Physiological
Measurement

IPEM
Institute of Physics and Engineering in Medicine

PAPER • OPEN ACCESS

Single-lead ECG biomarkers for cardiovascular and mortality risk prediction

To cite this article: Lisa Attali *et al* 2026 *Physiol. Meas.* **47** 095011

View the article online for updates and enhancements.

You may also like

- Integrating machine learning techniques for enhanced prognostic modeling of heart failure risk in the american population
Vishnumon S, Sayooj Aby Jose and Anuwat Jirawattanapanit
- Estimating blood pressure from the electrocardiogram: findings of a large-scale negative results study
Seyedeh Somayyeh Mousavi, Sajjad Karimi, Mohammadsina Hassannia et al.
- Machine learning-based atrial fibrillation detection and onset prediction using QT-dynamicity
Jean-Marie Grégoire, Cédric Gilon, Nathan Vaneberg et al.

This content was downloaded from IP address 90.165.163.198 on 30/09/2026 at 09:40

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

Check for updates

OPEN ACCESS

RECEIVED
5 April 2026

REVISED
3 September 2026

ACCEPTED FOR PUBLICATION
18 September 2026

PUBLISHED
30 September 2026

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

Single-lead ECG biomarkers for cardiovascular and mortality risk prediction

Lisa Attali[1], Yosef Solewicz[1], Shany Brimer Biton[1], Hagai Hamami[1], Eran Zvuloni[1], Ilan Green[3,7], Izhar Laufer[3], Pablo Laguna[5,6], Alba Martín-Yebra[5,6], Juan Pablo Martínez[5,6], Ronit Almog[1,4] and Joachim A Behar[1,2,*]

[1] Faculty of Biomedical Engineering, Technion-IIT, Haifa, Israel
[2] Faculty of Data and Decision Sciences, Technion-IIT, Haifa, Israel
[3] Leumit Health Services, Tel Aviv-Yafo, Tel Aviv-Yafo, Israel
[4] Epidemiology Unit, Rambam Health Care Campus, Haifa, Israel
[5] BSICoS Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, Zaragoza, Spain
[6] CIBER-BBN, Zaragoza, Spain
[7] Department of Family Medicine, Gray Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv-Yafo 6997801, Israel
* Author to whom any correspondence should be addressed.

E-mail: jbehar@technion.ac.il

Keywords: electrophysiology, digital biomarkers, atrial fibrillation burden, PVC burden, T-wave alternance

Supplementary material for this article is available online

Abstract

Objective. Cardiovascular disease remains one of the leading causes of morbidity and mortality worldwide, highlighting the need for accurate risk prediction. Atrial fibrillation burden (AFB), premature ventricular contraction burden (PVCB), and T-wave alternans (TWA) have been individually associated with adverse outcomes. We hypothesized that these three digital ECG biomarkers provide complementary prognostic information for major cardiovascular endpoints and all-cause mortality (ACM), and developed a machine learning approach combining these biomarkers for risk prediction. *Approach.* We analyzed 81 362 Holter recordings from 54 395 individuals across 20 primary care centers in Israel. A random forest model using AFB, PVCB, TWA, and age was trained to predict 5 year risk of heart failure (HF), ischemic stroke (IS), and ACM. *Main results.* On the test set, best AUROCs were for HF 0.75 [95% CI: 0.74–0.77] ($n_{pos} = 622$), for IS 0.69 [0.66–0.71] ($n_{pos} = 365$), and for ACM 0.79 [0.77–0.80] ($n_{pos} = 1060$). Combining the three ECG-derived biomarkers showed complementary predictive value and improved discrimination by up to 10% over age alone in individuals aged < 75 years. *Significance.* AFB, PVCB, and TWA contribute complementary prognostic information for risk stratification. When combined with age, these biomarkers improve predictive performance compared with age alone.

1. Introduction

The electrocardiogram (ECG) is a simple, noninvasive tool that records the heart's electrical activity and plays a central role in the diagnosis of cardiovascular disease and the prediction of future cardiac events. Atrial fibrillation burden (AFB), premature ventricular contraction burden (PVCB), and T-wave alternans (TWA) have each been investigated as individual markers of cardiovascular risk. However, to the best of our knowledge, their combined predictive value has not been evaluated.

Atrial fibrillation (AF) is the most common cardiac arrhythmia (Kamel *et al* 2016) and is associated with an increased risk of stroke (Chen *et al* 2018, Go *et al* 2018), mortality (Lee *et al* 2018, Hindricks *et al* 2021), and heart failure (HF) (Vinter *et al* 2024). Premature ventricular contractions (PVCs) are another common arrhythmia, originating from an ectopic focus in the ventricles. Both PVC frequency, also called PVCB, and morphology have been associated with reduced left ventricular ejection fraction (LVEF), HF (Dukes *et al* 2016), cardiomyopathy (Baman *et al* 2010), stroke (Agarwal *et al* 2010, 2015), and increased mortality (Lin *et al* 2014). TWA are defined as a consistent beat-to-beat variation in the

© 2026 The Author(s). Published on behalf of Institute of Physics and Engineering in Medicine by IOP Publishing Ltd

amplitude, duration, or morphology of the ST segment and/or T wave, reflecting underlying electrical instability (Verrier *et al* 2011a). The index of average T-wave alternans (IAA) is a quantitative measure of TWA burden, and it is considered a marker of ventricular vulnerability, providing important information for stratifying the risk of cardiovascular mortality and sudden cardiac death (SCD) (Goldberger *et al* 2008, Monasterio *et al* 2012, Martin-Yebra *et al* 2018). Together, AFB, PVCB, and IAA reflect complementary electrophysiological mechanisms involving atrial remodeling, ventricular ectopy, and ventricular repolarization instability.

Several studies have demonstrated that conventional ECG biomarkers contribute to cardiovascular risk stratification. Reduced heart rate variability has been associated with mortality after myocardial infarction and in patients with chronic HF (Kleiger *et al* 1987, Nolan *et al* 1998). In the Framingham Heart Study, Cheng *et al* reported that PR-interval prolongation was associated with an increased risk of all-cause mortality (ACM) (HR approximately 1.3–1.4) (2009). Prolonged QRS duration and QTc have similarly been associated with mortality and SCD (Kashani and Barold 2005, Straus *et al* 2006). More recently, Biton *et al* predicted 5 year AF risk by combining HRV measures, ECG morphology, deep-learning (DL) representations, and clinical metadata derived from 12-lead ECGs, achieving an AUROC of 0.90 with a random forest (RF) model (2021). These findings suggest that combining different ECG-derived features may improve cardiovascular risk prediction. DL studies using 10 s 12-lead ECGs have shown strong performance for cardiac detection and risk prediction. For prevalent disease detection, Attia *et al* identified reduced LVEF with an AUROC of approximately 0.93 (2019). However, detecting ventricular dysfunction present at the time of ECG acquisition is different from predicting future HF. For incident HF, Akbilgic *et al* reported an AUROC of approximately 0.75 using ECG data alone, increasing to approximately 0.82 after incorporating clinical variables (2021). Similarly, the AIRE model developed by Sau *et al* achieved C-indices of 0.787 internally and 0.768 externally for HF prediction (2024). These results suggest that ECG signals contain information relevant to future HF, while also showing the additional value of clinical variables and representations of the complete ECG signal. For mortality prediction, Raghunath *et al* achieved an AUROC of approximately 0.85 for 1 year ACM using a neural network applied to 12-lead ECGs (2020). Tsai *et al* subsequently reported C-indices of 0.858 internally and 0.836 externally for long-term mortality prediction (2023). The AIRE model achieved a C-index of 0.775 internally and values ranging from 0.63 to 0.77 across external cohorts (Sau *et al* 2024). The literature on risk prediction using long-duration Holter or ambulatory ECG recordings remains more limited. Kim *et al* reported an AUROC of 0.79 for predicting cardiac death over a 3 year follow-up using single-lead 24 h Holter recordings, increasing to 0.82 when combined with LVEF (2025). Zvuloni *et al* achieved an AUROC of 0.80 for predicting 5 year HF risk using DL applied to 24 h single-lead Holter recordings (2026). Singh *et al* reported AUROCs of 0.79 internally and 0.76 externally for predicting AF onset within approximately 15 d from ambulatory ECG recordings (2022). These findings indicate that long-duration recordings contain temporal information relevant to future cardiovascular outcomes. Direct comparison for IS is challenging because much of the ECG-AI literature concerns AF prediction or associations between predicted AF risk and subsequent stroke, rather than direct prediction of incident IS. For example, Raghunath *et al* found that 62 While previous research has primarily focused on risk prediction using standard 12-lead ECGs and DL , our study investigates the predictive value of specific digital biomarkers derived from long-term continuous ECG recordings, which capture cardiac characteristics over the course of an entire day. To the best of our knowledge, no previous study has evaluated the combined predictive value of AFB, PVCB, and TWA extracted from long-term single-lead Holter recordings for predicting HF, ACM, and IS using a machine learning (ML) approach.

We hypothesized that because AFB, PVCB, and IAA capture distinct electrophysiological processes, their combination would provide complementary prognostic information beyond any individual biomarker. Accordingly, our research objective is to investigate the combined predictive value of AFB, PVCB and IAA for incident HF, IS and ACM within 5 years. We developed and validated our models on a large cohort of 81 362 Holter ECG recordings from 54 395 unique patients. An overview of the research is shown in figure 1.

2. Methods

2.1. TLH10^2 dataset

The Technion Leumit Holter ECG 2 (TLH10^2) dataset is an updated version of the TLH10^1 dataset and was developed in collaboration with Leumit Health Services, an Israeli health maintenance organization (HMO). The acquisition, linkage, and curation of the original dataset (TLH10^1) was first presented in Zvuloni *et al* (Zvuloni *et al* 2026)., and subsequently used in Biton *et al* (2026).

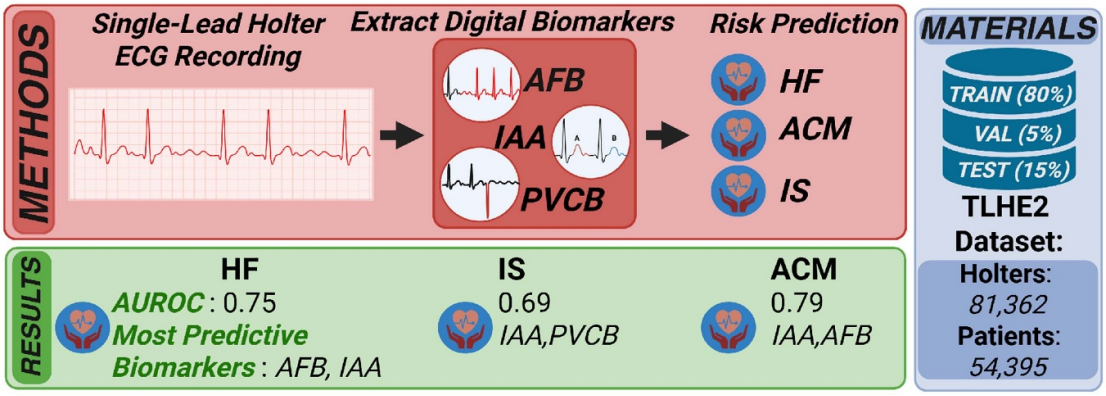


Figure 1. Research overview. A single-lead Holter ECG is processed to extract three digital biomarkers: atrial fibrillation burden (AFB), premature ventricular contraction burden (PVCB), and the index of average alternans (IAA). These biomarkers serve as inputs to machine-learning models predicting heart failure (HF), all-cause mortality (ACM), and ischemic stroke (IS). Analyses use the TLH 10^2 cohort ($n = 81{,}362$ recordings).

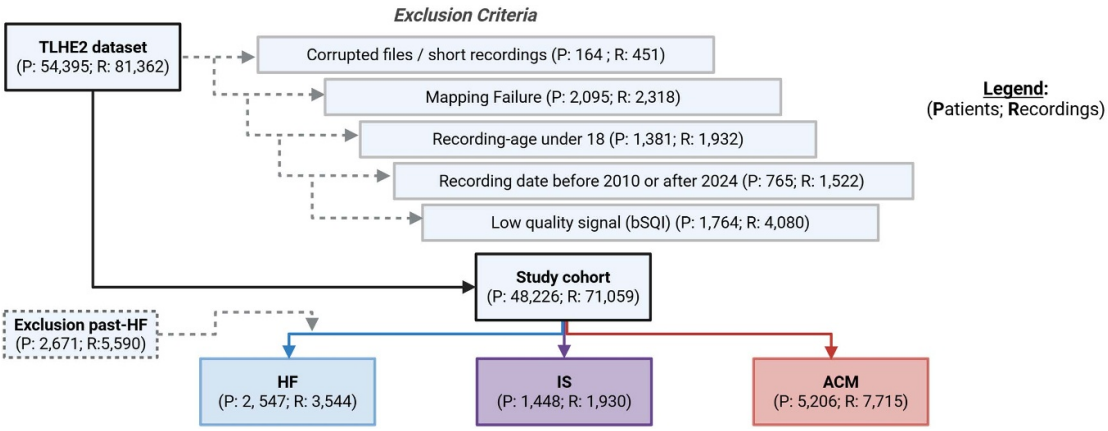


Figure 2. TLH10^2 cohort. Flowchart showing patient inclusion and exclusion criteria, and the definition of positive cases (events) for heart failure (HF), all-cause mortality (ACM), and ischemic stroke (IS) within five years of the Holter ECG. Counts are reported as the number of patients (P) and recordings (R).

The dataset includes long-term ambulatory Holter ECG recordings, generally lasting 24 h. The data was collected between 2010 and 2025, and is linked to longitudinal electronic medical records (EMR) and national mortality data, enabling follow-up up to 15 years. EMR data was last updated in January 2025. Exclusion criteria followed those of the original TLHE dataset (see figure 2). Recordings were excluded if they were corrupted, shorter than 20 h, not reliably linked to EMR data, from patients under 18 years, or had implausible timestamps (before 2010, when digital storage began, or after December 2024). Low-quality signals (bSQI) were also excluded. For the HF prediction analysis, recordings from patients with an HF diagnosis prior to the recording were excluded to ensure only incident HF events were considered. Recordings were acquired using a Lifecard CF device (Spacelabs Healthcare) with three leads, originally sampled at 1024 Hz and downsampled to 128 Hz (dynamic range 6 ± 5 mV; resolution 2.5 μV), with an average duration of approximately 24 h. For analysis, a single lead was selected according to the manufacturer-recommended electrode configuration (right sternal border at the second rib(-) and left anterior axillary line at the sixth rib(+)).

The study protocol was approved by the statutory Leumit Health Services Institutional Review Board (IRB) (approval number 0025–22 LEU). The IRB waived the requirement for informed consent given the retrospective nature of the study. All the analysis was performed within a secured Azure cloud instance.

2.2. Class definition

The first endpoint, HF, was defined as the first documented HF diagnosis in the EMR within five years following the Holter recording, based on physician-reported ICD-9 codes. The earliest recorded diagnosis date was considered the event date. These codes were selected in consultation with the Leumit director of family medicine (co-author IG) to ensure inclusion of all HF-related diagnostic codes used by

Leumit physicians (listed in Ext table 1). Recordings without HF within five years were labeled as non-HF (Zvuloni *et al* 2026).

The second endpoint, IS, was defined similarly, as the first documented IS event within five years of the Holter recording, with event dates identified through the mandatory stroke registry maintained by Leumit Health Services as part of its national reporting requirements. Recordings were labeled as IS or non-IS accordingly.

The third endpoint, ACM, was defined as death occurring within five years of the Holter recording, with event dates obtained from the national death registry. Cause-specific mortality was not available.

2.3. Preprocessing

The ECGs were upsampled from 128 Hz to 200 Hz to match the input sampling frequency used by the algorithms for biomarker extraction. Then they were bandpass filtered between 0.67 and 100 Hz (Biton *et al* 2023), followed by a 50 Hz notch filter. R-peaks were detected using the GQRS and XQRS R-peak detectors available as part of the WFDB toolkit (Goldberger *et al* 2000).

The recordings were then divided into non-overlapping windows of 60 consecutive beats based on the R-peaks. Then, a first signal-quality exclusion criterion was applied: windows containing more than 10 s of missing beat annotations were excluded. If more than 25% of a recording's 60-beat windows were excluded based on this criterion, the entire recording was discarded, similar to the procedure used in Biton *et al* (2023). From the remaining recordings, signal quality was then assessed at the beat level using the bSQI (Clifford *et al* 2012, Behar *et al* 2013). The bSQI was computed per window: a beat was considered valid when both detectors identified it within an agreement tolerance of 50 ms, otherwise, it was classified as poorly detected. Windows with fewer than 80% of well-detected beats were excluded. Finally, recordings in which more than 75% of the windows were excluded were considered severely corrupted and discarded. The final cohort is presented in figure 2.

2.4. ECG biomarkers

2.4.1. AFB

AFB is defined as the total time spent in atrial fibrillation over the entire recording. AFB is computed using the ArNet2 (Biton *et al* 2023) DL model, developed for atrial fibrillation detection in long-term ECG recordings. Following the methodology developed in Biton *et al* (2023), preprocessed ECG signals, segmented into non-overlapping windows of 60 consecutive beats, are imputed to ArNet2, which outputs window-level binary AF classifications. AFB is then computed as the ratio between the cumulative duration of windows labeled as AF and the total recording time. The equation for AFB is,

$$\mathrm{AFB} = \frac{\sum_{i=1}^{N} t_i \, \mathbb{I}_i}{\sum_{i=1}^{N} t_i} \tag{1}$$

where N represents the number of available windows, t_i the duration of the ith window (in ms), and $\mathbb{I}_i$ an indicator function equal to 1 if the i-th window is classified as AF, and 0 otherwise.

2.4.2. PVCB

PVCB is defined as the proportion of PVC beats relative to the total number of beats in the recording. PVC detection was performed using the uPVC-Net deep learning model (Hamami *et al* 2026), a PVC beat detector designed for single-lead ECG recordings and trained on multiple large ECG datasets using a multi-source, multi-lead strategy to improve generalization across lead configurations and populations. For each recording, the preprocessed non-overlapping 60-beat windows were concatenated to form a continuous signal representation, and detected R-peaks were provided to uPVC-Net to ensure that only valid signal segments were analyzed. The model outputs beat-level PVC probability scores, which are binarised using a default threshold of 0.5. PVCB is then computed as the proportion of beats classified as PVC relative to the total number of beats. The equation for PVCB is,

$$\mathrm{PVCB} = \frac{1}{B}\sum_{i=1}^{B} \hat{p}_i, \quad \hat{p}_i = \begin{cases} 1 & \text{if } p_i \geqslant 0.5, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where p_i is the PVC probability predicted for beat i, and B, the total number of beats.

2.4.3. TWA analysis (IAA)

The IAA extraction method for single-lead ECGs was adapted from the three-lead ECG method proposed by Monasterio *et al* (2012) and subsequently applied by Martin-Yebra *et al* (2018). After preprocessing, each beat was classified as normal (1) or abnormal (0) using the bSQI. Full recordings were then

divided into windows of $B_w = 128$ beats with 50% overlap. A segment was retained for TWA analysis if at least 80% of its beats satisfied three criteria: (i) the beats were labeled as normal, (ii) the absolute difference between the baseline voltage measured at the PQ segment of the current beat and that of the preceding beat was below 200 μV, (iii) and, the difference between the maximum and minimum RR intervals computed from consecutive normal beats, (ΔRR_N), was less than 300 ms. Finally, for each beat, the ventricular repolarization phase (ST-T complex) was extracted using a fixed 300 ms window after the QRS offset.

Let $x_{k,i}(n)$ denote the ST-T segment of the ith beat in the k:th window, where n indexes the N_s samples within the ST-T window and $i \in \{1, \ldots, B_k\}$ indexes the beats in the segment. Beat-to-beat alternans were computed as,

$$d_{k,i}(n) = x_{k,i}(n) - x_{k,i+1}(n), \qquad i = 1, \ldots, B_k - 1. \tag{3}$$

To preserve alternating polarity, these differences were demodulated,

$$\tilde{d}_{k,i}(n) = (-1)^{i-1} d_{k,i}(n). \tag{4}$$

The kth segment-level alternans waveform $y_k(n)$ was then obtained as the median across beats,

$$y_k(n) = \underset{i}{\text{median}} \left(\tilde{d}_{k,i}(n) \right). \tag{5}$$

To avoid cancellation between segment-level alternans waveforms of opposite polarity, the polarity of each $y_k(n)$ was aligned with a reference waveform before aggregation. After polarity alignment, a representative recording-level alternans waveform $\bar{y}(n)$ was obtained as the median across the K retained segments,

$$\bar{y}(n) = \underset{k}{\text{median}} \left(y_k(n) \right). \tag{6}$$

Finally, the IAA was computed as the mean absolute value of this aggregated waveform,

$$\text{IAA} = \frac{1}{N_s} \sum_{n=1}^{N_s} |\bar{y}(n)|. \tag{7}$$

In addition, we computed the segment ratio (SR), defined as the proportion of segments that passed exclusion, relative to the total number of segments in the recording. This metric reflects how much of the recording contributed to the IAA estimate. Let K be the number of valid segments and K_{total} the total number of extracted segments. The SR is defined as,

$$\text{SR} = \frac{K}{K_{\text{total}}}. \tag{8}$$

2.4.4. Other features

For the IS ML model, a past ischemic stroke (PIS) feature, indicating whether a patient had a history of ischemic stroke (IS) at the time of the Holter examination, was included in model training. In addition, models for all endpoints were evaluated both with and without age in order to quantify the incremental performance gain provided by the ECG-derived digital biomarkers.

2.5. Statistical analysis for individual feature

The distributions of the three digital biomarkers were compared between the positive and negative classes. Group differences were evaluated using the Mann–Whitney U test, with statistical significance defined as a two-sided p-value < 0.05.

Cox proportional hazards analyses were performed in the TLH10[2] cohort with incident HF, ACM, or IS as the outcome. PVCB and IAA were divided into quartiles (Q1–Q4). For AFB, a first group was created for the true zero AFB (G1), and the remaining non-zero values were divided into three equally populated groups(Q2–Q4). All models were adjusted for age and sex. Hazard ratios (HRs) were estimated for each quartile relative to the reference category, along with 95% confidence intervals (CI) and p-values. Separate Cox models were fitted for each metric, each adjusted for the same set of covariates. For AFB and PVCB, the lowest group (the G1 for AFB and the Q1 for PVCB) served as the reference category. For IAA, the third quartile (Q3) was used as the reference category, representing mid-range IAA

values in the cohort and allowing comparison of both lower and higher values against a central baseline, to align with survival analyses performed in prior IAA studies. Patients were followed until the occurrence of the endpoint of interest, death, or the last available follow-up. For HF and IS analyses, death and the last available follow-up were treated as censoring events. All survival analyses were performed in R version 4.5.0 using the survival package (version 3.8.3), with a patient-specific frailty term to account for the correlation between multiple Holter ECG recordings from the same individual.

All survival analyses were performed in R version 4.5.0 using the `survival` package (version 3.8.3). Because some patients contributed multiple Holter ECG recordings, Cox proportional hazards models included a shared patient-specific Gaussian frailty term. The hazard for recording j from patient i was modeled as

$$h_{ij}(t) = h_0(t) \exp\left(\boldsymbol{\beta}^{\mathsf{T}} \mathbf{X}_{ij} + b_i\right),$$

where $h_0(t)$ is the baseline hazard, $\mathbf{X}_{ij}$ represents the covariates for recording j from patient i, and b_i is a patient specific random effect. All recordings from the same patient shared the same b_i, whereas recordings from different patients had distinct random effects. This formulation accounts for within-patient dependence and unmeasured patient-level heterogeneity rather than treating repeated Holter ECG recordings as independent observations.

Kaplan–Meier (KM) curves were generated to visualize survival differences across biomarker quartiles for each endpoint. Group differences were quantified using the log-rank p-value. Curves were plotted up to 5 years of follow-up for HF and IS, and up to 10 years for ACM. KM curves complemented the Cox models by illustrating temporal separation between groups and the graded relationship between biomarker levels and clinical outcomes.

2.6. Risk prediction models

A RF classifier was trained. In our data, features varied in type and scale, and RF can accommodate this variability with minimal pre-processing. The test set consisted of all recordings from the first six months of 2018, the earliest interval with complete hospital diagnosis data, providing up to five years of follow-up. To prevent data leakage, any additional recordings from patients appearing in the test set were also assigned to the test set. The remaining data were split into training and validation subsets (95:5).The set splits for each endpoint can be found in Ext figure 1.

The RF models were trained with multiple input-feature combinations, including the digital ECG biomarkers (AFB, PVCB, IAA, SR) and age, in order to assess both the standalone predictive value of the digital biomarkers and the gains from combining them with demographic variables. Hyperparameters were tuned via Bayesian optimization to maximize the area under the receiver operating characteristic curve (AUROC). In addition to the standard RF parameters (e.g. number of trees, maximum depth, minimum samples per split), the optimization process also included class-balance settings and feature normalization strategies. The test set was kept at its natural prevalence to ensure an unbiased estimate of model performance.

Model performance was assessed on the independent test set using AUROC. Uncertainty was quantified through non-parametric bootstrapping with 1000 iterations, sampling with replacement to derive 95% CI. The same random seed was used across models to ensure identical bootstrap samples for fair comparison. Age-stratified analysis was performed for the best-performing models by computing the ΔAUROC between the combined and age-only models within predefined age groups ($<$65, 65–75, and $>$75 years), with discrimination estimated using bootstrap resampling. Feature importance was quantified using two complementary approaches. First, impurity-based feature importance was extracted directly from the fitted RF models, and expressed as percentage importance. Second, permutation importance was computed on the independent test set as the decrease in AUROC (ΔAUROC) after feature shuffling and was evaluated separately across age and sex strata.

3. Results

3.1. ECG Biomarkers distributions

A summary of biomarker distributions by class and endpoint is shown in Ext figure 2. For all three outcomes (HF, IS, and ACM), AFB, PVCB, and IAA are right-skewed, with most patients in both classes having near-zero values. AFB shows a bimodal pattern across endpoints, with peaks near 0% and 100%;

positive cases have higher densities above 0% and are more common at 100%, suggesting sustained AF is more frequent among patients who later develop the outcome. PVCB is concentrated near zero, but for HF and ACM, positive cases show higher densities in the low-burden range (about 2%–10%) and remain slightly more represented at higher values. IAA shows a similar pattern, with most values near zero and higher densities in positive cases for values above zero (1–15 μV), consistent with greater event risk.

3.2. Statistical analysis for individual features

Tables 1–3 summarize biomarker Cox regression analysis for the HF, ACM, and IS endpoints, while figure 3 provides a visual summary of the adjusted HRs and 95% CI. Across all outcomes, higher quartiles of AFB, PVCB, and IAA were associated with greater risk in unadjusted analyses.

HF (table 1, figure 3(A) showed the strongest associations. AFB demonstrated a clear dose-response relationship, with risk increasing across quartiles in both unadjusted and adjusted models (all $p < .001$). PVCB showed a similar but weaker gradient; associations were significant in unadjusted analyses and attenuated after adjustment. IAA showed a U-shaped pattern, with both low and high quartiles associated with higher HF risk relative to Q3, particularly in the higher quartile after adjustment.

For ACM (table 2, figure 3(B)), AFB remained strongly associated with risk after adjustment, though effect sizes were reduced. PVCB showed an increasing pattern in unadjusted analyses, but this was no longer significant after adjustment. IAA showed a weak U-shaped pattern, with high alternans levels significantly associated with increased risk before and after adjustment, while low levels were associated with higher risk than mid-range values but were not statistically significant.

For IS (table 3, figure 3(C)), AFB was also associated with higher risk, with significant but attenuated adjusted effects. PVCB showed a graded increase in unadjusted analyses, but not after adjustment. IAA showed a weak U-shaped pattern, with high alternans levels remaining significantly associated with higher risk after adjustment, while high levels showed a weaker trend

Overall, AFB showed the strongest and most consistent associations across all endpoints. PVCB showed a graded relationship, most clearly for HF, whereas IAA showed a U-shaped association, with both low and high alternans levels linked to higher risk than mid-range values.

KM curves supported these findings see Ext figure 3. AFB showed clear early separation across quartiles for all endpoints, with progressively worse survival from G1 to Q4. PVCB showed a similar pattern, especially for HF, while separation for ACM and IS was weaker but still significant. IAA again showed a U-shaped tendency, with reduced survival mainly in Q1 and Q4; however, for IS, only Q4 showed reduced survival, consistent with the survival analysis findings. Overall, AFB was the most discriminative biomarker, while PVCB and IAA provided more moderate but consistent risk stratification; all log-rank tests were significant ($p < .001$).

3.3. ML models for risk prediction

ML model performance was evaluated on the test set for HF, ACM, and IS using three models per endpoint: age alone, ECG biomarkers alone, and a combined Age+ECG model. Model inputs are summarized in Ext table 2, and AUROC results are shown in figure 4.

For HF, the ECG-only model achieved an AUROC of 0.68 (95% CI: 0.68-0.70), while the age-only model reached 0.72 (0.70–0.74). The combined model (HF-ECG+Age) achieved the highest performance, with an AUROC of 0.75 (0.74–0.77). For ACM, the age-only and combined models showed similar discrimination, with AUROCs of 0.78 (0.77–0.80) and 0.79 (0.77–0.80), respectively, whereas the ECG-only model achieved 0.65 (0.63–0.67). For IS, overall performance was lower. For this endpoint, the input features also included past PIS. The ECG-only model achieved an AUROC of 0.61 (0.58–0.64), which increased to 0.69 (0.66–0.71) in the combined model (IS-ECG+Age). The age-only model achieved 0.64 (0.62–0.67), while the IS-PIS+Age model reached 0.67 (0.65–0.69). Across all endpoints, the combined models consistently showed the best performance.

Age-stratified analyses comparing the combined model with the age-only model showed improved performance across all age groups for HF, with median AUROC gains of approximately 5% in the < 65 and > 75 groups and 8% in the 65–75 group. For ACM, the largest gain was observed in the 65–75 group (4%), with smaller improvements in the < 65 group (3%) and no improvement in the > 75 group (0%). For IS, the largest improvement was seen in younger individuals, with ΔAUROC values of 11% in the < 65 group, 6% in the 65–75 group, and 0% in the > 75 group.

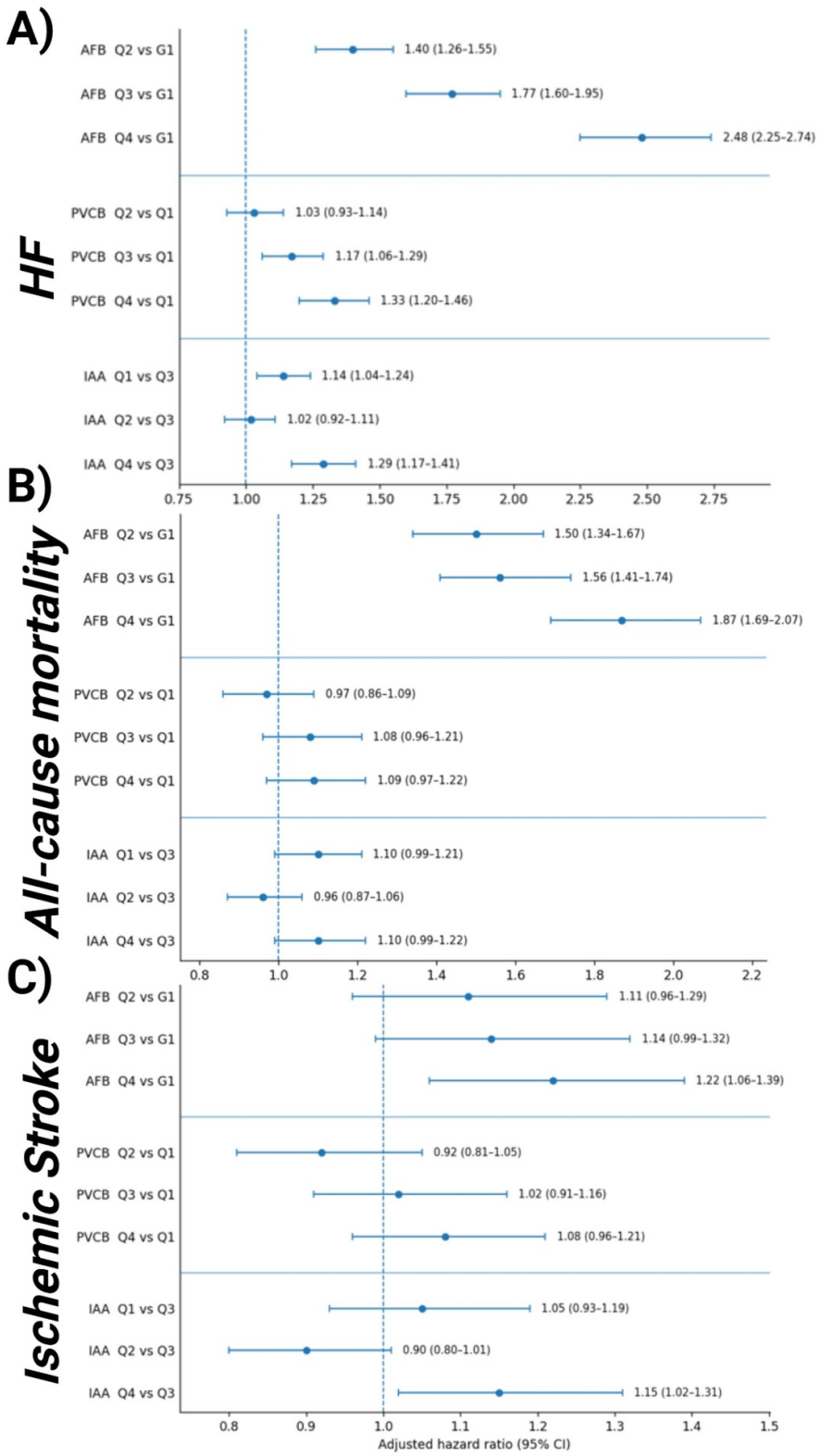


Figure 3. Forest plots of adjusted hazard ratios for incident heart failure, all-cause mortality, and ischemic stroke according to ECG-derived biomarkers. Forest plots showing hazard ratios (HRs) and 95% confidence intervals (CIs) from age and sex, adjusted Cox proportional hazards models for (A) incident heart failure (HF), (B) all-cause mortality (ACM), and (C) ischemic stroke (IS). Atrial fibrillation burden (AFB) quartiles 2–4 were compared with the reference group (G1, recordings without atrial fibrillation), premature ventricular contraction burden (PVCB) quartiles 2–4 were compared with the lowest quartile (Q1), and index of average alternans (IAA) quartiles 1, 2, and 4 were compared with the third quartile (Q3, reference). The dashed vertical line indicates the null value (HR = 1.0).

Table 1. Cox regression analysis. Results for heart failure (HF): associations with atrial fibrillation burden (AFB), premature ventricular contraction burden (PVCB), and index of average alternans (IAA), reported as hazard ratios (HR) with 95% confidence intervals (CI) and *p*-values.

Variable	Unadjusted HR (95% CI)	*p*-value	Adjusted* HR (95% CI)	*p*-value
AFB				
G1	Reference	—	Reference	—
Q2	1.89 (1.70–2.05)	<.001	1.40 (1.26–1.55)	<.001
Q3	2.68 (2.45–2.93)	<.001	1.77 (1.60–1.95)	<.001
Q4	4.14 (3.79–4.53)	<.001	2.48 (2.25–2.74)	<.001
PVCB				
Q1	Reference	—	Reference	—
Q2	1.17 (1.07–1.27)	<.001	1.03 (0.93–1.14)	0.54
Q3	1.47 (1.35–1.60)	<.001	1.17 (1.06–1.29)	0.002
Q4	2.00 (1.84–2.17)	<.001	1.33 (1.20–1.46)	<.001
IAA (TWA)				
Q1	1.36 (1.25–1.49)	<.001	1.14 (1.04–1.24)	0.03
Q2	1.01 (0.93–1.09)	0.08	1.02 (0.92–1.11)	0.75
Q3	Reference	—	Reference	—
Q4	1.39 (1.29–1.49)	<.001	1.29 (1.17–1.41)	<.001

* Adjusted for age and sex.

Table 2. Cox regression analysis. Results for all-cause mortality (ACM): associations with atrial fibrillation burden (AFB), premature ventricular contraction burden (PVCB), and index of average alternans (IAA), reported as hazard ratios (HR) with 95% confidence intervals (CI) and *p*-values.

Variable	Unadjusted HR (95% CI)	*p*-value	Adjusted* HR (95% CI)	*p*-value
AFB				
G1	Reference	—	Reference	—
Q2	2.13 (1.94–2.34)	<.001	1.50 (1.34–1.67)	<.001
Q3	2.45 (2.23–2.69)	<.001	1.56 (1.41–1.74)	<.001
Q4	3.12 (2.86–3.41)	<.001	1.87 (1.69–2.07)	<.001
PVCB				
Q1	Reference	—	Reference	—
Q2	1.10 (1.00–1.21)	0.04	0.97 (0.86–1.09)	0.63
Q3	1.37 (1.26–1.50)	<.001	1.08 (0.96–1.21)	0.15
Q4	1.64 (1.50–1.79)	<.001	1.09 (0.97–1.22)	0.15
IAA (TWA)				
Q1	1.24 (1.14–1.36)	<.001	1.10 (0.99–1.21)	0.07
Q2	0.97 (0.89–1.06)	0.53	0.96 (0.87–1.06)	0.48
Q3	Reference	—	Reference	—
Q4	1.36 (1.26–1.47)	<.001	1.10 (0.99–1.22)	0.05

* Adjusted for age and sex.

3.4. Models feature importance

Feature importance was assessed using both impurity-based and permutation-based methods (Ext table. 3, Ext figure 4, Ext figure 5)). Both approaches showed that all three ECG-derived biomarkers contributed to risk prediction across endpoints. Although age remained the strongest predictor overall (45%, 54%, and 23% for HF, ACM, and IS, respectively), the three biomarkers provided complementary information. AFB was the most important ECG feature for HF (25%), IAA contributed most strongly to ACM (15%) and IS (19%), and PVCB provided additional predictive value (9%, 9%, and 15%, respectively).

Age-stratified permutation analysis further highlighted this complementarity. For HF, age was the dominant predictor in individuals aged < 65 years (ΔAUROC $= 0.11$), followed by AFB (0.02). In the 65–75 year group, AFB had the largest effect (0.03), whereas age contributed minimally (< 0.01). In individuals aged > 75 years, age and AFB had comparable effects (0.04 each), while the remaining

Table 3. Cox regression analysis. Results for ischemic stroke (IS): associations with atrial fibrillation burden (AFB), premature ventricular contraction burden (PVCB), and index of average alternans (IAA), reported as hazard ratios (HR) with 95% confidence intervals (CI) and *p*-values.

Variable	Unadjusted HR (95% CI)	*p*-value	Adjusted* HR (95% CI)	*p*-value
AFB				
G1	Reference	—	Reference	—
Q2	1.50 (1.31–1.71)	<.001	1.11 (0.96–1.29)	0.15
Q3	1.55 (1.34–1.79)	<.001	1.14 (0.99–1.32)	<.001
Q4	1.67 (1.45–1.92)	<.001	1.22 (1.06–1.39)	<.001
PVCB				
Q1	Reference	—	Reference	—
Q2	1.05 (0.95–1.13)	0.71	0.92 (0.81–1.05)	0.20
Q3	1.23 (1.11–1.36)	<.001	1.02 (0.91–1.16)	0.67
Q4	1.32 (1.18–1.43)	<.001	1.08 (0.96–1.21)	0.19
IAA (TWA)				
Q1	1.09 (0.98–1.21)	0.09	1.05 (0.93–1.19)	0.41
Q2	1.07 (0.97–1.19)	0.17	0.90 (0.80–1.01)	0.09
Q3	Reference	—	Reference	—
Q4	1.28 (1.17–1.43)	<.001	1.15 (1.02–1.31)	0.02

* Adjusted for age and sex.

biomarkers showed smaller contributions (< 0.02). For ACM, age was the dominant predictor in the < 65 year group (0.17), with IAA and AFB each contributing 0.03. In the 65–75 year group, IAA showed the largest effect (0.03). In the > 75 year group, age again had the strongest influence (0.11), followed by IAA (0.03). For IS, in individuals aged < 65 years, PIS (0.10) and age (0.09) were the strongest contributors. In the 65–75 year group, PIS remained the leading feature (approximately 0.05), followed by IAA (0.015), whereas age contributed negligibly. In individuals aged > 75 years, age (0.04) and IAA (approximately 0.03) were the most influential features, while PIS no longer contributed.

Sex-stratified permutation analysis showed consistent feature importance rankings between males and females across all endpoints (figure 5). For HF, age remained the dominant predictor in both sexes, followed by AFB. Similarly, for ACM, age remained the most influential feature, with IAA and AFB showing comparable contributions in both males and females. For IS, age and PIS remained the principal predictors in both sexes, whereas the ECG biomarkers contributed substantially less. Overall, only minor sex-related differences in permutation importance were observed.

4. Discussion

The survival analysis showed that all three ECG-derived biomarkers were associated with HF, ACM, and IS. Among them, AFB showed the strongest and most consistent associations, with a graded increase in risk across quartiles that persisted after adjustment, supporting its prognostic value beyond age and sex. These findings are consistent with previous studies linking higher AFB to an increased risk of HF, IS and ACM (Go *et al* 2018, Steinberg *et al* 2021). PVCB also showed a relationship across outcomes. A previous meta-analysis of 20 studies (Kilic *et al* 2025) reported pooled unadjusted HRs of 1.73 for HF, 1.31 for ACM, and 1.28 for IS when comparing the highest versus lowest PVCB categories. These estimates are broadly consistent with our unadjusted quartile-based results (HF: 1.17–2.00; ACM: 1.10–1.64; IS: 1.05–1.32). However, the adjusted HR were more modest. For IAA, previous studies reported that higher alternans levels were associated with increased ACM risk (unadjusted HR = 1.62) (Monasterio *et al* 2012). This is consistent with our findings (table 3), although the adjusted HR was more modest for our population sample. To our knowledge, no previous studies have reported associations between IAA and HF or IS, which limits direct comparison for these outcomes. We found that higher IAA levels (Q4) were associated with increased risk after adjustment, reaching statistical significance for HF and IS and showing a borderline association for ACM, further supporting the potential of this biomarker for cardiovascular risk prediction.

In the ML analysis, the combined ECG+Age model achieved the highest discrimination for HF (AUROC 0.75 vs 0.72 for age alone), with improvements across all age groups (ΔAUROC 5%–8%). Age accounted for 45% of overall feature importance, while AFB was the most influential ECG biomarker

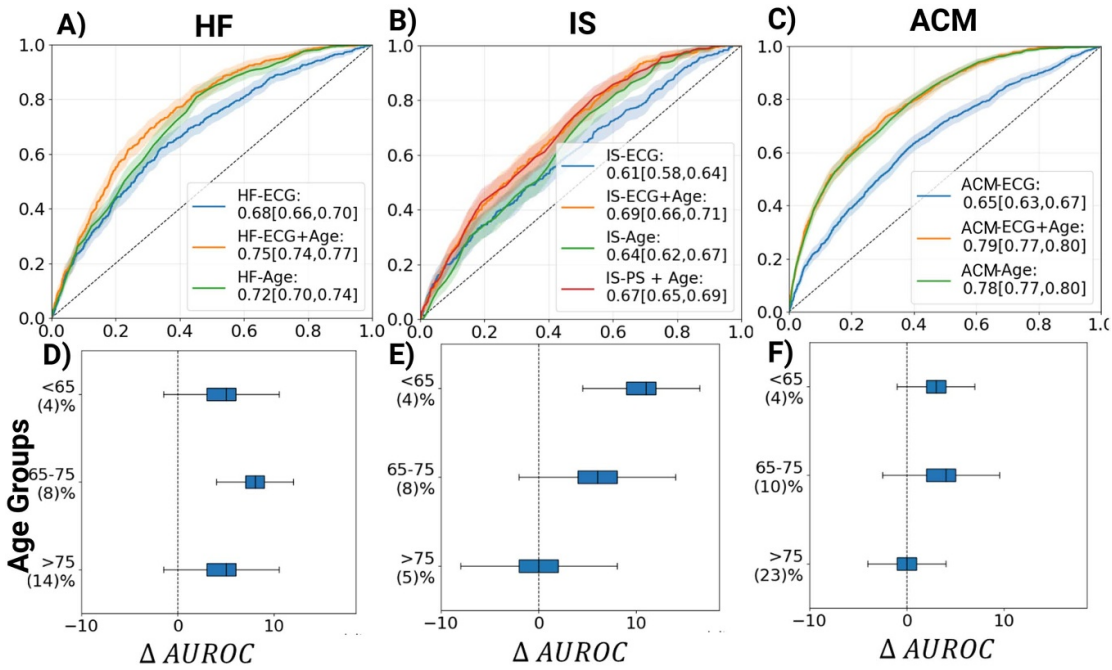


Figure 4. Model discrimination and age-stratified ΔAUROC in the TLH10[2] cohort. (A)–(C) Receiver operating characteristic (ROC) curves for heart failure (HF, (A), ischemic stroke (IS, (B), and all-cause mortality (ACM, (C) in the TLH10[2] test set, comparing ECG-only, Age-only, and combined Age+ECG models. Area under the ROC curve (AUROC) values with 95% confidence intervals (CI) are reported as *Model: AUROC [CI]*. Shaded regions indicate 95% CIs obtained from 1000 bootstrap iterations. (D)–(F) Age-stratified ΔAUROC (Age+ECG minus Age-only) for HF (D), IS (E), and ACM (F), computed within predefined age groups (<65, 65–75, and >75 years). Boxes represent the interquartile range (IQR) of the bootstrap distribution, with whiskers extending to 1.5×IQR. Percentages denote the proportion of individuals experiencing the respective endpoint.

(25%). Age-stratified permutation analysis showed that age contributed most in individuals aged < 65 years (ΔAUROC $= 0.11$), whereas AFB had the largest contribution in the 65–75 year group (0.03). In those aged > 75 years, age and AFB had comparable influence (0.04 each). Overall, AFB was the most consistent biomarker for HF prediction, particularly in middle-aged and older individuals (Steinberg *et al* 2021). PVCB and IAA provided smaller but non-redundant contributions (9%–11% importance; 1%–2% permutation), supporting complementary predictive value.

For ACM, discrimination was primarily driven by age, with similar AUROC values for the age-only and combined models (0.78 vs 0.79). Age-stratified analysis showed modest gains from ECG biomarkers in individuals aged 65–75 years (4%) and < 65 years (3%), but no improvement in those aged > 75 years. Age accounted for 54% of feature importance, while IAA was the leading ECG biomarker (15%). Permutation analysis confirmed age as the main contributor in those aged < 65 years (ΔAUROC $=$ 0.17), followed by IAA and AFB (0.03 each). In the 65–75 year group, IAA (0.03) contributed slightly more than other features, whereas in those aged > 75 years, age (0.11) remained dominant, with modest contributions from IAA (0.03).

For IS, the combined ECG+Age model achieved an AUROC of 0.69, compared with 0.64 for age alone and 0.67 for PIS+Age. The largest improvement was observed in individuals aged < 65 years (ΔAUROC 11%), followed by 6% in the 65–75 year group, with no improvement in those aged > 75 years. Age accounted for 23% of feature importance, while IAA (19%) and PVCB (15%) also contributed substantially. Permutation analysis showed that in individuals aged < 65 years, PIS (0.10) and age (0.09) were the strongest contributors. In the 65–75 year group, PIS (0.05) remained dominant, followed by IAA (0.015), whereas age showed no measurable contribution. In those aged > 75 years, age (0.04) and IAA (0.03) contributed most, and PIS no longer contributed. Overall, ECG-derived biomarkers provided complementary prognostic value beyond age, with their relative contributions varying across clinical endpoints and age groups.

Taken together, these findings support our hypothesis that AFB, PVCB, and IAA capture complementary electrophysiological processes whose relative importance depends on the clinical endpoint. AFB reflects atrial remodeling, PVCB ventricular ectopy, and IAA ventricular repolarization instability, explaining their differential contribution across HF, ACM, and IS prediction. The high importance of AFB for HF is biologically plausible, as increasing AFB reflects progressive atrial remodeling and structural heart disease while impairing ventricular filling through loss of atrial systole and irregular

ventricular activation, thereby promoting HF development (Verhaert *et al* 2021). Likewise, TWA reflects ventricular repolarization instability resulting from impaired intracellular calcium handling, electrical remodeling, and myocardial fibrosis, all characteristic of HF progression (Narayan *et al* 2008, Verrier *et al* 2011a). For ACM, the importance of AFB may reflect more advanced atrial cardiomyopathy and cardiovascular disease, whereas TWA reflects ventricular electrical instability and susceptibility to malignant ventricular arrhythmias, both of which are associated with increased mortality risk (Benjamin *et al* 1998, Verrier *et al* 2011a, Merchant *et al* 2012). The association between IAA and IS is also biologically plausible because ventricular repolarization instability and underlying structural heart disease indicate a greater overall cardiovascular disease burden (Verrier *et al* 2011a, Verhaert *et al* 2021). Finally, the consistent contribution of PVCB, particularly for IS, agrees with previous studies showing that frequent PVCs are markers of subclinical myocardial disease and ventricular dysfunction and are associated with an increased risk of IS (Agarwal *et al* 2010).

In the present study, the combination of AFB, PVCB, and IAA achieved AUROCs of approximately 0.75 for HF, 0.79 for ACM, and 0.69 for IS over a 5 year prediction horizon. The performance for HF was comparable to the ECG-only model reported by Akbilgic *et al* but lower than models using complete ECG signals, DL representations, or additional clinical information (Akbilgic *et al* 2021, Sau *et al* 2024, Zvuloni *et al* 2026). For ACM, our performance was comparable to that reported for AIRE but lower than some 12-lead DL models (Raghunath *et al* 2020, Tsai *et al* 2023, Sau *et al* 2024). Although end-to-end DL models generally achieved higher predictive performance, they learned complex representations directly from the complete ECG signal. In contrast, our objective was to evaluate the predictive value of three specific Holter-derived biomarkers using a simpler ML approach.

The age-stratified analysis further showed that, across all endpoints, age remained the most important predictor in patients aged < 65 and > 75 years but contributed substantially less in the 65–75 year group, where ECG biomarkers accounted for a larger proportion of the predictive information while maintaining a similar ranking to the overall feature importance. This may reflect the reduced discriminative value of chronological age within an intermediate-age population. In younger individuals, age strongly differentiates low- from high-risk patients, whereas in the oldest group risk is largely driven by advanced age itself. In contrast, among individuals aged 65–75 years, disease-specific electrophysiological biomarkers provide greater incremental information for distinguishing cardiovascular risk. Unlike age, sex had little influence on feature importance across all endpoints. Although small changes in the ranking of individual biomarkers were observed, the overall importance profile remained consistent between males and females. These findings suggest that the predictive information captured by AFB, PVCB, and IAA is largely independent of sex and generalizes similarly across both sexes.

Despite the large Holter cohort, this study has several limitations. Clinical outcomes were identified using ICD-9 diagnosis codes and may therefore be subject to coding errors or misclassification. The ECG biomarkers were derived using different validated extraction methods, whose methodological differences may have influenced biomarker values. HF was analyzed as a single endpoint without stratification according to ejection fraction phenotype (HFrEF, HFmrEF, or HFpEF), and only ACM was available, precluding analyses of cause-specific mortality. Performance was evaluated across age and sex but not ethnicity because ethnicity data were unavailable. Death was treated as a censoring event rather than modeled as a competing risk. Therefore, the reported hazard ratios apply to patients who remained alive and endpoint-free and do not represent competing-risk-adjusted outcome probabilities. Future studies could use Fine-Gray models to account explicitly for death before the outcome. In addition, external validation was not performed, limiting the generalizability of our findings. Future work should validate these models in independent and ethnically diverse cohorts. Feature importance reflects model-dependent predictive contributions and should not be interpreted as causal effects. Finally, because paired 12-lead ECGs were unavailable, direct comparison with contemporary deep-learning models based on standard 12-lead ECGs was not possible.

5. Conclusion

ECG-derived biomarkers were independently associated with incident HF, ACM, and IS over a 5 year follow-up. Combining these biomarkers with age improved discrimination for HF and IS, particularly in younger and middle-aged individuals, whereas ACM prediction remained largely driven by age. Together, these findings support the hypothesis that AFB, PVCB, and IAA provide complementary prognostic information and demonstrate that a small set of automatically extracted ECG biomarkers can enhance cardiovascular risk stratification using routinely acquired Holter recordings.

Because these biomarkers can be extracted automatically from routine ECG recordings, they could facilitate large-scale screening, earlier identification of high-risk individuals, and more targeted preventive strategies. As AI methods for long-duration single-lead ECG analysis continue to advance, similar biomarker extraction approaches could also be integrated into wearable devices for continuous cardiovascular risk assessment. However, prospective external validation, model calibration, and the definition of clinically meaningful decision thresholds will be required before clinical implementation.

Future research directions include developing deep learning strategies for analyzing continuous 24 h ECG recordings and comparing their performance with that of the current feature-engineering approach. They also include evaluating quantitatively the merit of using 24 h ECG recordings versus 12-lead ECGs for these risk-prediction tasks.

Acknowledgments

Manuscript submitted on September 3, 2026. Research was supported by Israel PBC-VATAT. We acknowledge the assistance of ChatGPT, an AI-based language model developed by OpenAI, in editing the manuscript.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Supplementary data 1 available at https://doi.org/10.1088/1361-6579/aea9ec/data1.

Funding

LA and JB acknowledges funding from the European Union's Horizon Europe programme under the Marie Skłodowska Curie Grant Agreement No. 101119941 (INSIDE-HEART). The research was supported by a cloud computing grant from the Israel Council of Higher Education, administered by the Israel Data Science Initiative. PL, AMY, JPM also report support from PID2022-140556OB-I00 by MICINN, Spain and T39-23 R and ARAID by Aragon government.

ORCID iDs

Lisa Attali 0009-0006-6003-2145
Yosef Solewicz 0000-0003-3987-1201
Hagai Hamami 0000-0002-2162-8513
Ilan Green 0000-0002-9435-7152
Pablo Laguna 0000-0003-3434-9254
Juan Pablo Martínez 0000-0002-7503-3339
Joachim A Behar 0000-0001-5956-7034

References

Agarwal S K *et al* 2010 Premature ventricular complexes and risk of incident stroke: the atherosclerosis risk in communities (ARIC) study *Stroke* **41** 588–93

Agarwal S K *et al* 2015 Premature ventricular complexes on screening electrocardiogram and risk of ischemic stroke *Stroke* **46** 1365–7

Akbilgic O, Butler L, Karabayir I, Chang P P, Kitzman D W, Alonso A, Chen L Y and Soliman E Z 2021 ECG-AI: electrocardiographic artificial intelligence model for prediction of heart failure *Eur. Heart J. Digit. Health* **2** 626–34

Attia Z I, Kapa S and Lopez-Jimenez F *et al* 2019 Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram *Nat. Med.* **25** 70–74

Baman T S *et al* 2010 Relationship between burden of premature ventricular complexes and left ventricular function *Heart Rhythm* **7** 865

Behar J, Oster J, Li Q and Clifford G D 2013 ECG signal quality during arrhythmia and its application to false alarm reduction *IEEE Trans. Biomed. Eng.* **60** 1660–6

Benjamin E J, Wolf P A, D'Agostino R B, Silbershatz H, Kannel W B and Levy D 1998 Impact of atrial fibrillation on the risk of death: the framingham heart study *Circulation* **98** 946–52

Biton S, Aldhafeeri M, Marcusohn E, Tsutsui K, Szwagier T, Elias A, Oster J, Sellal J M, Suleiman M and Behar J A 2023 Generalizable and robust deep learning algorithm for atrial fibrillation diagnosis across geography, ages and sexes *npj Digit. Med.* **6** 44

Biton S, Gendelman S, Ribeiro A H, Miana G, Ribeiro A L P and Behar J A 2021 Atrial fibrillation risk prediction from the 12-lead electrocardiogram using digital biomarkers and deep representation learning *Eur. Heart J. Digit. Health* **2** 576–85

Brimer Biton S *et al* 2026 Temporal phenotyping of paroxysmal atrial fibrillation reveals prognostic circadian subtypes *Biomed. Phys. Eng. Express* **2** 015009

Chen L Y, Chung M K, Allen L A, Ezekowitz M, Furie K L, McCabe P, Noseworthy P A, Perez M V and Turakhia M P 2018 Atrial fibrillation burden: moving beyond atrial fibrillation as a binary entity: a scientific statement from the American Heart Association *Circulation* **137** e623–44

Cheng S, Keyes M J, Larson M G, McCabe E L, Newton-Cheh C, Levy D, Benjamin E J, Vasan R S and Wang T J 2009 Long-term outcomes in individuals with prolonged pr interval or first-degree atrioventricular block *JAMA* **301** 2571–7

Clifford G D, Behar J, Li Q and Rezek I 2012 Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms *Physiol. Meas.* **33** 1419–33

Dukes J W *et al* 2016 Ventricular ectopy as a predictor of heart failure and death *J Am Coll Cardiol* **66** 101–9

Go A S, Reynolds K, Yang J, Gupta N, Lenane J, Sung S H, Harrison T N, Liu T I and Solomon M D 2018a Association of burden of atrial fibrillation with risk of ischemic stroke in adults with paroxysmal atrial fibrillation: the KP-RHYTHM study *JAMA Cardiol.* **3** 601–8

Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C-K and Stanley H E 2000 Physiobank, physiotoolkit and physionet: Components of a new research resource for complex physiologic signals *Circulation* **101** e215–20

Goldberger J J *et al* 2008 American Heart Association/American College of Cardiology foundation/Heart Rhythm Society scientific statement on noninvasive risk stratification techniques for identifying patients at risk for sudden cardiac death *Circulation* **118** 1497–518

Hamami H, Solewicz Y, Zur D, Kleerekoper Y and Behar J A 2026 UPVC-NET: a universal premature ventricular contraction detection deep learning algorithm *IEEE Trans. Biomed. Eng.* **20** 1–10

Hindricks G, Potpara T, Dagres N and *et al* 2021 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European association for cardio-thoracic surgery (EACTS) *Eur. Heart J.* **42** 373–498

Kamel H, Okin P M, Elkind M S V and *et al* 2016 Atrial fibrillation and mechanisms of stroke: Time for a new model *Stroke* **47** 895–900

Kashani A and Barold S S 2005 Significance of QRS duration in patients with heart failure *J. Am. College Cardiol.* **46** 2183–92

Kiliç M E *et al* 2025 Frequent premature ventricular complexes and risk of atrial fibrillation, heart failure, stroke and all-cause mortality: a meta-analysis *Heart* **111** 1175–86

Kim J-Y, Kim K-G, Joo S, Chang M, Kim J, Park K-M, On Y-K, Kim J-S, Lee Y-S and Park S-J *et al* 2025 Deep learning-based risk assessment and prediction of cardiac outcomes using single-lead 24-hour holter-ECG in patients with heart failure or myocardial infarction *J. Clin. Med.* **14** 7209

Kleiger R E, Miller J P, Bigger J T and Moss A J 1987 Decreased heart rate variability and its association with increased mortality after acute myocardial infarction *Am. J. Cardiol.* **59** 256–62

Lee E *et al* 2018 Mortality and causes of death in patients with atrial fibrillation: a nationwide population-based study *PLoS One* **13** e0209687

Lin C-Y *et al* 2014 Long-term outcome of multiform premature ventricular complexes in structurally normal heart *Int. J. Cardiol* **180** 80–85

Martin-Yebra A, Monasterio V, Cygankiewicz I, Bayes-de Luna A, Caiani E G, Laguna P and Martinez J P 2018 Post-ventricular premature contraction phase correction improves the predictive value of average T-wave alternans in ambulatory ECG recordings *IEEE Trans. Biomed. Eng.* **65** 635–44

Merchant F M, Ikeda T, Pedretti R F E, Salerno-Uriarte J A, Chow T, Chan P S, Bartone C, Hohnloser S H, Cohen R J and Armoundas A A 2012 Clinical utility of microvolt t-wave alternans testing in identifying patients at high or low risk of sudden cardiac death *Heart Rhythm* **9** 1256–1264.e2

Monasterio V, Laguna P, Cygankiewicz I and Martínez J P 2012 Average t-wave alternans activity in ambulatory ECG records predicts sudden cardiac death in patients with chronic heart failure *Heart Rhythm* **9** 383–9

Narayan S M, Bayer J D, Lalani G and Trayanova N A 2008 Action potential dynamics explain arrhythmic vulnerability in human heart failure: a clinical and modeling study implicating abnormal calcium handling *J. Am. College Cardiol.* **52** 1782–92

Nolan J, Batin P D, Andrews R, Lindsay S J, Brooksby P, Mullen M, Baig W, Flapan A D, Cowley A and Prescott R J *et al* 1998 Prospective study of heart rate variability and mortality in chronic heart failure *Circulation* **98** 1510–6

Raghunath S *et al* 2020 Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network *Nat. Med.* **26** 886–91

Sau A *et al* 2024 Artificial intelligence-enabled electrocardiogram for mortality and cardiovascular risk estimation: a model development and validation study *Lancet Digit. Health* **6** e791–802

Singh J P *et al* 2022 Short-term prediction of atrial fibrillation from ambulatory monitoring ECG using a deep neural network *Eur. Heart J. - Digit. Health* **3** 208–17

Steinberg B A, Li Z, O'Brien E C, Pritchard J, Chew D S, Bunch T J, Mark D B, Nabutovsky Y, Greiner M A and Piccini J P 2021 Atrial fibrillation burden and heart failure: data from 39,710 individuals with cardiac implanted electronic devices *Heart Rhythm* **18** 709–16

Straus S M, Kors J A, De Bruin M L, van der Hooft C S, Hofman A, Heeringa J, Deckers J W, Kingma J H, Sturkenboom M C and Stricker B H 2006 Prolonged QTC interval and risk of sudden cardiac death in a population of older adults *J. Am. College Cardiol.* **47** 362–7

Tsai D-J, Lou Y-S, Lin C-S, Fang W-H, Lee C-C, Ho C-L, Wang C-H and Lin C 2023 Mortality risk prediction of the electrocardiogram as an informative indicator of cardiovascular diseases *Digit. Health* **9** 1–16

Verhaert D V M, Brunner-La Rocca H-P, van Veldhuisen D J and Vernooy K 2021 The bidirectional interaction between atrial fibrillation and heart failure: consequences for the management of both diseases *Europace* **23** ii40-ii45

Verrier R L *et al* 2011a Microvolt t-wave alternans: physiological basis, methods of measurement and clinical utility—consensus guideline by international society for holter and noninvasive electrocardiology *J. Am. College Cardiol.* **58** 1309–24

Vinter N *et al* 2024 Temporal trends in lifetime risks of atrial fibrillation and its complications between 2000 and 2022: Danish, nationwide, population based cohort study *BMJ* **385** e077209

Zvuloni E, Almog R, Glikson M, Brimer Biton S, Green I, Laufer I, Amir O and Behar J A 2026 Modeling day-long ECG signals to predict heart failure risk with explainable AI *npj Digital Medicine* **9** 486 (arXiv: 2601.00014)