Applying Human-Centered Data Science to Healthcare: Hyperlocal Modeling of COVID-19 Hospitalizations

Victoria Chui
victoria.chui@mail.utoronto.ca
University of Toronto
Toronto, Canada

Jessica Pater
jessica.pater@parkview.com
Parkview Research Center
Fort Wayne, Indiana, United States

Tammy Toscos
tammy.toscos@parkview.com
Parkview Research Center
Fort Wayne, Indiana, United States

Shion Guha
shion.guha@utoronto.ca
University of Toronto
Toronto, Canada
Parkview Research Center
Fort Wayne, IN, USA

ABSTRACT

Algorithms as a component of decision-making in healthcare are becoming increasingly prevalent and AI in healthcare has become a topic of mass consideration. However, pursuing these methods without a human-centered framework can lead to bias, thus incorporating discrimination on behalf of the algorithm upon implementation. By examining each step of the design process from a human-centered perspective and incorporating stakeholder motivations, algorithmic implementation can become vastly useful, and more accurately tailored to stakeholder needs. We examine previous work in healthcare executed with a human-centered design, to analyze the multiple frameworks which effectively create human-centered application, as extended to healthcare.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → *Life and medical sciences.*

KEYWORDS

human-centered data science, healthcare, participatory design, speculative design

ACM Reference Format:
Victoria Chui, Jessica Pater, Tammy Toscos, and Shion Guha. 2023. Applying Human-Centered Data Science to Healthcare: Hyperlocal Modeling of COVID-19 Hospitalizations. In *The 2023 ACM International Conference on Supporting Group Work (GROUP '23) Companion (GROUP '23), January 8–11, 2023, Hilton Head, SC, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3565967.3570979

1 INTRODUCTION

As the sheer volume of new and existing data, coupled with expanding computing power, continues to evolve in an increasingly technical world [7], data science methods are progressively permeating many public facing industries, including justice, labour, and healthcare [4]. Yet, implementation of data science in healthcare can lead to biased and incomprehensible results, often due to black-box features or a lack of stakeholder input [1]. SIGCHI research has shown incorporating stakeholder ideas and contributions throughout an algorithm's design process can satisfy the technical task at hand, while also maintaining a standardized motivation grounded in specific human needs as heard directly from the source [9]. This is the foundation of human-centered data science (HCDS). HCDS aims to approach computational challenges while considering social factors on behalf of stakeholders and those who face the challenge at hand [1, 9].

What does it mean to extend HCDS to healthcare and how can HCDS can assist and advance the field of healthcare, elevating techniques currently applied on the ground? Human-centered algorithm design (HCAD) has been proposed to bring human-centered design to the space of algorithmic creation [2]. In this extended abstract, we analyze an HCDS application in healthcare using the HCAD framework through a case study of COVID-19 modeling at Parkview Health, a community healthcare system based in Northeastern Indiana, USA. By looking at this case study through a human-centered framework, we aim to analyze the multiple foundations of HCDS that allow algorithmic development to directly involve stakeholder input, expand past assumptions intrinsically found in algorithmic design, and use situational-specific contexts and limitations to ensure maximum efficacy.

In this extended abstract, we present a position of how human-centered algorithmic creation can lead to a more nuanced, precision style of computational implementation in healthcare. By examining work through the HCAD framework, other healthcare systems can advance their own design processes to include more stakeholder input, thus incorporating a more human-centered strategy throughout algorithmic design and implementation.

2 BACKGROUND

HCAD is a practice which serves to bridge the gap between human-centered decision-making efforts and the technological implementation of algorithms [2] and emphasizes the need to combine HCAD with more traditional evaluation metrics used in data science, but

GROUP '23, January 8–11, 2023, Hilton Head, SC, USA

ACM ISBN 978-1-4503-9945-6/23/01.
https://doi.org/10.1145/3565967.3570979

again through a more human-centered lens. Albert et al. (2020) agree that evaluation metrics for machine learning implemented in healthcare settings should be specific to the statistical task at hand. The clinical relevance of each of the common “metrics of success” (ROC-AUC, F1 score, confusion matrix) is weighted differently depending on the needs of the stakeholders [3].

While algorithms in healthcare are exceeding effective and beneficial, there has been much concern over their deployment [5]. The use of statistical models for predictive, clustering, and other quantitative tasks is becoming an increasingly standardized notion, as the volume of medical data and the capabilities of computers continue to advance [7]. Despite the advantageous nature of streamlined decision-making processes and healthcare decision support, algorithms bring an intrinsic nature of bias [5]. When using a human-centered lens, algorithmic decisions are approached with the stakeholders' views and needs as core foundations. The effectiveness of the algorithm is not altered, instead more precisely tuned to those who will interact with it.

3 IMPLEMENTATION CONTEXT

Parkview Health is a 13-facility hospital system serving parts of Indiana, Ohio, and Michigan, based in Fort Wayne, Indiana. During the COVID-19 pandemic, it was imperative for hospital leadership have reasonable forecast of projected patients as they serve as the health "safety net" in the region. In order to predict global case numbers, many pursued predictive simulation models, with SIR models appearing frequently in literature [10, 11]. Limitations of these models arose quickly, as they often assumed constant variables and parameters, temporally and regionally [10]. Authors noted that these models did not take into account reoccurring outbreaks, or unpredictable data sources and types [10]. As well, these models were often extrapolated across multiple countries [10, 12], which other researchers believe make models too inaccurate for implementation in hospital facilities [8]. Parkview realized that the national- and state-level models did not provide data that could be extrapolated into actionable results nor did it take into consideration local contexts.

Thus, Parkview Health began developing its own models, taking into consideration local population statistics, dynamic infection rates, and projected intake percentages related to other health systems in the region. Initially with limited local hospital data, this took the form of compartmental models (SEIR/SEIRR: a simulation model often used for infectious diseases). With the increased local data, more sophisticated, accurate models were introduced including those developed through the Gaussian Mixture technique and our current piecewise regression method, which predicts inpatient COVID hospitalizations for a 1-3 week window (See Figure 1). Piecewise models are widely deemed more explainable and more strongly predictive of a shorter-term timeframe, benefiting public health research.

This implementation was grounded within an HCDS approach as the team developing the models met weekly with hospital leadership, including the heads of infection and prevention, quality and safety, and those in charge of increasing hospital patient capacity. Thus, essential requirements were taken into consideration at every step of the model design. By pairing a human-centered

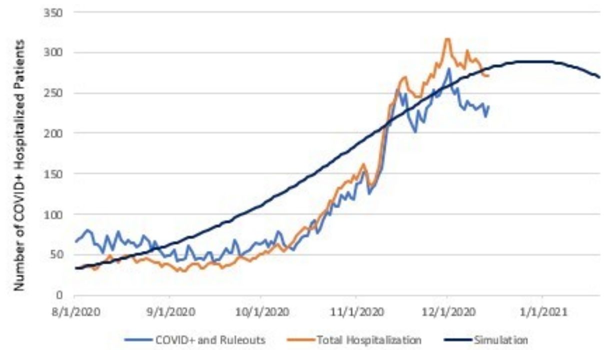

Figure 1a. SEIR Simulation Model (12/2020)

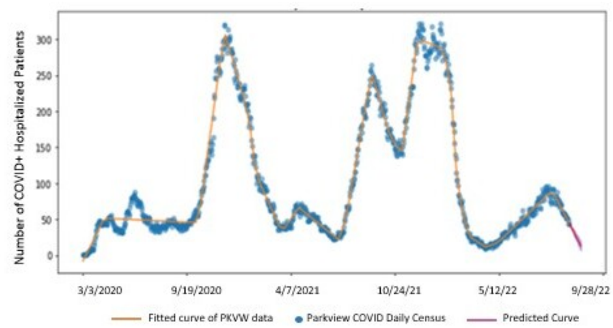

Figure 1b. Piecewise Regression Predictive Model (09/2022)

Figure 1: Evolution of HCDS COVID-19 Modeling

approach with complex computational modeling, a more intricate and effective model could be made for stakeholders [9]. Much of the analysis of COVID-19 spreading behavior strongly incorporated social factors that require local contexts (e.g., masking) [11]. By considering these concepts, thus expanding the algorithm past purely numeric ideas, the algorithm can provide more detailed and accurate predictions as seen through Figure 1 [9].

4 ANALYSIS USING A HUMAN CENTERED DATA SCIENCE APPROACH

We use HCAD's three dimensions [2] to analyze this example of human-centered algorithmic implementation, in order to bring to light the ways in which a human-centered approach can benefit decision-making in healthcare.

Theoretical Approaches to HCDS in Healthcare. *It is important to make sure that an HCDS approach is theoretically valid and not applied in a naive way.* For instance, Parkview administration directly asked for modeling for their population. It was important to understand the population to whom the algorithm was to apply, to center it on their demographics, infection rates and needs. As Jewell et al. (2020) mention, effectiveness of a predictive model for a large region (a country) can be limited due to variation among demographics and social aspects of COVID-19 spread. By centering model development around the population from whom the data is created, a more precise model can be created. Since Parkview's catchment concerns only a small area of the Midwest United States, it was not necessary to incorporate data from other regions or countries, which means human behavior can be more appropriately mapped [8].

Participatory Approaches to HCDS in Healthcare. *An effective HCDS process includes creating a model that prioritizes the benefit of the stakeholder, as well as their preferences.* Instead of anticipating totality of COVID-19 waves, Parkview leadership needed to focus on immediate needs. By anticipating a window of 1-3 weeks, extra staff could be allocated, resources requested, and facilities prepared. As Jewell et al. (2020) mention, focusing on long-term predictions early into the pandemic could prove futile, as the dynamics of COVID-19 had not been analyzed enough for a meaningful model. Nesteruk (2020) seems to prove this case as they predicted an end to COVID "no earlier than March 2021". They acknowledge that certain variables are dynamic and uncertain, supporting the case for shorter-term models. By rejecting the idea of a simulation model and focusing on tangible targets within a month's notice, administrators were able to to take action on the shorter-term piecewise regression model outputs. This was directly beneficial to stakeholders, as it aligned with their preferences, contributed to a realistic set of recommendations for their facilities, and allowed for reliable data-driven decision-making across the enterprise. Data quality is another issue concerning model validity [8], but by using Parkview's own patient data, the outputs were focused on the community and provided a way for us to immediately validate outputs with this data.

Speculative Approaches to HCDS in Healthcare. *HCDS tries to do what others have not previously done, and expand the motivation past intrinsic assumptions.* While other works pursued compartmental and epidemic spread models, which involved trying a simulation approach [12], the predictive model for Parkview took the form of a piecewise regression model. By splitting the window of prediction into shorter periods, the empirical data was more closely matched. Taking an alternative approach to statistical modelling for forecasting COVID-19 hospitalizations allowed the desired time frame for prediction to be accommodated, and this became the iterative unit for further modeling. Using a piecewise regression approach also allowed intrinsic assumptions associated with simulation modelling to be set aside. These assumptions include the consistency of input data variables and data types, where taking data from multiple countries and compiling it into a simulation model can lead to bias towards those countries contributing most greatly to class imbalance [6].

5 CONCLUSION

In this extended abstract, we presented a case study of data science applied to healthcare through a human-centered lens. We showed that implementing HCAD's multiple frameworks can lead to a precise, effective algorithm accounting for stakeholder needs right from the start. Incorporating an HCDS algorithmic design process in healthcare can lead to greater gains in terms of accuracy, utility, and prosperity of a model. With this work, we aim to encourage a human-centered lens for future work with stakeholders in the realm of algorithms in healthcare.

REFERENCES

[1] Cecilia Aragon, Shion Guha, Marina Kogan, Michael Muller, and Gina Neff. 2022. *Human-Centered Data Science: An Introduction.* MIT Press, Cambridge, MA, USA.

[2] Eric PS Baumer. 2017. Toward human-centered algorithm design. *Big Data & Society* 4, 2 (2017), 2053951717718854. https://doi.org/10.1177/2053951717718854

[3] Riyad Bin Rafiq, Francois Modave, Shion Guha, and Mark V. Albert. 2020. Validation Methods to Promote Real-World Applicability of Machine Learning in Medicine. In *2020 3rd International Conference on Digital Medicine and Image Processing* (Kyoto, Japan) *(DMIP '20)*. Association for Computing Machinery, New York, NY, USA, 13–19. https://doi.org/10.1145/3441369.3441372

[4] Longbing Cao. 2018. Data Science: A Comprehensive Overview. *Comput. Surveys* 50, 3, Article 43 (May 2018), 42 pages. https://doi.org/10.1145/3076253

[5] Irene Y. Chen, Emma Pierson, Sherri Rose, Shalmali Joshi, Kadija Ferryman, and Marzyeh Ghassemi. 2021. Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science* 4, 1 (2021), 123–144. https://doi.org/10.1146/annurev-biodatasci-092820-114757

[6] Sibel Eker. 2020. Validity and usefulness of COVID-19 models. *Humanities and Social Sciences Communications* 7, Article 54 (2020). https://doi.org/10.1057/s41599-020-00553-4

[7] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. 2019. A guide to deep learning in healthcare. *Nature Medicine* 25 (Jan. 2019), 24–29. https://doi.org/10.1038/s41591-018-0316-z

[8] Nicholas P. Jewell, Joseph A. Lewnard, and Britta L. Jewell. 2020. Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. *JAMA* 323, 19 (05 2020), 1893–1894. https://doi.org/10.1001/jama.2020.6585

[9] Marina Kogan, Aaron Halfaker, Shion Guha, Cecilia Aragon, Michael Muller, and Stuart Geiger. 2020. Mapping Out Human-Centered Data Science: Methods, Approaches, and Best Practices. In *Companion of the 2020 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '20)*. Association for Computing Machinery, New York, NY, USA, 151–156. https://doi.org/10.1145/3323994.3369898

[10] Igor Nesteruk. 2020. Simulations and Predictions of COVID-19 Pandemic With the Use of SIR Model. *Innovative Biosystems and Bioengineering* 4, 2 (May 2020), 110–121. https://doi.org/10.20535/ibb.2020.4.2.204274

[11] Pushpendra Singh and Anubha Gupta. 2022. Generalized SIR (GSIR) epidemic model: An improved framework for the predictive monitoring of COVID-19 pandemic. *ISA Transactions* 124 (2022), 31–40. https://doi.org/10.1016/j.isatra.2021.02.016

[12] Gang Xie. 2020. A novel Monte Carlo simulation procedure for modelling COVID-19 spread over time. *Scientific Reports* 10 (2020), 13120. https://doi.org/10.1038/s41598-020-70091-1