieva, D.P., Johari, R. & Fox, E.. (2024). Hybrid^2 Neural ODE Causal Modeling and an Application to Glycemic Response. Proceedings of the 41st International Conference on Machine Learning, in Proceedings of Machine Learning Research 235:62934-62963 Available from https://proceedings.mlr.press/v235/zou24b.html