ESC European Society of Cardiology

European Heart Journal - Digital Health (2025) **00**, 1–12
https://doi.org/10.1093/ehjdh/ztaf013

ORIGINAL ARTICLE

Unsupervised clustering of single-lead electrocardiograms associates with prevalent and incident heart failure in coronary artery disease

Josseline Madrid [1,*], William J. Young[2,3], Stefan van Duijvenboden[2,4,5], Michele Orini [5,6], Patricia B. Munroe[2,6], Julia Ramírez[1,7,2,†], and Ana Mincholé[1,7,†]

[1]Aragon Institute of Engineering Research, University of Zaragoza, Zaragoza, Spain; [2]William Harvey Research Institute, Barts and the London Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK; [3]Cardiac electrophysiology Department, St Bartholomew's Hospital, Barts Health NHS Trust, London, UK; [4]Nuffield Department of Population Health, University of Oxford, Oxford, UK; [5]Institute of Cardiovascular Science, University College London, London, UK; [6]National Institute of Health and Care Research, Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK; and [7]Centro de Investigación Biomédica en Red—Bioingeniería, Biomateriales y Nanomedicina, Zaragoza, Spain

Received 22 November 2024; revised 14 January 2025; accepted 22 January 2025; online publish-ahead-of-print 17 March 2025

Aims Clinical consequences of coronary artery disease (CAD) are varied [e.g. atrial fibrillation (AF) and heart failure (HF)], and current risk stratification tools are ineffective. We aimed to identify clusters of individuals with CAD exhibiting unique patterns on the electrocardiogram (ECG) in an unsupervised manner and assess their association with cardiovascular risk.

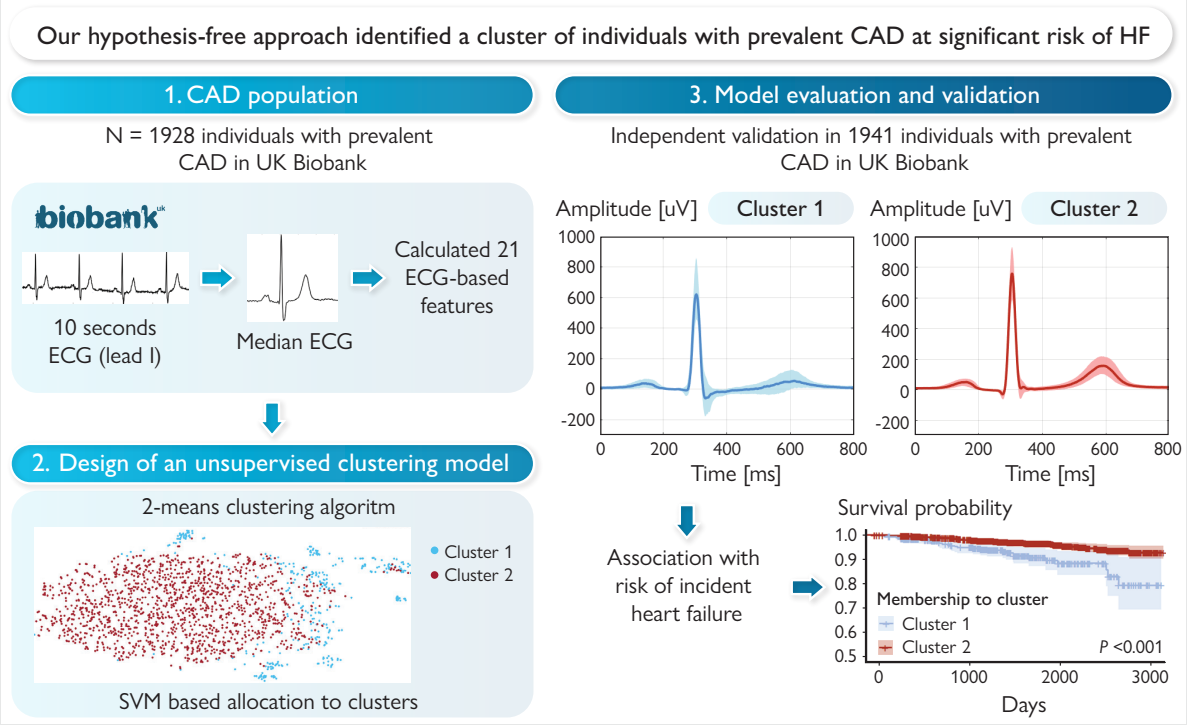
Methods and results Twenty-one ECG markers were derived from single-lead median-beat ECGs of 1928 individuals with CAD without a previous diagnosis of AF, HF, or ventricular arrhythmia (VA) from the imaging study in UK Biobank (CAD-IMG-UKB). An unsupervised clustering algorithm was used to group these markers into distinct clusters. We characterized each cluster according to their demographic and ECG characteristics, as well as their prevalent and incident risk of AF, HF, and VA (4-year median follow-up). Validation and association with prevalent diagnoses were performed in an independent cohort of 1644 individuals. The model identified two clusters within the CAD-IMG-UKB cohort. Cluster 1 ($n = 359$) exhibited prolonged QRS duration and QT intervals, along with greater morphological variations in QRS and T-waves, compared with Cluster 2 ($n = 1569$). Cluster 1, relative to Cluster 2, had a significantly higher risk of incident HF [hazard ratio (HR): 2.40, 95% confidence interval (CI): 1.51–3.83], confirmed by independent validation (HR: 1.77, CI: 1.31–2.41). It also showed a higher association with prevalent HF (odds ratio: 4.10, CI: 2.02–8.29), independent of clinical risk factors.

Conclusion Our approach identified a cluster of individuals with CAD sharing ECG characteristics indicating HF risk, holding significant implications for targeted treatment and prevention enabling accessible large-scale screening.

* Corresponding author. Tel: (+34) 876555460, Email: jmadrid@unizar.es

†The last two authors supervised this study equally.

Graphical Abstract



Keywords Unsupervised clustering model • Heart failure • Coronary artery disease • Electrocardiogram

Introduction

Coronary artery disease (CAD) is the main cause of cardiovascular mortality and morbidity worldwide.[1] Coronary artery disease manifests heterogeneously across individuals resulting in distinct adverse outcomes, which may be driven by cardiac ion channel remodelling in the context of acute or chronic ischaemia, or by myocardial scar tissue following infarction.[2,3] Heart failure (HF, incidence of 26%),[4] ventricular arrhythmias (VAs, 3–5%),[5] and atrial fibrillation (AF, 0.2–5%)[6] are common adverse outcomes in CAD, each requiring specific treatment strategies.[4,5,7] Considering the global burden of CAD, there is a need for accessible, cost-effective, and non-invasive tools that can be used for large-scale screening. Early risk stratification of patients is needed to optimize prevention strategies and ensure effective use of healthcare resources.

Electrocardiogram-based risk prediction strategies allow for affordable, non-invasive, and efficient ways to screen large populations.[8] In CAD, these commonly rely on supervised models derived from statistical[9–11] and machine learning techniques.[12–15] These models learn from prior labelled outcome data to make classifications in new unseen data supporting risk stratification in outcomes such as myocardial infarction.[14] However, supervised models may oversimplify the heterogeneous nature of CAD, highlighting the need for 'hypothesis-free' methodologies to uncover ECG patterns through unsupervised approaches.[16]

Unsupervised clustering models based on ECG data are able to differentiate clusters of individuals within a population by identifying unique ECG characteristics specific to each cluster.[16] These studies have demonstrated potential ability to risk stratify patients following acute coronary syndrome (ACS) at major risk of major adverse cardiovascular events[17] and in patients with hypertrophic cardiomyopathy at elevated arrhythmic risk.[18] However, the study in Syed and Guttag[17] focused on a post-ACS patient cohort, which inherently limits the scope to individuals with a recent coronary event. A broader population-level definition of CAD would encompass a wider range of patients, including those at earlier stages of CAD enhancing the generalizability of the findings. Furthermore, a longer follow-up period and focusing on specific outcomes would provide a more comprehensive and tailored long-term risk assessment potentially enabling the development of more accurate predictive models for diverse CAD populations. Two previous studies have performed unsupervised clustering using a multimodal approach for CAD, which included clinical data, ECG markers, and advanced imaging features.[19,20] These studies were able to identify distinct clinical and imaging profiles in CAD[19,20] and provide better risk stratification for all-cause mortality compared with stress total perfusion deficit alone[20]; however, their dependence on cardiac imaging means the models developed have limited utility in large-scale screening and healthcare centres without easy access to advanced imaging technologies.

Our hypothesis is that there are distinct clusters of individuals with CAD who share similar morphological ECG features, which are in turn associated with specific cardiovascular outcomes. Accordingly, this study aimed to identify unique clusters of individuals with CAD using advanced ECG features derived from single-lead 10 s ECG signals using unsupervised clustering and to investigate the association of these clusters with incident and prevalent AF, HF, or VA.

Materials and methods

Study population

The UK Biobank (UKB) study is a large-scale prospective cohort that contains half a million densely phenotyped individuals from the UK.[21] From the UKB study, two cohorts of individuals were included in this study: individuals from the imaging study (IMG-UKB, 4 years of median follow-up) that had a standard 10 s 12-lead ECG recorded in a supine position at rest and individuals who participated in an exercise stress test (EST-UKB, 13 years of median follow-up), during which a single-lead ECG (lead I) signal was recorded at rest in the first 15 s (pre-test). All individuals involved in the UKB study provided informed consent. The UKB study has approval from the North West Multi-Centre Research Ethics Committee.[21] This work was performed under UKB application number 8256. Additionally, available information included demographic data, cardiovascular risk factors, and health electronic records.

Electrocardiogram signal processing and characterization

Pre-processing of ECG signals involved high-pass filtering at 0.3 Hz, baseline wander correction through cubic splines interpolation,[22] high-frequency noise removal by low-pass filtering at 40 Hz, and exclusion of ectopic beats.[23] A median heartbeat ECG was derived, and a wavelet-based delineator[24] located ECG wave onset, peaks, and end timings. Characterization of ECG waveforms included standard and morphology-based indices. These ECG indices were considered due to prior evidence linking them to adverse cardiovascular outcomes.[11,25,26] Standard indices comprised RR-interval, QRS amplitude, QRS complex slopes,[27] QRS duration, ST segment amplitude, ST segment area, corrected QT interval (QTc), corrected T-peak-to-end interval (Tpec), T-wave polarity, and T/QRS amplitude ratio. QRS and T-wave morphologies were mathematically characterized using Hermite functions[28] to capture inter-subject variability. Additionally, morphologic variations in the T-waves compared with a normal reference were calculated through the T-wave morphologic variation (TMV) index.[11] Supplementary material online, *Table S1* provides details on the definition and calculation of each ECG feature. Supplementary material online, *Methods* provide further information on the methodology of the ECG signal processing and feature extraction. All processing of the ECG signals was performed using custom software in MATLAB R2022b (The MathWorks Inc.).

CAD-IMG-UKB cohort

We used a set of 2551 UKB individuals with prevalent CAD from the IMG-UKB cohort (CAD-IMG-UKB; *Figure 1*). Health electronic records were used to define CAD according to the World Health Organization International Classification of Diseases[29] ninth revision codes (ICD-9) as 410, 411, 414, and 429 and tenth revision codes (ICD-10) as I20 to I25 and the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures Fourth Revision codes (OPCS4) K40 to K42; K44 to K45; K49 to K50; and K75 (see Supplementary material online, *Table S2*). Individuals with a prior diagnosis of AF, HF, or VA before ECG acquisition ($n = 348$) were excluded. Atrial fibrillation was defined using ICD-10 codes I48 to I48.9 and OPCS4 codes K62.1 to K62.4. Heart failure was defined using ICD-10 codes I50, I50.0, I50.1, I13.0, I13.2, and I50.9 and OPCS4 codes K59.6, K61.7, and K60.7. Ventricular arrhythmia was defined using ICD-10 codes I46.0, I46.1, I46.9, I47.0, I47.2, and I49.0 and OPCS4 codes K57.6, K64.1, ×50.3 and ×50.4 (see Supplementary material online, *Table S2*). These diagnoses were defined within each study as self-reported or during admission to hospital.

Signals displaying significant electrode-driven artefacts, ectopic beats, and low correlation with their calculated median heartbeat ECG were excluded ($n = 275$). Ultimately, a total of 1928 individuals constituted the CAD-IMG-UKB cohort, which was used for the unsupervised clustering analysis (*Figure 1*).

CAD-EST-UKB cohort

From a total of 1941 independent individuals with prevalent CAD from the EST-UKB cohort (CAD-EST-UKB; *Figure 1*), two subsets were extracted: (i) To validate the association of the unsupervised model with incident AF, HF, and VA, the same exclusion criteria were used as described for CAD-IMG-UKB, leading to a total of 1644 individuals, and (ii) to find association with prevalent AF, HF, or VA diagnoses, we excluded ECGs of insufficient quality ($n = 7$) and individuals with AF, HF, or VA prior to a diagnosis of CAD ($n = 102$; *Figure 1*), leading to 1832 individuals.

Unsupervised identification of clusters

In the CAD-IMG-UKB cohort, to reduce multicollinearity, we removed those ECG features that exhibited a strong Spearman's correlation ($r^2 > 0.8$) with multiple other features. The remaining features were then standardized using *z*-score, centring the data around the mean and scaling it by the standard deviation.

We determined the optimal number of clusters using the 'Silhouette score' and the 'Gap statistic' for a range of 2–10 clusters (choosing the optimal number of clusters as the highest Gap statistic and Silhouette score; Supplementary material online, *Figure S1*).[13,27] Then, a *K*-means clustering algorithm was performed to derive the centroids. The squared Euclidean distance was used to evaluate the distance between neighbours. The clustering analysis was performed blindly to clinical data, only relying on the ECG features.

Allocation to clusters

A support vector machine model[30] was built to allocate individuals from the CAD-IMG-UKB and CAD-EST-UKB cohorts to each of the *K*-clusters identified with the *K*-means algorithm. We generated cluster plots to visualize the distance distribution of each individual to each cluster using the t-distributed Stochastic Neighbour Embedding algorithm (t-SNE), which embeds high-dimensional points into low dimensions[31] allowing for a graphical representation of the allocation into clusters of individuals in the CAD-EST-UKB cohort (*Figure 2*). Subsequently, a representative median heartbeat was calculated from each cluster to characterize the ECG morphology representing each cluster (*Figure 3*). t-SNE and median heartbeats plots were generated using MATLAB R2022b (The MathWorks Inc.).

Contribution of ventricular depolarization and repolarization features in the clustering process

To discern the individual contributions of ventricular depolarization and repolarization features to the *K*-clusters and to evaluate their specific impact on detecting and predicting HF, AF, or VA diagnoses in individuals with CAD, two analyses were conducted by repeating the main analyses using each group of features independently (the total number of features was 8 for the depolarization analysis and 11 for the repolarization analysis; Supplementary material online, *Table S1*).

Association analyses of clusters with incident and prevalent atrial fibrillation, heart failure, and ventricular arrhythmia risk

Features with less than 10% missing data were imputed using the R library 'mice' (multivariate imputation by chained equations).

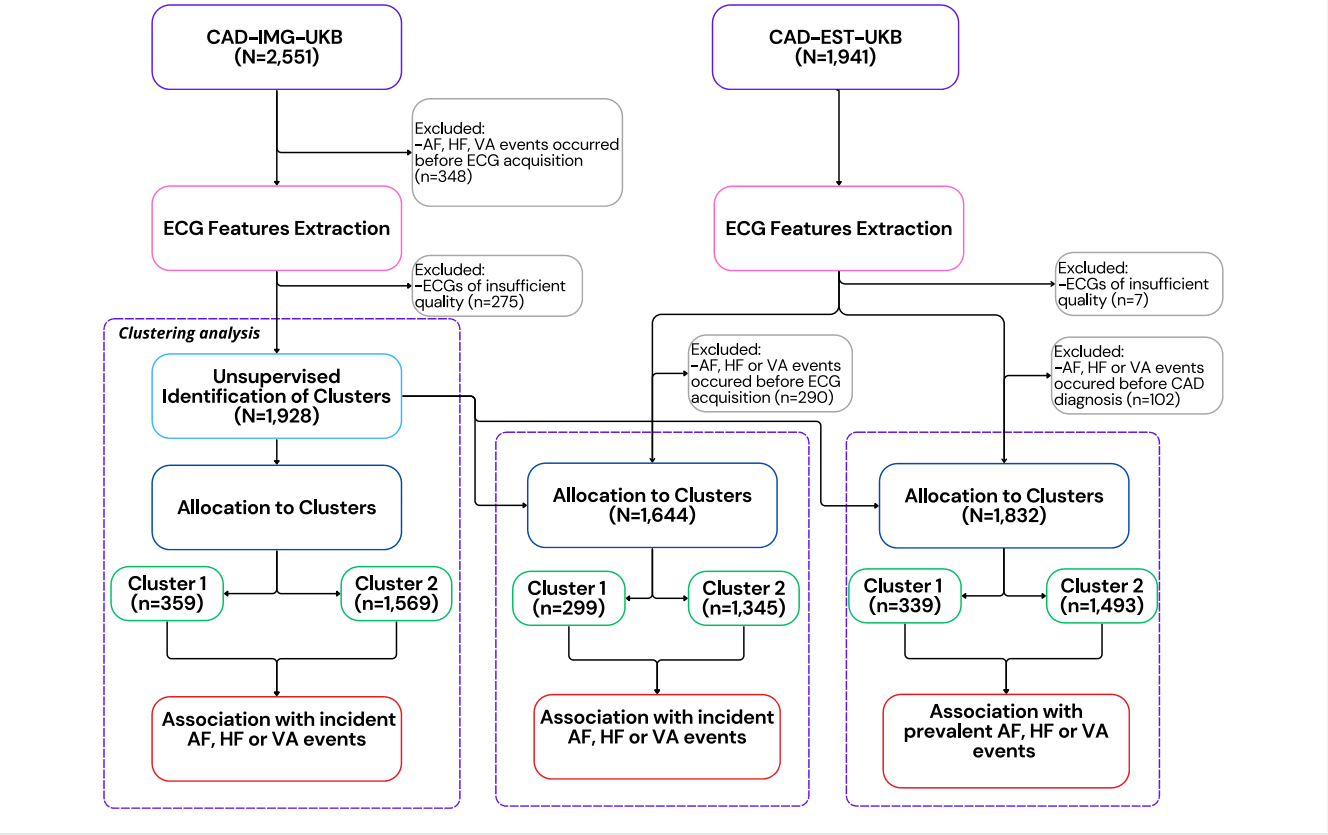


Figure 1 Outline of the study design. From the UK Biobank study, we derived two independent populations of individuals with prevalent coronary artery disease: the CAD-IMG-UKB and the CAD-EST-UKB cohorts. The CAD-IMG-UKB cohort was used to perform an unsupervised clustering model based solely on electrocardiogram features and test the association of the clusters with incident atrial fibrillation, heart failure, and ventricular arrhythmia diagnoses. The CAD-EST-UKB cohort was used to validate the association of incident diagnoses and to test the association of the clusters with prevalent ones. Inclusion and exclusion criteria for each analysis are presented. CAD, coronary artery disease; AF, atrial fibrillation; HF, heart failure; VA, ventricular arrhythmia; CAD-IMG-UKB, individuals with prevalent CAD in the UKB study who participated in the imaging study; CAD-EST-UKB, individuals with prevalent CAD in the UKB study who participated in the exercise stress test study.

The imputation was performed following the classification and regression trees method, with a total of 10 iterations to refine the estimates. Continuous variables are reported as medians and interquartile range (IQR), while categorical variables are presented as numbers and percentages. Differences in ECG, demographic (sex, age, alcohol, and smoking status), and clinical features [body mass index (BMI), systolic and diastolic blood pressure (SBP and DBP), left ventricular ejection fraction (LVEF), LDL, HDL, triglycerides, and diabetes] between clusters were compared using the Wilcoxon rank-sum test for continuous variables and the Fisher test for categorical variables.

Association with incident AF, HF, and VA risk in the CAD-IMG-UKB and CAD-EST-UKB cohorts was tested using univariable and multivariable Cox proportional hazards models. Age, sex, BMI, SBP, DBP, alcohol, and smoking status were considered as covariates for the multivariable Cox models. Additionally, HDL, LDL, and triglycerides were included as covariates in the association with incident diagnoses for the CAD-EST-UKB cohort, as these covariates were only available for this cohort. Hazard ratio (HR), 95% confidence interval (CI) and *P*-values are reported.

To test for association with prevalent AF, HF, and VA risk in the CAD-EST-UKB, we used univariable and multivariable binomial logistic regression models. Multivariate models included age, sex, BMI, SBP,

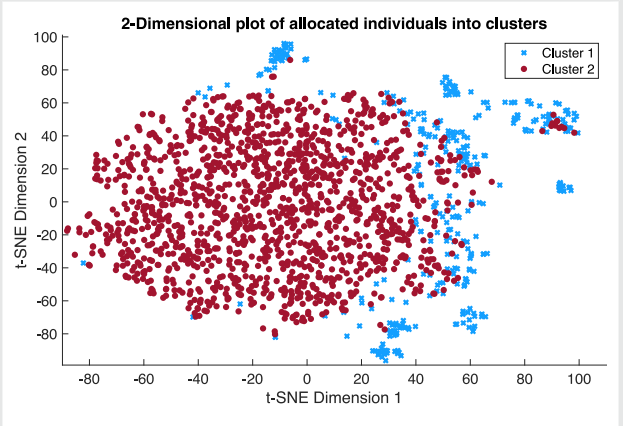


Figure 2 Two-dimensional representation of the clusters identified in the CAD-IMG-UKB cohort along with the allocation of individuals from the association analyses with incident and prevalent diagnoses in the CAD-EST-UKB cohort within each cluster. Each point represents an allocated individual.

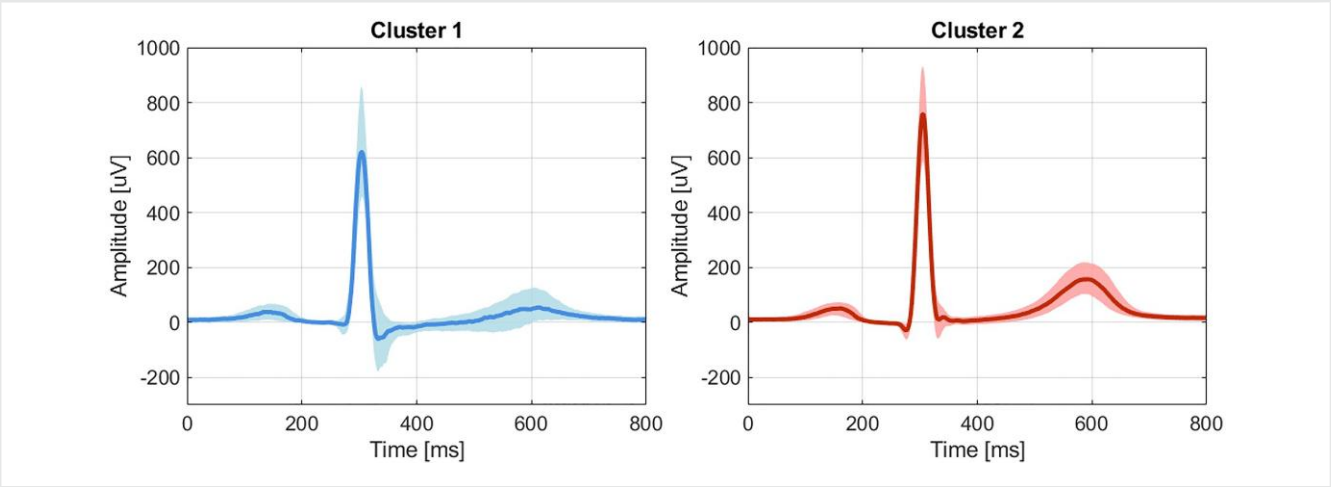


Figure 3 Representative median and 25–75% interquartile range (shaded) heartbeat ECG for each cluster in the CAD-IMG-UKB cohort.

DBP, HDL, LDL, triglycerides, diabetes, alcohol, and smoking status as covariates. We report the performance using odds ratio (OR), 95% CI, and *P*-values.

For both Cox proportional hazard and binomial logistic regression models, multivariable models were constructed adjusting for covariates with less than 10% of missing data. Subsequently, stepwise regression models were applied to identify the minimal group of features that optimally describe the model based on the Akaike information criterion. The Kaplan–Meier survival curves were plotted using the R library 'survminer'. Coordinate charts were plotted to represent differences in median clinical and ECG features using the R library 'ggplot'. The most significant ($P < 0.001$) and independent features are represented in each plot. Data points represent the median for each feature, and the scale is normalized using min–max scale (ranges from 0 to 1) for each group of features. Categorical features are represented in bar plots. The units for each variable are indicated in *Table 1*. Statistical significance was assumed when $P < 0.001$ after Bonferroni correction. Statistical analyses, survival curves, coordinate charts, and box plots were performed using R (version 4.2.2).

Results

Study population

The median age in the CAD-IMG-UKB cohort was 70 years (IQR 10), and 69.7% individuals were male (*Table 1*). The BMI levels showed individuals were in overweight, with a median of 27.18 kg/m^2 (IQR 5.45) and had elevated SBP levels, with a median of 140.5 mmHg (IQR 25.5). In the CAD-EST-UKB cohort, the median age was 64 years (IQR 7) and 72.1% were male (*Table 2*). Further demographic and clinical characteristics of both cohorts are shown in *Tables 1* and *2*.

Unsupervised clustering analysis

From the 24 standardized ECG features, the upward and downward slopes of the QRS complex were removed from the model due to their high correlation with its amplitude and duration. Additionally, the Hermite-based width for the reconstruction of the T-wave was withdrawn due to its high correlation (>0.8) with its Hermite Base 2. Therefore, the final unsupervised clustering model included 21 standardized features that represent components of ventricular depolarization and repolarization in the median heartbeat ECG of each individual (see Supplementary material online, *Table S1*).

The optimal number of clusters indicated by the Silhouette score was 2, with a corresponding coefficient of 0.47 (see Supplementary material online, *Figure S1*). The Gap statistic showed no significant increase for other number of clusters. Thus, the 2-means clustering algorithm identified two distinct clusters of ECG features in CAD.

Allocation to clusters

In the CAD-IMG-UKB cohort, Cluster 1 included 359 individuals and Cluster 2 included the remaining 1569 individuals (*Table 1* and *Figure 1*). In the association with incident diagnoses for the CAD-EST-UKB cohort, 299 individuals were allocated to Cluster 1, while the remaining 1345 individuals were allocated to Cluster 2 (*Table 2* and *Figure 1*). Finally, in the association analysis with prevalent diagnoses for the CAD-EST-UKB cohort, Cluster 1 comprised 339 individuals, while 1493 individuals were allocated to Cluster 2. *Figure 2* illustrates a 2D representation of the allocated individuals from the association analyses for incident and prevalent diagnoses for the CAD-EST-UKB cohort to each cluster.

Differences in standard and morphologic electrocardiogram indices between clusters

Figure 3 illustrates the representative ECG heartbeats of Clusters 1 and 2 in the CAD-IMG-UKB cohort. Individuals in Cluster 1 had significantly wider QRS complexes (*Table 1*; Supplementary material online, *Figure S2*). The morphological differences in the QRS complex between the two clusters were mainly characterized by the Hermite's Bases 1 and 2, exhibiting different Q and S waves, and QRS amplitudes as shown in *Table 1*. Specifically, Cluster 1 exhibited lower median values of these bases when compared with Cluster 2 (2.2 vs. 2.6 for Hermite's Bases 1 and 0.1 vs. 0.4 for Hermite's Base 2) and exhibited higher variability (IQR 1.37 vs. 0.49 for Hermite's Base 1 and 0.66 vs. 0.27 for Hermite's Base 2). The higher contribution of QRS Hermite's Bases 1 and 2 in the QRS reconstruction in Cluster 2 indicates more pronounced Q and R waves than in Cluster 1 (see Supplementary material online, *Figure S2*).

Additionally, significant differences in ventricular repolarization-related features were observed between the two clusters (*Table 1*). Cluster 1 exhibited a larger area under the ST segment and larger depression of the ST segment's amplitude (median of −8.7 μV), lower

Table 1 Baseline and electrocardiogram characteristics for all individuals and clusters in the CAD-IMG-UKB cohort

Characteristic	All ($n = 1928$)		Cluster 1 ($n = 359$)		Cluster 2 ($n = 1569$)		*P*-value
Demographic and clinical							
Male sex, no. [%]	1344	[69.7%]	287	[79.94%]	1057	[67.37%]	**<0.001**
Age, yr	70	[10]	70	[8]	69	[10]	0.01
BMI, kg/m^2	27.18	[5.45]	27.80	[5.57]	27.04	[5.32]	0.01
SBP, mmHg	140.5	[25.5]	143	[23.62]	140	[25.25]	**0.001**
DBP, mmHg	76	[13]	77	[13.62]	75.5	[13]	0.08
Diabetes, no. [%]	268	[13.90%]	63	[17.55%]	205	[13.07%]	0.03
Smoker, no. [%]	69	[3.58%]	13	[3.62%]	56	[3.57%]	1.00
Alcohol, no. [%]	315	[16.34%]	54	[15.04%]	261	[16.63%]	0.48
LVEF, %	58.93	[8.89]	56.85	[10.22]	59.38	[8.28]	**<0.001**
ECG indices							
RR interval, ms	1084.0	[220.5]	1064.0	[216]	1090.0	[220]	0.01
Depolarization ECG indices							
Standard indices							
QRS-Amplitude, μV	840.9	[389.92]	824.0	[410.56]	844.2	[384.14]	0.26
QRS-width, ms	82.0	[26]	104.0	[60]	78.0	[22]	**<0.001**
Morphology-based indices							
QRS-Hermite width, ms	12.7	[2.42]	14.6	[5.28]	12.5	[2.11]	**<0.001**
QRS-Hermite error	0.0	[0.01]	0.0	[0.03]	0.0	[0.01]	**<0.001**
QRS-Hermite Base 1	2.6	[0.6]	2.2	[1.37]	2.6	[0.49]	**<0.001**
QRS-Hermite Base 2	0.4	[0.31]	0.1	[0.66]	0.4	[0.27]	**<0.001**
QRS-Hermite Base 3	−0.5	[0.59]	−0.7	[1.36]	−0.5	[0.51]	0.01
QRS-Hermite Base 4	−0.1	[0.41]	−0.3	[0.52]	0.0	[0.38]	**<0.001**
Repolarization ECG indices							
Standard indices							
ST segment area, μV^2	23.4	[21.94]	25.0	[28.01]	23.0	[20.6]	**<0.001**
ST segment amplitude, μV	5.6	[30.91]	−8.7	[42.48]	7.5	[27.85]	**<0.001**
T-amplitude, μV	190.8	[113.43]	140.6	[120.83]	198.7	[106.05]	**<0.001**
QTc interval, ms	395.6	[34.72]	417.9	[50.07]	392.6	[31.47]	**<0.001**
Tpec interval, ms	76.7	[13.04]	83.1	[25.61]	76.0	[11.22]	**<0.001**
Morphology-based indices							
T-Hermite error	0.05	[0.04]	0.1	[0.07]	0.0	[0.03]	**<0.001**
T-Hermite Base 1	5.18	[0.88]	4.4	[6.77]	5.2	[0.69]	**<0.001**
T-Hermite Base 2	0.03	[0.22]	0.1	[0.55]	0.0	[0.2]	**<0.001**
TMV index	2.0	[1.91]	4.3	[4.65]	1.8	[1.37]	**<0.001**
TMV index (T-peak to T-end)	1.9	[1.8]	2.8	[3.11]	1.8	[1.59]	**<0.001**
Inversed T-polarity	91	[4.72%]	70	[19.50%]	21	[1.34%]	**<0.001**
Other ECG indices							
Ratio T-amplitude/QRS-amplitude	0.2	[0.15]	0.2	[0.16]	0.2	[0.14]	**<0.001**

Number [%]; median [interquartile range]; bold indicates $P < 0.001$.
BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; LVEF, left ventricular ejection fraction; ECG, electrocardiogram; QTc, corrected QT interval; Tpec, corrected T-peak to T-end interval; TMV, t-wave's morphological variations index.

T-wave amplitude, and a longer QTc interval (25 ms on average). Longer Tpec intervals in Cluster 1 (7 ms on average) suggest that the prolonged QTc interval is not driven by QRS duration increase alone. Morphological variations in the T-wave were characterized by the Hermite's Bases 1 and 2, TMV indices, and T-wave polarity. A larger contribution of the Hermite's Base 1 in Cluster 2 [5.2 (IQR 0.69)] is associated with positive and monophasic T-waves. In contrast, Cluster 1 showed more negative and biphasic T-waves associated with the larger contribution of the Hermite Base 2 [0.1 (IQR 0.55)] vs. 0.01 (IQR 0.20); Supplementary material online, *Figure S3*]. In the same line, the majority of inverted T-waves were found in Cluster 1 (19.5%). The TMV indices were two-fold higher in Cluster 1 [4.3 (IQR 4.65)], indicating larger morphological differences from a reference T-wave when compared with Cluster 2 [1.8 (IQR 1.37)]. *Figure 4* summarizes the ECG profiles for each cluster within the CAD-IMG-UKB cohort.

Findings observed for ECG features in the CAD-EST-UKB cohort were concordant with those in CAD-IMG-UKB (*Table 2*).

Table 2 Baseline and electrocardiogram characteristics for all individuals and clusters in the CAD-EST-UKB cohort

Characteristic	All ($n = 1644$)		Cluster 1 ($n = 299$)		Cluster 2 ($n = 1345$)		*P*-value
Demographic and clinical							
Male sex, no. [%]	1186	[72.1%]	250	[83.61%]	936	[69.59%]	**<0.001**
Age, yr	64	[7]	65	[6]	64	[7]	0.003
BMI, kg/m^2	28.15	[5.6]	28.95	[6.74]	28.04	[5.37]	0.03
SBP, mmHg	137	[22]	138	[20]	136.5	[22.5]	0.16
DBP, mmHg	78.5	[12.5]	78.5	[13]	78	[12.5]	0.86
LDL, mmol/L	2.66	[0.98]	2.55	[0.96]	2.67	[0.98]	0.01
HDL, mmol/L	1.22	[0.4]	1.18	[0.36]	1.23	[0.42]	0.01
Triglycerides, mmol/L	1.54	[1.11]	1.61	[1.09]	1.53	[1.12]	0.49
Diabetes, no. [%]	276	[16.79%]	66	[22.07%]	210	[15.61%]	0.01
Smoker, no. [%]	150	[9.12%]	29	[9.70%]	121	[9.00%]	0.66
Alcohol, no. [%]	337	[20.50%]	61	[20.40%]	276	[20.52%]	1.00
ECG indices							
RR interval, ms	970.0	[204]	928.0	[234]	982.0	[204]	**<0.001**
Depolarization ECG indices							
Standard indices							
QRS-amplitude, μV	1113.5	[514.09]	1091.7	[534.44]	1117.4	[512.17]	0.24
QRS-width, ms	86.0	[32]	106.0	[50]	82.0	[24]	**<0.001**
Morphology-based indices							
QRS-Hermite width, ms	12.4	[2.17]	13.9	[3.68]	12.2	[2]	**<0.001**
QRS-Hermite error	0.0	[0.01]	0.0	[0.02]	0.0	[0.01]	**<0.001**
QRS-Hermite Base 1	2.5	[0.71]	2.0	[1.37]	2.5	[0.61]	**<0.001**
QRS-Hermite Base 2	0.3	[0.38]	0.1	[0.9]	0.4	[0.32]	**<0.001**
QRS-Hermite Base 3	−0.6	[0.6]	−0.8	[1.05]	−0.6	[0.54]	**<0.001**
QRS-Hermite Base 4	−0.1	[0.4]	−0.3	[0.44]	−0.1	[0.39]	**<0.001**
Repolarization ECG indices							
Standard indices							
ST segment area, μV^2	34.5	[29.44]	39.5	[37.55]	33.5	[27.92]	**<0.001**
ST segment amplitude, μV	11.1	[40.58]	−1.9	[61.35]	12.9	[38.35]	**<0.001**
T-amplitude, μV	242.7	[143.15]	181.9	[125.74]	255.0	[138.39]	**<0.001**
QTc interval, ms	396.9	[35.23]	416.1	[39.7]	393.7	[32.73]	**<0.001**
Tpec interval, ms	78.6	[14.5]	86.4	[26.52]	77.8	[12.78]	**<0.001**
Morphology-based indices							
T-Hermite error	0.1	[0.04]	0.1	[0.06]	0.1	[0.03]	**<0.001**
T-Hermite Base 1	5.2	[0.82]	4.7	[3.62]	5.2	[0.69]	**<0.001**
T-Hermite Base 2	0.0	[0.24]	0.2	[0.56]	0.0	[0.2]	**<0.001**
TMV index	2.1	[1.86]	4.0	[4.36]	1.9	[1.35]	**<0.001**
TMV index (T-peak to T-end)	2.0	[1.92]	2.9	[2.91]	1.8	[1.73]	**<0.001**
Inversed T-polarity	65	[3.95%]	52	[17.39%]	13	[0.97%]	**<0.001**
Other ECG indices							
Ratio T-amplitude/QRS-amplitude	0.2	[0.15]	0.2	[0.14]	0.2	[0.14]	**<0.001**

Number [%]; median [interquartile range]; bold indicates $P < 0.001$.
BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; LDL, low-density lipoprotein; HDL, high-density lipoprotein; ECG, electrocardiogram; QTc, corrected QT interval; Tpec, corrected T-peak to T-end interval; TMV, T-wave's morphological variations index; AF, atrial fibrillation; HF, heart failure; VA, ventricular arrhythmia.

Differences in demographic and clinical features between clusters

As shown in *Table 1* and *Figure 4*, the distribution of demographic and clinical features was similar across the two clusters in the CAD-IMG-UKB cohort. The median age was 70 years (IQR 8) in Cluster 1 and 69 years (IQR 10) in Cluster 2, with 79.94% male in Cluster 1 and 67.37% male in Cluster 2 ($P < 0.001$). Among the cardiovascular risk factors, individuals in Cluster 1 exhibited higher SBP (143 mmHg vs. 140 mmHg) and had lower LVEF (56.85% vs. 59.38%) than those in Cluster 2. Additional details regarding differences in the characteristics among clusters are provided in *Table 1* and Supplementary material online, *Figure S4*.

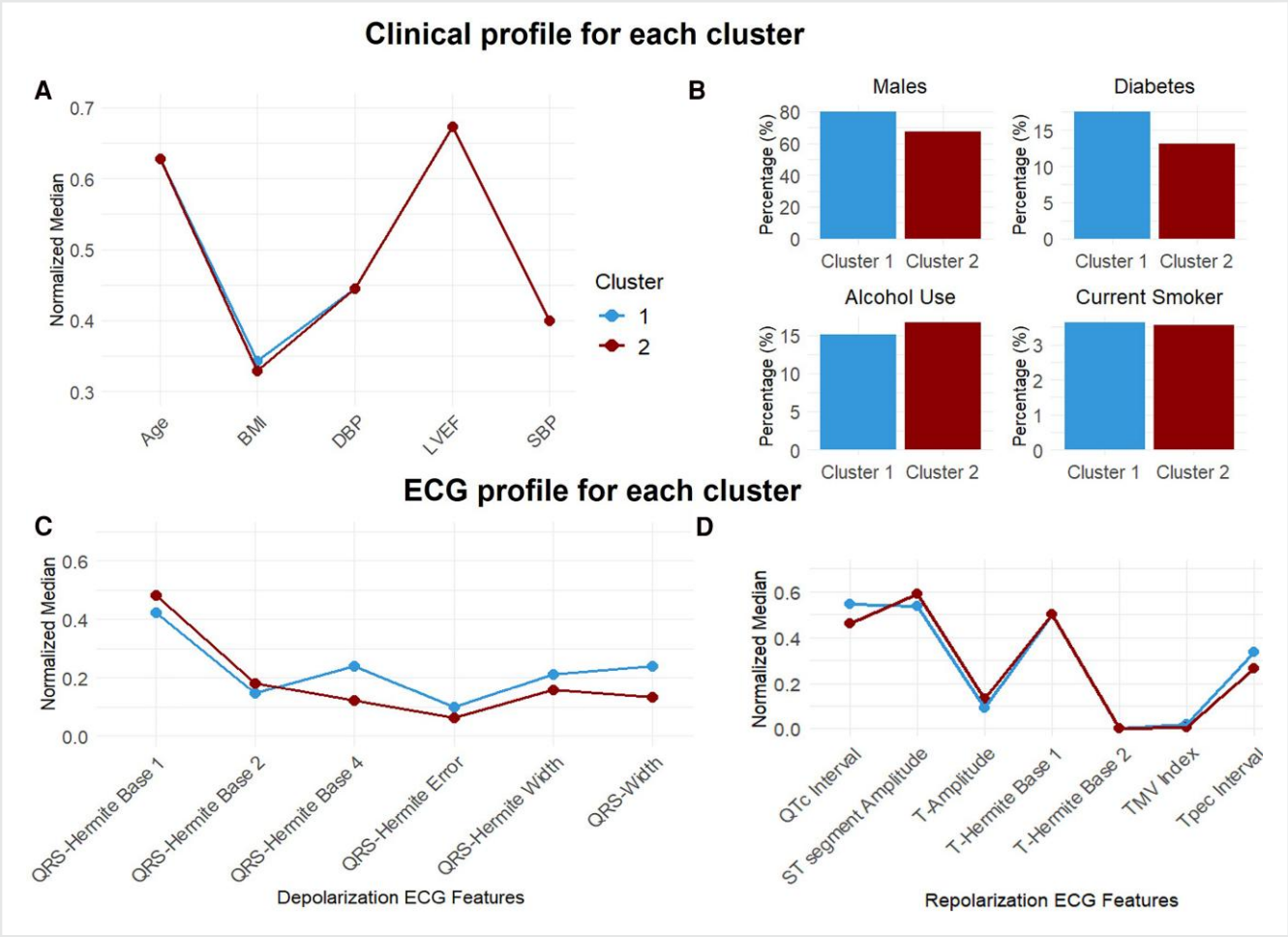


Figure 4 Clinical and electrocardiogram profiles for each cluster in the CAD-IMG-UKB cohort. (*A*) The quantitative clinical features. (*B*) Categorical clinical features are represented in bar plots. (*C*) The most significant and independent depolarization-related electrocardiogram features. (*D*) The most significant and independent repolarization-related electrocardiogram features. Data points represent the median for each feature. In each plot, the scale is normalized using min–max to the range of each group of features. Units for each variable correspond to those indicated in *Table 1*.

Regarding the CAD-EST-UKB cohort, in contrast to the CAD-IMG-UKB cohort, no significant differences were found in levels of SBP. The rest of the findings were consistent with those in the CAD-IMG-UKB cohort (*Table 2*).

Differences in prevalent coronary artery disease diagnoses of ICD9, ICD10, and OPCS-4 codes between clusters

In the CAD-IMG-UKB cohort, nominally significant differences were observed for angina pectoris (I20.9) [$n = 65$ (18%) vs. $n = 362$ (23%), $P = 0.03$] and saphenous vein graft replacement of one coronary artery (K40.1) [$n = 3$ (1%) vs. $n = 2$ (0.1%), $P = 0.05$] when comparing Cluster 1 vs. Cluster 2.

In the CAD-EST-UKB cohort, additional significant differences were found between clusters for saphenous vein graft replacement for two coronary arteries [$n = 5$ (2%) vs. $n = 5$ (0.4%), $P = 0.02$]. The remaining results aligned with those from the CAD-IMG-UKB cohort (see Supplementary material online, *Table S3*).

There was no significant difference in the number of diagnoses of myocardial infarction between clusters, in either cohort.

Cluster-based association with incident atrial fibrillation, heart failure, or ventricular arrhythmia risk

In the CAD-IMG-UKB cohort, there were a total of 101 incident AF [in Cluster 1: 33 (9.19%) vs. Cluster 2: 68 (4.33%)], 80 incident HF [29 (8.08%) vs. 51 (3.25%)], and 15 incident VA diagnoses [6 (1.67%) vs. 9 (0.57%)] out of 1928 individuals. The univariable Cox model demonstrated that belonging to Cluster 1 was significantly associated with higher risk of incident HF in 4 years of median follow-up [HR: 2.78 (CI: 1.76–4.39), $P < 0.001$], and this association remained significant in a multivariable Cox model [HR: 2.40 (CI: 1.51–3.83), $P < 0.001$; *Table 3*; Supplementary material online, *Table S4*]. Similarly, a univariable Cox model revealed a significant association between Cluster 1 and the risk of incident AF [HR: 2.32 (CI: 1.53–3.52), $P < 0.001$], which remained significant in the multivariable Cox model [HR: 1.99 (CI: 1.30–3.04), $P = 0.001$]. Regarding risk for incident VA, the univariable Cox model revealed a nominally significant association [HR: 3.06 (CI: 1.09–8.60), $P = 0.03$], which became non-significant in the multivariable model.

In the independent CAD-EST-UKB cohort, there were a total of 267 incident AF diagnoses [in Cluster 1: 63 (21.07%) vs. Cluster 2: 204

Table 3 Cox models for incident risk of atrial fibrillation, heart failure, and ventricular arrhythmia risk in the CAD-IMG-UKB and CAD-EST-UKB cohorts

Type of CVD	Univariate model			Multivariate model		
	HR	(95% CI)	*P*-value	HR	(95% CI)	*P*-value
Hazard ratio for AF, HF, and VA risk in the CAD-IMG-UKB cohort						
HF	2.78	(1.76–4.39)	**<0.001**	2.4	(1.51–3.83)	**<0.001**
AF	2.32	(1.53–3.52)	**<0.001**	1.99	(1.30–3.04)	**0.001**
VA	3.06	(1.09–8.60)	**0.03**	2.67	(0.94–7.63)	0.07
Hazard ratio for AF, HF, and VA risk in the CAD-EST-UKB cohort						
HF	2.2	(1.64–2.96)	**<0.001**	1.77	(1.31–2.41)	**<0.001**
AF	1.55	(1.16–2.05)	**0.003**	1.31	(0.98–1.75)	0.07
VA	1.48	(0.81–2.69)	0.2	—	—	—

Bold indicates $P < 0.05$.
AF, atrial fibrillation; HF, heart failure; VA, ventricular arrhythmia; CVD, cardiovascular disease; HR, hazard ratio; CI, confidence interval.

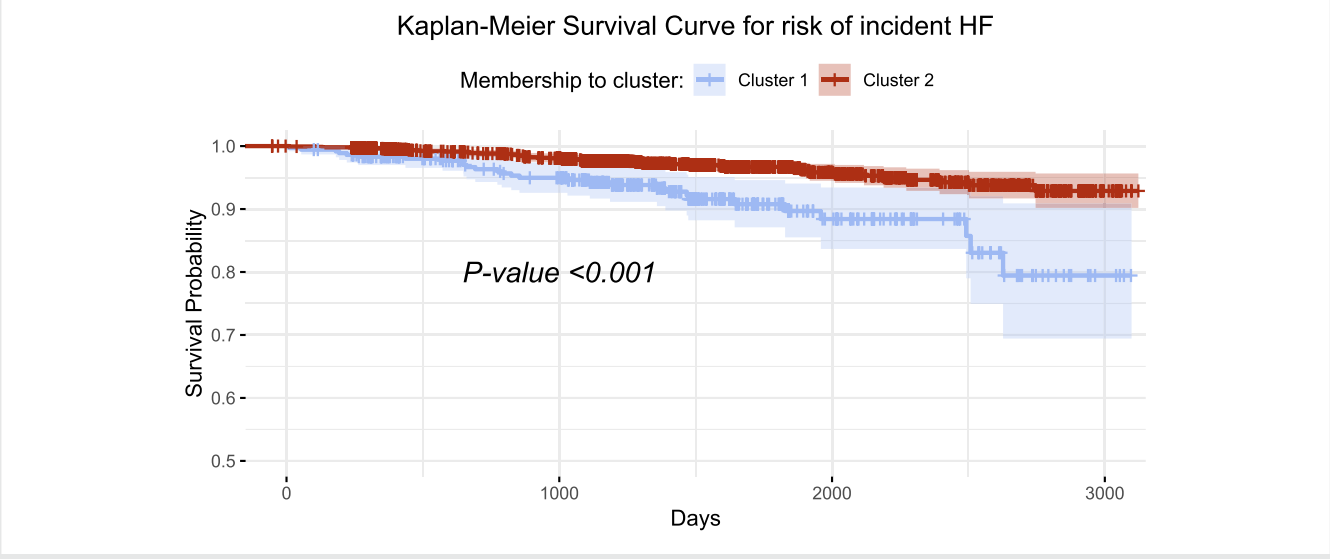


Figure 5 Survival curves for risk of incident heart failure by Clusters 1 and 2 demonstrate strong risk stratification in the CAD-IMG-UKB cohort.

(15.17%)] and 209 incident HF [63 (21.07%) vs. 146 (10.86%)] out of 1644 individuals. Both the univariable and multivariable Cox models supported the association of Cluster 1 with an increased risk of incident HF [HR: 2.20 (CI: 1.64–2.96), $P < 0.001$ and HR: 1.77 (CI: 1.31–2.41), $P < 0.001$, respectively; *Table 3*]. *Figure 5* shows the Kaplan–Meier survival curves for incident HF risk in the CAD-IMG-UKB cohort, demonstrating a decreased survival probability for individuals in Cluster 1. The association with incident AF risk was not validated in the CAD-EST-UKB cohort (*Table 3*).

Cluster-based association with prevalent atrial fibrillation, heart failure, or ventricular arrhythmia risk

In the CAD-EST-UKB cohort, when including prevalent cases, there were a total of 85 [in Cluster 1: 20 (5.90%) vs. Cluster 2: 65 (4.35%)] prevalent AF diagnoses, 34 prevalent HF diagnoses [17 (5.01%) vs. 17 (1.14%)], and 10 prevalent VA diagnoses [3 (0.88%) vs. 7 (0.47%)]. The univariate binomial logistic regression model revealed a significant association with prevalent HF diagnoses, with an OR and CI of 4.57 (2.30–9.08) and $P < 0.001$, which remained significant in the multivariable model [OR: 4.10 (CI: 2.02–8.29), $P < 0.001$; *Table 4*; Supplementary material online, *Table S5*). No association with AF or VA was observed.

Contribution of depolarization and repolarization features in the clustering process

The clustering analysis using only the repolarization-related features revealed a significant association with risk for incident AF [HR: 1.43 (CI: 1.04–1.96), $P = 0.03$], incident HF [HR: 1.59 (CI: 1.13–2.24), $P = 0.008$], and incident VA [HR: 2.05 (CI: 1.12–3.76), $P = 0.02$] when validated in the CAD-EST-UKB cohort. Regarding the association

Table 4 Binomial logistic regression for prevalent atrial fibrillation, heart failure, and ventricular arrhythmia risk in the CAD-EST-UKB cohort

Type of CVD	Univariate model			Multivariate model		
	OR	(95% CI)	*P*-value	OR	(95% CI)	*P*-value
HF	4.57	(2.30–9.08)	**<0.001**	4.10	(2.02–8.29)	**<0.001**
AF	1.38	(0.83–2.29)	0.22	—	—	—
VA	1.90	(0.49–7.33)	0.36	—	—	—

Bold indicates $P < 0.05$.
AF, atrial fibrillation; HF, heart failure; VA, ventricular arrhythmia; CVD, cardiovascular disease; Std; standard error.

with prevalent HF diagnoses in the CAD-EST-UKB cohort, Cluster 1 showed a significant association [OR: 2.75 (CI: 1.28–5.90), $P = 0.009$] (see Supplementary material online, *Table S6*). Most of the individuals classified in the HF-risk cluster (81.54%) matched those from the original Cluster 1 using all features, suggesting that similar clusters are obtained using only repolarization features. When only depolarization features were considered, no association with HF, AF, or VA was found (see Supplementary material online, *Table S7*).

Discussion

In this study, we developed, tested, and validated a model to identify distinct ECG-based morphological clusters of individuals with CAD in an unsupervised manner using short 10 s single-lead ECGs. The main finding of this study is the identification of individuals with CAD mapping to two distinct ECG morphological clusters, of which Cluster 1, relative to Cluster 2, showed a stronger association with both incident and prevalent HF, independently of age, sex, and other clinical variables.

Given the configuration of the clusters in our unsupervised model, the evaluation metrics indicated moderate separation between clusters with minimal overlapping between the groups, while no significant improvement was observed when tested for other configurations (3–10 clusters). Thus, suggesting that this configuration of clusters adequately captures the underlying structure of the data without adding unnecessary complexity to the model.

Individuals in Cluster 1 exhibited significant associations with incident and prevalent HF, suggesting that the shared ECG features in this cluster reflect an underlying mechanical and/or electrophysiological predisposition to HF. These ECG features included wider QRS complexes, longer QTc and Tpec intervals, a greater prevalence of inverted T-waves in lead I, and higher T-wave morphological variations, all of which are known clinical biomarkers associated with increased risk in the general population.[27,32–34] Individuals in Cluster 1 also showed higher prevalence of established clinical risk factors in CAD, including a lower end of normal LVEF (56.85% compared with 59.38% in Cluster 2) and higher blood pressure.[1] Additionally, 43 out of the 235 individuals in Cluster 1 with available LVEF information had a LVEF below 50%, compared with 85 out of the 1015 in Cluster 2. The larger proportion of participants in Cluster 1 with a LVEF < 50% indicates that individuals in this cluster share ECG features reflecting LV dysfunction that may predispose to developing the clinical syndrome of HF.

Coronary artery disease is one of the main causes of HF, and individuals diagnosed with both CAD and concomitant HF exhibit a significant increase in mortality rate.[35,36] So far, no association with HF has been declared in previous unsupervised models in CAD.[19,20,37,38] These models had offered prognostic value in typical outcomes in CAD (i.e. mortality, myocardial infarction, stroke, and stenosis) using as input demographic, biochemical, standard ECG, imaging, and genetic data.[19,20,37,38] Our ECG-based model was able to identify a subgroup of individuals with CAD at a higher risk of having or developing HF based solely on advanced ECG features, suggesting that the inclusion of these features is capturing relevant information on the potential structural and electrophysiological substrate leading to HF, thereby enhancing its association with incident and prevalent HF diagnoses. These standard and advanced ECG features were extracted from short single-lead ECGs at rest, with low computational complexity, enabling an affordable, fast, and non-invasive method for screening HF risk in CAD populations.

Considering ECG markers may reflect undiagnosed HF, to determine whether the association with HF in CAD is primarily driven by abnormalities in the depolarization or repolarization phase, we conducted additional analyses incorporating ECG features specific to each of these phases. We discovered that among the features used in our model, the parameters related to ventricular repolarization formed the primary component in determining the clusters, and specifically the association of Cluster 1 with incident risk of AF, HF, and VA. Unlike our study including all features, the clusters derived from only repolarization features have a stronger association with VA, suggesting that VA may be more closely linked to abnormalities in the repolarization process as reported in previous studies.[11,39] Although the depolarization features used in this study did not indicate a higher incidence or prevalence of HF on their own, some of these features provided additional risk information when combined together with the repolarization features (i.e. QRS width, Hermite's Base 2, and error of the reconstruction). In our main study, Cluster 1 exhibited an increased QRS duration, which is common in HF.[32] However, studies have shown prolongation of QRS is typically observed in more advanced stages of HF.[40,41] Therefore, when predicting the risk of incident HF in CAD, depolarization features may be less relevant in the earlier stages of the disease progression. Our findings suggest risk for HF in CAD could primarily manifest as abnormalities in duration and dispersion of ventricular repolarization. This is in agreement with the previous studies showing HF is frequently associated with repolarization features.[34,42–44]

We did not observe any association of our clusters with either incident or prevalent risk of AF and VA. Coronary artery disease is an important risk factor for AF[45]; however, we did not find a definitive association in this study. This highlights the need for further research, potentially incorporating additional ECG features. Future studies should explore this using longer ECG recordings with availability of the 12-leads and inclusion of *P*-wave parameters.[7] Regarding VA, our cohorts encompassed a limited number of VA events as found in the general population, which limited statistical power to conduct this study.

One of the strengths of our methodology is its robustness to ECGs recorded under different postural changes. While the unsupervised model was developed using 10 s resting ECGs (lead I) in a supine position (CAD-IMG-UKB cohort), it was validated with stress test ECGs recorded on a bike (CAD-EST-UKB). Variations in body posture can

affect ECG signal amplitude and morphology, though less so in lead I, as shown in Mincholé *et al.*[46] Even some studies suggest that such variations may be significant enough to either mask or mimic signs of ischaemia in the ECG.[47] Additionally, individuals in CAD-EST-UKB were younger and had longer follow-up period. Therefore, validating our findings in the CAD-EST-UKB cohort enhances the model's generalizability and robustness, reducing overfitting risks and improving clinical applicability. This approach can potentially be further validated in any dataset that includes ECG raw signals from lead I and minimum 10 s duration at rest.

Among the limitations, first we recognize some clinical variables exhibited a considerable amount of missing data (>10%), which limited their inclusion in the analyses. Second, our study was constrained by a limited population size of individuals with CAD. Finally, although this study included individuals from multiple ancestries, the majority (>90%) were of European ancestry. Future research should replicate similar studies in cohorts with diverse ancestries to enhance the generalizability of the findings.

Conclusions

Our analysis has identified in an unsupervised manner a group of individuals with CAD at risk of HF (Cluster 1) based on 10 s single-lead ECGs, allowing affordable and fast risk assessment with potential for application in large populations. Cluster 1 showed abnormalities in the ECG associated with delayed ventricular conduction and increased repolarization time compared with Cluster 2. Moreover, individuals in Cluster 1, relative to Cluster 2, were at a higher prevalent and incident HF risk independent of age, gender, or differences in other risk factors. Our findings highlight the potential utility of our model to identify individuals with CAD at increased risk of HF, crucial step to optimize effective therapeutic measures and prioritize targeted prevention measures.

Lead author biography

Josseline Madrid graduated with a M.Sc. in Biomedical Engineering from the Polytechnic University of Madrid, Spain (2022). Currently, she is a PhD researcher at the Aragon Institute of Engineering Research, University of Zaragoza, where her work focuses on advancing data science, machine learning, and cardiac electrophysiology for cardiovascular disease management. Her research aims to improve diagnosis, prognosis, and treatment options through computational approaches. Alongside her doctoral studies, she is a research collaborator at Queen Mary University of London.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Acknowledgements

The computations were performed using ICTS NANBIOSIS (HPC Unit at the University of Zaragoza). P.B.M. and W.J.Y. acknowledged the support of the National Institute for Health and Care Research Barts Biomedical Research Centre (NIHR203330), a delivery partnership of Barts Health NHS Trust, Queen Mary University of London, St George's University Hospitals NHS Foundation Trust, and St George's University of London. W.J.Y. recognized the National Institute for Health and Care Research (NIHR) Integrated Academic Training programme, which supports his Academic Clinical Lectureship post.

Author contributions

Design and conceptualization of the project: A.M., J.R., J.M. ECG signal processing, machine learning analysis, and statistical analysis: J.M., with supervision from A.M. and J.R. Interpretation of results, writing, and editing the manuscript: J.M., A.M., J.R., W.J.Y. All authors read, revised, and approved the manuscript.

Funding

This work was supported by projects PID2021-128972OA-I00, PID2023-148975OB-I00, CNS2023-143599, CNS2022-135899, and TED2021-130459B-I00 and by fellowships RYC2019-027420-I and RYC2021-031413-I, funded by the Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033).

Conflict of interest: none declared.

Data availability

Anonymized data and materials generated in this work will be returned to the UK Biobank and can be accessed upon request.

References

1. Martin SS, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, *et al.* 2024 heart disease and stroke statistics: a report of US and global data from the American Heart Association. *Circulation* 2024;**149**:e347–e913.
2. Herring N, Paterson DJ. ECG diagnosis of acute ischaemia and infarction: past, present and future. *QJM* 2006;**99**:219–230.
3. Sörnmo L, Laguna P. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. San Diego, CA, USA: Elsevier Inc.; 2005.
4. Gheorghiade M, Sopko G, De Luca L, Velazquez EJ, Parker JD, Binkley PF, *et al.* Navigating the crossroads of coronary artery disease and heart failure. *Circulation* 2006;**114**:1202–1213.
5. Benito B, Josephson ME. Ventricular tachycardia in coronary artery disease. *Rev Esp Cardiol (Engl Ed)*. 2012;**65**:939–955.
6. Kralev S, Schneider K, Lang S, Süselbeck T, Borggrefe M. Incidence and severity of coronary artery disease in patients with atrial fibrillation undergoing first-time coronary angiography. *PLoS One* 2011;**6**:e24964.
7. National Institute for Health and Care Excellence (NICE). Atrial fibrillation: diagnosis and management. NICE guideline NG196. 2021 Apr 27 [updated 2021 Jun 30]. Available from: https://www.nice.org.uk/guidance/ng196. 2021.
8. Rjoob K, Bond R, Finlay D, McGilligan V, Leslie SJ, Rababah A, *et al.* Machine learning and the electrocardiogram over two decades: time series and meta-analysis of the algorithms, evaluation metrics and applications. *Artif Intell Med* 2022;**132**:102381.
9. Siva Kumar S, Al-Kindi S, Tashtish N, Rajagopalan V, Fu P, Rajagopalan S, *et al.* Machine learning derived ECG risk score improves cardiovascular risk assessment in conjunction with coronary artery calcium scoring. *Front Cardiovasc Med* 2022;**9**:976769.
10. Sobiczewski W, Wirtwein M, Trybala E, Gruchala M. Severity of coronary atherosclerosis and stroke incidence in 7-year follow-up. *J Neurol* 2013;**260**:1855–1858.
11. Ramírez J, Kiviniemi A, van Duijvenboden S, Tinker A, Lambiase PD, Junttila J, *et al.* ECG T-wave morphologic variations predict ventricular arrhythmic risk in low- and moderate-risk populations. *J Am Heart Assoc* 2022;**11**:e025897.
12. Dong H, Wang X, Li Y, Sun C, Jiao Y, Zhao L, *et al.* Non-destructive detection of CAD stenosis severity using ECG-PCG coupling analysis. *Biomed Signal Process Control* 2023;**86**:105328.
13. Krittayaphong R, Muenkaew M, Chiewvit P, Ratanasit N, Kaolawanich Y, Phrommintikul A. Electrocardiographic predictors of cardiovascular events in patients at high cardiovascular risk: a multicenter study. *J Geriatr Cardiol* 2019;**16**:630–638.
14. Awasthi S, Sachadeva N, Abbou R, Gupta Y, Anto A, Asfahan S, *et al.* Risk stratification of coronary artery disease using the artificial intelligence-enabled electrocardiogram. *Eur Heart J* 2023;**44**:ehad655.2935.
15. Akbilgic O, Butler L, Karabayir I, Chang PP, Kitzman DW, Alonso A, *et al.* ECG-AI: electrocardiographic artificial intelligence model for prediction of heart failure. *Eur Heart J Digit Health* 2021;**2**:626–634.

16. Nezamabadi K, Sardaripour N, Haghi B, Forouzanfar M. Unsupervised ECG analysis: a review. *IEEE Rev Biomed Eng* 2023;**16**:208–224.
17. Syed Z, Guttag J. Unsupervised similarity-based risk stratification for cardiovascular events using long-term time-series data. *J Mach Learn Res* 2011;**12**:999–1024.
18. Lyon A, Ariga R, Mincholé A, Mahmod M, Ormondroyd E, Laguna P, *et al.* ECG phenotypes identified in hypertrophic cardiomyopathy using machine learning associate with arrhythmic risk markers. *Front Physiol* 2018;**9**:213.
19. Pezel T, Unterseeh T, Hovasse T, Asselin A, Lefèvre T, Chevalier B, *et al.* Clustering of patients with newly diagnosed coronary artery disease using cardiovascular magnetic resonance and coronary computed tomography angiography. *Front Cardiovasc Med* 2021;**8**:760120.
20. Williams MC, Bednarski BP, Pieszko K, Miller RJH, Kwiecinski J, Shanbhag A, *et al.* Unsupervised learning to characterize patients with known coronary artery disease undergoing myocardial perfusion imaging. *Eur J Nucl Med Mol Imaging* 2023;**50**:2656–2668.
21. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;**12**:e1001779.
22. Jane R, Laguna P, Thakor NV, Caminal P. Adaptive baseline wander removal in the ECG: comparative analysis with cubic spline technique. Proceedings Computers in Cardiology, IEEE, Durham, NC, USA, 1992:143–146.
23. Mateo J, Laguna P. Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Trans Biomed Eng* 2003;**50**:334–343.
24. Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Trans Biomed Eng* 2004;**51**:570–581.
25. Welten SJGC, van der Heijden AA, Remmelzwaal S, Blom MT, Nijpels G, Rutters F, *et al.* Prolongation of the QTc interval is associated with an increased risk of cardiovascular diseases: the hoorn study. *J Electrocardiol* 2023;**80**:133–138.
26. Chua KCM, Rusinaru C, Reinier K, Uy-Evanado A, Chugh H, Gunson K, *et al.* Tpeak-to-tend interval corrected for heart rate: a more precise measure of increased sudden death risk? *Heart Rhythm* 2016;**13**:2181–2185.
27. Pueyo E, Sornmo L, Laguna P. QRS slopes for detection and characterization of myocardial ischemia. *IEEE Trans Biomed Eng* 2008;**55**:468–477.
28. Sörnmo L, Börjesson PO, Nygårds ME, Pahlm O. A method for evaluation of QRS shape features using a mathematical model for the ECG. *IEEE Trans Biomed Eng* 1981;**28**:713–717.
29. World Health Organization. *International statistical classification of diseases and related health problems 10th revision*. Geneva, Switzerland: World Health Organization; 2011.
30. Yeh Y-C, Wang W-J, Chiou CW. Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals. *Measurement* 2009;**42**:778–789.
31. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**:2579–2605.
32. Lund LH, Jurga J, Edner M, Benson L, Dahlström U, Linde C, *et al.* Prevalence, correlates, and prognostic significance of QRS prolongation in heart failure with reduced and preserved ejection fraction. *Eur Heart J* 2013;**34**:529–539.
33. Erikssen G, Liestøl K, Gullestad L, Haugaa KH, Bendz B, Amlie JP. The terminal part of the QT interval (T peak to T end): a predictor of mortality after acute myocardial infarction. *Ann Noninvasive Electrocardiol* 2012;**17**:85–94.
34. Zègre-Hemsey JK, Hogg M, Crandell J, Pelter MM, Gettes L, Chung EH, *et al.* Prehospital ECG with ST-depression and T-wave inversion are associated with new onset heart failure in individuals transported by ambulance for suspected acute coronary syndrome. *J Electrocardiol* 2021;**69S**:23–28.
35. Lala A, Desai AS. The role of coronary artery disease in heart failure. *Heart Fail Clin* 2014;**10**:353–365.
36. John JE, Claggett B, Skali H, Solomon SD, Cunningham JW, Matsushita K, *et al.* Coronary artery disease and heart failure with preserved ejection fraction: the ARIC study. *J Am Heart Assoc* 2022;**11**:e021660.
37. Flores AM, Schuler A, Eberhard AV, Olin JW, Cooke JP, Leeper NJ, *et al.* Unsupervised learning for automated detection of coronary artery disease subgroups. *J Am Heart Assoc* 2021;**10**:e021976.
38. Yoon YE, Baskaran L, Lee BC, Pandey MK, Goebel B, Lee S-E, *et al.* Differential progression of coronary atherosclerosis according to plaque composition: a cluster analysis of PARADIGM registry data. *Sci Rep* 2021;**11**:17121.
39. Patel RB, Ng J, Reddy V, Chokshi M, Parikh K, Subacius H, *et al.* Early repolarization associated with ventricular arrhythmias in patients with chronic coronary artery disease. *Circ Arrhythm Electrophysiol* 2010;**3**:489–495.
40. Berger M, Kumowski N, Straw S, Verket M, Marx N, Witte KK, *et al.* Clinical implications and risk factors for QRS prolongation over time in heart failure patients. *Clin Res Cardiol* 2023;**112**:312–322.
41. Wang NC, Maggioni AP, Konstam MA, Zannad F, Krasa HB, Burnett JC, *et al.* Clinical implications of QRS duration in patients hospitalized with worsening heart failure and reduced left ventricular ejection fraction. *JAMA* 2008;**299**:2656–2666.
42. Tomaselli GF, Beuckelmann DJ, Calkins HG, Berger RD, Kessler PD, Lawrence JH, *et al.* Sudden cardiac death in heart failure. The role of abnormal repolarization. *Circulation* 1994;**90**:2534–2539.
43. Prenner SB, Swat SA, Ng J, Baldridge A, Wilcox JE. Parameters of repolarization heterogeneity are associated with myocardial recovery in acute heart failure. *Int J Cardiol* 2020;**301**:147–151.
44. O'Neal WT, Mazur M, Bertoni AG, Bluemke DA, Al-Mallah MH, Lima JAC, *et al.* Electrocardiographic predictors of heart failure with reduced versus preserved ejection fraction: the multi-ethnic study of atherosclerosis. *J Am Heart Assoc* 2017;**6**:e006023.
45. Batta A, Hatwal J, Batta A, Verma S, Sharma YP. Atrial fibrillation and coronary artery disease: an integrative review focusing on therapeutic implications of this relationship. *World J Cardiol* 2023;**15**:229–243.
46. Mincholé A, Zacur E, Ariga R, Grau V, Rodriguez B. MRI-Based Computational torso/biventricular multiscale models to investigate the impact of anatomical variability on the ECG QRS Complex. *Front Physiol* 2019;**10**:1103.
47. Swenson DJ, Geneser SE, Stinstra JG, Kirby RM, MacLeod RS. Cardiac position sensitivity study in the electrocardiographic forward problem using stochastic collocation and boundary element methods. *Ann Biomed Eng* 2011;**39**:2900–2910.