

LEVERAGING SUPERVISED MACHINE LEARNING TO PREDICT HEALTH BEHAVIORS IN SPATIAL AGENT-BASED MODELS OF DISEASE SPREAD

Emma Marie Von Hoene
evonhoen@gmu.edu

INTRODUCTION AND BACKGROUND

As infectious diseases threaten global health, policymakers rely on epidemiological models, such as agent-based models (ABMs), to predict disease outcomes and improve public health response. ABMs capture the behavior and interactions among heterogeneous individuals or “agents” within their environment, making them ideal for simulating disease spread, given the key role individual behavior plays in disease outcomes. However, traditional ABMs of disease spread ignore or oversimplify health behaviors, such as mask-wearing or vaccination. This was one of several COVID-19 model shortcomings that led to inaccurate forecasts, highlighting the need to improve the representation of health behaviors in disease spread models [1].

Despite advancements in developing behavioral frameworks for ABMs of disease spread, particularly using theories like game theory and social contagion, these models often lack support from empirical data [2]. While theory is important, ABMs are designed to simulate disease spread in specific populations, and data is essential to parameterize, calibrate, and validate agent behaviors [1]. A potential approach for addressing this gap involves leveraging supervised machine learning (ML) algorithms that can incorporate both behavioral data and theory to predict individual behaviors. In various applications, ABMs have integrated supervised ML techniques to simulate agent decision-making. Given data sources like surveys, a supervised ML model can learn how a set of determinants, guided by theory or domain knowledge, are associated with a behavior. Then, agents can apply the trained ML model based on their characteristics and perceptions.

Using supervised ML to model agent behaviors is less common in ABMs of infectious diseases, possibly due to limited collaboration and data sharing between epidemiology and modeling communities. In one example, Karimi et al. [3] calculates vaccination and social distancing probabilities using logistic regression, informed by survey data and based on determinants from the Health Belief Model (HBM). In another study, Beerman et al. [4] used decision trees trained on various data sources to predict if an agent vaccinates based on characteristics. Such examples predict an agent’s health behavior as a one-time decision, overlooking the dynamic nature of health decisions. Further research is necessary to incorporate supervised ML in a way that allows for changes in individual behavior over time, considering social or environmental feedback. Efforts are also needed to validate emergent behaviors across space and time, and at different scales.

Therefore, the overarching goal of my Ph.D. is to develop and implement a framework that combines data-driven and theoretical approaches through leveraging supervised ML algorithms to empirically derive and implement dynamic individual health behaviors. This will involve creating a network-based ABM that focuses on the case study of yearly vaccination uptake behavior in Fairfax County, Virginia, USA, and its impact on the spread of COVID-19. To achieve this goal and address these challenges within the ABM community, I will answer the following research questions: 1) How do different supervised machine learning techniques compare in performance when predicting individual health behaviors within ABMs of infectious disease spread?, 2) Which public health theories are well-suited to guide the selection of determinants, and

how can these determinants be dynamically encoded into an ABM of infectious disease spread (e.g. perception of severity from the HBM)?, 3) What aspects of ABMs of infectious disease spread are important for empirically validating health behaviors, and how can the development and validation of these models be effectively communicated to demonstrate how they perform sufficiently for decision support? The contribution of my research aims to improve the representation of health behaviors in ABMs of infectious disease spread, ultimately offering the public and policymakers support with producing accurate predictions and intervention strategies for future outbreaks.

PRELIMINARY AND FUTURE WORK

Previous work consisted of developing a proof-of-concept that integrates a data-driven agent behavioral framework in a geospatial ABM of disease spread [5]. The framework is implemented to simulate COVID-19 spread and dynamic masking behavior among students at George Mason University. Using open survey data, the framework leverages empirically derived odds ratios through a binary logistic regression model. This model estimates the daily likelihood of an agent adopting masking behavior, considering an individual's characteristics and perception, specifically perceived vulnerability to the disease. When the framework is applied, the masking behaviors emerge based on agent characteristics and their evolving perception of the outbreak. Results reveal that the framework creates a feedback between the behaviors and the disease outcomes, where behaviors impact the disease outcomes, and the outcomes further impact behaviors. Despite promising results in preliminary work, there are limitations that will be addressed in my future research.

Future work involves expanding upon my existing ABM, which consists of transitioning to a network-based structure, focusing on Fairfax County at the census-tract level, and generating a synthetic population. This model will serve as an essential tool for addressing my dissertation research questions. To answer these questions, my objectives are: 1) assess the performance and efficiency of different supervised ML techniques for agent health behaviors in the scaled-up ABM of disease spread, 2) investigate available data and select and compare the predictive power of different determinants that are theorized to explain health behaviors, 3) validate the spatial patterns of behavior and disease outcomes that emerge from the model. The ultimate goal of my dissertation is to create a user-friendly framework that leverages ML algorithms that derives individual health behaviors from theory and data throughout an infectious disease simulation. It is intended that users have the flexibility to modify parameters such as population, determinants, and disease types, which promotes the usability of the model for various research applications related to infectious diseases.

ACKNOWLEDGMENTS

I am grateful to my advisor, X, for I also appreciate the X for funding this research.

REFERENCES

- [1] V. Iranzo and S. Pérez-González, "Epidemiological models and covid-19: a comparative view." *HPLS*, vol. 43, 2021.
- [2] T. Anderson and et al., "Health behavior in agent-based models of respiratory infectious disease spread: A systematic review (in revision)," *JASSS*.
- [3] E. Karimi and et al., "Effect of individual protective behaviors on influenza transmission: an agent-based model," *Health Care Manag Sci*, vol. 18, 2015.
- [4] J. Beerman and et al., "On the necessity of human decision-making errors to explain vaccination rates for covid-19: an agent-based modeling study," in *2023 ANNSIM*, 2023.
- [5] ... and et al., "Citation removed for double blind," in *6th ACM SIGSPATIAL GeoSpatial Simulation Workshop*, 2023.