

Augmented Intelligence to Identify Patients with Advanced Heart Failure

Baljash Cheema, MD, MSCI,^{1,2,3} R. Kannan Mutharasan, MD,^{1,2,3} Firas H. Wehbe, MD, PhD,^{1,2} Aditya Sharma,^{1,3} Jason Ronald,¹ Susan Lehrer, BSN, RN,¹ Kaleigh Powers, MSN, RN,¹ Lindsay Pifer, RN, BSN,¹ Maia Jacobs, PhD,^{2,4} Jonathan D. Rich, MD,^{1,2} Kambiz Ghafourian, MD,^{1,2} Amit Pawale, MD,^{1,2} Anjan Tibrewala, MD, MS,^{1,2} Patrick McCarthy, MD,^{1,2,3} Duc T. Pham, MD,^{1,2} Jane E. Wilcox, MD, MSc,^{1,2} Faraz S. Ahmad, MD, MS^{1,2,3}

¹Northwestern Medicine; ²Northwestern University Feinberg School of Medicine; ³Bluhm Cardiovascular Institute Center for Artificial Intelligence, Northwestern Medicine; ⁴Northwestern University McCormick School of Engineering

Background

- Heart failure (HF) is a highly morbid and costly condition affecting over 6.2 million US adults.
- Clinical guidelines recommend timely referral of patients with advanced HF for specialist evaluation, which includes optimizing medical treatment, discussing prognosis and goals of care, and, if indicated, evaluating for heart transplantation and heart pumps.
- However, the transition from stage C HF to advanced or stage D HF often goes undetected in routine care, resulting in delayed referral and higher mortality rates.

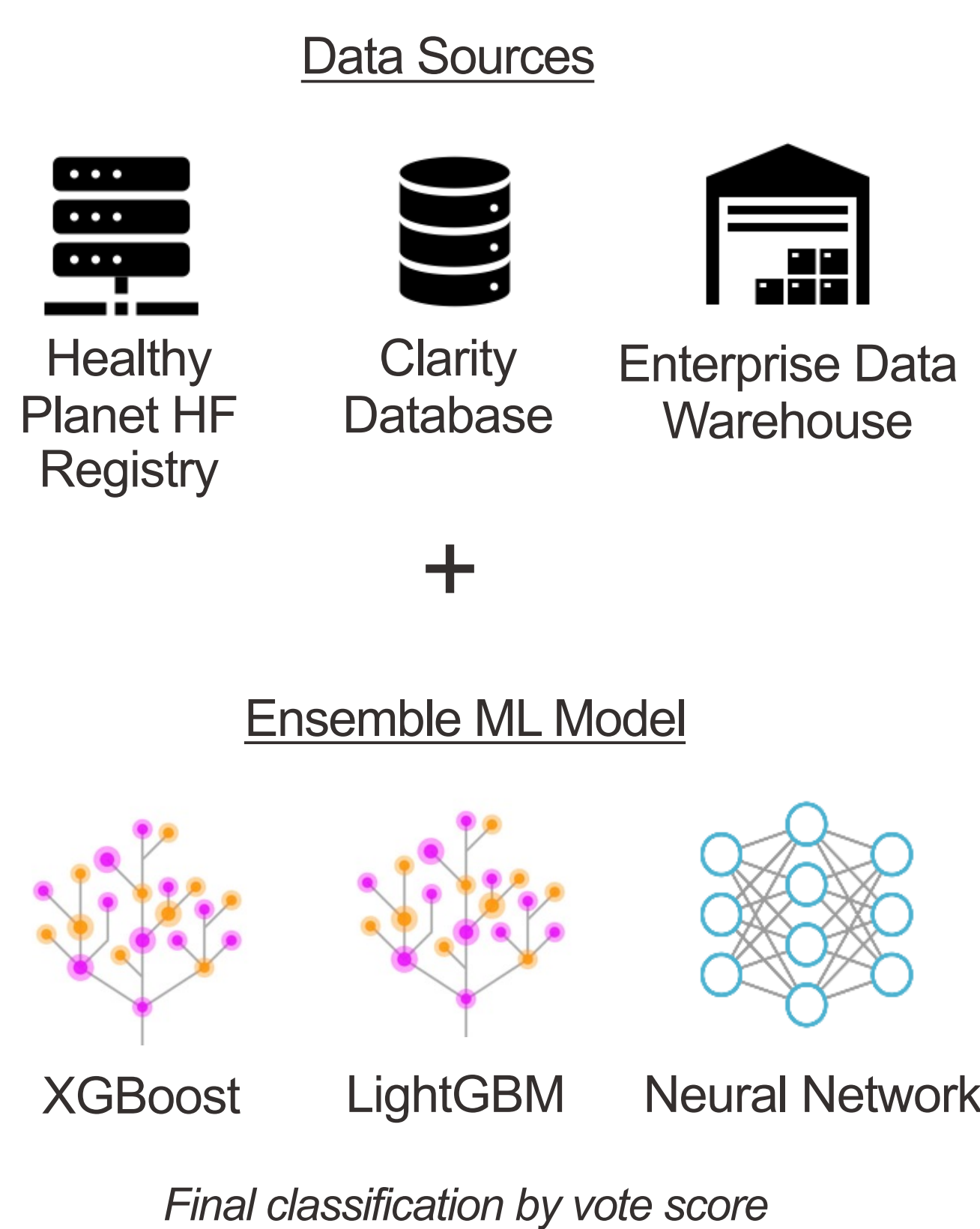
Objective

- Develop an augmented intelligence-enabled workflow to identify patients with advanced HF and streamline referral.

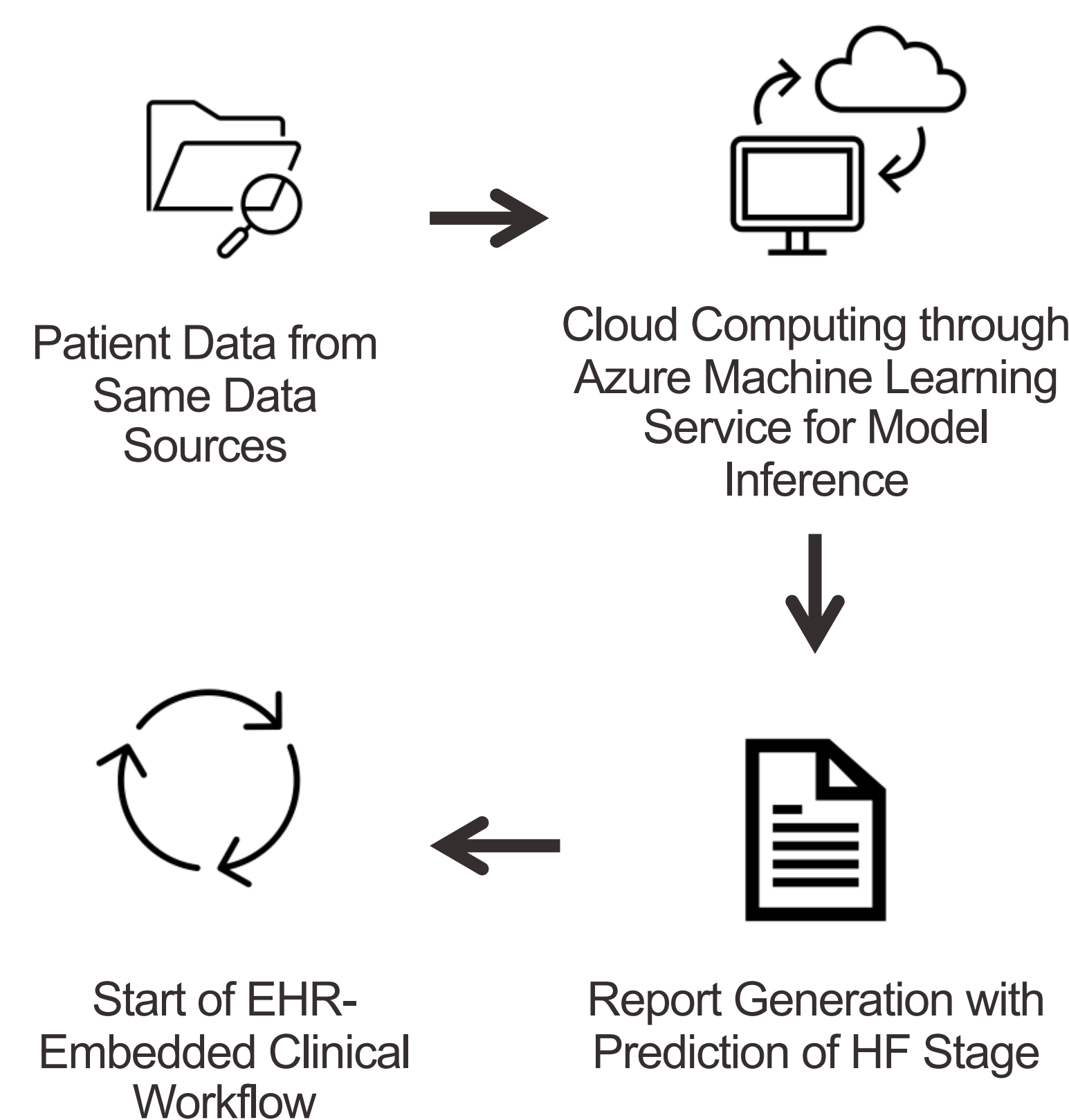
Methods

- We extracted data on HF patients with encounters from 01/01/07-11/30/20 from a HF registry at a regional, integrated health system.
- We created an ensemble ML model to predict stage C or stage D HF and integrated the results within the EHR and clinical operations.

Model Development



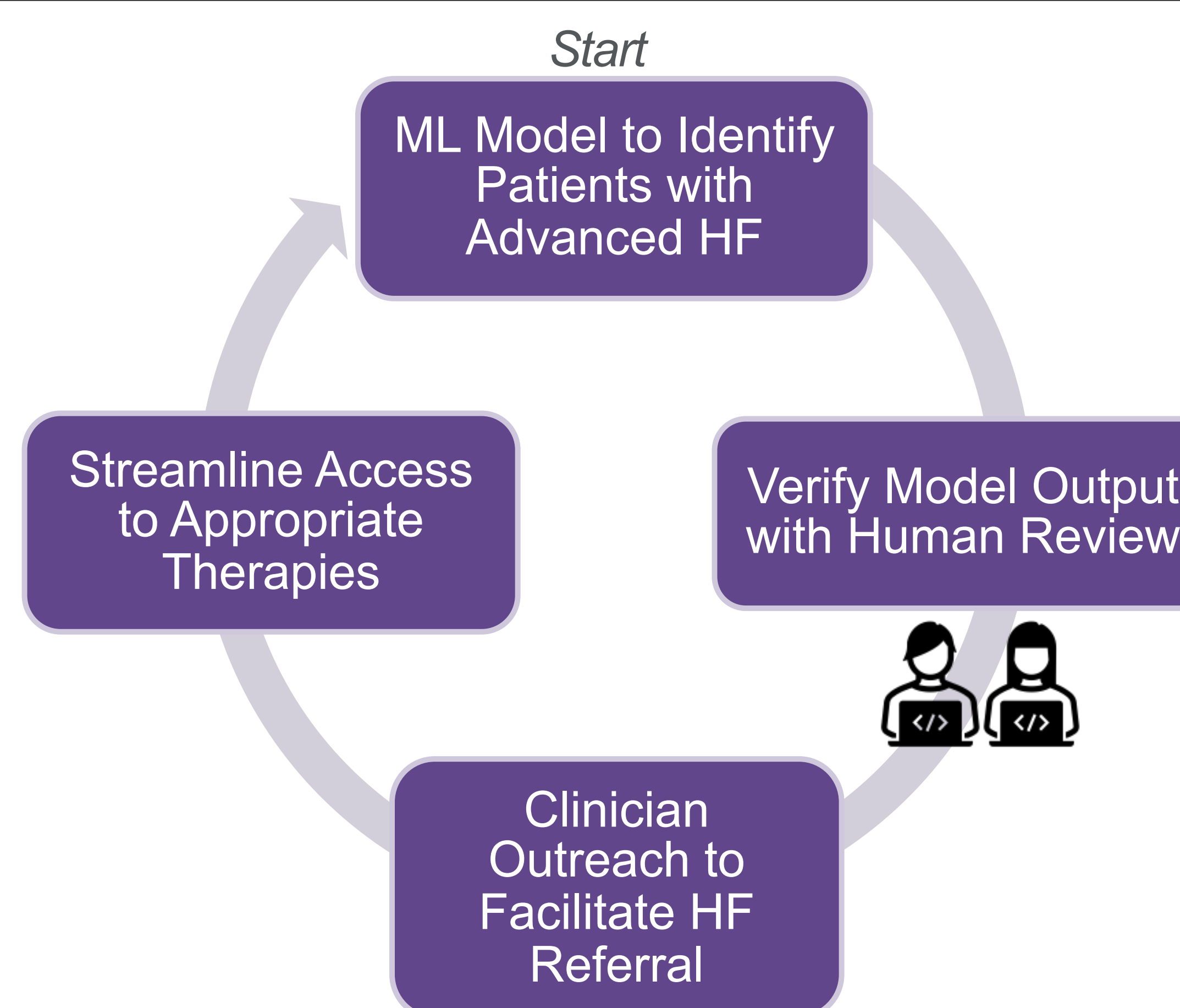
Model Implementation



Key Lessons Learned:

- ✓ Implementing augmented intelligence-enabled workflows within the EHR requires a high-performing, diverse, adaptable team with support from leadership to allow for ongoing performance evaluation and improvement.
- ✓ The workflow for successful model deployment may require more resources than creating the model itself.
- ✓ Ongoing evaluation is required to track model performance and workflow outcomes and ensure the workflow advances health equity and does not perpetuate structural racism and bias.

Augmented Intelligence-Enabled Clinical Workflow



With Continuous Quality Improvement for Model and Workflow

Results

- In a retrospective dataset of 14,846 patients, the model had good precision (0.58) and low recall (0.30) for identifying stage D HF in a 100-person, physician-reviewed, holdout test set.

TEST SET	Precision	Recall	F1 Score	Accuracy	N
NOT HF	0.87	0.87	0.87		1500
STAGE C	0.73	0.81	0.77		1189
STAGE D	0.60	0.34	0.44		303
ACCURACY				0.79	
Total = 2992					
PHYSICIAN-REVIEWED SET	Precision	Recall	F1 Score	Accuracy	N
NOT HF	-	-	-		-
STAGE C	0.81	0.90	0.85		77
STAGE D	0.58	0.30	0.40		23
ACCURACY				0.76	
Total = 100					

- During prospective implementation of the workflow from 4/1/21 to 2/15/22, 416 patients were reviewed by a clinical coordinator, with a positive predictive value of 50.3% for Stage D prediction.
- 24 patients have been scheduled for evaluation in a HF clinic, 4 patients started an evaluation for advanced therapies, and 1 patient received a left ventricular assist device.

Model Performance (n=416)	
	Precision
Stage C (n=120)	63.1%
Stage D (n=293)	50.3%
"Next Steps" (n=401)	
	Count
Consider evaluation in Advanced Heart Failure New Access Clinic	56
Review in 3 Months	58
Follow-up in HF clinic*	77
Consider evaluation in general cardiology clinic	3
No additional recommendations at this time	207

*Includes patients already established in HF clinic or new patients with Stage C HF who can be seen in HF clinics through the standard referral workflow.

Conclusions and Future Directions

- It is feasible to implement an augmented intelligence-enabled workflow embedded into the EHR and integrated into clinical operations to identify patients with advanced HF.
- Our next steps include improving model performance by using unstructured (imaging, notes) and longitudinal data and evaluating performance across diverse populations.