HHS Public Access

Author manuscript

JACC Adv. Author manuscript; available in PMC 2023 January 13.

Published in final edited form as:

JACC Adv. 2022 October ; 1(4): . doi:10.1016/j.jacadv.2022.100123.

Augmented Intelligence to Identify Patients With Advanced Heart Failure in an Integrated Health System

Baljash Cheema, MD, MSCI, MSAI[a,b], **R. Kannan Mutharasan, MD**[a,b], **Aditya Sharma, BS**[a,c], **Maia Jacobs, PHD**[d], **Kaleigh Powers, MSN, RN**[c], **Susan Lehrer, BSN, RN**[c], **Firas H. Wehbe, MD, PHD**[a,e], **Jason Ronald, MA**[c], **Lindsay Pifer, RN, BSN**[c], **Jonathan D. Rich, MD**[b], **Kambiz Ghafourian, MD**[b], **Anjan Tibrewala, MD, MS**[b], **Patrick McCarthy, MD**[a,e], **Yuan Luo, PHD**[f], **Duc T. Pham, MD**[e], **Jane E. Wilcox, MD, MSC**[b], **Faraz S. Ahmad, MD, MS**[a,b,f]

[a]Bluhm Cardiovascular Institute Center for Artificial Intelligence, Northwestern Medicine, Chicago, Illinois, USA

[b]Division of Cardiology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

[c]Northwestern Medicine, Chicago, Illinois, USA

[d]Department of Computer Science, Northwestern University McCormick School of Engineering, Evanston, Illinois, USA

[e]Division of Cardiac Surgery, Department of Surgery, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

[f]Division of Health and Biomedical Informatics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.

Abstract

BACKGROUND—Timely referral for specialist evaluation in patients with advanced heart failure (HF) is a Class 1 recommendation. 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.

OBJECTIVES—The authors sought to develop an augmented intelligence-enabled workflow using machine learning to identify patients with stage D HF and streamline referral.

METHODS—We extracted data on HF patients with encounters from January 1, 2007, to November 30, 2020, from a HF registry within a regional, integrated health system. We created an ensemble machine learning model to predict stage C or stage D HF and integrated the results within the electronic health record.

ADDRESS FOR CORRESPONDENCE: Dr Faraz S. Ahmad, Northwestern University Feinberg School of Medicine, 676 North Saint Clair Street, Suite 600, Chicago, Illinois 60611, USA. faraz.ahmad@northwestern.edu.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate.

RESULTS—In a retrospective data set of 14,846 patients, the model had a good positive predictive value (60%) and low sensitivity (25%) for identifying stage D HF in a 100-person, physician-reviewed, holdout test set. During prospective implementation of the workflow from April 1, 2021, to February 15, 2022, 416 patients were reviewed by a clinical coordinator, with agreement between the model and the coordinator in 50.3% of stage D predictions. Twenty-four 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.

CONCLUSIONS—An augmented intelligence-enabled workflow was integrated into clinical operations to identify patients with advanced HF. Endeavors such as this require a multidisciplinary team with experience in design thinking, informatics, quality improvement, operations, and health information technology, as well as dedicated resources to monitor and improve performance over time.

Keywords

advanced heart failure; artificial intelligence; augmented intelligence; electronic health record; integrated healthcare system; machine learning

Heart failure (HF) is a chronic, heterogeneous, and highly morbid condition affecting approximately 6 million U.S. adults.[1] Each year, approximately 4.5% of patients with chronic HF (referred to as stage C HF) progress to advanced or stage D HF, resulting in an estimated prevalence of stage D HF ranging from 5% to 15% among the entire HF population.[2–4] Timely referral of patients with advanced HF is a Class I recommendation in the 2022 American Heart Association/American College of Cardiology/Heart Failure Society of America guideline for the management of HF.[5] In addition to optimization of guideline-directed medical therapy, this referral may lead to an evaluation for advanced therapies, such as heart transplantation, left ventricular assist device implantation (LVAD), or palliative home inotropes.[6] These interventions may improve quality of life, increase survival, or both. The median survival duration after heart transplantation is 12 to 13 years while the survival rate after LVAD implantation is 51% through 7 years.[7,8] In patients who warrant advanced therapies but do not receive them, mortality estimates are near 50% or greater annually.[2]

The challenge, however, is that patients are often referred too late for these therapies to help. This was shown in a recent retrospective review of referral patterns for advanced HF care in 9 academic medical centers. In patients who were evaluated for advanced therapies but did not receive them, 50% were felt to be “too sick” to benefit or died before their evaluation was completed.[9] Multiple factors likely contribute to delayed referral, including the need for frontline clinicians to identify and act on the at-times-subtle signs and symptoms that mark the transition from chronic to advanced HF.

Increases in computing power and advances in digital infrastructure have enabled the creation of machine learning (ML) and artificial intelligence (AI) models that perform well on isolated predication tasks across a wide range of clinical domains.[10,11] However, addressing real-life problems that arise when caring for patients using such models has

been a significantly greater challenge, limiting the impact of ML and AI applications on day-to-day clinical workflows.[12]

With this context in mind, we share our efforts to streamline the identification of patients with advanced HF to facilitate referral for specialist evaluation across a large, diverse, integrated health system using augmented intelligence.[13] Augmented intelligence is a subsection of AI that aims to enhance human performance and decision-making capability by emphasizing cooperation between humans and AI model output. We first sought guidance on the structure of an electronic health record (EHR)-based clinical workflow that would allow for care coordination of patients with advanced HF. Next, we used ML to create a model that identifies such patients, embedding those results within the workflow. We used iterative feedback from the clinical team using the workflow for continued process improvement.[14] Here, we report on developing, testing, and deploying the ML model, detail the creation and integration of the EHR-based clinical workflow, highlight aspects of decision-making in the care team, and share early results and lessons learned from using this process in daily clinical operations.

METHODS

STUDY POPULATION AND DATA SOURCES.

This is a study carried out at a large, integrated health system with 11 hospitals and over 100 clinical sites spanning from urban to suburban locations including some rural area coverage. Retrospective data were obtained from the health system's comprehensive enterprise data warehouse, which collects diverse types of health data including demographics, laboratory values, prescription data, administrative claims data, procedure codes from inpatient and outpatient encounters, and patient-reported outcomes from multiple recording systems. This study was not considered human subjects research, thus was granted exemption from review by the Northwestern University Institutional Review Board. This study adheres to guidance put forward by the MI-CLAIM (minimum information about clinical artificial intelligence modeling) checklist for providing "minimum information" criterion for ML/AI research in clinical medicine (Supplemental Table 1).[15]

TRAINING SET AND LABEL GENERATION.

The final data set for model training was created from the Heart Failure Registry within the Healthy Planet module of Epic (Epic Systems Corp) on March 21, 2021, from encounters dated between January 1, 2007, to November 30, 2020. Patients are entered into this registry if they have had 2 or more International Classification of Diseases codes for HF identified in inpatient, outpatient, or claims encounters within an 18-month period.

We considered multiple strategies to assign the advanced HF label. Advanced HF is defined by a constellation of features, many of which exist as unstructured data within the EHR and are not consistently available across a significant number of patients. One consideration was to assign the label to patients who underwent evaluation for heart transplantation or LVAD implantation previously within the system, but we were concerned that this approach could miss patients from racial and ethnic minoritized groups and from under-resourced

communities who have been shown to have lower rates of timely referrals for HF specialist evaluation in the literature. We decided to assign labels using natural language processing to identify text phrases coinciding with the stages of HF within clinical notes. This terminology is based off the widely utilized consensus guidance from the American College of Cardiology and the American Heart Association.[16] In this schema, stage C consists of patients who currently have or have ever had symptoms of HF. Stage D consists of patients with advanced HF, who have persistent symptoms in spite of optimized medical and device therapies and should be considered for advanced therapies including palliative care.

Within the given time window, 52,713 patients were identified in the registry. Among those, a text search using regular expressions for the phrases “stage C” or “stage D” was completed on clinical notes of History and Physical, Ambulatory History and Physical, Progress Note, or Ambulatory Progress Note, written by a cardiology clinician. The corresponding label was assigned to the patient. If both stage C and stage D were identified, the label of stage D was assigned. If multiple notes with matching phrases were found, the earliest note for the given label was used. A total of 7,346 patients with disease labels were identified (6,034 stage C; 1,312 stage D). To balance the data set, 7,500 patients who had an echocardiogram within the health system during the same time window but were not in the HF registry were chosen at random and assigned the label of “not HF”. The final cohort contains 14,846 patients with 1 of 3 disease labels (Figure 1).

FEATURE SELECTION.

Candidate features for the ML model were selected from 3 established HF risk scores: the Seattle Heart Failure Model, the Meta-Analysis Global Group in Chronic Heart Failure Risk Calculator, and the MARKER-HF (Machine learning Assessment of RisK and EaRly mortality in Heart Failure), in addition to expert opinion from the HF team. These risk scores were chosen for their individual performance as well as the unique aspects of each that made them appealing. The Seattle Heart Failure Model and Meta-Analysis Global Group in Chronic Heart Failure are 2 of the most well-known risk models for patients with HF and have been included as a tool that can be used for risk stratification in HF consensus statements and guidelines.[2,5,17,18] MARKER-HF is a boosted decision tree ML risk model that relies upon readily available and low-cost tests.[19]

All data used for feature generation were captured in routine clinical care and were extracted as structured data from the EHR. The final list contains 32 features (Table 1). Data were pulled from the Heart Failure Registry or the enterprise data warehouse. All features were transformed into categorical values with cut points based on laboratory cutoffs or expert guidance (Supplemental Table 2). For features with more than 2 classes, one-hot encoding was performed. We excluded age as a feature due to the high correlation between advanced age and the target label, a finding that may mask the ability of the model to identify the underlying relationships between other features and the target label. Furthermore, this was particularly important in this use case because advanced age would disqualify many patients from receiving several forms of advanced therapies, limiting the utility of the model in practice. Features were extracted nearest to the date of the clinical note used to provide the patient label for those with stage C or stage D or the date of the echocardiogram for

patients of class not HF. Missing data were categorized as unknown. As our data are linked to processes done sporadically in routine care, some values had a moderate to high rate of missingness (Supplemental Table 3). No patients were excluded due to the degree of missingness of any of the data elements.

MODEL DEVELOPMENT AND PERFORMANCE.

The ML model was tasked with a multiclass classification problem, specifically determining whether a given patient had 1 of the 3 possible labels (stage D HF, stage C HF, or not HF). The training set was separated to allow for an 80/20 train/test split, after 100 patients were held out and reserved for manual chart review. Those 100 were chosen randomly from all patients with an ejection fraction (EF) of <40%, a group we felt was most likely to benefit from advanced HF therapies. A cardiology fellow (B.C.) assigned a label of stage C or stage D HF to these 100 using consensus criteria and guideline recommendations.[2,6] This set of patients is referred to as the physician-reviewed set. Model performance on the test set was stratified by sex and race.

The ML model is an ensemble of a fully connected feedforward neural network and 2 variants of decision trees with gradient boosting.[20] Network architecture and hyperparameters for the neural network were chosen using AutoKeras, an open-source library for automatically discovering the best-performing model given a data set. The neural network consisted of an input layer with 27 nodes, 2 hidden layers with 32 and 512 nodes, respectively, and a final classification layer with 3 output nodes. Hyperparameters were chosen using grid search with cross-validation for the decision trees. The final hyperparameters used are given in Supplemental Table 4, and the final classification of the ensemble was made by vote score.

MODEL DEPLOYMENT AND AUGMENTED INTELLIGENCE WORKFLOW.

Input on how best to utilize the ML predictions to facilitate coordination of care in patients with advanced HF was sought from key stakeholders, including the HF clinical team and members from the information services team, prior to workflow creation. The approach decided upon is discussed below.

The ML model output was embedded into an augmented intelligence-enabled workflow within the EHR using the Epic Cognitive Computing Platform powered by Microsoft Azure (Microsoft Corp). The model screened patients within the Heart Failure Registry for the presence of advanced HF each time the Heart Failure Registry was triggered to refresh, and the results were delivered to a nurse coordinator via an Epic Workbench Report (Figure 2). The nurse coordinator subsequently performed a chart review and, in an Epic SmartForm, documented an assessment of likely HF stage, providing a check on model performance and generating data for future fine-tuning of the model. Criteria for making this determination were created by the HF team and are in conjunction with guideline statements.[2,6] Recognizing the inherent challenges with assessing a patient's clinical status through chart review alone, we erred on the side of assigning the stage D label when some but not all criteria were met, to avoid missing patients with a potentially advanced disease. All patients assigned the stage D label were reviewed with an advanced HF cardiologist.

The nurse coordinator then documented next steps for the patient, such as an expedited evaluation in an advanced HF clinic, re-review patient in 3 months, follow-up in HF clinic as an established patient or as a new patient nonurgently, or no additional recommendations at this time. Patients who are predicted to be stage D with EF <40% and younger than 75 years were given the highest priority in the initial stages of the review, followed by patients with stage C HF with the youngest age and lowest EF. Patients who were identified by the nurse coordinator as candidates for expedited evaluation in HF clinic were further reviewed with at least one other member of the advanced HF team, which included 2 HF cardiologists, a cardiology fellow, and an advanced HF program manager. A letter via Epic was then sent to the patient's primary outpatient cardiologist or primary care physician if there is no cardiologist, and with approval pending from these individuals, the patient was invited to be seen in the HF clinic. The augmented intelligence-enabled clinical workflow is depicted in Figure 3. The full process from data collection, to model training and deployment, to the real-time clinical workflow is depicted in the Central Illustration.

MACHINE LEARNING MODEL GENERATION AND STATISTICAL ANALYSES.

The ML algorithm was created with Python v3.8.5 using commonly employed computational libraries including NumPy, Pandas, scikit-learn, and Matplotlib.[21–24] The neural network was created with Autokeras within the TensorFlow framework, and the 2 decision trees with gradient boosting were each created with XGBoost and LightGBM.[25–27] During model development, performance was assessed using sensitivity, positive predictive value, F1 score (the harmonic mean of sensitivity and positive predictive value), and overall accuracy of predicting stage C or stage D in the test set.

QUALITY IMPROVEMENT AND CLINICAL OUTCOMES.

The DMAIC (define, measure, analyze, improve, control) strategy was used for quality improvement.[28] After the launch of the augmented intelligence-enabled workflow, we measured our ability to improve advanced HF detection and referrals by calculating the positive predictive value for patients predicted to have stage D HF by the ML classifier relative to the disease labels from clinical documentation as well as agreement of disease class by the nurse coordinator. Additionally, we quantified the clinical recommendations for each patient documented within the workflow, the number of patients referred and seen by the HF team, and the number of advanced therapy evaluations opened as part of this program.

RESULTS

Demographic data for the study population are provided in Table 2. At the time of model creation, 7,346 of the 52,713 patients (13.9%) in the registry had a disease label of stage C or stage D documented in a clinical note within the EHR and were included in model training. For those with stage C HF, positive predictive value, sensitivity, and F1 score were 78%, 85%, and 81% in the test set and 79%, 91%, and 85% in the physician-reviewed set, respectively. For those with stage D HF, positive predictive value, sensitivity, and F1 score were 74%, 43%, and 54% in the test set and 60%, 25%, and 35% in the physician-reviewed set, respectively. The overall model accuracy was 83% in the test set and 75%

in the physician-reviewed set. Model performance is given in Table 3. Model performance was similar when stratified by sex or race and is given in Supplemental Tables 5 and 6, respectively.

From April 1, 2021, to February 15, 2022, the nurse coordinator prospectively reviewed 508 patients. Ninety-two patients were excluded from further review due to prior LVAD or transplant or deceased status, leaving 416 patients for a more in-depth chart review. Of the 294 patients predicted as stage D, the coordinator agreed with 50.3% (148/294) of the predictions (Table 4). Of the 122 patients predicted as stage C, the coordinator agreed with 63.1% (77/122).

As shown in Table 4, a total of 56 patients were recommended for an expedited evaluation in HF clinic. Fifty-eight patients with high-risk features who may benefit from expedited HF specialist evaluation in the future were flagged for re-review in 3 months. The remaining patients were recommended to either follow up with their current care team or consider nonurgent evaluation in a general cardiology or HF clinic. This included patients who were receiving their cardiology care primarily outside of our institution or had recent comorbidities such as active cancer or stroke, precluding an evaluation for advanced therapies. Twenty-four patients have been scheduled for evaluation in a HF clinic, 4 patients started an evaluation for advanced therapies, and 1 patient received an LVAD.

DISCUSSION

We created and deployed an ML model within an augmented intelligence-enabled workflow embedded in the EHR to identify patients with advanced HF and facilitate specialist evaluation for optimization of guideline-directed medical therapy and an evaluation for potentially lifesaving, time-sensitive therapies. We did this by leveraging structured data generated through routine clinical care to train an ML algorithm to predict the presence of advanced HF using a cloud computing service. The model was then used for inference to predict the severity of HF among patients within an HF registry, with its predictions routed to the EHR via an Epic Workbench report for care coordination by the clinical team. The predictions were subsequently reviewed, and patients were contacted and arranged for clinic appointments as appropriate. We quantified aspects of this process to measure our ability to improve the detection of advanced HF using augmented intelligence. This process went live in clinical operations by April 2021, and through the first 9 months of its use, 4,104 patients have been labeled stage C HF or stage D HF by the ML classifier, with 508 of these patients reviewed by the coordinator and 416 labeled as stage C or stage D. The coordinator agreement for a stage D label by the classifier was 50.3%. Fifteen patients have been seen in clinic or more urgently in the hospital, with 4 advanced therapies evaluations opened and 1 patient receiving an LVAD.

Our work was motivated by a gap in timely referrals for advanced HF patients to specialty care. Of the 6 million people in the United States with HF, 4% to 5% transition each year to advanced HF, and there are at least 300,000 people with advanced HF in the United States right now.[1,2] Heart transplantation can be a life-saving therapy for this patient population, with 1-year survival of roughly 90% and median survival of about 12 to 13 years. Without

advanced therapies including heart transplantation and LVAD, the mortality rate is very high, with about 50% of patients living with chronic inotrope therapy likely to die within 1 year.[29] In current clinical care models, the onus for identifying patients with advanced HF and completing a timely HF specialist referral is on the clinicians caring for these patients. This is a challenging task for several reasons including the at-times-subtle signs and symptoms of advanced HF, the slow and progressive nature of the illness, competing responsibilities for busy providers, and the fact that the majority of HF patients are cared for by non-HF specialists. Thus, the optimal timing for specialist referral, termed the "golden window" where patients show signs and symptoms of progressive disease before irreversible, multifocal end-organ damage sets in, is often missed.[2]

We learned several lessons in completing this work. First, developing and implementing augmented intelligence-enabled workflows within the EHR requires a high-performing, diverse, and adaptable team with support from leadership. Incorporating the variety of roles, experiences, and viewpoints on caring for patients is key to designing a truly human-centered process. Further, upfront investment in evaluation metrics and dashboards is critical to real-time improvement. In our experience, the workflow for successful model deployment may require more resources than creating the model itself. Additionally, EHR upgrades can impact embedded workflows when using cognitive computing platforms and must be considered. Lastly, promising results will take time, require rigorous evaluation and re-evaluation, necessitate ongoing resource allocation, and need continuous quality improvement.

Several of these points are related to specific design choices for the creation of a tool that will integrate well into a human-centered workflow and requires collaborating with key stakeholders early in the process, evaluating progress repeatedly along the way, and iterating on the initial decisions. This is in line with the recent work by both Li et al[12] and Marwaha et al[30] discussing the science of implementing or delivering AI tools within complex health systems. Li et al[12] highlight the flaws in the pipeline of AI generation at present, namely that these technologies are created in isolation from the environment of their application, focusing instead on narrow prediction tasks. Thus, they often suffer in the implementation stages, when the focus shifts to how and when to use these tools and how they will affect the lives of patients, as opposed to their prediction metrics using well-crafted data sets.[12] Marwaha et al[30] highlight the various resources needed for the implementation of digital health tools, pointing out the diverse sets of individuals needed, and the long-term considerations once these tools are up and running.

STUDY STRENGTHS AND LIMITATIONS.

There are several strengths to our approach. First, this HER-embedded augmented intelligence-enabled workflow addresses a key gap in clinical care affecting patients with advanced HF by using data science and cloud computing, and it is currently running live in clinical operations. Second, the workflow itself allows us to capture the patients who have an incorrect prediction by the ML model, thus generating data that can be used to retrain the model in the future to improve its performance. Lastly, our pipeline focuses on increasing referrals with less burden on clinicians recognizing those with advanced HF, but it maintains

clinician involvement to ensure model output is appropriate, thus relying on cooperation between humans and machines to make better decisions.

STUDY LIMITATIONS.

Our work, however, has several limitations. The agreement between our coordinator and the model for stage D was lower than the positive predictive value of the model during training (50.3% vs 60%) but sufficiently high to facilitate screening for a disease with an estimated prevalence of 5% to 15% in the overall HF population.[3] Furthermore, our method for assigning disease labels relied on clinicians documenting “stage D” in clinical notes. This may have resulted in finding patients who are already past the “golden window” or mislabeled due to taking the phrase “stage D” out of context. This is reflected in lower model performance in the physician-reviewed set vs the test set, as the test set relied on text search for label generation where the physician-reviewed set was labeled based on prospective physician review. This was necessary as there is no gold standard data set for this diagnosis; however, we plan to evaluate several alternative approaches in the future. Additionally, our workflow still requires a significant amount of manual chart review, an aspect we hope will improve with incremental progress in model performance. Some of our model inputs had a moderate to high rate of missingness, which may have impacted model performance. Lastly, as of February 15, 2022, due to small numbers of patients from racial and ethnic minority groups in our population, we are unable to report reliable model performance by subgroup. As the workflow continues and more patients are reviewed, our team will monitor performance in racial and minority subgroups to ensure that the workflow is improving access to care to all patients, especially those impacted by structural racism.[31]

Our initial work has identified several important next steps, both in model development and in advancing the clinical workflow, that we will use to “control” the process within the DMAIC framework. In terms of model development, an important task will be to improve the label-generation process. We anticipate evaluating additionally methodologies to identify patients with advanced HF through the EHR, using previously published approaches including that by Dunlay et al.[3] Additionally, we will test if the incorporation of features from unstructured data in clinical notes and use of longitudinal changes in vital signs, laboratory work, and cardiac function measured by echocardiography can improve model performance. Lastly, we anticipate validating the next versions of our model at external sites to ensure reproducibility of model performance, as well as to improve our capabilities for real-time model performance monitoring and updating.

As for the workflow, we plan to further risk stratify those classified as stage C using the MARKER-HF risk model, an externally validated ML model to predict mortality in HF using readily available and low-cost tests. We are also in the process of a mixed-method evaluation of the augmented intelligence-enabled workflow, which includes interviews with key stakeholders and a more detailed analysis of the outcomes of this workflow, including evaluation for bias against minority groups and other vulnerable populations. In many ways, creation and optimization of the workflow was the more challenging and resource-intensive part. Now that the workflow is operational, resources will be devoted to developing a

high-performing model to identify patients in the “golden window” who may benefit from expedited evaluation by a HF specialist for advanced therapies.

CONCLUSIONS

We utilized augmented intelligence to create an ML model and embed the results within a clinical workflow in EHR to streamline identification of patients with advanced HF and who need referral for evaluation by HF specialists, as patients are routinely referred too late in the process once they are too ill to qualify for life-saving therapies. Our work presented here focuses on both the first operational version of our model as well as the creation of a human-centered workflow that allows the use of this tool successfully in practice. We learned several key lessons, namely the prominence of design thinking when creating a clinical pipeline which uses an AI tool and the need for real-time monitoring of model performance and workflow metrics. We also highlight avenues for further research. Augmented intelligence-enabled workflows have the potential to improve clinical care and benefit patients in need; however, more work is needed on developing best practices to ensure prospective implementation has the best chance for success.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Jonna Peterson (librarian) for her efforts in completing a thorough literature review to support this work. The authors also thank the leadership and staff from Northwestern Medicine Information Systems and the Bluhm Cardiovascular Institutes who supported this program.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Dr Ahmad was supported by grants from the National Institutes of Health/National Heart, Lung, and Blood Institute (K23HL155970) and the American Heart Association (AHA number 856917); and has received consulting fees from Teladoc Livongo and Pfizer unrelated to this manuscript. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ABBREVIATIONS AND ACRONYMS

AI	artificial intelligence
EF	ejection fraction
EHR	electronic health record
HF	heart failure
LVAD	left ventricular assist device
ML	machine learning

REFERENCES

1. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. Circulation. 2022;145:e153–e639. [PubMed: 35078371]
2. Morris AA, Khazanie P, Drazner MH, et al. Guidance for timely and appropriate referral of patients with advanced heart failure: a scientific statement from the American Heart Association. Circulation. 2021;144(15):e238–e250. [PubMed: 34503343]
3. Dunlay SM, Roger VL, Killian JM, et al. Advanced heart failure epidemiology and outcomes. J Am Coll Cardiol HF. 2021;9:722–732.
4. Kalogeropoulos AP, Samman-Tahhan A, Hedley JS, et al. Progression to stage D heart failure among outpatients with stage C heart failure and reduced ejection fraction. J Am Coll Cardiol HF. 2017;5:528–537.
5. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure. J Am Coll Cardiol. 2022;79(17):e263–e421. [PubMed: 35379503]
6. Fang JC, Ewald GA, Allen LA, et al. Advanced (stage D) heart failure: a statement from the Heart Failure Society of America Guidelines Committee. J Card Fail. 2015;21:519–534. [PubMed: 25953697]
7. Khush KK, Cherikh WS, Chambers DC, et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: thirty-sixth adult heart transplantation report - 2019; focus theme: donor and recipient size match. J Heart Lung Transplant. 2019;38:1056–1066. [PubMed: 31548031]
8. Zimpfer D, Fiane AE, Larbalestier R, et al. Long-term survival of patients with advanced heart failure receiving an left ventricular assist device intended as a bridge to transplantation. Circ Heart Fail. 2020;13:e006252.
9. Herr JJ, Ravichandran A, Sheikh FH, et al. Patient referral practices to advanced heart failure centers. J Card Fail. 2021;27:1251–1259. [PubMed: 34146684]
10. Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature. 2020;580:252–256. [PubMed: 32269341]
11. Tomasev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature. 2019;572:116–119. [PubMed: 31367026]
12. Li RC, Asch SM, Shah NH. Developing a delivery science for artificial intelligence in healthcare. NPJ Digit Med. 2020;3:107. [PubMed: 32885053]
13. Maddox TM, Albert NM, Borden WB, et al. The learning healthcare system and cardiovascular care: a scientific statement from the American Heart Association. Circulation. 2017;135:e826–e857. [PubMed: 28254835]
14. Bazoukis G, Hall J, Loscalzo J, Antman EM, Fuster V, Armoundas AA. The inclusion of augmented intelligence in medicine: a framework for successful implementation. Cell Rep Med. 2022;3:100485.
15. Norgeot B, Quer G, Beaulieu-Jones BK, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med. 2020;26:1320–1324. [PubMed: 32908275]
16. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure. J Am Coll Cardiol. 2013;62(16):e147–e239. [PubMed: 23747642]
17. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle heart failure model: prediction of survival in heart failure. Circulation. 2006;113:1424–1433. [PubMed: 16534009]
18. Sartipy U, Dahlström U, Edner M, Lund LH. Predicting survival in heart failure: validation of the MAGGIC heart failure risk score in 51,043 patients from the Swedish Heart Failure Registry. Eur J Heart Fail. 2014;16:173–179. [PubMed: 24464911]
19. Adler ED, Voors AA, Klein L, et al. Improving risk prediction in heart failure using machine learning. Eur J Heart Fail. 2020;22:139–147. [PubMed: 31721391]
20. Deo RC. Machine learning in medicine. Circulation. 2015;132:1920–1930. [PubMed: 26572668]

21. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature. 2020;585:357–362. [PubMed: 32939066]
22. McKinney W. Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference. 2010. Accessed September 20, 2022. https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf
23. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–2830.
24. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng. 2007;9:90–95.
25. Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. arXiv. Posted online May 27, 2016. http://arxiv.org/abs/1605.08695
26. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. arXiv. Posted online August 13, 2016. http://arxiv.org/abs/1603.02754
27. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc; 2017:3149–3157.
28. Ahmed S. Integrating DMAIC approach of Lean Six Sigma and theory of constraints toward quality improvement in healthcare. Rev Environ Health. 2019;34:427–434. [PubMed: 31314742]
29. Hashim T, Sanam K, Revilla-Martinez M, et al. Clinical characteristics and outcomes of intravenous inotropic therapy in advanced heart failure. Circ Heart Fail. 2015;8:880–886. [PubMed: 26179184]
30. Marwaha JS, Landman AB, Brat GA, Dunn T, Gordon WJ. Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. NPJ Digit Med. 2022;5:13. [PubMed: 35087160]
31. Piña IL, Jimenez S, Lewis Eldrin F, et al. Race and ethnicity in heart failure. J Am Coll Cardiol. 2021;78:2589–2598. [PubMed: 34887145]

PERSPECTIVES

COMPETENCY IN PRACTICE-BASED LEARNING AND IMPROVEMENT:

ML and AI tools can be integrated into practice to help clinicians make accurate and timely diagnoses, incorporating rapidly changing clinical values as well as changes in guideline recommendations.

COMPETENCY IN SYSTEMS-BASED PRACTICE:

Integrating health informatics tools into practice requires a reliance on diverse, multitalented teams that bring unique perspectives to patient care. Each of these team members plays a valuable role and can aid in providing the best patient care.

TRANSLATIONAL OUTLOOK:

Augmented intelligence-enabled workflows have the potential to increase the efficiency and timeliness of clinical care and improve patient outcomes. However, significant work needs to be done to develop best practices to implement ML tools and facilitate ongoing, prospective evaluation of their use over time. Objective evaluation of these tools in the form of randomized controlled trials is needed.

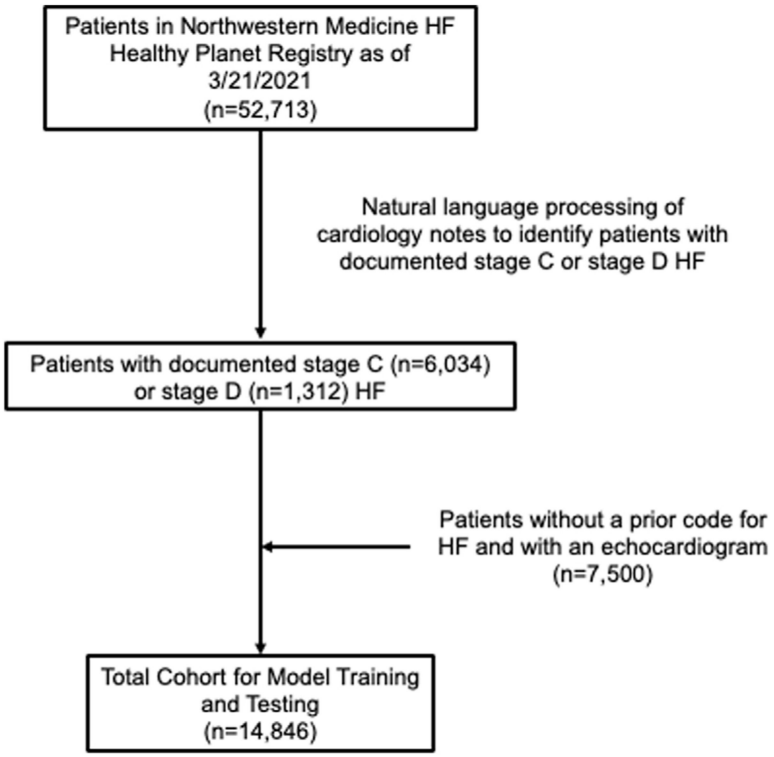


FIGURE 1. Cohort Generation

The schematic is shown for the generation of the cohort used to train the machine learning algorithm. HF = heart failure.

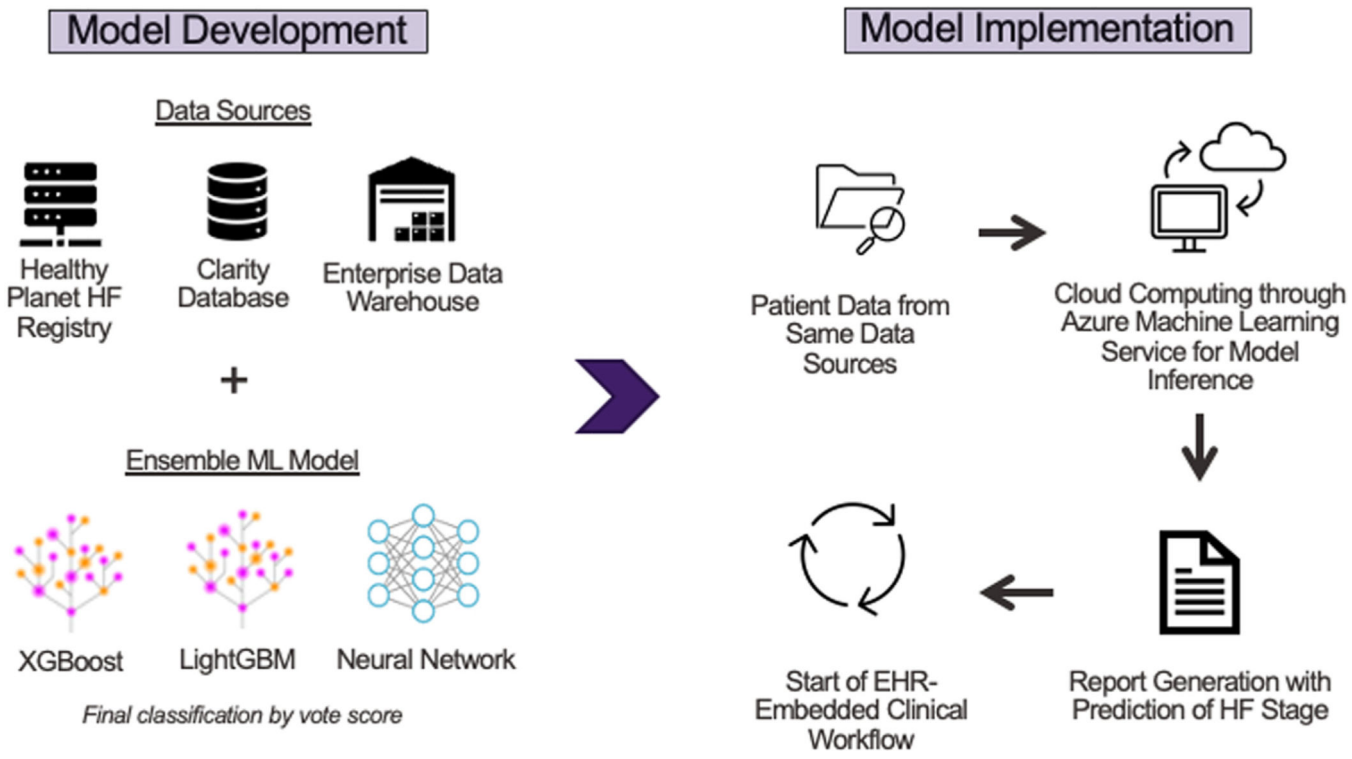


FIGURE 2. Model Development and Implementation

A description of the model development and implementation using cloud computing via Microsoft Azure. EHR = electronic health record; HF = heart failure.

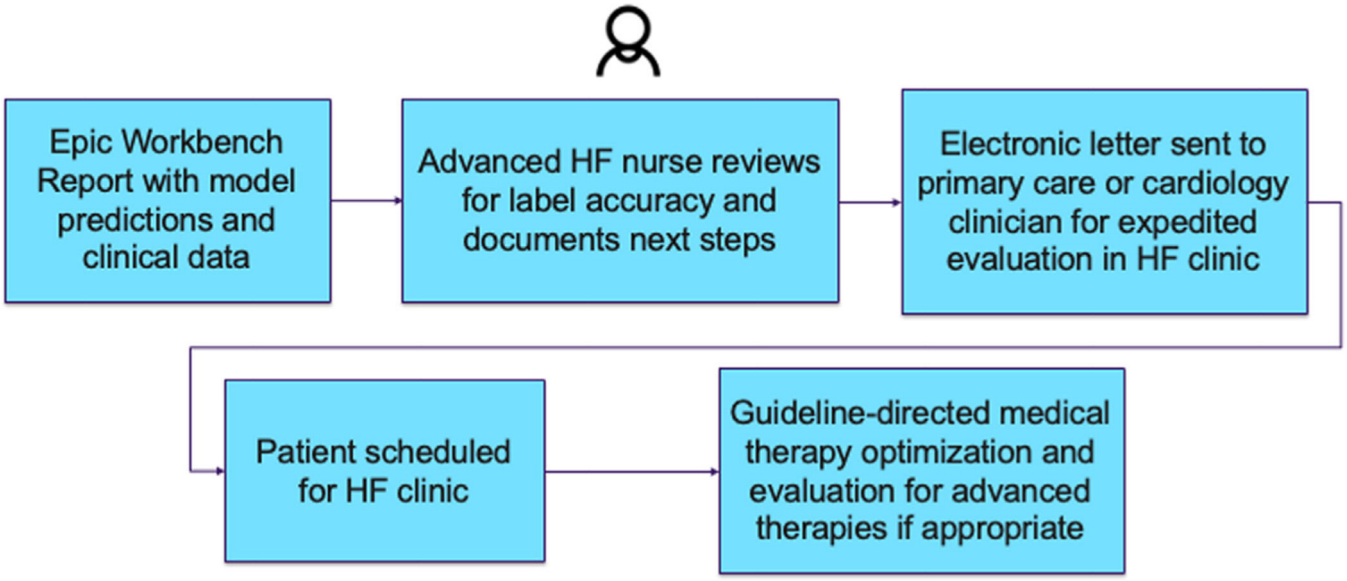
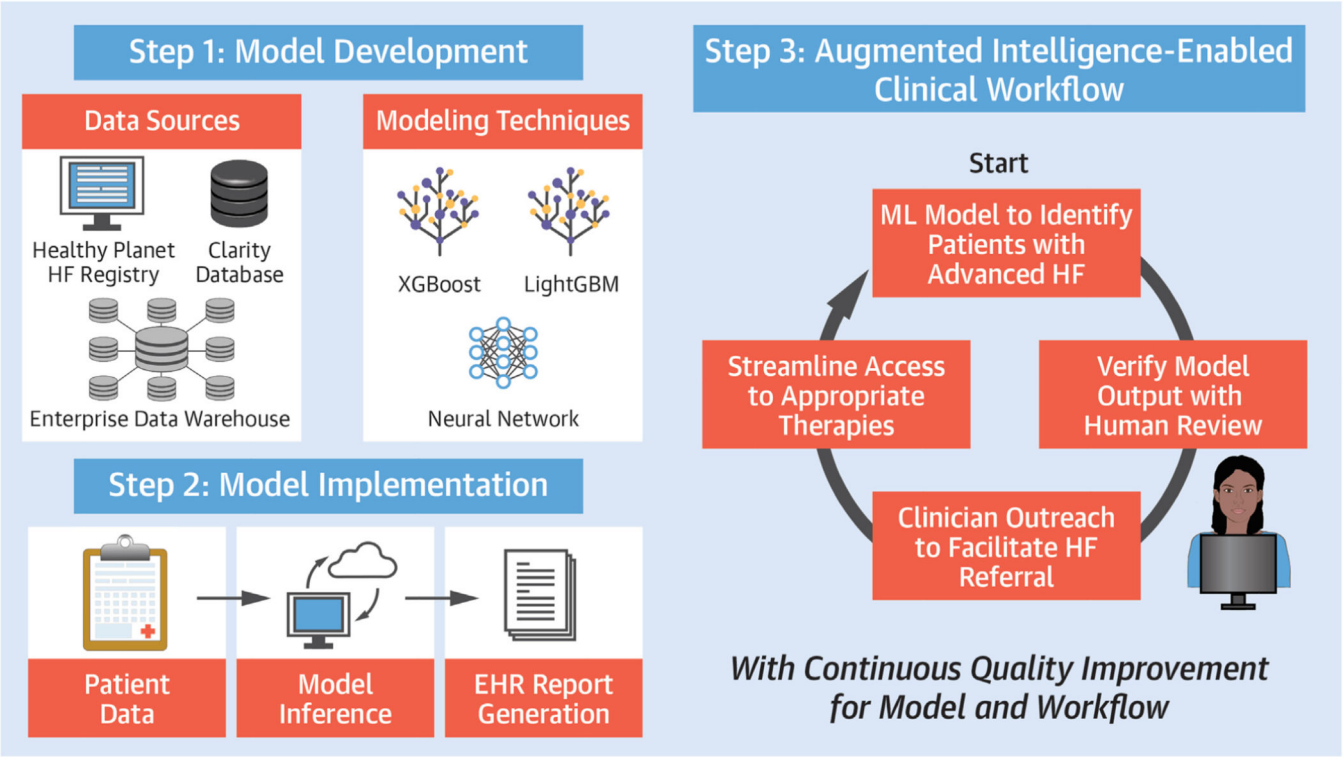


FIGURE 3. Augmented Intelligence-Enabled Clinical Workflow

A description of the augmented intelligence-enabled clinical workflow within the electronic health record. HF = heart failure.



CENTRAL ILLUSTRATION. The Use of Augmented Intelligence to Identify Patients With Stage D Heart Failure

A depiction of the steps involved in data acquisition, creation of the machine learning model, and the augmented intelligence-based clinical workflow utilizing the model is shown. Patients for training set generation were identified within the Healthy Planet Heart Failure Registry within epic. Data were collected from an enterprise data warehouse associated with a large, integrated health system, as well as from the electronic health record itself. An ensemble machine learning model was created, consisting of 2 gradient boosting trees as well as a feedforward neural network, with the final classification made by vote score. At the time of inference, the model predicts whether a patient has stage D HF and that prediction is embedded within a workflow in the electronic health record. This is subsequently reviewed by a nurse coordinator, and clinician outreach for appropriate patients is performed to streamline access to heart failure clinic for further evaluation. The labels generated by the nurse coordinator are saved for future iterations of model training. EHR = electronic health record; HF = heart failure; ML = machine learning.

TABLE 1

Composite Feature List

Clinical characteristics
Gender[a,b]
BMI[b]
Current smoker[b]
Vital signs
Systolic BP[a,b]
Diastolic BP[c]
Medications
Beta-blocker[a,b]
ACE/ARB[a,b]
Angiotensin receptor-neprilysin inhibitor[d]
SGLT2 inhibitor[d]
Mineralocorticoid receptor antagonist[a]
Thiazide diuretic[a]
Loop diuretic[a]
Statin[a]
Xanthine oxidase inhibitor[a]
Inotrope[a]
Diagnostic testing
White blood cell count[c]
RDW[c]
Hemoglobin[a,c]
Lymphocytes[a]
Platelet count[c]
Sodium[a]
BUN[a,b]
Creatinine[b,c]
Albumin[c]
BNP[d]
Hemoglobin A1c[b]
Total cholesterol[a]
Ejection fraction[a,b]

[a] Seattle Heart Failure Model.

[b] MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure Risk Calculator).

[c] MARKER-HF (Machine learning Assessment of RisK and EaRly mortality in Heart Failure).

[d] Expert consensus. Features were treated as categorical. Thresholds for continuous variables were selected based on established guidelines or team consensus.

ARB = angiotensin II receptor blocker; ACE = angiotensin-converting enzyme inhibitor; BMI = body mass index; BNP = brain natriuretic peptide; BP = blood pressure; BUN = blood-urea nitrogen; RDW = red cell distribution width; SGLT2 = sodium glucose cotransporter-2.

TABLE 2

Patient Characteristics for Model Development, Testing, and Validation Data Set

	Full Cohort (N = 14,830)	Not HF (n = 7,485)	Stage C (n = 6,033)	Stage D (n = 1,312)
Demographics				
Age, y	65.3 ± 16.6	61.5 ± 16.8	70.7 ± 14.9	61.8 ± 15.2
Female	6,953 (46.9)	3,096 (52.1)	2,665 (44.1)	373 (28.4)
Race/ethnicity				
White	9,743 (64.7)	5,329 (71.2)	3,572 (59.2)	678 (51.7)
Black	2,681 (17.8)	801 (10.7)	1,472 (24.4)	362 (27.6)
Hispanic	1,205 (8.0)	636 (8.5)	440 (7.3)	106 (8.1)
Other	1,430 (9.5)	719 (9.6)	549 (9.1)	165 (12.6)
Comorbidities				
Hypertension	10,841 (73.1)	4,521 (60.4)	5,249 (87.0)	1,051 (80.1)
Hyperlipidemia	9,298 (62.7)	4,072 (54.4	4,464 (74.0	792 (60.4)
Diabetes	4,894 (33.0)	1,579 (21.1)	2,703 (44.8)	628 (47.9)
Obstructive sleep apnea	3,811 (25.7)	1,272 (17.0)	2,099 (34.8	457 (34.8)
Atrial fibrillation	4,983 (33.6)	1,183 (15.8)	2,992 (49.6	816 (62.2)
Chronic kidney disease or dialysis	6,644 (44.8)	1,744 (23.3)	3,837 (63.6)	1,077 (82.1)
Ischemic heart disease	7,163 (48.3)	1,916 (25.6)	4,199 (70.0)	1,056 (80.5)
ICD present	1,157 (7.8)	97 (1.3)	712 (11.8)	353 (26.9)
	(N = 14,846)	**(n = 7,500)**	**(n = 6,034)**	**(n = 1,312)**
Clinical characteristics				
BMI, kg/m^2	29.6 ± 7.8	29.0 ± 7.0	30.6 ± 8.5	28.2 ± 7.2
SBP, mmHg	122.8 ± 20.1	126.6 ± 18.2	121.2 ± 20.6	108.3 ± 20.5
DBP, mmHg	71.0 ± 12.1	73.3 ± 11.4	69.2 ± 12.1	67.1 ± 13.4
LVEF, %	54.5 ± 15.7	62.4 ± 7.3	47.6 ± 16.3	36.4 ± 20.2
Medication use				
ACE/ARB/ARNI	1,867 (12.4)	270 (3.6)	1,271 (21.1)	359 (23.4)
Beta blocker	4,397 (29.2)	,695 (22.6)	2,204 (36.6)	499 (32.5)
MRA	2,741 (18.2)	278 (3.7)	1,843 (30.6)	668 (43.5)

	Full Cohort (N = 14,830)	Not HF (n = 7,485)	Stage C (n = 6,033)	Stage D (n = 1,312)
SGLT2i	226 (1.5)	68 (0.9)	139 (2.3)	29 (1.9)
Inotrope use within 1 y	1,114 (7.4)	270 (3.6)	361 (6.0)	558 (36.3)

Values are mean ± SD or n (%). Demographic and comorbidity data were not available for 16 patients.

ACE = angiotensin-converting enzyme inhibitor; ARB = angiotensin II receptor blocker; ARNI = angiotensin receptor-neprilysin inhibitor; BMI = body mass index; DBP = diastolic blood pressure; HF = heart failure; ICD = implantable cardioverter-defibrillator; LVEF = left ventricular ejection fraction; MRA = mineralocorticoid receptor antagonist; SBP = systolic blood pressure; SD = standard deviation; SGLT2i = sodium/glucose cotransporter-2 inhibitor.

TABLE 3

Model Performance in Test Set and Physician-Reviewed Set

	PPV	Sensitivity	F1 Score	Accuracy	N
Test set					
Not HF	0.88	0.89	0.88		1,500
Stage C	0.78	0.85	0.81		1,189
Stage D	0.74	0.43	0.54		303
Accuracy				0.83	
					Total = 2,992
Physician-reviewed set					
Not HF	-	-	-		-
Stage C	0.79	0.91	0.85		76
Stage D	0.60	0.25	0.35		24
Accuracy				0.75	
					Total = 100

HF = heart failure; PPV = positive predictive value.

TABLE 4

Coordinator Agreement and Documented “Next Steps” From April 1, 2021, to February 15, 2022

Model Performance (n = 416)	Coordinator Agreement
Stage C (n = 122)	63.1%
Stage D (n = 294)	50.3%
“Next Steps” (n = 401)	**Count**
Consider evaluation in Advanced Heart Failure New Access Clinic	56
Review in 3 mo	58
Follow-up in HF clinic[a]	77
Consider evaluation in general cardiology clinic	3
No additional recommendations at this time	207

[a]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.

HF = heart failure.