Artificial Intelligence-assisted automated heart failure detection and classification from electronic health records

Oo, Mon Myat; Gao, Chuang; Cole, Christian; Hummel, Yoran M.; Guignard-Duff, Magalie; Jefferson, Emily

Published in:
ESC Heart Failure

DOI:
10.1002/ehf2.14828

Publication date:
2024

ESC HEART FAILURE
ESC Heart Failure (2024)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.14828

ORIGINAL ARTICLE

Artificial intelligence-assisted automated heart failure detection and classification from electronic health records

Mon Myat Oo[1], Chuang Gao[2], Christian Cole[2,3], Yoran Hummel[4], Magalie Guignard-Duff[2], Emily Jefferson[2,3], James Hare[4], Adriaan A. Voors[5], Rudolf A. de Boer[6], Carolyn S.P. Lam[7,8], Ify R. Mordi[1], Jasper Tromp[7,8,9]* and Chim C. Lang[1]*

[1]*Division of Molecular and Clinical Medicine, School of Medicine, University of Dundee, Dundee, UK;* [2]*Health Informatics Centre, School of Medicine, University of Dundee, Dundee, UK;* [3]*Division of Population Health and Genomics, School of Medicine, University of Dundee, Dundee, UK;* [4]*Us2.ai, Singapore;* [5]*University Medical Centre Groningen, Groningen, The Netherlands;* [6]*Department of Cardiology, Erasmus University Medical Centre, Rotterdam, The Netherlands;* [7]*National Heart Centre Singapore, Singapore;* [8]*Duke-NUS Medical School, Singapore; and* [9]*Saw Swee Hock School of Public Health, National University of Singapore & National University Health System, Singapore*

Abstract

Aims Electronic health records (EHR) linked to Digital Imaging and Communications in Medicine (DICOM), biological specimens, and deep learning (DL) algorithms could potentially improve patient care through automated case detection and surveillance. We hypothesized that by applying keyword searches to routinely stored EHR, in conjunction with AI-powered automated reading of DICOM echocardiography images and analysing biomarkers from routinely stored plasma samples, we were able to identify heart failure (HF) patients.

Methods and results We used EHR data between 1993 and 2021 from Tayside and Fife (~20% of the Scottish population). We implemented a keyword search strategy complemented by filtering based on International Classification of Diseases (ICD) codes and prescription data to EHR data set. We then applied DL for the automated interpretation of echocardiographic DICOM images. These methods were then integrated with the analysis of routinely stored plasma samples to identify and categorize patients into HF with reduced ejection fraction (HFrEF), HF with preserved ejection fraction (HFpEF), and controls without HF. The final diagnosis was verified through a manual review of medical records, measured natriuretic peptides in stored blood samples, and by comparing clinical outcomes among groups. In our study, we selected the patient cohort through an algorithmic workflow. This process started with 60 850 EHR data and resulted in a final cohort of 578 patients, divided into 186 controls, 236 with HFpEF, and 156 with HFrEF, after excluding individuals with mismatched data or significant valvular heart disease. The analysis of baseline characteristics revealed that compared with controls, patients with HFrEF and HFpEF were generally older, had higher BMI, and showed a greater prevalence of co-morbidities such as diabetes, COPD, and CKD. Echocardiographic analysis, enhanced by DL, provided high coverage, and detailed insights into cardiac function, showing significant differences in parameters such as left ventricular diameter, ejection fraction, and myocardial strain among the groups. Clinical outcomes highlighted a higher risk of hospitalization and mortality for HF patients compared with controls, with particularly elevated risk ratios for both HFrEF and HFpEF groups. The concordance between the algorithmic selection of patients and manual validation demonstrated high accuracy, supporting the effectiveness of our approach in identifying and classifying HF subtypes, which could significantly impact future HF diagnosis and management strategies.

Conclusions Our study highlights the feasibility of combining keyword searches in EHR, DL automated echocardiographic interpretation, and biobank resources to identify HF subtypes.

Keywords Deep learning algorithms; Electronic health record data; Epidemiology; Heart failure; Preserved ejection fraction; Validation

Received: 3 April 2024; Accepted: 8 April 2024
**Correspondence to: Chim Lang, Division of Molecular and Clinical Medicine, School of Medicine, University of Dundee, Dundee DD1 9SY, UK. Email: monmyat85@gmail.com, moo001@dundee.ac.uk; Jasper Tromp, National Heart Centre Singapore, Singapore. Email: jasper_tromp@nus.edu.sg*
Mon Myat Oo, Chuang Gao, Jasper Tromp, and Chim C. Lang have equal contribution,

Introduction

Heart failure (HF) is a highly prevalent yet underdiagnosed clinical syndrome with high mortality and morbidity.[1] Echocardiography is a foundational investigation to diagnose HF and differentiate HF with reduced (HFrEF) and preserved (HFpEF) ejection fraction.[2,3]

Electronic health records (EHRs) are an increasingly high-quality data source that can be used for the creation of pragmatic cohort studies[4] disease surveillance, case selection for pragmatic randomized clinical trials (RCTs),[5] and quality improvement initiatives.[6] The quality and quantity of EHR data are expanding and increasingly include EHR-linked biobanks[7–9] and EHR-linked imaging data.[10]

Validation of HF diagnosis using imaging data is critical to unlocking the full potential of EHR data for HF surveillance and research. However, retrospective selection and analysis of echocardiographic images are time-consuming and prohibitively expensive. Furthermore, clinically reported echocardiographic values are commonly incomplete and can be of differing quality due to time constraints in routine clinical practice and differences in clinical indications and readers. Deep learning (DL) algorithms can automate the reading of echocardiographic images.[11,12] Combined with EHR data, these algorithms can help identify HF and classify patients according to their left ventricular ejection fraction (LVEF) subtypes.[13,14] Measurement of plasma biomarkers, like N-terminal-pro-brain natriuretic peptide (NT-proBNP), in bio-banked blood samples can further enhance diagnosis and offer opportunities for pragmatic patient selection for RCTs or translational cohort studies.

This study aimed to identify and classify patients with HF from routinely stored EHR data, linked to Scottish Health Research Register (SHARE)[9] bioresource and echocardiographic data collected from the Tayside and Fife region of Scotland using a deep learning-based approach. By leveraging keyword searches in Electronic Health Records (EHRs), automated reading of DICOM echocardiography images through AI, and analysis of biomarkers from routinely stored plasma samples, we aim to refine the identification and differentiation process of HF patients, particularly distinguishing between HF with preserved ejection fraction (HFpEF) and HFrEF.

Method

Data sources

This study used EHR data from the National Health Service (NHS) via the Health Informatic Centre (HIC). HIC is a third-party data provider with NHS data custodians to host collated EHRs. De-identified, anonymized, and linkable clinical data was provided for the scope of the study in a Trusted Research Environment.[15] Data included demographic characteristics, co-morbidities, laboratory test results, community-dispensed prescriptions, DICOM images, hospital admission records, and mortality data.[15] HIC is a partner with SHARE, the Scottish Health Research Register, and Biobank. NHS Research Scotland supports SHARE, Universities in Scotland, and the Chief Scientists Office.[9] SHARE Biobank collects and stores consented blood samples from registered individuals via an interception method with NHS blood sciences laboratories that have been taken for NHS diagnostic purposes but are surplus to these requirements. These samples were collected before disposal as clinical waste. Access to the anonymized clinical datasets was administered by HIC at the University of Dundee using established protocols. For this study, access to SHARE was approved by Research Ethic Committee in Tayside (Study reference number 119).

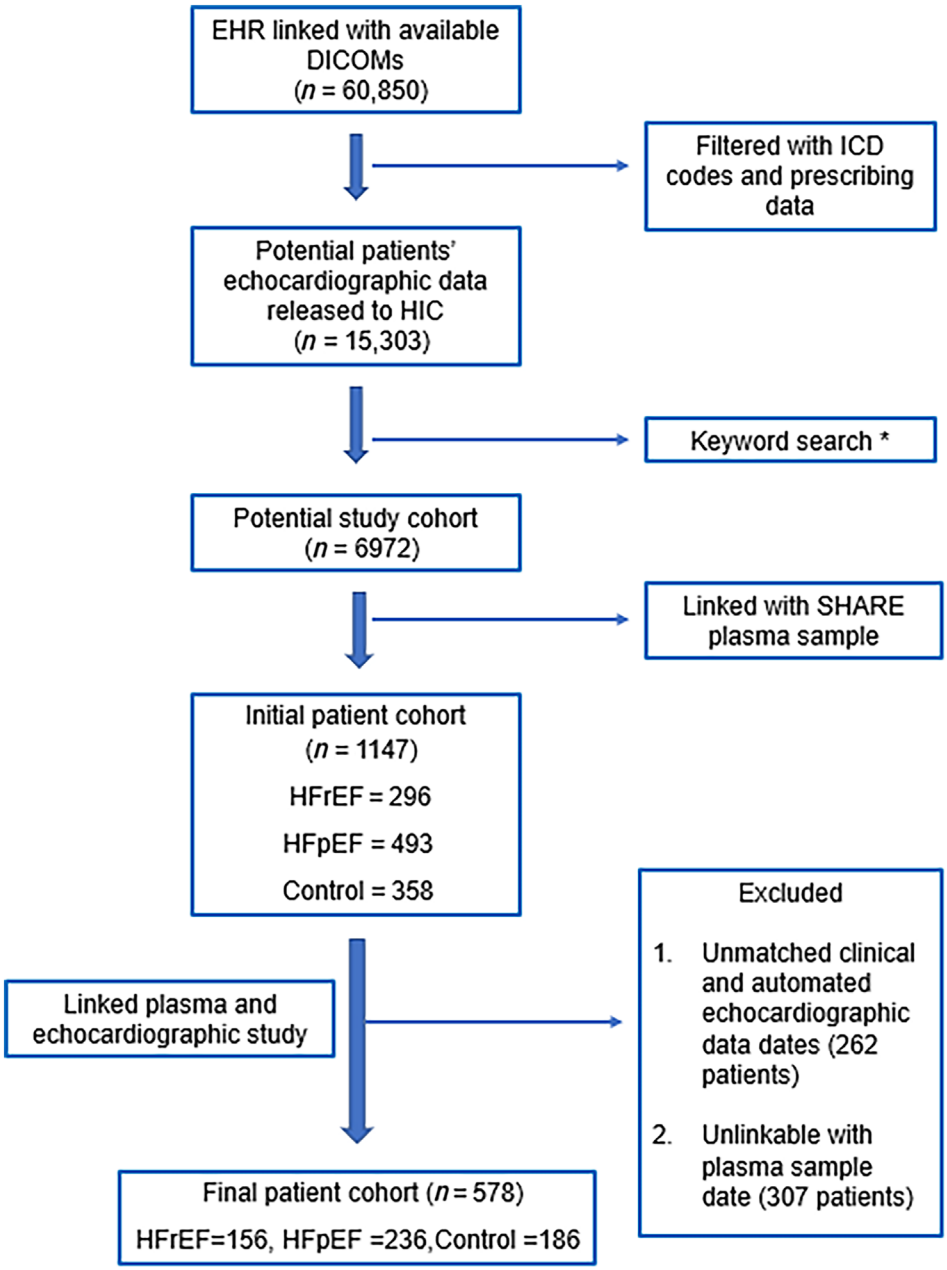
Study definitions and clinical outcomes

Patients with type 2 diabetes mellitus (T2DM) were identified by linkage to the DARTS register that identifies patients with T2DM based on having T2DM according to primary and/or secondary care data sources.[16,17] The combination of electronic prescribing records and ICD-10 coding for atrial fibrillation (AF) identification. This method has a positive predictive value of 97%, as previously described.[18] We used ICD-10 coding for coronary artery disease identification. Chronic obstructive pulmonary disease (COPD) was defined by linkage to the community COPD registry (the Tayside Allergy and Respiratory Disease Information System)[19] and those previously hospitalized for COPD based on ICD-10 coding.[20] For the identification of chronic kidney disease (CKD), we used the NHS CKD diagnosis[21] and National Institute for Health and Care Excellence (NICE) guideline on CKD assessment and management.[22] This means patients with eGFR less than 60 mL/min were considered as CKD in our study. Our defined clinical outcomes were all-cause mortality and HF related hospitalization.

Algorithm and study cohort selection

Figure 1 illustrates the patient cohort identification and data linkage workflow. First, we retained patients with echocardiogram records and available corresponding DICOMs. We filtered medical records to identify HF cases and non-HF controls based on the International Classification of Diseases-10 (ICD-10) code of HF (I50.x) or a history of continuous (>6 months) prescriptions of furosemide. HF patients were further categorized based on left ventricular ejection fraction (LVEF): those with LVEF below 40% were designated as having heart failure with reduced ejection fraction (HFrEF), whereas individuals with LVEF above 50% were identified as

Figure 1 Patient cohort identification and data linkage utilizing available EHR, DICOM, and plasma samples. *n* = number of records. *Keyword search as per *Table S1*.



having heart failure with preserved ejection fraction (HFpEF). Controls were defined as individuals without a history of HF hospitalization, furosemide prescription, or HF ICD-10 code, and with an echocardiogram indicating normal or preserved LVEF. Participants with HF or controls with evidence of valvular heart disease, such as severe aortic and/or mitral stenosis, were excluded. We conducted a manual screening combination with keyword search to clinical data set released to HIC. *Table S1* shows the exact keywords used to select controls and patients with HFrEF or HFpEF.

We manually reviewed and validated clinical and echocardiographic parameters for the final cohort selection. We filtered out different study cohorts according to LVEF cut-off values according to the hierarchy of (i) Simpson's LVEF available; (ii) If Simpson's LVEF was not available, an LVEF was derived from 'eyeball' LVEF in the clinical report. Patients were classified as HFrEF if they had an LVEF $\leq$40%, HFmrEF if LVEF was $>$40% but below 50%, and HFpEF if the LVEF was $\geq$50%, in keeping with clinical guidelines.[2,3] For descriptive left ventricular function, we mapped LVEF terminology with LVEF nu-

merical values and different types of heart failure as previously described and shown in *Table 1*.[23] Because we identified few patients with HFmrEF and patients with HFmrEF are closer to those with HFrEF, we merged patients with HFmrEF and HFrEF and considered patients with LVEF <50% as HFrEF.[24]

In the subsequent phase, the selection was refined to include only those participants who had available plasma samples from the SHARE database. These samples were then allocated for biomarker analysis, and the corresponding DICOM files were processed through automated analysis. Individuals lacking plasma samples were omitted from further analysis.

In the third step of our methodology, depicted in *Figure 2*, we synchronized the plasma samples with echocardiographic data. Given that heart failure (HF) patients may undergo multiple echocardiograms at different times, there was a need to methodically match these with the corresponding plasma samples collected for the SHARE project. To achieve this, we adopted a hierarchical approach to identify the echocardiographic study that was most temporally proximate to the plasma sample collection. Our primary criterion was to select the echocardiogram conducted closest to, and preferably before, the date of plasma collection. If no such echocardiogram was available, we then selected the nearest study performed after the plasma collection, provided it was within a 180-day timeframe. Any echocardiographic data collected more than 180 days following the plasma sample were not considered for analysis. This careful matching process resulted in a refined cohort of 578 patients, which was divided into 186 controls, 236 patients diagnosed with heart failure with preserved ejection fraction (HFpEF), and 156 patients with heart failure with reduced ejection fraction (HFrEF), following the exclusion of any unmatched plasma and echocardiographic data.

Table 1 Mapping LVEF numerical values from report terminology

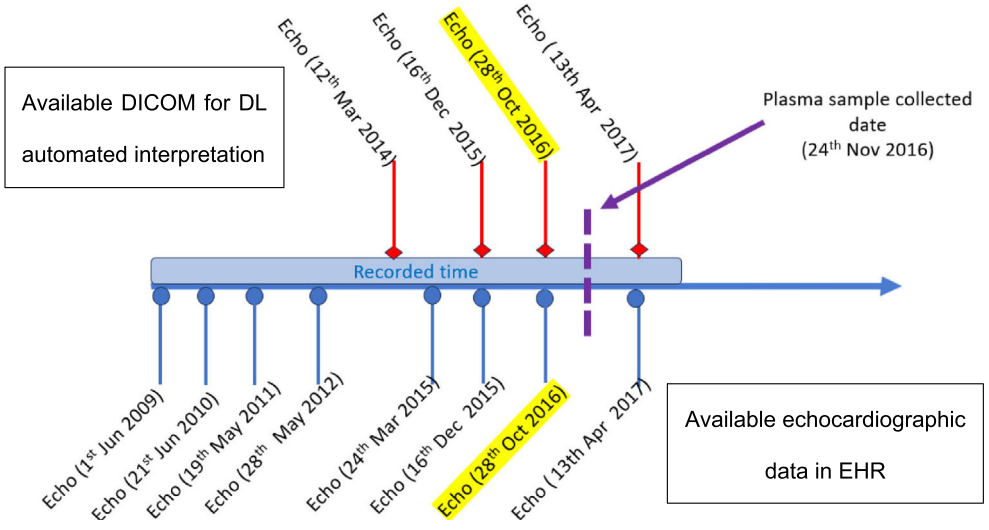
LVSD report terminology	LVEF assignment	HF subtype
'Severe'	25%	HFrEF
'Moderate to severe'	30%	HFrEF
'Moderate'	35%	HFrEF
'Mild to moderate'	40%	HFmrEF[a]
'Mild'	45%	HFmrEF[a]
'Normal'	55%	HFpEF

[a]Heart failure with mildly reduced ejection fraction.

Analysis of Digital Imaging and Communications in Medicine data and biomarker levels

The training and external validation of the deep learning workflow to analyse the DICOM images have been extensively described elsewhere.[11,25] The fully automated DL-based workflow uses various supervised and unsupervised convolutional neural networks (CNN) to classify echocardiographic 2D videos and doppler modalities according to their respective views (e.g. A4C or A2C) or modality without the need for human intervention. The workflow has two separate pipelines for 2D videos and Doppler modalities, which use a CNN to produce segmentation masks and/or heatmaps to annotate 2D images and Doppler modalities. These algorithms were previously validated in various external cohorts[11,25] and against core-lab measurements,[25] leading to the United States Food and Drug Administration (FDA) approval of the workflow as artificial intelligence (AI) based decision support tool.[26]

Figure 2 Exemplar patient regarding selection and linkage of echocardiographic data, DICOM files and SHARE plasma samples.



Plasma concentrations of NT-proBNP were quantified utilizing a commercially available electro chemiluminescent sandwich immunoassay (Roche Modular E170, Roche Diagnostics, Mannheim, Germany), employing plasma samples sourced from SHARE.[27] Moreover, exploratory biomarker analyses performed with plasma samples derived from the study cohort will be detailed in a forthcoming scholarly article.

Statistical analysis

Continuous variables were summarized as median ± standard deviation, while categorical variable was represented a n with a percentages. We used the one-way analysis of variance, Wilcoxon rank sum test, analysis of variance (ANOVA), or chi-squared test to test differences between controls, HFrEF and HFpEF, depending on the nature and distribution (i.e. normal vs. non-normal, categorical vs. continuous) of the variable. A Cox proportional hazards regression model was used to explore the relationship between HF diagnosis and clinical outcomes. We performed Cohen's kappa (κ) analysis to determine the agreement of the final diagnosis between manual validation and the proposed algorithm.

Results

Baseline characteristics

Table 2 shows the participants' baseline characteristics. Compared with controls, patients with HFrEF and HFpEF were older, had a higher body mass index (BMI), and had a higher prevalence of diabetes, COPD, and CKD. Compared with HFrEF, patients with HFpEF were more likely women and had less diabetes, AF, and CAD than those with HFrEF. Patients with HF had higher concentrations of NT-proBNP than controls. NT-proBNP concentrations were higher in HFrEF than HFpEF; 93% of patients with HFrEF were on beta-blockers, 82% on ACE inhibitors ad 31% on spironolactone compared with 77%, 68%, and 19% of patients with HFpEF on beta-blockers, ACE inhibitors, and spironolactone, respectively.

Table 2 Demographic and clinical characteristics of the study groups

	Control	HFrEF	HFpEF	*P*-value[a]
Patients (*n*)	186	156	236	
Demography				
Age, years (median ± SD)	59.5 ± 18	74 ± 10	77.5 ± 13	<0.001
Sex, female (%)	61%	37%	61%	<0.001
BMI[b]	25.8 ± 6.4	27.0 ± 0.4[c]	28.9 ± 5.8	0.87
Medical history (%)				
Diabetes	9%	39%	31%	<0.001
AF	NA	40%	30%	<0.001
CAD	NA	46%	29%	<0.001
COPD	32%	53%	56%	<0.001
CKD	3%	42%	39%	<0.001
Laboratory				
Serum creatinine	67 ± 15.3	93 ± 98.1	83 ± 88.4	<0.001
eGFR (MDRD)	94 ± 27.7	63.6 ± 23.6	66.9 ± 27.2	<0.001
Potassium	4.3 ± 0.4	4.4 ± 0.4	4.3 ± 0.5	0.68
Sodium	139 ± 3.2	139 ± 3.3	139 ± 3.7	0.36
NT-proBNP (pg/mL)	77.2 [32.8–155.8]	1651 ± 6302.4	686.3 ± 4925.3	<0.001
Medication (%)				
Furosemide	NA	90%	83%	<0.001
Bumetanide	NA	13%	11%	<0.001
Spironolactone	2%	31%	19%	<0.001
Eplerenone	NA	17%	3%	<0.001
Aspirin	24%	68%	59%	<0.001
Statin	25%	78%	69%	<0.001
SGLT2i	1%	2%	2%	0.48
Beta-blocker	31%	93%	77%	<0.001
ARB	14%	33%	33%	<0.001
ACEI	20%	82%	68%	<0.001

ACEI, angiotensin-converting enzyme inhibitors; AF, atrial fibrillation; ARB, angiotensin receptor blockers; BMI, body mass index; CAD, coronary artery disease; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; eGFR, estimated glomerular filtration rate; MDRD, Modification of Diet in Renal Disease; NT-proBNP, N terminal pro B type natriuretic peptide; SGLT2i, sodium-glucose co-transporter-2 inhibitors.

[a]*P*-value is calculated using either ANOVA or chi-squared test. Chi-squared test is used when the data is categorical, for example, sex, comorbidity, and medication.

[b]Based on 43 patients with weight and height available.

[c]Based on 2 patients.

Table 3 Echocardiographic characteristics of the study groups (parameters derived from the Us2.ai software analyses) (median ± SD)

	Us2.ai features	Coverage	Control	HFpEF	HFrEF	*P*-value
LV dimensions (linear)	IVSd (mm)	93%	8.7 ± 1.8	9.8 ± 2.3	9.8 ± 2.2	<0.001
	LVIDd (mm)	93%	45.8 ± 5.7	44.9 ± 6.1	51.5 ± 7.3	<0.001
	LVPWd (mm)	92%	8.4 ± 1.2	9.4 ± 1.6	9.3 ± 1.6	<0.001
	LVIDs (mm)	91%	30.9 ± 5.7	32.2 ± 7.3	42.8 ± 8.5	<0.001
	RWT (mm)	93%	0.4 ± 0.1	0.4 ± 0.1	0.4 ± 0.1	<0.001
	LV mass (g)	93%	123 ± 40.4	150.8 ± 43.4	183 ± 50.4	<0.001
LV (volume and) systolic function	LVEF (any single plane) (%)	92%	62.7 ± 9.3	59.6 ± 11	41.6 ± 15.5	<0.001
	LVEF biplane (%)	73%	62.5 ± 8.5	59.8 ± 10.2	41.4 ± 14.3	<0.001
	LVEDV biplane (mL)	74%	79.6 ± 23.9	81.9 ± 30.7	109.6 ± 42.7	<0.001
	LVESV biplane (mL)	73%	30.3 ± 12.3	33.7 ± 17.7	63.7 ± 36.1	<0.001
LVGLS	LVGLS (overall)	60%	−21.2 ± 3.4	−18.4 ± 4.3	−12.1 ± 4.3	<0.001
	LVGLS (apical 3-chamber)	73%	−21.6 ± 5.1	−17.8 ± 6.5	−11.9 ± 5.7	<0.001
	LVGLS (apical 4-chamber)	91%	−20.9 ± 4.1	−19.1 ± 4.9	−12.2 ± 4.8	<0.001
	LVGLS (apical 2-chamber)	78%	−21.5 ± 4.4	−19 ± 5.7	−12.8 ± 5.1	<0.001
LV diastolic function	MV E (cm/s)	88%	71.2 ± 17.5	79 ± 33.7	88.5 ± 30.3	<0.001
	MV A (cm/s)	71%	69.7 ± 20.9	82.8 ± 28.4	73.7 ± 29.2	<0.001
	E/A ratio	70%	1 ± 0.4	0.8 ± 0.5	1 ± 0.6	0.059
	DT (ms)	76%	223.2 ± 46.6	209.8 ± 64.6	180.4 ± 65	<0.001
	e' septal (cm/s)	52%	7.4 ± 3	5.9 ± 2	4.8 ± 2.1	<0.001
	e' lateral (cm/s)	52%	10.6 ± 4	8.1 ± 2.6	7.7 ± 2.8	<0.001
	e' mean (cm/s)	47%	9 ± 3.1	7.2 ± 2	6.3 ± 2	<0.001
	E/e' mean (cm/s)	46%	7.7 ± 2.9	10.3 ± 4.9	12.6 ± 5.6	<0.001
	LAESV biplane (mL)	59%	35.9 ± 15.2	52.4 ± 27.2	59.3 ± 25.6	<0.001
RV function	TAPSE (mm)	75%	22.7 ± 5	21 ± 5.1	18.7 ± 4.7	<0.001
	RVFAC (%)	57%	36.2 ± 13	30.4 ± 13.2	26.3 ± 15.4	<0.001
	RVIDd (mm)	76%	33.4 ± 5.6	35.4 ± 6.3	36.2 ± 7.4	<0.001
	TR Vmax (m/s)	68%	2.3 ± 0.4	2.5 ± 0.6	2.7 ± 0.6	<0.001
	TR Pmax (mmHg)	73%	17.9 ± 6.6	21.9 ± 14.9	23.7 ± 9.1	<0.001
RA size	RAESV (mL)	80%	29.9 ± 18.6	38.1 ± 23.7	41.1 ± 33.6	<0.001
	RA area (cm^2/m^2)	72%	12.8 ± 3.8	15.6 ± 5	15.7 ± 6.4	<0.001

DT, deceleration time; E/A, mitral valve early diastolic velocity to late diastolic velocity ratio; E/e', ratio between early mitral inflow velocity and mitral annular early diastolic velocity; IVSd, interventricular septum thickness in diastole; LAESV, left atrial end-systolic volume; LV, left ventricle; LVGLS, left ventricle global longitudinal strain; LVEDV, LV end-diastolic volume; LVEF, left ventricular ejection fraction; LVESV, left ventricular end-systolic volume; LVIDd, left ventricular internal. Diameter at end-diastole; LVIDs, left ventricular internal diameter end-systole; LVMass, left ventricle mass; LVPWd, end-diastolic left ventricular posterior wall thickness; MV A, mitral valve late diastolic velocity; MV E, mitral valve early diastolic velocity; PASP, pulmonary arterial systolic pressure; RA area, right atrial area; RAESV, right atrial end-systolic volume; RV.FAC, right ventricular fractional area changes; RVIDd, right ventricular internal dimension at end-diastole; RWT, relative wall thickness; TAPSE, tricuspid annular plane systolic excursion; TR Pmax, tricuspid regurgitation maximum pressure; TR Vmax, tricuspid regurgitation maximum velocity.

Echocardiographic parameters

Table 3 shows the coverage and median values with interquartile range (IQR) of the echocardiographic parameters measured by the deep learning workflow. Coverage was generally high, ranging from 46% to 93%. Patients with HFrEF had a higher left ventricular diameter in diastole (LVIDd) than the control and HFpEF group. Patients with HFrEF had a lower LVEF than those with HFpEF and controls. Patients with HFrEF more often had an abnormal left ventricular global longitudinal strain (LVGLS) than those with HFpEF or controls. Left ventricular filling pressure index E/e' and left atrial end-systolic volume (LAESV) biplane values were higher in HFrEF and HFpEF compared with controls. Compared with controls, patients with HFpEF and HFrEF had greater LV mass, higher mitral E/e' ratio, higher pulmonary artery systolic pressure, more impaired LV strain and right ventricle (RV) dysfunction.

Table 4 shows the coverage of echocardiographic parameters for diagnosis of HFpEF according to ESC guidelines.[2] Deep learning analysis enhanced the availability of echocardiographic parameters used to aid in the diagnosis of HFpEF.

Clinical outcomes

In total, 119 patients died, and 116 patients were hospitalized for HF during a median follow-up time of 1089 days. Compared with 5% of controls, 37% of HFpEF and 58% of HFrEF experienced heart failure hospitalization or all cause death during follow-up. Compared with controls, patients with HFrEF (hazard ratio [HR]: 7.27, 95% CI 3.58–14.78) or HFpEF (HR 5.44, 95% CI 2.69–10.98) were at a higher risk of death. Patients with HFrEF (HR 16.13, 95% CI 8.12–32.07) or HFpEF (HR 9.2, 95% CI 4.63–18.28) were at a higher risk of HF hospitalization during follow-up than controls.

Table 4 Coverage of echocardiographic parameters for diagnosis of HFpEF according to ESC guideline by EHR data and Us2.ai software analyses

Echocardiographic parameters	Data available from EHR (percentage)	Data available from deep learning analyses (percentage)
LV mass	51.04	92.56[a]
Relative wall thickness	77.85	92.56
LA volume	0.00	58.65[b]
E/e' ratio at rest	0.00	45.50
TR velocity at rest	74.57	67.99
PASP	3.11	54

LV, left ventricle; RWT, relative wall thickness; LVMass, left ventricle mass; E/e', ratio between early mitral inflow velocity and mitral annular early diastolic velocity; LA volume, left atrial volume; TR Vmax, tricuspid regurgitation maximum velocity; PASP, pulmonary artery systolic pressure.
[a]As body surface area cannot be obtained from the DICOM image, the percentage shown here is LV mass.
[b]As body surface area cannot be obtained from the DICOM image, the percentage shown here is LA volume.

Proposed algorithm concordance with clinical records

We have taken 150 patients from our study cohort to test the agreement between patients identified using the proposed algorithm and diagnosis by manual validation. The positive predictive value was 86%, 100% sensitivity, 94% specificity with a kappa value of 0.891 was recorded in HFrEF patient group. The positive predictive value of the automated versus manual diagnosis of HFpEF was 80%, 100% sensitivity, 90% specificity with a kappa value of 0.842. A 100% concordance rate was noted in control group.

Discussion

In our research, we explored the integration of routinely stored EHRs with DL automated interpretation of DICOM and biobank resources to refine the identification and classification of heart failure (HF) subtypes. Through a detailed process beginning with a dataset of 15 000 patient records and narrowing down to a carefully curated cohort of 578 patients, our study underscores the critical role of data curation and the application of AI in enhancing the precision and relevance of health data analysis. This meticulous selection process, involving keyword searches, linkage with plasma samples, and exclusion criteria for data quality, enabled the effective utilization of deep learning (DL) algorithms in our study.

Despite the inherent challenges in curating HF diagnosis and subtypes from large EHR databases such as the often-limited availability of left ventricular ejection fraction (LVEF) data,[28–30] our approach demonstrates a novel method to overcome these obstacles. Patients can be classified based on their ICD-10 coding, but the specificity is limited. In a previous study, identifying patients with HFrEF or HFpEF based on ICD-10 coding only had a specificity of 63–68% for HFrEF and 86%–93% for HFpEF.[31] Advances in natural language processing (NLP) to extract HF subtypes have shown promise.[32,33] However, NLP algorithms might be challenging to validate externally and provide limited information on parameters other than LVEF. We highlighted the limitations of relying solely on ICD-10 coding for HF classification due to its limited specificity, and we addressed the potential of natural language processing (NLP) and AI in extracting more nuanced HF subtypes and echocardiographic parameters as using AI algorithms can significantly improve the availability of echocardiographic data in EHRs that can help diagnose HFpEF according to guidelines.[2,3] These findings suggest that AI algorithms coupled with EHR data can improve efficiency, reduce the time for patient selection for pragmatic clinical trials, and improve HF surveillance and early diagnosis across hospital systems.

Our research represents an advancement in the utilization of deep learning (DL) to automatically interpret echocardiographic DICOM images, thus streamlining the identification of patients with heart failure with reduced ejection fraction (HFrEF) and heart failure with preserved ejection fraction (HFpEF) within electronic health record (EHR) datasets. We previously demonstrated the use of these algorithms to automatically interpret systolic and diastolic parameters in five external cohorts.[11,25] In this study, we have applied a machine learning algorithm which automatically generated echocardiographic parameters resulting in >90% of patients having available LV linear dimensions, LV volume and systolic function. Our data also support the use of machine learning to help enhance HFpEF diagnosis. HFpEF is characterized by its heterogeneity, which extends to its definition. Over the years, HFpEF has previously been referred to as 'diastolic' heart failure, or heart failure with normal ejection fraction (HFnEF). However, diastolic dysfunction has been shown not to be unique to HFpEF, as evidence of diastolic dysfunction may also be found in systolic heart failure. Moreover, the term 'normal ejection fraction' is not correct either as some patients with HFpEF may in fact have supranormal function, such as in hypertrophic cardiomyopathy or cardiac amyloidosis. The updated European Society of Cardiology guidelines include the objective evidence of cardiac structural and/or functional abnormalities consistent with the presence of LV diastolic dysfunction or raised LV filling pressure or raised natriuretic peptides.[2] These structural and/or functional abnormalities help to define and diagnose HFpEF.

By leveraging AI, we not only facilitate the detection of these structural and functional cardiac anomalies but also enrich the dataset with critical parameters, such as relative wall thickness. This enhancement is crucial for diagnosing conditions like left ventricular hypertrophy and evaluating diastolic function. The comprehensive echocardiographic parameters we have made accessible, including left ventricular ejection fraction (LVEF), interventricular septal diameter (IVSd), LV mass, and indices of diastolic function, support a more re-

fined diagnosis of HFrEF or HFpEF in line with contemporary clinical guidelines.

Our approach has potential clinical implications, especially in the precision required in HFpEF clinical trials and the broader context of heart failure diagnosis and surveillance. The automation our DL algorithms provide not only makes the diagnosis process more efficient than traditional methods but also paves the way for identifying heart failure cohorts more pragmatically. When combined with biobank data, such as that from the SHARE project, our methods hold the promise of accelerating biomarker validation and fostering innovations in drug discovery for heart failure treatment.

Limitations

Our study has several limitations. Firstly, we have found that the final number of patients identified from our selection process was limited. The key limiting factor was that we had to have the most temporally proximate plasma sample to the echocardiogram conducted closest to, and preferably before, the date of plasma collection. We believe that this was a useful learning process for future record linkage studies that wish to link bioresources to echocardiography data. Second, our study was a retrospective study, which might have introduced selection bias. Thirdly, in our study process, we were not able to differentiate between patients with HFpEF and those with heart failure who had recovered EF. Finally, our study did not include direct records of visits with cardiologists or other specific clinical interactions that could serve as additional independent verification of HF diagnoses. It should be noted that our reliance on furosemide prescription data for more than 6 months as a criterion for identifying HF patients had been shown previously to reliably identify HF cases from EHRs.

Conclusions

Our study demonstrated the potential of integrating a keyword search of routinely stored electronic health records with AI-based machine learning algorithms and biobank resources for identifying heart failure (HF) subtypes. While our approach shows promise in enhancing the efficiency and speed of patient selection for pragmatic clinical trials, as well as in improving HF surveillance and early diagnosis across hospital systems, it is important to acknowledge the challenges posed by limited data availability. The effectiveness of this method, particularly in terms of early disease detection and its capability to increase the availability of biomarkers and echocardiographic parameters, remains an area for further exploration. Consequently, while our findings are encouraging, they also underscore the necessity of further studies to assess the diagnostic effectiveness of this approach fully.

Conflict of interest

The institution of Dr. De Boer has received research grants and/or fees from AstraZeneca, Abbott, Boehringer Ingelheim, Cardior Pharmaceuticals GmbH, Novo Nordisk, and Roche; Dr. de Boer has had speaker engagements with and/or received fees from and/or served on an advisory board for Abbott, AstraZeneca, Bristol Myers Squibb, Cardior Pharmaceuticals GmbH, NovoNordisk, and Roche, and received travel support from Abbott, Cardior Cardior Pharmaceuticals GmbH, and NovoNordisk. Carolyn SP Lam is supported by a Clinician Scientist Award from the National Medical Research Council of Singapore; has received research support from Bayer, NovoNordisk and Roche Diagnostics; has served as consultant or on the Advisory Board/Steering Committee/Executive Committee for Actelion, Alleviant Medical, Allysta Pharma, Amgen, AnaCardio AB, Applied Therapeutics, AstraZeneca, Bayer, Boehringer Ingelheim, Boston Scientific, Cytokinetics, Darma Inc., EchoNous Inc, Eli Lilly, Impulse Dynamics, Intellia Therapeutics, Ionis Pharmaceutical, Janssen Research & Development LLC, Medscape/WebMD Global LLC, Merck, Novartis, Novo Nordisk, Prosciento Inc, Radcliffe Group Ltd., Recardio Inc, ReCor Medical, Roche Diagnostics, Sanofi, Siemens Healthcare Diagnostics and Us2.ai; and serves as co-founder & non-executive director of Us2.ai. JT is supported by the National University of Singapore Start-up grant, the tier 1 grant from the ministry of education and the CS-IRG New Investigator Grant from the National Medical Research Council; has received consulting or speaker fees from Daiichi-Sankyo, Boehringer Ingelheim, Roche diagnostics and Us2.ai, owns patent US-10702247-B2 unrelated to the present work.

Funding

This study was conducted by a research collaboration between Roche Diagnostics International Ltd, us2.ai (Singapore), the University Medical Center Groningen and the University of Dundee. The research collaboration was funded by Roche Diagnostics International Ltd. [Correction added on 27 May 2024, after first online publication: Funding section has been added in this version.]

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1: keywords to identify controls without heart failure, patients with HFrEF and patients with HFpEF.

References

1. Lam CSP, Voors AA, de Boer RA, Solomon SD, van Veldhuisen DJ. Heart failure with preserved ejection fraction: from mechanisms to therapies. *Eur Heart J* 2018;**39**:2780-2792. doi:10.1093/eurheartj/ehy301
2. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, *et al*. Corrigendum to: 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2021;**42**:4901. doi:10.1093/eurheartj/ehab670
3. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, *et al*. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* 2022;**145**:e895-e1032.
4. Hébert HL, Shepherd B, Milburn K, Veluchamy A, Meng W, Carr F, *et al*. Cohort profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). *Int J Epidemiol* 2018;**47**:380-381j.
5. Jones WS, Mulder H, Wruck LM, Pencina MJ, Kripalani S, Muñoz D, *et al*. Comparative effectiveness of aspirin dosing in cardiovascular disease. *N Engl J Med* 2021;**384**:1981-1990.
6. Hemingway H, Asselbergs FW, Danesh J, Dobson R, Maniadakis N, Maggioni A, *et al*. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J* 2018;**39**:1481-1495.
7. Bowton E, Field JR, Wang S, Schildcrout JS, van Driest SL, Delaney JT, *et al*. Biobanks and electronic medical records: enabling cost-effective research. *Sci Transl Med* 2014;**6**:234cm3.
8. Allen NE, Sudlow C, Peakman T, Collins R, on behalf of UK Biobank. UK biobank data: come and get it. *Sci Transl Med* 2014;**6**:224ed4.
9. McKinstry B, Sullivan FM, Vasishta S, Armstrong R, Hanley J, Haughney J, *et al*. Cohort profile: the Scottish research register SHARE. A register of people interested in research participation linked to NHS data sets. *BMJ Open* 2017;**7**:e013351.
10. Nind T, Sutherland J, McAllister G, Hardy D, Hume A, MacLeod R, *et al*. An extensible big data software architecture managing a research resource of real-world clinical radiology data linked to other health data from the whole Scottish population. *Gigascience* 2020;**9**:
11. Tromp J, Seekings PJ, Hung CL, Iversen MB, Frost MJ, Ouwerkerk W, *et al*. Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *Lancet Digit Health* 2022;**4**:e46-e54.
12. Tromp J, Bauer D, Claggett BL, Frost M, Iversen MB, Prasad N, *et al*. A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *Nat Commun* 2022;**13**:6776.
13. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, *et al*. Fully automated echocardiogram interpretation in clinical practice. *Circulation* 2018;**138**:1623-1635.
14. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, *et al*. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 2020;**580**:252-256.
15. Gao C, McGilchrist M, Mumtaz S, Hall C, Anderson LA, Zurowski J, *et al*. A national network of safe havens: Scottish perspective. *J Med Internet Res* 2022;**24**:e31684.
16. Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, *et al*. The Diabetes Audit and Research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO collaboration. *BMJ* 1997;**315**:524-528.
17. (MFT), M.U.N.F.T. Diagnosis of DM. 2019; Available from: https://mft.nhs.uk/the-trust/other-departments/laboratory-medicine/biochemistry/specimen-requirements-and-reference-ranges/reporting-of-results/diagnosis-of-diabetes-mellitus/. Accessed 30 January 2024
18. Moore C. Enhancing the identification of atrial fibrillation using prescribing records in electronic medical record research. *European Journal of Arrythmia & Electrophysiology e-journal* 2022;**8**:abstr62.
19. Dicker AJ, Crichton ML, Pumphrey EG, Cassidy AJ, Suarez-Cuartin G, Sibila O, *et al*. Neutrophil extracellular traps are associated with disease severity and microbiota diversity in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 2018;**141**:117-127. doi:10.1016/j.jaci.2017.04.022
20. NHS. NHS Data Model and Dictionary; Chronic obstructive pulmonary disease. Available from: https://www.datadictionary.nhs.uk/Covid19PRA/COPD.html. Accessed 21 December 2023
21. NHS. Chronic Kidney Disease; Diagnosis. Available from: https://www.nhs.uk/conditions/kidney-disease/diagnosis/. Accessed 15 November 2023
22. Forbes A, Gallagher H. Chronic kidney disease in adults: assessment and management. *Clin Med (Lond)* 2020;**20**:128-132.
23. Voors AA, Anker SD, Cleland JG, Dickstein K, Filippatos G, van der Harst P, *et al*. A systems BIOlogy study to TAilored treatment in chronic heart failure: rationale, design, and baseline characteristics of BIOSTAT-CHF. *Eur J Heart Fail* 2016;**18**:716-726.
24. Tromp J, Khan MAF, Mentz RJ, O'Connor CM, Metra M, Dittrich HC, *et al*. Biomarker profiles of acute heart failure patients with a mid-range ejection fraction. *JACC Heart Fail* 2017;**5**:507-517.
25. Tromp, J.E.A. A prospective validation of a deep learning-based automated workflow for the interpretation of the echocardiogram, in nature communications. 2022.
26. US2.AI. FDA clears fully automated cardiac ultrasound solution to measure 2D and Doppler. 2022; Available from: https://us2.ai/. Accessed 15 November 2023
27. Suthahar N, Meijers WC, Ho JE, Gansevoort RT, Voors AA, van der Meer P, *et al*. Sex-specific associations of obesity and N-terminal pro-B-type natriuretic peptide levels in the general population. *Eur J Heart Fail* 2018;**20**:1205-1214. doi:10.1002/ejhf.1209
28. Stålhammar J, Stern L, Linder R, Sherman S, Parikh R, Ariely R, *et al*. The burden of preserved ejection fraction heart failure in a real-world Swedish patient population. *J Med Econ* 2014;**17**:43-51. doi:10.3111/13696998.2013.848808
29. Boman K, Lindmark K, Stålhammar J, Olofsson M, Costa-Scharplatz M, Fonseca AF, *et al*. Healthcare resource utilisation and costs associated with a heart failure diagnosis: a retrospective, population-based cohort study in Sweden. *BMJ Open* 2021;**11**:e053806. doi:10.1136/bmjopen-2021-053806
30. Mahesri M, Chin K, Kumar A, Barve A, Studer R, Lahoz R, *et al*. External validation of a claims-based model to predict left ventricular ejection fraction class in patients with heart failure. *PLoS ONE* 2021;**16**:e0252903. doi:10.1371/journal.pone.0252903
31. Goyal P, Bose B, Creber RM, Krishnan U, Yang M, Brady J, *et al*. Performance of electronic health record diagnosis codes for ambulatory heart failure encounters. *J Card Fail* 2020;**26**:1060-1066. doi:10.1016/j.cardfail.2020.07.015
32. Jonnalagadda SR, Adupa AK, Garg RP, Corona-Cox J, Shah SJ. Text mining of the electronic health record: an information extraction approach for automated identification and subphenotyping of HFpEF patients for clinical trials. *J Cardiovasc Transl Res* 2017;**10**:313-321. doi:10.1007/s12265-017-9752-2
33. Patel YR, Robbins JM, Kurgansky KE, Imran T, Orkaby AR, McLean RR, *et al*. Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records. *BMC Cardiovasc Disord* 2018;**18**:128. doi:10.1186/s12872-018-0866-5