ESC European Society of Cardiology

European Journal of Heart Failure (2019) **21**, 74–85
doi:10.1002/ejhf.1333

RESEARCH ARTICLE

Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy

Maja Cikes[1†], Sergio Sanchez-Martinez[2†], Brian Claggett[3], Nicolas Duchateau[4], Gemma Piella[2], Constantine Butakoff[2], Anne Catherine Pouleur[5], Dorit Knappe[6], Tor Biering-Sørensen[3,7], Valentina Kutyifa[8], Arthur Moss[8], Kenneth Stein[9], Scott D. Solomon[3*†], and Bart Bijnens[2,10†]

[1]Department of Cardiovascular Diseases, University of Zagreb School of Medicine, and University Hospital Center Zagreb, Zagreb, Croatia; [2]Department of Information and Communication Technologies, University Pompeu Fabra, Barcelona, Spain; [3]Brigham and Women's Hospital, Boston, MA, USA; [4]Creatis, CNRS UMR5220, INSERM U1206, Université Lyon 1, France; [5]Division of Cardiology, Cliniques Saint-Luc UCL, Brussels, Belgium; [6]University Heart Center Hamburg, Hamburg, Germany; [7]Herlev & Gentofte Hospital - Copenhagen University, Copenhagen, Denmark; [8]University of Rochester, Rochester, NY, USA; [9]Boston Scientific, Minneapolis, MN, USA; and [10]ICREA, Barcelona, Spain

Received 30 May 2018; revised 30 August 2018; accepted 11 September 2018; online publish-ahead-of-print 17 October 2018

Aims We tested the hypothesis that a machine learning (ML) algorithm utilizing both complex echocardiographic data and clinical parameters could be used to phenogroup a heart failure (HF) cohort and identify patients with beneficial response to cardiac resynchronization therapy (CRT).

Methods and results We studied 1106 HF patients from the Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT) (left ventricular ejection fraction ≤ 30%, QRS ≥ 130 ms, New York Heart Association class ≤ II) randomized to CRT with a defibrillator (CRT-D, $n = 677$) or an implantable cardioverter defibrillator (ICD, $n = 429$). An unsupervised ML algorithm (Multiple Kernel Learning and K-means clustering) was used to categorize subjects by similarities in clinical parameters, and left ventricular volume and deformation traces at baseline into mutually exclusive groups. The treatment effect of CRT-D on the primary outcome (all-cause death or HF event) and on volume response was compared among these groups. Our analysis identified four phenogroups, significantly different in the majority of baseline clinical characteristics, biomarker values, measures of left and right ventricular structure and function and the primary outcome occurrence. Two phenogroups included a higher proportion of known clinical characteristics predictive of CRT response, and were associated with a substantially better treatment effect of CRT-D on the primary outcome [hazard ratio (HR) 0.35; 95% confidence interval (CI) 0.19–0.64; $P = 0.0005$ and HR 0.36; 95% CI 0.19–0.68; $P = 0.001$] than observed in the other groups (interaction $P = 0.02$).

Conclusions Our results serve as a proof-of-concept that, by integrating clinical parameters and full heart cycle imaging data, unsupervised ML can provide a clinically meaningful classification of a phenotypically heterogeneous HF cohort and might aid in optimizing the rate of responders to specific therapies.

Keywords Machine learning • Heart failure • Personalized medicine • Echocardiography • Cardiac resynchronization therapy

*Corresponding author. Cardiovascular Division, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02115, USA. Tel: +1 857 3071960, Fax: +1 857 3071944, Email: ssolomon@rics.bwh.harvard.edu
†These authors contributed equally.

Introduction

The goal of personalized medicine is to optimize the tailoring of treatments to specific patients in order to maximize the treatment response, which, as a prerequisite, requires accurate patient phenogrouping. The syndrome of heart failure (HF) comprises particularly heterogeneous patient groups, burdened by limited success of some treatment options. Machine learning (ML) approaches have been applied in the diagnosis, classification, assessment of readmissions and medication adherence of HF patients,[1] as well as to identify distinct phenogroups in several disorders, including HF with preserved ejection fraction (HFpEF),[2,3] and to predict mortality in patients with suspected coronary artery disease.[4] Supervised ML involves using iterative algorithms that 'learn' from a large accurately labelled training dataset[5]; while often diagnostically 'accurate', it is generally impossible to infer the 'diagnostic reasoning' employed in these algorithms. Unsupervised approaches, however, do not attempt to identify a diagnostic or prognostic 'truth' but instead group (or cluster) patients together based on multiple characteristics, which could be demographic, historical, or measured. By grouping similar patients together in multiple dimensions, it is then possible to analyse the characteristics of similarly grouped individuals and relate them to outcomes or therapeutic responses. We have previously shown that unsupervised Multiple Kernel Learning (MKL) can be applied to find similarities among patients, based on a wide range of heterogeneous data, such as complex imaging-based descriptors of ventricular structure and function, in an 'agnostic' manner[2].

One such area where more accurate phenogrouping could improve selection of patients is cardiac resynchronization therapy (CRT) which, despite clear guidelines for which patients should be treated, a substantial proportion of patients do not respond to this therapy.[6–9] We hypothesized that novel approaches based on ML, integrating clinical parameters with complex echocardiographic data on myocardial deformation and left ventricular (LV) volume changes measured over the entire cardiac cycle might be able to overcome some of the limitations of traditional approaches to patient selection for CRT, and provide an example of how ML can be utilized to better phenogroup patients with HF with respect to both outcomes and response to therapy. We therefore utilized data from Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT), a large randomized clinical trial of 1820 patients with New York Heart Association (NYHA) functional class $\leq$ II symptoms, LV ejection fraction (LVEF) $\leq$ 30% and QRS $\geq$ 130 ms,[10] to determine whether unsupervised ML could aid in the identification of patients likely to respond to CRT.

Methods

Study population

The design and results of MADIT-CRT have been published previously.[10,11] In brief, the MADIT-CRT trial enrolled 1820 patients from December 2004, through April 2008, at 110 centres in the United States, Canada, and Europe. These were mildly symptomatic patients with ischaemic heart disease (in NYHA class I or II) or patients with non-ischaemic heart disease (in NYHA class II) in sinus rhythm with an LVEF $\leq$ 30%, and a QRS duration $\geq$ 130 ms, who were randomly assigned in a 3:2 ratio to receive a CRT-D or an ICD alone. All recruited subjects met guideline indications for ICD therapy.[12] The main objective was to determine whether CRT-D reduces the risk of death or HF events compared with ICD. The average follow-up period was 2.4 years. The protocol was approved by the institutional review board at each of the participating centres, and each subject gave written informed consent.

Echocardiography

Two-dimensional (2D) echocardiography was performed before device implantation (baseline) and at 1-year follow-up, following a study-specific protocol.[13] The echocardiographic core laboratory at Brigham and Women's Hospital performed the screening of the echocardiograms for quality, and the echocardiographic measurements relevant to the study. Left ventricular and atrial volumes were assessed by the biplane Simpson's method. LVEFs were calculated according to standard methods.[13] Reproducibility of the primary volumetric measurements has been previously demonstrated.[14]

The echocardiographic images of 1106 patients in this MADIT-CRT analysis (CRT-D, $n = 677$; ICD-only, $n = 429$) were analysed using the TomTec Arena software (v1.0, TomTec Imaging Systems, Unterschleissheim, Germany). Endocardial borders were traced in the end-systolic frame of the apical 4- and 2-chamber views, and automatically propagated over the course of two cardiac cycles. We stored 49 segmental LV longitudinal strain and one volume curves for *a posteriori* ML analysis. Previous studies report excellent reproducibility of the estimated LV myocardial deformation.[15] The reasons to exclude patients from the analysis included: images in non-DICOM format, frame rate < 30 Hz, missing of 4- or 2-chamber images, unacceptable 2D image quality, use of echocardiographic contrast agent, presence of endocardial dropout, or out-of-plane images.

Outcome measures

The primary endpoint of the trial was death from any cause or a non-fatal HF event, whichever came first.[10] The adjudication of the endpoints was carried out by an independent endpoint committee, unaware of patient randomization status.[10] In addition to determining the treatment effect on the primary outcome over an average follow-up of 2.3 years, we have also assessed the benefit on echocardiographic response at 1-year follow-up.

Baseline characteristics, data pre-processing and unsupervised machine learning

Seventy-seven baseline variables, consisting of clinical and echocardiographic parameters with $< 20\%$ missing data were identified. After filtering correlated variables using a cut-off Pearson's coefficient > 0.8, 50 variables including demographic and laboratory data, ECG and echocardiography measurements, data on medication use and recruitment centre were selected at baseline and were used as input for the ML algorithm (*Table 1*). These variables included both categorical and continuous data, with the continuous variables converted to ordinal by dividing their range into 10 uniform bins.[16] Missing data for

Table 1 Baseline characteristics of the study patients by phenogroups

	Overall average	Phenogroup 1 ($n = 157$)	Phenogroup 2 ($n = 370$)	Phenogroup 3 ($n = 291$)	Phenogroup 4 ($n = 288$)	Group *P*-value
Age, years	64 ± 11	62 ± 11	64 ± 11	67 ± 11	63 ± 11	< 0.001
Female sex	274 (25%)	35 (22.3%)	30 (8.1%)	195 (67.0%)	14 (4.9%)	< 0.001
Race, white	1006 (91%)	151 (96%)	324 (88%)	274 (94%)	257 (89%)	0.002
Ischaemic CMP	622 (56%)	71 (45.2%)	264 (71.4%)	124 (42.6%)	163 (56.6%)	< 0.001
NYHA class II	934 (84%)	141 (89.8%)	287 (77.6%)	261 (89.7%)	245 (85.1%)	< 0.001
Hypertension	687 (62%)	77 (49.0%)	254 (68.7%)	174 (59.8%)	182 (63.2%)	< 0.001
Diabetes	311 (28%)	36 (22.9%)	118 (31.9%)	62 (21.3%)	95 (33.0%)	0.002
Smoking	134 (12%)	27 (17.2%)	44 (11.9%)	30 (10.3%)	33 (11.5%)	0.18
Prior CABG	312 (28%)	39 (24.8%)	129 (34.9%)	53 (18.2%)	91 (31.6%)	< 0.001
Prior non-CABG revascularization	314 (28%)	30 (19.1%)	146 (39.5%)	60 (20.6%)	78 (27.1%)	< 0.001
Prior MI	497 (45%)	52 (33.1%)	215 (58.1%)	101 (34.7%)	129 (44.8%)	< 0.001
Prior CVA	66 (6%)	9 (5.7%)	33 (8.9%)	9 (3.1%)	15 (5.2%)	0.016
Prior HF hospitalization	409 (37%)	56 (35.7%)	108 (29.2%)	125 (43.0%)	120 (41.7%)	< 0.001
No. of hospitalizations prior to enrolment						0.73
0	590 (53%)	88 (56%)	195 (53%)	160 (55%)	147 (51%)	
1	374 (34%)	48 (31%)	130 (35%)	96 (33%)	100 (35%)	
2	97 (9%)	18 (12%)	29 (8%)	24 (8%)	26 (9%)	
≥ 3	45 (4%)	3 (2%)	16 (4%)	11 (4%)	15 (5%)	
Prior ventricular arrhythmias	71 (6%)	17 (10.8%)	23 (6.2%)	14 (4.8%)	17 (5.9%)	0.09
Prior atrial arrhythmias	115 (10%)	15 (9.6%)	44 (11.9%)	19 (6.5%)	37 (12.9%)	0.06
SBP, mmHg	123 ± 18	117 ± 16	125 ± 18	123 ± 18	123 ± 17	< 0.001
DBP, mmHg	72 ± 10	71 ± 11	73 ± 10	71 ± 10	72 ± 11	0.07
Heart rate, b.p.m.	63 ± 11	66 ± 11	62 ± 11	64 ± 11	64 ± 12	< 0.001
Height, cm	173 ± 9.6	171 ± 8	177 ± 7	163 ± 8	178 ± 7	< 0.001
BMI, kg/m^2	28.3 ± 5.0	27.4 ± 4.0	29.6 ± 4.7	25.5 ± 4.7	29.8 ± 5.0	< 0.001
BSA, m^2	2.01 ± 0.24	1.95 ± 0.20	2.13 ± 0.17	1.75 ± 0.15	2.14 ± 0.19	< 0.001
QRS duration, ms	157 ± 19	172 ± 22	152 ± 16	154 ± 16	159 ± 20	< 0.001
LBBB	782 (71%)	135 (86%)	204 (55.1%)	235 (80.8%)	208 (72.2%)	< 0.001
RBBB	133 (12%)	7 (4.5%)	80 (21.6%)	18 (6.2%)	28 (9.7%)	< 0.001
Interventricular conduction delay	184 (17%)	14 (8.9%)	86 (23.2%)	33 (11.3%)	51 (17.7%)	< 0.001
Six-minute walk distance, m	363 ± 103	378 ± 100	367 ± 106	346 ± 102	366 ± 99	0.006
Blood urea nitrogen, mg/dL	22 ± 9	22 ± 9	22 ± 10	21 ± 8	23 ± 9	0.014
Creatinine, mg/dL	1.2 ± 0.3	1.1 ± 0.3	1.2 ± 0.3	1.1 ± 0.3	1.3 ± 0.3	< 0.001
ACE inhibitor/ARB	1055 (95%)	151 (96.2%)	348 (94.1%)	284 (97.6%)	272 (94.4%)	0.14
Beta-blocker	1031 (93%)	145 (92.4%)	343 (92.7%)	276 (94.9%)	267 (92.7%)	0.64
Diuretic	814 (74%)	132 (84.1%)	242 (65.4%)	215 (73.9%)	225 (78.1%)	< 0.001
Aldosterone antagonist	326 (30%)	56 (35.7%)	98 (26.5%)	76 (26.1%)	96 (33.3%)	0.04
Calcium channel blocker	88 (8%)	6 (3.8%)	46 (12.4%)	16 (5.5%)	20 (6.9%)	< 0.001
Amiodarone	74 (7%)	13 (8.3%)	20 (5.4%)	11 (3. 8%)	30 (10.4%)	0.007
Digitalis	282 (26%)	56 (35.7%)	77 (20.8%)	83 (28.5%)	66 (22.9%)	0.002
Statin	743 (67%)	88 (56.1%)	292 (78.9%)	166 (57.1%)	197 (68.4%)	< 0.001
Antiarrhythmic medication	7 (1%)	1 (0.6%)	4 (1.1%)	0 (0%)	2 (0.7%)	0.38
LVEDVi, mL/m^2	124 ± 28	172 ± 31	105 ± 12	119 ± 16	128 ± 14	< 0.001
LVESVi, mL/m^2	88 ± 23	128 ± 26	72 ± 9	83 ± 12	93 ± 10	< 0.001
Regional wall thickness	0.25 ± 0.03	0.22 ± 0.03	0.26 ± 0.02	0.26 ± 0.02	0.25 ± 0.02	< 0.001
LV mass, g	211 ± 38	249 ± 43	197 ± 25	186 ± 22	233 ± 32	< 0.001
LVMi, g/m^2	106 ± 18	128 ± 18	93 ± 12	107 ± 14	109 ± 14	< 0.001
LA width, cm	4.0 ± 0.2	4.2 ± 0.2	3.9 ± 0.1	3.8 ± 0.1	4.1 ± 0.1	< 0.001
LAVi, mL/m^2	46 ± 10	59 ± 11	39 ± 6	44 ± 8	50 ± 8	< 0.001
LVEF, %	29 ± 3	26 ± 3	31 ± 3	31 ± 3	28 ± 3	< 0.001
12-segment GLS, %	−9.8 ± 2.8	−7.8 ± 2.4	−10.8 ± 2.8	−10.4 ± 2.7	−9.0 ± 2.4	< 0.001
RV diameter, mm	28.3 ± 2.3	30.4 ± 1.9	27.7 ± 1.7	26.7 ± 1.7	29.8 ± 1.6	< 0.001
RV FAC, %	43 ± 6	39 ± 5	44 ± 5	44 ± 6	40 ± 4	< 0.001

ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; BMI, body mass index; BSA, body surface area; CABG, coronary artery bypass grafting; CMP, cardiomyopathy; CVA, cerebrovascular accident; DBP, diastolic blood pressure; FAC, fractional area change; GLS, global longitudinal strain; HF, heart failure; LA, left atrial; LAVI, left atrial volume index; LBBB, left bundle branch block; LV, left ventricular; LVEDVi, left ventricular end-diastolic volume index; LVEF, left ventricular ejection fraction; LVESVi, left ventricular end-systolic volume index; LVMi, left ventricular mass index; MI, myocardial infarction; NYHA, New York Heart Association; RBBB, right bundle branch block; RV, right ventricular; SBP, systolic blood pressure.

input variables ranged from 0% to 15.6% in the case of right ventricular fractional area change. They were imputed using the imputeFAMD function within the missMDA package in R,[17] which allows imputing mixed datasets (with continuous and categorical variables) using a principal component method adapted for mixed data.

In addition to these common baseline characteristics, baseline LV strain and volume traces throughout the entire cardiac cycle as well as a temporal deformation vector (used to keep the relative changes in duration of the cardiac phases, relevant for improved HF characterization[2]) were included in the further processing by the ML algorithm. In order to retain the wealth of data on LV geometry and deformation over a cardiac cycle contained in the traces, each one of them was defined and inputted to the algorithm as a set of data points (specifically, 102 variables per trace), instead of e.g. only a peak value (such as end-diastolic and end-systolic volume or peak systolic strain). Prior to analysis, these traces need to be referenced to a common temporal framework[18] (see online supplementary *Methods S1*). The 49 segmental strain traces available from the 2-chamber and 4-chamber views were converted into two basal, two mid-LV and two apical segments, by isolating and averaging groups of eight consecutive traces. The most apical trace was discarded.

The final input to the algorithm is shown in *Figure 1* (*left panel*). For the 2-chamber and 4-chamber views of every patient, a total of eight echocardiographic descriptors (traces) were analysed per view (*Figure 1*, *left panel*): one volume trace, six strain traces, and one temporal deformation vector, which results from the temporal alignment step. Fifty clinical parameters inputted to the algorithm are listed in *Table 1*. These echocardiographic descriptors (full traces) and baseline clinical parameters provided a total of 1682 input variables: each echocardiographic trace contains 102 data points, making a total of 1632 echocardiographic trace data points (8 traces x 2 views x 102 time instants) to which 50 clinical parameters were added. We then used unsupervised MKL (*Figure 1*, *right panel*), an ML algorithm already validated and extensively tested to combine cardiac motion data,[2] to convert the input dataset consisting of 1682 variables into a compact representation space where subjects are positioned according to their similarity, while blinded to the patient's outcome status with respect to both clinical events and volume response.

Once positioned in the compact representation space, subjects were clustered with the K-means algorithm (*Figure 1*, *right panel*) to identify phenotypically-distinct categories of CRT candidates. We ran the clustering algorithm with increasing number of predefined groups (from 3 to 8), however, the clinical interpretation of the clusters remained stable; ultimately, we chose the configuration that maximizes the statistical significance (minimizing *P*-value for trend, adjusted for multiple testing) of the treatment effect on the primary outcome among clusters (henceforth referred as phenogroups).

Further details about our unsupervised ML method can be found in the online supplementary *Methods S1*.

Comparison of clinical and echocardiographic characteristics; survival and treatment effect on primary outcome and left ventricular reverse remodelling

Categorical variables are expressed as counts and percentages, and differences among phenogroups were assessed using the chi-square test. Continuous variables are presented as mean ± standard deviation, and inter-group differences were calculated using ANOVA. A *P*-value of < 0.05 was considered statistically significant. The previous comparison was complemented with a physiologic interpretation of the found phenogroups in the form of a variability analysis of strain and volume patterns in the 2-chamber and 4-chamber views, using advanced regression techniques.[19] Kaplan–Meier estimates for HF or death in each phenogroup were determined and statistically compared with the log-rank test. Cox proportional hazards regression analyses were performed on each phenogroup to estimate the treatment effect on the primary endpoint. The treatment effect on volumetric response was expressed for every phenogroup as the difference between treated and untreated patients in LV end-diastolic volume index (LVEDVi) per cent change (from baseline to 1-year follow-up).

Stability and internal validation of the unsupervised machine learning model

We evaluated the generalizability of our dimensionality reduction solution assessing the correlation among low-dimensional space distributions obtained by analysing populations with an increasing number of subjects in common. We also checked the consistency among the K-means clustering configurations by computing the membership agreement when increasingly partitioning the space, from 3 to 8 clusters.

We assessed the stability of these results through internal validation, which involved running our ML algorithm in a randomly-selected portion of the database (75% = training set) to create clusters, finding the corresponding cluster for the remaining subjects (25% = validation set), and comparing both the training and the validation clustering solutions in terms of clinical characteristics and outcome. Further details on both the stability experiments and the internal validation can be found in the online supplementary *Methods S1*.

The ML algorithm as well as the regression technique used to analyse the variability of echocardiographic patterns among phenogroups were implemented using MATLAB (R2016b, 2016, The MathWorks Inc., Natick, MA, USA). Survival and treatment effect analyses were performed in Stata version 13 (StataCorp, College Station, TX, USA).

Results

Results of machine learning

The MKL algorithm reduced the dimensionality of the input data to equal the number of input subjects minus 1. However, only the first two dimensions of the output (low-dimensional) space were considered for clustering, as they encoded the most salient characteristics of these data.[2] Furthermore, they presented the highest standard deviations on the coordinates of subjects (with further dimensions showing a linear decay up to the sixth dimension, from which the standard deviation is $> 98\%$ smaller than the first dimension), and thus contributed to a higher extent to the cluster assignment computed by the K-means algorithm.

Baseline characteristics of patients by phenogroups

Baseline characteristics of the patients included in this analysis were comparable to the remainder of the MADIT-CRT study, as reported previously.[15]

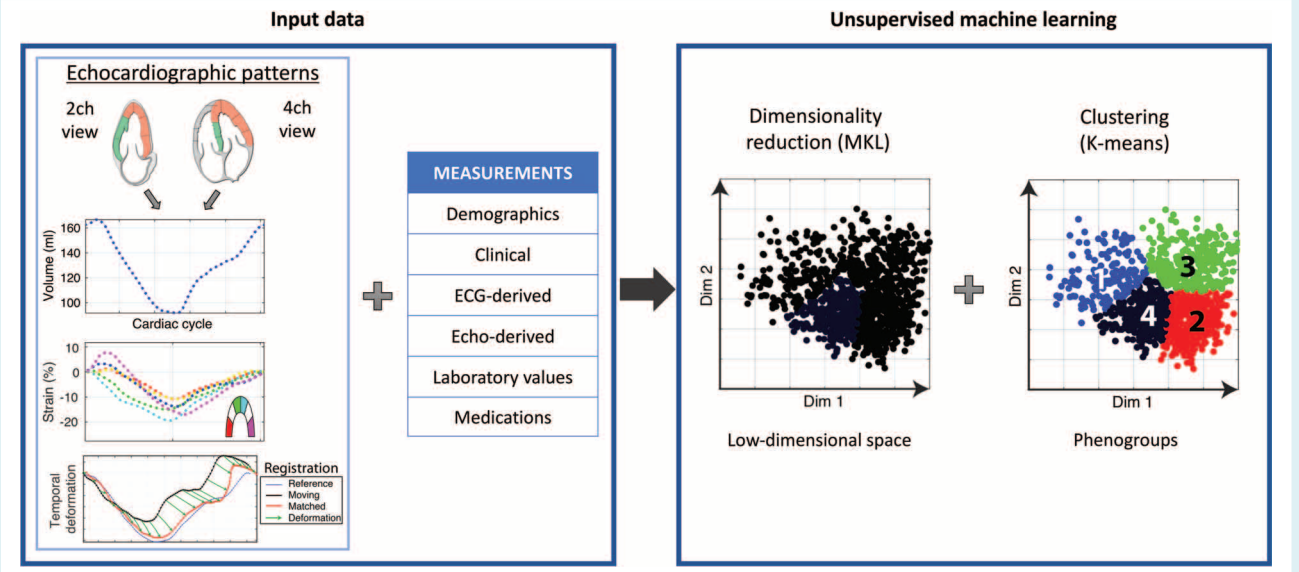


Figure 1 Overview of the proposed analysis. Input data consist of both complex descriptors from echocardiography and clinical parameters (*left panel*), which are used by the unsupervised machine learning algorithm to position subjects according to their similarity through dimensionality reduction, and eventually propose coherent subgroups of subjects using clustering (*right panel*). 2ch, 2-chamber; 4ch, 4-chamber; ECG, electrocardiogram; MKL, Multiple Kernel Learning.

The most statistically significant clustering solution categorized the overall patient population into four clusters, i.e. phenogroups (*Figure 1*, *right panel*) with distinct clinical and echocardiographic characteristics (*Table 1*, *Figure 2*). This solution was better at identifying CRT responders than those obtained by independently analysing clinical parameters or complex echocardiographic descriptors alone (see online supplementary *Methods S1*). Phenogroups 1 and 3 were associated with the highest proportion of clinical characteristics known to be predictive of volumetric response to CRT[14]: phenogroups 1 and 3 comprised the highest proportion of patients with non-ischaemic cardiomyopathy (54.8% and 57.4%, respectively) and left bundle branch block (LBBB, 86.0% and 80.8%, respectively), the QRS duration was the longest in phenogroup 1, which was also the phenogroup with the lowest median age, while phenogroup 3 consisted of the largest proportion of female patients. Conversely, phenogroups 2 and 4 were associated with the highest proportion of male patients and ischaemic origin of HF, as well as the lowest proportion of patients with LBBB morphology on the elctrocardiogram.

The values of systolic blood pressure were the lowest and heart rate was the highest in phenogroup 1; this was also the phenogroup with the highest proportion of patients receiving diuretics and aldosterone antagonists. There was no significant difference in the proportion of patients receiving beta-blockers or angiotensin-converting enzyme inhibitors/angiotensin receptor blockers among the four phenogroups.

Furthermore, echocardiography measurements revealed that the patients in phenogroup 1 had the most remodelled left ventricles at baseline (the largest LV end-diastolic and end-systolic volume index, LV mass index and left atrial volume index) and the lowest LVEF and 12-segment global longitudinal strain, while the same was observed for right ventricular size and function in this phenogroup (largest right ventricular diameter and lowest fractional area change), with phenogroup 4 having similarly remodelled right ventricles. Conversely, these measurements of left atrial and LV structure demonstrated the lowest severity of remodelling in phenogroup 2. Right ventricular size was the smallest and fractional area change and LVEF were the highest in phenogroups 2 and 3 (group *P*-values for all mentioned echo parameters < 0.001).

In addition to the clinical and echocardiographic characteristics of the studied patients, the MKL algorithm also included data on LV volume traces and longitudinal strain traces. Representative 'fingerprints' of such traces are shown for each phenogroup in *Figure 3*. In phenogroup 1, the LV strain curves show late systolic stretch of the apical septal segment and a mirrored contraction of the apical lateral wall — a feature described as a part of the LBBB-related septal flash pattern. The volume trace shows a delayed peak, i.e. tardily achieved end-systolic volume. These patients had the largest LV end-diastolic volumes and the lowest LVEF values. The strain curves in phenogroup 2 show nearly absent deformation only in the apical anterolateral region, with a normal shape of the strain trace and lower peak values in the septal and inferior regions. Along with phenogroup 3, these were the least dilated ventricles with the highest LVEF values. In phenogroup 3, there is early deformation of the apical septum while some early stretch is present in the lateral traces, mirroring an early deformation of the apical septum. Phenogroup 4 exhibits nearly absent deformation in the basal inferoseptum with very low deformation in all apical regions and somewhat preserved deformation in the basal anterolateral wall — this pattern is indicative of large apical infarcts extending

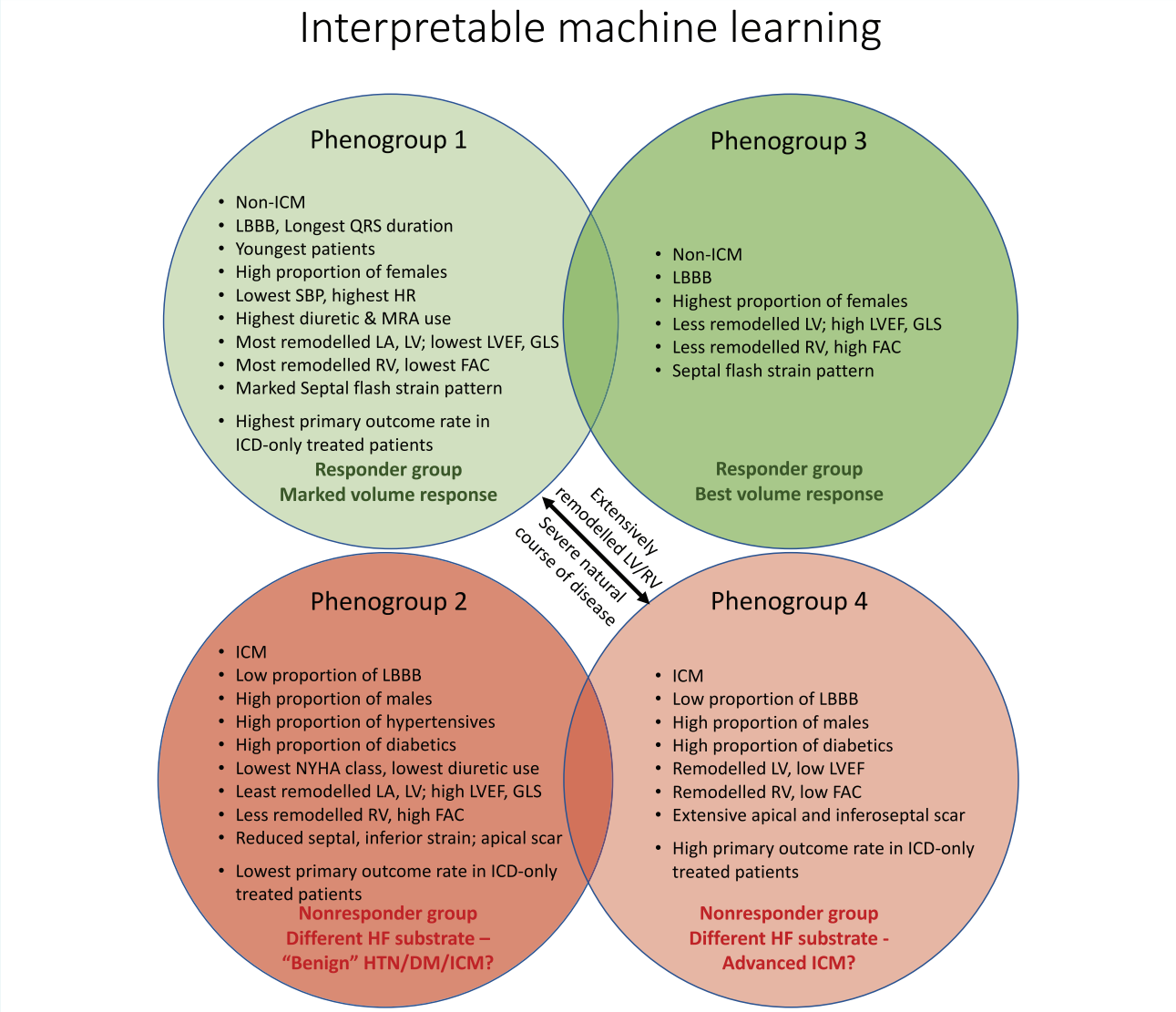


Figure 2 Typical clinical characteristics, features of left ventricular deformation patterns and outcome rates of the four phenogroups. The green circles represent the phenogroups more likely to respond to cardiac resynchronization therapy, as opposed to the red circles. The image summarizes the clinical interpretability of the results obtained by the utilized unsupervised machine learning algorithm. DM, diabetes mellitus; FAC, fractional area change; GLS, global longitudinal strain; HF, heart failure; HR, heart rate; HTN, hypertension; ICD, implantable cardioverter defibrillator; ICM, ischaemic cardiomyopathy; LA, left atrium; LBBB, left bundle branch block; LV, left ventricle; LVEF, left ventricular ejection fraction; MRA, mineralocorticoid receptor antagonist; NYHA, New York Heart Association; RV, right ventricle; SBP, systolic blood pressure.

to the inferoseptum. Although the volume curve in phenogroup 4 peaks early, these patients also have markedly remodelled left ventricles.

Comparison of survival among phenogroups

The natural course of disease, as assessed in the ICD-only subgroup of patients, varied among the phenogroups (*Figure 4*, *left panel*): the Kaplan–Meier estimate of the probability of survival free of HF revealed a less severe disease course in phenogroup 2 in which the primary event occurred in 15.4% of the patients in the ICD-only subgroup (2.1% of the patients died, 9.8% were hospitalized for HF and the remaining 3.5% had an out-of-hospital HF event). Conversely, the untreated patients in phenogroup 1 had the highest incidence of the primary event, occurring in 38% of patients (1.4% had an all-cause death, 33.8% were hospitalized for HF and 2.8% had a HF event not requiring hospitalization). Overall, the primary outcome occurred in 220 patients from the current analysis, and differed significantly among phenogroups: it occurred most frequently in phenogroups 1 and 4 [41 patients (26.1%) and 79 patients (27.4%), respectively] and was least represented

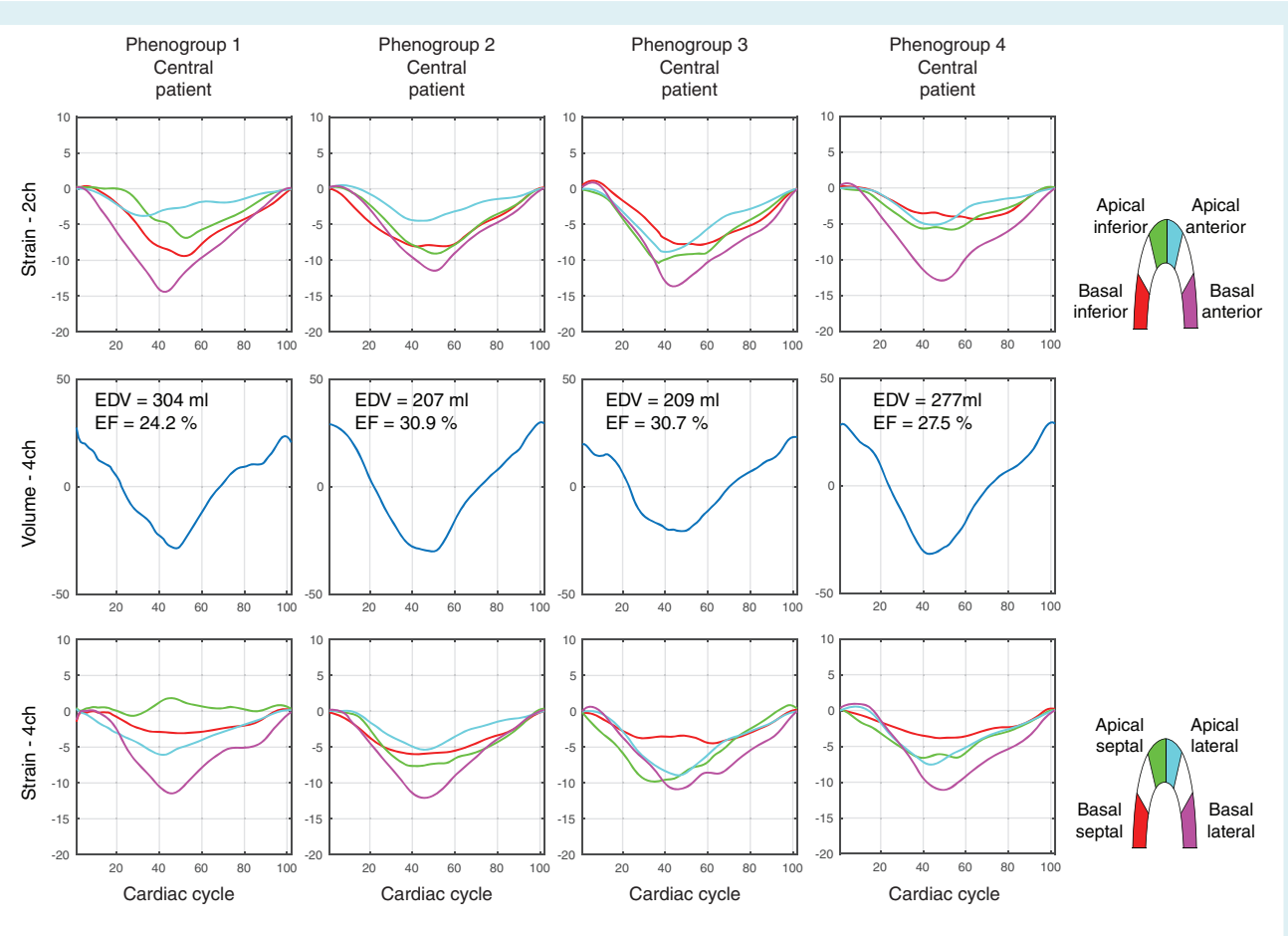


Figure 3 Volume and strain traces corresponding to the representative patient of each phenogroup, i.e. those located at the barycentre of the phenogroup distribution. 2ch, 2-chamber; 4ch, 4-chamber; EDV, end-diastolic volume; EF, ejection fraction.

in phenogroups 2 and 3 [55 patients (14.9%) and 45 patients (15.5%), respectively] (*Figure 4*, *left panel*). All-cause death did not differ significantly among the phenogroups; the difference in the primary endpoint was mainly driven by a significant difference in the occurrence of HF requiring hospitalization, occurring most often in phenogroups 1 and 4 (21.0% and 20.5% of patients, respectively) and least frequently in phenogroups 2 and 3 (10.3% and 11.3% of patients, respectively).

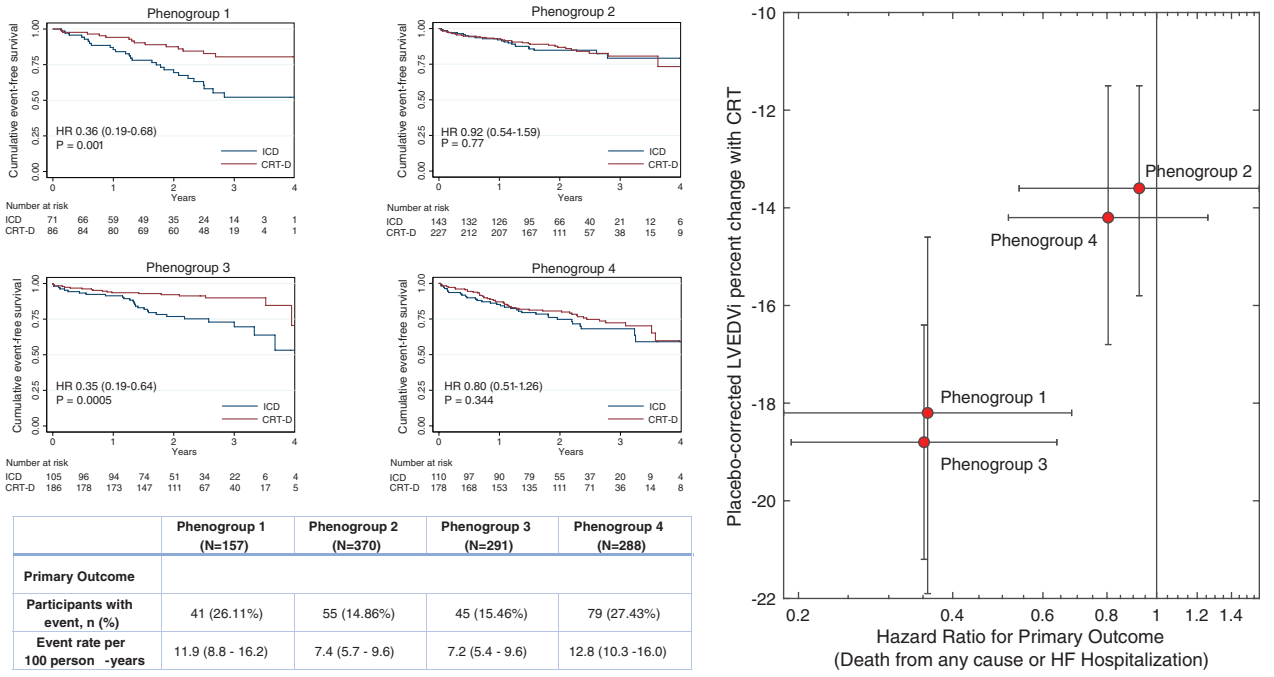
Effect of treatment on primary outcome and left ventricular reverse remodelling

The effect of CRT-D treatment, compared to ICD-only, on the primary outcome of death or HF event assessed among the four phenogroups by Cox proportional hazard analysis is depicted on *Figure 4* (*right panel*): patients categorized to phenogroups 1 and 3 exhibited a 64% and 65% reduction in the risk of HF or death, respectively [hazard ratio (HR) 0.36, 95% confidence interval (CI) 0.19–0.68; $P = 0.001$ and HR 0.35, 95% CI 0.19–0.64; $P = 0.0005$, respectively], which was a substantially higher treatment benefit than observed in the other groups (interaction $P = 0.02$). Phenogroups 2 and 4 benefited from CRT-D therapy to a lesser extent compared to the overall cohort; however, the non-response did not reach statistical significance.

A significant treatment effect on LV reverse remodelling, defined as LVEDVi per cent change, was noted in all phenogroups (*Figure 4*, *right panel*). However, phenogroup 3, characterized by a lower severity of ventricular remodelling at baseline, was identified to be associated with a substantially better volume response: in this phenogroup, CRT-D treatment was associated with an average 18.8% decrease in LVEDVi, when corrected for ICD-only treatment (95% CI −21.2 to −16.4; $P < 0.0001$). A marked volume response was also detected in phenogroup 1 with an average 18.2% decrease in LVEDVi (95% CI −21.9 to −14.6; $P < 0.0001$), while patients in phenogroups 2 (HR −13.6, 95% CI −15.8 to −11.5; $P < 0.0001$) and 4 (HR −14.2, 95% CI −16.8 to −11.5; $P < 0.0001$) showed the lowest amount of LVEDVi per cent change within 12 months.

Stability and internal validation

The similarity among low-dimensional space distributions increased with the number of subjects in common at the input of



	Phenogroup 1 (N=157)	Phenogroup 2 (N=370)	Phenogroup 3 (N=291)	Phenogroup 4 (N=288)
Primary Outcome				
Participants with event, n (%)	41 (26.11%)	55 (14.86%)	45 (15.46%)	79 (27.43%)
Event rate per 100 person -years	11.9 (8.8 - 16.2)	7.4 (5.7 - 9.6)	7.2 (5.4 - 9.6)	12.8 (10.3 -16.0)

Figure 4 Kaplan–Meier estimates of the probability of survival free of heart failure (HF) according to treatment arm in each of the phenogroups. The table shows the incidence rates for the primary outcome by phenogroup (*left panel*). The combined effect of cardiac resynchronization therapy-defibrillator (CRT-D) treatment on the primary outcome of death or HF event (x-axis) and implantable cardioverter defibrillator (ICD)-only corrected per cent change in left ventricular end-diastolic volume index (LVEDVi, y-axis) assessed among the four phenogroups (*lower panel*). $P = 0.02$ and $P = 0.005$ for interaction of primary outcome and volume response, respectively (*right panel*). HR, hazard ratio.

the ML analysis, resulting in excellent correlation when the subjects in common were above 500 (Pearson correlation coefficient > 0.90). We also observed a high degree of consistency across clustering configurations from 3 to 8 clusters. Lastly, similar trends were observed in the clinical parameters and the treatment effect when comparing the training and validation clustering configurations, emphasizing the capacity of our model to predict outcomes for new, unseen data. All these results are detailed in the online supplementary *Methods S1*.

Discussion

In this analysis we have shown that unsupervised ML allows for a novel integration of entire cycle-wide LV volume and deformation traces from echocardiography, rather than only single data points, which can be combined with extensive clinical and medication parameters to phenotype patients with complex diseases such as HF. We have also demonstrated the added value of combining both sets of descriptors to find subjects that are more likely to respond to CRT, compared to the results obtained by independently analysing clinical parameters or complex echocardiographic descriptors alone. Our results serve as a proof-of-concept that unsupervised ML-based approaches can be used to combine both standard clinical parameters and complex echocardiographic data to provide a clinically interpretable and meaningful classification of a phenotypically heterogeneous HF cohort and to identify patients most likely to respond to specific therapies.

Integrating echocardiographic tracings to address the heterogeneity of a heart failure population

Heart failure is a multifaceted syndrome and response to therapies is based on multiple clinical and imaging parameters as well as biomarkers. Traditional methods to define phenotypes and predict outcomes within groups of individuals with HF rely on the elucidation of individual phenotypic subgroups that focus on isolated characteristics (i.e. aetiology of HF, QRS morphology, presence or absence of specific co-morbidities, cardiac structure and function, etc.). Furthermore, while assessment of cardiac structure and function using current echocardiographic analysis tools can identify subgroups of HF patients at higher risk for adverse outcomes,[20] standard approaches ascribe risk to a limited amount of individual measurements in a unidimensional fashion. Namely, data on cardiac structure and function provided by echocardiography contain a plethora of information representing multiple time points in a cardiac cycle (the number of data points corresponds to the frame rate of the acquired images), but are typically under-exploited in standard quantitative data analyses and replaced by single measurements, thus failing to summarize the complexity of events over

the cardiac cycle. Unlike previous studies that aimed, but failed, at finding a single echocardiographic measure of dyssynchrony to improve patient selection for CRT beyond current guidelines,[21] we integrated echocardiographic imaging in a more comprehensive and novel manner by integrating entire LV volume and strain patterns throughout the cardiac cycle rather than utilizing single measures such as LV end-diastolic and systolic volume or global longitudinal strain. By integrating over 1600 data points per cardiac cycle, this method also is able to incorporate complex patterns of regional cardiac function that are impossible to describe parametrically. These algorithms thus combine a detailed analysis of cardiac dynamics over an entire cardiac cycle with an extensive set of clinical parameters.

It is often emphasized that the management of HF patients requires improved integration of clinical data with echocardiography — the most widely used and accessible diagnostic tool for a comprehensive assessment of cardiac structure and function. Indeed, ML allows for the integration of very large amounts of continuous and discrete variables pertinent to clinical characteristics, laboratory values, electrocardiographic parameters and commonly analysed echocardiographic variables, as has been applied in HF and other cardiovascular diseases. Unsupervised ML techniques, such as the MKL version that we have utilized in this study, offer the advantage of exploiting these full acquired datasets to compare similarities amongst patients without assumptions on which single measurements (data points) are most relevant for the studied patient population. In addition to tissue Doppler trace analysis in HF with preserved ejection fraction patients,[22] the analysis of LV strain traces has previously been performed in a study of 60 patients with acute myocardial infarction by applying principal component analysis.[23] Furthermore, the prediction of CRT response was also attempted in a smaller cohort of 34 CRT candidates.[24] The strengths of our analysis as compared to both of these studies is in the use of a non-linear (more adequate to process cardiac motion patterns, compared to principal component analysis[2]) and unsupervised analysis technique (compared to supervised MKL[24]), thus better suited to agnostically partition a population into homogeneous groups. Importantly, the richness of the analysed data obtained by integrating entire volume/strain traces provides enough information to enable identification of phenogroups (which have been 'agnostically' defined by the K-means algorithm) of patients with similar (but not identical) properties without prior assumptions on outcomes. The performed dimensionality reduction aids in extracting the relevant clinical characteristics of the phenogroups, providing (patho)physiologically relevant and interpretable results. Indeed, ML has previously been successfully employed in the diagnosis, classification and prognostication of HF cohorts.[1,25] In addition to the achieved advancements and ongoing efforts in the field,[26] we believe that the approach proposed in this analysis contributes to this growing field by providing novelty and strengthening the integration of detailed imaging data with standardly utilised clinical variables in an ML analysis dedicated to providing clinically interpretable results. Namely, our analysis was superior at identifying CRT responders compared to independently analysing clinical parameters or complex echocardiographic descriptors alone, which did not provide phenogroups with statistically significant differences in the treatment effect (see online supplementary *Methods S1*).

Positioning unsupervised learning in the spectrum of machine learning approaches and the utility of interpretability

The current data analysis trend is towards powerful approaches such as deep learning, which uses neural networks to solve complex pattern recognition problems[27] such as object[28] and speech[29] recognition, but requires immense collections of data (often lacking in clinical medicine) to make reliable predictions.[27] Furthermore, the 'black-box' nature of this methodology often provides results difficult to interpret.[27] Human interpretability is increasingly recognized as a highly relevant feature of ML methodologies, crucial in efforts towards data-driven precision medicine, based on informed and auditable decisions. Thus, we opted for a 'simpler' and less data-demanding analysis approach, for which we reinforce aspects of interpretability (*Figure 5*). This approach was specifically designed to combine heterogeneous data in an unsupervised way, which ultimately allows finding groups of patients with similar characteristics and therapy response. Our unsupervised analysis approach, rather than classifying based on *a priori* knowledge as done in a recent study targeting the same clinical problem,[30] allows for natural clustering of patients, and results in the identification of patient subgroups with defined treatment effects. Unlike the work by Kalscheur *et al.*,[30] where the authors used a Random Forest regression model to predict outcome, we emphasized the interpretability of our model, which allows exploring the computed data 'universe', and highlights the data features that are relevant to the clinical hypothesis under study. This provides for a more meaningful description and distinction of specific patient groups within the cohort. Specifically, phenogroups 1 and 3 showed marked response to CRT (both in primary outcome and volume response) and shared similar clinical attributes known to be predictive of (volume) response to CRT while the LV strain traces revealed an LBBB-related strain pattern. In contrast, phenogroups 2 and 4 represented the non-responder groups, characterized by a low proportion of LBBB, high proportion of ischaemic heart disease and LV strain patterns consistent with ischaemia/scar. Phenogroup 2 consisted of patients with the least severe course of disease — those with the lowest NYHA class, lowest diuretic use and the least remodelled left ventricles. We postulate that, in conjunction with a different HF substrate (predominantly ischaemic heart disease and other co-morbidities), this lead to a lack of response to CRT. Conversely, those in phenogroup 4 exhibited a larger amount of biventricular remodelling and extensive scarring on the LV strain traces with a high primary outcome rate in the ICD-only subgroup, possibly inferring a more advanced stage of ischaemic heart disease, too advanced to respond to CRT (*Figure 2*). In summary, a combination of beneficial clinical parameters and strain patterns (some known to be typically associated with LBBB and describing LV mechanics in CRT responders) appear to predict a beneficial treatment effect of CRT, superior to echocardiographic and clinical

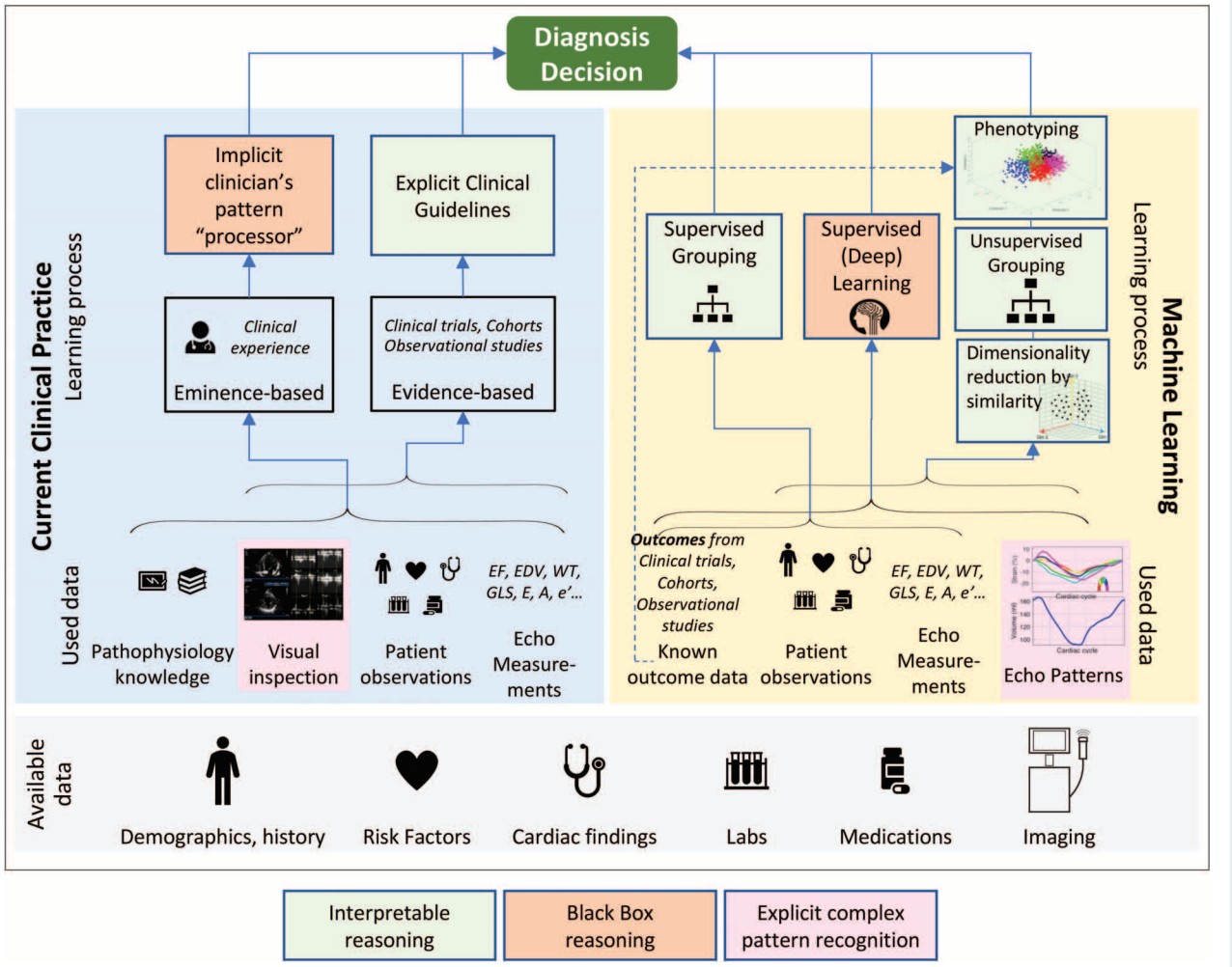


Figure 5 The comparison of current clinical practice and machine learning in the approach to diagnosis and clinical decision making. Both approaches utilize similar 'input' sources, and depending on clinical experience or machine learning approach chosen, can use more or less complex data. Also, both approaches can be based on either interpretable reasoning or black box reasoning. While outcome data for a specific patient are known only *a posteriori* in clinical practice, most machine learning approaches integrate these data *a priori*. The proposed phenotyping approach based on dimensionality reduction of complex patterns and unsupervised grouping is agnostic to outcomes, allowing for phenogroup interpretation based on the integration of outcomes data *a posteriori* (dashed line). EDV, end-diastolic volume; GLS, global longitudinal strain; WT, wall thickness; EF, ejection fraction.

parameters alone. Contrasting clinical features and strain patterns revealing more non-deforming regions suggest less successful treatment by CRT.

In the current manuscript, we did not aim to set out a specific 'model' or scoring system for the prediction of response to CRT, which we believe requires further tool development as well as external validation. Rather, we aimed to ascertain the potential of unsupervised learning approaches in novel phenogrouping of a HF cohort, extended by its application in the prediction of response to a specific therapy, and to demonstrate the benefit of integrating complex imaging data and clinical parameters to accomplish robust phenogrouping. We believe that the novelty predominantly lies in the described methodology, and perhaps less so in the features detected to be associated with CRT response: while this agnostic approach identified features that were previously shown to predict response to CRT,[14,21] we were able to accomplish this in a multivariable manner using both clinical and imaging-based data, rather than by comparison of unidimensional subgroups. Our study is timely, since our echo-based analyses could be relatively easily programmed into echocardiographic post-processing equipment that already does extract the kind of deformation descriptors that we have used.

Limitations

Several limitations of this study should be acknowledged. The results are confined to a selected population of patients with mild HF enrolled in a clinical trial with robust inclusion and

exclusion criteria, which have thus determined the input data to the algorithm. A longer follow-up time than the average of 2.3 years available in our cohort may have been beneficial. Furthermore, an inherent limitation of echocardiographic studies applies to our study as well: the quality of data relies on acquired images and their quality, which was however minimized by excluding echo studies with unacceptable 2D image quality. Although our analysis approach is unsupervised, some human intervention in the form of specification of the most meaningful clustering configuration was required. We demonstrated the overall stability of our results with different database sizes and different sets of descriptors; however, as with all statistical modelling, the results are dependent on the input data, and careful interpretation is needed to guarantee the generalizability of the results. Due to the overlapping between phenogroups, our findings lose power in areas close to the frontier between clusters ('grey zone'). However, if more subjects and clinical descriptors were available, our implementation would allow a more detailed phenotyping, enabling a more patient-specific approach. In such scenario, for every new case the algorithm could suggest similar subjects from its records and provide statistics on the likeliness that a certain subject may develop a disease (diagnosis) or may evolve in a determined way with time or therapy (prognosis). Finally, while external validation would be optimal, a comparable dataset is difficult to obtain, particularly in view of the detailed baseline characteristics and outcomes of the cohort, as well as in respect to the completeness of the dataset. However, we have assessed the stability of our data through internal validation (online supplementary *Methods S1*).

Conclusion

This analysis confirms the utility of unsupervised ML for a novel approach to the integration of complex echocardiographic data (data on LV volume and deformation throughout the cardiac cycle instead of single data points) with clinical parameters to phenotype patients with HF with reduced ejection fraction. Our results serve as a proof-of-concept that fully unsupervised ML approaches can provide an interpretable and clinically meaningful classification of a heterogeneous cohort of HF patients, creating a basis of a data-driven platform that might aid in identifying patient subgroups most likely to respond to specific therapies. The feasibility and novelty of the proposed model for patient phenogrouping in HF and its added value in clinical decision making should be evaluated in a prospective controlled trial.

Supplementary Information

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

Methods S1. Supplementary methods.

Figure S1. Example of non-rigid registration of a left ventricular (LV) volume curve (black) to the reference LV volume curve (blue) via currents, with the resulting aligned LV curve shown in red.

Figure S2. Forest plots for clustering configurations from 3 to 8 clusters. They show the effect of CRT-D treatment, compared to ICD only, on the primary outcome of death or heart failure event for each of the clusters.

Figure S3. Correlation between pairs of output distributions (subsets of the first 6 dimensions).

Figure S4. Correlation between the coordinates of 200 subjects in common to output distributions of increasing size, for the subspace dimensionality varying from 2 to 8.

Figure S5. Consistency among different clustering configurations. The plots show the first two dimensions considered for this work.

Figure S6. Description of the clinical characteristics among the clusters for different clustering configurations (from 3 to 8), using a single spider plot superimposing the signature of each cluster. The values depicted are normalized to the entire population.

Figure S7. Hazard ratios for primary outcome among the phenogroups in the training and validation sets.

Figure S8. Kaplan–Meier curves showing the ICD arms for all phenogroups.

Figure S9. This figure is equivalent to *Figure 4* (*right panel*) in the manuscript, but replacing end-diastolic volume by the left ventricular ejection fraction per cent change in the y-axis.

Table S1. Selected group of clinical characteristics by phenogroup in the training set.

Table S2. Selected group of clinical characteristics by phenogroup in the validation set.

Funding

The work of S. Sanchez-Martinez was supported by a fellowship from 'la Caixa' Banking Foundation. C. Butakoff was supported by a grant from the Fundació La Marató de TV3 (n. 20154031), Spain. N. Duchateau was supported by 'Programme Avenir Lyon Saint-Etienne' (PALSE-IMPULSION-2016, Lyon, France). MADIT-CRT was sponsored by Boston Scientific, while no additional funding was provided for this analysis. This study was also partially supported by the Spanish Ministry of Economy and Competitiveness (grant TIN2014-52923-R; Maria de Maeztu Units of Excellence Programme - MDM-2015-0502) and FEDER.

Conflict of interest: V.K. has received research grant support and speaker honoraria from Zoll and Boston Scientific. A.M. and S.D.S. has received research grants from Alnylam, Amgen, AstraZeneca, Bellerophon, BMS, Celladon, Gilead, GSK, Ionis, Lone Star Heart, Mesoblast, MyoKardia, NIH/NHLBI, Novartis, Sanofi Pasteur, Theracos, and has consulted for Akros, Alnylam, Amgen, AstraZeneca, Bayer, BMS, Corvia, Gilead, GSK, Ironwood, Merck, Novartis, Pfizer, Roche, Takeda, Theracos, Quantum Genetics, Cardurion, AoBiome, Janssen, Cardiac Dimensions. This work was not supported by any of these grants or contracts. All other authors have no conflict of interests to declare.

References

1. Awan SE, Sohel F, Sanfilippo FM, Bennamoun M, Dwivedi G. Machine learning in heart failure: ready for prime time. *Curr Opin Cardiol* 2018;**33**:190–195.
2. Sanchez-Martinez S, Duchateau N, Erdei T, Fraser AG, Bijnens BH, Piella G. Characterization of myocardial motion patterns by unsupervised multiple kernel learning. *Med Image Anal* 2017;**35**:70–82.
3. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, Bonow RO, Huang CC, Deo RC. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation* 2015;**131**:269–279.

4. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, Andreini D, Budoff MJ, Cademartiri F, Callister TQ, Chang HJ, Chinnaiyan K, Chow BJ, Cury RC, Delago A, Gomez M, Gransar H, Hadamitzky M, Hausleiter J, Hindoyan N, Feuchtner G, Kaufmann PA, Kim YJ, Leipsic J, Lin FY, Maffei E, Marques H, Pontone G, Raff G, Rubinshtein R, Shaw LJ, Stehli J, Villines TC, Dunning A, Min JK, Slomka PJ. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J* 2017;**38**:500–507.
5. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;**316**:2402–2410.
6. Abraham WT, Fisher WG, Smith AL, Delurgio DB, Leon AR, Loh E, Kocovic DZ, Packer M, Clavell AL, Hayes DL, Ellestad M, Trupp RJ, Underwood J, Pickering F, Truex C, McAtee P, Messenger J; MIRACLE Study Group. Cardiac resynchronization in chronic heart failure. *N Engl J Med* 2002;**346**:1845–1853.
7. Young J, Abraham W, Smith A, Leon A, Lieberman R, Wilkoff B, Canby R, Schroeder J, Liem L, Hall S, Wheelan K; Multicenter InSync ICD Randomized Clinical Evaluation (MIRACLE ICD) Trial Investigators. Combined cardiac resynchronization and implantable cardioversion defibrillation in advanced chronic heart failure: the MIRACLE ICD trial. *JAMA* 2003;**289**:2685–2694.
8. Saxon LA, Boehmer JP, Hummel J, Kacet S, De Marco T, Naccarelli G, Daoud E. Biventricular pacing in patients with congestive heart failure: two prospective randomized trials. *Am J Cardiol* 1999;**83**:120D–123D.
9. Packer M. Proposal for a new clinical end point to evaluate the efficacy of drugs and devices in the treatment of chronic heart failure. *J Card Fail* 2001;**7**:176–182.
10. Moss AJ, Hall WJ, Cannom DS, Klein H, Brown MW, Daubert JP, Estes NA, Foster E, Greenberg H, Higgins SL, Pfeffer MA, Solomon SD, Wilber D, Zareba W; MADIT-CRT Trial Investigators. Cardiac-resynchronization therapy for the prevention of heart-failure events. *N Engl J Med* 2009;**361**:1329–1338.
11. Moss AJ, Brown MW, Cannom DS, Daubert JP, Estes M, Foster E, Greenberg HM, Hall WJ, Higgins SL, Klein H, Pfeffer M, Wilber D, Zareba W. Multicenter Automatic Defibrillator Implantation Trial-Cardiac Resynchronization Therapy (MADIT-CRT): design and clinical protocol. *Ann Noninvasive Electrocardiol* 2005;**10**:34–43.
12. Epstein AE, Dimarco JP, Ellenbogen KA, Estes NA, Freedman RA, Gettes LS, Gillinov AM, Gregoratos G, Hammill SC, Hayes DL, Hlatky MA, Newby LK, Page RL, Schoenfeld MH, Silka MJ, Stevenson LW, Sweeney MO, Smith SC, Jacobs AK, Adams CD, Anderson JL, Buller CE, Creager MA, Ettinger SM, Faxon DP, Halperin JL, Hiratzka LF, Hunt SA, Krumholz HM, Kushner FG, Lytle BW, Nishimura RA, Ornato JP, Page RL, Riegel B, Tarkington LG, Yancy CW. ACC/AHA/HRS 2008 Guidelines for device-based therapy of cardiac rhythm abnormalities: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the ACC/AHA/NASPE 2002 Guideline update for implantation of cardiac pacemakers and antiarrhythmia devices): developed in collaboration with the American Association for Thoracic Surgery and Society of Thoracic Surgeons. *Circulation* 2008; **117**:e350–408.
13. Solomon SD, Foster E, Bourgoun M, Shah A, Viloria E, Brown MW, Hall WJ, Pfeffer MA, Moss AJ; MADIT-CRT Investigators. Effect of cardiac resynchronization therapy on reverse remodeling and relation to outcome: Multicenter Automatic Defibrillator Implantation Trial: Cardiac Resynchronization Therapy. *Circulation* 2010;**122**:985–992.
14. Goldenberg I, Moss AJ, Hall WJ, Foster E, Goldberger JJ, Santucci P, Shinn T, Solomon S, Steinberg JS, Wilber D, Barsheshet A, McNitt S, Zareba W, Klein H; MADIT-CRT Executive Committee. Predictors of response to cardiac resynchronization therapy in the Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT). *Circulation* 2011;**124**:1527–1536.
15. Knappe D, Pouleur AC, Shah AM, Cheng S, Uno H, Hall WJ, Bourgoun M, Foster E, Zareba W, Goldenberg I, McNitt S, Pfeffer MA, Moss AJ, Solomon SD; Multicenter Automatic Defibrillator Implantation Trial-Cardiac Resynchronization Therapy Investigators. Dyssynchrony, contractile function, and response to cardiac resynchronization therapy. *Circ Heart Fail* 2011;**4**:433–440.
16. Dougherty J, Kohavi R, Sahami M. Supervised and unsupervised discretization of continuous features. In: Prieditis A, Russell S, eds. *Machine Learning: Proceedings of the Twelfth International Conference*. San Francisco, CA: Morgan Kaufmann Publishers; 1995. pp194–202.
17. Josse J, Husson F. missMDA: a package for handling missing values in multivariate data analysis. *J Stat Softw* 2016;**70**:1–31.
18. Duchateau N, Giraldeau G, Gabrielli L, Fernández-Armenta J, Penela D, Evertz R, Mont L, Brugada J, Berruezo A, Sitges M, Bijnens BH. Quantification of local changes in myocardial motion by diffeomorphic registration via currents: application to paced hypertrophic obstructive cardiomyopathy in 2D echocardiographic sequences. *Med Image Anal* 2015;**19**:203–219.
19. Duchateau N, Craene M De, Sitges M, Caselles V. Adaptation of multiscale function extension to inexact matching. Application to the mapping of individuals to a learnt manifold. In: Nielsen F, Barbaresco F, eds. *Geometric Science of Information. GSI 2013*. Berlin, Heidelberg: Springer; 2013. pp578–586.
20. Cikes M, Solomon SD. Beyond ejection fraction: an integrative approach for assessment of cardiac structure and function in heart failure. *Eur Heart J* 2016;**37**:1642–1650.
21. Chung ES, Leon AR, Tavazzi L, Sun JP, Nihoyannopoulos P, Merlino J, Abraham WT, Ghio S, Leclercq C, Bax JJ, Yu CM, Gorcsan J, Sutton MS, De Sutter J, Murillo J. Results of the predictors of response to CRT (PROSPECT) trial. *Circulation* 2008;**117**:2608–2616.
22. Sanchez-Martinez S, Duchateau N, Erdei T, Kunszt G, Aakhus S, Degiovanni A, Marino P, Carluccio E, Piella G, Fraser AG, Bijnens BH. Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction. *Circ Cardiovasc Imaging* 2018;**11**:e007138.
23. Tabassian M, Alessandrini M, Herbots L, Mirea O, Pagourelias ED, Jasaityte R, Engvall J, De Marchi L, Masetti G, D'hooge J. Machine learning of the spatio-temporal characteristics of echocardiographic deformation curves for infarct classification. *Int J Cardiovasc Imaging* 2017;**33**:1159–1167.
24. Peressutti D, Sinclair M, Bai W, Jackson T, Ruijsink J, Nordsletten D, Asner L, Hadjicharalambous M, Rinaldi CA, Rueckert D, King AP. A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: application to cardiac resynchronisation therapy response prediction. *Med Image Anal* 2017;**35**:669–684.
25. Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, Dahlström U, O'Connor CM, Felker GM, Desai NR. Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. *J Am Heart Assoc* 2018;**7**:1–15.
26. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart* 2018;**104**: 1156–1164.
27. Castelvecchi D. Can we open the black box of AI? *Nature* 2016;**538**:20–23.
28. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, eds. *Advances in Neural Information Processing Systems 25*. Red Hook, NY: Curran Associates, Inc.; 2012. pp1097–1105.
29. Hinton GE, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 2012;**29**:82–97.
30. Kalscheur MM, Kipp RT, Tattersall MC, Mei C, Buhr KA, DeMets DL, Field ME, Eckhardt LL, Page CD. Machine learning algorithm predicts cardiac resynchronization therapy outcomes. *Circ Arrhythm Electrophysiol* 2018;**11**:e005499.