SUPPLEMENTARY MATERIAL

Section I

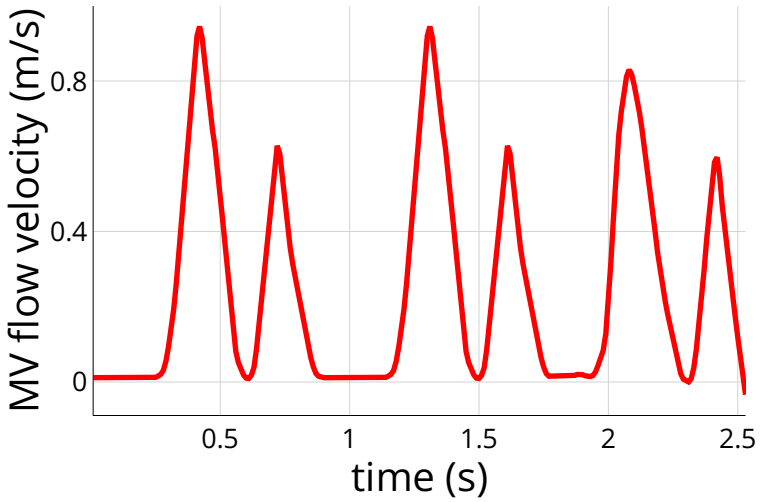
1. Left atrial segmentation and mesh processing

To minimize segmentation bias, an independent researcher not involved in the study segmented the LA geometries from the provided CT scans. LA blood-pool segmentation was carried out using Slicer 4.10.1[1], an open-source software tool, with semi-automatic region-growing techniques. Surface meshes were generated using the Marching Cubes algorithm applied to the binary segmentations, followed by smoothing with a Taubin filter (lambda = 0.5). Manual adjustments were made to correct self-intersecting faces and non-manifold edges. For CFD simulation preparation, the pulmonary veins were manually cut just before the first bifurcation from the LA cavity, perpendicular to the main flow direction, ensuring sufficient tube length without adding extensions. Meshlab 2016.121[2] and Autodesk Meshmixer 3.52[3] software tools were employed to build the surface meshes and apply the required mesh manipulations. Tetrahedral meshing was performed using the Delaunay algorithm in Gmsh 4.0.4[4], with an edge length of 1 mm. Final volumetric mesh optimization was conducted using Netgen[5], resulting in meshes containing 7 to 9 $\times 10^5$ elements, depending on the LA volume.

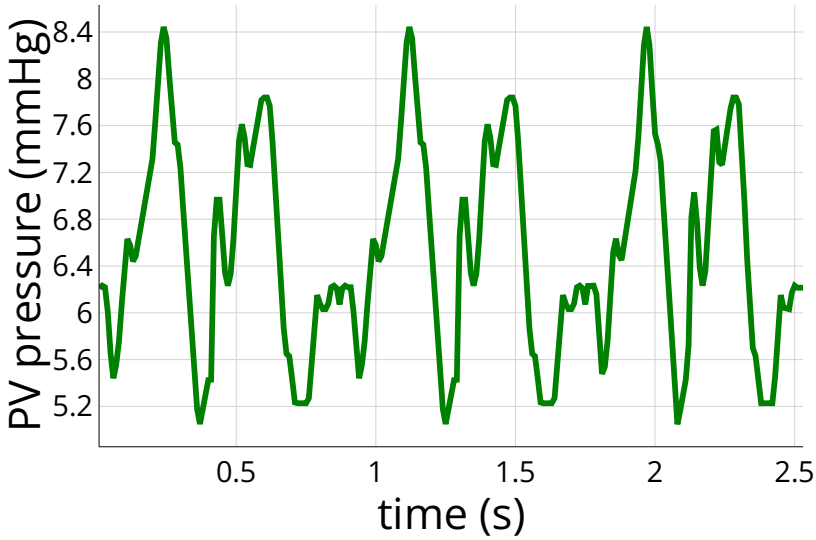
2. Boundary conditions computational fluid dynamic simulations

The following section illustrates the boundary conditions applied in the CFD simulations , which were derived from a representative AF patient within the cohort. Boundary conditions are shown for 3 cardiac beats (the span of the simulation).

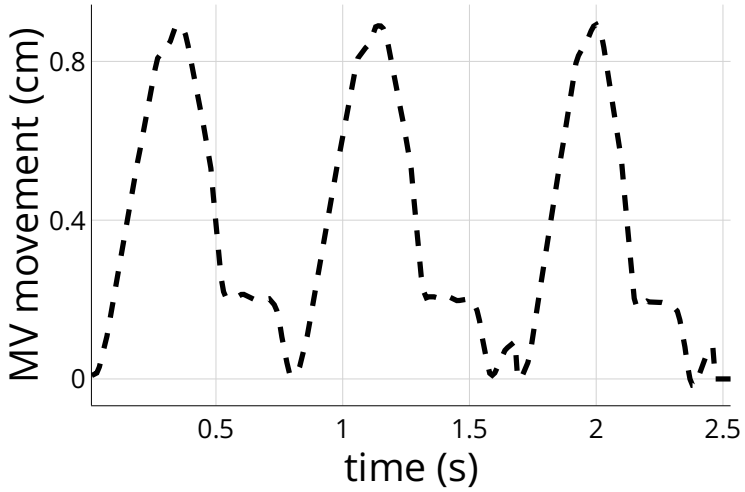
-
[1] https://www.slicer.org/
[2] https://www.meshlab.net/
[3] https://meshmixer.com/
[4] https://gmsh.info/
[5] https://ngsolve.org/



Supplementary Figure 1. Mitral valve flow curve used as the pressure outlet boundary condition. Mitral valve (MV) curve shown for 3 cardiac cycles.



Supplementary Figure 2. Pulmonary vein pressure profile used as the pressure outlet boundary condition. Pulmonary vein (PV) pressure profile shown for 3 cardiac cycles.



Supplementary Figure 3. Mitral valve movement curve used for the spring-based mesh movement. Longitudinal motion of the mitral valve (MV) annulus ring shown for 3 cardiac cycles.

3. Missing data imputation and feature selection

The following section explains the steps for missing data imputation and feature selection for unsupervised ML and clustering analysis. Missing values were identified in the clinical features (Table 1), which are considered missing at random (MAR). Variables that were measurable in all patients (i.e., PV-specific parameters such as PV ostium diameter and orientation excluded due to variability in the number of PVs) and that were not ratio of others (i.e., ratio ostium area/ LAA volume) were employed to construct the missing data imputation model (Table 2). For this, a Random Forest for mixed data types algorithm from the missForest package v1.4 in R was used. The missForest algorithm is a non-parametric method for imputing missing data. It employs an iterative approach, where a Random Forest (RF) model is initially trained on observed values to predict missing values, repeating this process iteratively. The use of a Random Forest allows the algorithm to handle mixed data types and capture complex, non-linear interactions. Additional details about the algorithm are provided in [1]. missForest was implemented with n_trees = 100 and took 5 iterations for convergence. Only the imputation of clinical variables with <40% missing values were considered, to lower bias. Supplementary Figure 1 compares the distributions of the raw and imputed variables used in the imputation model to assess imputation bias. The Jaccard similarity of the correlation matrices between the raw and imputed dataset (considering input variables to cluster analysis) was 85.48%. Categorical variables were excluded as input to the clustering analysis to lower feature weight imbalances in the MKL dimensionality reduction algorithm. The correlation matrix was computed on morphological, hemodynamic, and clinical features variables with Pearson's correlation exceeding 0.7 were excluded. Additionally, morphological features that were entirely computed from other included variables were not considered for clustering input due to redundancy. Furthermore, the hemodynamic index "particle

age" was excluded from machine learning input due to its lack of significant findings in prior studies [2].

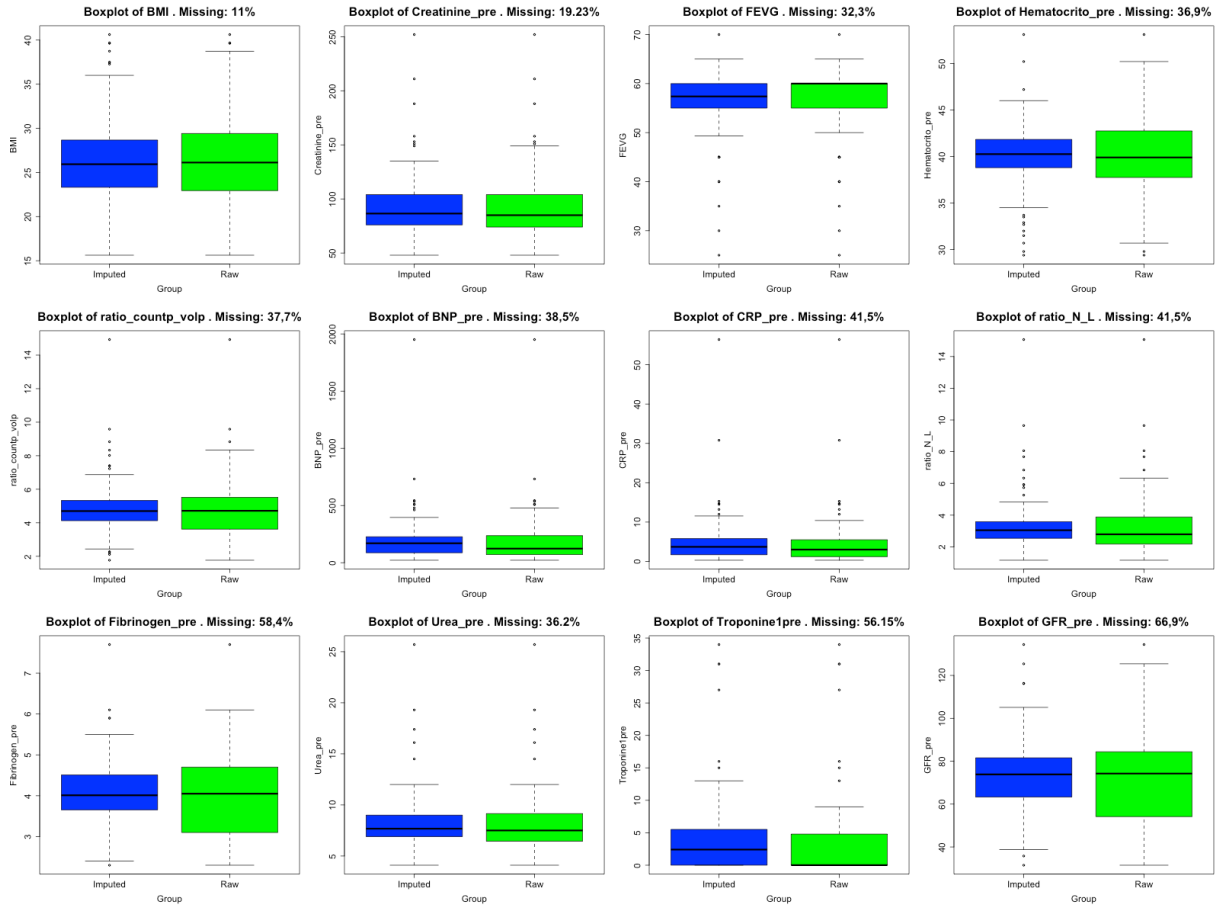
Supplementary Table 1. Missing values in clinical variables. BMI: body mass index; BNP: B-type natriuretic peptide; CRP: C-reactive protein; GFR: glomerular filtration rate; LVEF: left ventricular ejection fraction.

Clinical variables	% of missing values
Age	0
BMI	11
BNP	38.5
CRP	41.5
Fibrinogen	58.4
Creatinine	19.23
Troponine	56.15
Hematocrite	36.9
Ratio_neutrophyl_lymphocyte	41.5
GFR	66.9
ratio_platelet_volume_platelet_count	37.7
LVEF	32.3
Urea	36.2

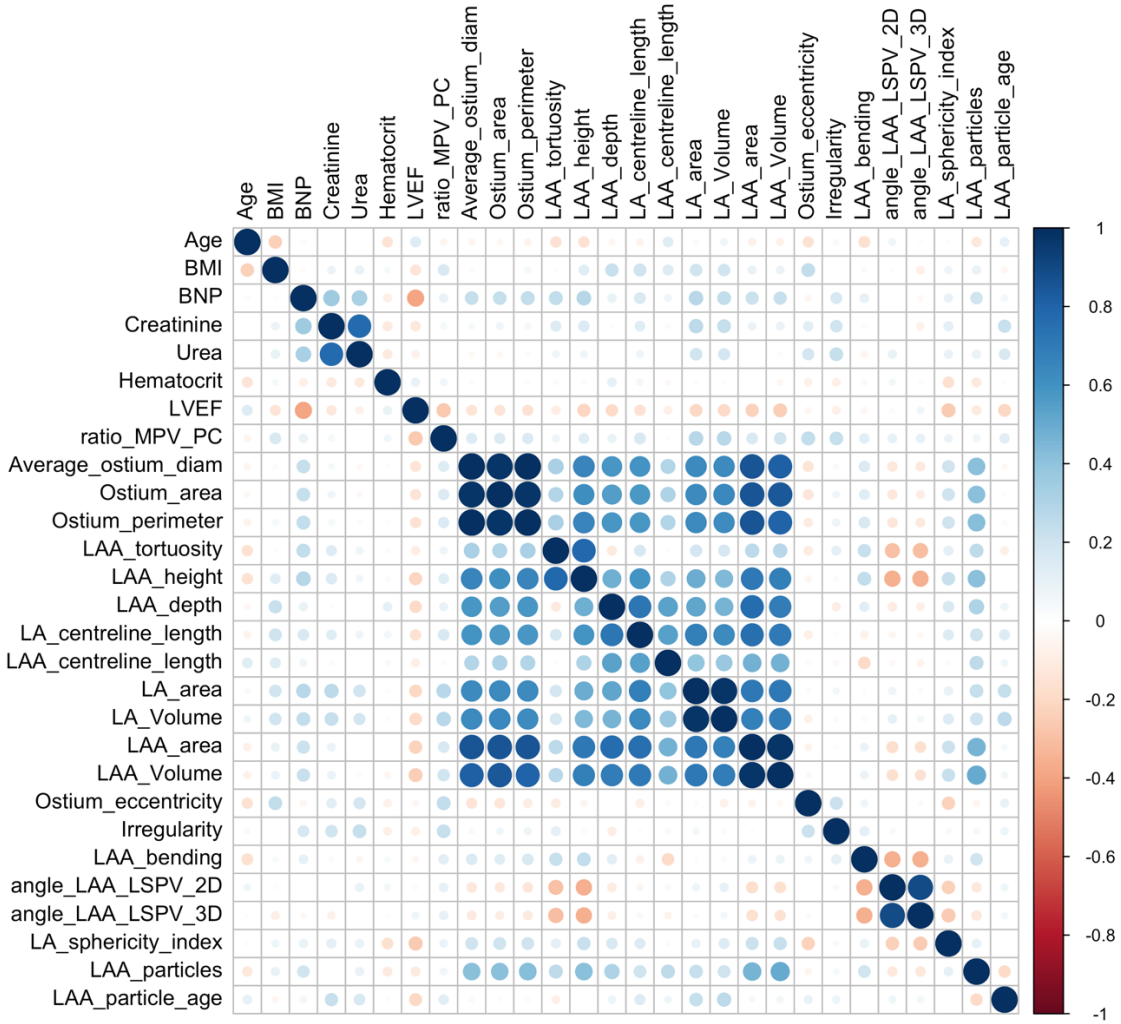
Supplementary Table 2. Overview of missing data imputation and feature selection. The clinical variables with missing data imputation and input to the machine learning are underlined. AF: atrial fibrillation; BMI: body mass index; BNP: B-type natriuretic peptide; CRP: C-reactive protein; GFR: glomerular filtration rate; HT: hypertension; LAA: left atrial appendage; LA: left atrium; LSPV: left superior pulmonary vein; LVEF: left ventricular ejection fraction; MPV: mean platelet volume; PC: platelet count.

	Variables extracted	Variables used for missing data imputation model	After exclusion of categorical variables	After exclusion of variables with >40% missing	After exclusion of highly correlated variables (r>0.7)/redundancies	Input of clustering analysis
	Clinical Variables - Continuous					
1	Age	X	X	X	X	X
2	BMI	X	X	X	X	X
3	BNP levels	X	X	X	X	X
4	CRP	X	X			
5	Fibrinogen	X	X			
6	Creatinine	X	X	X	X	X
7	Urea	X	X	X		
8	Troponine	X	X			
9	Hematocrit level	X	X	X	X	X
10	Neutrophyl/Lymphocite	X	X			
11	GFR	X	X			
12	LVEF	X	X	X	X	X
13	Ratio MPV/PC	X	X	X	X	X
	Clinical Variables - Categorical					
14	Gender	X				
15	Diabetes mellitus	X				
16	Cardiac insufficiency	X				
17	Vascular disease	X				
18	HT	X				
19	Dyslipidemia	X				
20	Type of AF	X				
21	HAS_BLED	X				
	Morphological Variables - Continuous					
22	Mean ostium	X	X	X		
23	Ostium area	X	X	X	X	X
24	Ostium perimeter	X	X	X		
25	LAA tortuosity	X	X	X	X	X
26	LAA height	X	X	X		
27	LAA depth	X	X	X		
28	LA centreline length	X	X	X		
29	LAA centreline length	X	X	X	X	X
30	LA area	X	X	X		
31	LA volume	X	X	X	X	X

32	LAA area	X	X	X		
33	LAA volume	X	X	X		
34	Ostium eccentricity	X	X	X		
35	Ostium irregularity	X	X	X	X	X
36	LAA bending	X	X	X	X	X
37	Angle lspv/laa alignment 2D	X	X	X	X	X
38	Angle lspv/laa alignment 3D	X	X	X		
39	LA sphericity index	X	X	X	X	X
	Hemodynamic variables					
40	Total LAA particles	X	X	X	X	X
41	Particle age	X	X	X	X	



Supplementary Figure 4. Comparison of imputed and raw clinical variable distributions. Boxplots showing raw distributions in green and imputed distributions in blue for different features.



Supplementary Figure 5. Correlation of morphological, hemodynamic and clinical variables. The figure shows the correlation matrix heatmap among variables. AF: atrial fibrillation; BMI: body mass index; BNP: B-type natriuretic peptide; CRP: C-reactive protein; GFR: glomerular filtration rate; HT: hypertension; LAA: left atrial appendage; LA: left atrium; LSPV: left superior pulmonary vein; LVEF: left ventricular ejection fraction; MPV: mean platelet volume; PC: platelet count.

4. SHAP feature importance analysis

For each cluster analysis, a Random Forest classifier (*RandomForestClassifier*, scikit-learn), with n_trees = 100, was trained with the cluster labels as outcomes to determine the most influential variables for cluster discrimination. The classifier was then analyzed to assess the relative importance of predictor attributes using the Shapley Additive Explanations (SHAP) method [3], implemented in Python with the 'shap' library[6]. For each cluster, SHAP ranks the attributes based on their relative importance in determining patient cluster membership. Bar plots of the average absolute SHAP value for each feature were obtained.

-

[6] https://pypi.org/project/shap/

5. Comparison with supervised models

The results of the unsupervised analysis were compared with a supervised approach to assess its association with thrombus formation history and enhance its interpretability. The seven most important variables influencing cluster formation were selected based on the SHAP feature importance analysis described in Section 4 of the main text. The selected variables include: the number of LAA fluid particles, LAA ostium area, LA volume, LAA centerline length, LAA tortuosity, BMI, and BNP. These features were used as input to a Random Forest (RF) classifier to evaluate their relationship with thrombus formation, using history of thrombus as the classification label. The dataset was split into training and testing subsets (80%/20%), and the RF model was trained with 100 estimators to enhance robustness.

6. Internal evaluation of model stability and generalization capability

MKL and clustering stability. To assess the robustness of our model, we performed a comparative analysis between the model trained on the full dataset and those trained on different data subsets. For this, the dataset was divided into 4 folds, with one fold used for testing and the remaining three for training. MKL was applied on the training subset for dimensionality reduction and K-means clustering for phenogroup definition. The holdout fold (25% of the dataset) was then projected into the low-dimensional space and assigned to the corresponding phenogroups.

K-means stability evaluation. Additionally, the stability of the K-means clustering configuration selected (3 clusters in a low-dimensional space of 2 dimensions) was evaluated with a non-parametric bootstrap-based approach. The 'bootcluster' package in R was employed, with scheme 1 selected, to assess the confidence around clustering from the original data based on bootstrap sample replications. To ensure robustness, 1000 bootstrap replicates were employed. A Jaccard-based estimate of stability was employed to compare the clusterings within each bootstrapping scheme, from which the average was computed. As recommended by the authors of the 'bootcluster' method [4], a stability threshold of 0.8 was used to determine whether the clustering solution could be considered reliable.

Internal validation. We evaluated the generalization ability of MKL and clustering model through an internal validation, where a portion of the data was used to train the model and fine-tune its parameters, while the remaining hold-out data was reserved to test the model's performance on previously unseen samples. For this, we performed MKL on a randomly selected training set (75%

of the database) to reduce the dimensionality of the input data and define phenogroups. The subjects from the validation set (comprising the remaining 25% of the database) were subsequently projected into this low-dimensional space to determine their respective phenogroups. We then conducted a comparative analysis of the phenogroups between the training and validation sets, focusing on the input ML parameters and the history of thrombus outcome.

7. Output space and phenogroup stability analysis

We evaluated the stability of the MKL output space following the inclusion of clinical variables. Specifically, we examined patient reclassifications, assessing whether the new phenogroup assignments maintained a similar distribution regarding the history of thrombus formation. For instance, patients originally classified in the phenogroup with the lowest proportion of thrombus history (Phenogroup 0) would again be assigned to a phenogroup with the lowest proportion of thrombus history, thus forming the new Phenogroup 0. Furthermore, we computed the k-Nearest Neighbors (KNN) (k = 10) of each patient in the output space generated with morphological and hemodynamic factors as input (output space 1), and in the output space generated by morphological, hemodynamic, and clinical factors as input (output space 2), and assessed their similarities by computing the following ratio:

$$ratio = \frac{nº\ of\ coinciding\ KNNs\ (output\ space 1,\ output\ space\ 2)}{k}$$

Finally, we analyzed the morphological, hemodynamic, and clinical characteristics of reclassified patients after the inclusion of clinical variables. For this, we compared the morphological, hemodynamic and clinical features of the reclassified patients with the phenogroups they left or entered.

Section II

1. Population baseline characteristics

Supplementary Table 3. Clinical baseline features. AF: atrial fibrillation; BMI: body mass index; BNP: B-type natriuretic peptide; CRP: C-reactive protein; CT: computed tomography; GFR: glomerular filtration rate; HF: heart failure; LAAO: left atrial appendage occlusion; LVEF: left ventricular ejection fraction; MPV: mean platelet volume; OAC: oral anticoagulation; PC: platelet count; TIA: transient ischemic attack.

Variable	Value (n = 130)
Age (years)	75 (70-80.75)
BMI (kg/m^2)	26.47 (5.08)
Female	48 (37%)
Non-paroxysmal AF	63 (56.8%) (n = 111)
HF (%)	15 (13.9%) (n = 108)
Hypertension (%)	84 (79.2%) (n = 106)
Diabetes type 2 (%)	26 (24.5%) (n = 106)
Vascular disease (%)	29 (27.4%) (n = 106)
Dyslipidemia (%)	65 (73.9%) (n = 88)
CHADS2VASC	3.97 (1.46)
HAS_Bled	3.15 (0.93)
History of stroke/TIA	54
Thrombus detected on pre-LAAO CT	6
History of peripheral embolism	1
History of bleeding	92 (90.2%) (n = 102)
History of bleeding under OAC	69 (75%) (n = 92)
Laboratory data	
CRP (mg/l)	3 (1.2-5.5) (n = 78)
BNP (pg/ml)	123.5 (71.5-234) (n = 80)
Hematocrit (%)	39.97 (4.34) (n = 83)
Creatinine (µmol/l)	85 (74-104) (n = 105)
GFR (ml/min)	74.14 (24.30) (n = 44)
Neutrophyl / Lymphocite	2.79 (2.2-3.88) (n = 76)
MPV (fL)	9.03 (1.15) (n = 80)
MPV / PC ratio	4.71 (3.62-5.51) (n = 80)
Troponin (ng/ml)	0.01 (0.01-4.80) (n = 57)
Fibrinogen (g/l)	4.02 (1.11) (n = 54)
Echocardiography data	
LVEF (%)	60 (55-60) (n = 88)9

Supplementary Table 4. Morphological baseline features. LA: left atrium; LAA: left atrial appendage; LCPV: left central pulmonary vein; LIPV: left inferior pulmonary vein; LSPV: left superior pulmonary vein; RCIPV: right central inferior pulmonary vein; RCSPV: right central superior pulmonary vein; RCPV: right central pulmonary vein; RIPV: right inferior pulmonary vein; RSPV: right superior pulmonary vein.

Variable	Value (n = 130)
Maximal ostium diameter (mm)	28.15 (24.01-33.18)
Minimal ostium diameter (mm)	20.62 (18.01-24.35)
Average ostium diameter (mm)	24.38 (21.27 – 28.60)
LAA tortuosity	0.39 (0.34-0.45)
LAA height (mm)	18.11 (5.54)
LAA depth (mm)	36.31 (7.18)
LA volume (ml)	157.39 (125.85 – 189.55)
LAA volume (ml)	10.43 (7.02-15.77)
Ratio LAA / LA volume	0.07 (0.05-0.09)
LA area (mm^2)	19903.98 (4352.37)
LAA area (mm^2)	3383.13 (2638.76 – 4389.98)
Ostium perimeter (mm)	78.23 (67.38 – 90.88)
Ostium area (mm^2)	462.14 (343.79-627.42)
Ostium irregularity	0.02 (0.01-0.04)
Ostium eccentricity	0.26 (0.10)
LA centreline length (mm)	99.15 (13.25)
LAA centreline length (mm)	45.38 (40.47-51.53)
LAA bending angle (deg)	126.38 (105.97-137.20)
LA sphericity index	0.76 (0.04)
Ratio ostium area / LAA volume	0.04 (0.04-0.05)
Number of lobes (%)	
1	44
2	48
>= 3	38
Shape & alignment label (%)	
Chicken-wing LAA	34
LAA/LSPV fully aligned	55
LAA/LSPV moderately aligned	11
LAA/LSPV not aligned	30
Pulmonary vein configuration	
Number of pulmonary veins	
3	4
4	69
5	41
6	13
7	4
2D angle LAA/LSPV alignment	32.48 (15.66-45.12)
LSPV average ostium diameter (mm)	17.86 (25.54-21.05) (n = 121)
LCPV average ostium diameter (mm)	22.94 (9.36) (n = 11)
LIPV average ostium diameter (mm)	16.10 (13.33-17.93) (n = 121)
RSPV average ostium diameter (mm)	18.14 (16.28-22.29) (n = 130)
RCSPV average ostium diameter (mm)	7.93 (6.58-9.85) (n = 19)
RCPV average ostium diameter (mm)	9.47 (2.80) (n = 44)
RCIPV average ostium diameter (mm)	7.73 (1.35) (n = 19)
RIPV average ostium diameter (mm)	16.68 (14.18-20.26) (n = 130)
LSPV orientation angle (deg)	124.52 (18.13) (n = 121)
LCPV orientation angle (deg)	106.32 (21.59) (n = 11)
LIPV orientation angle (deg)	107.80 (21.41) (n = 121)
RSPV orientation angle (deg)	63.64 (19.04) (n = 130)
RCSPV orientation angle (deg)	72.27 (23.55) (n = 19)
RCPV orientation angle (deg)	55.84 (23.47) (n = 44)
RCIPV orientation angle (deg)	29.12 (15.15) (n = 19)
RIPV orientation angle (deg)	31.30 (22.78 – 41.14) (n = 130)

Supplementary Table 5. In-silico hemodynamic baseline features. LAA: left atrial appendage.

Variable	Value
LAA total number of particles	164 (106.5-222.25)
LAA particle age (ms)	1.31 1.17-1.41)

2. Univariate analysis results

Supplementary Table 6. Baseline demographic and clinical characteristics of non-valvular atrial fibrillation (AF) patients with or without history of thrombus. BMI: body mass index; HF: heart failure; C-reactive protein; BNP: B-type natriuretic peptide; GFR: glomerular filtration rate; MPV: mean platelet volume; n: number of patients; PC: platelet count. Variables presented as mean (standard deviation), median (interquartile range) or n of patients (%). Significant p-values (< 0.05) in bold.

Variable	History of non-thrombus (n = 72)	History of thrombus (n = 58)	p-value
Age (years)	75 (69 - 79)	75.50 (71 - 81.75)	0.353
BMI (kg/m^2)	26.49 (5.15) (n = 64)	26.49 (5.03) (n = 51)	0.967
Female (%)	26 (36%)	22 (38%)	0.853
Non – paroxysmal AF (%)	33 (52%) (n = 64)	30 (64%) (n = 47)	0.248
HF (%)	9 (15%) (n = 60)	5 (11%) (n = 46)	0.592
Hypertension (%)	47 (78%) (n = 60)	37 (80%) (n = 46)	0.822
Diabetes type 2 (%)	14 (23%) (n = 60)	12 (26%) (n = 46)	0.823
Vascular disease (%)	18 (30%) (n = 60)	11 (24%) (n = 46)	0.517
Dyslipidemia (%)	30 (62%) (n = 48)	35 (88%) (n = 40)	0.015
Laboratory data			
CRP (mg/l)	3.20 (1.20 – 6.30) (n = 45)	2.50 (1.40 - 4.10) (n = 33)	0.404
BNP (pg/l)	123.50 (72.50 – 210.50) (n = 46)	128 (70.50 – 247.50) (n = 34)	0.888
Hematocrit (%)	40.07 (4.70) (n = 46)	39.84 (3.91) (n = 37)	0.812
Creatinine (µmol/l)	84 (70.50 – 114.50) (n = 59)	86.50 (75.25 – 97.00) (n = 46)	0.769
GFR (ml/min)	71.23 (27.16) (n = 26)	78.34 (19.41) (n = 18)	0.345
Neutrophyl / Lymphocite	2.60 (2.16 – 3.81) (n = 42)	3.01 (2.28 – 4.28) (n = 34)	0.361
MPV (fL)	8.81 (1.03) (n = 45)	9.3 (1.25) (n = 35)	0.06
Ratio MPV/platelet count	4.71 (3.62 – 5.47) (n = 45)	4.83 (3.83 – 5.69) (n = 35)	0.528
Troponin (ng/ml)	0.01 (0.01 – 5.65) (n = 30)	0.01 (0.01 – 4.80) (n = 27)	0.987
Fibrinogen (g/l)	4.09 (1.28) (n = 29)	3.94 (0.89) (n = 25)	0.635
Echocardiography data			
LVEF (%)	60 (55-60) (n = 48)	60 (55-60) (n = 40)	0.858

Supplementary Table 7. Morphological indices of non-valvular atrial fibrillation (AF) patients with or without history of thrombus. n: number of patients. diam.: diameter; LA: left atrium; LAA: left atrial appendage; LCPV: left central pulmonary vein; LIPV: left inferior pulmonary vein; LSPV: left superior pulmonary vein; RCPV: right central pulmonary vein; RCIPV: right central-inferior pulmonary vein; RCSPV: right central-superior pulmonary vein; RIPV: right inferior pulmonary vein; RSPV: right superior pulmonary vein. Significant p-values (< 0.05) in bold.

Variable	No history of thrombus (n = 72)	History of thrombus (n = 58)	p-value
Maximal ostium diameter (mm)	26.95 (23.63-32.08)	29.47 (25.71-34.14)	0.045
Minimal ostium diameter (mm)	19.42 (17.64-22.95)	22.52 (18.48-25.26)	0.009
Average ostium diameter (mm)	23.17(21.22-27.13)	26.45 (21.90-28.78)	0.024
LAA tortuosity	0.38(0.33-0.43)	0.40 (0.35-0.48)	0.052
LAA height (mm)	17.40 (5.22)	18.99 (5.85)	0.103
LAA depth (mm)	36.57 (6.46)	35.99 (8.04)	0.652
LA volume (ml)	152.85 (126.66-180.46)	164.93(125.03-214.60)	0.252
LAA volume	9.77 (6.43-14.52)	11.84 (9.11-16-27)	0.070
Ratio LAA/LA volume	0.07 (0.06-0.09)	0.06 (0.05-0.09)	0.159
LA area (mm^2)	19465.82 (3880.11)	20448 (4855)	0.202
LAA area (mm^2)	3605.64 (3010.33-4562.76)	33.18.97(2569.43-4223.37)	0.164
Ostium perimeter (mm^2)	74.27 (67.09-88.20)	85.10 (69.57-93.32)	0.023
Ostium area (mm^2)	412.08(338.72-565.77)	543.58(371.33-646.57)	0.020
Ostium irregularity	0.022(0.011-0.041)	0.019 (0.009-0.038)	0.541
Ostium eccentricity	0.27(0.10)	0.25 (0.10)	0.481
LA centerline length (mm)	77.12 (10.70)	77.42 (12.26)	0.880
LAA centerline length (mm)	49.06(42.13-56.19)	46.10 (39.17-61.38)	0.864
LAA bending angle (deg)	125.26 (104.41-136.73)	128.21 (110.14-136.93)	0.508
LA sphericity index	0.76 (0.04)	0.76 (0.04)	0.387
Ratio ostium area /LAA volume	0.04 (0.04-0.05)	0.04 (0.03-0.05)	0.703
Number of lobes (%)			
1	43 (n = 31)	22 (n = 13)	0.048
2	32 (n = 23)	43 (n = 25)	
>= 3	25 (n = 18)	34 (n = 20)	
Shape & alignment label (%)			
Chicken-wing LAA	31 (n = 22)	21 (n = 12)	0.057
Fully aligned LAA/LSPV	33 (n = 24)	53 (n = 31)	
Moderately aligned LAA/LSPV	7 (n = 5)	10 (n = 6)	
Not aligned LAA/LSPV	29 (n = 21)	16 (n = 9)	
Pulmonary vein configuration			
Number of pulmonary veins	4 (4-5)	4 (4-5)	0.951
2D LSPV/LAA alignment angle (deg)	32.21 (19.55-49.7)	33.16 (13.69-44.33)	0.628
LSPV average ostium diam. (mm)	18.00 (15.77-20.84) (n = 66)	17.15(15.29-22.02) (n = 55)	0.555
LCPV average ostium diam. (mm)	20.62 (10.07) (n = 8)	29.11 (1.95) (n = 3)	0.194
LIPV average ostium diam. (mm)	16.40 (13.43-17.64) (n = 66)	15.56 (13.28-18.15) (n = 55)	0.981
RSPV average ostium diam. (mm)	18.31 (16.24-22.41)	17.85 (16.50-21.98)	0.931
RCSPV average ostium diam. (mm)	9.02 (6.96-10.19) (n = 10)	7.00 (6.52-8.63) (n = 9)	0.549
RCPV average ostium diam. (mm)	9.21 (2.48) (n = 26)	9.84 (3.24) (n = 18)	0.466
RCIPV average ostium diam. (mm)	7.77 (1.59) (n = 10)	7.68 (1.14) (n = 9)	0.890
RIPV average ostium diam. (mm)	17.40 (14.94-20.60)	16.16 (13.49-18.98)	0.119
LSPV orientation angle (deg)	123.11 (19.88) (n = 66)	126.21 (15.79) (n = 55)	0.351
LCPV orientation angle (deg)	104.10 (14.89) (n = 8)	112.24 (38.49) (n = 3)	0.604
LIPV orientation angle (deg)	109.74 (20.96) (n = 66)	105.48 (21.90) (n = 55)	0.277
RSPV orientation angle (deg)	64.54 (18.41)	62.53 (19.90)	0.553
RCSPV orientation angle (deg)	75.52 (21.35) (n = 10)	68.44 (26.52) (n = 9)	0.516
RCPV orientation angle (deg)	57.58 (24.48) (n = 26)	53.34 (22.39) (n = 18)	0.562
RCIPV orientation angle (deg)	30.61 (17.98) (n = 10)	27.46 (12.11) (n = 9)	0.664
RIPV orientation angle (deg)	32.73 (22.67-43.09)	30.22 (23.05-38.22)	0.466

Supplementary Table 8. Hemodynamic indices of non-valvular atrial fibrillation (AF) patients with or with- out history of thrombus. LAA, left atrial appendage; n: number of patients. Variables presented as mean (standard deviation), median (interquartile range) or n of patients (%). Significant p-values (< 0.05) in bold.

Variable	No history of thrombus (n = 72)	History of thrombus (n = 58)	p-value
LAA total number of particles	130.0 (83.0 – 180.3)	207.5 (157.5 – 253.8)	<0.01
LAA particle age (ms)	1.265 (1.168-1.400)	1.325 (1.178-1.425)	0.452

3. Cluster analysis results

1. Morphological factors as input

Table 9 shows the ML training and additional descriptive variables characterizing phenogroups obtained with solely morphological features. Patient stratification driven solely by LA morphological features led to three phenogroups with significant differences in the total number of in-silico fluid LAA. Phenogroup 2 patients (n = 44), showing indications of high LAA stasis by exhibiting the largest number of LAA particles, were characterized by large atrial dimensions (ostium area, LA and LAA volume, ratio LAA/LA volume), high LAA complexity (LAA tortuosity and LAA centreline length) and the highest proportion of non-paroxysmal AF cases (74%). Interestingly, after adjusting the ostium area by LAA volume (ratio ostium area / LAA volume), Phenogroup 2 patients did not exhibit the largest ostium dimensions among the different clusters. Regarding pulmonary vein configuration, significantly high LAA/LSPV alignment (including and excluding CW morphologies), as well as increased LSPV and RSPV ostium diameters, were observed. On the other hand, Phenogroup 0 patients (n = 56) presented the lowest number of LAA particles, coupled with the lowest atrial dimensions, LAA complexity, and low LAA/LSPV alignment. Phenogroup 1 patients (n = 30) presented lower number of particles than Phenogroup 2, as well as a significantly larger ostium area adjusted by LAA volume.

Supplementary Table 9. ML training and inference variables characterizing phenogroups obtained with morphological variables. AF: atrial fibrillation; LA: left atrium; LAA: left atrial appendage; LSPV: left superior pulmonary vein; RSPV: right superior pulmonary vein.

Variables	Phenogroup 0 (n = 56)	Phenogroup 1 (n = 30)	Phenogroup 2 (n = 44)	p-value
Input ML variables				
Ostium area (mm^2)	374.6 (292.3 - 425.6) ‡ ⋄	526.6 (326.2 - 606.0) † ⋄	625.5 (530.5 - 721.5) † ‡	<0.01
LA Volume (ml)	136.4 (102.7 -167.7) ⋄	137.5 (126.6 - 185.0)⋄	182.1 (166.8 - 213.2) † ‡	<0.01
LAA tortuosity	0.37 (0.32 - 0.40) ⋄	0.36 (0.34 - 0.39) ⋄	0.45 (0.41 - 0.48) † ‡	<0.01
LAA centerline length (mm)	40.5 (34.8 - 46.5) ⋄	42.9 (41.9 – 44.6) ⋄	52.7 (49.7 - 55.6) † ‡	<0.01
2D LSPV/LAA alignment angle (deg)	37 (27 - 58) ⋄	30 (14 - 44)	25 (10 - 36) †	<0.01
Ostium irregularity	0.02 (0.01 – 0.03)	0.03 (0.02 - 0.05)	0.02 (0.01 – 0.04)	0.059
LA sphericity index	0.76 (0.73 - 0.78)	0.76 (0.72 - 0.79)	0.77 (0.75 - 0.80)	0.16
LAA bending angle (deg)	127 (96 - 138)	126 (113 - 145)	126 (115 - 131)	0.40
Descriptive variables - clinical				

Non-paroxysmal AF (%)	38% (n = 50) ‡ ⋄	69% (n = 26) †	74% (n = 35) †	<0.01
Descriptive variables - morphological				
Maximal ostium diameter (mm)	24.98 (22.23 - 27.55) ⋄	28.39 (23.03 - 32.16)	33.02 (30.19 - 35.77) †	<0.01
Minimal ostium diameter (mm)	18.45 (16.45 - 20.30) ‡ ⋄	22.57 (18.09 - 24.82) † ⋄	33.02 (30.19 - 35.77) †	<0.01
Average ostium diameter (mm)	21.76 (19.59 - 23.42) ‡ ⋄	26.16 (20.74 - 28.32) † ⋄	23.34 (21.12 - 27.10) † ‡	<0.01
Ostium perimeter (mm)	70.20 (61.97 - 75.18) ‡ ⋄	84.68 (64.65 - 90.37) † ⋄	28.53 (26.22 - 30.59) † ‡	<0.01
LAA height (mm)	15.1 (13.0 – 17.4) ⋄	15.5 (13.7 – 19.9) ⋄	90.29 (84.66 - 97.82) † ‡	<0.01
LAA depth (mm)	32.9 (28.7 – 37.4) ⋄	34.4 (31.8 – 38.3) ⋄	22.71 (20.9 – 24.7) † ‡	<0.01
LAA volume (ml)	8.71 (5.80 – 10.49) ⋄	9.30 (6.54 – 12.50) ⋄	40.9 (36.9 – 43.9) † ‡	<0.01
LA centerline length (mm)	91.3 (84.1 - 99.6) ⋄	95.1 (90.0 - 99.3) ⋄	107.39 (102.5 - 113) † ‡	<0.01
Ratio LAA / LA volume	0.06 (0.05 - 0.08) ⋄	0.06 (0.05 - 0.07) ⋄	0.08 (0.07-0.10) † ‡	<0.01
Ratio ostium area / LAA volume	0.04 (0.04 - 0.06)	0.05 (0.04 - 0.06) ⋄	0.04 (0.04 - 0.05) ‡	0.019
LSPV average ostium diam. (mm)	17.26 (15.16-22.72) ⋄	16.69 (15.11 - 21.16) ⋄	20.11 (17.57 - 22.28) † ‡	<0.01
RSPV average ostium diam. (mm)	18.26 (16.20 - 21.93)	17.58 (16.04 - 18.50) ⋄	21.11 (17.04 - 24.65) ‡	0.041
Shape & alignment label (%)	⋄		†	<0.01
LAA/LSPV not aligned	29	20	18	
LAA/LSPV moderately aligned	16	3	2	
LAA/LSPV fully aligned	23 ‡ ⋄	53 ‡	59 †	
Chicken-wing	32	23	20	
Descriptive variables - hemodynamic				
LAA total number of particles	149 (81 – 185) ⋄	170 (130.5 – 212.3)	192.5 (128.75 – 267) †	0.03

† : Statistically different from cluster 0 (p-value < 0.05)
‡ : Statistically different from cluster 1 (p-value < 0.05)
⋄ : Statistically different from cluster 2 (p-value < 0.05)

2. Morphological and in-silico hemodynamic factors as input

Table 10 displays the ML training and additional descriptive variables characterizing phenogroups obtained with morphological and in-silico hemodynamic characteristics. After the inclusion of the number of in-silico LAA, three phenogroups were obtained with significant differences in the number of cases with history of thrombus formation. The highest prevalence of history of thrombus formation (65%) was found in Phenogroup 2 (n = 31), which was characterized by the highest number of LAA particles, atrial and ostium dimensions, LAA complexity and LAA/LSPV alignment. Moreover, significant differences were observed in the bending angle of the LAA main lobe, being higher in Phenogroup 2. Clinically, significant differences were observed in the BNP levels (highest in Phenogroup 2, median value = 231 pg/ml) and LVEF (lowest in Phenogroup 2, median value = 55%). Also, a higher proportion of non-paroxysmal AF cases was observed in Phenogroup 2 (79%). On the other hand, Phenogroup 0 (n = 64) had the lowest history of thrombus formation (36%, p-value = 0.037 with respect to Phenogroup 2), and presented with the lowest number of LAA particles, atrial dimensions, LAA complexity and LAA/LSPV alignment with statistical significance. After adjusting the ostium area to LAA volume, Phenogroup 0 patients showed the highest ostium dimensions. Furthermore, lower BNP levels and higher LVEF were observed compared with Phenogroup 2. Phenogroup 1 patients (n = 35, 43% history of thrombus) presented larger atrial dimensions, LAA/LSPV alignment, LSPV and LIPV ostium diameters, and steeper LIPV entry angle compared to Phenogroup 0, coupled with a higher number of LAA particles, although this number was significantly lower compared to Phenogroup 2.

Supplementary Table 10. ML training and inference variables characterizing phenogroups obtained with morphological and in-silico hemodynamic variables. AF: atrial fibrillation; BMI: body mass index; BNP: B-type natriuretic peptide; LA: left atrium; LAA: left atrial appendage; LIPV: left inferior pulmonary vein; LSPV: left superior pulmonary vein; LVEF: left ventricular ejection fraction.

Variables	Phenogroup 0 (n = 56)	Phenogroup 1 (n = 30)	Phenogroup 2 (n = 44)	p-value
Input ML variables				
Ostium area (mm^2)	345.1 (283.6 - 395.7) ‡ ⋄	526.2 (440.5 - 568.7) † ⋄	676.3 (609.5 – 757.0) † ‡	<0.01
LA Volume (ml)	129.60 (107.58 - 154.62) ‡ ⋄	173.16 (165.41 - 192.26) †	189.33 (151.07 - 216.08) †	<0.01
LAA tortuosity	0.36 (0.32 - 0.40) ‡ ⋄	0.39 (0.37 - 0.46) †	0.44 (0.41 - 0.48) †	<0.01
LAA centerline length (mm)	40.8 (34.5 - 45.4) ‡ ⋄	49.9 (46.7 – 54.2) †	51.3 (44.0 - 54.3) †	<0.01
2D LSPV/LAA alignment angle (deg)	37 (24 - 57) ⋄	26.4 (13 - 44)	29 (10 - 37) †	<0.01
Ostium irregularity	0.02 (0.02 – 0.04)	0.02 (0.01 - 0.03)	0.03 (0.01 – 0.06)	0.23
LA sphericity index	0.76 (0.72 - 0.78)	0.76 (0.74 - 0.78)	0.77 (0.74 - 0.81)	0.11
LAA bending angle (deg)	121 (96 - 136) ⋄	128 (112 - 132)	132 (120 - 146) †	0.046
LAA total number of particles	121 (80.75 - 173.5) ‡ ⋄	157 (130 - 278.5) †	214 (194 - 258) †	<0.01
Descriptive variables - clinical				
BMI (kg/m^2)	24.9 (23.2 - 27.7) (n = 56) ‡	28.4 (23.1 - 33.2) (n = 31) †	25.9 (22.7 - 28.7) (n = 28)	0.03
BNP(pg/ml)	83 (54 - 156) (n = 45) ‡ ⋄	207 (131 - 267) (n = 20) †	231 (122 - 361) (n = 15) †	<0.01
LVEF (%)	60 (55 - 60) (n = 44) ⋄	60 (53.75 - 60) (n = 20)	55 (45 - 60) (n = 24)†	<0.01
Heart failure (%)	0% (n = 55) ‡ ⋄	22% (n = 27) †	35% (n = 26) †	<0.01
Dyslipidemia (%)	61% (n = 44)	90% (n = 20)	83% (n = 24)	0.032
Non-paroxysmal AF (%)	42% (n = 55) ⋄	64% (n = 28)	79% (n = 28) †	<0.01
Descriptive variables - morphological				
Maximal ostium diameter (mm)	24.05 (21.64 - 27.11) ‡ ⋄	30.18 (27.32 - 32.25) † ⋄	34.73 (32.16 - 36.25) † ‡	<0.01
Minimal ostium diameter (mm)	18.23 (16.17 - 19.43) ‡ ⋄	22.20 (20.39 - 23.10) †	25.13 (23.34 - 27.25) †	<0.01
Average ostium diameter (mm)	21.31 (19.13 - 22.92) ‡ ⋄	26.02 (23.95 - 27.39) † ⋄	29.47 (28.41 - 31.33) † ‡	<0.01
Ostium perimeter (mm)	67.45 (61.32 - 73.98) ‡ ⋄	83.08 (76.77 - 88.29) † ⋄	96.04 (90.29 - 99.76) † ‡	<0.01
LAA height (mm)	14.6 (12.4 – 17.1) ‡ ⋄	20.7 (17.0 – 23.0) †	22.8 (19.8 – 24.5) †	<0.01
LAA depth (mm)	32.9 (28.7 – 36.6) ‡ ⋄	40.6 (35.0 – 43.3) †	40.4 (34.4 – 44.0) †	<0.01
LAA volume (ml)	7.21 (5.68 – 9.81) ‡ ⋄	13.30 (11.12 – 16.59) †	15.99 (12.97 – 19.41) †	<0.01
LA centerline length (mm)	90.6 (85.6 - 96.7) ‡ ⋄	105.2 (99.8 - 110.7)†	105.5 (99.6 - 110.6) †	<0.01
Ratio ostium area / LAA volume	0.05 (0.04 - 0.06) ‡	0.04 (0.03 - 0.05) †	0.04 (0.04 - 0.05)	0.019
Ratio LAA/ LA volume	0.06 (0.05 - 0.07) ‡ ⋄	0.08 (0.06 - 0.10) †	0.08 (0.07 - 0.10) †	<0.01
LSPV average ostium diam. (mm)	16.78 (15.24-18.87) ‡	19.85 (17.42 - 21.28) †	18.15 (15.87 - 22.60)	<0.01
LIPV average ostium diam. (mm)	15.21 (12.65 - 16.97)	16.68 (14.64 - 18.00)	16.63 (13.75 - 18.12)	0.047
LIPV orientation angle	113 (99 - 124) ‡	103.68 (86.59 - 113) †	106.08 (89.57 - 121.30)	0.046
Shape & alignment label (%)		⋄	‡	<0.01
LAA/LSPV not aligned	30	20	13	
LAA/LSPV moderately aligned	12	9	0	
LAA/LSPV fully aligned	23 ‡ ⋄	46 † ⋄	77 † ‡	
Chicken-wing	34 ⋄	26	10 †	
Outcome variable				
History of thrombus (%)	36 ⋄	43	65 †	0.024

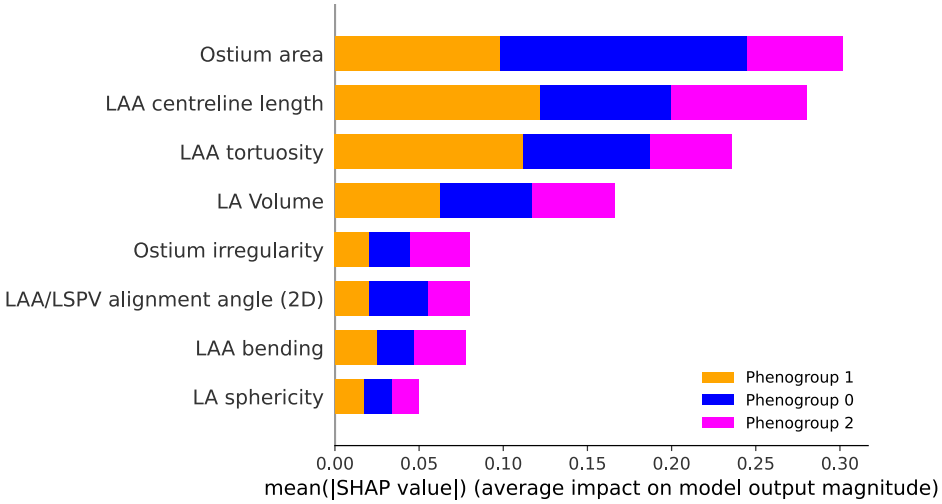
† : Statistically different from cluster 0 (p-value < 0.05)

‡ : Statistically different from cluster 1 (p-value < 0.05)

⋄ : Statistically different from cluster 2 (p-value < 0.05

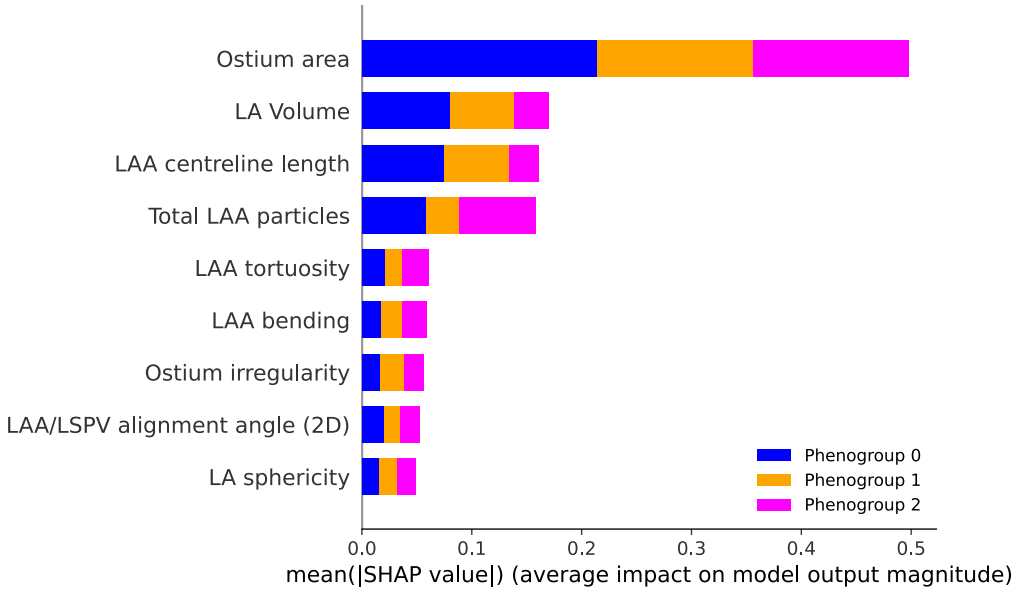
4. SHAP analysis bar plots

1. Clustering with morphological features as input



Supplementary Figure 6. SHAP-based feature importance using morphological inputs. Bar plot showing the average absolute SHAP values per feature, with taller bars indicating a greater contribution to predicting specific phenogroups. Features are ranked from highest to lowest importance. LA: left atrium; LAA: left atrial appendage; LSPV: left superior pulmonary vein.

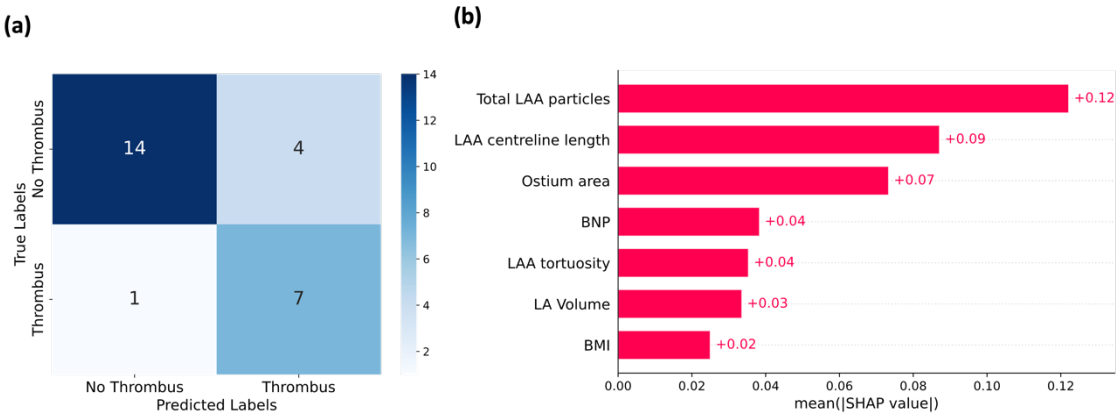
2. Clustering with morphological and hemodynamic as input



Supplementary Figure 7. SHAP-based feature importance using morphological and hemodynamic inputs. Bar plot showing the average absolute SHAP values per feature, with taller bars indicating a greater contribution to predicting specific phenogroups. Features are ranked from highest to lowest importance. LA: left atrium; LAA: left atrial appendage; LSPV: left superior pulmonary vein.

5. Phenogroup validation with supervised models

The model achieved a weighted average precision of 84%, recall of 81%, and F1-score of 81%, with the confusion matrix shown in Figure 8(a). SHAP analysis, using the 'shap' library in Python, quantified feature importance, with mean absolute SHAP values illustrated in Figure 8(b). The results highlight the association between the selected features and thrombus formation, identifying the number of LAA particles, LAA centerline length, and ostium area as the most important predictors, ranked 1st, 2nd, and 3rd, respectively.



Supplementary Figure 8. Results of phenogroup validation with supervised models. The figure presents (a) the confusion matrix of the supervised Random Forest model and (b) the average absolute SHAP values indicating the impact of specific features on predicting the 'history of thrombus' class label.

6. Analysis of intra-cluster history of thrombus differences

The following section presents the results of an intra-cluster analysis comparing the characteristics of history and non-history of thrombus patients within each phenogroup. The phenogroups obtained with morphological and in-silico hemodynamic factors as input, and those obtained with morphological, in-silico hemodynamic and clinical factors as input, were analyzed.

1. Morphological and in-silico hemodynamic factors as input

Table 11 displays the statistically significant intra-cluster differences between patients with and without history of thrombus in the phenogroup with the highest proportion of thrombus history cases (Phenogroup 2). Conversely, Table 12 presents these differences for the phenogroup with the lowest proportion of thrombus history cases (Phenogroup 0).

Supplementary Table 11. Analysis of history of thrombus differences in the phenogroup with highest proportion of history of thrombus (Phenogroup 2).

Variable	No history of thrombus (n = 11)	History of thrombus (n = 20)	p-value
LAA height (mm)	24.3 (22.8-26.0)	21.6 (19.6-23.1)	0.049
Age (years)	73 (62-75.5)	78 (73.25-81.5)	0.021

Supplementary Table 12. Analysis of history of thrombus differences in the phenogroup with lowest proportion of history of thrombus (Phenogroup 0).

Variable	No history of thrombus (n = 41)	History of thrombus (n = 23)	p-value
Total LAA particles	112.3 (55)	178.3 (96)	0.001
LAA centerline length (mm)	42.37 (36.13-46.54)	39.19 (33.23 – 40.86)	0.011
RSPV orientation angle (deg)	66.73 (13.39)	55.54 (17.44)	0.006
LAA depth (mm)	33.93 (30.24-37.42)	29.44 (27.43-34.87)	0.038
Shape & alignment label (%)			0.02
Not aligned	39	13	
Moderately aligned	7	22	
Fully aligned	15	39	
Chicken-wing	39	26	
Dyslipidemia (%)	50 (n = 30)	86 (n = 14)	0.0499

2. Clinical, morphological and in-silico hemodynamic factors as input

Table 13 displays the statistically significant intra-cluster differences between patients with and without history of thrombus in the phenogroup with the highest proportion of thrombus history cases after the inclusion of clinical variables (Phenogroup 2). Table 14 presents these differences for the phenogroup with the lowest proportion of thrombus history cases (Phenogroup 0).

Supplementary Table 13. Analysis of history of thrombus differences in phenogroup with highest proportion of history of thrombus (Phenogroup 2). Orient.: orientation.

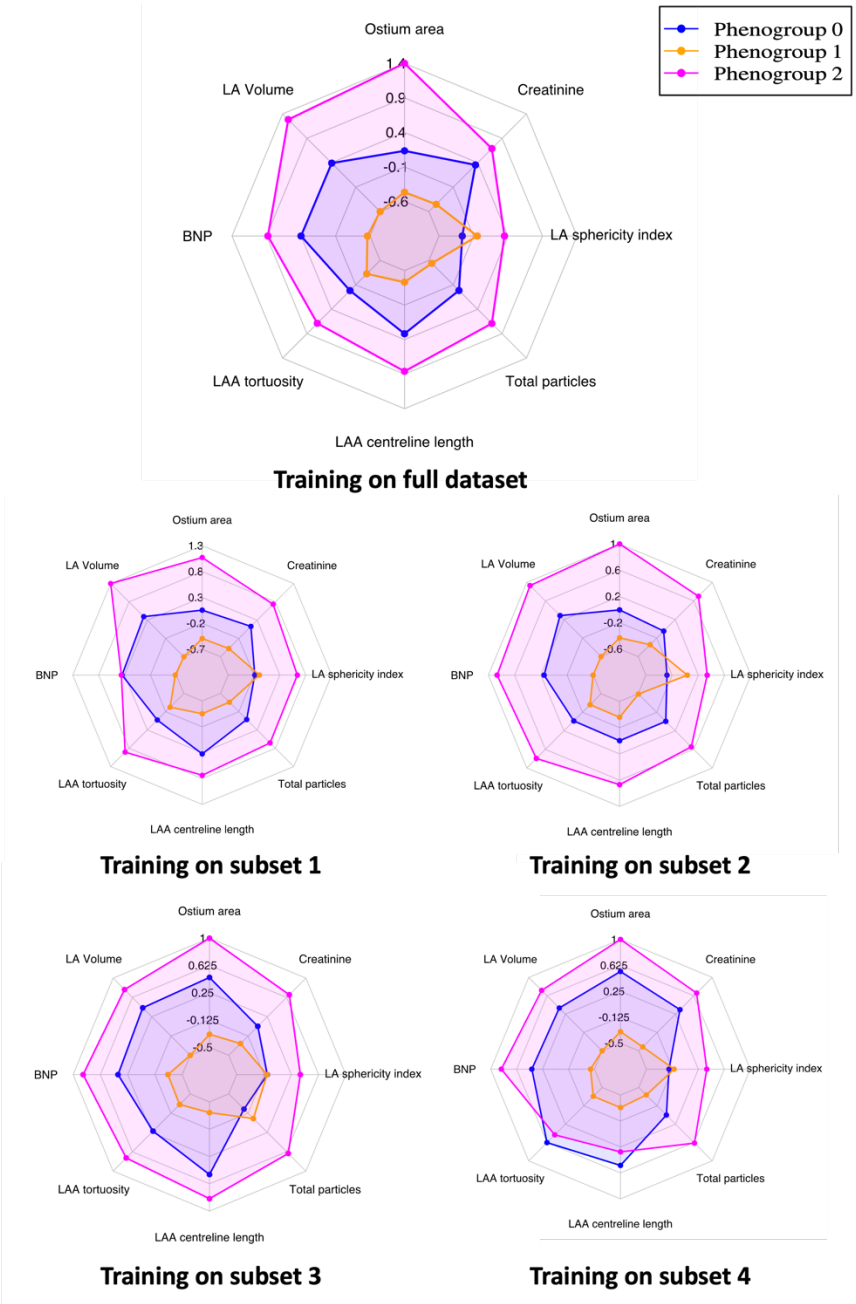
Variable	No history of thrombus (n = 9)	History of thrombus (n = 21)	p-value
LA Volume (ml)	175.40 (31.81)	230.69 (59.50)	0.014
BMI (kg/m^2)	23.62 (4.48)	27.21 (4.09)	0.041
Hematocrit (%)	38.30 (37.50-39.24)	39.80 (39.05-40.20)	0.025
LIPV orient. angle (deg)	117.15 (26.30) (n = 8)	94.02 (24.81) (n = 20)	0.038
Hypertension (%)	50 (n = 8)	92 (n = 13)	0.044

Supplementary Table 14. Analysis of history of thrombus differences in phenogroup with lowest proportion of history of thrombus (Phenogroup 0).

Variable	No history of thrombus (n = 28)	History of thrombus (n = 14)	p-value
MPV (fL)	8.68 (1.16) (n = 18)	9.69 (0.90) (n = 10)	0.025

7. Internal evaluation of model stability and generalization capability

1. Stability: MKL and clustering model



Supplementary Figure 9. Phenogroup characteristics from different training subsets. Radar charts displaying median normalized values for clinical, morphological, and hemodynamic features across phenogroups obtained from various training subsets or folds, including the projection of the unseen subset.

2. Internal validation

Supplementary Table 15. Morphological, hemodynamic and clinical differences between phenogroups of the training set (n = 97).

	Phenogroup 0 (n = 45)	Phenogroup 1 (n = 23)	Phenogroup 2 (n =30)	p-value
Ostium area (mm^2)	363.8 (291.9-449.0)	547.7 (510.1-648.2)	632.1 (445-736)	<0.01
LA Volume (ml)	124.6 (99.8-139.8)	174.6 (156.3-189.5)	201.2 (162.6-222.2)	<0.01
LAA tortuosity	0.37 (0.32-0.4)	0.44 (0.38-0.49)	0.43 (0.38-0.49)	<0.01
LAA bending angle (deg)	113 (95-132)	130 (121-136)	125 (116-142)	0.078
LAA centreline length (ml)	41.8 (37.2-45.4)	49.9 (46.0-52.6)	50.18 (42.2-55.1)	<0.01
LA sphericity index	0.76 (0.72-0.79)	0.75 (0.73-0.77)	0.78 (0.75-0.8)	0.069
2D LAA/LSPV angle (deg)	36 (14.6-57.4)	28.7 (16-53)	29.8 (13.7-35.5)	0.169
Ostium irregularity	0.018 (0.005-0.033)	0.019 (0.015-0.032)	0.035 (0.011-0.050)	0.17
Total particles	132 (80-179)	164 (115.5-268.5)	215.5(186-262)	<0.01
Age (years)	74 (68-82)	74 (71-78)	76.5 (69-80)	0.92
BMI (kg/m^2)	24.8 (23.7-26.6)	30.3 (26.3-32)	26.7 (23-28.3)	<0.01
BNP (pg/ml)	98 (67-140)	200.5 (82-218)	250.8 (191.2-360.7)	<0.01
LVEF (%)	59.5 (55-60)	57 (55-60)	56 (52-58)	0.032
Creatinine (µmol/l)	76.6 (70-88)	94 (83.5-105)	98 (87-112)	<0.01
Hematocrit (%)	40.3 (38.9-42.6)	41.4 (40.5-43.1)	39.2 (38.2-40.1)	<0.01
Ratio MPV/PC	4.22 (3.74-4.73)	5.47(4.88-5.89)	4.73 (4.41-4.94)	<0.01
History of thrombus (%)	42.22	43.48	63.33	0.167

Supplementary Table 16. Morphological, hemodynamic and clinical differences between phenogroups of the training set (n = 32).

	Phenogroup 0 (n = 18)	Phenogroup 1 (n = 6)	Phengroup 2 (n = 8)	p-value
Ostium area (mm^2)	410.5 (343.8-560.7)	455.6 (408.2-548.7)	529 (366.1-841.4)	0.49
LA Volume (ml)	147.3 (127.2-179.3)	162.6 (152.9-170.2)	172.7 (149.3-191.3)	0.58
LAA tortuosity	0.36 (0.29-0.39)	0.42 (0.40-0.48)	0.41 (0.36-0.46)	0.027

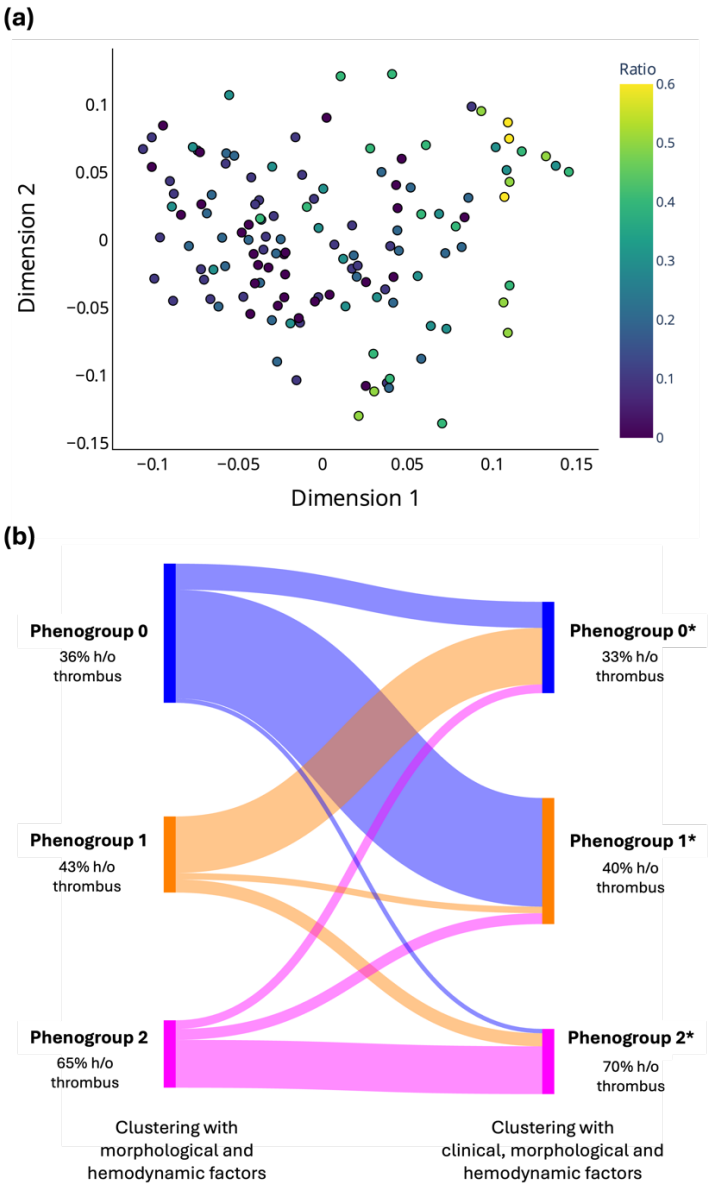
LAA bending (deg)	125 (98-143)	130 (12-135)	131 (126-139)	0.43
LAA centreline length (mm)	47.3 (41.6-55.3)	49.9 (46.4-52.1)	43.0 (40.7-51.6)	0.73
LA sphericity index	0.76 (0.75-0.76)	0.77 (0.77-0.79)	0.78 (0.76-0.82)	0.094
2D LAA/LSPV angle (deg)	37.5 (26.7-54.9)	27.3 (14.6-36.4)	40.3 (24.5-45.4)	0.5
Ostium irregularity	0.028 (0.007-0.04)	0.017 (0.017-0.022)	0.022 (0.008-0.035)	0.98
Total particles	128 (54-154)	232 (97-469)	200 (188-209)	0.011
Age (years)	77 (71-81)	73 (69-75)	79 (72-83)	0.33
BMI (kg/m^2)	24.6 (21.7-28.1)	28.4 (23.6-35.7)	27.4 (26.4-30.1)	0.20
BNP (pg/ml)	125 (68.75-159.1)	220.6 (187.8-252)	246.5 (212.6-284)	<0.01
LVEF (%)	58.94 (56.4-60)	54.8(54.1-56.4)	55 (53-60)	0.083
Creatinine (μmol/l)	82 (72.3-92)	88.3 (83.4-120.5)	109.5 (86.3-124.5)	0.30
Hematocrit (%)	40.9 (38.9-42)	38 (31.6-39.7)	39.6 (39.4-40.3)	0.126
Ratio MPV/PC	4.66 (4.06-5.1)	5.46 (4.47-5.86)	4.83 (4.5-5.2)	0.41
History of thrombus (%)	22.2	33.3	50	0.367

8. Output space and phenogroup stability analysis

Figure 10 shows the results of the output space stability analysis. Figure 10(a) illustrates the distribution of the proportion (i.e., ratio) of similar k nearest neighbours (k = 10) before and after the inclusion of clinical variables. Upon observation, the upper right-hand region of the output space (corresponding to Phenogroup 2) exhibited the highest neighbouring patient similarities, with the highest proportion of neighbouring patient similarities of 60%. In addition, the analysis of patient phenotype reclassifications (Figure 10(b)) showed a high stability in Phenogroup 2, as 71% of patients initially classified in the Phenogroup 2 (or the phenogroup with the highest proportion of a thrombus history) remained in that group after the inclusion of clinical variables. On the other hand, a high instability was observed in phenogroups 0 (low proportion of thrombus history) and 1 (intermediate proportion of thrombus history).

We further analyzed the characteristics of reclassified patient, following the inclusion of clinical features. After incorporating these features, 9 patients left (4 with no previous history of thrombus formation, see Table 17 for characteristic comparison) and 8 patients entered (6 with previous history of thrombus formation, see Tables 18 and 19 for characteristic comparison) the

phenogroup with highest number of patients with history of thrombus, leading to the improvement of history of thrombus proportion from 65% to 70%. The reclassified patients had 'borderline' (i.e., less similar to their group) morphological and in-silico hemodynamic characteristics and the main cause for reclassification, regarding their clinical characteristics, was their BNP levels, and to a lesser extent their creatinine values. On the other hand, two patients with previous history of thrombus were reclassified from a high to a low history of thrombus proportion phenogroup, due to a lower LA volume and BNP levels compared to Phenogroup 2 (Table 20). After the analysis of the remaining clinical variables, a high sum of CHADS2VASC score factors (score of 5, not including previous stroke/TIA) and a presence of dyslipidemia was observed in both patients.



Supplementary Figure 10. Low-dimensional output space stability analysis results. The figure presents (a) the output space (obtained using morphological, hemodynamic and clinical input features) colored by the ratio of coinciding k-nearest neighbors (k=10) before and after the inclusion of clinical variables, and (b) a Sankey diagram assessing phenogroup stability before and after the inclusion of clinical data.

Supplementary Table 17. Most significant clinical, morphological and hemodynamic characteristics of non-history of thrombus patients reclassified away from the high history of thrombus group (after inclusion of clinical variables) compared to patients within the high history of thrombus group (Phenogroup 2).

Variable	Non-thrombus patients reclassified (n = 4)	Phenogroup 2* (n = 30)	p-value
LA volume (ml)	176.33 (138.81-214.51)	211.64 (164.40-247.94)	0.22
LAA centerline length (mm)	47.58 (44.91-50.23)	52.73 (47.43-55.63)	0.16
Total number LAA particles	164.00 (133.75-190.25)	214.00 (193.00-246.75)	0.082
BNP levels (pg/ml)	163.05 (129.84-219.48)	250.81 (211.78-318.94)	0.087
Hematocrit levels (%)	42.08 (38.65-43.60)	39.35 (38.30-39.98)	0.18

Supplementary Table 18. Most significant clinical, morphological and hemodynamic characteristics of history of thrombus patients reclassified into the high history of thrombus group (after inclusion of clinical variables) compared to the phenogroup they left (middle history of thrombus group or Phenogroup 1).

Variable	Thrombus patients reclassified (n = 5)	Phenogroup 1 (n =30)	p-value
Ostium area (mm^2)	617.00 (552.54 - 835.09)	497.87 (436.82 - 556.32)	0.01
LA volume (ml)	256.20 (217.67 – 279.43)	171.55 (165.40 – 181. 26)	0.059
LA sphericity index	0.78 (0.77 – 0.8)	0.75 (0.74 – 0.77)	0.048
LAA tortuosity	0.46 (0.45 – 0.49)	0.39 (0.35 – 0.44)	0.048
BNP levels (pg/ml)	250.27 (211.71 – 264.28)	195.50 (143.66 – 243.60)	0.13

Supplementary Table 19. Most significant clinical, morphological and hemodynamic characteristics of history of thrombus patients reclassified into the high history of thrombus group (after inclusion of clinical variables) compared to the phenogroup they left (low history of thrombus group or Phenogroup 0). Due to low sample number no statistical analyses are performed.

Variable	Thrombus patients reclassified (n = 1)	Phenogroup 0 (n =63)
Ostium area (mm^2)	1797.28	342.47 (280.79 – 391.00)
LA volume (ml)	389.82	128.62 (106.74 – 152.41)
LAA centreline length (mm)	65	40.74 (35.28 – 44.65)
LAA total particles	407	121 (80.5 – 172.5)
BNP levels (pg/ml)	326.52	114 (68 – 169.22)
LVEF (%)	51.21	59.62 (55.15 – 60)
Creatinine (µmol/l)	122	85 (72 – 105.50)
MPV/platelet count	6.19	4.55 (3.96 – 5.10)

Supplementary Table 20. Case analysis of 2 patients with history of thrombus, initially assigned in the high history of thrombus group but reclassified to the low history of thrombus group, following the inclusion of clinical variables. BNP: B-type natriuretic peptide; TIA: transient ischemic attack. F: female.

Variable	Phenogroup 2 * (n = 30)	Patient 1	Patient 2
Ostium area (mm^2)	701.63 (598.81-784.30)	648.48	271.04
LA volume (ml)	211.64 (164.40-247.94)	118.95	126.46
LA sphericity	0.77 (0.76-0.8)	0.74	0.82
LAA tortuosity	0.44 (0.40-0.48)	0.51	0.36
Total particles	214 (193-246.75)	343	205
BNP levels (pg/ml)	250 (211.78-318.94)	189.71	144.62
Hematocrit level (%)	39.35 (38.30-39.98)	41.52	41.50
Baseline clinical characteristics (not included in cluster formation)			
Gender		F	F
Type of AF		non-paroxysmal	non-paroxysmal
CHADS2VASC (not including stroke/TIA)		5	5
Dyslipidemia		yes	yes
Hypertension		yes	yes
Heart Failure		no	yes
Vascular disease		yes	no

-

References

[1] Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics, 28(1), 112-118.

[2] Mill, J., Harrison, J., Saiz-Vivo, M., Albors, C., Morales, X., Olivares, A. L., ... & Camara, O. (2024). The role of the pulmonary veins on left atrial flow patterns and thrombus formation. *Scientific Reports*, *14*(1), 5860.

[3] Lundberg, S.: A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874 (2017)

[4] Bootstrapping estimates of stability for clusters, observations and model selection. Han Yu, Brian Chapman, Arianna DiFlorio, Ellen Eischen, David Gotz, Matthews Jacob and Rachael Hageman Blair.