Phenotypes of atrial fibrillation in a Taiwanese longitudinal cohort: Insights from an Asian perspective

Jung-Chi Hsu, MD, PhD,[1,2] Yen-Yun Yang, MD,[3] Shu-Lin Chuang, PhD,[3] Lian-Yu Lin, MD, PhD[2,4,5]

From the [1]Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Jinshan Branch, New Taipei City, Taiwan, [2]Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan, [3]Department of Medical Research, National Taiwan University Hospital, Taipei, Taiwan, [4]Cardiovascular Center, National Taiwan University Hospital, Taipei, Taiwan, and [5]Department of Internal Medicine, National Taiwan University Hospital, Yunlin Branch, Yunlin, Taiwan.

BACKGROUND Atrial fibrillation (AF) is a condition with heterogeneous underlying causes, often involving multiple cardiovascular comorbidities. Large-scale studies examining the heterogeneity of patients with AF in the Asian population are limited.

OBJECTIVES The purpose of this study was to identify distinct phenotypic clusters of patients with AF and evaluate their associated risks of ischemic stroke, heart failure hospitalization, cardiovascular mortality, and all-cause mortality.

METHODS We analyzed 5002 adult patients with AF from the National Taiwan University Hospital between 2014 and 2019 using an unsupervised hierarchical cluster analysis based on the CHA_2DS_2-VASc score.

RESULTS We identified 4 distinct groups of patients with AF: cluster I included diabetic patients with heart failure preserved ejection fraction as well as chronic kidney disease (CKD); cluster II comprised older patients with low body mass index and pulmonary hypertension; cluster III consisted of patients with metabolic syndrome and atherosclerotic disease; and cluster IV comprised patients with left heart dysfunction, including reduced ejection fraction. Differences in the risk of ischemic stroke across clusters (clusters I, II, and III vs cluster IV) were statistically significant (hazard ratio [HR] 1.87, 95% confidence interval [CI] 1.00–3.48; HR 2.06, 95% CI 1.06–4.01; and HR 1.70, 95% CI 1.02–2.01). Cluster II was independently associated with the highest risk of hospitalization for heart failure (HR 1.19, 95% CI 0.79–1.80), cardiovascular mortality (HR 2.51, 95% CI 1.21–5.22), and overall mortality (HR 2.98, 95% CI 1.21–4.2).

CONCLUSION A data-driven algorithm can identify distinct clusters with unique phenotypes and varying risks of cardiovascular outcomes in patients with AF, enhancing risk stratification beyond the CHA_2DS_2-VASc score.

KEYWORDS Atrial fibrillation; Stroke; Heart failure; Phenotyping; Cluster analysis; Machine learning

(Heart Rhythm O^2 2025;6:129–138)

Introduction

Atrial fibrillation (AF) is associated with significant risks of cardiovascular mortality and morbidity, making the identification of individuals at high risk of stroke and thromboembolism crucial for therapeutic decision making.[1] Risk scores such as CHA_2DS_2-VASc are commonly used to identify high-risk individuals; however, their predictive accuracy can be limited. While integrating biomarkers into these risk scores might enhance their precision, it also adds complexity and reduces their practical applicability in clinical settings.[2] The onset and progression of AF are influenced by a range of comorbidities, each affecting patient outcomes differently. The interplay among these conditions can further complicate the disease, suggesting that standard classifications may fall short in adequately defining patients with AF.

In recent years, machine learning has emerged as a tool to improve disease classification, diagnosis, prognosis, and treatment selection, with potential applications in precision medicine.[3–5] Cluster analysis, a common unsupervised, data-driven method, is an exploratory statistical technique that organizes data into natural clusters without relying on predefined assumptions. This method

Address reprint requests and correspondence: Dr Jung-Chi Hsu, Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Jinshan Branch, No.7, Yulu Rd, Wuhu Village, Jinshan District, New Taipei City 20844, Taiwan. E-mail address: cactus146@gmail.com.

2666-5018/

https://doi.org/10.1016/j.hroo.2024.11.009

KEY FINDINGS

- The study identified 4 phenotypic clusters of patients with atrial fibrillation: (1) diabetes with heart failure and chronic kidney disease, (2) higher age with low body mass index and pulmonary hypertension, (3) metabolic syndrome with atherosclerotic disease, and (4) left heart dysfunction with reduced ejection fraction.
- Significant differences in ischemic stroke risk were observed across clusters. Cluster II had the highest rates of ischemic stroke, heart failure hospitalization, cardiovascular mortality, and all-cause mortality, with approximately twice the stroke risk and thrice the mortality rates compared to other clusters.
- In our highest-risk cluster (cluster II), approximately 23.5% of patients did not comply with anticoagulant treatment guidelines. This noncompliance could potentially affect clinical outcomes and underscore the need for better adherence to treatment protocols.

can reveal meaningful differences between disease phenotypes and outcomes that may not be captured by traditional classification systems. In the cardiovascular field, cluster analysis has successfully grouped heterogeneous patients into distinct clusters on the basis of various clinical attributes, revealing disease diversity that conventional methods might miss.[6–9]

Our research aims to enhance the understanding of AF by clustering individuals with similar risk factors and outcomes. This approach captures disease diversity that standard classifications might overlook. By using advanced phenotyping methods, including biomarkers and cardiovascular imaging, we seek to gain deeper insights into the complex relationships between risk factors and outcomes in patients with AF.

Methods

Study design and cohort

Between January 1, 2014, and December 31, 2019, we evaluated data from a tertiary medical facility. Our previous research explored the association between diabetes and AF using the National Taiwan University Hospital database.[10,11] External validation was conducted using data from the Yunlin branch hospital in central southern Taiwan. The data were sourced from the National Taiwan University Hospital-integrated Medical Database (NTUH-iMD), which includes information on *International Classification of Diseases, Tenth Revision* codes, Anatomical Therapeutic Chemical Classification drug codes, and regulated examination codes in Taiwan. The research reported in this article adhered to the Helsinki Declaration. The Institutional Review Board of the National Taiwan University Hospital approved the research (202007138RIND and 202403114RINC). We excluded patients lost to follow-up and included those 50 years or older (defined as no outpatient clinic visits for >3 months). From electronic health records, we extracted baseline characteristics such as hypertension, diabetes, hyperlipidemia, gout, heart failure, coronary artery disease, valvular heart disease, CKD, chronic obstructive pulmonary disease, and peripheral arterial occlusive disease. In addition, we collected data on estimated glomerular filtration rate, transient ischemic attack (TIA)/ischemic stroke history, medications, blood test results, and echocardiographic examinations. The outcomes of interest include TIA/ischemic stroke, hospitalization due to decompensated heart failure, cardiovascular mortality, and all-cause mortality.

Statistical analysis

Hierarchical cluster analysis was performed using the components of the CHA_2DS_2-VASc score, with the average linkage approach and Euclidean distance.[12] Smaller Euclidean distances between patients or clusters indicate greater similarity, meaning that patients with closer distances share

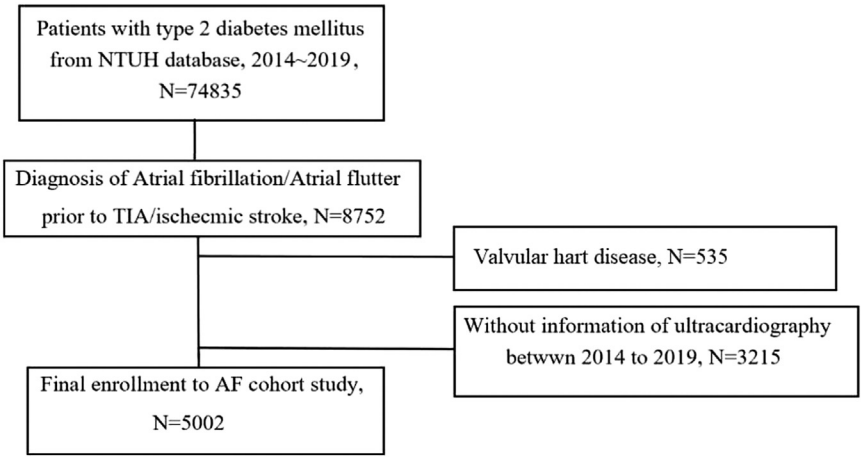

Figure 1 Study flowchart. AF = atrial fibrillation; NTUH = National Taiwan University Hospital; TIA = transient ischemic attack.

Table 1 Patients' characteristics of 4 clusters

Characteristic	Cluster I (n=1918)	Cluster II (n=1006)	Cluster III (n=1731)	Cluster IV (n=347)	*P*
Age (y)	77.2±3.2*	87.2±3.4*	65.5±3.9*	53.8±2.7*	<.001
Male	1017 (53.0)	441 (43.8)	1164 (67.2)	271 (78.1)	<.001
BMI (kg/m^2)	25.1±4.6*	24.3±4.2*	26.1±4.8*	26.9±5.2*	<.001
CHA_2DS_2-VASc score	4.0±1.4*	4.4±1.3*	2.7±1.4*	1.6±1.3*	<.001
HF group					<.001
Normal	218 (11.4)	111 (11.0)	264 (15.3)	86 (24.8)	
HFpEF	1398 (72.9)	743 (73.9)	1114 (64.4)	162 (46.7)	
HFmEF	135 (7.0)	72 (7.2)	134 (7.7)	41 (11.8)	
HFrEF	167 (8.7)	80 (7.9)	219 (12.6)	58 (16.7)	
History of HF hospitalization	86 (4.5)	73 (7.3)	92 (5.3)	21 (6.1)	.017
History of TIA/stroke	95 (4.9)	49 (6.5)	69 (4.0)	6 (1.7)	.001
HTN	1377 (71.8)	763 (75.8)	1256 (72.6)	216 (62.2)	<.001
DM	1232 (64.2)	647 (64.3)	1103 (63.7)	183 (52.7)	<.001
Hyperlipidemia	428 (22.3)	248 (24.7)	403 (23.3)	49 (14.1)	.001
Gout	96 (5.0)	67 (6.7)	81 (4.7)	10 (2.9)	.026
COPD	131 (6.8)	90 (8.9)	74 (4.3)	6 (1.7)	<.001
PAOD	137 (7.1)	100 (9.9)	128 (7.4)	11 (3.2)	<.001
CAD	405 (21.1)	231 (22.9)	381 (22.0)	47 (32.0)	.002
AMI	31 (1.6)	19 (1.9)	38 (2.2)	10 (2.9)	.356
CABG	8 (0.4)	4 (0.4)	9 (0.5)	0 (0.0)	.596
CKD	207 (10.8)	140 (13.9)	176 (10.2)	28 (8.1)	.004
Stage 3–5 CKD	807 (42.1)	537 (52.4)	463 (26.7)	67 (19.3)	<.001
Hemodialysis	40 (2.1)	15 (1.5)	67 (3.9)	15 (4.3)	<.001
Carotid stenosis	7 (0.4)	8 (0.8)	3 (0.2)	0 (0.0)	.041
FG (mg/dL)	133.7±101.3	130.4±57.7	131.2±58.9	139.8±130.9	.343
HbA1C (%)	6.8±1.3	6.7±1.1*	6.8±1.3*	6.8±1.3*	.001
TCHO (mg/dL)	155.8±41.1*	151.6±43.7#	161.7±43.0*#	157.9±42.6*#	<.001
TG (mg/dL)	129.8±96.7*	122.1±73.9#	143.4±110.7*#	135.5±104.6*#	<.001
LDL (mg/dL)	90.9±31.4*	95.9±32.4#	99.7±36.3*#	92.7±31.9*#	<.001
HDL (mg/dL)	43.2±13.0*	43.0±13.6#	43.1±13.0&	40.3±12.2*#&	.012
ALT (IU/L)	24.7±41.3*	20.7±24.8#	28.9±53.9#	34.2±112.5*#	<.001
eGFR (mL/min/1.73 m^2)	45.1±23.0*	34.8±17.9*	61.6±31.7*	75.3±36.1*	<.001
hsCRP (mg/dL)	6.0±7.0*	5.7±6.4*#	6.3±7.4#	5.4±6.5#	.255
NT-pro-BNP (pg/mL)	4697.4±6735.5*	6132.7±7029.9*#	4288.5±6777.3#	3566.1±5938.1#	<.001
IVSd (cm)	1.2±0.2	1.2±0.2	1.2±0.2	1.2±0.2	.448
LVIDd (cm)	4.8±0.7*	4.6±0.7*	4.9±0.7*	5.1±0.8*	<.001
LVIDs (cm)	3.1±0.8*	3.0±0.7*	3.3±0.8*	3.5±0.9*	<.001
LVPWd (cm)	1.1±0.2	1.1±0.2	1.1±0.2	1.1±0.2	.462
FS (%)	34.9±8.7*	35.2±8.5#	33.6±9.6*#	30.9±9.8*#	<.001
EDV (ml)	109.9±37.3*	102.4±35.6*	117.3±39.9*	125.7±48.2*	<.001
ESV (ml)	43.0±29.2*	38.9±24.5*	48.8±33.6*	58.1±42.2*	<.001
LVAd (cm^2)	29.3±8.6*	26.7±7.9*	31.2±8.4*	33.0±9.4*	<.001
LVAs (cm^2)	20.4±7.7*	18.6±6.8#	22.5±7.7*#	24.1±7.5*#	<.001
LVLd (cm)	7.5±0.9*	7.2±0.9*	7.8±1.0*	8.1±1.1*	<.001
LVLs (cm)	6.7±1.1*	6.5±0.9#	7.1±1.1*#	7.3±1.2*#	<.001
LVEF (%)	63.0±12.7	63.7±12.0*	60.9±14.4*	57.1±15.2*	<.001
LVM (g)	207.3±64.1*	196.6.1±60.3*	215.7±68.3*	230.6±82.6*	<.001
TRPG max velocity (cm/s)	279.9±51.3*	289.5±48.6*	269.9±50.4*	261.1±60.6*	<.001
TRPG max PG (mmHg)	32.4±11.9*	34.5±11.6*	30.2±11.7*	28.7±14.0*	<.001
LA size (cm)	4.3±0.8*	4.2±0.8#	4.3±0.8#	4.4±0.9*#	<.001
MV E max vel (cm/s)	96.4±30.2	99.6±30.7*	94.9±29.9*	94.7±33.1	.002
MV A max vel (cm/s)	89.9±28.1*	96.3±32.4*	84.9±28.4*	77.6±28.7*	<.001
MV E/A	1.1±1.9*	2.0±2.1*#	1.1±0.8#	1.2±5.7	.031
MV dec time (sec)	0.21±0.07*	0.20±0.07#	0.19±0.06*	0.18±0.07*#	<.001
Lateral E' (cm/s)	8.4±3.3	7.9±3.2*	8.5±3.2*	9.2±3.8*	.003
Lateral A' (cm/s)	8.5±3.2	9.8±3.9	8.8±3.8	9.5±3.2	.226
Antiplatelet	980 (51.1)	540 (53.7)	862 (49.8)	168 (48.4)	.185
Anticoagulation	992 (51.7)	438 (43.5)	921 (53.2)	181 (52.2)	<.001
CCB	1337 (69.7)	697 (69.3)	1086 (62.7)	201 (57.9)	<.001
β-Blocker	1207 (62.9)	563 (56.0)	1112 (64.2)	239 (68.9)	<.001
ACEI/ARB	1175 (61.3)	582 (57.9)	1027 (59.3)	219 (63.1)	.177
Diuretics	1280 (66.7)	766 (76.1)	933 (53.9)	186 (53.6)	<.001
Statin	817 (42.6)	334 (33.2)	826 (47.7)	171 (49.3)	<.001

(*Continued*)

Table 1 (*Continued*)

Characteristic	Cluster I (n=1918)	Cluster II (n=1006)	Cluster III (n=1731)	Cluster IV (n=347)	*P*
Propafenone	327 (17.0)	130 (12.9)	350 (20.2)	78 (22.4)	<.001
Amiodarone	986 (51.4)	469 (46.6)	935 (54.0)	183 (52.7)	.003
Overall mortality	701 (36.5)	533 (53.0)	422 (24.4)	57 (16.4)	<.001
Cardiovascular mortality	198 (10.3)	160 (15.9)	125 (7.2)	23 (6.6)	<.001
HF hospitalization	526 (27.4)	373 (37.1)	372 (21.5)	93 (26.8)	<.001
TIA/ischemic stroke	244 (12.7)	152 (15.1)	172 (9.9)	22 (6.3)	<.001

Values are presented as mean ± SD or n (%).

A' = late diastolic mitral annular velocity; ACEI/ARB = angiotensin-converting enzyme inhibitor/angiotensin receptor blocker; ALT = alanine aminotransferase; AMI = acute myocardial infarction; BMI = body mass index; CABG = coronary artery bypass graft; CAD = coronary artery disease; CCB = calcium channel blocker; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; DM = diabetes mellitus; E' = early diastolic mitral annular velocity; E/A = early diastolic transmitral flow velocity/late diastolic transmitral flow velocity; EDV = end-diastolic volume; eGFR = estimated glomerular filtration rate; ESV = end-systolic volume; FG = fasting glucose; FS = fractional shortening; HbA1C = glycated hemoglobin A1C; HDL = high-density lipoprotein; HF = heart failure; HFmEF = heart failure with mid-range ejection fraction; HFpEF = heart failure with preserved ejection fraction; HFrEF = heart failure with reduced ejection fraction; hsCRP = high-sensitivity C-reactive protein; HTN = hypertension; IVSd = interventricular septal end diastole; LA = left atrial; LDL = low-density lipoprotein; LV = left ventricular; LVAd = left ventricular area in diastole; LVAs = left ventricular area in systole; LVEF = left ventricular ejection fraction; LVIDd = left ventricular internal diameter end diastole; LVIDs = left ventricular internal diameter end systole; LVLd = left ventricular length in diastole; LVLs = left ventricular length in systole; LVM = left ventricular mass; LVPWd = left ventricular posterior wall end diastole; MV = mitral valve; MV A max vel = maximum velocity of mitral valve A wave; MV E max vel = maximum velocity of mitral valve E wave; NT-pro-BNP = N-terminal pro–brain natriuretic peptide; PAOD = peripheral arterial occlusive disease; TCHO = total cholesterol; TG = triglyceride; TIA = transient ischemic attack; TRPG max = tricuspid regurgitation peak gradient maximum; TRPG max PG = tricuspid regurgitation peak gradient maximum pressure gradient.

*, #, &: Intragroup statistical significance $P < .05$ by Bonferroni test.

more similar CHA_2DS_2-VASc score profiles. Missing values were excluded without imputation to ensure that the analysis was based solely on complete cases. The optimal number of clusters was determined using the elbow method, which identifies the point at which the addition of more clusters yields diminishing returns in within-cluster variance. The hierarchical clustering model was developed using the primary data set and subsequently applied to an external validation set to assess the stability and generalizability of the identified clusters. This external validation set comprised patients with AF from a separate cohort in Taiwan. We compared the clinical features and outcomes of the clusters between the primary and external data sets to evaluate the consistency and relevance of the clusters across different patient populations.

Continuous variables were summarized using means and standard deviations, while categorical variables were presented as frequencies and percentages. Differences between clusters were analyzed using the χ^2 test for categorical data and 1-way analysis of variance for continuous variables. The Bonferroni correction was applied to adjust for multiple

Table 2 Multivariable Cox regression stratified by clusters

Outcome	Crude	Model 1	Model 2
TIA/ischemic stroke			
Cluster I	2.21 (1.43–3.41)	1.97 (1.10–3.50)	1.87 (1.00–3.48)
Cluster II	2.79 (1.78–4.36)	2.24 (1.20–4.20)	2.06 (1.06–4.01)
Cluster III	1.68 (1.08–2.62)	1.76 (1.09–2.8)	1.70 (1.02–2.84)
Cluster IV (reference)			
HF hospitalization			
Cluster I	0.82 (0.65–1.03)	0.60 (0.48–0.98)	0.79 (0.54–1.16)
Cluster II	1.73 (1.38–2.17)	0.95 (0.65–1.40)	1.19 (0.79–1.80)
Cluster III	1.10 (0.88–1.38)	0.68 (0.52–0.89)	0.77 (0.57–1.03)
Cluster IV (reference)			
Cardiovascular mortality			
Cluster I	1.84 (1.19–2.83)	1.36 (0.71–2.60)	1.57 (0.79–3.13)
Cluster II	3.33 (2.15–5.16)	2.20 (1.10–4.40)	2.51 (1.21–5.22)
Cluster III	1.22 (0.78–1.90)	1.26 (0.77–2.10)	1.32 (0.78–2.24)
Cluster IV (reference)			
Overall mortality			
Cluster I	2.64 (2.02–3.46)	1.99 (1.38–2.88)	1.89 (1.28–2.80)
Cluster II	4.52 (3.44–5.94)	3.13 (2.13–4.62)	2.98 (1.98–4.50)
Cluster III	1.67 (1.26–2.20)	1.65 (1.22–2.23)	1.63 (1.18–2.25)
Cluster IV (reference)			

Values are presented as hazard ratio (95% confidence interval).

Model 1***: adjusting for age (<65, 65–74, ≥75 y), sex, hypertension, hyperlipidemia, gout, history of HF, CAD, COPD, PAOD, prior TIA/ischemic stroke, and CKD. Model 2***: adjusting for model 1, plus baseline LA size, baseline LVEF, and baseline LVM. ***$P < .001$.

Abbreviations as in Table 1.

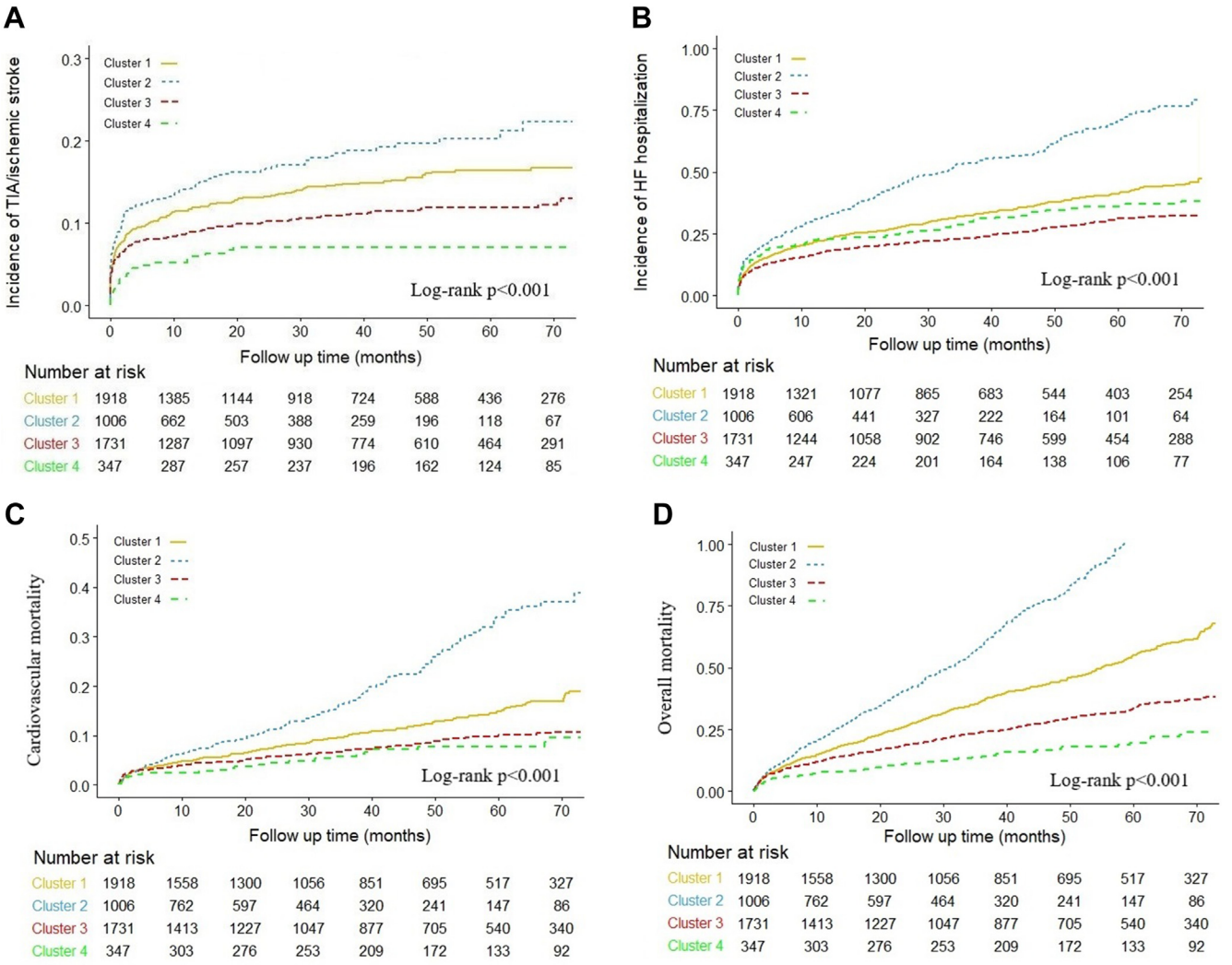

Figure 2 Kaplan-Meier analyses for (**A**) transient ischemic attack (TIA)/ischemic stroke, (**B**) heart failure (HF) hospitalization, (**C**) cardiovascular mortality, and (**D**) overall mortality.

comparisons. Cumulative incidences for outcomes were estimated using the Kaplan-Meier method, and differences across survival curves were evaluated using the log-rank test. Hazard ratios (HRs) and 95% confidence intervals (CIs) were determined using multivariable Cox regression models, with stepwise adjustments for confounding variables to ensure robust associations across progressively complex models. All statistical analyses were performed using R version 4.2 (The R Foundation for Statistical Computing, Vienna, Austria) and SPSS version 25.0 (IBM Corp, Armonk, NY). A *P* value of .05 or less was considered statistically significant.

Results

Baseline characteristics

Figure 1 illustrates the cohort flowchart of the study. Between 2014 and 2019, a total of 174,835 patients were initially enrolled. Of these patients, 8752 (5.0%) were identified as having AF. The cohort was further refined by excluding 535 individuals with valvular heart disease and 3215 patients without echocardiographic data. Consequently, the final AF cohort consisted of 5002 individuals.

Table 1 summarizes the clinical characteristics of patients of the 4 identified clusters. Cluster I, comprising 38.3% of the population with AF, had a mean age of 77.2 years. Patients in this cluster were more likely to have diabetes mellitus (DM; 64.2%), severe stage III CKD (42.1%), and heart failure with preserved ejection fraction (HFpEF) (72.9%). Although these rates were lower than those in cluster II, they remained higher than those in clusters III and IV. Cluster II, accounting for 20.1% of the population with AF, had the highest mean age (87.2 years) and the highest prevalence of hypertension (75.8%), chronic obstructive pulmonary disease (8.9%), and pulmonary hypertension (mean tricuspid regurgitation peak gradient maximum pressure gradient 34.5 mm Hg). This cluster also had the lowest mean body mass index (BMI; 24.3 kg/m^2) and the highest CHA_2DS_2-VASc score (mean score 4.4) and history of heart failure hospitalization (7.3%). Cluster III, representing 34.6% of the population

A

Subgroup		Patient No.	Adjusted HR	(95%CI)	p-value
Cluster	cluster 4	347		reference	
	cluster 1	1,918	1.87	(1.00- 3.48)	0.050
	cluster 2	1,006	2.06	(1.06- 4.01)	0.033 *
	cluster 3	1,731	1.70	(1.02- 2.84)	0.044 *
AGE	age<65	1,003		reference	
	65<=age<75	1,537	0.93	(0.67- 1.28)	0.644
	age>=75	2,426	1.11	(0.70- 1.74)	0.663
SEX		5,002	0.93	(0.77- 1.11)	0.406
HTN		5,002	0.77	(0.61- 0.96)	0.023 *
DM		5,002	0.83	(0.67- 1.02)	0.080
HLP		5,002	0.85	(0.67- 1.07)	0.158
COPD		5,002	0.67	(0.44- 1.03)	0.070
CKD		5,002	0.84	(0.62- 1.14)	0.264
Gout		5,002	0.68	(0.42- 1.11)	0.125
HFEF_group		5,002	0.88	(0.73- 1.06)	0.179
PAOD		5,002	0.86	(0.61- 1.20)	0.371
CAD		5,002	0.86	(0.67- 1.10)	0.241
EF_Teich_		5,002	1.00	(0.99- 1.01)	0.692
Lvmass		5,002	1.00	(1.00- 1.00)	0.618
LA		5,002	1.08	(0.97- 1.22)	0.172

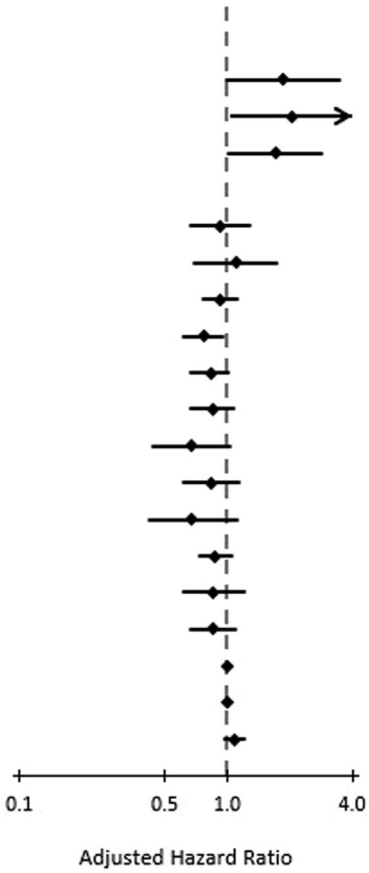

Figure 3 Subgroup analyses for (**A**) transient ischemic attack/ischemic stroke, (**B**) heart failure hospitalization, (**C**) cardiovascular mortality, and (**D**) overall mortality. CAD = coronary artery disease; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; DM = diabetes mellitus; EF = ejection fraction; HFEF = heart failure ejection fraction; HTN = hypertension; HDL = high-density lipoprotein; TIA = transient ischemic attack; LA = left atrial; LV = left ventricular; PAOD = peripheral arterial occlusive disease.

with AF, had a mean age of 65.5 years and was characterized by hyperlipidemia. This cluster had the highest average levels of total cholesterol (161.7 mg/dL), triglycerides (143.4 mg/dL), and low-density lipoprotein (99.7 mg/dL). Cluster IV, which comprised 6.9% of the population with AF, had the lowest mean age (53.8 years), the highest proportion of male patients (78.1%), and the highest mean BMI (26.9 kg/m^2). This cluster also had the highest incidence of heart failure with reduced ejection fraction (HFrEF) at 16.7%, the lowest mean left ventricular ejection fraction (LVEF; 57.1%), the largest mean left atrial size (4.4 cm), and the largest mean left ventricular internal dimension in diastole (5.1 cm).

Clinical outcomes of each cluster

Table 2 presents the association between clusters and clinical outcomes, adjusted for various covariates. Compared to cluster IV, patients in clusters II, I, and III exhibited a higher risk of TIA/ischemic stroke (HR 2.79, 95% CI 1.78–4.36; HR 2.21, 95% CI 1.43–3.41; and HR 1.68, 95% CI 1.08–2.62), even after adjusting for baseline risk factors and echocardiographic parameters (HR 2.06, 95% CI 1.06–4.01; HR 1.87, 95% CI 1.00–3.48; and HR 1.70, 95% CI 1.02–2.01). Both cardiovascular mortality and overall mortality followed similar trends. The likelihood of hospitalization due to heart failure did not show a consistent trend across clusters.

Figure 2 displays the cumulative incidence of TIA/ischemic stroke, heart failure hospitalization, cardiovascular mortality, and overall mortality for each cluster. Cluster II had the highest incidence, followed by clusters I, III, and IV, which had the lowest incidence. The log-rank test was significant for all comparisons ($P<.001$). Figure 3 shows the forest plots of the HRs for the subgroup analysis.

External validation

To validate our findings, we used an independent cohort of patients with AF from Yunlin, an agricultural region in central southern Taiwan. The data set spans from 2014 to 2019 and includes 6822 individuals. After excluding 325 patients with valvular heart disease and 3606 patients without an echocardiogram, the final cohort comprised 2891 individuals. Nearly one-third of these patients were more than 80

B

Subgroup		Patient No.	Adjusted HR	(95%CI)	p-value	
Cluster	cluster 4	347		reference		
	cluster 1	1,918	0.79	(0.54- 1.16)	0.231	
	cluster 2	1,006	1.19	(0.79- 1.80)	0.396	
	cluster 3	1,731	0.77	(0.57- 1.03)	0.078	
AGE	age<65	1,003		reference		
	65<=age<75	1,537	1.42	(1.13- 1.79)	0.003	*
	age>=75	2,426	1.80	(1.30- 2.48)	<0.001	*
SEX		5,002	0.62	(0.55- 0.70)	<0.001	*
HTN		5,002	0.92	(0.79- 1.07)	0.256	
DM		5,002	1.08	(0.94- 1.25)	0.250	
HLP		5,002	1.12	(0.97- 1.30)	0.132	
COPD		5,002	1.44	(1.17- 1.78)	<0.001	*
CKD		5,002	1.60	(1.35- 1.90)	<0.001	*
Gout		5,002	1.12	(0.87- 1.44)	0.386	
history_TIAstroke		5,002	0.83	(0.63- 1.10)	0.197	
HFEF_group		5,002	1.26	(1.12- 1.43)	<0.001	*
PAOD		5,002	1.19	(0.97- 1.45)	0.091	
CAD		5,002	1.08	(0.93- 1.25)	0.337	
EF_Teich_		5,002	0.97	(0.96- 0.97)	<0.001	*
Lvmass		5,002	1.00	(1.00- 1.00)	<0.001	*
LA		5,002	1.06	(0.98- 1.14)	0.154	

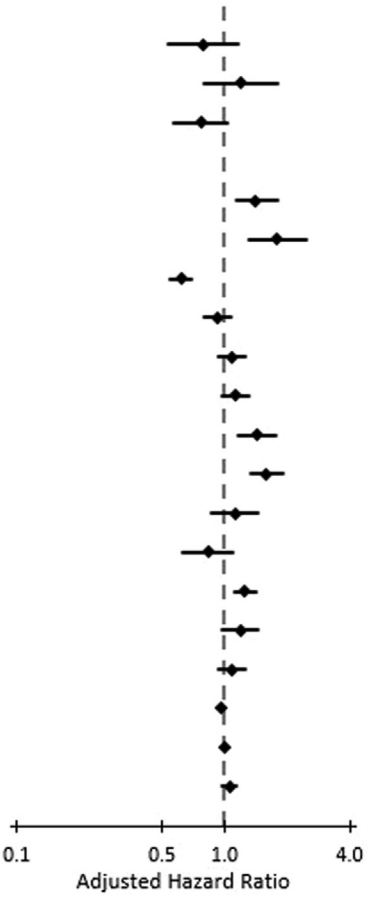

Figure 3 (*continued*).

years old, with some reaching their centenary, highlighting a significant aging trend (Online Supplemental Table 1). Despite the presence of extreme values, the cluster characteristics in Yunlin were consistent with those in the primary cohort (Online Supplemental Table 2). Elderly patients in external cluster II had the lowest BMI, aligning with previous research linking low BMI to a higher risk profile.

External cluster I corresponds to cluster II in the training set, showing the highest risk profile. External cluster II aligns with cluster I in the training set, reflecting an intermediate risk level. External cluster III matches with cluster IV in the training set and demonstrates less pronounced risk differences, while external cluster IV corresponds to cluster III in the training set. The stability of cluster characteristics, including BMI, HFpEF, and CKD, across different populations reinforces the robustness of our clustering methodology. The Kaplan-Meier results for the external clusters, as shown in Online Supplemental Figure 1, are consistent with the risk trends of their corresponding clusters in the training set. External clusters I and II exhibit significant differences in risk, while external clusters III and IV display more subtle variations. This consistency enhances the generalizability and reliability of our findings across diverse settings.

Discussion

Our study elucidates 4 distinct phenotypes of patients with AF, each characterized by specific clinical features:

- Cluster I: comprising patients with type 2 diabetes, chronic renal disease, and HFpEF
- Cluster II: encompassing elderly individuals with low BMI, chronic obstructive pulmonary disease, and pulmonary hypertension
- Cluster III: consisting of patients with hyperlipidemia and atherosclerosis
- Cluster IV: predominantly men with reduced systolic function and larger left ventricular and left atrial chambers.

Previous research has shown variability in AF and heart failure phenotypes across different regions, including within Asian populations. Studies of heart failure in Asian cohorts have identified variations in prognosis based on factors such as diabetes, age, and AF. In Taiwan, for example, a metabolic phenotype is predominantly characterized by a higher prevalence of diabetic HFpEF.[13] This suggests that even within Asia, different countries and regions may exhibit distinct disease profiles and characteristics because of the heterogeneous nature of these conditions.

C

	Subgroup	Patient No.	Adjusted HR	(95%CI)	p-value	
Cluster	cluster 4	347		reference		
	cluster 1	1,918	1.57	(0.79- 3.13)	0.201	
	cluster 2	1,006	2.51	(1.21- 5.22)	0.013	*
	cluster 3	1,731	1.32	(0.78- 2.24)	0.309	
AGE	age<65	1,003		reference		
	65<=age<75	1,537	0.92	(0.62- 1.35)	0.660	
	age>=75	2,426	1.66	(0.95- 2.89)	0.074	
SEX		5,002	0.93	(0.76- 1.14)	0.474	
HTN		5,002	0.91	(0.70- 1.18)	0.482	
DM		5,002	1.57	(1.23- 2.00)	<0.001	***
HLP		5,002	0.90	(0.70- 1.15)	0.379	
COPD		5,002	0.82	(0.55- 1.23)	0.331	
CKD		5,002	1.01	(0.74- 1.37)	0.944	
Gout		5,002	1.26	(0.82- 1.95)	0.293	
history_TIAstroke		5,002	1.29	(0.88- 1.91)	0.192	
HFEF_group		5,002	1.41	(1.15- 1.73)	0.001	**
PAOD		5,002	1.76	(1.31- 2.36)	<0.001	***
CAD		5,002	1.35	(1.06- 1.71)	0.014	*
EF_Teich_		5,002	0.97	(0.95- 0.98)	<0.001	***
Lvmass		5,002	1.00	(1.00- 1.00)	0.571	
LA		5,002	0.95	(0.84- 1.08)	0.472	

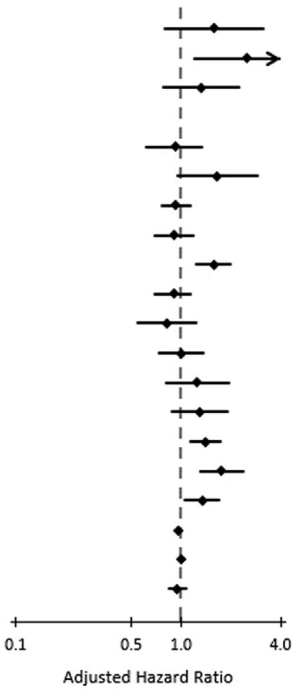

Figure 3 (*continued*).

D

	Subgroup	Patient No.	Adjusted HR	(95%CI)	p-value	
Cluster	cluster 4	347		reference		
	cluster 1	1,918	1.89	(1.28- 2.80)	0.001	*
	cluster 2	1,006	2.98	(1.98- 4.50)	<0.001	*
	cluster 3	1,731	1.63	(1.18- 2.25)	0.003	*
AGE	age<65	1,003		reference		
	65<=age<75	1,537	1.02	(0.83- 1.25)	0.886	
	age>=75	2,426	1.59	(1.19- 2.12)	0.002	*
SEX		5,002	1.08	(0.97- 1.20)	0.182	
HTN		5,002	0.85	(0.74- 0.97)	0.018	*
DM		5,002	1.83	(1.61- 2.08)	<0.001	*
HLP		5,002	0.78	(0.68- 0.90)	<0.001	*
COPD		5,002	1.18	(0.97- 1.43)	0.099	
CKD		5,002	1.46	(1.24- 1.71)	<0.001	*
Gout		5,002	1.08	(0.84- 1.38)	0.552	
history_TIAstroke		5,002	1.05	(0.85- 1.31)	0.635	
HFEF_group		5,002	1.20	(1.08- 1.33)	<0.001	*
PAOD		5,002	1.53	(1.29- 1.82)	<0.001	*
CAD		5,002	0.92	(0.80- 1.05)	0.204	
EF_Teich_		5,002	0.98	(0.97- 0.98)	<0.001	*
Lvmass		5,002	1.00	(1.00- 1.00)	0.677	
LA		5,002	0.78	(0.72- 0.83)	<0.001	*

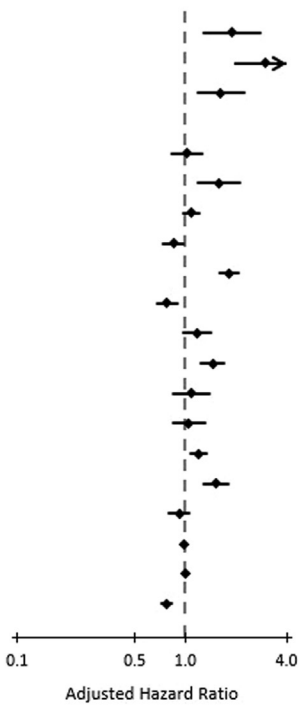

Figure 3 (*continued*).

Our Taiwanese cohort's inclusion of echocardiographic data provides detailed perspective on symptomatic cases, although this focus on potentially more severe symptoms or detectable AF may introduce bias. Nevertheless, the comprehensive analysis made by echocardiographic data offers valuable insights. The single-center design of our study might also reflect a more severe disease profile compared to broader multicenter analyses. The CHA_2DS_2-VASc score alone may not fully encompass all risk factors influencing AF outcomes. By focusing specifically on CHA_2DS_2-VASc score factors, our approach differs from broader machine learning methods, such as random survival forest algorithms, which integrate a wider range of variables and handle high-dimensional data effectively.[14] This focused methodology simplifies analysis, reduces data input requirements, and maintains clinical relevance.

The Outcomes Registry for Better Informed Treatment of Atrial Fibrillation II (ORBIT-AF II) registry, a multicenter study in the United States, identified 4 clusters, with the "low comorbidity" cluster exhibiting the lowest risk of major adverse cardiovascular events.[15] Conversely, even the lowest-risk cluster in our Taiwanese cohort presented a significant burden of conditions such as reduced LVEF and enlarged left atrial diameter. Similarly, the Fushimi AF Registry in Japan categorized patients into 6 clusters, underscoring the role of AF subtype and left atrial size in cluster formation, rather than solely focusing on behavioral risk factors.[16–18] Our analysis, which identified 4 clusters, provides a more detailed stratification of conditions such as heart failure and CKD, with distinct outcomes for stroke, heart failure hospitalization, and cardiovascular mortality. While the Evaluating the Use of SR34006 Compared to Warfarin or Acenocoumarol in Patients With Atrial Fibrillation (AMADEUS) trial and Borealis-Atrial Fibrillation (BOREALIS-AF) trial found significance only in the highest-risk cluster after multivariable Cox model adjustments,[19] our analysis showed statistically significant differences across all 4 clusters. These differences remained significant even after incorporating CHA_2DS_2-VASc parameters and echocardiographic variables, enhancing the precision of stroke risk assessment.

Although age is a well-established risk factor for stroke, our analysis suggests that the CHA_2DS_2-VASc score may not effectively differentiate stroke risk among various patient clusters. Patients with HFpEF appear to have a higher stroke risk than do those with HFrEF. This finding aligns with other Taiwanese studies advocating for a closer examination of HFpEF within the CHA_2DS_2-VASc framework.[20] Despite being a CHA_2DS_2-VASc risk factor, heart failure's predictive accuracy, especially for specific subtypes, remains limited. Biomarkers such as N-terminal pro–brain natriuretic peptide have not significantly improved stroke prediction in AF, and the role of LVEF is still uncertain.[21,22] While LVEF is not an independent predictor of thromboembolism, anticoagulation is recommended for patients with concurrent heart failure and AF.[23] The similar stroke incidence in patients with HFpEF and HFrEF, coupled with HFpEF's association with subclinical cerebral infarction, suggests that stroke risk is influenced by heart failure type.[24–27] The presence of HFpEF with AF may indicate a complex interplay of inflammation and thromboembolic processes.[28,29]

While atherosclerosis is a recognized stroke risk factor, our results indicate that stroke risk in patients with AF, DM, CKD, or HFpEF may surpass that associated with atherosclerosis alone. This underscores the need to incorporate HFpEF and potentially CKD into risk assessments beyond the CHA_2DS_2-VASc score. For younger patients with AF and DM, vigilant monitoring for comorbidities such as CKD and HFpEF is essential, even if their CHA_2DS_2-VASc score does not meet current anticoagulation criteria. Addressing these comorbid conditions could significantly influence stroke risk, indicating a need for more proactive and individualized management. Identifying these risk clusters facilitates nuanced risk stratification and patient counseling, prioritizing comorbidities over age-related factors. Furthermore, adherence to the ABC pathway—a comprehensive framework for AF management—has shown varying effects across different AF clusters.[30] Cluster analysis of patients with AF undergoing cryoablation has identified age, persistent AF, and left atrial diameter as significant predictors of AF recurrence.[31] These insights suggest that treatment efficacy may differ among clusters, underscoring the need for tailored therapeutic approaches that consider specific cluster characteristics.

Limitations

Some limitations in our study should be addressed. First, we were unable to identify AF subtype (paroxysmal vs nonparoxysmal), smoking status, alcohol consumption, socioeconomic status, physical activity, and obstructive sleep apnea because of data constraints. Nevertheless, we estimated that approximately 48.2% of the population with AF in our sample had paroxysmal AF on the basis of the echocardiographic mitral early diastolic transmitral flow velocity/late diastolic transmitral flow velocity ratio. We also lacked data on ablation, international normalized ratio levels, and alcohol consumption, which prevented us from calculating the HAS-BLED score for this population. Second, as an observational cohort study, we did not intervene with additional anticoagulation once patients were assigned to high-risk clusters. We observed a noncompliance rate of approximately 23.5% with anticoagulant treatment guidelines in the highest-risk cluster (cluster II), which may introduce bias. Our study does not assess the effects of targeted treatments on different AF risk clusters. Future research should examine how changes in comorbidities over time affect stroke incidence and whether tailored interventions based on cluster characteristics can improve clinical outcomes.

Conclusion

Our study demonstrates that a data-driven algorithm can effectively identify distinct patient phenotypes within the

population with AF, independent of CHA_2DS_2-VASc risk scores. The algorithm reveals heterogeneous risk profiles across different clusters, each associated with varying risks of cardiovascular events. This highlights the potential of such algorithms to improve patient stratification and facilitate more individualized therapeutic strategies in clinical practice.

Acknowledgments

We thank the staff of the Department of Medical Research for providing clinical data from the National Taiwan University Hospital-integrated Medical Database (NTUH-iMD).

Funding Sources: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosures: The authors have no conflicts to disclose.

Authorship: All authors attest they meet the current ICMJE criteria for authorship.

Patient Consent: Patient consent was waived because of the retrospective nature of the study, which was based on de-identified data collection.

Ethics Statement: The research reported in this article adhered to the Helsinki Declaration. The Institutional Review Board of the National Taiwan University Hospital approved the research (202007138RIND and 202403114RINC).

Appendix

Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.hroo.2024.11.009.

References

1. Boriani G, Vitolo M, Diemberger I, et al. Optimizing indices of atrial fibrillation susceptibility and burden to evaluate atrial fibrillation severity, risk and outcomes. Cardiovasc Res 2021;117:1–21.
2. Nicholls M. The anticoagulation grey zone. Eur Heart J 2021;42:4412–4414.
3. Al'Aref SJ, Anchouche K, Singh G, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur Heart J 2019;40:1975–1986.
4. Quer G, Arnaout R, Henne M, Arnaout R. Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review. J Am Coll Cardiol 2021;77:300–313.
5. Feeny AK, Chung MK, Madabhushi A, et al. Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. Circ Arrhythm Electrophysiol 2020;13:e007952.
6. Kwak S, Lee Y, Ko T, et al. Unsupervised cluster analysis of patients with aortic stenosis reveals distinct population with different phenotypes and outcomes. Circ Cardiovasc Imaging 2020;13:e009707.
7. Ahmad T, Pencina MJ, Schulte PJ, et al. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. J Am Coll Cardiol 2014;64:1765–1774.
8. Lindman BR, Dávila-Román VG, Mann DL, et al. Cardiovascular phenotype in HFpEF patients with or without diabetes: a RELAX trial ancillary study. J Am Coll Cardiol 2014;64:541–549.
9. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation 2015;131:269–279.
10. Hsu JC, Yang YY, Chuang SL, Yu CC, Lin LY. Higher long-term visit-to-visit glycemic variability predicts new-onset atrial fibrillation in patients with diabetes mellitus. Cardiovasc Diabetol 2021;20:148.
11. Hsu JC, Yang YY, Chuang SL, Chung YW, Wang CH, Lin LY. Underweight is a major risk factor for atrial fibrillation in Asian people with type 2 diabetes mellitus. Cardiovasc Diabetol 2021;20:226.
12. Ultsch A, Lötsch J. Euclidean distance-optimized data transformation for cluster analysis in biomedical data (EDOtrans). BMC Bioinformatics 2022;23:233.
13. Tromp J, Tay WT, Ouwerkerk W, et al. Multimorbidity in patients with heart failure from 11 Asian regions: a prospective cohort study using the ASIAN-HF registry. PLoS Med 2018;15:e1002541.
14. Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. Biostatistics 2014;15:757–773.
15. Inohara T, Shrader P, Pieper K, et al. Association of atrial fibrillation clinical phenotypes with treatment patterns and outcomes: a multicenter registry study. JAMA Cardiol 2018;3:54–63.
16. Inohara T, Piccini JP, Mahaffey KW, et al. A cluster analysis of the Japanese multicenter outpatient registry of patients with atrial fibrillation. Am J Cardiol 2019;124:871–878.
17. Ogawa H, An Y, Nishi H, et al. Characteristics and clinical outcomes in atrial fibrillation patients classified using cluster analysis: the Fushimi AF Registry. Europace 2021;23:1369–1379.
18. Watanabe E, Inoue H, Atarashi H, et al. Clinical phenotypes of patients with nonvalvular atrial fibrillation as defined by a cluster analysis: a report from the J-RHYTHM registry. Int J Cardiol Heart Vasc 2021;37:100885.
19. Vitolo M, Proietti M, Shantsila A, Boriani G, Lip GYH. Clinical phenotype classification of atrial fibrillation patients using cluster analysis and associations with trial-adjudicated outcomes. Biomedicines 2021;9:843.
20. Cheng CC, Huang PS, Chen JJ, et al. Both HFpEF and HFmrEF should be included in calculating CHA_2DS_2-VASc score: a Taiwanese longitudinal cohort. Heart Rhythm 2024;21:1500–1506.
21. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC). Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021;42:373–498.
22. Rivera-Caravaca JM, Marín F, Vilchez JA, et al. Refining stroke and bleeding prediction in atrial fibrillation by adding consecutive biomarkers to clinical risk scores. Stroke 2019;50:1372–1379.
23. Sobue Y, Watanabe E, Lip GYH, et al. Thromboembolisms in atrial fibrillation and heart failure patients with a preserved ejection fraction (HFpEF) compared to those with a reduced ejection fraction (HFrEF). Heart Vessels 2018;33:403–412.
24. Abdul-Rahim AH, Perez AC, Fulton RL, et al. Risk of stroke in chronic heart failure patients without atrial fibrillation: analysis of the Controlled Rosuvastatin in Multinational Trial Heart Failure (CORONA) and the Gruppo Italiano per lo Studio della Sopravvivenza nell'Insufficienza Cardiaca-Heart Failure (GISSI-HF) trials. Circulation 2015;131:1486–1494. discussion 1494.
25. Abdul-Rahim AH, Perez AC, MacIsaac RL, et al. Risk of stroke in chronic heart failure patients with preserved ejection fraction, but without atrial fibrillation: analysis of the CHARM-Preserved and I-Preserve trials. Eur Heart J 2017;38:742–750.
26. Uhm JS, Kim J, Yu HT, et al. Stroke and systemic embolism in patients with atrial fibrillation and heart failure according to heart failure type. ESC Heart Fail 2021;8:1582–1589.
27. Cogswell RJ, Norby FL, Gottesman RF, et al. High prevalence of subclinical cerebral infarction in patients with heart failure with preserved ejection fraction. Eur J Heart Fail 2017;19:1303–1309.
28. Packer M. HFpEF is the substrate for stroke in obesity and diabetes independent of atrial fibrillation? JACC Heart Fail 2020;8:35–42.
29. Kotecha D, Lam CS, Van Veldhuisen DJ, Van Gelder IC, Voors AA, Rienstra M. Heart failure with preserved ejection fraction and atrial fibrillation: vicious twins. J Am Coll Cardiol 2016;68:2217–2228.
30. Krittayaphong R, Treewaree S, Wongtheptien W, Kaewkumdee P, Lip GYH. Clinical phenotype classification to predict risk and optimize the management of patients with atrial fibrillation using the Atrial Fibrillation Better Care (ABC) pathway: a report from the COOL-AF registry. QJM 2024;117:16–23.
31. Bisson A, Fawzy AM, El-Bouri W, et al. Clinical phenotypes and atrial fibrillation recurrences after catheter ablation: an unsupervised cluster analysis. Curr Probl Cardiol 2023;48:101732.