



HHS Public Access

Author manuscript

JACC Heart Fail. Author manuscript; available in PMC 2023 December 04.

Published in final edited form as:

JACC Heart Fail. 2023 July ; 11(7): 852–854. doi:10.1016/j.jchf.2023.02.012.

Natural Language Processing for Adjudication of Heart Failure in the Electronic Health Record

Jonathan W. Cunningham, MD, MPH[†],

Pulkit Singh, BA[†],

Christopher Reeder, PhD,

Emily S. Lau, MD, MPH,

Shaan Khurshid, MD, MPH,

Xin Wang, MPH,

Patrick T. Ellinor, MD, PhD,

Steven A. Lubitz, MD, MPH,

Puneet Batra, PhD,

Jennifer E. Ho, MD*

Beth Israel Deaconess Medical Center, 330 Brookline Avenue, CLS 945, Boston, Massachusetts 02215, USA

Electronic health record (EHR) data repositories have the potential to accelerate research in heart failure (HF) epidemiology, biology, and policy because of large sample size, long-term follow-up, and granular clinical details. However, imprecise ascertainment of clinical endpoints such as HF hospitalization by International Classification of Disease (ICD) diagnosis codes has limited research using EHR data sets. The gold standard for endpoint adjudication, physician review of medical records, is not feasible at the scale of EHR cohorts. Innovation in natural language processing (NLP) now enables training of complex text models with few labeled examples. In this study, we trained and validated an NLP model for automated adjudication of HF hospitalization from raw discharge summary text.

We studied hospitalizations in the Community Care Cohort Project (C3PO), an EHR cohort of 520,868 patients (mean age: 48 years, 61% women) receiving longitudinal primary care between 2001 and 2018 at Mass General Brigham.¹ Analysis of C3PO was approved by the Mass General Brigham Institutional Review Board. Possible HF hospitalizations were identified by ICD codes in the primary or secondary position as defined by the American Heart Association “Get with the Guidelines” registry. Board-certified cardiologists adjudicated medical records to determine whether hospitalizations met the Standardized Data Collection for Cardiovascular Trials Initiative criteria for HF: 1

* jho@bidmc.harvard.edu .

[†]Dr Cunningham and Ms Singh contributed equally to this paper.

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. For more information, visit the [Author Center](#).

symptom, 2 signs, and 1 treatment. From these adjudications, training (n = 1,268), test (n = 214), and validation (n = 452) sets were created. Candidate NLP models based on 6 publicly available pretrained transformer-based models were trained using discharge summary text and cardiologist labels in the training set, and their performance was compared in the test set. A model developed from Clinical-Longformer,² which features clinical text pretraining and accommodates longer text (average precision [AP]: 0.88), outperformed regular Longformer (AP: 0.79) and models with clinical pretraining but shorter attention windows (maximum AP: 0.79), and was therefore selected for validation.

In the held-out validation set of 452 hospitalizations with HF ICD codes, 162 (36%) were confirmed as HF by cardiologists conducting full-chart review. Common reasons for negative adjudication were iatrogenic volume overload following surgery or critical illness, or patients with chronic HF admitted for reasons other than acute HF. Relative to the cardiologist gold standard, adding the NLP model to ICD codes improved accuracy (AP: 0.88; area under the receiver-operating characteristic curve [AUROC]: 0.93). The ICD+NLP model achieved precision of 81% (95% CI: 74%-86%) and recall of 82% (95% CI: 75%-88%), which exceeded strategies using structured data alone: HF ICD code alone (precision: 36% [95% CI: 31%-40%]; recall 100%; AUROC: 0.50), HF ICD code in primary position (precision: 60% [95% CI: 53%-66%]; recall 78% [95% CI: 71%-84%]; AUROC: 0.74 [95% CI: 0.70-0.78]), and the combination of HF ICD code in the primary position and intravenous loop diuretic administration (precision: 72% [95% CI: 63%-79%]; recall 61% [95% CI: 53%-69%]; AUROC: 0.74 [95% CI: 0.70-0.78]) (Figure 1). The NLP model performed consistently across age, sex, and calendar year subgroups, and among the 211 patients with primary diagnosis HF ICD code (AUROC: 0.89 [95% CI: 0.84-0.93]).

Among hospitalizations without HF ICD codes, true HF was rare: just 1 of 250 randomly selected cases were adjudicated as HF by cardiologist reviewers. The NLP model also adjudicated this case as HF. We applied the NLP model to 100,000 hospitalizations without HF ICD codes; the model adjudicated 895 (0.9%) as HF. On cardiologist review of a sample of these ICD⁻/NLP⁺ cases, 18 of 40 (45%) were true HF. Thus, among hospitalizations without HF ICD codes, the NLP model performs well, identifying cases markedly enriched for true HF. However, because HF is very rare in the absence of HF ICD codes, implementing the NLP model together with ICD codes minimizes false-positive results.

We applied the ICD+NLP model in the full C3PO cohort. 21,611 patients had 1 hospitalization with an HF ICD code, of which 9,690 (45%) were deemed HF by NLP. Compared with patients with HF by ICD codes but not NLP, patients with ICD+NLP HF more commonly underwent natriuretic peptide testing (82% vs 47%), and had higher natriuretic peptide levels (median N-terminal pro-brain natriuretic peptide 4,006 vs 1,250 pg/mL), as expected for acute HF. Age- and sex-stratified rates of incident HF in C3PO were compared to a published pooled analysis from the Framingham Heart Study, MESA (Multi-Ethnic Study of Atherosclerosis), and PREVENT (Prevention of Renal and Vascular Endstage Disease).³ Rates of incident HF defined by the combination of ICD codes and NLP in C3PO were consistent with these physician-adjudicated community studies; HF ICD codes alone overestimated community-based incidence rates. For example, for men 75

years of age, incident HF rates per 1,000 patient-years were 22.2 (95% CI: 18.9-25.5) in the pooled cohorts, 18.1 (95% CI: 17.1-19.1) in C3PO by ICD+NLP, and 38.5 (95% CI: 37.0-40.0) in C3PO by ICD codes alone.

Several limitations must be noted. Generalizability to other EHR systems remains to be studied. The NLP model was applied to hospitalizations with HF ICD codes to minimize false-positive results, and may miss some HF cases without codes. The NLP model interprets only inpatient discharge summaries; including admission, consultation, or clinic notes may improve performance and enable identification of outpatient HF. Median N-terminal pro-brain natriuretic peptide in NLP-negative patients is subject to ascertainment bias because more than half of patients had missing values.

In conclusion, fewer than half of hospitalizations with HF ICD codes in any position, and fewer than two-thirds with HF ICD code as primary diagnosis, met accepted criteria for HF hospitalization by cardiologist review. We developed a transformer-based NLP model that interprets raw discharge summary text and improved the accuracy of HF adjudications when added to ICD codes. These results highlight the potential of NLP as a scalable approach to improve the accuracy of clinical endpoints in data sets too large for manual adjudication, including EHR cohort studies and pragmatic randomized trials.

Acknowledgments

This work was supported by Bayer AG. Dr Cunningham has received the KL2/Harvard Catalyst Medical Research Investigator Training program and the American College of Cardiology/Association of Black Cardiologists Bristol Myers Squibb Research Fellowship; and has received consulting fees from Roche Diagnostics, Occlutech, and KCK. Dr Lau has received grants from the National Institutes of Health (NIH) (K23-HL159243) and the American Heart Association (AHA) (853922); and has received consulting fees from Roche Diagnostics. Dr Ellinor has received grants from NIH (1R01HL092577, K24HL105780), AHA (18SFRN34110082), Foundation Leducq (14CVD01), and MAESTRIA (965286); has received sponsored research support from Bayer AG and IBM Health; and has served on advisory boards or consulted for Bayer AG, Quest Diagnostics, MyoKardia, and Novartis. Dr Lubitz has received NIH grants 1R01HL139731, R01HL157635, and AHA 18SFRN34250007 during this work; has been a full-time employee of Novartis as of July 18, 2022; has received sponsored research support from Bristol Myers Squibb, Pfizer, Boehringer Ingelheim, Fitbit, Medtronic, Premier, and IBM; and has received consulting fees from Bristol Myers Squibb, Pfizer, Blackstone Life Sciences, and Invitae. Dr Batra has been a full-time employee of Flagship Pioneering as of January 2023; has received sponsored research support from Bayer AG and IBM; and has received consulting fees from Novartis and Prometheus Biosciences. Dr Ho has received grants from the NIH (R01 HL134893, R01 HL140224, R01HL160003, and K24 HL153669); and has received sponsored research support from Bayer AG.

REFERENCES

1. Khurshid S, Reeder C, Harrington LX, et al. Cohort design and natural Language processing to reduce bias in electronic health records research. *NPJ Digit Med.* 2022;5:47. [PubMed: 35396454]
2. Li Y, Wehbe RM, Ahmad FS, Wang H, Luo Y. Clinical-Longformer and Clinical-BigBird: transformers for long clinical sequences. Accessed July 1, 2022. <https://arxiv.org/abs/2201.11838>
3. Tromp J, Paniagua SMA, Lau ES, et al. Age dependent associations of risk factors with heart failure: pooled population based cohort study. *BMJ.* 2021;372:n461. [PubMed: 33758001]

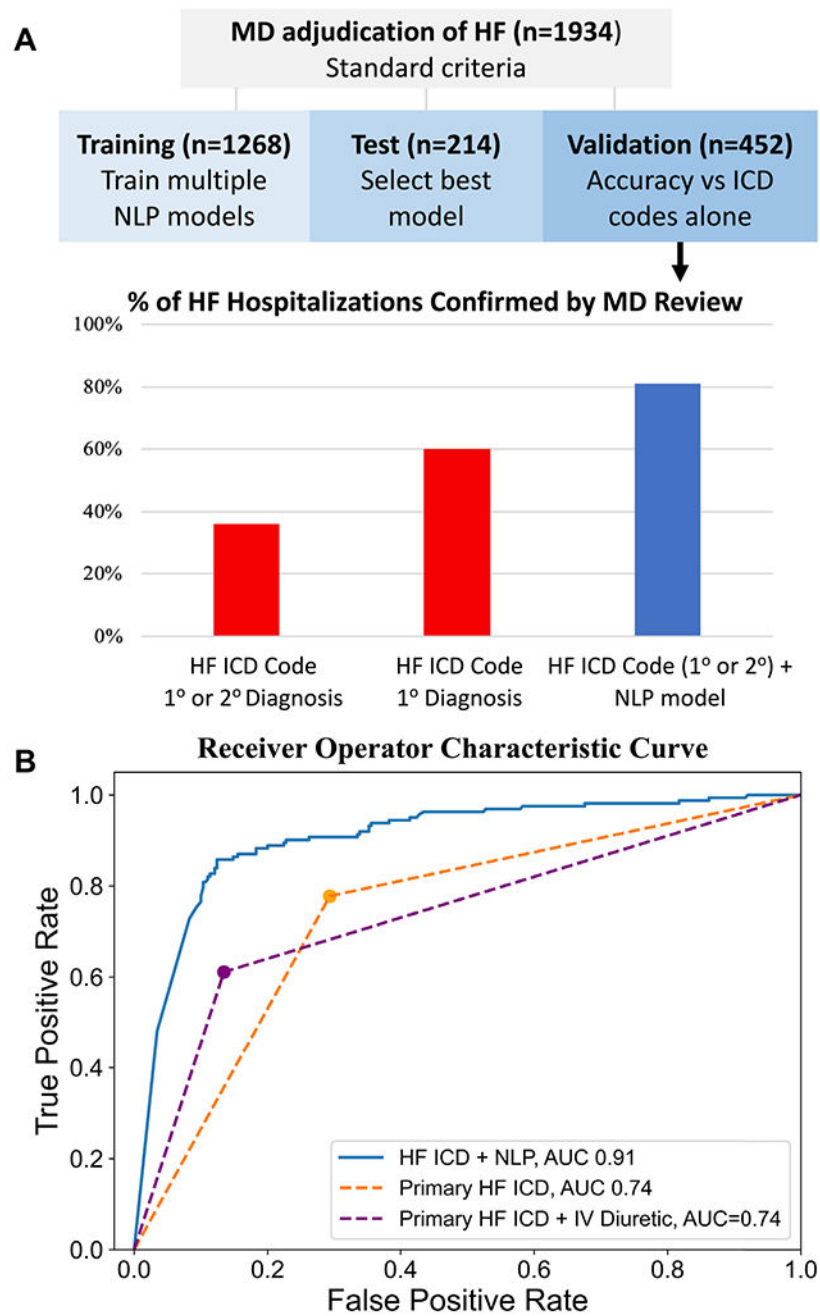


FIGURE 1. Development and Accuracy of the Natural Language Processing Model
(A) Adding natural language processing to International Classification of Diseases (ICD) codes increases the proportion of heart failure (HF) hospitalizations confirmed by gold-standard physician review. **(B)** Natural language processing (NLP) model has greater discrimination than definitions of HF based on ICD code in the primary position alone or together with intravenous (IV) loop diuretic. AUC = area under the curve; MD = medical doctor.