Interdisciplinary Research in Medical Sciences Specialty
ISSN: 3062-4401
2025, Volume 5, Issue 2, Page No: 167-186

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

Deep Learning–Driven Phenogroup Identification in Heart Failure with Preserved Ejection Fraction and Associated Treatment Effects

Sara Al-Fahd[1]*, Noura Al-Khalifa[1], Omar Al-Turki[2]

[1]Department of Cardiology and Machine Learning Phenotyping, College of Medicine, King Saud University, Riyadh, Saudi Arabia.
[2]Department of Heart Failure Research and Deep Learning, College of Medicine, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

***E-mail** ✉ s.alfahd@ksu.edu.sa

Received: 05 July 2025; **Revised:** 09 September 2025; **Accepted:** 16 September 2025

ABSTRACT

Heart failure with preserved ejection fraction (HFpEF) comprises over half of heart failure presentations, although the availability of proven treatments is restricted owing to the condition's considerable clinical variability. The current investigation sought to uncover discrete HFpEF subtypes by applying a machine learning algorithm and to assess how treatment outcomes vary among these phenogroups. Data from 2147 patients admitted with heart failure and left ventricular ejection fraction (LVEF) ≥50% at Peking University Third Hospital between 2014 and 2023 formed the training set. Phenogroups were derived using a two-stage DeepCluster model built around a fully connected neural network, incorporating 107 demographic and clinical variables obtained from electronic medical records (EMR). Associations with prognosis and treatment effects were examined through Cox proportional hazards models. Model robustness was tested internally with leave-one-out cross-validation and externally using the TOPCAT trial dataset (n = 1696) together with a detailed HFpEF cohort from the University of Michigan Health System (UMHS, n = 128). Analysis revealed three clearly differentiated HFpEF phenogroups. Phenogroup 1 (n = 815) was marked by an extensive load of metabolic comorbidities, left ventricular hypertrophy, and abnormalities in both systolic and diastolic performance. Phenogroup 2 (n = 608) primarily involved female patients who had atrial fibrillation and notable structural alterations in the atria and right ventricle, with diastolic dysfunction as the dominant feature. Phenogroup 3 (n = 724) consisted of younger male individuals with detrimental lifestyle patterns, elevated rates of hyperlipidemia and liver dysfunction, yet relatively intact cardiac structure and function. Phenogroup 1 carried the highest overall risk of death from any cause. Standard heart failure medications did not produce clear survival improvements across the whole study population. Within phenogroup 1, however, the post-diagnosis administration of sodium-glucose cotransporter 2 inhibitors (SGLT2i) corresponded to a 55% lower likelihood of HF rehospitalization (HR = 0.45, 95% CI 0.20–0.99). Angiotensin receptor-neprilysin inhibitors (ARNIs) were associated with a 66% decrease in all-cause mortality risk (HR = 0.34, 95% CI 0.14–0.79). The impact of ARNIs on mortality varied meaningfully between phenogroups (p for interaction = 0.048). In phenogroup 2, treatment with calcium channel blockers linked to diminished risks of both all-cause mortality (HR = 0.62, 95% CI 0.38–0.99) and HF rehospitalization (HR = 0.58, 95% CI 0.38–0.88). The DeepCluster model achieved 96.0% consistency on internal validation. External testing in the TOPCAT and UMHS populations confirmed comparable clinical and pathophysiological patterns for each of the three phenogroups. A machine learning algorithm can successfully distinguish HFpEF phenogroups that exhibit unique clinical characteristics and respond differently to therapies. The results highlight potential benefits of SGLT2i and ARNI for individuals burdened by metabolic comorbidities and dual systolic–diastolic dysfunction, and of calcium channel blockers for those with atrial fibrillation. Additional prospective validation of these observations is warranted.

Keywords: Heart failure with preserved ejection fraction, Phenomapping, Treatment responses, Machine learning

How to Cite This Article: Al-Fahd S, Al-Khalifa N, Al-Turki O. Deep Learning–Driven Phenogroup Identification in Heart Failure with Preserved Ejection Fraction and Associated Treatment Effects. Interdiscip Res Med Sci Spec. 2025;5(2):167-86. https://doi.org/10.51847/8mSfGAZh5v

Introduction

Heart failure (HF) with preserved ejection fraction (HFpEF) constitutes a multifaceted clinical entity distinguished by varied symptom profiles, extensive comorbidity loads, and involvement of systemic pathophysiological mechanisms across multiple organs [1, 2]. Driven by demographic shifts toward older populations and the rising incidence of predisposing conditions, the prevalence of HFpEF is projected to increase further [3–5]. Worldwide, this syndrome accounts for approximately 50% of all HF cases [6]. Within China, affected individuals commonly display a substantial comorbidity burden and adverse long-term outcomes, including 67% of patients presenting with three or more comorbidities and a 1-year cardiovascular mortality rate of 3.1% [7]. The scarcity of proven therapies identified in randomized trials stems largely from the condition's inherent heterogeneity.

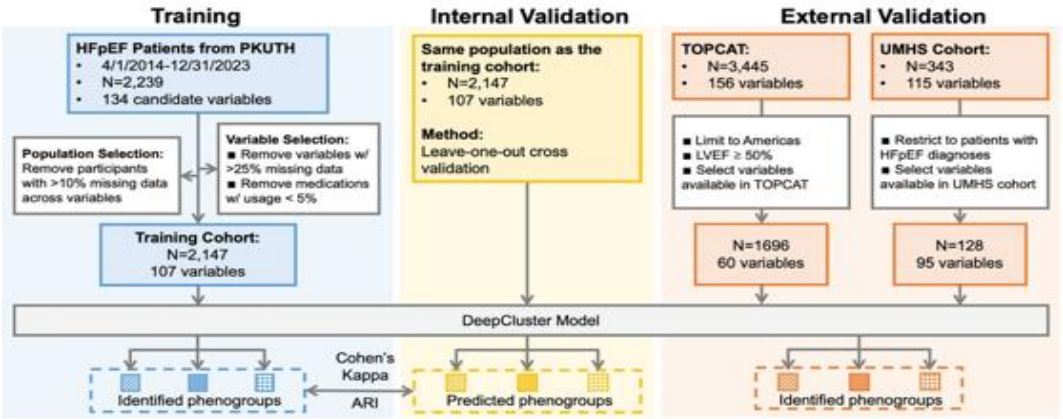
Recent large-scale trials have established the benefits of sodium–glucose cotransporter 2 (SGLT2) inhibitors and the mineralocorticoid receptor antagonist (MRA) finerenone in decreasing the composite risk of HF hospitalization and cardiovascular disease (CVD) death [8–10]. However, the selected study populations in these trials may not encompass the entire clinical spectrum of HFpEF. Trial enrollment typically mandated elevated N-terminal pro–B-type natriuretic peptide (NT-proBNP) concentrations alongside evidence of cardiac remodeling, such as left ventricular hypertrophy or left atrial enlargement. By comparison, prior analyses indicate that nearly 60% of HFpEF patients exhibit normal NT-proBNP levels (NT-proBNP <125 ng/L or <375 ng/L among those with atrial fibrillation) [11]. Additionally, some trials enrolled participants with left ventricular ejection fraction (LVEF) below 50%, a threshold inconsistent with standard HFpEF definitions. HFpEF further spans a wider array of phenotypic presentations—including obese or cardiometabolic, ischemic, and pulmonary vascular subtypes—beyond those captured in these studies. This diversity underscores the necessity for phenotype-informed, individualized management strategies.

In recent years, phenomapping approaches applied to HFpEF have proliferated, leveraging diverse datasets, input variables, and methodological frameworks [12–38]. Such research has commonly delineated between two and seven phenogroups, albeit with considerable feature overlap. The bulk of existing work has emphasized differences in baseline clinical attributes and survival outcomes, leaving real-world insights into treatment responsiveness across phenogroups comparatively underdeveloped. To address this gap, the present study developed and externally validated a machine learning algorithm intended to delineate distinct HFpEF phenotypes, while also evaluating their clinical profiles, prognostic trajectories, and differential responses to therapy across three independent, comprehensively characterized patient cohorts.

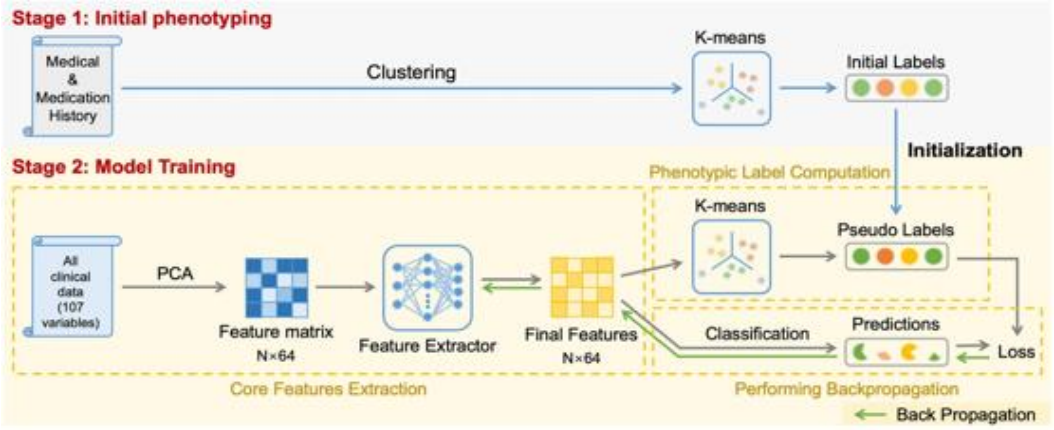
Materials and Methods

Study population

For model development, 6763 patients hospitalized with a diagnosis of HF and suspected HFpEF at Peking University Third Hospital (PKUTH) were screened between April 1, 2014, and December 31, 2023. HFpEF was defined according to the presence of characteristic HF symptoms combined with an LVEF of 50% or above during the index admission (the initial hospitalization for HF) [4, 39]. Following application of eligibility criteria, the final training cohort included 2239 patients with confirmed HFpEF **(Figure 1a)**.



a)



b)

Figure 1. (a) Study flowchart and (b) schematic overview of the two-stage unsupervised DeepCluster model.

(a) Flowchart illustrating the selection of HFpEF patients and phenotypic variables in the training cohort and the two external validation cohorts. (b) The phenotyping procedure employed a two-stage unsupervised DeepCluster model designed to emulate the stepwise diagnostic reasoning used in clinical practice. The framework comprises an initialization phase and an iterative optimization phase, both intended to improve the precision and refinement of phenotypic representations in HFpEF, thereby facilitating more accurate subgroup identification. In the first stage, K-means clustering was performed on medical and medication history data to derive initial phenotypic pseudo-labels. These pseudo-labels served as preliminary classifications to direct the subsequent optimization of the DeepCluster model. In the second stage, the DeepCluster model conducted iterative refinement of the classifications. The model incorporates two primary elements: a deep neural network for feature extraction and a classifier. This stage encompasses three key processes: (1) extraction of core features, (2) computation of phenotypic labels, and (3) backpropagation to reduce prediction loss. HFpEF, heart failure with preserved ejection fraction; PCA, principal component analysis; PKUTH, Peking University Third Hospital; TOPCAT, Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist Trial; UMHS, University of Michigan Health System.

External validation was conducted using two independent cohorts. The first was the TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist Trial) [40] cohort. Details of its design and participant profile have been published previously [40]. The trial originally enrolled 3445 individuals older than 50 years who had symptomatic HFpEF with LVEF ≥45% and were randomized to spironolactone or placebo. Analyses were restricted to participants from the Americas to account for established regional differences in baseline traits, outcomes, and spironolactone responsiveness relative to Russia/Georgia [41]. In total, 1696 TOPCAT participants meeting the criterion of baseline LVEF ≥50% with available echocardiographic information were retained **(Figure 1a)**.

The second external validation cohort consisted of 343 HF patients evaluated at the University of Michigan Health System (UMHS) between June 1, 2009, and November 30, 2023. Inclusion and exclusion standards for this cohort have been described elsewhere [42]. Only cases with HFpEF confirmed by a heart failure specialist cardiologist were considered, yielding 128 eligible patients for analysis.

Phenotypic variables selection

A total of 134 variables were initially evaluated for inclusion in the clustering process within the training cohort. These spanned multiple categories drawn from electronic medical records (EMR), such as demographic details, anthropometric measures, laboratory results, historical medical and pharmacologic information, and echocardiographic indices. Extraction of narrative medical and medication data relied on natural language processing (NLP) methods, incorporating a validated large language model [43] and a specialized text mining procedure. To confirm extraction reliability, a cardiologist performed manual chart review on a random sample of 100 cases, demonstrating a sensitivity of 0.985 and specificity of 0.995.

When multiple measurements existed during the index hospitalization, the initial record was preferentially used. Variables missing in more than 25% of cases were generally excluded, except for selected echocardiographic parameters. Medications prescribed to fewer than 5% of the population were omitted. Patients with greater than

10% overall missingness were also excluded. Following these steps, phenomapping proceeded with 107 variables across 2147 HFpEF patients **(Figure 1a)**.

Machine learning-based clustering

Prior to analysis, missing values were handled via multivariate imputation by chained equations (MICE) [44], after which all features underwent Z-score standardization. Optimal cluster count was ascertained by jointly considering Silhouette coefficients [45], Gap statistics [46], Dunn index [47], and the Elbow method [48]. The analysis supported selection of three phenogroups.

Phenotype derivation was achieved through a two-stage unsupervised DeepCluster model engineered to approximate clinical diagnostic workflows **(Figure 1b)**. The model incorporated an initialization phase followed by iterative optimization, both directed toward greater precision and depth in representing HFpEF phenotypes. In the initialization phase, K-means clustering generated starting pseudo-labels exclusively from medical and medication history data. These served as preliminary assignments to facilitate model refinement [49]. The second phase involved iterative optimization using a DeepCluster architecture composed of a fully connected neural network feature extractor and a classifier. To enhance computational stability and mitigate high dimensionality, principal component analysis (PCA) was performed on input features; the resulting transformation matrix derived from the training set was then applied uniformly to all participants. Optimization proceeded through three repeated steps: (1) extraction of core features, (2) computation of phenotypic labels, and (3) backpropagation aimed at reducing predictive loss.

Internal and external validation

The reliability of the DeepCluster model was assessed internally in the training set by means of leave-one-out cross-validation. Concordance between phenogroup assignments generated through this leave-one-out procedure and those from the complete training dataset was measured using the adjusted Rand index (ARI) [50], Cohen's kappa coefficient [51], and c-statistics. External validation proceeded in two distinct cohorts: TOPCAT and UMHS. These analyses incorporated 1696 HFpEF patients featuring 60 variables from TOPCAT and 128 patients with 95 variables from UMHS **(Figure 1a)**. Variables showing over 80% missing values were not imputed; missing entries were substituted with −9999 to exclude their influence within the neural network while retaining the original data structure matching the training cohort. For variables with up to 80% missingness, imputation was performed via MICE. The identical DeepCluster model developed in the training phase was applied to assign phenogroups in each validation cohort, enabling subsequent comparisons of clinical attributes and prognostic patterns.

Outcome ascertainment

Follow-up for the training cohort extended until July 15, 2024, with information collected either through direct telephone contact by research staff or during outpatient evaluations. Key endpoints comprised all-cause mortality and readmission for HF, with confirmation provided by the responsible clinical team. Event timing was computed as the number of years elapsed from discharge after the index hospitalization until the initial occurrence of the outcome, the final follow-up contact, or study closure. In the TOPCAT cohort, endpoints encompassed the trial's primary composite outcome (aborted cardiac arrest, HF hospitalization, or CVD death), cardiovascular mortality, HF hospitalization, and overall mortality. All events underwent independent adjudication by the TOPCAT endpoint review committee following a predefined protocol [40]. For the UMHS cohort, the main endpoint was all-cause mortality tracked over a 5-year period, extracted from electronic health records and confirmed through blinded review by clinical personnel.

Statistics

Demographic information, physical measurements, laboratory markers, medical histories, medication records, and echocardiographic results were contrasted among the three phenogroups. Differences in continuous variables were tested with one-way Analysis of Variance (ANOVA) or Kruskal–Wallis tests, depending on data distribution, whereas categorical variables were evaluated using the Chi–Square test.

Survival differences across phenogroups were illustrated via Kaplan–Meier plots and examined with the log-rank test. Multivariable Cox proportional hazards regression was applied to determine the independent association of phenogroup membership with all-cause mortality and HF rehospitalization, incorporating adjustment for the

Meta-Analysis Global Group in Chronic (MAGGIC) [52] risk score. The proportional hazards assumption was verified by assessing the scaled Schoenfeld residuals for time-dependent trends.

Treatment responsiveness across phenogroups was investigated by fitting Cox proportional hazards models to calculate hazard ratios (HRs) and corresponding 95% confidence intervals (CIs). These models examined the link between post-diagnosis exposure to each medication—angiotensin-converting enzyme inhibitor/angiotensin receptor blockers [ACE-I/ARB], angiotensin receptor-neprilysin inhibitors [ARNIs], beta blockers, calcium channel blockers, spironolactone, and SGLT2 inhibitors—and the study outcomes, stratified according to phenogroup. To address immortal time bias, follow-up for medication recipients began at the prescription date, while non-recipients were followed from the discharge date. Confounders were chosen prospectively to control for confounding by indication and known prognostic factors. Adjustments included age, sex, BMI, smoking status, alcohol use, diuretic dosage, and drug-specific indications (For ACE-I/ARB: hypertension, coronary artery disease [CAD], myocardial infarction [MI], and chronic kidney disease [CKD]; For ARNI: hypertension, CAD, MI, CKD, New York Heart Association [NYHA] class, and prior ACE-I/ARB use; For beta blockers: hypertension, CAD, MI, and atrial fibrillation [AF]; For calcium channel blockers: hypertension, CAD, and MI; For Spironolactone: hypertension, CKD, and NYHA class; For SGLT2 inhibitors: hypertension, diabetes, and CKD). Interaction effects were tested via the likelihood ratio statistic from models containing the phenogroup-by-medication interaction term. Sensitivity analyses employed Fine–Gray subdistribution hazard models for HF rehospitalization, accounting for all-cause mortality as a competing risk.

In the external validation sets, three phenogroups were identified using the same procedure, and parallel comparisons of clinical features and outcomes were conducted. Within TOPCAT, the impact of spironolactone was additionally explored through stratified Cox models by phenogroup, with formal testing of modification via an interaction term between phenogroup and treatment assignment. Associations involving calcium channel blockers, ACE-I/ARB, and beta blockers with survival were also evaluated in this cohort. For the UMHS cohort, predicted hemodynamic and physiological metrics derived from a mechanistic computational model [42] were contrasted across phenogroups.

Statistical significance was set at $p < 0.05$. Analyses were executed in R (version 4.3.2, R Foundation for Statistical Computing). Development and execution of the machine learning clustering procedures were carried out in Python (version 3.8.16, Python Software Foundation) with the PyTorch framework (version 1.10.0, Facebook AI Research).

Ethics

This research followed the fundamental ethical tenets described in the Declaration of Helsinki. Analysis of clinical data originating from Peking University Third Hospital (PKUTH) received approval from the hospital's Institutional Review Board (IRB number: IRB00006761-M2022577). The external validation dataset from the University of Michigan Health System (UMHS) was authorized by the University of Michigan Institutional Review Board (IRB protocol number: HUM00243399). All PKUTH participants gave their informed consent prior to inclusion. Consent was not required for the UMHS cohort in view of its retrospective character and reliance on anonymized records, as permitted by relevant institutional regulations.

Owing to policies protecting patient confidentiality and data security mandated by the institution and regulatory authorities, the PKUTH dataset remains unavailable for open access. De-identified data can nevertheless be shared with other researchers for academic inquiry when a valid request is made, suitable data transfer agreements are signed, and necessary ethical permissions are granted.

Role of funding source

The sponsoring organizations exerted no influence over any aspect of the project, including study design, acquisition or processing of data, interpretation of findings, preparation of the manuscript, or the decision regarding submission for publication.

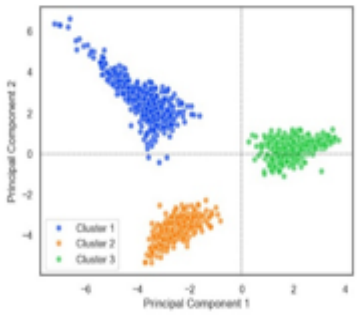
Results and Discussion

A cohort of 2147 patients diagnosed with HFpEF, together with 107 selected variables, formed the basis of the training dataset. At the time of hospital admission, the median age was 73.0 years (interquartile range [IQR] 65.0–81.0), and female patients represented 47.5% of the group. More than half the participants (51.4%) were

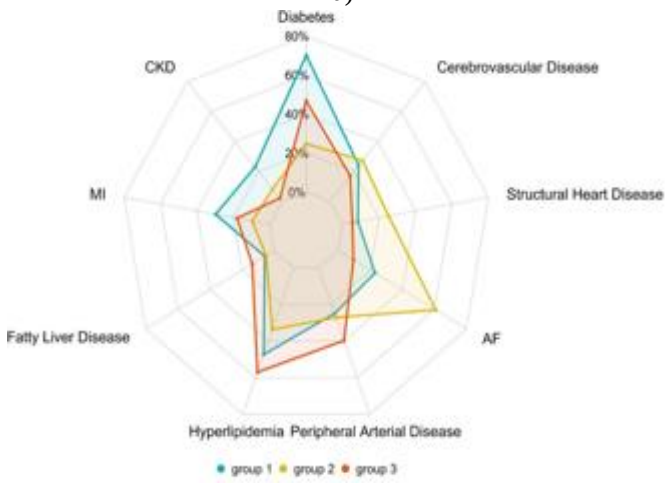
overweight or obese, with an average BMI of 25.4 ± 4.2 kg/m². Hypertension stood out as the most common associated condition (78.2%), followed closely by diabetes mellitus (49.5%) and hyperlipidemia (46.8%). Previous myocardial infarction and atrial fibrillation were recorded in 20.1% and 29.2% of cases, respectively.

Two-stage unsupervised DeepCluster model

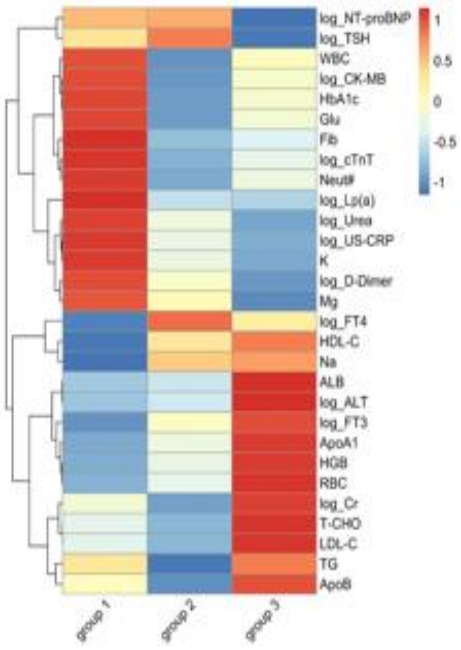
The derivation of phenogroups was accomplished via a two-stage unsupervised DeepCluster algorithm designed to emulate the stepwise reasoning employed in everyday clinical practice. The first stage drew upon patients' medical backgrounds and medication profiles to create an initial framework comparable to an early clinical evaluation. To address intercorrelations among features and support stable model training, principal component analysis (PCA) was conducted for dimensionality reduction **(Figure 2a)**. Retention of 64 principal components allowed explanation of 90% of the total variance observed in the data, with the first two components contributing 10.04%. A fully connected neural network was then introduced to detect complex nonlinear associations, generating 64-dimensional core representations that sharpened distinctions between the resulting phenotypes. Experiments removing specific components showed that restricting analysis to medical history and medication data alone impaired classification quality. Likewise, performance suffered when the PCA preprocessing step was eliminated or when the neural network classifier was substituted with plain K-means clustering that lacked iterative refinement.



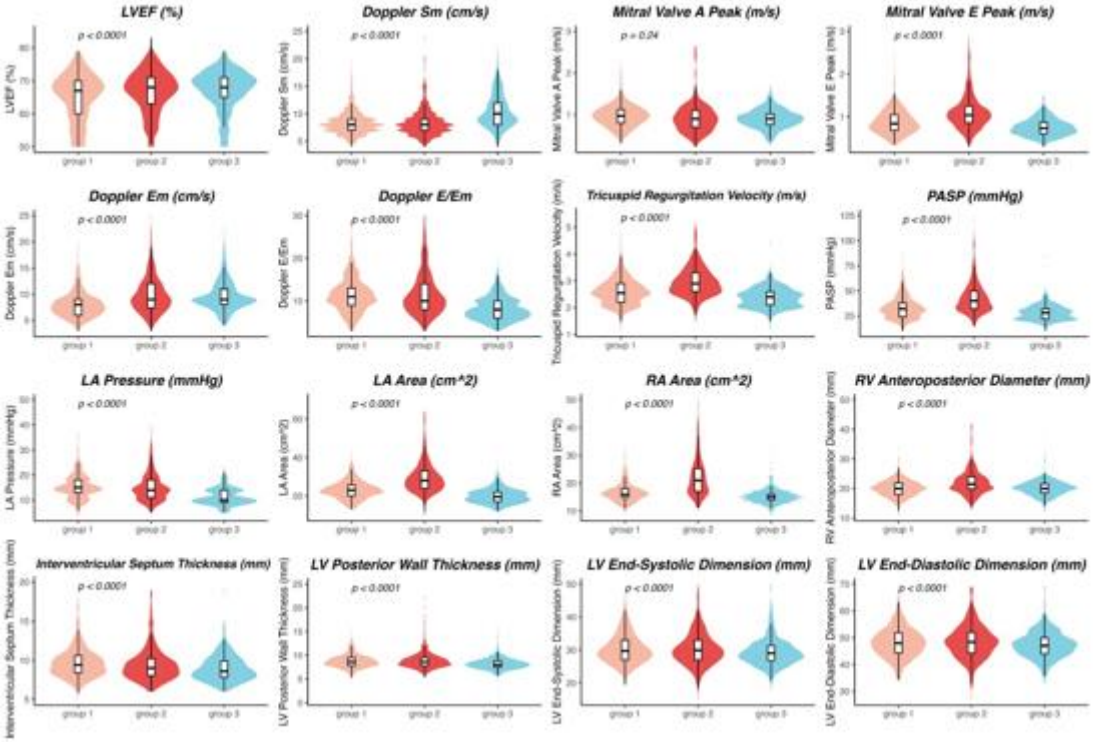
a)



c)



b)



d)

Figure 2. Clustering patterns observed in the training cohort: (a) PCA plot displaying cluster separation together with differences in (b) biomarkers, (c) comorbidities, and (d) echocardiographic variables among the clinical phenogroups generated by the DeepCluster algorithm.

(a) Principal component analysis scatterplot illustrating the spatial distribution of the identified clusters in the training set. (b) Heatmap depicting standardized and log-transformed biomarker profiles across phenogroups, where intensified red shading corresponds to elevated concentrations and deeper blue shading denotes reduced levels. (c) Radar plot summarizing the relative frequencies of various comorbidities within each phenogroup. (d) Violin plots showing the distribution of major echocardiographic parameters stratified according to phenogroup, including p-values obtained from one-way ANOVA testing. AF, atrial fibrillation; ALB, albumin; ALT, alanine aminotransferase; ANOVA, Analysis of Variance; ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; CKD, chronic kidney disease; CK-MB, creatine kinase-MB; COPD, chronic obstructive pulmonary disease; Cr, creatinine; cTnT, cardiac troponin T; Fib, Fibrinogen; FT3, free triiodothyronine; FT4, free thyroxine; Glu, glucose; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; HGB, hemoglobin; K, potassium; LA, left atrium; LDL-C, low-density lipoprotein cholesterol; Lp(a), lipoprotein(a); LV, left ventricle; LVEF, left ventricular ejection fraction; MI, myocardial infarction; Mg, magnesium; Na, sodium; Neut#, neutrophils count; NT-proBNP, N-terminal pro-B-type natriuretic peptide; PASP, pulmonary artery systolic pressure; PCA, principal component analysis; RA, right atrium; RBC, red blood cell count; RV, right ventricle; T-CHO, total cholesterol; TG, triglycerides; TSH, thyroid-stimulating hormone; Urea, blood urea nitrogen; US-CRP, ultra-sensitive C-reactive protein; WBC, white blood cell count.

Clinical characteristics of identified phenogroups

Implementation of the two-stage unsupervised DeepCluster model resulted in the recognition of three separate phenogroups. Key demographic and clinical attributes for these groups are detailed in **Table 1**. Individuals in phenogroup 1 were generally older (median age 76.0 [68.5–82.0] years), presented with elevated systolic blood pressure (SBP) and reduced diastolic blood pressure (DBP), and showed higher proportions of overweight (40.4%) and obese (12.4%) status. Phenogroup 2 contained the oldest participants (median age 79.0 [71.0–83.0] years), the largest fraction of female patients (53.3%), more severe heart failure symptoms according to NYHA classification, and a lower rate of obesity. Phenogroup 3 was distinguished by younger age (median age 67.0 [60.0–73.0] years), male predominance (61.3%), increased smoking (25.6%) and alcohol consumption (28.5%), and comparatively milder heart failure manifestations. Medication histories further differentiated the groups, with

greater prior exposure to ARBs, beta blockers, calcium channel blockers, diuretics, and antiplatelet agents in phenogroup 1, more frequent anticoagulant prescriptions in phenogroup 2, and higher SGLT2 inhibitor utilization in phenogroup 3.

Table 1. Baseline demographic and clinical characteristics across the phenogroups identified using the DeepCluster model.a,b

Variable	p-value	Phenogroup 3 (N = 724)	Phenogroup 2 (N = 608)	Phenogroup 1 (N = 815)	Overall (N = 2147)
Demographics					
Age, years (median [IQR])	<0.001	67.0 (60.0–73.0)	79.0 (71.0–83.0)	76.0 (68.5–82.0)	73.0 (65.0–81.0)
Female sex, n (%)	<0.001	280 (38.7)	324 (53.3)	415 (50.9)	1019 (47.5)
Smoking status, n (%)	<0.001				
Never		400 (55.2)	459 (75.5)	540 (66.2)	1399 (65.2)
Former		139 (19.2)	92 (15.1)	180 (22.1)	411 (19.1)
Current		185 (25.6)	57 (9.4)	95 (11.7)	337 (15.7)
Alcohol consumption, n (%)	<0.001				
Never		474 (65.4)	499 (82.1)	639 (78.4)	1612 (75.1)
Former		44 (6.1)	30 (4.9)	69 (8.5)	143 (6.6)
Current		206 (28.5)	79 (13.0)	107 (13.1)	392 (18.3)
Anthropometrics & vitals					
BMI, kg/m² (mean [SD])	0.280	25.54 (4.03)	25.19 (4.41)	25.55 (4.15)	25.44 (4.19)
Underweight, n (%)	0.070	17 (2.7)	31 (6.2)	19 (3.5)	67 (4.0)
Normal weight, n (%)		284 (44.4)	229 (45.9)	239 (43.7)	752 (44.6)
Overweight, n (%)		261 (40.8)	180 (36.1)	221 (40.4)	662 (39.3)
Obese, n (%)		77 (12.1)	59 (11.8)	68 (12.4)	204 (12.1)
Heart rate, bpm (mean [SD])	0.945	76.55 (13.71)	76.27 (16.63)	76.48 (15.57)	76.44 (15.28)
Systolic BP, mmHg (mean [SD])	<0.001	135.47 (19.54)	135.41 (20.32)	139.58 (22.02)	137.01 (20.82)
Diastolic BP, mmHg (mean [SD])	<0.001	78.05 (12.04)	74.47 (13.26)	71.09 (12.29)	74.39 (12.82)
NYHA class, n (%)	<0.001				
Class I		80 (13.4)	62 (12.8)	120 (21.0)	262 (15.8)
Class II		474 (79.4)	224 (46.3)	240 (42.0)	938 (56.7)
Class III		35 (5.9)	138 (28.5)	147 (25.6)	320 (19.5)
Class IV		8 (1.3)	60 (12.4)	65 (11.4)	133 (8.0)
Medication history, n (%)					
ACE inhibitor	0.140	33 (4.6)	24 (3.9)	50 (6.1)	107 (5.0)
ARB	<0.001	203 (28.0)	155 (25.5)	307 (37.7)	665 (31.0)
ARNI	0.084	33 (4.6)	41 (6.7)	35 (4.3)	109 (5.1)
Beta-blocker	<0.001	127 (17.5)	138 (22.7)	266 (32.6)	531 (24.7)
Calcium channel blocker	<0.001	227 (31.4)	210 (34.5)	444 (54.5)	881 (41.0)
Diuretics	<0.001	57 (7.9)	69 (11.3)	118 (14.5)	244 (11.4)
SGLT2 inhibitor	<0.001	70 (9.7)	8 (1.3)	36 (4.4)	114 (5.3)
Anticoagulants	<0.001	4 (0.6)	124 (20.4)	27 (3.3)	155 (7.2)
Antiplatelet agents	<0.001	33 (4.6)	51 (8.4)	181 (22.2)	265 (12.3)

ACE-I, angiotensin-converting enzyme inhibitor; AF, atrial fibrillation; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor-neprilysin inhibitor; BMI, body mass index; DBP, diastolic blood pressure; HF, heart failure; HR, heart rate; IQR, interquartile range; NYHA, New York Heart Association; SBP, systolic blood pressure; SD, standard deviation; SGLT2, sodium–glucose cotransporter 2.

a Values are reported as mean (SD) or median (IQR) for continuous variables and as N (%) for categorical variables.
b p-values were determined using one-way ANOVA or Kruskal–Wallis tests for continuous data and Chi-square tests for categorical data.

Biomarker differences across phenogroups are shown in **Figure 2b**. Phenogroup 1 displayed markedly raised levels of indicators reflecting renal injury (such as blood urea nitrogen), hyperglycemia, myocardial damage (including cardiac troponin T [cTnT] and creatine kinase (CK)-MB), inflammatory activity (such as C-reactive protein [CRP], white blood cells [WBC], and neutrophils), and prothrombotic state (including fibrinogen and D-dimer). Conversely, this cluster had reduced concentrations of thyroid hormones (free T3 [FT3] and free T4 [FT4]) and anemia markers (hemoglobin and red blood cells [RBC]). Phenogroup 2 recorded the peak NT-proBNP values. Phenogroup 3 was distinguished by higher lipid-related parameters (such as total cholesterol, triglycerides, low-density lipoprotein cholesterol [LDL-C], and Apolipoprotein B [Apo B]) along with altered liver enzymes and proteins (such as albumin and alanine transaminase [ALT]).

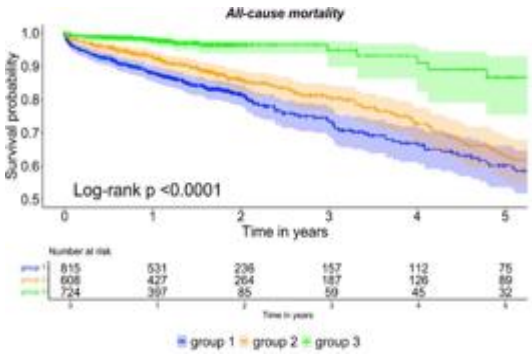
Clear differences in comorbidity patterns emerged among the phenogroups **(Figure 2c)**. Phenogroup 1 showed the greatest burden of metabolic conditions, notably diabetes mellitus (70.2%), CKD (22.9%), and prior MI (29.7%). Phenogroup 2 had the largest representation of AF (60.9%) and structural heart disease (25.2%). Phenogroup 3 was more often linked with hyperlipidemia (57.0%), fatty liver disease (13.7%), and peripheral arterial disease (39.8%).

Echocardiographic parameters across the phenogroups are shown in **Figure 2d**. Phenogroup 1 presented a unique profile marked by left ventricular (LV) hypertrophy, including increased wall thickness and expanded chamber volumes, corresponding to the largest LV end-systolic and end-diastolic dimensions. This group further demonstrated impaired diastolic (reduced Doppler Em velocity and elevated E/e' ratio) and systolic (decreased Doppler Sm and LVEF) performance. Phenogroup 2 featured the most enlarged right ventricle (RV) and atrial chambers, together with the highest peak E wave velocity, tricuspid regurgitation velocity, pulmonary artery systolic pressure, and estimated LV filling pressure (E/e'). In comparison, phenogroup 3 maintained largely preserved cardiac structure and contractile function relative to the remaining clusters.

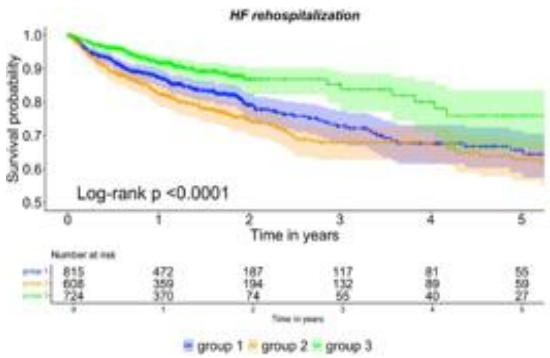
Echocardiographic parameters across the phenogroups are shown in **Figure 2d**. Phenogroup 1 was notable for LV hypertrophy with thickened walls, increased ventricular volumes, and evidence of both diastolic (lower Doppler Em and elevated E/e') and systolic (lower Doppler Sm and reduced LVEF) impairment. Phenogroup 2 exhibited the greatest right ventricular (RV) and atrial enlargement as well as the highest values of peak E wave, tricuspid regurgitation velocities, pulmonary artery systolic pressure, and LV filling pressure (E/e'). Phenogroup 3, however, displayed comparatively normal cardiac anatomy and performance.

Association of phenogroups with HF prognosis

Over a median observation period of 1.28 years, 330 patients (15.4%) died and 353 (16.4%) required HF rehospitalization. Substantial variation in long-term outcomes was apparent, with statistically significant differences in 5-year all-cause mortality and HF rehospitalization rates across the three phenogroups **(Figures 3a and 3b)**, (all log-rank p-values <0.0001). In fully adjusted Cox proportional hazards models, phenogroup 1 (compared with phenogroup 3) was linked to the greatest hazard of death from any cause (hazard ratio [HR] = 3.34, 95% confidence interval [CI] 2.06–5.50). Phenogroup 2, meanwhile, carried the highest risk of HF rehospitalization (HR = 1.85, 95% CI 1.33–2.56).



a)



b)

	All-cause mortality			HF rehospitalization		
	Events/Total	HR (95% CI)	P-value	Events/Total	HR (95% CI)	P-value
Phenogroup 1	181/815	3.34 (2.06-5.50)	<0.0001	144/815	1.46 (1.05-2.04)	0.027
Phenogroup 2	126/608	2.67 (1.63-4.37)	<0.0001	145/608	1.85 (1.33-2.56)	<0.0001
Phenogroup 3	23/724	Ref.		64/724	Ref.	

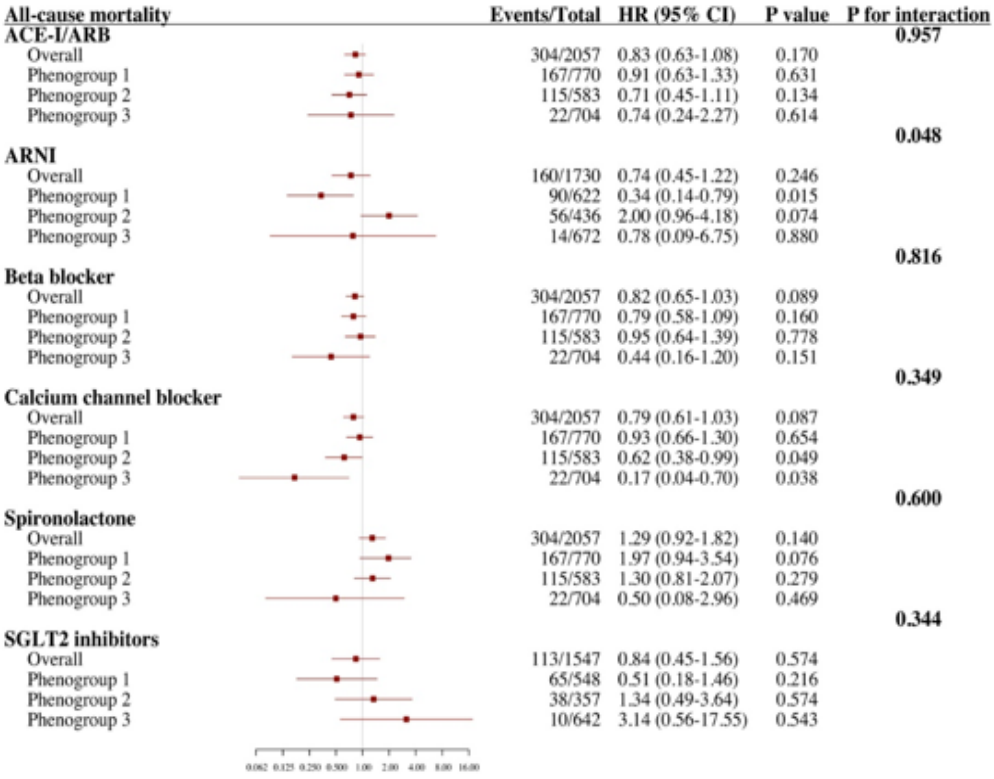
c)

Figure 3. Association between phenogroups and clinical outcomes during follow-up.

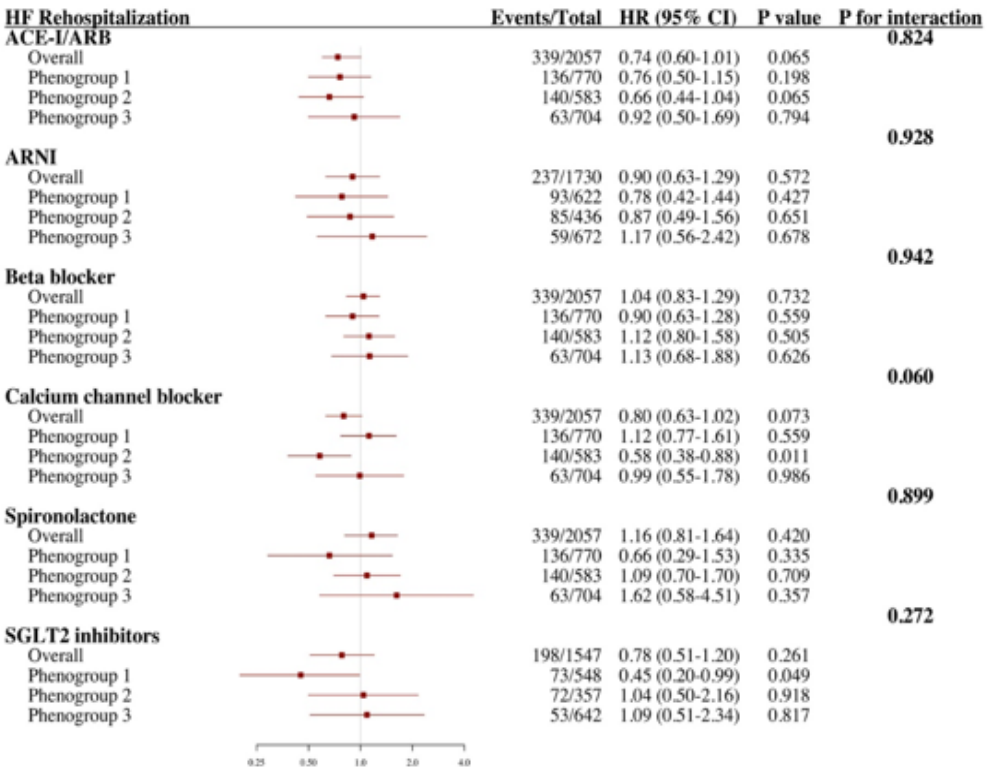
(a) Kaplan–Meier survival curves depicting all-cause mortality. (b) Kaplan–Meier curves for heart failure (HF) rehospitalization. (c) Forest plot summarizing multivariable-adjusted hazard ratios for adverse events according to phenogroup. Kaplan–Meier analyses compared event-free survival across phenotypic clusters using the log-rank test. Multivariable Cox regression models provided hazard ratios (HRs) and 95% confidence intervals (CIs) for all-cause mortality and HF rehospitalization, using phenogroup 3 as the reference. All models were adjusted for the MAGGIC risk score. The assumption of proportional hazards was satisfied. CI, confidence interval; HF, heart failure; HR, hazard ratio; MAGGIC, Meta-Analysis Global Group in Chronic.

Association of post-diagnostic medication use with prognosis in HFpEF phenogroups

Across the entire study population, standard heart failure treatments did not confer clear overall survival advantages. However, when examined within individual phenogroups, certain therapies appeared beneficial. In phenogroup 1, administration of SGLT2 inhibitors was related to a 55% decrease in HF rehospitalization risk (HR = 0.45, 95% CI 0.20–0.99, approaching statistical significance), and ARNI therapy was associated with a 66% reduction in all-cause mortality (HR = 0.34, 95% CI 0.14–0.79). The impact of ARNI on mortality risk varied significantly by phenogroup (p for interaction = 0.048). Among patients in phenogroup 2, calcium channel blocker use correlated with a 38% lower all-cause mortality risk (HR = 0.62, 95% CI 0.38–0.99, borderline significance) and a 42% reduction in HF rehospitalization (HR = 0.58, 95% CI 0.38–0.88). In phenogroup 3, calcium channel blockers were linked to a pronounced decrease in all-cause mortality (HR = 0.17, 95% CI 0.04–0.70), **(Figure 4)**.



a)



b)

Figure 4. Association of post-diagnostic medications with clinical prognosis, stratified according to phenogroup membership.

Hazard ratios (HRs) with 95% confidence intervals (CIs) illustrate the relationship of each medication class with (A) all-cause mortality and (B) HF rehospitalization within each phenogroup. The multivariable Cox model evaluating ACE-I/ARB was adjusted for age, sex, BMI, smoking status, alcohol drinking status, diuretic dose, hypertension, CAD, MI, and CKD. For ARNI analyses, the cohort was limited to patients admitted after January 1 2019 (when ARNI became available at PKUTH), comprising 622, 436, and 672 individuals in phenogroups 1, 2, and 3, respectively. The ARNI model was further adjusted for age, sex, BMI, smoking status, alcohol drinking status, diuretic dose, hypertension, CAD, MI, CKD, NYHA class, and ACE-I/ARB use. The beta blocker model was adjusted for age, sex, BMI, smoking status, alcohol drinking status, diuretic dose, hypertension, CAD, MI, and AF. The calcium channel blocker model was adjusted for age, sex, BMI, smoking status, alcohol drinking status, diuretic dose, hypertension, CAD, and MI. The spironolactone model was adjusted for age, sex, BMI, smoking status, alcohol drinking status, diuretic dose, hypertension, CKD, and NYHA class. For SGLT2 inhibitors, the analysis was restricted to admissions after January 1 2021, including 548, 357, and 642 patients in groups 1, 2, and 3, respectively. This model was adjusted for age, sex, BMI, smoking status, alcohol drinking status, diuretic dose, hypertension, diabetes, and CKD. Interaction p-values were calculated via likelihood ratio testing of the phenogroup-by-medication product term. Statistical significance was defined as $p < 0.05$. ACE-I, angiotensin-converting enzyme inhibitor; AF, atrial fibrillation; ARB, angiotensin II receptor blocker; ARNI, angiotensin receptor-neprilysin inhibitor; BMI, body mass index; CAD, coronary artery disease; CI, confidence interval; CKD, chronic kidney disease; HF, heart failure; HR, hazard ratio; MI, myocardial infarction; NYHA, New York Heart Association; PKUTH, Peking University Third Hospital; SGLT2, sodium-glucose co-transporter 2.

Internal and external validation

The robustness of the two-stage DeepCluster model was assessed through internal validation employing a leave-one-out methodology. This process yielded excellent stability, evidenced by 96.0% consistency, an adjusted Rand index (ARI) of 0.87, Cohen's kappa coefficient of 0.93, and a c-statistic of 0.965. Such metrics underscore the reliability of the proposed DeepCluster framework.

External validation was conducted in the independent TOPCAT cohort, where three separate phenogroups were successfully reproduced. Visualization of the clustering results revealed strong concordance between the TOPCAT cohort and the original training set, with corresponding cluster centroids located in close proximity **(Figure 5a)**. Mirroring the patterns seen in the derivation cohort, phenogroup 1 within TOPCAT was characterized by the highest body mass index (BMI) and a substantial burden of cardiometabolic comorbidities, including diabetes (82.8%) and hyperlipidemia (84.7%). This cluster also exhibited the highest frequency of prior revascularization (51.0%), enlarged left ventricular (LV) volumes, and moderate diastolic dysfunction. Phenogroup 2 comprised the oldest participants and featured the greatest prevalence of atrial fibrillation (AF) (90.0%) and significant valvular disease (26.7%), along with the largest left atrial (LA) volume, the highest E/A ratio, elevated tricuspid regurgitation velocity, and the most pronounced diastolic impairment. In contrast, phenogroup 3 included relatively younger individuals who presented with comparatively better diastolic function. With regard to clinical prognosis, phenogroup 1 experienced the least favorable outcomes for the primary composite endpoint, showing a markedly increased hazard compared with phenogroup 3 (HR = 2.82, 95% CI 2.27–3.51) **(Figure 5b)**. Evaluation of therapeutic responses indicated that spironolactone treatment was associated with an 18% risk reduction in the primary composite outcome (HR = 0.82, 95% CI 0.69–0.98), a 31% decrease in cardiovascular mortality (HR = 0.69, 95% CI 0.53–0.91), and a 19% reduction in HF hospitalization (HR = 0.81, 95% CI 0.66–0.99) across the full TOPCAT population. While formal interaction testing did not reach statistical significance (all p for interaction > 0.05), the protective effect of spironolactone appeared strongest in phenogroup 3 **(Figure 5c)**. In addition, calcium channel blocker therapy correlated with a decreased risk of all-cause mortality, with more prominent associations observed in phenogroups 2 and 3 **(Figure 5d)**. Notably, beta-blocker use in the overall TOPCAT population was linked to an increased risk of the primary composite endpoint (HR = 1.29, 95% CI 1.02–1.64) and HF hospitalization (HR = 1.58, 95% CI 1.18–2.12).

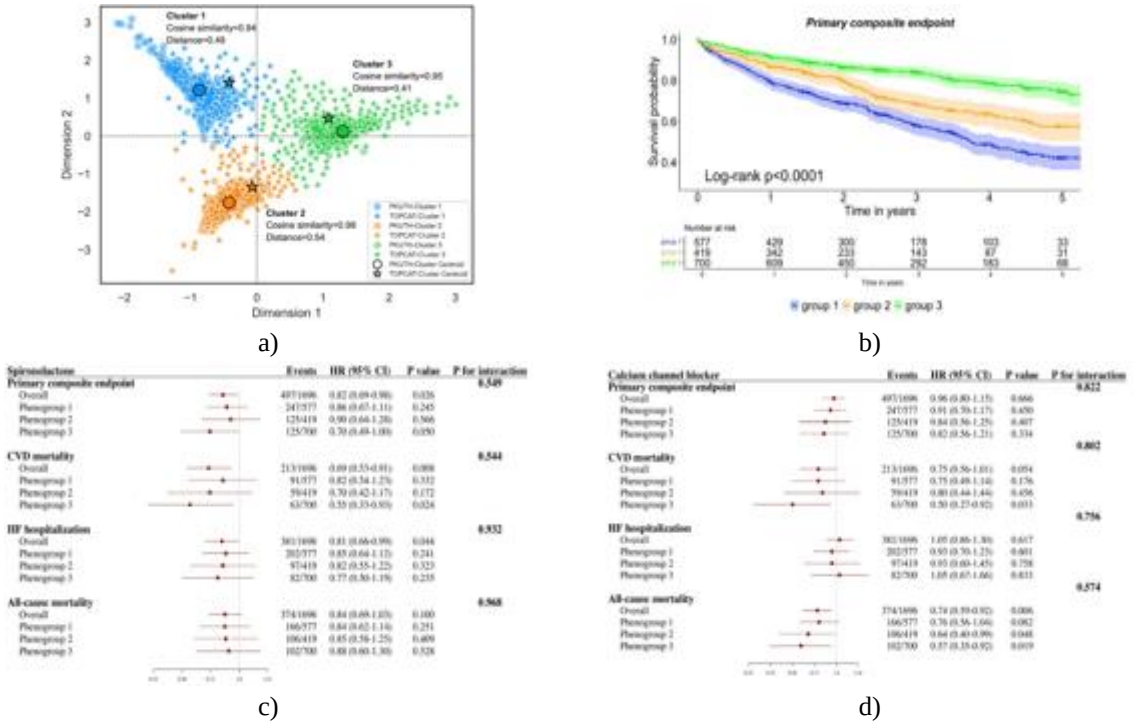


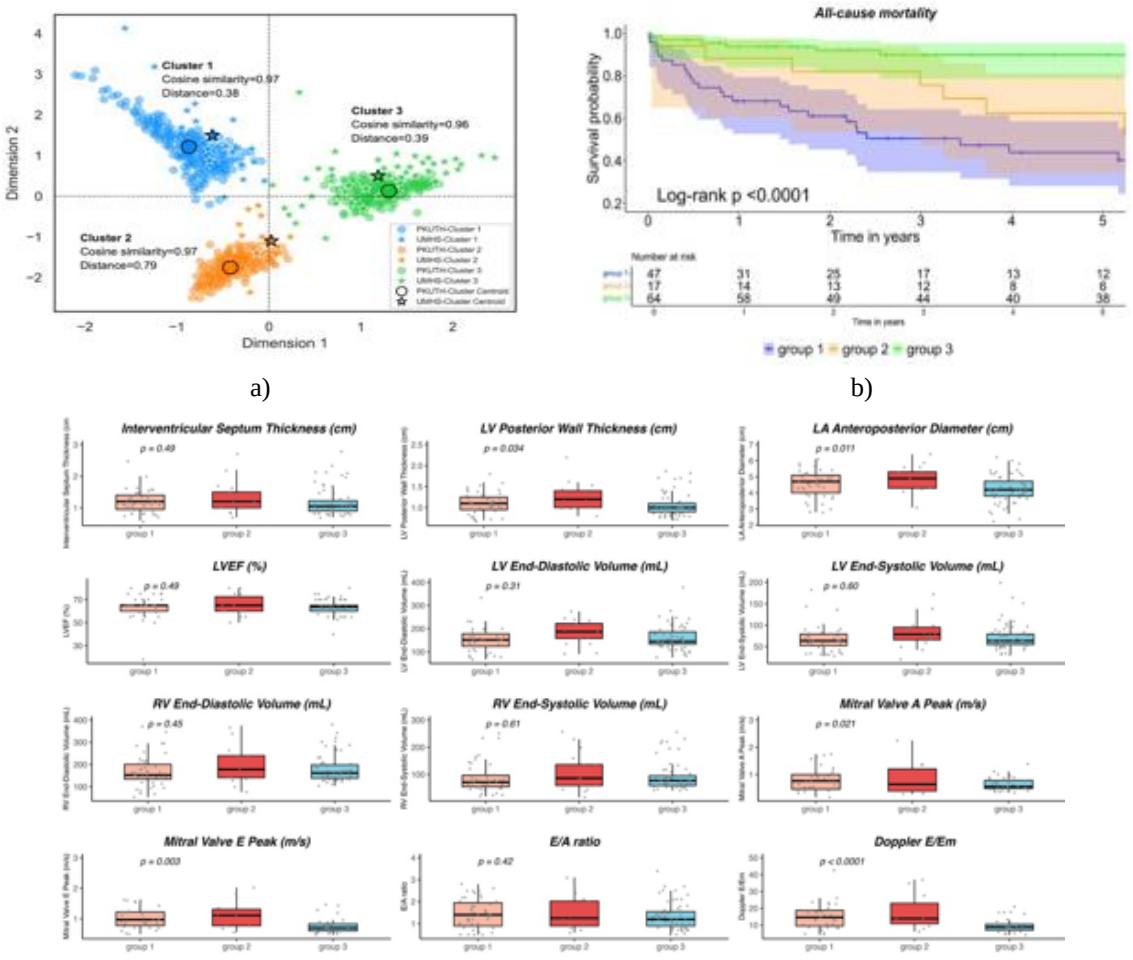
Figure 5. External validation results from the TOPCAT cohort: (a) cluster visualization, (b) Kaplan–Meier survival curve for the primary composite endpoint, (c) impact of spironolactone treatment across the three phenogroups, and (d) association of calcium channel blocker utilization with clinical prognosis according to phenogroup.

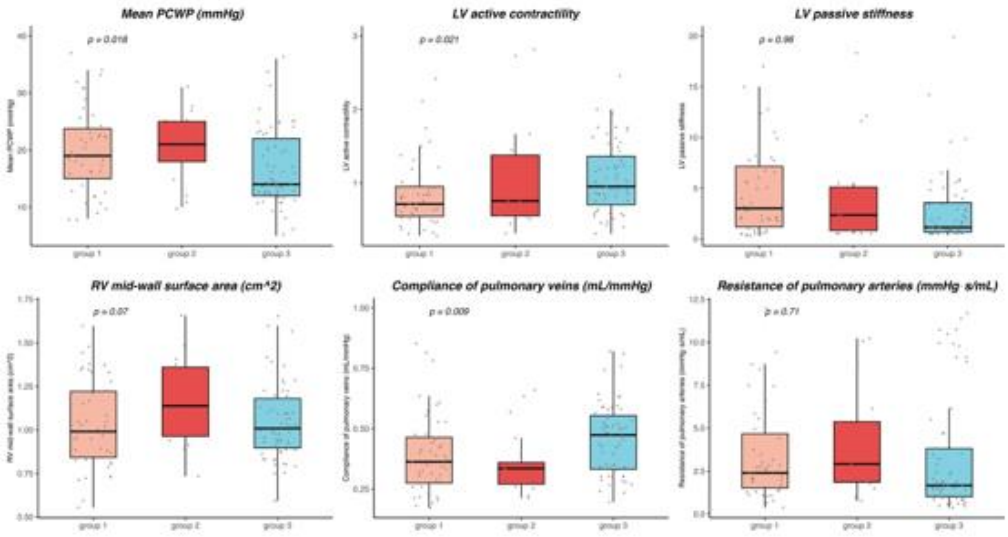
(a) Clustering distribution observed in the TOPCAT cohort in comparison with the training cohort. Cosine similarity between centroids belonging to the same cluster was determined by computing the dot product of their respective centroid vectors and dividing by the product of their vector norms. Euclidean distance between centroids of each cluster was also calculated to assess the geometric separation of the vectors in Euclidean space. (b) Kaplan–Meier curves depicting the primary composite outcome, which encompassed cardiovascular mortality,

aborted cardiac arrest, or admission for heart failure management. (c) Hazard ratios (HRs) along with their 95% confidence intervals (CIs) for the effect of spironolactone on the composite outcome, cardiovascular death, heart failure hospitalization, and all-cause mortality, presented separately for each phenogroup. (d) Hazard ratios (HRs) with 95% confidence intervals (CIs) for the association between calcium channel blocker therapy and the composite outcome, cardiovascular death, heart failure hospitalization, and all-cause mortality, stratified by phenogroup. Multivariable Cox proportional hazards models were adjusted for age, sex, race, BMI, hypertension, CAD, MI, and spironolactone randomization status. BMI, body mass index; CAD, coronary artery disease; CI, confidence interval; HF, heart failure; HR, hazard ratio; MI, myocardial infarction; TOPCAT, Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist.

Similarly, three distinct phenogroups emerged in the UMHS cohort **(Figure 6a)**. Phenogroup 1 carried the heaviest load of metabolic comorbidities — including diabetes (44.7%), hyperlipidemia (72.3%), and CKD (63.8%) — and demonstrated the most adverse clinical trajectory (HR = 3.81, 95% CI 1.69–8.59), **(Figure 6b)**. Phenogroup 2 comprised the eldest patients and exhibited the greatest frequency of AF (58.8%) and valvular heart disease (58.8%), with a higher likelihood of more severe aortic and valvular regurgitation. Phenogroup 3 tended to include younger individuals who presented with less severe symptomatology.

Furthermore, notable differences in echocardiographic, hemodynamic, and physiologic parameters were observed across the phenogroups. Phenogroup 1 showed the most advanced degree of systolic impairment, evidenced by decreased LVEF and reduced LV contractility. Phenogroup 2 was characterized by more severe diastolic dysfunction, featuring the highest E/e' ratio, larger ventricular volumes, increased LA anteroposterior dimensions, and elevated pulmonary capillary wedge pressure (PCWP). This group additionally displayed the lowest LV passive stiffness and pulmonary vein compliance, coupled with the highest pulmonary artery resistance **(Figures 6c and 6d)**. The leading ten variables ranked by SHAP values within each phenogroup in the external validation cohorts corresponded closely with the key clinical attributes of the respective groups.





d)

Figure 6. External validation results in the UMHS cohort: (a) cluster distribution plot, (b) Kaplan–Meier survival analysis for all-cause mortality, (c) violin plots depicting echocardiographic variables across phenogroups, and (d) violin plots showing key hemodynamic and predicted physiologic measures by phenogroup.

(a) Clustering visualization obtained from the UMHS cohort in parallel with the training cohort. Cosine similarity for centroids belonging to identical clusters was derived by computing the dot product of their vectors and dividing this value by the product of the individual vector norms. Euclidean distance between the centroids of each cluster was determined to evaluate the actual geometric separation of these vectors in Euclidean space. (b) Kaplan–Meier curve for all-cause mortality. (c) Violin plots illustrating echocardiographic parameters stratified according to phenogroup, with accompanying p-values calculated via one-way ANOVA. (d) Violin plots presenting selected hemodynamic and predicted physiologic parameters stratified by phenogroup, with p-values calculated via one-way ANOVA. ANOVA, Analysis of Variance; LA, left atrium; LV, left ventricle; LVEF: left ventricular ejection fraction; PCWP, pulmonary capillary wedge pressure; RA, right atrium; RV, right ventricle; UMHS, University of Michigan Health System.

In the current investigation, a two-stage DeepCluster algorithm was applied to a well-characterized cohort of patients with heart failure with preserved ejection fraction (HFpEF), resulting in the identification of three distinct phenogroups defined by clinical variables. External validation of this model was performed successfully in two separate cohorts (TOPCAT and UMHS), which supported the reliability and generalizability of the phenotyping strategy. In addition to pronounced differences in baseline clinical profiles and long-term outcomes, the analysis revealed notable variation in therapeutic responses among the phenogroups. Phenogroup 1 in particular showed the most substantial improvements with SGLT2 inhibitors and ARNI treatment, whereas phenogroups 2 and 3 demonstrated comparatively better responses to calcium channel blockers.

The analysis delineated three phenogroups that differed markedly in their clinical features and prognostic trajectories. Phenogroup 1 was distinguished by the greatest metabolic burden, prominent left ventricular hypertrophy, combined systolic and diastolic impairment, and the least favorable clinical outcomes. Phenogroup 2 included the oldest participants and was characterized by the highest rates of atrial fibrillation together with prominent atrial and right ventricular remodeling as well as advanced diastolic dysfunction. In contrast, phenogroup 3 consisted of younger patients who maintained relatively normal cardiac structure and function. These observations are generally consistent with earlier phenotyping research in HFpEF populations [15, 18, 23, 31, 33]. For instance, Kyodo and colleagues utilized unsupervised machine learning techniques in a Japanese coh ort of 365 HFpEF patients and described three analogous clusters: one dominated by atherosclerotic disease and chronic kidney disease, a second featuring older women with atrial fibrillation, and a third involving younger patients with left ventricular hypertrophy [23]. Likewise, Shah *et al*. stratified 397 symptomatic HFpEF individuals into three subtypes, including a group with extensive metabolic comorbidities, one with frequent atrial

fibrillation and right ventricular dysfunction, and a younger cluster displaying the lowest BNP concentrations [33].

Comparable phenotypes have been documented across multiple prior studies [15, 18, 20, 21, 23, 30, 33]. Drawing on the I-PRESERVE trial, Kao *et al*. described a subgroup with substantial metabolic and renal comorbidities (including diabetes and chronic kidney disease) that carried the poorest prognosis, closely resembling our phenogroup 1 [21]. Cohen and co-workers identified an older phenotype marked by arterial stiffness, reduced left ventricular size, diastolic dysfunction, and a high prevalence of atrial fibrillation, which parallels phenogroup 2 in the present work [15]. This subtype also featured elevated pulmonary artery systolic pressure, right ventricular impairment, and atrial myopathy—features recognized as typical of atrial fibrillation-related HFpEF [33, 35]—and these characteristics were replicated in both the derivation and validation datasets. The younger, lower NT-proBNP phenotype observed here has likewise been reported in several earlier investigations [15, 20, 30, 33].

This study represents the first comprehensive examination of differential treatment effects stratified by HFpEF phenogroups. The results demonstrate that individuals in phenogroup 1, who exhibit high levels of obesity, diabetes, and chronic kidney disease, experience the largest reduction in heart failure rehospitalization risk with SGLT2 inhibitor therapy. These observations are in line with major randomized trials such as EMPEROR-Preserved and DELIVER, which established the ability of SGLT2 inhibitors to decrease the combined endpoint of heart failure hospitalization or cardiovascular death while supporting renal preservation [8, 9]. Similar risk reductions of 29–33% for this composite outcome were noted in dedicated chronic kidney disease trials, irrespective of diabetes status [53–55]. Overall, the present findings add to the body of evidence indicating that HFpEF patients affected by cardiovascular-kidney-metabolic syndrome stand to gain the most from SGLT2 inhibitors.

The data additionally indicate potential advantages of ARNI therapy for phenogroup 1 patients. This is supported by a meta-analysis reporting a 21% relative risk reduction in heart failure hospitalization or cardiovascular death with ARNI among HFpEF individuals with chronic kidney disease (eGFR < 60 mL/min/1.73 m²) [56]. Earlier clinical and observational data have underscored the favorable effects of ARNI on both renal progression and metabolic parameters, including delayed deterioration of kidney function and better glycemic regulation [57–62]. Furthermore, ARNI benefits appear enhanced among those with reduced left ventricular ejection fraction [63], a profile that matches the lower Sm and LVEF values observed in phenogroup 1. Such results strengthen the rationale for using ARNI in HFpEF patients burdened by renal and metabolic conditions.

Finally, the analysis suggests that calcium channel blockers may offer particular value for HFpEF patients with atrial fibrillation, corresponding to phenogroup 2. While earlier literature on calcium channel blockers in HFpEF has produced mixed conclusions [64–66], validation analyses within the TOPCAT Americas cohort indicated an overall 26% reduction in all-cause mortality risk associated with their use. This survival advantage was especially evident in TOPCAT phenogroup 2 participants with atrial fibrillation, mirroring patterns seen in the primary cohort. These observations accord with previous reports linking calcium channel blocker therapy to lower rates of mortality and heart failure hospitalization in atrial fibrillation populations [67, 68]. Contemporary guidelines endorse non-dihydropyridine agents (e.g., diltiazem or verapamil) for rate control in HFpEF patients with atrial fibrillation [4, 69]. In the present cohorts, prescription rates for non-dihydropyridine calcium channel blockers were highest in phenogroup 2 (5.3% in phenogroup 1, 11.2% in phenogroup 2, and 9.2% in phenogroup 3). This pattern implies that such medications could alleviate atrial fibrillation symptoms and contribute to improved prognosis in this subgroup, although the modest number of users calls for cautious interpretation and additional confirmatory research. The study also noted possible benefits of calcium channel blockers among patients with milder symptoms and dyslipidemia, potentially linked to the coronary vasodilatory properties of dihydropyridine agents that may relieve ischemic stress; however, these preliminary signals require validation in future prospective studies.

By evaluating treatment responses according to phenogroup membership, this study offers real-world evidence that underscores the value of SGLT2 inhibitors and ARNI for HFpEF patients burdened by metabolic comorbidities, as well as the utility of calcium channel blockers for those with atrial fibrillation. Larger-scale investigations with extended follow-up periods will be required to corroborate these phenogroup-specific therapeutic benefits.

The present study possesses several notable strengths. First, it systematically examines the responses to commonly used heart failure medications across phenogroups in HFpEF, thereby generating real-world data that can inform personalized management strategies for this heterogeneous population. Second, the phenomapping model

integrates a broad array of variables, encompassing demographic details, anthropometric measures, medical and pharmacological histories, and comprehensive echocardiographic assessments. This multifaceted inclusion improves the accuracy of subgroup delineation relative to approaches relying on narrower sets of clinical parameters. Third, the adoption of a two-stage DeepCluster framework incorporating unsupervised iterative training enables dynamic discovery of underlying data structures. This approach outperforms conventional machine learning techniques by simultaneously refining feature representation and cluster assignment. Moreover, the two-stage design emulates the stepwise nature of clinical diagnosis, thereby enhancing the stability and clinical interpretability of the phenotyping algorithm. Finally, external validation was conducted using two independent HFpEF cohorts that differed in patient profiles. This validation not only affirms the robustness of the DeepCluster methodology but also yields valuable hemodynamic and physiologic insights into the structural and functional variations among HFpEF subgroups, deepening the overall understanding of disease heterogeneity.

Several limitations of the study should be acknowledged. First, the training cohort identified HFpEF on the basis of an LVEF ≥50% and the presence of heart failure symptoms, which may have led to some degree of misclassification. Application of the 2023 ESC diagnostic criteria confirmed that 89.0% of these patients fulfilled the requirements. Nevertheless, the study deliberately sought to encompass a wider HFpEF spectrum, using the phenotyping strategy to detect a subgroup potentially representing atypical or HFrEF-like presentations. Second, given the observational design, residual confounding cannot be entirely excluded despite statistical adjustment when assessing relationships between phenogroups, clinical outcomes, and treatment effects. Third, substantial missingness (>80%) of certain echocardiographic and biomarker data in the external validation cohorts restricted detailed subgroup profiling in those datasets. Nonetheless, the comparable distribution of phenogroups and clinical features across cohorts supports the generalizability of the main findings. Fourth, participants in the training set were hospitalized patients recruited from a tertiary referral center, whose characteristics may not fully represent those of community-dwelling HFpEF individuals. Although validation was achieved in two ethnically diverse external cohorts, the possibility of selection bias persists. Subsequent research should therefore evaluate the findings in outpatient populations. Lastly, some confidence intervals were relatively wide, likely due to modest sample sizes within certain subgroups. A few observed associations reached only borderline statistical significance and lost significance after correction for multiple comparisons. Consequently, these results should be viewed with appropriate caution, and larger studies are needed to confirm them.

Conclusion

In summary, we developed and externally validated a machine learning-driven clustering approach that identified three distinct phenogroups of HFpEF patients exhibiting clear differences in cardiac performance and comorbidity profiles. Phenogroup 1 is defined by a high burden of metabolic comorbidities, left ventricular hypertrophy, combined systolic and diastolic dysfunction, and the least favorable prognosis. Phenogroup 2 is composed primarily of older patients with atrial fibrillation, atrial and right ventricular structural changes, and pronounced diastolic dysfunction. Phenogroup 3 comprises younger individuals with lifestyle-related risk factors who display preserved cardiac morphology and function.

The results further indicate that SGLT2 inhibitors and ARNI may be particularly advantageous for HFpEF patients presenting with metabolic comorbidities and biventricular dysfunction, while calcium channel blockers appear more suitable for those with atrial fibrillation. Overall, these findings supply real-world evidence regarding differential treatment responses across HFpEF phenogroups and advocate for a phenotype-guided strategy to tailor therapy. Dedicated clinical trials will be essential to assess the efficacy and safety of such individualized therapeutic approaches within specific HFpEF subgroups.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Campbell P, Rutten FH, Lee MM, Hawkins NM, Petrie MC. Heart failure with preserved ejection fraction: Everything the clinician needs to know. Lancet. 2024;403(10431):1083-92. doi:10.1016/S0140-6736(23)02756-3
2. Redfield MM, Borlaug BA. Heart failure with preserved ejection fraction: A review. JAMA. 2023;329(10):827-38.
3. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. N Engl J Med. 2006;355(3):251-9. doi:10.1056/NEJMoa052256
4. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;42(36):3599-726. doi:10.1093/eurheartj/ehab368
5. Hamo CE, DeJong C, Hartshorne-Evans N, Lund LH, Shah SJ, Solomon S, et al. Heart failure with preserved ejection fraction. Nat Rev Dis Primers. 2024;10(1):55.
6. Borlaug BA, Sharma K, Shah SJ, Ho JE. Heart failure with preserved ejection fraction. J Am Coll Cardiol. 2023;81(18):1810-34.
7. Cai A, Qiu W, Zhou Y, Feng Y, Chen J, Xia S, et al. Clinical characteristics and 1-year outcomes in hospitalized patients with heart failure with preserved ejection fraction: Results from the China Cardiovascular Association Database-Heart Failure Center Registry. Eur J Heart Fail. 2022;24(11):2048-62. doi:10.1002/ejhf.2654
8. Anker SD, Butler J, Filippatos G, Ferreira JP, Bocchi E, Böhm M, et al. Empagliflozin in heart failure with a preserved ejection fraction. N Engl J Med. 2021;385(16):1451-61. doi:10.1056/NEJMoa2107038
9. Solomon SD, McMurray JJV, Claggett B, de Boer RA, DeMets D, Hernandez AF, et al. Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction. N Engl J Med. 2022;387(12):1089-98.
10. Docherty KF, Henderson AD, Jhund PS, Claggett BL, Desai AS, Mueller K, et al. Efficacy and safety of finerenone across the ejection fraction spectrum in heart failure with mildly reduced or preserved ejection fraction: A prespecified analysis of the FINEARTS-HF trial. Circulation. 2025;151(1):45-58.
11. Verbrugge FH, Omote K, Reddy YNV, Sorimachi H, Obokata M, Borlaug BA. Heart failure with preserved ejection fraction in patients with normal natriuretic peptide levels is associated with increased morbidity and mortality. Eur Heart J. 2022;43(20):1941-51. doi:10.1093/eurheartj/ehab911
12. Arévalo-Lorido JC, Carretero-Gómez J, Aramburu-Bodas O, Grau-Amorós J, Torres-Cortada G, Camafort-Babkowski M. Blood pressure, congestion and heart failure with preserved ejection fraction among patients with and without type 2 diabetes mellitus. A cluster analysis approach from the observational registry DICUMAP. High Blood Press Cardiovasc Prev. 2020;27(5):399-408. doi:10.1007/s40292-020-00405-x
13. Casebeer A, Horter L, Hayden J, Simmons J, Evers T. Phenotypic clustering of heart failure with preserved ejection fraction reveals different rates of hospitalization. J Cardiovasc Med (Hagerstown). 2021;22(1):45-52. doi:10.2459/JCM.0000000000001116
14. Choy M, Liang W, He J, Fu M, Dong Y, He X, et al. Phenotypes of heart failure with preserved ejection fraction and effect of spironolactone treatment. ESC Heart Fail. 2022;9(4):2567-75.
15. Cohen JB, Schrauben SJ, Zhao L, Basso MD, Cvijic ME, Li Z, et al. Clinical phenogroups in heart failure with preserved ejection fraction: Detailed phenotypes, prognosis, and response to spironolactone. JACC Heart Fail. 2020;8(3):172-84.
16. Dong B, Yao Y, Xue R, Liang W, He J, Wei F, et al. Distinct implications of body mass index in different subgroups of nonobese patients with heart failure with preserved ejection fraction: A latent class analysis of data from the TOPCAT trial. BMC Med. 2022;20(1):423. doi:10.1186/s12916-022-02626-4
17. Fayol A, Wack M, Livrozet M, Carves JB, Domengé O, Vermersch E, et al. Aetiological classification and prognosis in patients with heart failure with preserved ejection fraction. ESC Heart Fail. 2022;9(1):519-30.
18. Gu J, Pan JA, Lin H, Zhang JF, Wang CQ. Characteristics, prognosis and treatment response in distinct phenogroups of heart failure with preserved ejection fraction. Int J Cardiol. 2021;323:148-54.
19. Harada D, Asanoi H, Noto T, Takagawa J. Different pathophysiology and outcomes of heart failure with preserved ejection fraction stratified by K-means clustering. Front Cardiovasc Med. 2020;7:607760.

20. Hedman ÅK, Hage C, Sharma A, Brosnan MJ, Buckbinder L, Gan LM, et al. Identification of novel phenogroups in heart failure with preserved ejection fraction using machine learning. Heart. 2020;106(5):342-9.
21. Kao DP, Lewsey JD, Anand IS, Massie BM, Zile MR, Carson PE, et al. Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response. Eur J Heart Fail. 2015;17(9):925-35.
22. Kobayashi Y, Tremblay-Gravel M, Boralkar KA, Li X, Nishi T, Amsallem M, et al. Approaching higher dimension imaging data using cluster-based hierarchical modeling in patients with heart failure preserved ejection fraction. Sci Rep. 2019;9(1):10431.
23. Kyodo A, Kanaoka K, Keshi A, Nogi M, Nogi K, Ishihara S, et al. Heart failure with preserved ejection fraction phenogroup classification using machine learning. ESC Heart Fail. 2023;10(3):2019-30.
24. Murray EM, Greene SJ, Rao VN, Sun JL, Alhanti BA, Blumer V, et al. Machine learning to define phenotypes and outcomes of patients hospitalized for heart failure with preserved ejection fraction: Findings from ASCEND-HF. Am Heart J. 2022;254:112-21.
25. Nouraei H, Nouraei H, Rabkin SW. Comparison of unsupervised machine learning approaches for cluster analysis to define subgroups of heart failure with preserved ejection fraction with different outcomes. Bioengineering (Basel). 2022;9(4):175.
26. Nouraei H, Rabkin SW. A new approach to the clinical subclassification of heart failure with preserved ejection fraction. Int J Cardiol. 2021;331:138-43.
27. Pandey A, Kagiyama N, Yanamala N, Segar MW, Cho JS, Tokodi M, et al. Deep-learning models for the echocardiographic assessment of diastolic dysfunction. JACC Cardiovasc Imaging. 2021;14(10):1887-900.
28. Perry A, Loh F, Adamo L, Zhang KW, Deych E, Foraker R, et al. Unsupervised cluster analysis of patients with recovered left ventricular ejection fraction identifies unique clinical phenotypes. PLoS One. 2021;16(3):e0248317.
29. Przewlocka-Kosmala M, Marwick TH, Dabrowski A, Kosmala W. Contribution of cardiovascular reserve to prognostic categories of heart failure with preserved ejection fraction: A classification based on machine learning. J Am Soc Echocardiogr. 2019;32(5):604-15.e6.
30. Sabbah MS, Fayyaz AU, de Denus S, Felker GM, Borlaug BA, Dasari S, et al. Obese-inflammatory phenotypes in heart failure with preserved ejection fraction. Circ Heart Fail. 2020;13(8):e006414.
31. Schrub F, Oger E, Bidaut A, Hage C, Charton M, Daubert JC, et al. Heart failure with preserved ejection fraction: A clustering approach to a heterogenous syndrome. Arch Cardiovasc Dis. 2020;113(6-7):381-90.
32. Segar MW, Patel KV, Ayers C, Basit M, Tang WHW, Willett D, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. Eur J Heart Fail. 2020;22(1):148-58.
33. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation. 2015;131(3):269-79.
34. Sotomi Y, Tamaki S, Hikoso S, Nakatani D, Okada K, Dohi T, et al. Pathophysiological insights into machine learning-based subphenotypes of acute heart failure with preserved ejection fraction. Heart. 2024;110(6):441-7.
35. Stienen S, Ferreira JP, Kobayashi M, Preud'homme G, Dobre D, Machu JL, et al. Enhanced clinical phenotyping by mechanistic bioprofiling in heart failure with preserved ejection fraction: Insights from the MEDIA-DHF study (The Metabolic Road to Diastolic Heart Failure). Biomarkers. 2020;25(2):201-11.
36. Uijl A, Savarese G, Vaartjes I, Dahlström U, Brugts JJ, Linssen GCM, et al. Identification of distinct phenotypic clusters in heart failure with preserved ejection fraction. Eur J Heart Fail. 2021;23(6):973-82.
37. Woolley RJ, Ceelen D, Ouwerkerk W, Tromp J, Figarska SM, Anker SD, et al. Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction. Eur J Heart Fail. 2021;23(6):983-91.
38. Jones E, Randall EB, Hummel SL, Cameron DM, Beard DA, Carlson BE. Phenotyping heart failure using model-based analysis and physiology-informed machine learning. J Physiol. 2021;599(22):4991-5013.
39. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2023 focused update of the 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2023;44(37):3627-39.
40. Pitt B, Pfeffer MA, Assmann SF, Boineau R, Anand IS, Claggett B, et al. Spironolactone for heart failure with preserved ejection fraction. N Engl J Med. 2014;370(15):1383-92.

41. Pfeffer MA, Claggett B, Assmann SF, Boineau R, Anand IS, Clausell N, et al. Regional variation in patients and outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) trial. Circulation. 2015;131(1):34-42.
42. Gu F, Meyer AJ, Ježek F, Zhang S, Catalan T, Miller A, et al. Identification of digital twins to guide interpretable AI for diagnosis and prognosis in heart failure. NPJ Digit Med. 2025;8(1):110.
43. Zhang J, Gan R, Wang J, Zhang Y, Zhang L, Yang P, et al. Fengshenbang 1.0: Being the foundation of Chinese cognitive intelligence. arXiv. 2022;preprint arXiv:220902970.
44. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work?. Int J Methods Psychiatr Res. 2011;20(1):40-9.
45. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53-65.
46. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J Roy Stat Soc B Stat Methodol. 2002;63(2):411-23.
47. Dunn JC. Well-separated clusters and optimal fuzzy partitions. J Cybern. 1974;4(1):95-104.
48. Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theor. 1982;28(2):129-37.
49. Tian K, Zhou S, Guan J. Deepcluster: A general clustering framework based on deep learning. Springer; 2017. p. 809-25.
50. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2:193-218.
51. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.
52. Rich JD, Burns J, Freed BH, Maurer MS, Burkhoff D, Shah SJ. Meta-analysis Global Group in Chronic (MAGGIC) heart failure risk score: Validation of a simple tool for the prediction of morbidity and mortality in heart failure with preserved ejection fraction. J Am Heart Assoc. 2018;7(20):e009594.
53. Bhatt DL, Szarek M, Pitt B, Cannon CP, Leiter LA, McGuire DK, et al. Sotagliflozin in patients with diabetes and chronic kidney disease. N Engl J Med. 2021;384(2):129-39.
54. Heerspink HJL, Oshima M, Zhang H, Li J, Agarwal R, Capuano G, et al. Canagliflozin and kidney-related adverse events in type 2 diabetes and CKD: Findings from the randomized CREDENCE trial. Am J Kidney Dis. 2022;79(2):244-56.e1.
55. Heerspink HJL, Stefánsson BV, Correa-Rotter R, Chertow GM, Greene T, Hou FF, et al. Dapagliflozin in patients with chronic kidney disease. N Engl J Med. 2020;383(15):1436-46.
56. Yang L, Ye N, Bian W, Cheng H. Efficacy of medication therapy for patients with chronic kidney disease and heart failure with preserved ejection fraction: A systematic review and meta-analysis. Int Urol Nephrol. 2022;54(6):1435-44.
57. Pieske B, Wachter R, Shah SJ, Baldridge A, Szeczoedy P, Ibram G, et al. Effect of sacubitril/valsartan vs standard medical therapies on plasma NT-proBNP concentration and submaximal exercise capacity in patients with heart failure and preserved ejection fraction: The PARALLAX randomized clinical trial. JAMA. 2021;326(19):1919-29.
58. Solomon SD, Zile M, Pieske B, Voors A, Shah A, Kraigher-Krainer E, et al. The angiotensin receptor neprilysin inhibitor LCZ696 in heart failure with preserved ejection fraction: A phase 2 double-blind randomised controlled trial. Lancet. 2012;380(9851):1387-95.
59. Solomon SD, McMurray JJV, Anand IS, Ge J, Lam CPS, Maggioni AP, et al. Angiotensin-neprilysin inhibition in heart failure with preserved ejection fraction. N Engl J Med. 2019;381(17):1609-20.
60. Wijkman MO, Claggett B, Vaduganathan M, Cunningham JW, Rørth R, Jackson A, et al. Effects of sacubitril/valsartan on glycemia in patients with diabetes and heart failure: The PARAGON-HF and PARADIGM-HF trials. Cardiovasc Diabetol. 2022;21(1):110.
61. Kang H, Zhang J, Zhang X, Qin G, Wang K, Deng Z, et al. Effects of sacubitril/valsartan in patients with heart failure and chronic kidney disease: A meta-analysis. Eur J Pharmacol. 2020;884:173444.
62. Jia R, Zhang X, Xu Y, Zheng Z, Jiang L, Zhang X, et al. Effect of sacubitril/valsartan on renal function in patients with chronic kidney disease and heart failure with preserved ejection fraction: A real-world 12-week study. Eur J Pharmacol. 2022;928:175053.
63. Solomon SD, Vaduganathan M, Claggett BL, Packer M, Zile M, Swedberg K, et al. Sacubitril/valsartan across the spectrum of ejection fraction in heart failure. Circulation. 2020;141(5):352-61.

64. Patel K, Fonarow GC, Ahmed M, Morgan C, Kilgore M, Love TE, et al. Calcium channel blockers and outcomes in older patients with heart failure and preserved ejection fraction. Circ Heart Fail. 2014;7(6):945-52.
65. Wu CK, Lee JK, Chiang FT, Lin LY, Lin JW, Hwang JJ, et al. Prognostic factors of heart failure with preserved ejection fraction: A 12-year prospective cohort follow-up study. Int J Cardiol. 2014;171(3):331-7.
66. Wang X, Ju J, Chen Z, Lin Q, Zhang Z, Li Q, et al. Associations between calcium channel blocker therapy and mortality in heart failure with preserved ejection fraction. Eur J Prev Cardiol. 2022;29(9):1343-51.
67. Chao TF, Liu CJ, Tuan TC, Chen SJ, Wang KL, Lin YJ, et al. Rate-control treatment and mortality in atrial fibrillation. Circulation. 2015;132(17):1604-12.
68. Meyer M, Wetmore JB, Weinhandl ED, Roetker NS. Association of nondihydropyridine calcium channel blockers versus β-adrenergic receptor blockers with risk of heart failure hospitalization. Am J Cardiol. 2023;197:68-74.
69. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: A report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2022;145(18):e895-1032.