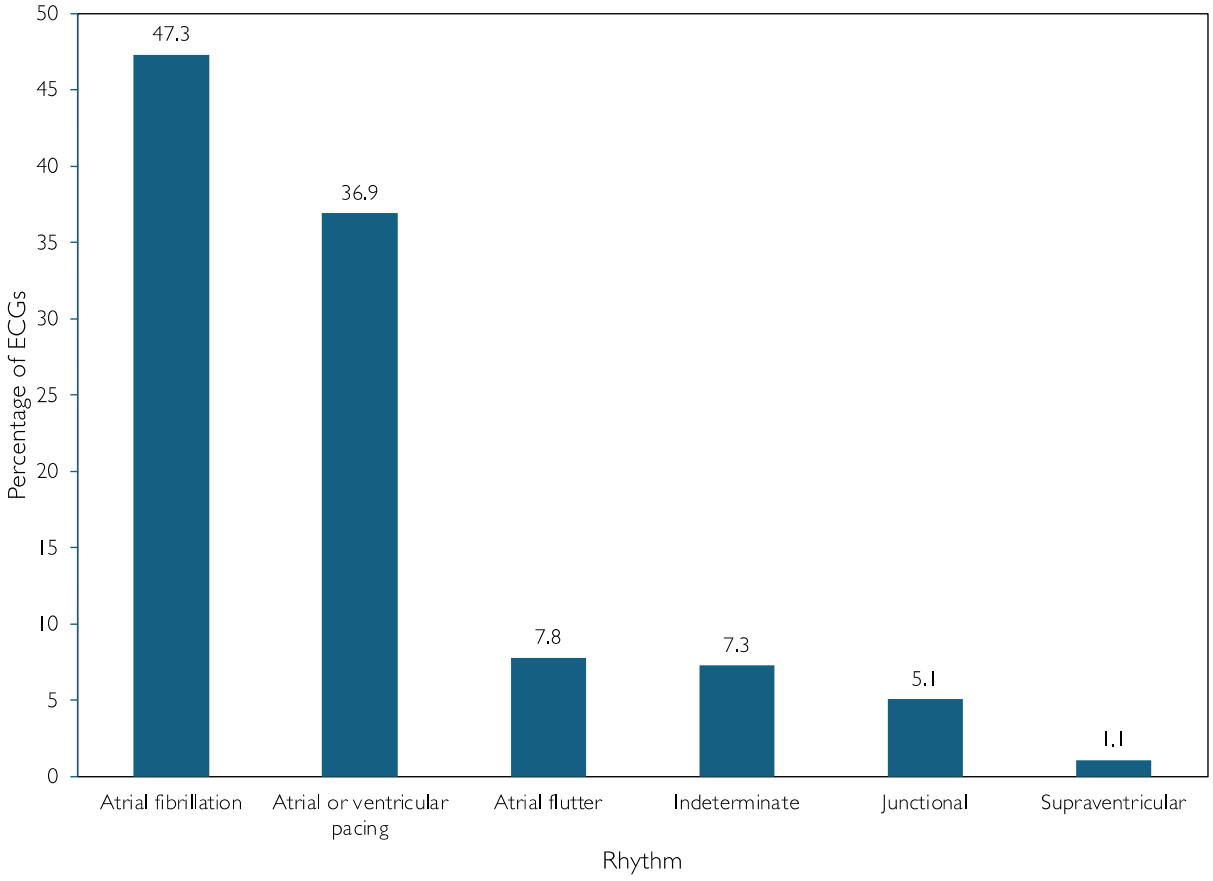
Prediction of incident atrial fibrillation using deep learning, clinical models and polygenic scores

Gilbert Jabbour et al.

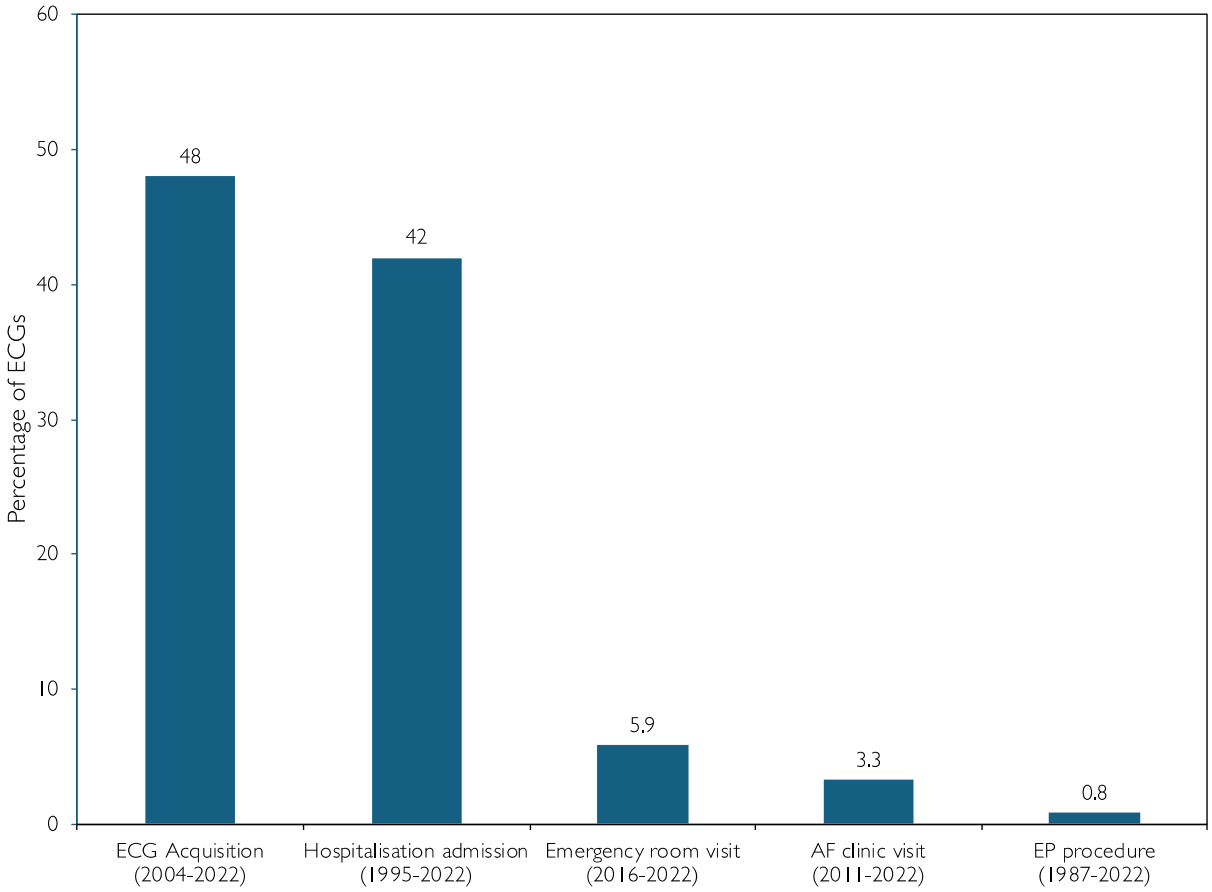
Supplementary Material included in this file:

- **Supplementary Figures 1-15**
- **Supplementary Tables 1-10**
- **Supplementary Note on ECG-AI development details, including Supplementary Table 11 and Supplementary Figure 16**

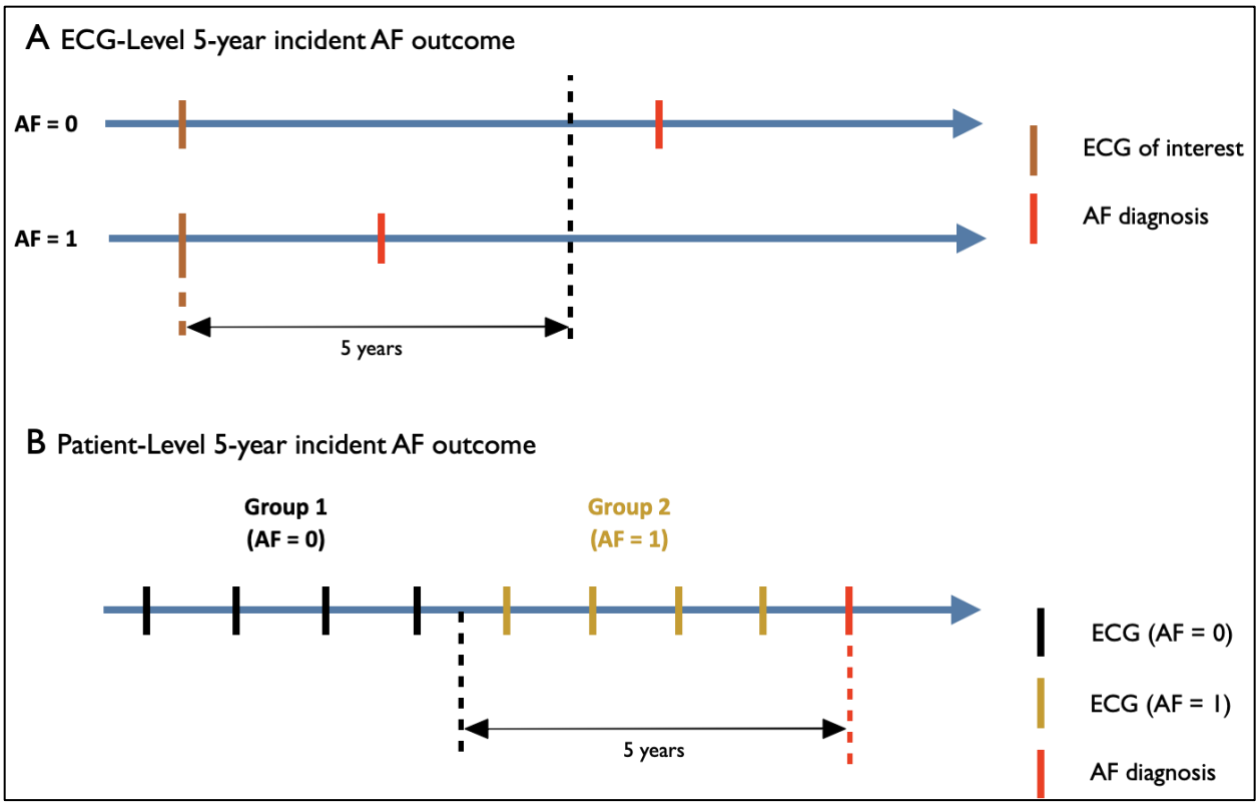
Supplementary Figures



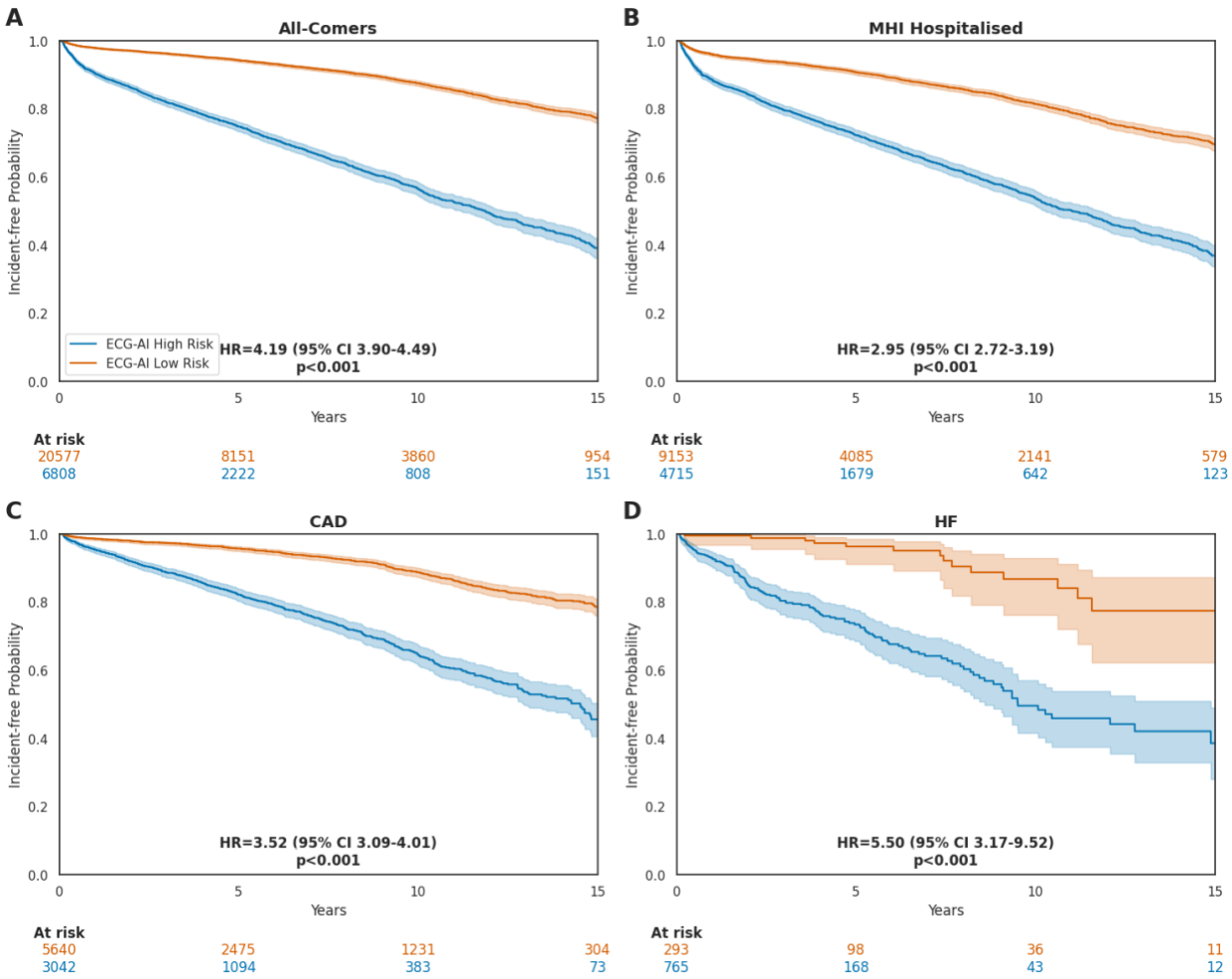
Supplementary Figure 1: Percentage of excluded ECGs which are not sinus rhythm. The sum does not amount to 100% since some categories overlap e.g. atrial fibrillation and ventricular pacing. ECGs with PAC (premature atrial contractions) or PVC (premature ventricular contractions) were included if a sinus rhythm is also present in the ECG.



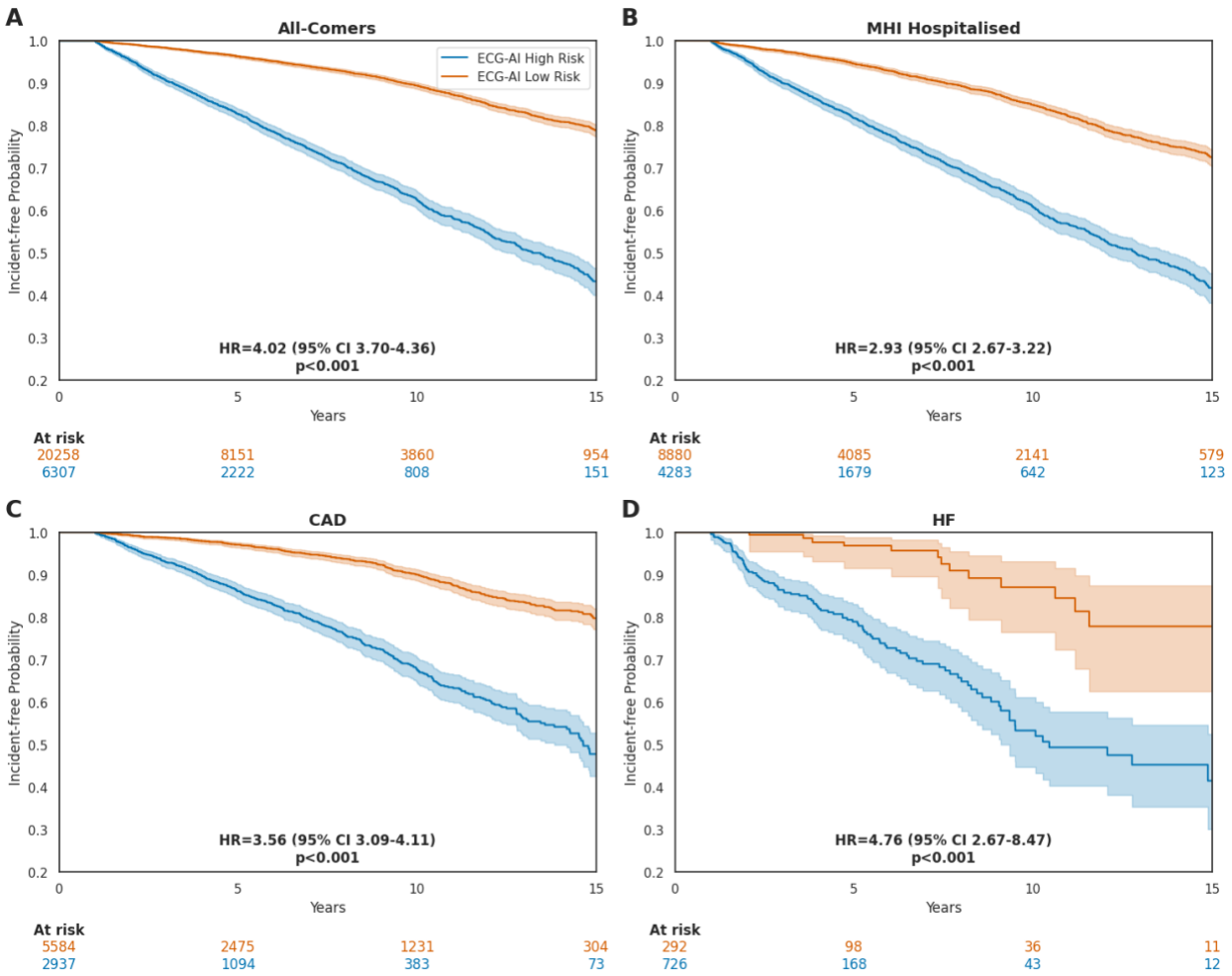
Supplementary Figure 2: Distribution of AF diagnosis date source amongst included ECGs. The numbers, such as (2004-2022), represent the range of years for the available data. For hospitalisations records including an AF diagnosis, the admission date is used as AF diagnosis date.



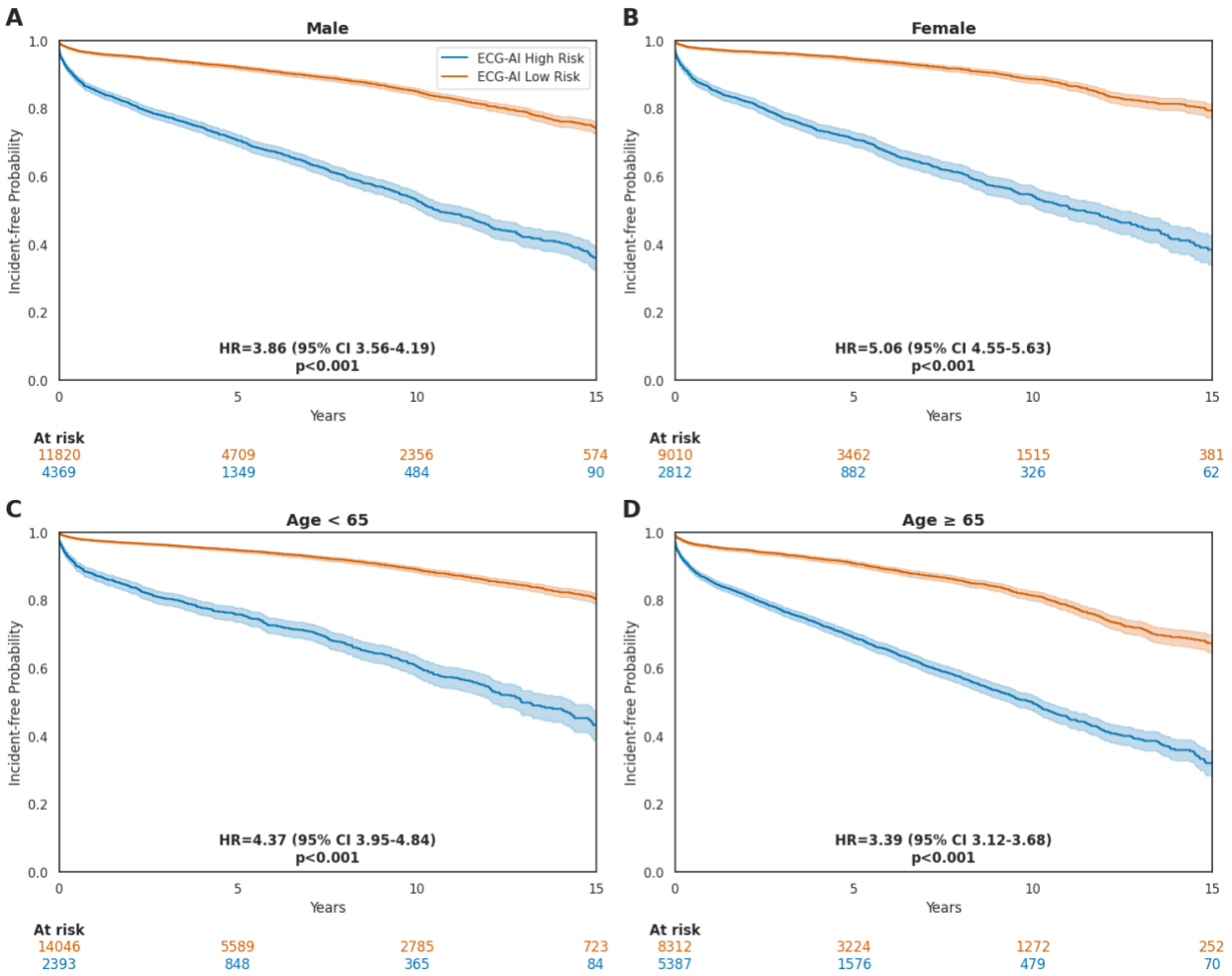
Supplementary Figure 3: Incident AF at 5 years was modelled as a binary outcome and was determined on the basis of available outpatient and inpatient clinical and medico-administrative databases and ECG diagnoses at MHI. The ECG-AI prediction at the patient level was derived by averaging the model's probability outputs for ECGs grouped according to both their 5-year AF outcome (AF=0 or AF=1) and the patient's identity.



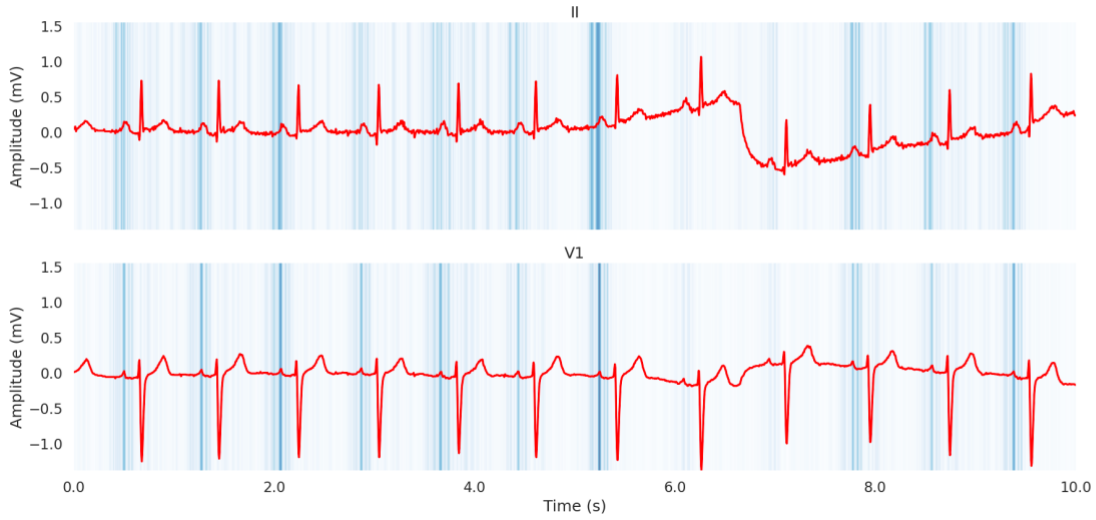
Supplementary Figure 4: Incident AF-Free Probability: Kaplan-Meier Curves using ECG-AI to stratify patients at classification threshold of 12%. Index ECGs with calculated time to AF diagnosis less than 30 days were removed. Hazard ratios were calculated by fitting a Cox proportional hazards model. P-values are calculated using the log-rank test. Panel A) KM curves patients in the “MHI All-Comers” group. Only the first ECG of each patient was used. Panel B) KM curves of patients in the “MHI Hospitalised” group. Only the first ECG of each patient was used. Panel C) KM curves of patients with a prior history of CAD. Only the first ECG acquired after the earliest record of CAD (coronary artery disease) diagnosis was used. Panel D) KM curves of patients with a prior history of HF (heart failure). Only the first ECG acquired after the earliest record of HF diagnosis was used.



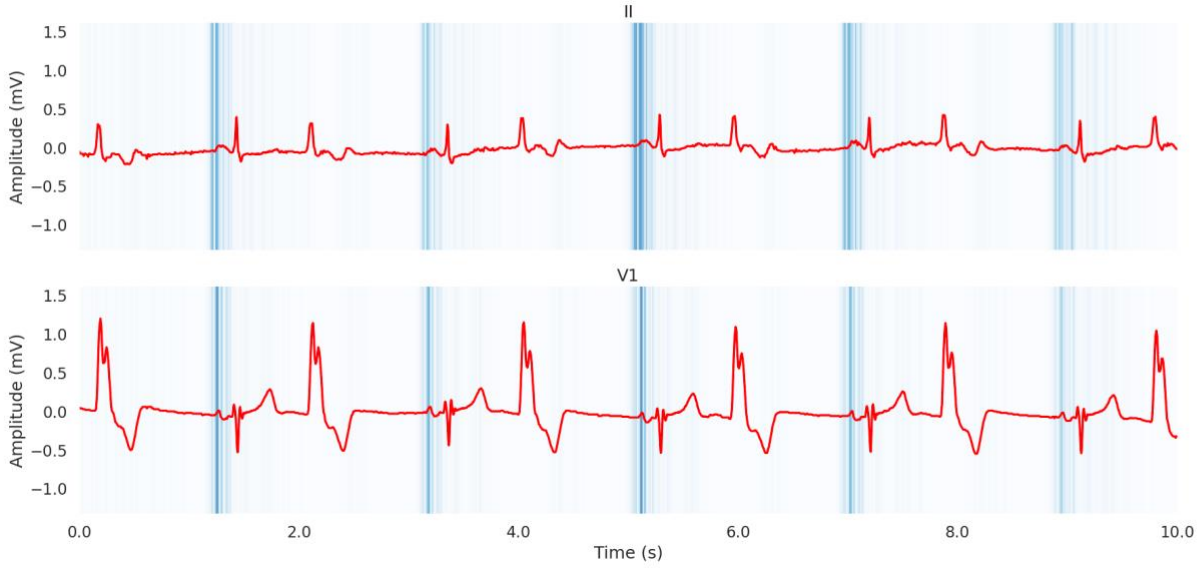
Supplementary Figure 5: Incident AF-Free Probability: Kaplan-Meier Curves using ECG-AI to stratify patients at classification threshold of 12%. ECGs with calculated time to AF diagnosis less than one year were removed. Hazard ratios were calculated by fitting a Cox proportional hazards model. P-values are calculated using the log-rank test. Panel A) KM curves patients in the “MHI All-Comers” group. Only the first ECG of each patient was used. Panel B) KM curves of patients in the “MHI Hospitalised” group. Only the first ECG of each patient was used. Panel C) KM curves of patients with a prior history of CAD. Only the first ECG acquired after the earliest record of CAD (coronary artery disease) diagnosis was used. Panel D) KM curves of patients with a prior history of HF (heart failure). Only the first ECG acquired after the earliest record of HF diagnosis was used.



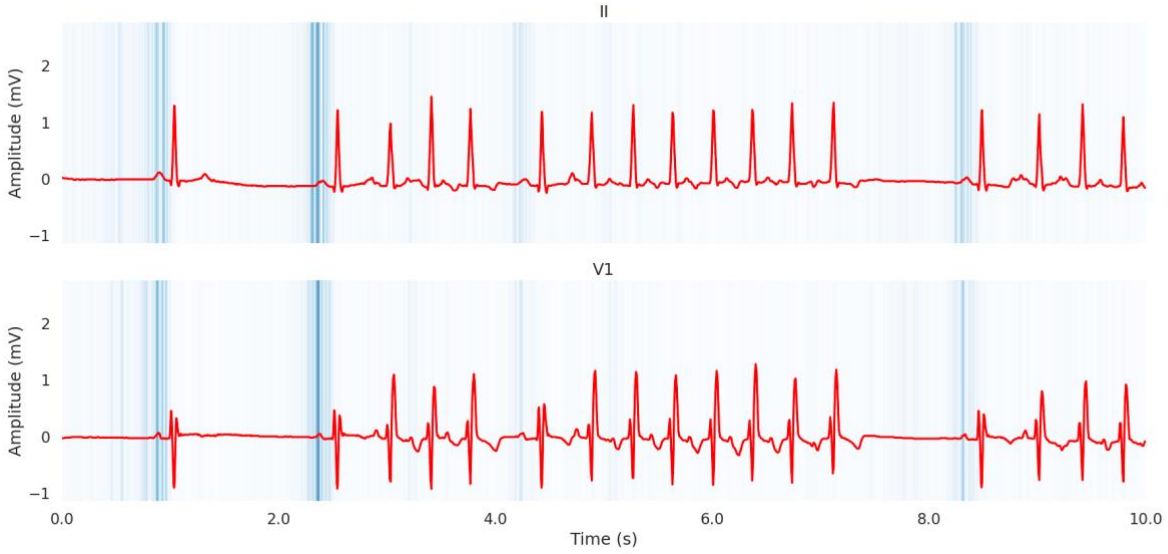
Supplementary Figure 6: Incident AF-Free Probability: Kaplan-Meier (KM) Curves using ECG-AI to stratify MHI all-comers patients at classification threshold of 12%. KM curve for subpopulations: Panel A) male sex; Panel B) female sex; Panel C) Age < 65; Panel D) Age ≥ 65. Only the first ECG of each patient was used. Index ECGs with calculated time to AF diagnosis of zero were removed. Hazard ratios were calculated by fitting a Cox proportional hazards model. P-values are calculated using the log-rank test.



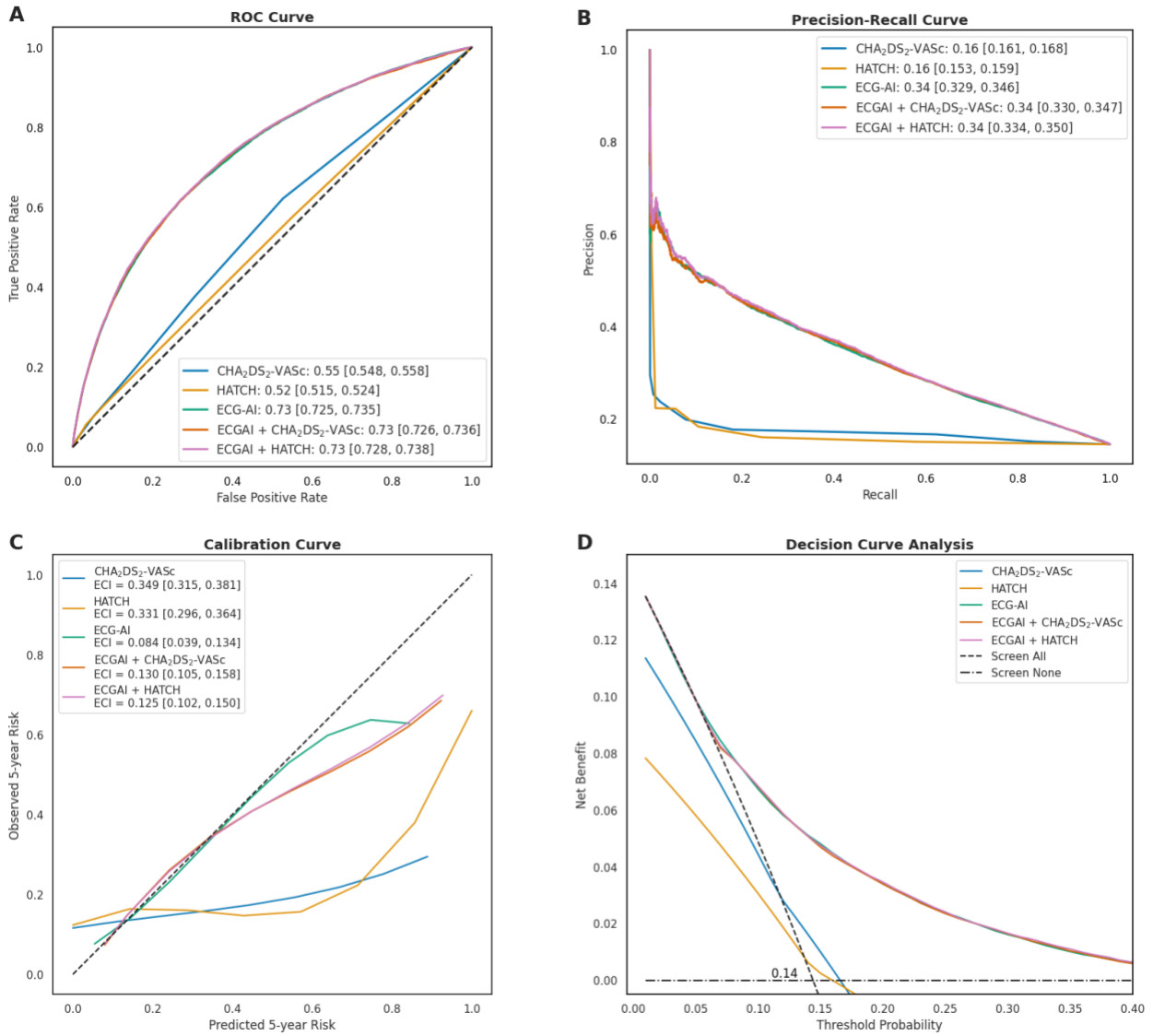
Supplementary Figure 7: ECG-AI saliency map applied on an ECG waveform including an artifact, which does not appear to be associated with a saliency.



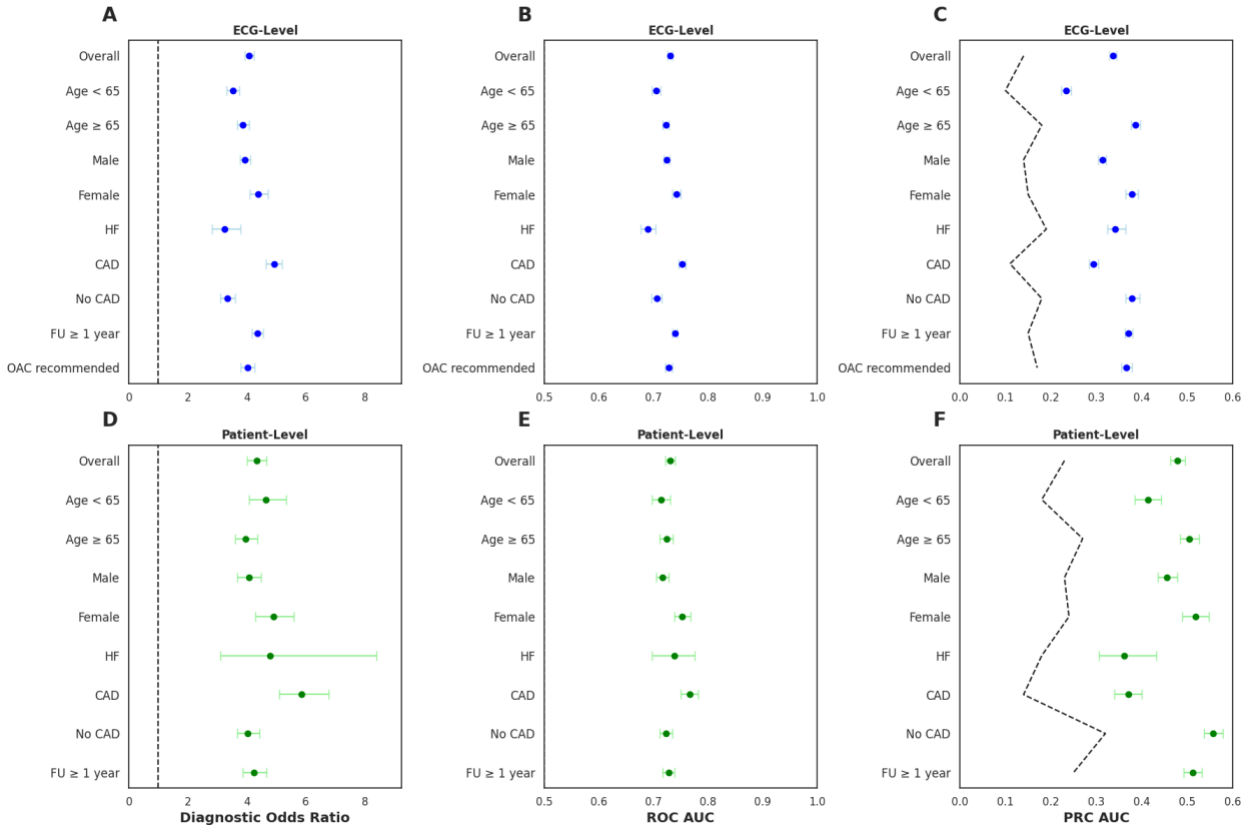
Supplementary Figure 8: ECG-AI saliency map applied on an ECG waveform with a ventricular bigeminy. Premature ventricular beats do not appear to be associated with a saliency.



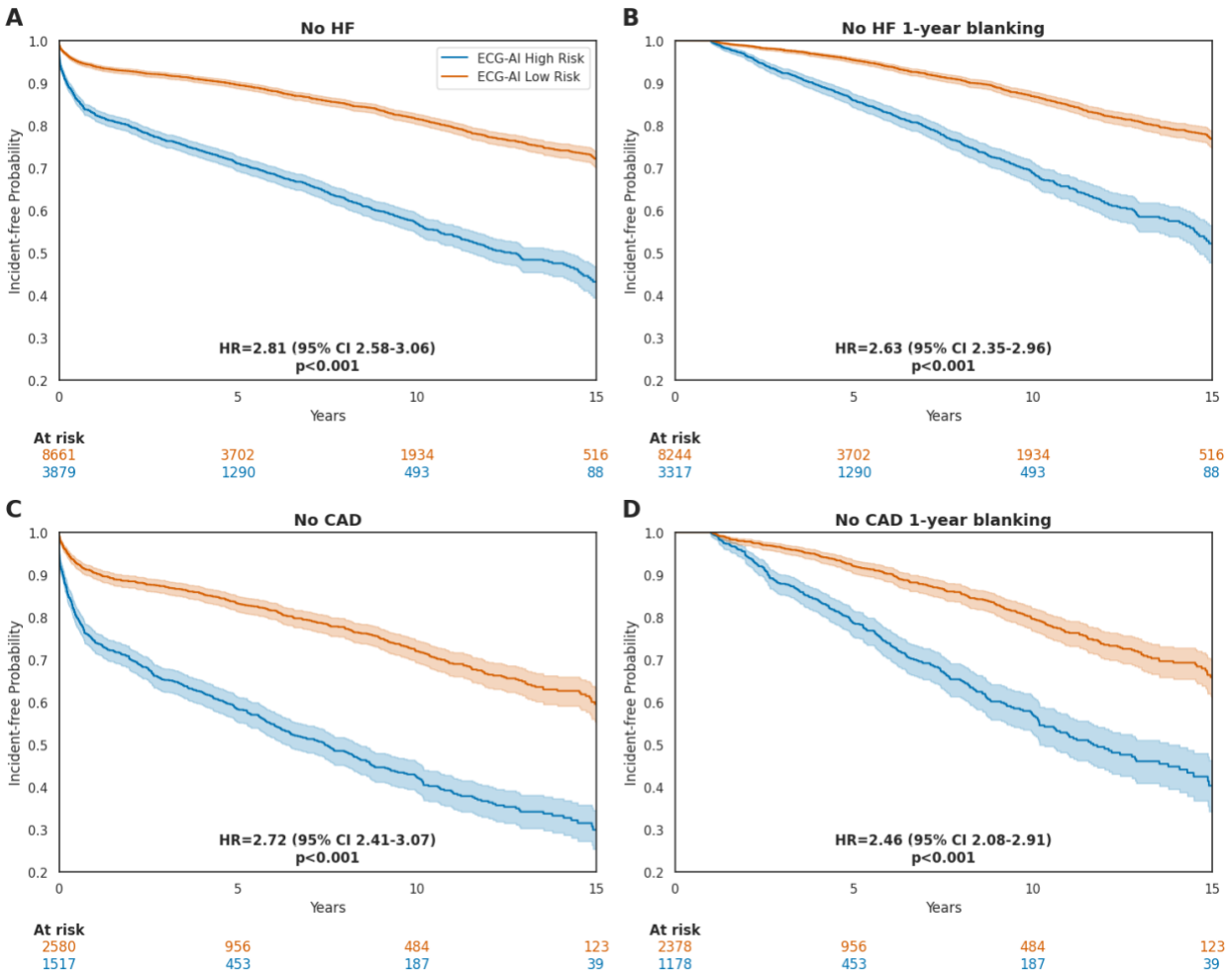
Supplementary Figure 9: ECG-AI saliency map applied on an ECG waveform including a normal sinus rhythm and an atrial tachycardia conducted with aberrancy. The atrial tachycardia does not appear to be associated with a saliency.



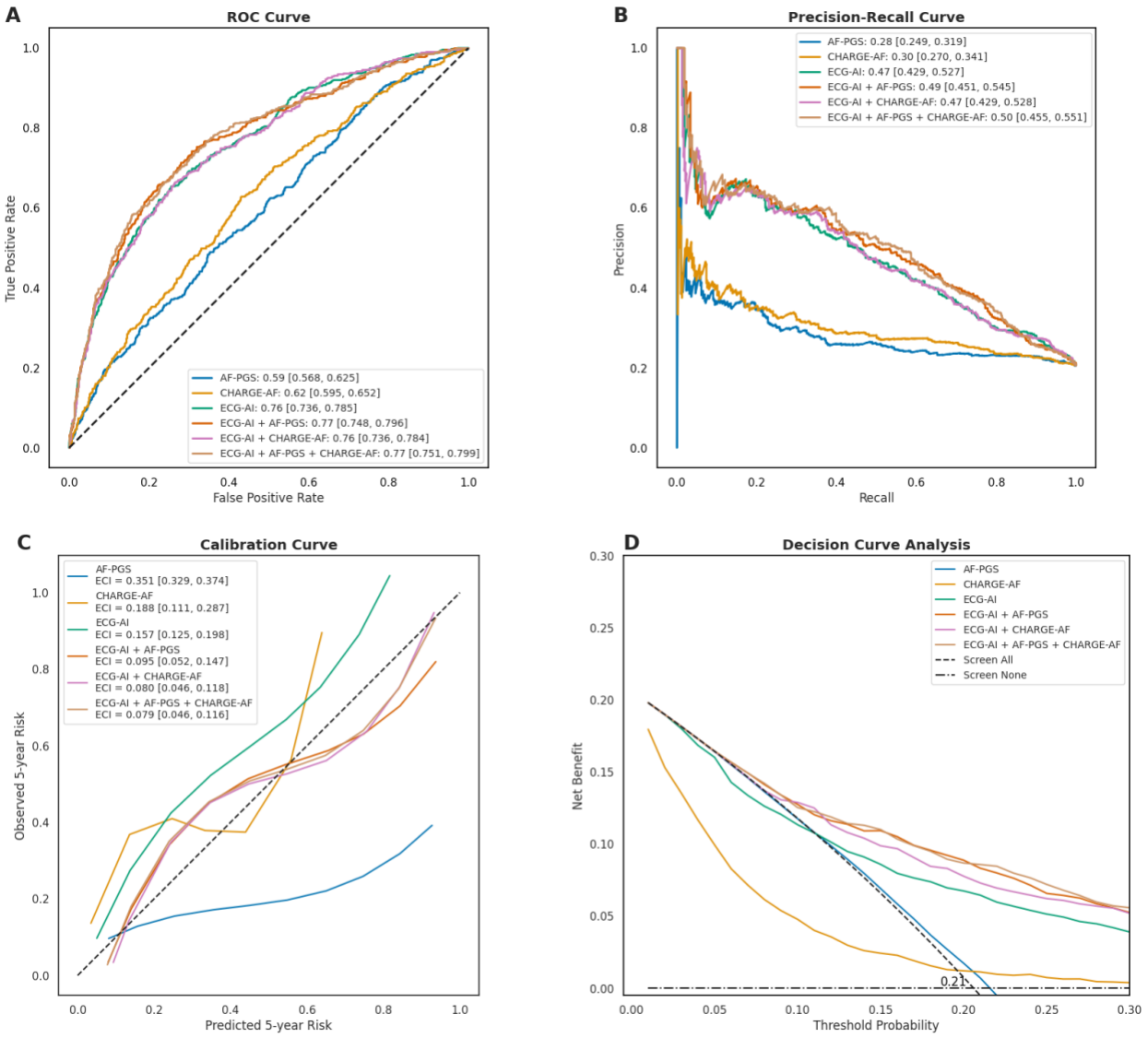
Supplementary Figure 10: MHI Hospitalised test set (15,271 patients, 96,578 ECG) performance assessment of five models: 1) CHA_2D_2-VASc, 2) HATCH, 3) ECG-AI, 4) ECG-AI + CHA2D2-VASc, and 5) ECGAI + HATCH. Panel A shows the ROC Curve, plotting the True Positive Rate against the False Positive Rate for each model, with the area under the curve (AUC) indicating discriminatory power. Panel B displays the Precision-Recall Curve, plotting precision against recall, with the AUC in the legend. Panel C presents the Calibration Curve, showing the relationship between predicted and observed 5-year AF risk; the slope and intercept are calculated using linear regression, and the curve is plotted using a univariate spline with smoothing factor of 1. The Estimated Calibration Index (ECI) is the root mean squared difference between the mean predicted probabilities and the spline-fitted calibration curve. Panel D illustrates the Decision Curve Analysis, plotting net benefit against threshold probability.



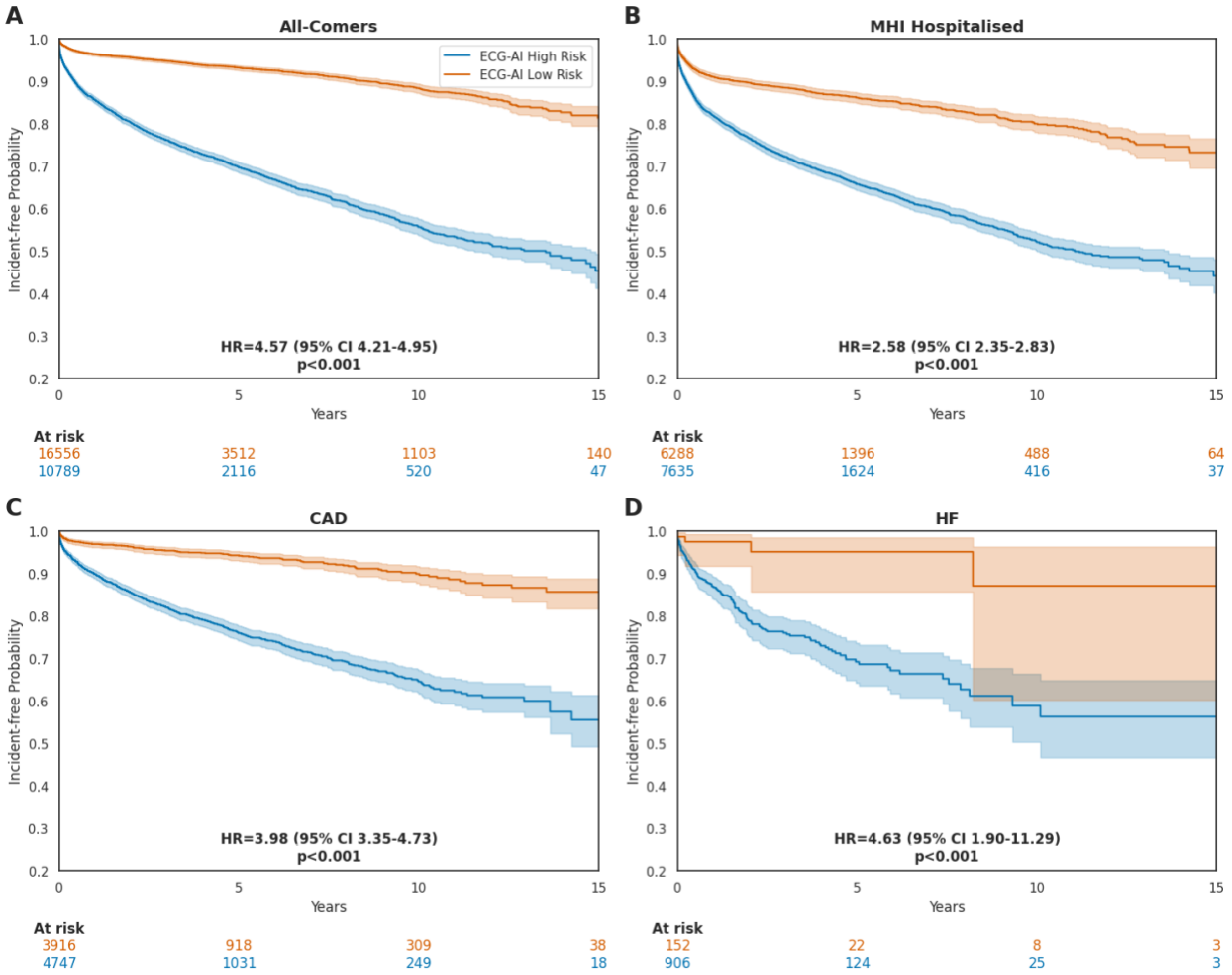
Supplementary Figure 11: ECG-AI discrimination performance metrics for the MHI Hospitalised test set (15,271 patients, 96,578 ECG) at the ECG-level (panels A, B, C) and patient-level (panels D, E, F). Panels A and D show the Diagnostic Odds Ratio (DOR) which is calculated at an optimal threshold of 12% for ECG-level and 15% for patient-level. Panels B and E display the ROC AUC (Receiver Operating Characteristic Area Under the Curve). Panels C and F present the PRC AUC (Precision-Recall Curve Area Under the Curve). The dashed lines in the PRC AUC panels represent prevalence, indicating the proportion of true positive cases within the population. Only the first ECG acquired after the earliest record of CAD (coronary artery disease) diagnosis or HF (heart failure) were respectively included in the CAD and HF subgroups. Confidence intervals for all metrics were derived from 1000 bootstrap iterations. Oral Anticoagulation (OAC) is recommended for CHA_2DS_2-VASc ≥ 2 for males and CHA_2DS_2-VASc ≥ 3 for females.(49, 50). Acronyms: FU (Follow-Up).



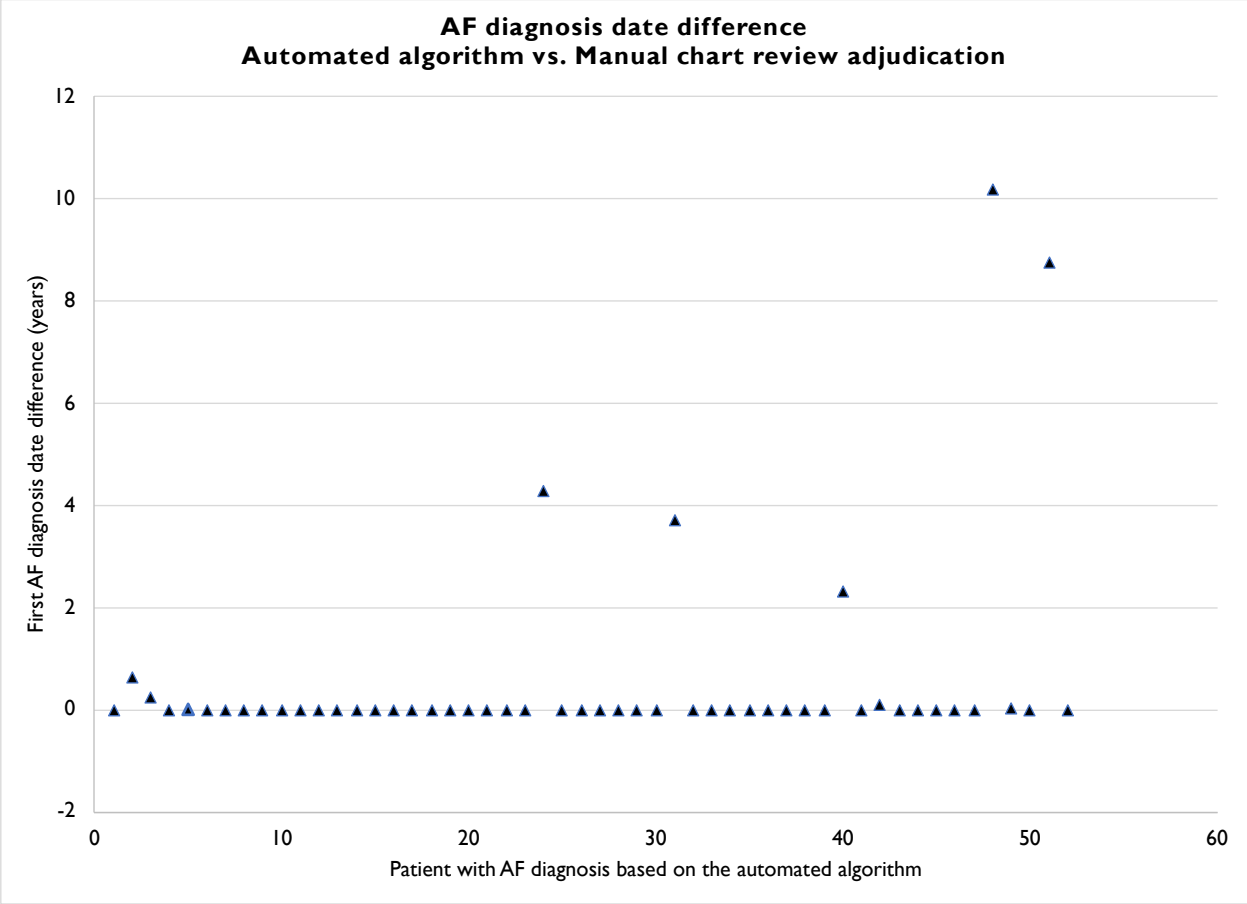
Supplementary Figure 12: Incident Free AF Probability: Kaplan-Meier Curves using ECG-AI to stratify patients at classification threshold of 12%. Hazard ratios were calculated by fitting a Cox proportional hazards model. Only the first ECG of each patient was used. P-values are calculated using the log-rank test. Panel A) KM curves patients in patients without a documented diagnosis of HF. Panel B) KM curves patients in patients without a documented diagnosis of HF. ECGs with calculated time to AF diagnosis less than one year were removed. Panel C) KM curves of patients without a documented diagnosis of CAD. Panel D) KM curves of patients with a prior history of CAD. ECGs with calculated time to AF diagnosis less than one year were removed.



Supplementary Figure 13: MHI Biobank test set (2,301 patients) performance assessment of six models: 1) AF Polygenic risk score (PGS), 2) CHARGE-AF score, 3) ECG-AI, 4) ECG-AI + AF-PGS, 5) ECG-AI + CHARGE-AF, and 6) ECG-AI + AF-PGS + CHARGE-AF. Panel A) shows the ROC Curve, plotting the True Positive Rate against the False Positive Rate for each model, with the area under the curve (AUC) indicating discriminatory power. Panel B) displays the Precision-Recall Curve, plotting precision against recall. Panel C) presents the Calibration Curve, showing the relationship between predicted and observed 5-year AF risk; the slope and intercept are calculated using linear regression, and the curve is plotted using a univariate spline with smoothing factor of 1. The Estimated Calibration Index (ECI) is the root mean squared difference between the mean predicted probabilities and the spline-fitted calibration curve. Panel D) illustrates the Decision Curve Analysis, plotting net benefit against threshold probability.



Supplementary Figure 14: Incident Free AF Probability: Kaplan-Meier Curves using ECG-AI to stratify patients at classification threshold of 12%. Index ECGs with calculated time to AF diagnosis of zero were removed. Hazard ratios were calculated by fitting a Cox proportional hazards model. P-values are calculated using the log-rank test. Panel A) KM curves patients in the “MHI All-Comers” group. The ECG with the highest predicted probability of AF was chosen. Panel B) KM curves of patients in the “MHI Hospitalised” group. Only the first ECG of each patient was used. Panel C) KM curves of patients with a prior history of CAD. Only the first ECG acquired after the earliest record of CAD (coronary artery disease) diagnosis was used. Panel D) KM curves of patients with a prior history of HF (heart failure). Only the first ECG acquired after the earliest record of HF diagnosis was used.



Supplementary Figure 15: Automated Incident AF Algorithm Validation. Gold standard = Manual review of patient records by a medical resident. A positive difference indicated the AF diagnosis date as determined by the automated algorithm is more recent than the date determined using manual chart review. In all five cases, the actual AF diagnosis date preceded the first time it was captured in the MHI EHR because the AF was diagnosed at different centres (2 patients) or the diagnosis occurred before the time range covered by the EHR examination (3 patients) (Supplementary Figure 2). The AF diagnosis date comparison was only performed for the 52 patients with AF diagnosis in the validation study (Supplementary Table 7)

Supplementary Tables

Supplementary Table 1: Description of the MHI All-Comers dataset overall and by split

-	Overall	Training	Validation	Test
ECG	669782 (100.0%)	467638 (69.8%)	66600 (9.9%)	135544 (20.2%)
Patients	145323 (100.0%)	101726 (70.0%)	14532 (10.0%)	29065 (20.0%)
ECG per Patient	2.0 (Q1: 1.0, Q3: 5.0)	2.0 (Q1: 1.0, Q3: 5.0)	2.0 (Q1: 1.0, Q3: 5.0)	2.0 (Q1: 1.0, Q3: 5.0)
ECG Level Data				
Age [years]	62.8 (± 14.8)	62.8 (± 14.8)	62.8 (± 14.7)	62.8 (± 14.8)
Male	415956 (62.1%)	290969 (62.2%)	40820 (61.3%)	84167 (62.1%)
CIMD	3.2 (± 1.4)	3.2 (± 1.4)	3.2 (± 1.4)	3.2 (± 1.4)
Hospitalization	475986 (71.1%)	331963 (71.0%)	47445 (71.2%)	96578 (71.3%)
Follow-up [years]	4.2 (Q1: 1.2, Q3: 8.3)	4.2 (Q1: 1.2, Q3: 8.3)	4.2 (Q1: 1.2, Q3: 8.4)	4.2 (Q1: 1.2, Q3: 8.3)
5-year incident AF [years]	80183 (12.0%)	56281 (12.0%)	8071 (12.1%)	15831 (11.7%)
Time to incident AF [years]	2.0 (Q1: 0.1, Q3: 5.6)	2.0 (Q1: 0.1, Q3: 5.5)	2.0 (Q1: 0.1, Q3: 5.6)	2.2 (Q1: 0.2, Q3: 5.9)
Patient Level Data				
Age [years]	61.3 (± 15.2)	61.3 (± 15.2)	61.1 (± 15.1)	61.3 (± 15.2)
Male	84087 (57.9%)	58861 (57.9%)	8371 (57.6%)	16855 (58.0%)
CIMD	3.2 (± 1.4)	3.2 (± 1.3)	3.2 (± 1.4)	3.2 (± 1.4)
Hospitalization	76680 (52.8%)	53724 (52.8%)	7685 (52.9%)	15271 (52.5%)
Follow-up [years]	3.2 (Q1: 0.3, Q3: 8.6)	3.3 (Q1: 0.3, Q3: 8.6)	3.2 (Q1: 0.3, Q3: 8.6)	3.2 (Q1: 0.3, Q3: 8.6)
5-year incident AF [years]	22695 (15.6%)	15855 (15.6%)	2245 (15.4%)	4595 (15.8%)

Supplementary Table 2: Description of the MHI Hospitalised dataset overall and by split

MHI Hospitalised	Overall	Training	Validation	Test
ECG	475986 (100.0%)	331963 (69.7%)	47445 (10.0%)	96578 (20.3%)
Patients	76680 (100.0%)	53724 (70.1%)	7685 (10.0%)	15271 (19.9%)
ECG per Patient	3.0 (Q1: 2.0, Q3: 8.0)	3.0 (Q1: 2.0, Q3: 8.0)	3.0 (Q1: 2.0, Q3: 8.0)	4.0 (Q1: 2.0, Q3: 8.0)
ECG Level Data				
Age [years]	64.2 (± 14.1)	64.2 (± 14.1)	64.2 (± 14.0)	64.2 (± 14.1)
Male	317725 (66.8%)	222376 (67.0%)	31112 (65.6%)	64237 (66.5%)
CIMD	3.2 (± 1.4)	3.2 (± 1.4)	3.2 (± 1.4)	3.2 (± 1.4)
Hospitalization	475986 (100.0%)	331963 (100.0%)	47445 (100.0%)	96578 (100.0%)
Follow-up [years]	4.3 (Q1: 1.1, Q3: 8.6)	4.3 (Q1: 1.1, Q3: 8.6)	4.4 (Q1: 1.1, Q3: 8.7)	4.4 (Q1: 1.2, Q3: 8.6)
5-year incident AF [years]	70230 (14.8%)	49176 (14.8%)	7144 (15.1%)	13910 (14.4%)
Time to incident AF [years]	1.9 (Q1: 0.1, Q3: 5.5)	1.9 (Q1: 0.1, Q3: 5.4)	1.8 (Q1: 0.1, Q3: 5.4)	2.1 (Q1: 0.1, Q3: 5.9)
Patient Level Data				
Age [years]	64.3 (± 13.7)	64.4 (± 13.7)	64.0 (± 13.7)	64.4 (± 13.7)
Male	50114 (65.4%)	35098 (65.3%)	4999 (65.0%)	10017 (65.6%)
CIMD	3.2 (± 1.3)	3.2 (± 1.3)	3.2 (± 1.4)	3.2 (± 1.4)
Hospitalization	76680 (100.0%)	53724 (100.0%)	7685 (100.0%)	15271 (100.0%)
Follow-up [years]	4.0 (Q1: 0.3, Q3: 9.7)	4.0 (Q1: 0.3, Q3: 9.7)	4.0 (Q1: 0.3, Q3: 9.9)	4.0 (Q1: 0.3, Q3: 9.7)
5-year incident AF [years]	18492 (24.1%)	12909 (24.0%)	1849 (24.1%)	3734 (24.5%)

Supplementary Table 3: Description of the MHI Biobank dataset overall and by split

MHI Biobank	Overall	Training	Validation	Test
ECG	104850 (100.0%)	72851 (69.5%)	10406 (9.9%)	21593 (20.6%)
Patients	11622 (100.0%)	8097 (69.7%)	1155 (9.9%)	2370 (20.4%)
ECG per Patient	6.0 (Q1: 3.0, Q3: 12.0)	6.0 (Q1: 3.0, Q3: 12.0)	6.0 (Q1: 2.0, Q3: 12.0)	6.0 (Q1: 3.0, Q3: 12.0)
ECG Level Data				
Age [years]	64.0 (± 12.0)	63.9 (± 12.1)	64.7 (± 11.9)	64.2 (± 12.0)
Male	70496 (67.2%)	49671 (68.2%)	6694 (64.3%)	14131 (65.4%)
CIMD	3.3 (± 1.4)	3.3 (± 1.4)	3.3 (± 1.3)	3.3 (± 1.4)
Hospitalization	85617 (81.7%)	59348 (81.5%)	8559 (82.3%)	17710 (82.0%)
Follow-up [years]	6.6 (Q1: 3.2, Q3: 10.5)	6.7 (Q1: 3.2, Q3: 10.6)	6.6 (Q1: 3.1, Q3: 10.3)	6.6 (Q1: 3.1, Q3: 10.5)
5-year incident AF [years]	13772 (13.1%)	9727 (13.4%)	1253 (12.0%)	2792 (12.9%)
Time to incident AF [years]	3.3 (Q1: 0.7, Q3: 7.0)	3.1 (Q1: 0.7, Q3: 6.8)	3.3 (Q1: 0.7, Q3: 7.1)	3.7 (Q1: 0.8, Q3: 7.2)
Patient Level Data				
Age [years]	63.2 (± 11.6)	63.2 (± 11.6)	63.2 (± 11.9)	63.3 (± 11.6)
Male	7326 (63.0%)	5106 (63.1%)	737 (63.8%)	1483 (62.6%)
CIMD	3.2 (± 1.4)	3.2 (± 1.4)	3.2 (± 1.4)	3.2 (± 1.4)
Hospitalization	7919 (68.1%)	5527 (68.3%)	790 (68.4%)	1602 (67.6%)
Follow-up [years]	9.6 (Q1: 4.6, Q3: 13.6)	9.6 (Q1: 4.7, Q3: 13.7)	9.4 (Q1: 4.2, Q3: 13.7)	9.4 (Q1: 4.7, Q3: 13.6)
5-year incident AF [years]	2846 (24.5%)	1976 (24.4%)	258 (22.3%)	612 (25.8%)

Supplementary Table 4: Medical conditions alongside their corresponding ICD-9 and ICD-10 codes for patient data extraction from the hospitalisation's records. ICD-9 codes are from the International Classification of Diseases, 9th Revision, while ICD-10 codes are from the 10th Revision. Some codes may differ from the original ICD codes because of local practice.

Condition	ICD-9 and ICD-10 codes
Atrial fibrillation	['427.31','I48.0','I48.00','I48.01','I48.90','I48.02','427.3Z','I97.13']
Atrial flutter	['427.32','I48.1','I48.2','I48.3','I48.4','I48.91']
Heart Failure	['428','I50']
Coronary artery disease	['410','411','412','413','414','I20','I21','I22','I25']
Chronic obstructive pulmonary disease	['491','492','496','J44']
Hypertension	['401','402','403','404','405','I10','I11','I12','I13','I15']
Diabetes	['250','E10','E11','E12','E13','E14']
Stroke	['434','435','436','I63','I64']
Dyslipidemia	['272','E78']
Obesity	['278.0','E66']
Chronic kidney disease	['585','586','N18','N19']
Sleep apnea	['327.2','780.57','G47.3']
Hyperthyroidism	['242','E05']
Vascular disease	['440','433','434','I65','I66','I70']

Supplementary Table 5: Clinical Risk Models

Model	Covariates	Description
Age & Sex	Age, Sex	Age and Sex separately added in the logistic regression
HATCH	(1) Hypertension (1) Age>75 (2) Stroke or TIA (1) COPD (2) Heart failure	Score 1 to 7
CHA_2DS_2-VASc	(1) Congestive heart failure (1) HTN (1-2) Age (>75=2 points, >65=1 point) (1) DM (2) Stroke or TIA (1) Vascular disease (1) Sex (female)	Score 1 to 9
CHARGE-AF	Age, race, height, weight, blood pressure, current smoking, antihypertensive medication use, diabetes, heart failure, myocardial infarction	Supplementary Table 6

Supplementary Table 6: The 5-year risk for the simple CHARGE-AF model is calculated as $1-0.9718412736^{\exp(\Sigma\beta X - 12.5815600)}$ where β is the regression coefficient and X is the level for each risk factor. When calculating the 5-year risk, estimated β for age, height, weight, systolic and diastolic blood pressure must be divided by the number of presented units. Adapted from (11).

Variable	β
Age (5 years)	0.508
Race (white)	0.465
Height (10 cm)	0.248
Weight (15 kg)	0.115
Systolic BP (20 mm Hg)	0.197
Diastolic BP (10 mm Hg)	−0.101
Smoking (current)	0.359
Antihypertensive medication use (Yes)	0.349
Diabetes (Yes)	0.237
Heart failure (Yes)	0.701
Myocardial infarction (Yes)	0.496

Supplementary Table 7: Automated Incident AF Algorithm Validation. Gold standard = Manual review of patient records by a medical resident. Specificity = 100%, Sensitivity = 91% (95% CI: 83.9-98.6). CI=95% Confidence interval calculated using the normal approximation method. The AF diagnosis dates in the 52 patients with AF determined by the automated algorithm considering selected databases, as described in the “Outcome” section, were compared with those from manual chart review adjudication and provided in Supplementary Figure 15.

		Automated Algorithm	
		No AF	AF
Manual Adjudication	No AF	143	0
	AF	5	52

Supplementary Table 8: Discrimination performance for the tested models. The metrics are reported with a 95% confidence interval derived using bootstrapping with 1000 iterations. Acronyms: ROC=Area under the Receiver Operating Characteristic; PRC=Area under the Precision-Recall Curve; ECI=Estimated Calibration Index; LR=Logistic Regression; N=Sample size.

Cohort	Model	N	Events	5-year AF outcome	ROC	PRC	ECI
All-Comers (Patient-level)	ECG-AI	29,065 patients	4,595	15.8%	0.78 (0.768 – 0.783)	0.42 (0.405 – 0.436)	0.095 (0.070 – 0.136)
All-Comers (ECG-level)	ECG-AI	135,544 ECG	15,831	11.7%	0.75 (0.745 – 0.753)	0.31 (0.303 – 0.318)	0.086 (0.050 – 0.125)
All-Comers (ECG-level)	Age & Sex	135,544 ECG	15,831	11.7%	0.63 (0.627 – 0.636)	0.17 (0.168 – 0.176)	0.349 (0.349 – 0.349)
All-Comers (ECG-level)	ECG-AI + Age & Sex	135,544 ECG	15,831	11.7%	0.75 (0.746 – 0.754)	0.31 (0.306 – 0.320)	0.124 (0.095 – 0.159)
Hospitalised (ECG-level)	ECG-AI	96,578 ECG	13,910	14.4%	0.73 (0.725 – 0.735)	0.34 (0.329 – 0.346)	0.084 (0.039 – 0.134)
Hospitalised (ECG-level)	HATCH	96,578 ECG	13,910	14.4%	0.52 (0.515 – 0.524)	0.16 (0.153 – 0.159)	0.331 (0.296 – 0.364)
Hospitalised (ECG-level)	CHA_2DS_2-VASc	96,578 ECG	13,910	14.4%	0.55 (0.548 – 0.558)	0.16 (0.161 – 0.168)	0.349 (0.315 – 0.381)
Hospitalised (ECG-level)	ECG-AI + HATCH	96,578 ECG	13,910	14.4%	0.73 (0.728 – 0.738)	0.34 (0.334 – 0.350)	0.125 (0.102 – 0.150)
Hospitalised (ECG-level)	ECG-AI + CHA_2DS_2-VASc	96,578 ECG	13,910	14.4%	0.73 (0.726 – 0.736)	0.34 (0.330 – 0.347)	0.130 (0.105 – 0.158)
Biobank (Patient-level)	ECG-AI	2,301 patients	474	20.6%	0.76 (0.736	0.47 (0.429	0.157 (0.125

Cohort	Model	N	Events	5-year AF outcome	ROC	PRC	ECI
					– 0.785)	– 0.527)	– 0.198)
Biobank (Patient-level)	CHARGE-AF	2,301 patients	474	20.6%	0.62 (0.595 – 0.652)	0.30 (0.270 – 0.341)	0.188 (0.111 – 0.287)
Biobank (Patient-level)	AF-PGS	2,301 patients	474	20.6%	0.59 (0.568 – 0.625)	0.28 (0.249 – 0.319)	0.351 (0.329 – 0.374)
Biobank (Patient-level)	ECG-AI + CHARGE-AF	2,301 patients	474	20.6%	0.76 (0.736 – 0.784)	0.47 (0.429 – 0.528)	0.080 (0.046 – 0.118)
Biobank (Patient-level)	ECG-AI + AF-PGS	2,301 patients	474	20.6%	0.77 (0.748 – 0.796)	0.49 (0.451 – 0.545)	0.095 (0.052 – 0.147)
Biobank (Patient-level)	ECG-AI + AF-PGS + CHARGE-AF	2,301 patients	474	20.6%	0.77 (0.751 – 0.799)	0.50 (0.455 – 0.551)	0.079 (0.046 – 0.116)
MIMIC-IV (Patient-level)	ECG-AI	109,870 patients	16,610	15.1%	0.77 (0.765 – 0.773)	0.39 (0.386 – 0.402)	0.071 (0.060 – 0.099)
MIMIC-IV (ECG-level)	ECG-AI	437,323 ECG	65,301	14.9%	0.71 (0.704 – 0.708)	0.30 (0.296 – 0.303)	0.141 (0.126 – 0.158)

Supplementary Table 9: MHI-All comers performance results at the patient level on the test set calculated for different classification thresholds. PPV=Positive predictive value; NPV=Negative predictive value.

Probability Threshold	Sensitivity	Specificity	PPV	NPV
0.09	0.75	0.65	0.27	0.94
0.11	0.71	0.70	0.30	0.93
0.12	0.66	0.75	0.32	0.93
0.14	0.60	0.80	0.35	0.92
0.17	0.53	0.85	0.39	0.91
0.21	0.43	0.90	0.44	0.90
0.29	0.30	0.95	0.51	0.88

Supplementary Table 10: Discrimination performance sensitivity analysis as function of maximum follow-up. The area under the ROC is reported with a 95% confidence interval derived using boostrapping with 1000 iterations. Eligible follow-up encounters encompass ECG acquisitions, hospitalisations, emergency room visits, AF clinic visits, and electrophysiology procedures. ECG-AI performance are reported at the ECG-level.

Model	MHI group	Overall	Follow-up > 1 year
ECG-AI	All-Comers	0.749 (0.745 - 0.753)	0.755 (0.751 - 0.760)
CHA_2DS_2-VASc	Hospitalised	0.55 (0.548 – 0.558)	0.56 (0.553 – 0.564)
HATCH	Hospitalised	0.52 (0.515 – 0.524)	0.53 (0.519 – 0.531)
CHARGE-AF	Biobank	0.62 (0.595 – 0.652)	0.63 (0.599 – 0.661)
AF-PGS	Biobank	0.59 (0.568 – 0.625)	0.59 (0.564 – 0.625)

Supplementary Note

The ResNet-50 model was implemented using TensorFlow (version 2.9.1) with various features and optimizations to ensure effective, reproducible and replicable training. Firstly, a Conda environment was defined to ensure the compatibility of library versions. Mixed precision training was enabled by setting the global policy to “mixed_float16”, leveraging the benefits of reduced memory usage and improved performance. To ensure reproducibility of the results, the random seed was set to 42 using “tf.random.set_seed(42)”, “np.random.seed(42)”, and “random.seed(42)”.
The model architecture was built using the “Classifiers” module from “classification_models_1D” library with an Adam optimizer and binary cross-entropy loss function. The model hyperparameters were dynamically set based on the YAML configuration.

The logic for reducing the learning rate (LR) and early stopping was implemented to enhance training efficiency and prevent overfitting. The “ReduceLROnPlateau” callback from “tensorflow.keras.callbacks” was configured to monitor the validation loss and reduce the learning rate by a factor of 0.1 if no improvement was observed for 3 consecutive epochs. This approach helps in fine-tuning the learning process by lowering the learning rate when the model hits a plateau, allowing for more precise weight updates.

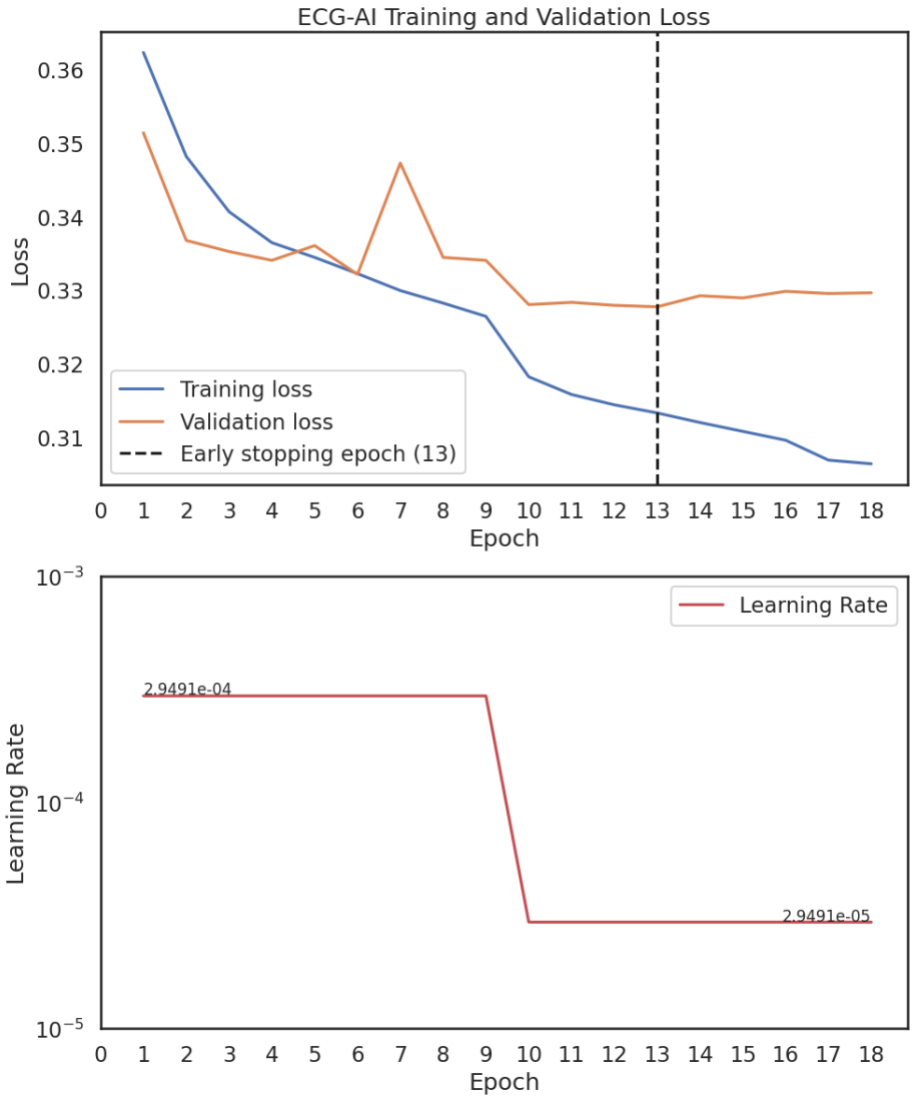
Early stopping was implemented using the Keras (version 2.9.0) “EarlyStopping” callback, which monitored the validation loss and stopped the training process if no improvement was observed for 5 consecutive epochs, with the “restore_best_weights”

parameter set to “True”. This minimizes overfitting and saves computational resources by halting unnecessary training iterations, ensuring the best model weights are retained. To initialize the training process, a YAML configuration file was used to define hyperparameters comprising kernel size, stride size, optimizer, learning rate, batch size, loss function, and activation function. Weights and Biases (wandb.ai, San Francisco, CA, United States) platform was employed to track the training runs and optimize hyperparameters using a Bayesian optimization approach.

The hyperparameter grid and optimal parameters are provided in Supplementary Table 11. The loss curves for the training of the model with the optimal hyperparameters are provided in Supplementary Figure 16. The optimal model was trained using four A6000 GPUs (NVIDIA, Santa Clara, CA, USA) during a total runtime of 4h 57min.

Supplementary Table 11: Hyperparameter Tuning Grid

Hyperparameter	Grid	Optimum
Kernel Size	[5, 7, 11]	11
First Kernel Size	[11, 15, 17, 21, 25]	21
Stride Size	[2, 4]	2
Learning Rate Exponent	[-5, -4, -3, -2]	-4.53
Batch Size	[32, 64, 128]	64
Pooling	[max, avg]	avg
First Kernel Size	[15, 21, 33, 45]	21
Activation	ReLU	ReLU
Optimizer	Adam	Adam
Dropout	0.2	0.2
Loss function	Binary Cross-Entropy	Binary Cross-Entropy



Supplementary Figure 16: ECG-AI training and validation loss with adaptive learning rate and early stopping.