A graphical toolkit for longitudinal dataset maintenance and predictive model training in healthcare

Eric Bai [1], Sophia Song [1], Hamish SF Fraser MBChB, MSc [2], Megan Ranney MD, MPH [2,3]
[1] Brown Alpert Medical School, [2] Brown University Center for Biomedical Informatics, [3] Brown-Lifespan Center for Digital Health

1 Background & Significance

Models used in real-world clinical practice

34.6% hospital readmission models

44 models published between 2010 and 2019

1. A large body of literature exists on the development of predictive models for healthcare, but best-practices for local validation, implementation, and long-term maintenance in clinical settings remain unclear
2. Only a **small fraction of the codebase of ML projects is actual ML code**; the rest is infrastructure that can be factored out of individual projects into a reusable framework
3. Reducing codebase size = smaller the footprint susceptible to accruing **technical debt**
4. **Machine learning operations (MLOps)** = applying well-established development and operations (DevOps) practices in software engineering to scaling ML in production
5. While DevOps tools are robust and widely available, **few comprehensive, publicly-available end-to-end platforms for MLOps exist**

Our work: A conceptual framework, in the form of a **graphical prototype**, for a programming language and model agnostic MLOps tool designed to **support the entire lifecycle of predictive model development using EHR data**

2 Methods

1. **Identify key user personas**
   (1) the health informatician, (2) the chief information officer (CIO), and (3) the organization's clinical and data stakeholders (physicians, clinical staff, and healthcare researchers)
2. **Define workflow parameters**
3. **Create graphical prototypes**
   (1) creating and implementing a new model and (2) maintaining production models deployed over a few months
4. **Conduct semi-structured user interviews**

Figure 1 The machine learning workflow

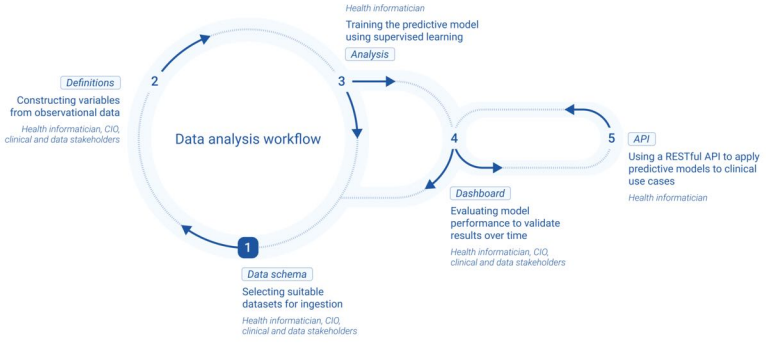

Figure 2 Design overview

3 Results & Discussion

71% successful completion of prompted tasks

75.8/100 System Usability Scale score (n=4)

- Convenience sample of two health informaticians, two CIO, and two healthcare researchers
- Using the Curved Grading Scale (CGS) interpretation for SUS scores, this translates to a **percentile range of 70 to 79, a letter grade of B, and an adjective rating of "good"**
- Users found the overall value of the system to be streamlining the continued monitoring and upkeep of production models

The initial graphical prototypes of this MLOps system are **preliminarily usable** and demonstrate an unmet need within the clinical informatics landscape

4 Future Work

- Conducting summative user research on our design and refining the core feature set and graphical interface through successive iterations of design and evaluation
- Exploring mobile interfaces, trialing custom performance metrics and visualizations to better support convergent and divergent external validation approaches, and integrating with third-party performance visualization tools

Full development of this tool will pave the way for effective application of advanced techniques such as transfer learning and encourage further advancements in **harnessing big data to improve the safety and efficacy of clinical care**

We would like to thank Brown Physicians Inc and Lifespan staff for their valuable feedback during the design and user testing phases of this project.