


# Machine learning approach to stratify complex heterogeneity of chronic heart failure: A report from the CHART-2 study

Kenji Nakano<sup>1</sup>, Kotaro Nochioka<sup>2</sup>, Satoshi Yasuda<sup>2</sup>, Daito Tamori<sup>1</sup>, Takashi Shiroto<sup>2</sup>, Yudai Sato<sup>1</sup>, Eichi Takaya<sup>3</sup>, Satoshi Miyata<sup>4</sup>, Eiryō Kawakami<sup>1,5,6,7</sup>, Tetsuo Ishikawa<sup>1,5,6,8</sup>, Takuya Ueda<sup>1,3\*</sup>  and Hiroaki Shimokawa<sup>2,9</sup>

<sup>1</sup>Department of Clinical Imaging, Tohoku University Graduate School of Medicine, Sendai, Japan; <sup>2</sup>Department of Cardiovascular Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; <sup>3</sup>AI Lab, Tohoku University Hospital, Sendai, Japan; <sup>4</sup>Teikyo University Graduate School of Public Health, Tokyo, Japan; <sup>5</sup>Advanced Data Science Project (ADSP), RIKEN Information R&D and Strategy Headquarters, Yokohama, Japan; <sup>6</sup>Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan; <sup>7</sup>Institute for Advanced Academic Research (IAAR), Chiba University, Chiba, Japan; <sup>8</sup>Department of Extended Intelligence for Medicine, The Ishii-Ishibashi Laboratory, Keio University School of Medicine, Tokyo, Japan; and <sup>9</sup>International University of Health and Welfare, Otawara, Japan

## Abstract

**Aims** Current approaches to classify chronic heart failure (HF) subpopulations may be limited due to the diversity of pathophysiology and co-morbidities in chronic HF. We aimed to elucidate the clusters of chronic patients with HF by data-driven approaches with machine learning in a hospital-based registry.

**Methods and results** A total of 4649 patients with a broad spectrum of left ventricular ejection fraction (LVEF) in the CHART-2 (Chronic Heart Failure Analysis and Registry in the Tohoku District-2) study were enrolled to this study. Chronic HF patients were classified using random forest clustering with 56 multiscale clinical parameters. We assessed the influence of the clusters on cardiovascular death, non-cardiovascular death, all-cause death, and free from hospitalization by HF. Latent class analysis using random forest clustering identified 10 clusters with four primary components: cardiac function (LVEF, left atrial and ventricular diameters, diastolic blood pressure, and brain natriuretic peptide), renal function (glomerular filtration rate and blood urea nitrogen), anaemia (red blood cell, haematocrit, haemoglobin, and platelet count), and nutrition (albumin and body mass index). All 11 significant clinical parameters in the four primary components and two disease aetiologies (ischaemic heart disease and valvular heart disease) showed statistically significant differences among the 10 clusters ( $P < 0.01$ ). Cluster 1 (26.7% of patients), which is characterized by preserved LVEF ( $< 59%$ , 37% of the total) with lowest brain natriuretic peptide ( $> 111.3$  pg/mL, 0.9%) and lowest left atrial diameter ( $> 42$  mm, 37.4%), showed the best 5 year survival rate of 98.1% for cardiovascular death, 95.9% for non-cardiovascular death, 92.9% for all-cause death, and 91.7% for free from hospitalization by HF. Cluster 10 (6.0% of the total), which is co-morbid disorders of all four primary components, showed the worst survival rate of 39.1% for cardiovascular death, 68.9% for non-cardiovascular death, 23.9% for all-cause death, and 28.1% for free from hospitalization by HF.

**Conclusions** These results suggest the potential applicability of the machine learning approach, providing useful clinical prognostic information to stratify complex heterogeneity in patients with HF.

**Keywords** Heart failure; Cohort study; Clustering; Machine learning; Prognosis

Received: 15 September 2022; Revised: 26 November 2022; Accepted: 9 January 2023

\*Correspondence to: to Takuya Ueda, MD, PhD, Department of Clinical Imaging, Tohoku University Graduate School of Medicine, Seiryō-machi 2-1, Aoba-ku, Sendai, Miyagi 980-8575, Japan. Tel: +81-22-717-7481; Fax: +81-22-717-7944.

Email: [takuya.ueda.d3@tohoku.ac.jp](mailto:takuya.ueda.d3@tohoku.ac.jp)

Drs. Kenji Nakano and Kotaro Nochioka equally contributed to this work.

## Introduction

Heart failure (HF) is a global burden affecting 640 million people and is regarded as the leading cause of death and

morbidity worldwide.<sup>1</sup> Current guidelines provide three chronic HF categories based on left ventricular ejection fraction (LVEF): heart failure with reduced (HFrEF), mildly reduced (HFmrEF), and preserved ejection fraction (HFpEF).<sup>1</sup>

Although LVEF has been used as a major parameter for the categorization and management of patients with HF, this approach for HF risk stratification may have fundamental limitations.<sup>2</sup> LVEF above 45% is not further considered for prognostic assessment in patients with chronic HF.<sup>3</sup> Furthermore, several studies suggested that LVEF alone is not sufficient to stratify the multi-factorial and heterogeneous nature of HF.<sup>4–6</sup> A recent large cohort study suggested that overall, adjusted hazard ratios for mortality showed a U-shaped relationship for LVEF with a nadir of risk noted at 60–65%.<sup>7</sup> An improved phenotypic classification of chronic HF with an integration of clinical parameters and biomarkers would provide useful information for optimal patient care and management strategies, providing much better predictive value than LVEF alone.<sup>6,7</sup>

Recently, machine learning (ML) algorithm such as random forest (RF) has been introduced to provide precise risk stratification beyond existing classification in patients with ovarian cancer,<sup>8</sup> breast cancer,<sup>9</sup> and hypokalaemia.<sup>10</sup> This approach is entirely hypothesis free and has the potential for discovery of novel insights in chronic HF.<sup>8</sup>

In this study, we aimed to derive the clusters of chronic HF patients by data-driven ML approach with 56 parameters, including physical data, aetiology, blood examination, echocardiography, urinalysis, and medication, in order to reveal the long-term prognostic relevance of clustering for death and free from hospitalization by HF in our CHART-2 (Chronic Heart Failure Analysis and Registry in the Tohoku District-2) study, one of the largest multicentre prospective observational studies on chronic HF patients.<sup>11</sup>

## Methods

### Study setting and subjects

The CHART-2 study is a hospital-based prospective observational study with 23 hospitals in six prefectures in Japan.<sup>11</sup> The design and methods have been previously described in detail.<sup>11</sup> In brief, between October 2006 and March 2010, we enrolled consecutive patients older than 20 years with significant coronary artery disease and those in Stage B (structural heart disease but without signs or symptoms of HF), Stage C (structural heart disease with early or current symptoms of HF), and Stage D enumerated by the current guidelines.<sup>1,12</sup> Subjects in Stage B must meet at least one of the following structural disorders and must not have signs, symptoms, or history of hospitalization for HF: (i) enlarged left ventricular (LV) end-diastolic dimension ( $\geq 55$  mm) measured by echocardiography; (ii) impaired LV ejection fraction (LVEF  $\leq 50\%$ ) measured by echocardiography; (iii) thickened interventricular septum ( $>12$  mm) and/or thickened LV posterior wall ( $>12$  mm) measured by echocardiography; (iv)

significant valvular stenosis/insufficiency; (v) significant myocardial abnormalities; (vi) congenital abnormalities, or (vii) history of cardiac surgery.<sup>11,13</sup> The diagnosis of Stage C was made by attending cardiologists based on the criteria of the Framingham study.<sup>14</sup> Enrolment began in October 2006 and ended in March 2010.<sup>11</sup> All information, including medical history, laboratory data, and echocardiography data, were recorded in a computer database at the time of enrolment. Annual follow-up was made by clinical research coordinators by means of review of medical records, surveys, and telephone interviews.<sup>11</sup>

After excluding 5343 patients in Stage A/B and 227 with missing echocardiographic data, we included a total of 4649 chronic HF patients with Stage C/D (*Figure S1*). The study outcomes included cardiovascular death, non-cardiovascular death, all-cause death, and free from hospitalization by HF. All outcomes were reviewed and adjudicated by consensus of three independent physicians, the members of the Tohoku Heart Failure Association.<sup>11</sup> They reviewed case reports, death certificates, medical records, and summaries provided by the investigators. This study conformed to the Declaration of Helsinki, and the study protocol was approved by institutional review boards at each institution. All participants provided written informed consent.

### Clinical parameters and random forest model

For RF clustering, we included 56 multiscale clinical parameters with  $<30\%$  missing rates, as displayed in *Table S1*. From the RF modelling, we excluded age and sex as they are clinically neither intervenable nor treatable. Missing values were complemented with the missForest algorithm, one of the major imputation algorithms utilizing RF.<sup>15</sup> After excluding 901 patients who died of non-cardiovascular death or censoring, we built a RF model to predict cardiovascular death within 5 years in 3748 patients (RF modelling dataset, *Figure S1*). We randomly split 3748 patients into 3423 patients for training data and 358 patients for test data. Based on the fixed training and test dataset, we built 10 supervised RF models by changing initial random seeds. RF classifier is composed from an ensemble of decision trees, bagging, and random feature selection. In bagging, each tree is trained based on a bootstrap sample of training data. In the training process, each tree grows from a particular bootstrap sample. During the training process, the performance of RF prediction for cardiovascular death was evaluated from out-of-bag samples, which are not selected in the bootstrap sample. The area under the receiver operating characteristic curve (AUC) was used to assess the performance of the RF models in 358 patients. The variables that contribute to predict cardiovascular death above LVEF were selected. After the RF models were established, we input 4649 patients overall into RF models without outcome data to archive RF proximity, which is

defined as the frequency with which two cases are classified into the same leaf in the decision trees of the RF model. Based on the proximity matrix from the branch information, 2D embedding was performed with densMAP.<sup>16</sup> Clustering was carried out by applying k-means to the 2D coordinates. The optimal number of clusters was selected from 2 to 15, considering the silhouette score and the silhouette plot.<sup>17</sup> After clustering, we compared baseline characteristics of each cluster by providing ratio to exceed median value of each significant parameter. In overall 4649 patients, Kaplan–Meier survival curves for cardiovascular death, non-cardiovascular death, all-cause death, and free of hospitalization by HF of the 10 clusters were described and 5 year survival rates were calculated according to the previously reported method.<sup>8</sup>

## Statistical analysis

The differences in clinical parameters were compared among clusters obtained from combining multiple ML algorithms. For statistical analysis, ANOVA was performed to compare continuous variables and chi-squared test for categorical variables. A two-sided *P* value of <0.05 was considered statistically significant. All analysis were performed with R version 4.0.2 (R Foundation for Statistical Computing, <https://www.R-project.org/>) and Python Version 3.9.7 (Python Software Foundation at <http://www.python.org>).

## Results

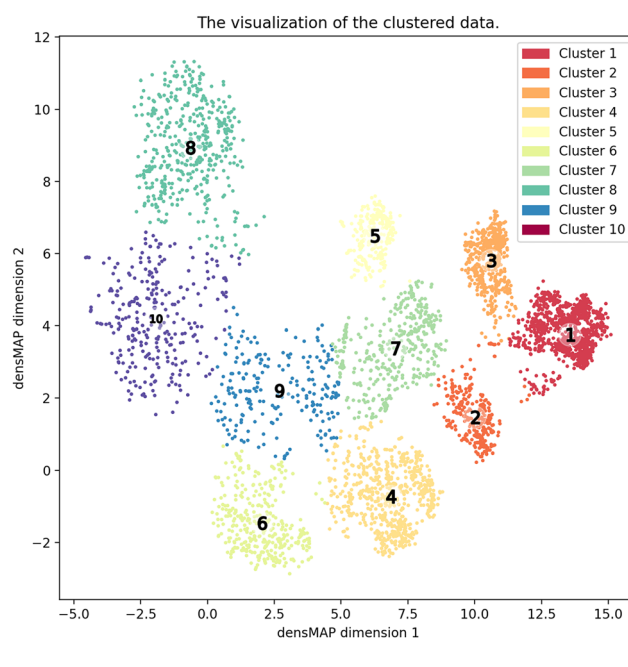
### Supervised RF model to predict cardiovascular death within 5 years

The mean age of the 4649 patients was  $68.9 \pm 12.3$  years, 32.0% were women, 2.0% were in Stage D, and 68.3% had HF with preserved EF (LVEF  $\geq 50\%$ ). The median [IQR] level of brain natriuretic peptide (BNP) was 111.3 [49.9 to 223.2] pg/mL. When checking the performance of the RF models to predict cardiovascular death within 5 years, the highest, lowest, and average values of AUC were 0.815, 0.811, and 0.813, respectively (Figure S2). The relative importance of variables (RI) for cardiovascular death within 5 years was calculated by averaging the importance of variables from each of the RF models. In this model, 13 variables showed higher RI than LVEF, including BNP, blood urea nitrogen (BUN), estimated glomerular filtration rate (eGFR), red blood cell count, haematocrit, albumin, haemoglobin, left atrial diameter (LAD), diastolic blood pressure (dbp), platelet count, body mass index (BMI), and LV end-systolic diameter (LVDs) (Figure S3).

### Clusters by RF model

The RF clustering divided 4649 Stage C/D patients into 10 clusters based on the silhouette score (Figure 1, Table S2). Of these, 1239 patients (26.7%) were clustered as Cluster 1, 286 (6.2%) Cluster 2, 533 (11.5%) Cluster 3, 664 (14.3%) Cluster 4, 247 (5.3%) Cluster 5, 356 (7.7%) Cluster 6, 372 (8.0%) Cluster 7, 451 (9.7%) Cluster 8, 223 (4.8%) Cluster 9, and 278 (6.0%) Cluster 10. Table S2 shows the baseline characteristics across the clusters, and Table 1 shows the baseline characteristics of the 10 clusters by displaying percentage of patients above median of each significant parameter. Cluster 1 (reference) was characterized by lowest BNP (BNP > 113 pg/mL, 0.9% of total) and lowest LAD diameter (LADim > 42 mm, 37.4%). When stratified by LVEF, Cluster 1 (mean LVEF, 61.9%), Cluster 2 (62%), Cluster 3 (60.5%), Cluster 4 (60.7%), Cluster 6 (58%), Cluster 7 (54.2%), and Cluster 9 (55.6%) were categorized as HFpEF, whereas Cluster 5 (31.6%) as HFrEF and Cluster 8 (48%) and Cluster 10 (50.1%) as HFmrEF. In detail, Cluster 1 also had preserved diastolic pressure (dbp < 70 mmHg, 18.4%), preserved LVEF, and the highest BMI (>23.5 pg/mL, 62.4%) among the 10 clusters. Cluster 2 had similar prevalence of preserved LVEF but higher BNP and lower dbp than did Cluster 1. Cluster 3 had higher BNP level and LADim than did Cluster 1. Cluster 4 was characterized by impaired renal function. Cluster 5 had impaired LV function (LVEF < 59%, 99.6%) but preserved renal function

**Figure 1** 2D visualization of the relative distances among all patients with chronic HF. Colours indicate different cluster assignment using k-means clustering (k = 10).



**Table 1** Percentage gradient scale (PGS) of primary components among the 10 clusters by random-forest clustering

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
# of patients (%)	1239 (26.7)	286 (6.2)	533 (11.5)	664 (14.3)	247 (5.3)	356 (7.7)	372 (8.0)	451 (9.7)	223 (4.8)	278 (6.0)
Cardiac functions										
PGS	0–20%	0–20%	20–40%	46.2	40–60%	57.3	60–80%	100.0	80–100%	
BNP (>111.3 pg/mL)	0.9	35.0	85.2	46.2	53.8	57.3	58.6	100.0	68.2	98.9
LADim (>42 mm)	37.4	42.3	60.8	51.8	55.9	55.6	52.4	62.7	54.7	67.6
dBp (<70 mmHg)	18.4	96.5	38.5	46.2	51.0	52.2	57.5	57.9	65.5	67.3
LVDs (>35 mm)	37.4	34.3	39.4	35.8	98.0	44.1	54.8	69.6	46.2	65.5
LVEF (<59%)	37.0	36.4	42.3	40.2	99.6	45.8	59.4	69.6	49.3	65.1
Renal functions										
PGS	0–20%	0–20%	20–40%	67.2	40–60%	100.0	60–80%	51.4	80–100%	
BUN (>17.5 mg/dL)	26.2	34.3	29.8	67.2	34.4	100.0	39.2	51.4	90.1	96.8
eGFR (<61.1 mL/min/1.73 m <sup>2</sup> )	15.1	29.7	15.6	100.0	22.7	100.0	46.0	49.7	97.8	97.1
Anaemia										
PGS	0–20%	0–20%	20–40%	53.9	40–60%	71.1	60–80%	56.8	80–100%	
RBC (<428 10 <sup>4</sup> /μL)	27.8	45.1	37.8	53.9	27.5	71.1	66.1	56.8	90.6	89.6
Ht (<39.9%)	27.4	46.5	35.6	55.7	28.3	71.3	68.5	54.3	92.8	89.2
Hb (<13.3 g/dL)	24.5	48.3	36.5	55.1	25.1	70.8	67.7	55.7	92.8	89.2
Plt (<198 000/μL)	39.4	51.0	55.3	51.2	42.5	56.5	46.0	52.8	58.7	68.7
Nutrition										
PGS	0–20%	0–20%	20–40%	43.1	40–60%	57.0	60–80%	70.5	80–100%	
Alb (<4.11 g/dL)	24.2	45.1	41.6	43.1	40.5	57.0	89.5	70.5	75.8	90.6
BMI (>23.5 kg/m <sup>2</sup> )	62.4	41.6	49.7	57.5	58.7	48.6	39.2	33.0	39.0	28.4
HF aetiology										
PGS	0–20%	0–20%	20–40%	52.11	40–60%	48.31	60–80%	34.81	80–100%	
IHD	55.13	39.16	40.34	52.11	44.13	48.31	45.43	34.81	53.36	49.64
VHD	23.65	44.06	33.96	31.63	16.6	33.43	30.91	33.7	40.36	38.85

Note: The numbers indicate percentage of patients who present above median of each significant parameter.

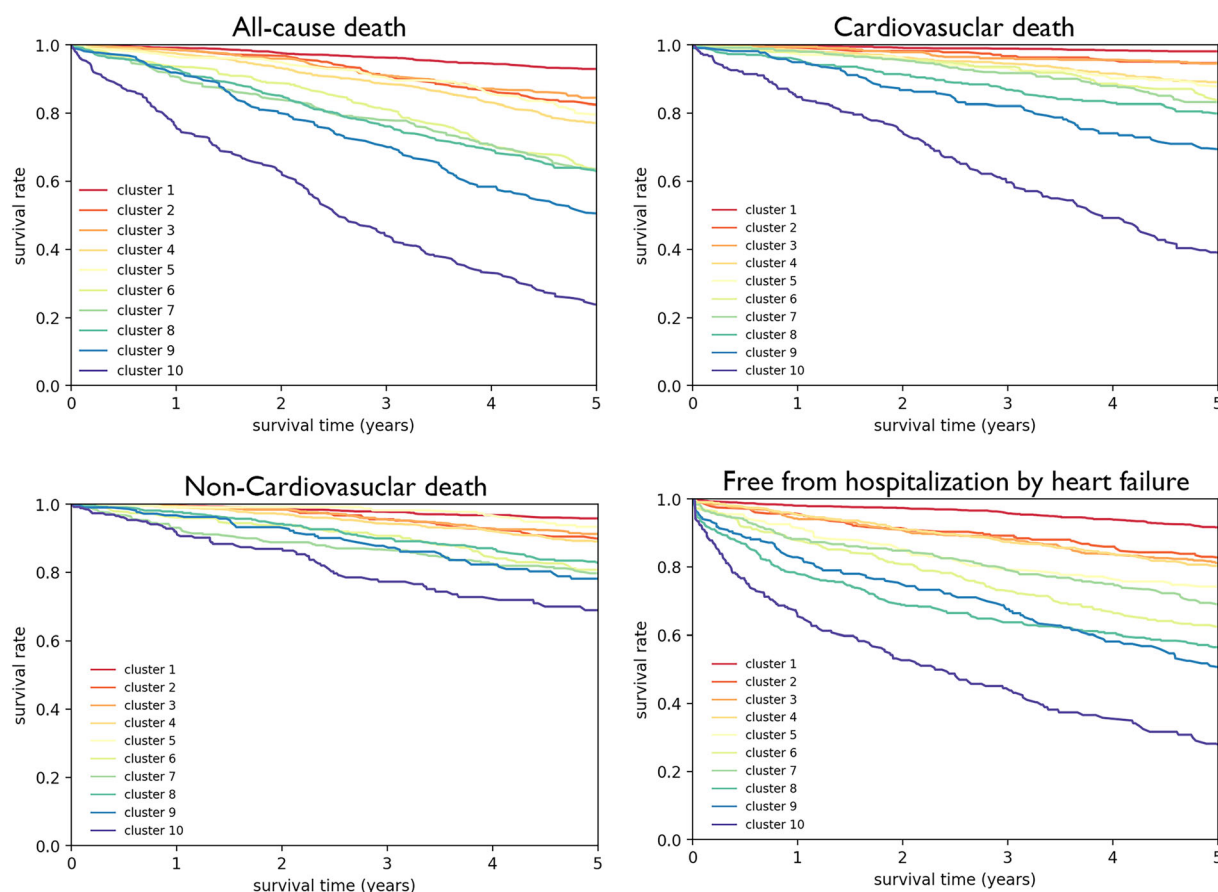
(eGFR < 61.1 mL/min/1.73 m<sup>2</sup>, 22.7%) and highest Hb level (Hb < 13.3 g/dL, 25.1%) among the 10 clusters. Cluster 6 had preserved LV function (LVEF < 59%, 59.4%) but impaired renal function (eGFR < 61.1 mL/min/1.73 m<sup>2</sup>, 100%) and anaemia (Hb < 13.3 g/dL, 70.8%). Cluster 7 was characterized by lower Hb (Hb < 13.3 g/dL, 67.7%) and Alb (< 4.11 g/dL, 89.5%) levels. Cluster 8 and Cluster 10 shared clinical characteristics in cardiac functions. However, Cluster 10 had higher prevalence of anaemia compared with Cluster 8. Cluster 9 and Cluster 10 had similar characteristics in renal dysfunction and anaemia, but Cluster 9 had better cardiac function than Cluster 10 (Table 1).

### Prognostic relevance of clustering

Median follow-up period was 5.9 years. We assessed the association between clusters and outcomes (cardiovascular death, non-cardiovascular death, all-cause death, and free from hospitalization by HF). Figure 2 shows Kaplan–Meier curves for all-cause death, cardiovascular death,

non-cardiovascular death, and free from hospitalization by HF among the 10 clusters. The 5 year survival rates for cardiovascular death were sequentially decreased from Cluster 1 to Cluster 10 as follows: 98.1% in Cluster 1, 94.7% in Cluster 2, 94.5% in Cluster 3, 89.1% in Cluster 4, 87.8% in Cluster 5, 83.9% in Cluster 6, 83.2% in Cluster 7, 80.0% in Cluster 8, 69.7% in Cluster 9, and 39.1% in Cluster 10. Similar trends were noted in non-cardiovascular death (95.9% in Cluster 1, 90.1% in Cluster 2, 91.3% in Cluster 3, 89.1% in Cluster 4, 93.4% in Cluster 5, 80.8% in Cluster 6, 79.8% in Cluster 7, 82.9% in Cluster 8, 78.2% in Cluster 9, and 68.9% in Cluster 10) and all-cause death (92.9% in Cluster 1, 82.7% in Cluster 2, 84.5% in Cluster 3, 77.2% in Cluster 4, 79.6% in Cluster 5, 63.7% in Cluster 6, 63.6% in Cluster 7, 63.0% in Cluster 8, 50.8% in Cluster 9, and 23.9% in Cluster 10). Furthermore, rates for free from hospitalization by HF clearly decreased from Cluster 1 to Cluster 10 as follows: 91.7% in Cluster 1, 83.0% in Cluster 2, 81.2% in Cluster 3, 80.3% in Cluster 4, 74.3% in Cluster 5, 62.5% in Cluster 6, 69.1% in Cluster 7, 56.5% in Cluster 8, 51.0% in Cluster 9, and 28.1% in Cluster 10 (Figure 2).

**Figure 2** Kaplan–Meier survival curves for (A) all-cause death, (B) cardiovascular death, (C) non-cardiovascular death, and (D) free from hospitalization by HF among the clusters. Curves are truncated at 5 years.



## Discussion

The present study is one of the largest studies that clustered 4649 chronic HF patients with a broad spectrum of LVEF, demonstrating that our data-driven ML approach is able to identify 10 distinct clinical clusters of patients with four primary components: cardiac function, renal function, anaemia, and nutrition. These results demonstrate that the ML approach is useful to stratify complex heterogeneity of chronic HF, suggesting its potential applicability for prognostic assessment of chronic HF patients.

Given by the limitation of LVEF for risk prediction, the current classifications of chronic HF by LVEF need to be improved.<sup>12,18</sup> By clustering disease, ML approach can reduce the dimensionality of features in a dataset with multimodal variables to understand and characterize the real-world manifestation of HF. Previous ML studies focused on patients with preserved LVEF given by the heterogeneous clinical syndrome of HFpEF in order to improve phenotypic classification.<sup>19,20</sup> Shah *et al.* studied 397 patients with HFpEF and performed detailed clinical, laboratory, ECG, and echocardiographic phenotyping of the patients. Using several statistical learning algorithms, they were able to classify study participants into three distinct groups that differed markedly in clinical characteristics, cardiac structure/function, invasive haemodynamics, and outcomes.<sup>19</sup> Uijl *et al.* also studied two large contemporary HF registries with over 9000 HFpEF patients between 2013 and 2016.<sup>20</sup> They identified five distinct clinical clusters of patients in HFpEF, including a young-low co-morbidity burden cluster, an atrial fibrillation-hypertensive cluster, an older-atrial fibrillation cluster, an obese-diabetic cluster, and a cardio-renal cluster.<sup>20</sup> These findings indicate that HFpEF is indeed a heterogeneous disorder.

By demonstrating the prognostic importance of clustering in patients with chronic HF and a broad spectrum of LVEF, our findings extend on the clinical utility of ML approaches showing the association of clusters and the outcomes. There are several clinically useful risk prediction models in HF patients that are limited by linear assumption between baseline characteristics and outcomes.<sup>21</sup> However, our ML clustering approaches enable non-linear stratification of disease status, considering various background pathological conditions, and allow stratification of HF phenotypes with different prognoses.

Of note, Clusters 1, 2, 3, 5, and 8 were mainly stratified by cardiac function biomarkers, suggesting a series of progression of HF without any non-cardiovascular co-morbidities. Clusters 4 and 6 were stratified by renal function biomarkers as well as cardiac function, suggesting the importance of cardio-renal relationships in HF progression.<sup>22</sup> Clusters 9 and 10 were characterized by multimorbidity, and as expected, they had poor prognosis. The present study provides further understanding of the complex HF pathophysiology and may provide chances of more personalized treatment

of HF patients. Importantly, we demonstrate that our ML approach is able to produce an automated and scalable understanding of a large population of patients with chronic HF. Our approach also identified 13 important parameters (*Figure S3*) associated with cardiac function (LVEF, LVDim, LVDs, dbp, and BNP), renal function (BUN and eGFR), presence of anaemia (RBC, Ht, and Hb), and nutrition (Alb) that can serve as a foundation for practice-based medicine for clinicians when considering various pathological conditions. It remains to be examined in future studies how HF pathology changes over time and leads to outcomes to establish personalized care and preventive medicine for HF patients.

Several studies have reported using unsupervised ML to cluster patients with HFpEF.<sup>23,24</sup> Unsupervised ML is useful to understand unrecognized patterns and trends within unlabelled data. The advantage of our CHART-2 study dataset, when compared with previous studies, is its inclusion of long-term prognostic data.<sup>23,24</sup> In the present study, we used supervised ML because it is fundamentally specialized to predict outcomes for unseen data and the definitive purpose of our research was to stratify patient groups with different long-term outcomes. The identified number of clusters, in our study, was larger than that in previously published papers because a wide range of patients with HF, including those with HFpEF and HFrEF with more complicated clinical backgrounds, were included in our datasets.<sup>23,24</sup>

Although further evidence is needed to determine patient management strategy, the identification of mutually exclusive phenotypes in patients with HF increases the reasoning of clinical benefit. For example, although anaemia is a known prognostic factor in patients with HF, it is still controversial as to what type of patients with HF and anaemia would benefit from anaemia treatment.<sup>25</sup> Recently, sodium–glucose cotransporter 2 inhibitors (SGLT2i) have been proven to be effective for patients with HF over a wide range of LVEF.<sup>26</sup> The identification of the HF phenotype that will benefit from specific target treatment (e.g. SGLT2i) may also aid in future clinical trials to determine treatment options. The detailed risk stratification of the HF phenotype may also provide insights in tailor-made follow-up strategies for patients with HF.

Several limitations of the present study should be mentioned. First, the data-driven ML approach for phenotypic clustering is highly influenced by cohort characteristics. Our cohort included only Asian patients and a relatively high prevalence of HFpEF. Thus, our findings need to be confirmed in other populations. Second, in this study, 56 clinical parameters had missing values that may cause selection bias. However, we complemented the values with the missForest algorithm, one of the major imputation algorithms for RF classification that would minimize the bias. Third, our analysis only used the baseline data and did not consider the transitional changes of the patients. For a better prediction of the long-term outcome, a time-course analysis of the data

needs to be considered. Fourth, in this study, the silhouette score is used to determine the optimal number of clusters as it is the most widely used method for evaluating the performance of clustering. Application of other indices may be considered to obtain other results regarding the optimal number of clusters.

## Conclusions

In a large hospital-based cohort of chronic HF patients, the CHART-2 study, we were able to demonstrate the novel clustering of chronic HF with four primary components (cardiac function, renal function, presence of anaemia, and nutrition) that had a diverse range of mortality rate. The ML approach provides clinical information to stratify complex heterogeneity of chronic HF, suggesting its potential applicability for prognostic assessment in chronic HF patients. Further clinical validation and longitudinal analysis are warranted.

## Acknowledgements

The authors wish to thank the members of the Tohoku Heart Failure Society and the staff and participants of the CHART-2 study for their important contributions.

## Conflict of interest

Nothing to disclose.

## References

- Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, Deswal A, Drazner MH, Dunlay SM, Evers LR, Fang JC, Fedson SE, Fonarow GC, Hayek SS, Hernandez AF, Khazanie P, Kittleson MM, Lee CS, Link MS, Milano CA, Nwacheta LC, Sandhu AT, Stevenson LW, Vardeny O, Vest AR, Yancy CW. 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.
- 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.
- Solomon SD, Anavekar N, Skali H, McMurray JJ, Swedberg K, Yusuf S, Granger CB, Michelson EL, Wang D, Pocock S, Pfeffer MA. Candesartan in heart failure reduction in mortality I. Influence of ejection fraction on cardiovascular outcomes in a broad spectrum of heart failure patients. *Circulation*. 2005; **112**: 3738–3744.
- Shah SJ, Katz DH, Deo RC. Phenotypic spectrum of heart failure with preserved ejection fraction. *Heart Fail Clin*. 2014; **10**: 407–418.
- Pfeffer MA, Shah AM, Borlaug BA. Heart failure with preserved ejection fraction in perspective. *Circ Res*. 2019; **124**: 1598–1617.
- Triposkiadis F, Butler J, Abboud FM, Armstrong PW, Adamopoulos S, Atherton JJ, Backs J, Bauersachs J, Burkhoff D, De Keulenaer GW, et al. The continuous heart failure spectrum: moving beyond an ejection fraction classification. *Eur Heart J*. 2019; **40**: 2155–2163.
- Wehner GJ, Jing L, Haggerty CM, Suever JD, Leader JB, Hartzel DN, Kirchner HL, Manus JNA, James N, Ayar Z, Gladding P, Good CW, Cleland JGF, Fornwalt BK. Routinely reported ejection fraction and mortality in clinical practice: where does the nadir of risk lie? *Eur Heart J*. 2020; **41**: 1249–1257.
- Kawakami E, Tabata J, Yanaihara N, Ishikawa T, Koseki K, Iida Y, Saito M, Komazaki H, Shapiro JS, Goto C, Akiyama Y, Saito R, Saito M, Takano H, Yamada K, Okamoto A. Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. *Clin Cancer Res*. 2019; **25**: 3006–3015.
- Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, West RB, van de Rijn M, Koller D. Systematic analysis of breast cancer morphology

## Funding

The work for this manuscript was supported in part by Japan Agency for Medical Research and Development (AMED) grants 22ek0210136h0003 and 22ek0109543h0002 (K.N.); the Japan Society for the Promotion of Science (JSPS) KAKENHI grant JP20K21837 (T.I.); the CREST Program of the Japan Science and Technology Agency (JST), CREST Grant Number JPMJCR15D1 (T.U.); and Phillips Co-creation grant J220000243 (T.U.).

## Supporting information

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

**Table S1.** Missing value rates in the 56 parameters for random forest clustering

**Table S2.** Baseline characteristics of the 10 clusters

**Figure S1.** Flow diagram of patient enrollment to this study

**Figure S2.** Area under curves (AUC) for assessment of the validity of the 10 RF models. The highest, lowest, and average values of AUC were 0.815, 0.811 and 0.813, respectively.

**Figure S3.** The relative importance of variables (RI) for 5-year CVD death among identified 13 parameters

**Figure S4.** Silhouette plots and scores of the 2–15 clusters. The silhouette plot of the 10 clusters (\*) showed the highest silhouette score (0.587).

- uncovers stromal features associated with survival. *Sci Transl Med.* 2011; **3**: 108ra113.
10. Thongprayoon C, Mao MA, Kattah AG, Keddis MT, Pattharanitima P, Erickson SB, Dillon JJ, Garovic VD, Cheungpasitporn W. Subtyping hospitalized patients with hypokalemia by machine learning consensus clustering and associated mortality risks. *Clin Kidney J.* 2022; **15**: 253–261.
  11. Shiba N, Nochioka K, Miura M, Kohno H, Shimokawa H. Trend of westernization of etiology and clinical characteristics of heart failure patients in Japan. First report from the CHART-2 study. *Circ J.* 2011; **75**: 823–833.
  12. Bozkurt B, Coats AJS, Tsutsui H, Abdelhamid CM, Adamopoulos S, Albert N, Anker SD, Atherton J, Bohm M, Butler J, Drazner MH, Michael Felker G, Filippatos G, Fiuzat M, Fonarow GC, Gomez-Mesa JE, Heidenreich P, Imamura T, Jankowska EA, Januzzi J, Khazanie P, Kinugawa K, Lam CSP, Matsue Y, Metra M, Ohtani T, Francesco Piepoli M, Ponikowski P, Rosano GMC, Sakata Y, Seferovic P, Starling RC, Teerlink JR, Vardeny O, Yamamoto K, Yancy C, Zhang J, Zieroth S. Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. *Eur J Heart Fail.* 2021; **23**: 352–380.
  13. Nochioka K, Yasuda S, Sakata Y, Shiroto T, Hayashi H, Takahashi J, Takahama H, Miyata S, Shimokawa H. Prognostic impact of a history of cancer and atrial fibrillation in antithrombotic therapy for chronic heart failure. *ESC Heart Fail.* 2022; **9**: 2445–2454.
  14. McKee PA, Castelli WP, McNamara PM, Kannel WB. The natural history of congestive heart failure: the Framingham study. *N Engl J Med.* 1971; **285**: 1441–1446.
  15. Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics.* 2012; **28**: 112–118.
  16. Narayan A, Berger B, Cho H. Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nat Biotechnol.* 2021; **39**: 765–774.
  17. Rousseeuw J, Silhouettes P. A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Mathematics.* 1987; **20**: 53–65.
  18. Lam CSP, Yancy C. Universal definition and classification of heart failure: is it universal? Does it define heart failure? *J Card Fail.* 2021; **27**: 509–511.
  19. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghide M, Bonow RO, Huang CC, Deo RC. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation.* 2015; **131**: 269–279.
  20. Ujil A, Savarese G, Vaartjes I, Dahlstrom U, Brugts JJ, Linssen GCM, van Empel V, Brunner-La Rocca HP, Asselbergs FW, Lund LH, Hoes AW, Koudstaal S. Identification of distinct phenotypic clusters in heart failure with preserved ejection fraction. *Eur J Heart Fail.* 2021; **23**: 973–982.
  21. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, Woodward M, Patel A, McMurray J, MacMahon S. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail.* 2014; **2**: 440–446.
  22. Ronco C, McCullough P, Anker SD, Anand I, Aspromonte N, Bagshaw SM, Bellomo R, Berl T, Bobek I, Cruz DN, Daliento L, Davenport A, Haapio M, Hillege H, House AA, Katz N, Maisel A, Mankad S, Zanco P, Mebazaa A, Palazzuoli A, Ronco F, Shaw A, Sheinfeld G, Soni S, Vescovo G, Zamperetti N, Ponikowski P. Acute dialysis quality initiative consensus group. Cardio-renal syndromes: report from the consensus conference of the acute dialysis quality initiative. *Eur Heart J.* 2010; **31**: 703–711.
  23. Segar MW, Patel KV, Ayers C, Basit M, Tang WHW, Willett D, Berry J, Grodin JL, Pandey A. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail.* 2020; **22**: 148–158.
  24. Woolley RJ, Ceelen D, Ouwerkerk W, Tromp J, Figarska SM, Anker SD, Dickstein K, Filippatos G, Zannad F, Metra M, Ng L, Samani N, van Veldhuisen DJ, Lang C, Lam CS, Voors AA. Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction. *Eur J Heart Fail.* 2021; **23**: 983–991.
  25. Maurer MS, Teruya S, Chakraborty B, Helmke S, Mancini D. Treating anemia in older adults with heart failure with a preserved ejection fraction with epoetin alfa: single-blind randomized clinical trial of safety and efficacy. *Circ Heart Fail.* 2013; **6**: 254–263.
  26. Zelniker TA, Wiviott SD, Raz I, Im K, Goodrich EL, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Furtado RHM, Bhatt DL, Leiter LA, McGuire DK, Wilding JPH, Sabatine MS. SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. *Lancet.* 2019; **393**: 31–39.