Bridging Public Health with Clinical Decisions from a Data Centric Perspective

Jiaming Cui

Virginia Tech
jiamingcui@vt.edu

Abstract

Public health and clinical decisions are intertwined. Public health crises place a high burden on healthcare facilities, forcing them to make decisions such as maintaining quality verses treating more people. Meanwhile, sub-optimal clinical decisions also cause downstream effects on communities. For example, discharging patients too early may increase disease spread. Motivated by this, we bring a data-centric perspective to bridge clinical decisions within the context of infectious diseases for public health. This work addresses multiple challenges arising from effectively utilizing rich clinical datasets and issues stemming from the complexity of disease spread dynamics in healthcare facilities. We will cover methods developed to address these challenges with better designed models to optimize disease surveillance and control policies and new techniques for end-to-end learning with mechanistic models. We will conclude by discussing emerging challenges and opportunities at the intersection of machine learning, scientific modeling, and clinical decision-making for computer scientists, epidemiologists, and computational biologists.

Better Modeling for Clinical Decisions

Healthcare-associated infections (HAIs) such as Methicillin-resistant *Staphylococcus aureus* (MRSA) and *Clostridioides difficile* (*C. difficile*) are infections acquired by patients during treatment in healthcare facilities such as hospitals and long-term nursing homes. They pose a significant threat to patient safety and burden the healthcare system with increased costs due to longer hospital stays and more expensive therapies. Note that HAIs can spread not only via direct patient-patient contacts but also through contaminated locations such as rooms or medical devices. Existing models that only account for patient contact networks are less suitable. Therefore, I propose utilizing environmental factor-mediated epidemiological models to better model the HAI spread dynamics. Here, the environmental factors refer to the pathogen load at locations (e.g., patient room, doctors' office, nurse station) contaminated by patients and healthcare workers. Such models could help solve the real-world HAI surveillance (Anand et al. 2024) and control (Cui et al. 2023, 2024) problems raised by clinicians.

ML for Better Modeling

Epidemiological models are widely used in infectious disease surveillance and control. For example, Ordinary Differential Equation (ODE) models were used for forecasting during the COVID-19 pandemic (Rodríguez et al. 2023), and agent-based models (ABMs) based on contact networks are used in HAI control (Cui et al. 2025). However, these models may suffer from suboptimal calibration and cannot leverage rich clinical datasets. Therefore, I proposed several hybrid frameworks that integrate epidemiological models with machine learning (ML) to achieve better calibration and modeling. Unlike methods that rely solely on one paradigm, such hybrid frameworks can leverage mechanistic models for interpretability while maintaining the flexibility of ML to incorporate rich electronic health records (EHRs). Experimental results show that such hybrid frameworks significantly outperform existing methods, including current clinical practices, marking a breakthrough in infectious disease surveillance and control in clinical practice.

ML for Better Decisions

Although deep neural networks have shown remarkable capabilities in clinical decision-making and public health, they encounter significant limitations in accounting for uncertainty: many deep neural networks can only provide point estimates while ignoring the uncertainty in their predictions. However, uncertainty quantification is crucial for encoding potential risks and inherent randomness to achieve robust decision-making (Kong et al. 2022). In addition, accurate prediction of clinical outcomes using Electronic Health Records (EHRs) is critical for early intervention, efficient resource allocation, and improved patient care. EHRs contain multimodal data, including both structured data and unstructured clinical notes that provide rich, context-specific information. Integrating these diverse modalities and leveraging all relevant available information could be helpful for more accurate prediction of clinical outcomes (Liu et al. 2024; Datta et al. 2025). Experimental results show that our framework achieves strong performance and outperforms all existing baselines and clinical practices, including established risk scoring systems. To the best of our knowledge, this is one of the first frameworks for healthcare prediction by combining unstructured clinical notes with structured data for improved clinical outcome prediction.

References

Anand, V.; Cui, J.; Heavey, J.; Vullikanti, A.; and Prakash, B. A. 2024. H2ABM: Heterogeneous agent-based model on hypergraphs to capture group interactions. In *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)*, 280–288. SIAM.

Cui, J.; Cho, S.; Kamruzzaman, M.; Bielskas, M.; Vullikanti, A.; and Prakash, B. A. 2023. Using spectral characterization to identify healthcare-associated infection (hai) patients for clinical contact precaution. *Scientific Reports*, 13(1): 16197.

Cui, J.; Heavey, J.; Klein, E.; Madden, G. R.; Sifri, C. D.; Vullikanti, A.; and Prakash, B. A. 2025. Identifying and forecasting importation and asymptomatic spreaders of multi-drug resistant organisms in hospital settings. *NPJ Digital Medicine*, 8(1): 147.

Cui, J.; Heavey, J.; Lin, L.; Klein, E. Y.; Madden, G. R.; Sifri, C. D.; Lewis, B.; Vullikanti, A. K.; and Prakash, B. A. 2024. Modeling relaxed policies for discontinuation of methicillin-resistant staphylococcus aureus contact precautions. *Infection Control & Hospital Epidemiology*, 45(7): 833–838.

Datta, R.; Cui, J.; Guan, Z.; Silwal, R.; Eby, J. C.; Madden, G.; and Vullikanti, A. 2025. Improving Hospital Risk Prediction with Knowledge-Augmented Multimodal EHR Modeling. *arXiv preprint arXiv:2508.01970*.

Kong, L.; Cui, J.; Zhuang, Y.; Feng, R.; Prakash, B. A.; and Zhang, C. 2022. End-to-end stochastic optimization with energy-based model. *Advances in neural information processing systems*, 35: 11341–11354.

Liu, H.; Xu, S.; Zhao, Z.; Kong, L.; Prabhakar Kamarthi, H.; Sasanur, A.; Sharma, M.; Cui, J.; Wen, Q.; Zhang, C.; et al. 2024. Time-mmd: Multi-domain multimodal dataset for time series analysis. *Advances in Neural Information Processing Systems*, 37: 77888–77933.

Rodríguez, A.; Cui, J.; Ramakrishnan, N.; Adhikari, B.; and Prakash, B. A. 2023. Einns: epidemiologically-informed neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 14453–14460.