Adventures in Biomedical Data Science

Research Vision

Harry Hochheiser, PhD
`harry@pitt.edu`
Professor
Biomedical Informatics
Intelligent Systems Program
Clinical and Translational Science Institute
Member, UPMC Cancer Center, Cancer Epidemiology and Prevention Program

Cross-cutting challenges in Biomedical Informatics

- Interpretability
    - How to best make sense of data and predictions?
- Reproducibility
- Team Science
    - Supporting community & collaborative science

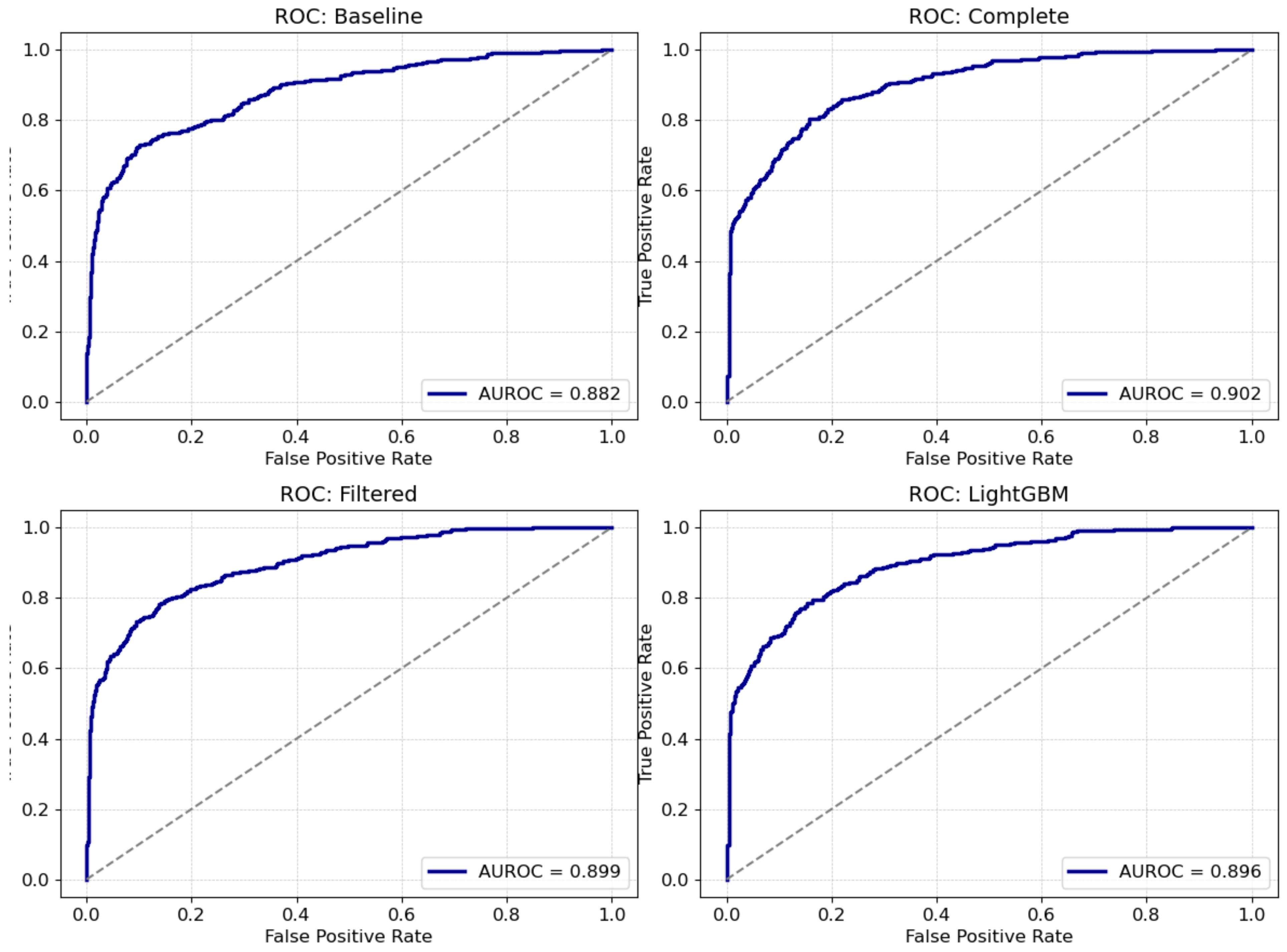
Automated and interoperable methods for generalizable development of clinical machine-learning models for predicting neuromorbidity in critically ill children

Ruoting Li, Christopher M Horvat, Mehdi Nourelahi, Eddie Pérez Claudio, Jason Hammett, Mark S Wainwright, Robert SB Clark, Alicia K Au, Harry Hochheiser medRxiv 2025 August 5 https://doi.org/10.1101/2025.08.01.25332805

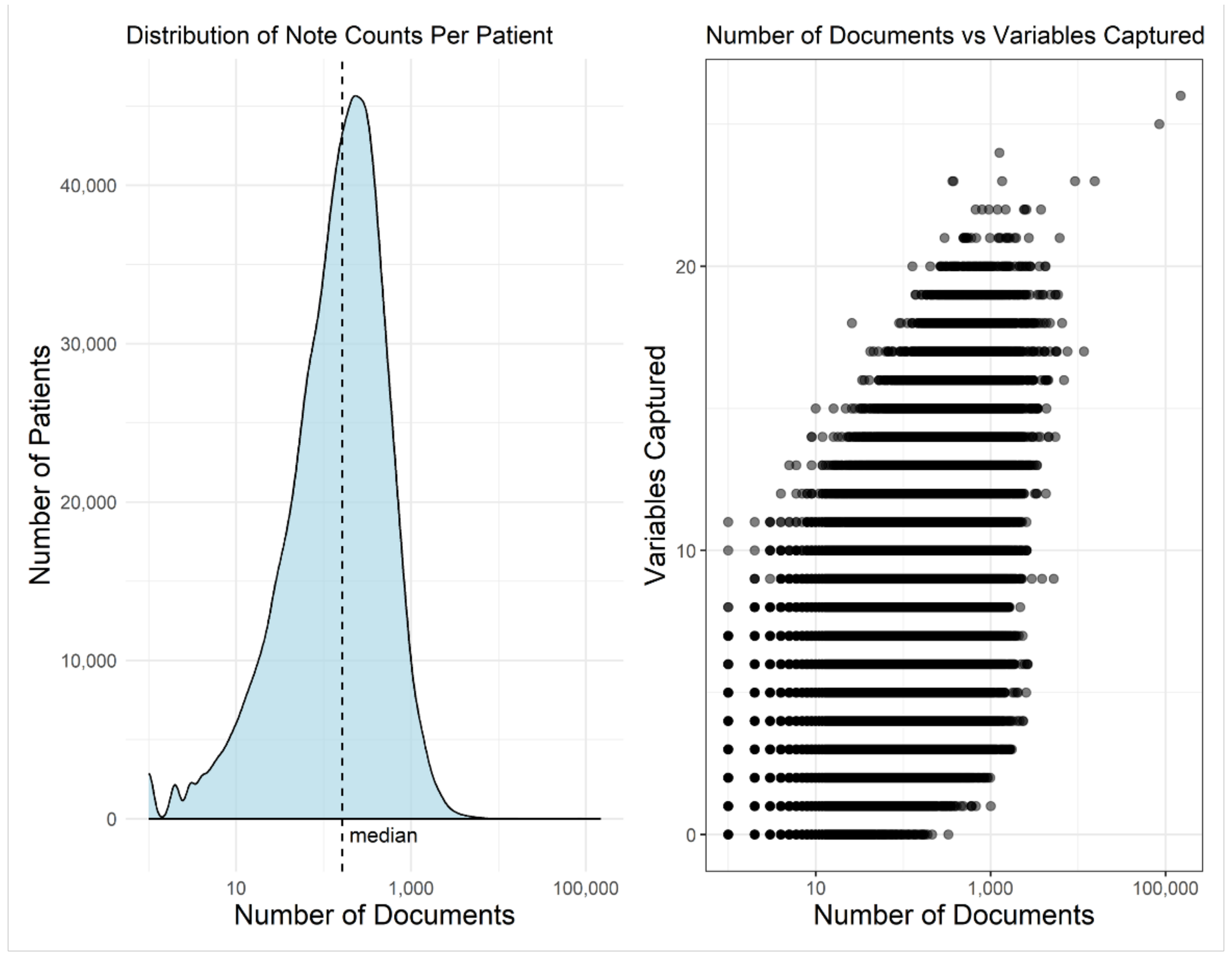
Declining mortality in the field of pediatric critical care medicine has shifted practicing clinicians' attention to preserving patients' neurodevelopmental potential as a main objective. Earlier identification of critically ill children at risk for incurring neurological morbidity would facilitate heightened surveillance that could lead to timelier clinical detection earlier interventions, and preserved neurodevelopmental trajectory.

- Challenge: building reproducible, generalizable models
- AutoML + FHIR
- AWS infrastructure (later used with open-source AutoML)
- Performance improves, even with smaller feature set.

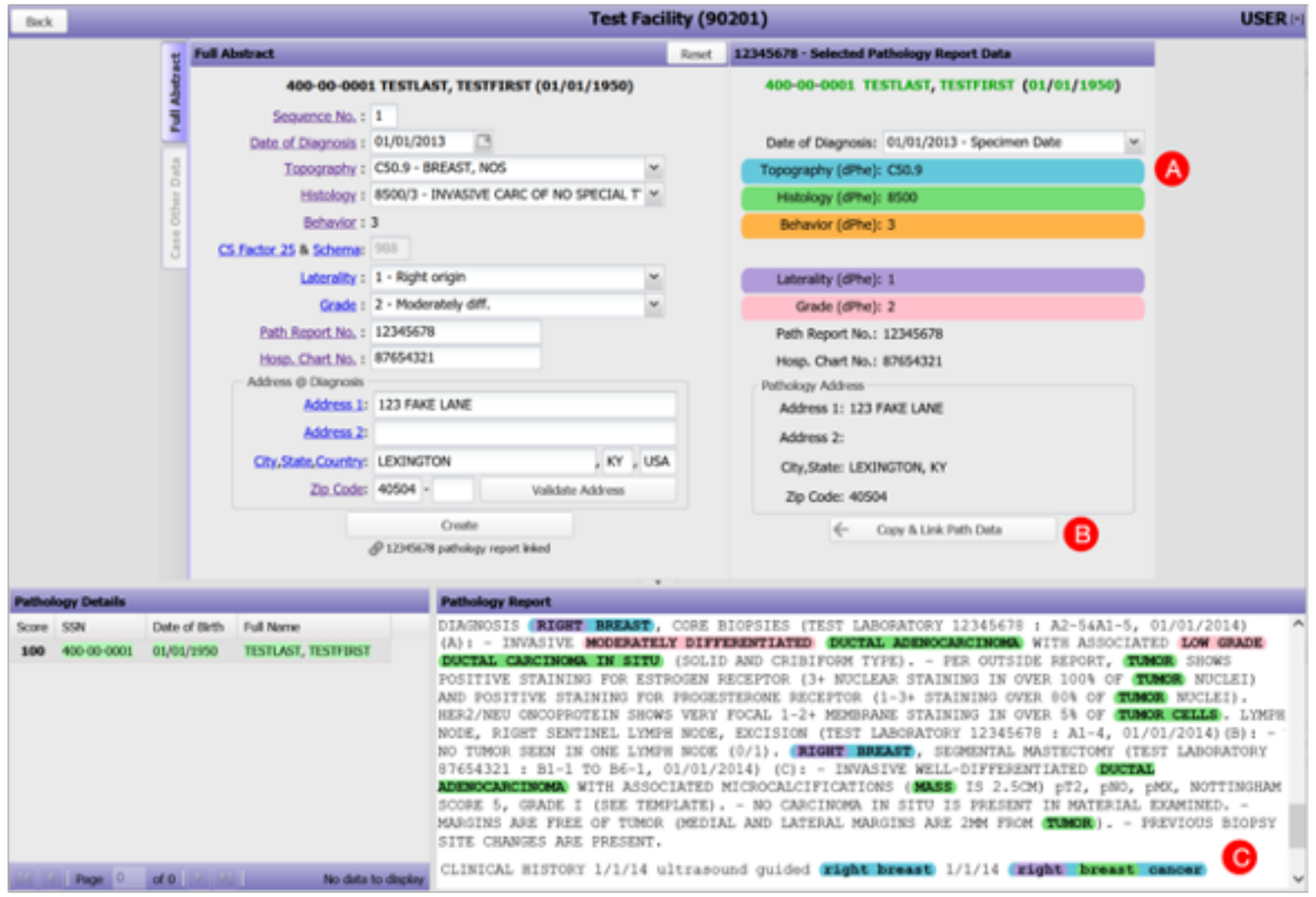
Funding: R01NS118716 Bio-Digital Rapid Alert to Identify Neuromorbidity (MPIs- Clark, Au, Horvat, Hochheiser)



DeepPhe: Cancer NLP and Visualization



Large variation in # of documents/patient and concepts/documents



NLP extraction embedded in cancer registry tools

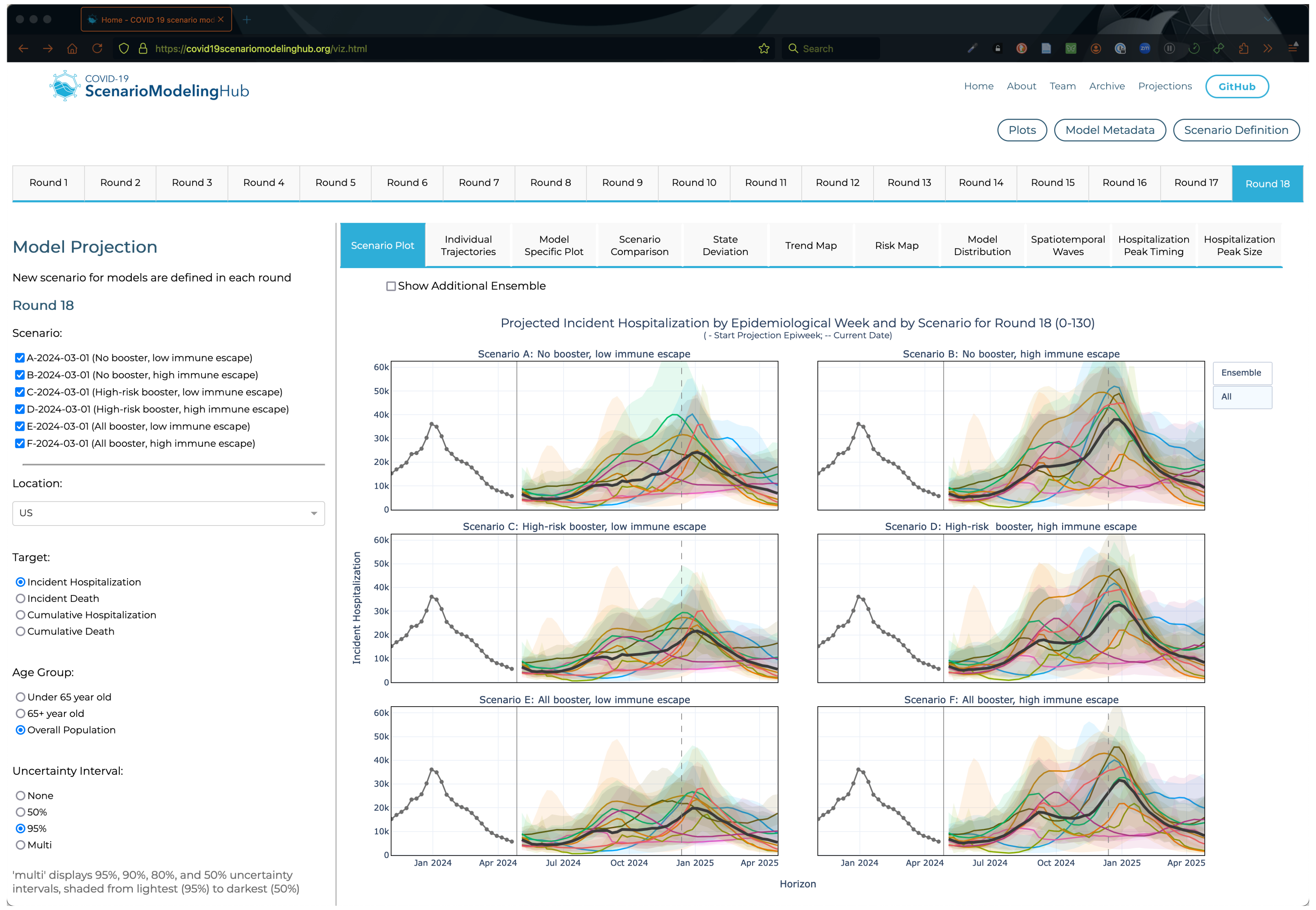
- Natural Language Processing, Knowledge Representation, and Visualization to extract and visualize cancer patient histories

- U24CA248010 Cancer DeepPhenotype Extraction from Electronic Medical Records (2020-2030)
- UH3CA243120 Natural Language Processing Platform for Cancer Surveillance (MPI, 2019-2024)

Models of Infectious Diseases Agent Study (MIDAS) Coordination Center

midasnetwork.us U24GM132013 (2019-2024) and R24GM153920 (2024-2029)

- The MIDAS network aims to advance science to improve global preparedness and response against infectious disease threats through research, training, promotion, and service.
- Community support
- Team Science -Scenario Modeling Hub collaborative projection efforts
- Research: reproducibility of infectious disease modeling efforts



An implementation framework to improve the transparency and reproducibility of computational models of infectious diseases

Darya Pokutnaya, Bruce Childers, Alice E. Arcury-Quandt, Harry Hochheiser, Willem G. Van Panhuis

Published: March 16, 2023 • https://doi.org/10.1371/journal.pcbi.1010856

Inter-rater reliability of the infectious disease modeling reproducibility checklist (IDMRC) as applied to COVID-19 computational modeling research

Darya Pokutnaya, Willem G Van Panhuis, Bruce Childers, Marquis S Hawkins, Alice E Arcury-Quandt, Meghan Matlack, Kharlya Carpio & Harry Hochheiser