Scalable Risk Stratification for Heart Failure Using Artificial Intelligence applied to 12-lead Electrocardiographic Images: A Multinational Study

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY METHODS

Data Sources

The YNHHS is the largest referral center in southern New England, comprising a network of 5 hospitals and a large outpatient medical group, serving areas of Connecticut, New York, and Rhode Island. The electronic health record (EHR) data represent an extract from the Clarity database, comprising the corpus of healthcare information acquired during patient care at YNHHS using Epic.[1,2] Data between 2014 through 2023 was included in the study.

We used data from the UKB under research application #71033. The UKB is a prospective observational study of 502,468 people aged 40-69 years recruited in 2006-2010.[3] These participants underwent ECGs starting 2014 and were followed through till 2021. The UKB represents the largest population-based cohort in the United Kingdom with protocolized imaging as a part of the study and EHR linkage from the National Health Service.[4,5]

The ELSA-Brasil study is a large multicenter prospective cohort from Brazil, enrolling 15,105 volunteering participants aged 35-74 years at recruitment between the years 2008-2010.[6,7] These participants represent active and retired civil servants from six higher education and research institutions in Brazilian state capitals in three geographical regions of the country: Southeast (Belo Horizonte, Rio de Janeiro, São Paulo and Vitória), South (Porto Alegre) and Northeast (Salvador).[8] ELSA's main objectives are to investigate the development and progression of chronic diseases and their determinants in the Brazilian adult population. Baseline data were collected at the time of recruitment using interviews with previously validated questionnaires and clinical, laboratory and imaging exams including protocolized ECGs and

echocardiograms.[6,7] In-person follow-up visits were conducted every three to four years to ascertain exposure status and to identify changes in baseline subclinical and clinical parameters. In addition, all participants were interviewed yearly via telephone to obtain information on new diagnoses, hospitalization, and death with adjudicated clinical events based on expert medical record review.[6]

Study Population – Cohort Identification at YNHHS

To identify patients with prevalent HF at the time of ECG, we identified the first recorded encounter for all patients within the EHR and followed for 1 year. Patients with prevalent HF based on either a diagnosis code for HF or an echocardiogram with reduced LVEF (defined as LVEF<50%) or left ventricular diastolic dysfunction (defined as “moderate” or “severe” left ventricular diastolic dysfunction) we excluded from the study. The baseline ECG for patients was defined after this 1-year blanking period to exclude prevalent HF (**Figure S1**). The YNHHS cohort also excluded patients previously included in the development of the AI-ECG algorithm and the small proportion of individuals who opted out of research participation (<0.01% of all YNHHS patients).

We further constructed subpopulations in the YNHHS cohort with more consistent follow-up to ensure the completeness of outcome assessments. This included patients with (1) ≥3 years of follow-up in YNHHS and (2) ≥1 encounter within the healthcare system every 2 years. As a sensitivity analysis, we also identified a single random ECG (instead of the first ECG) for each patient after the 1-year blanking period following cohort entry.

Study Exposure: AI-ECG-based HF Risk

The image-based AI-ECG model had 89% sensitivity and 77% specificity for identifying patients with concomitant left ventricular systolic dysfunction on cardiac imaging.[9]

In this study, the model was deployed on ECG images, plotted in standard clinical layout from signal waveform data**,** with a voltage calibration of 10 mm/mV, with the limbs and precordial leads arranged in four columns of 2.5-second each, representing leads I, II, and III; aVR, aVL, and aVF; V1, V2, and V3; and V4, V5, and V6 (**Figure S3**). A 10-second recording of the lead I signal was included as a rhythm strip. These images were converted to greyscale and down-sampled to 300x300 pixels using Python Image Library.[10] As a sensitivity analysis, we also evaluated the model on ECG images plotted in 3 novel formats that were not encountered by the model during training, including (a) Three-rhythm: 3 rhythm strips of 10 seconds each (leads I, II, and V1) below the limb and precordial lead columns, (b) No-rhythm: the 12 limb and precordial ECG leads without any rhythm lead, and (c) Rhythm-on-top: a single 10-second rhythm strip (lead I) above the 12 leads (**Figure Sx3.5**)."

While the study exposure was a positive-screen ECG (model output probability > 0.1), we further defined graded thresholds based on AI-ECG probabilities of 0-0.1, 0.1-0.3, 0.3-0.5, 0.5-0.7, and 0.7-1 to evaluate the association of a higher risk score with future HF.

Study Outcomes and Covariates

In ELSA-Brasil, HF was identified either by in-person interview or the annual telephonic surveillance and investigated by a designated committee that contacted health providers and requested copies of medical records for all hospitalizations. After investigation, the cardiovascular events were adjudicated by an independent

review of two cardiologists. A third senior cardiologist defined the event in case of disagreement.[11] New-onset HF was identified from hospitalization records, based on the presence of a clinical diagnosis of HF, with the individual receiving pharmacological therapy for HF, in addition to any of the following: (1) pulmonary congestion on chest X-ray, (2) reduced ejection fraction or systolic dysfunction observed on cardiac imaging, or (3) preserved ejection fraction with evidence of moderate to severe diastolic dysfunction.

Information about all-cause death was available in the YNHHS EHR, with in-hospital mortality data supplemented from the Connecticut death index to improve capture of out-of-hospital patient mortality. Similarly, information about mortality was available in the UKB via linkage to the EHR and the UK national death registries. Information about death in the ELSA-Brasil study was recorded via telephonic surveillance and confirmed using the national mortality database and death certificates.

Among patients in the YNHHS, we also identified echocardiograms with an LVEF less than 50% performed after the ECG. Further, to evaluate the AI-ECG model for a different definition for new-onset HF, we also identified hospitalizations with any HF diagnosis in YNHHS and UKB.

Study Comparator

The PCP-HF represent sex- and race-specific equations for estimating 10-year risk of HF and include a variety of demographic and clinical features such as age, body mass index, systolic blood pressure, total cholesterol, high-density cholesterol, fasting blood glucose, current smoking status, antihypertensive medication use, antihyperglycemic medication use, and electrocardiogram measurement of QRS

duration. To align with the score development, across cohorts, the PCP-HF score was calculated for White and Black individuals between 30 and 80 years of age with complete documentation of the score covariates. The calculated 10-year risk score was adjusted based on the length of follow-up for each individual to estimate the risk of heart failure over the study period.

In YNHHS, PCP-HF features were extracted from the EHR. Body mass index, systolic blood pressure, and laboratory measurements closest to and within two years of the ECG acquisition date were used for calculation. In ELSA-Brasil, all PCP-HF features, including the ECG, were captured at the baseline visit using established study protocols.[12,13]

SUPPLEMENTARY REFERENCES

1. Chishtie J, Sapiro N, Wiebe N, Rabatach L, Lorenzetti D, Leung AA, et al. Use of Epic Electronic health record system for health care research: Scoping review. *J Med Internet Res* JMIR Publications Inc.; 2023;**25**:e51003.

2. Dhingra LS, Shen M, Mangla A, Khera R. Cardiovascular Care Innovation through Data-Driven Discoveries in the Electronic Health Record. *Am J Cardiol* 2023;**203**:136–148.

3. Palmer LJ. UK Biobank: bank on it. *Lancet* Elsevier BV; 2007;**369**:1980–1982.

4. Littlejohns TJ, Sudlow C, Allen NE, Collins R. UK Biobank: opportunities for cardiovascular research. *Eur Heart J* 2019;**40**:1158–1166.

5. Raisi-Estabragh Z, Petersen SE. Cardiovascular research highlights from the UK Biobank: opportunities and challenges. *Cardiovasc Res* 2020;**116**:e12–e15.

6. Aquino EML, Barreto SM, Bensenor IM, Carvalho MS, Chor D, Duncan BB, et al. Brazilian Longitudinal Study of Adult Health (ELSA-Brasil): objectives and design. *Am J Epidemiol* 2012;**175**:315–324.

7. Schmidt MI, Duncan BB, Mill JG, Lotufo PA, Chor D, Barreto SM, et al. Cohort profile: Longitudinal study of adult health (ELSA-brasil). *Int J Epidemiol* 2015;**44**:68–75.

8. Aquino EML, Araujo MJ, Almeida M da CC, Conceição P, Andrade CR de, Cade NV, et al. Recrutamento de participantes no Estudo Longitudinal de Saúde do Adulto. *Rev Saúde Pública* Faculdade de Saúde Pública da Universidade de São Paulo; 2013;**47**:10–18.

9. Sangha V, Nargesi AA, Dhingra LS, Khunte A, Mortazavi BJ, Ribeiro AH, et al. Detection of Left Ventricular Systolic Dysfunction From Electrocardiographic Images. *Circulation* 2023;

10. Pillow. PyPI. https://pypi.org/project/pillow/ (February 29, 2024)

11. Barreto SM, Ladeira RM, Bastos M do SCB de O, Diniz M de FHS, Jesus EA de, Kelles SMB, et al. ELSA-Brasil strategies for outcome identification, investigation and ascertainment. *Rev Saude Publica* 2013;**47 Suppl 2**:79–86.

12. Mill JG, Pinto K, Griep RH, Goulart A, Foppa M, Lotufo PA, et al. Medical assessments and measurements in ELSA-Brasil. *Rev Saude Publica* 2013;**47 Suppl 2**:54–62.

13. Bensenor IM, Griep RH, Pinto KA, Faria CP de, Felisbino-Mendes M, Caetano EI, et al. Routines of organization of clinical tests and interviews in the ELSA-Brasil investigation center. *Rev Saude Publica* 2013;**47 Suppl 2**:37–47.

Figure S1. Overview of Cohort Creation at the Yale New Haven Health System.

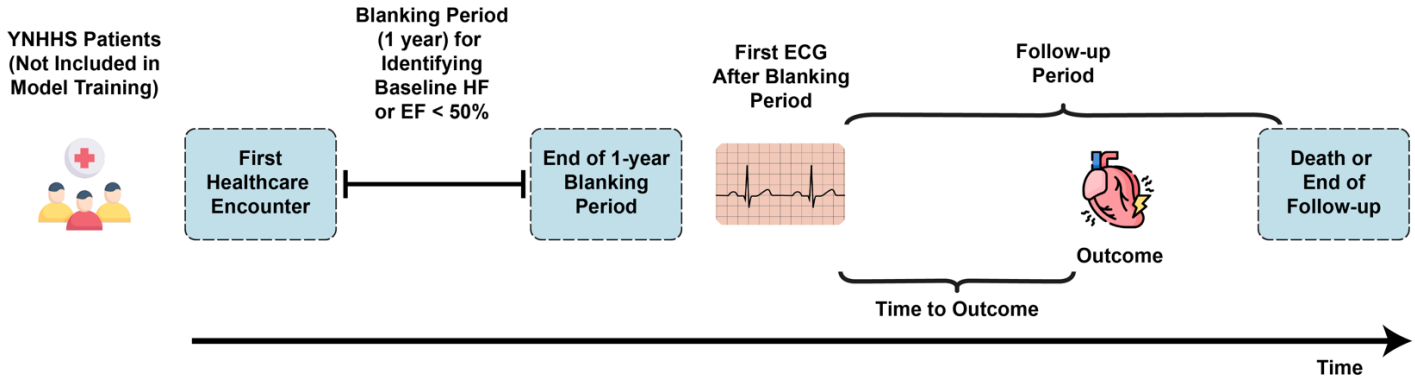

Figure S2. Consort Diagram for (A) Yale New Haven Health System Cohort and (B) UK Biobank Cohort. Abbreviations: AI, Artificial Intelligence; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; ECG, Electrocardiogram; HF, Heart Failure

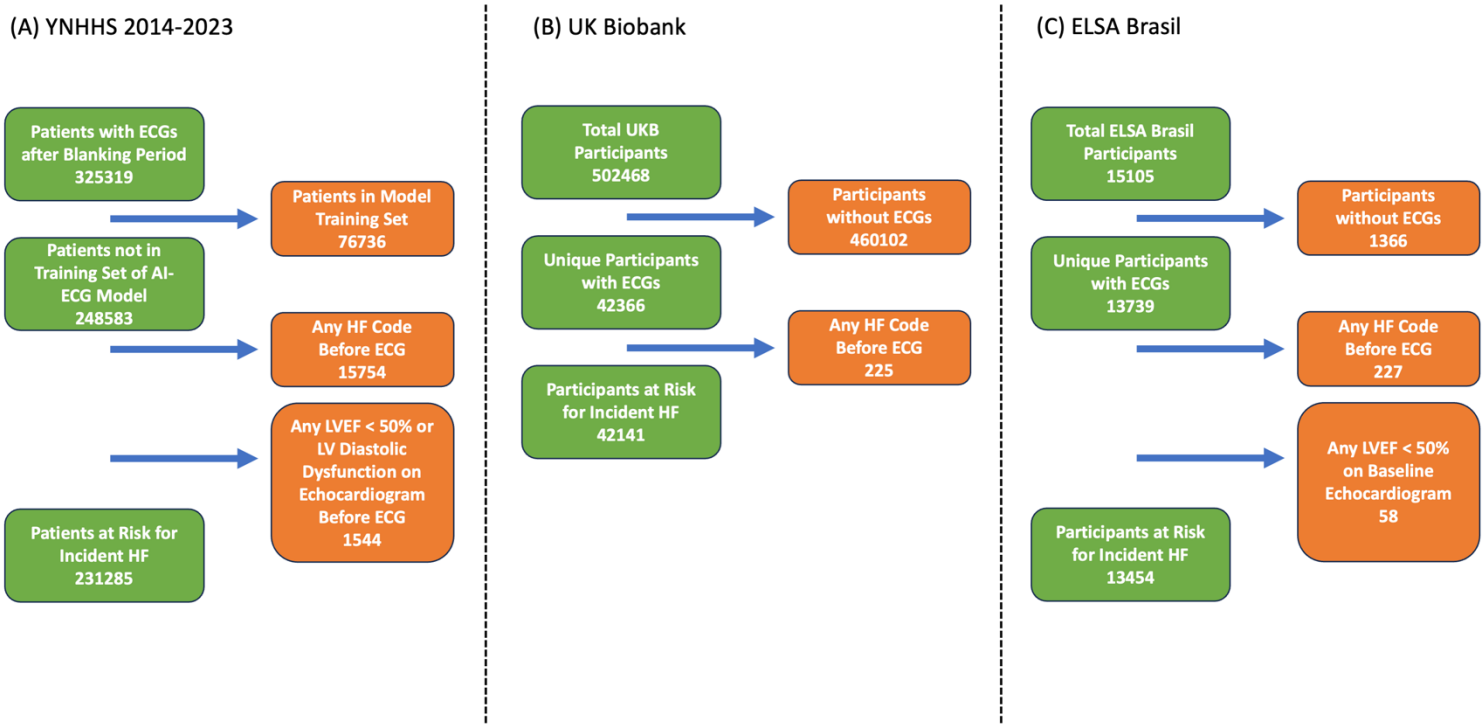

Figure S3. Examples of electrocardiogram images used for model evaluation.

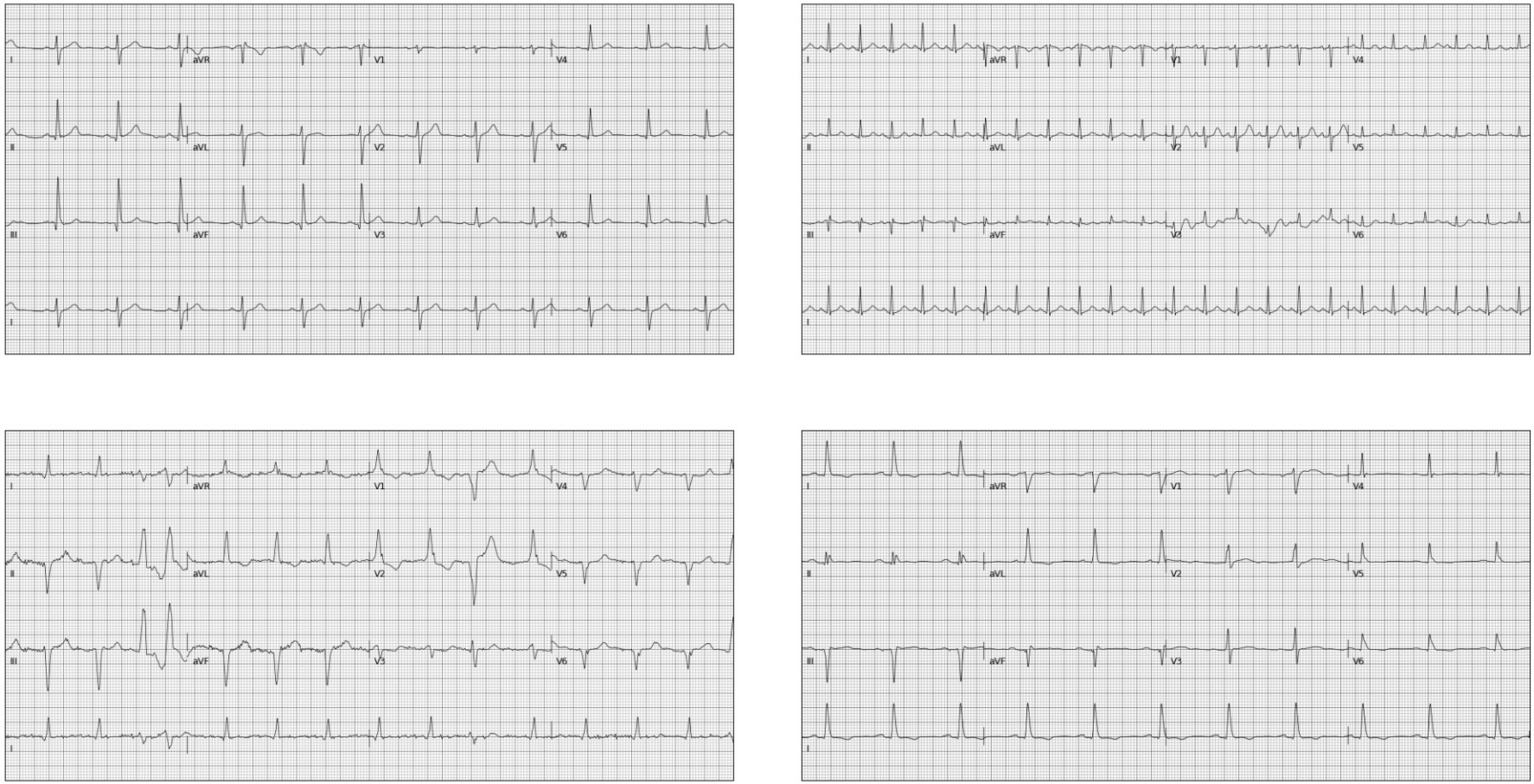

Figure S4. Novel electrocardiogram layouts not encountered during model development: (A) Three-rhythm, (B) No-rhythm, and (C) Rhythm-on-top.

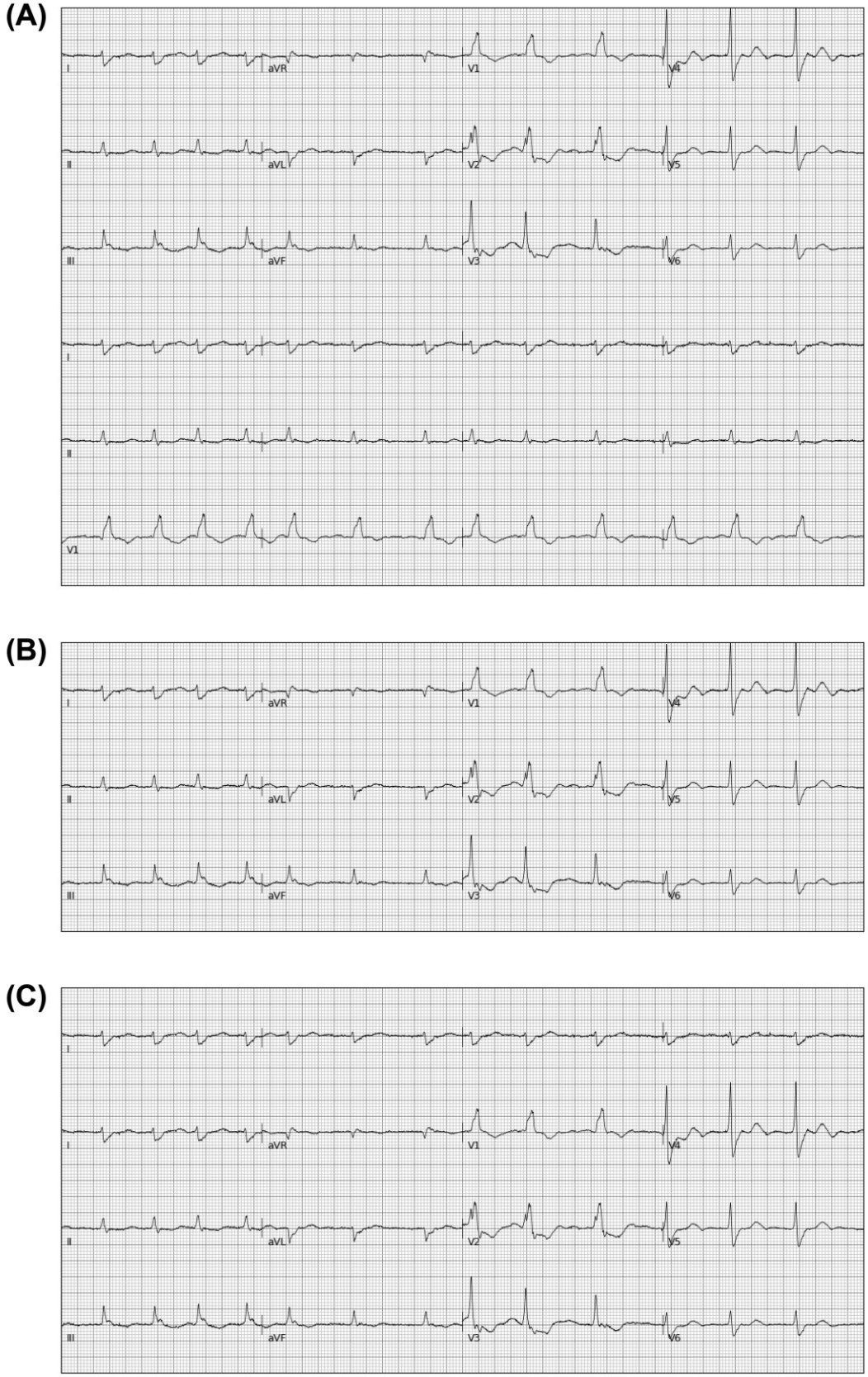

Figure S5. Age- and Sex-adjusted Cumulative Hazard Curve for Primary Heart Failure Hospitalization following a 3-month Blanking Period after the Electrocardiogram. Abbreviations: AI-ECG, Artificial Intelligence-enhanced Electrocardiograms

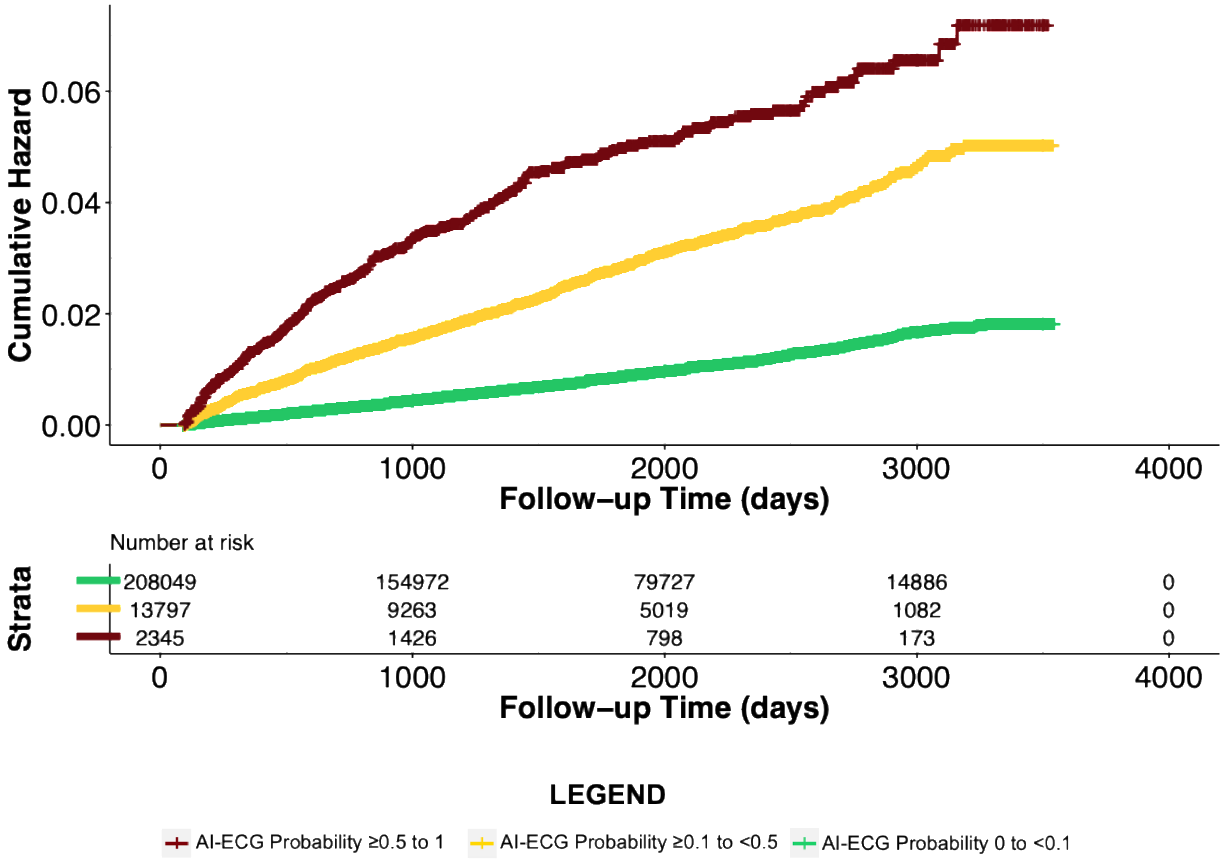

Figure S6. Age- and Sex-adjusted Cumulative Hazard Curve for Primary Heart Failure Hospitalization within the Yale New Haven Health System Cohort, including a Random Electrocardiogram following the One-year Blanking Period. Abbreviations: AI-ECG, Artificial Intelligence-enhanced Electrocardiograms

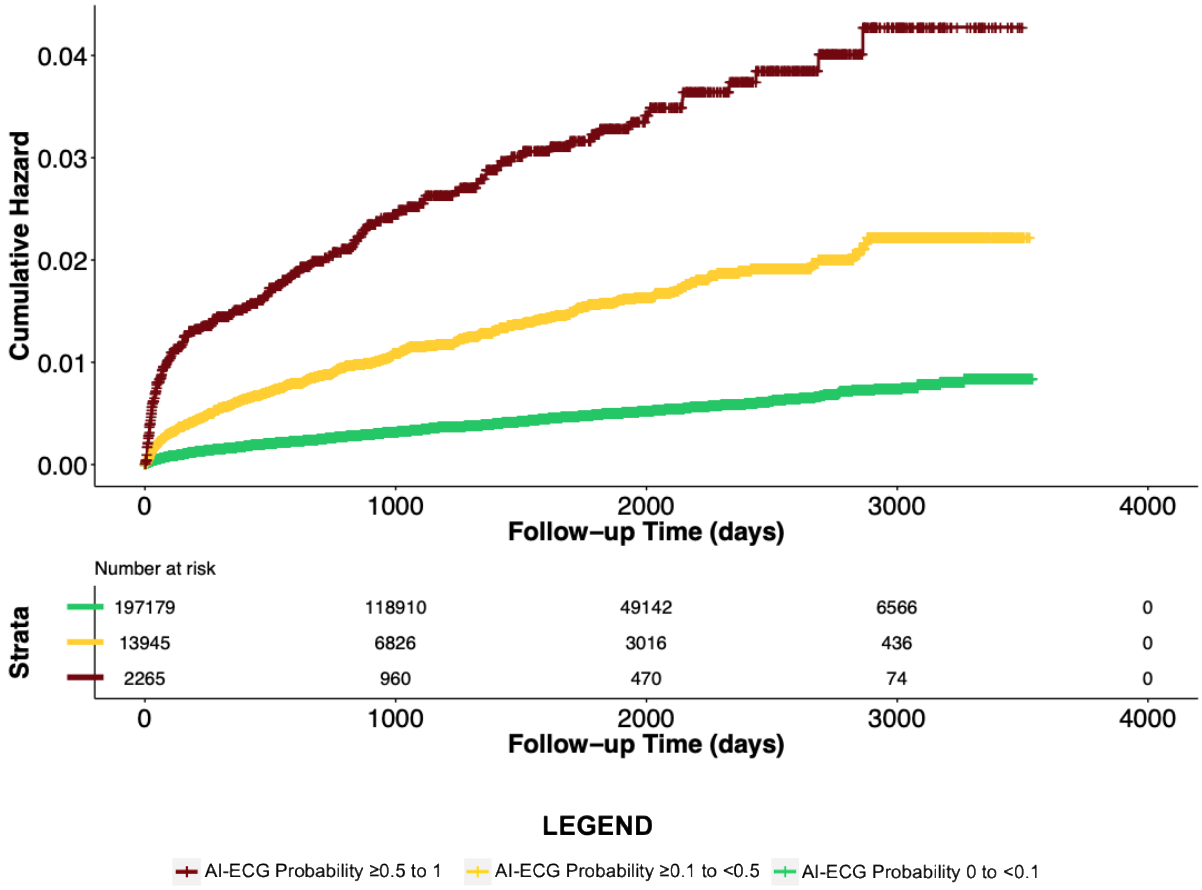

Figure S7. Age- and Sex-adjusted Cumulative Hazard Curve for Primary Heart Failure Hospitalization in the Pooled Population-based Study Cohorts.
Abbreviations: AI-ECG, Artificial Intelligence-enhanced Electrocardiograms

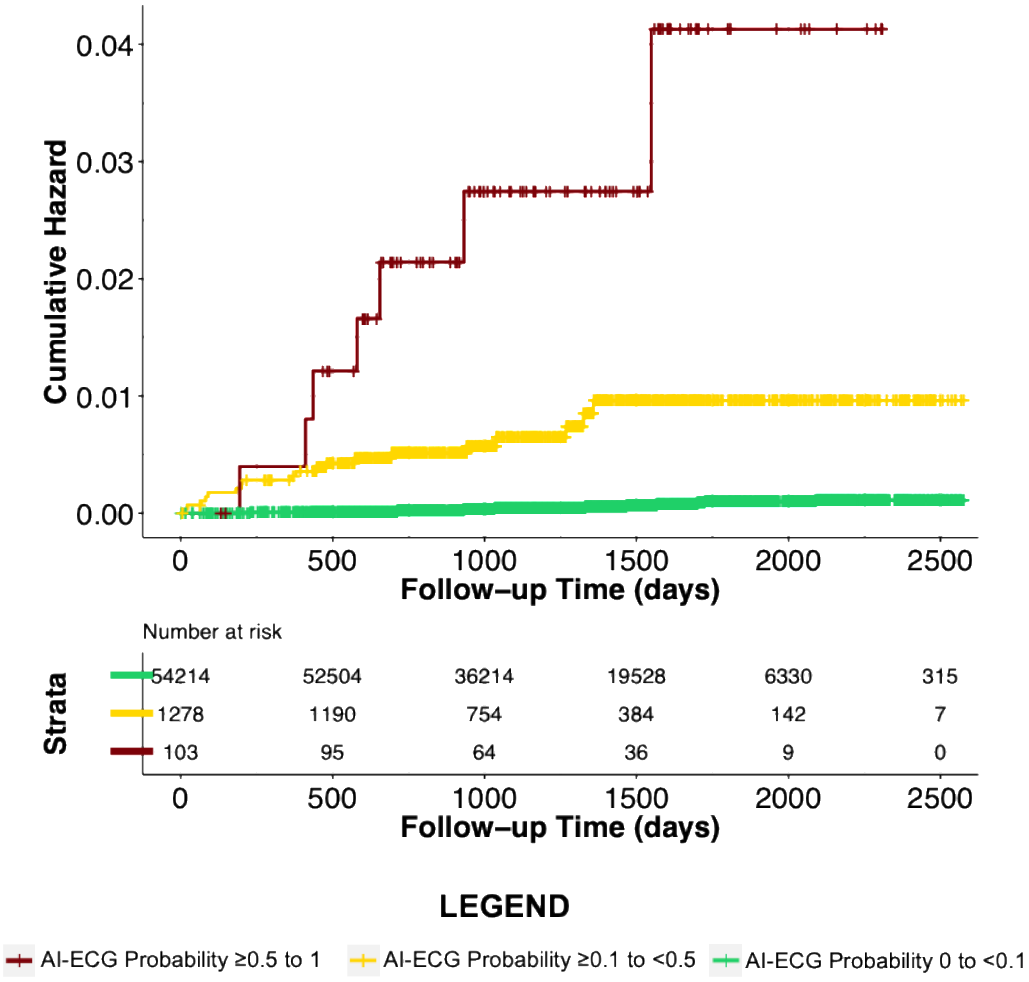

Table S1. International Classification of Disease Tenth Revision Codes for the Identification of Comorbidities and Outcomes. Abbreviations: ICD-10-CM, International Classification of Disease Tenth Revision Clinical Modification Codes.

Condition	ICD-10-CM codes
Heart Failure	'I11.0','I13.0','I13.2','I50','I50.0','I50.1','I50.9','Z95.81','I09.81'
Acute Myocardial Infarction	'I21', 'I22', 'I23', 'I24.0', 'I24.8', 'I24.9'
Stroke	'G45','G45.0','G45.1','G45.2','G45.3','G45.4','G45.8','G45.9', 'I63','I63.0','I63.1','I63.2','I63.3','I63.4','I63.5','I63.8','I63.9','I64', 'I65','I65.0','I65.1','I65.2','I65.3','I65.8','I65.9','I66','I66.0','I66.1', 'I66.2','I66.3','I66.4','I66.8','I66.9','I67.2','I69.3','I69.4'
Type 2 Diabetes Mellitus	'E11','E11.0','E11.1','E11.2','E11.3','E11.4','E11.5','E11.6', 'E11.7','E11.8','E11.9','O24.1'
Hypertension	'I10','I11','I11.0','I11.9','I12','I12.0','I12.9', 'I13','I13.0','I13.1','I13.2','I13.9','I67.4', 'O10','O10.0','O10.1','O10.2','O10.3','O10.9','O11'

Table S2. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis using Artificial Intelligence (TRIPOD + AI) Checklist.

Section/Topic	Item	Development / evaluation[1]	Checklist item	Reported on page
TITLE				
Title	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Pg 1
ABSTRACT				
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	Pg 2
INTRODUCTION				
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	Pg 5
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	Pg 6
	3c	D;E	Describe any known health inequalities between sociodemographic groups	Pg 16
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	Pg 6
METHODS				
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	Pg 6
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	Pg 7 and suppl.
Participants	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres	Pg 7
	6b	D;E	Describe the eligibility criteria for study participants	Pg 7-8
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	N/A
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	Supp.
Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	Pg 8-9
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	N/A
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	N/A
Predictors	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	Pg 8
	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	N/A
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	Pg 8
Sample size	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	Pg 6
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	Pg 6
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	Pg 9-10
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	Pg 9-10
	12c	D	Specify the type of model, rationale[2], all model-building steps, including any hyperparameter tuning, and method for internal validation	Pg 8-10
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations[3]	N/A
	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	Pg 10
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	N/A
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	Pg 9-10
Class imbalance	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	N/A
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	Pg 16
Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	Pg 10

[1] D=items relevant only to the development of a prediction model; E=items relating solely to the evaluation of a prediction model; D;E=items applicable to both the development and evaluation of a prediction model

[2] Separately for all model building approaches.

[3] TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (eg, at different hospitals or centres). Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018]

Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	N/A
Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	Pg 6
OPEN SCIENCE				
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	Pg 20
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	Pg 20-22
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	Not prepared
Registration	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	N/A
Data sharing	18e	D;E	Provide details of the availability of the study data	Pg 22
Code sharing	18f	D;E	Provide details of the availability of the analytical code[4]	Pg 22
PATIENT & PUBLIC INVOLVEMENT				
Patient & Public Involvement	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	N/A
RESULTS				
Participants	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Pg 11
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	Pg 11
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	Pg 14-15
Model development	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	Pg 13-14
Model specification	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)[5]	Pg 22
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	Pg 14-15
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details[3].	N/A
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	N/A
DISCUSSION				
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	Pg 15
Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	Pg 18
Usability of the model in the context of current care	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	Pg 19
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	N/A
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	Pg 20

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378

[4] This relates to the analysis code, for example, any data cleaning, feature engineering, model building, evaluation.
[5] This relates to the code to implement the model to get estimates of risk for a new individual.

Table S3. Model Performance for Prediction of Heart Failure Risk Across Demographic Subgroups. Abbreviations: CI, Confidence Interval; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; HF, Heart Failure; UKB, UK Biobank; YNHHS, Yale New Haven Health System

Subgroup		YNHHS			UKB			ELSA-Brasil		
		Total Number of Individuals at Risk	Number of Incident HF Events	Age- and Sex-Adjusted Cox Proportional Hazard Ratios (95% CI)	Total Number of Individuals at Risk	Number of Incident HF Events	Age- and Sex-Adjusted Cox Proportional Hazard Ratios (95% CI)	Total Number of Individuals at Risk	Number of Incident HF Events	Age- and Sex-Adjusted Cox Proportional Hazard Ratios (95% CI)
Age < 65		151199 (65.4)	1352	8.00 (7.12-8.99)	20802 (49.4)	9	25.63 (6.34-103.61)	12038 (89.4)	21	25.35 (10.01-64.21)
Age ≥ 65		80086 (34.6)	3120	3.00 (2.78-3.24)	21345 (50.6)	37	11.1 (5.54-22.23)	1416 (10.6)	10	19.94 (5.60-70.99)
Female		130941 (56.6)	2225	3.41 (3.09-3.76)	21795 (51.7)	11	13.41 (3.53-50.96)	7348 (54.6)	11	16.30 (6.20-42.89)
Male		100341 (43.4)	2247	4.34 (3.98-4.74)	20346 (48.3)	35	12.87 (6.33-26.15)	6106 (45.4)	20	43.74 (12.98-147.42)
Race/Ethnicity	**White**	145726 (63.0)	3343	3.38 (3.13-3.64)	40691 (96.6)	46	12.85 (6.87-24.02)	6920 (51.4)	15	27.18 (9.34-79.07)
	Black	36605 (15.8)	624	5.43 (4.59-6.42)	304 (0.7)	0	-	2130 (15.8)	9	26.18 (6.65-103.05)
	Hispanic	36298 (15.7)	358	5.73 (4.51-7.29)	0	-	-	-	-	-
	Asian	4221 (1.8)	47	9.63 (5.17-17.96)	600 (1.4)	0	-	332 (2.5)	0	-
	Other	2565 (1.1)	35	4.41 (2.03-9.57)	546 (1.3)	0	-	305 (2.3)	0	-
	Brazilian "Pardo"	-	-	-	-	-	-	3767 (28.0)	7	7.89 (0.94-66.19)
	Missing	5870 (2.5)	65	5.44 (3.16-9.39)	0	-	-	-	-	-

Table S4. Performance Characteristics across Cohorts following a 3-month Blanking Period after the Electrocardiogram. Abbreviations: AI-ECG, Artificial Intelligence-enabled Electrocardiography; YNHHS, Yale New Haven Health System

Metric	YNHHS
Number of Individuals at Risk	224191
Number of Heart Failure Hospitalizations	3768
Age- and Sex-adjusted Hazard Ratio for Positive AI-ECG Screen	3.45 (3.21-3.71)
Harrell's C-statistic for Model Probability	0.704 (0.694-0.714)

Table S5. Model Performance (Hazard Ratios [95% Confidence Intervals]) for Prediction of Future Heart Failure Across Key Subsets of Interest.
Abbreviations: YNHHS, Yale New Haven Health System

Characteristic		YNHHS with minimum 3-year follow-up	YNHHS with at least one encounter every two years	YNHHS Random ECG after Blanking Period
Total Number of Individuals at Risk		94848	128466	213389
Number of Heart Failure Hospitalizations		2898	3749	1685
Cox Proportional Hazard Model	Positive Screen + Age + Sex	3.75 (3.46-4.08)	3.49 (3.25-3.75)	3.76 (3.38-4.18)

Table S6. Age- and Sex- Adjusted Cox Proportional Hazard Models for the Prediction of Clinical Outcomes. Abbreviations: AMI, acute myocardial infarction; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; HF, heart failure; UKB, UK Biobank; YNHHS, Yale New Haven Health System

Outcome	Age- and Sex- Adjusted Cox Proportional Hazard Models		
	YNHHS Hazard Ratio (95% CI)	UKB Hazard Ratio (95% CI)	ELSA-Brasil Hazard Ratio (95% CI)
Primary HF Hospitalization	3.88 (3.63-4.14)	12.85 (6.87-24.02)	23.50 (11.09-49.81)
Primary HF Hospitalization or an Echocardiogram with LVEF < 50%	5.05 (4.83-5.27)	-	-
Primary AMI Hospitalization	1.44 (1.04-2.00)	3.16 (1.98-5.02)	3.53 (1.4-8.85)
Primary Stroke Hospitalization	1.05 (0.95-1.17)	2.30 (1.36-3.9)	5.74 (2.59-12.72)
All-cause Death	1.19 (1.15-1.24)	2.13 (1.41-3.24)	3.64 (2.27-5.83)
Major Adverse Cardiovascular Events	2.10 (2.04-2.17)	2.79 (2.17-3.6)	4.04 (2.77-5.89)
Any Hospitalization with HF	3.11 (3.00-3.22)	7.32 (5.3-10.09)	-
Any Hospitalization with HF or an Echocardiogram with LVEF < 50%	3.48 (3.37-3.59)	-	-

Table S7. Model Performance (Hazard Ratios [95% Confidence Intervals]) for Prediction of Future Heart Failure Across Key Subsets of Interest.
Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Novel image format	Age- and Sex- Adjusted Cox Proportional Hazard Models		
	YNHHS Hazard Ratio (95% CI)	UKB Hazard Ratio (95% CI)	ELSA-Brasil Hazard Ratio (95% CI)
Three-rhythm	3.20 (3.03 - 3.41)	7.92 (4.25 - 14.78)	11.10 (5.32 - 23.10)
No-rhythm	2.84 (2.67 - 3.00)	7.00 (3.64 - 13.40)	17.93 (8.63 - 37.29)
Rhythm-on-top	3.17 (2.98 - 3.37)	7.15 (3.74 - 13.69)	11.16 (5.42 - 23.00)

Table S8. Performance Characteristics for Predicting New-onset Heart Failure in the Pooled Population-based Study Cohorts. Abbreviations: ECG, Electrocardiogram; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; HF, heart failure; UKB, UK Biobank.

Metric	Pooled Cohort (UKB and ELSA-Brasil)
Number of Individuals at Risk	55595
Number of HF Hospitalizations	77
Age- and Sex-adjusted Hazard Ratio for Positive AI-ECG Screen	17.07 (10.54-27.65)
Harrell's C-statistic for Model Probability	0.782 (0.723-0.842)

Table S9. Age- and Sex- Adjusted Hazard Ratios for New-onset Heart Failure across Model Output Probabilities. Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; UKB, UK Biobank; YNHHS, Yale New Haven Health System

Age- and Sex- Adjusted Cox Proportional Hazard Models		YNHHS Hazard Ratio (95% CI)	UKB Hazard Ratio (95% CI)	ELSA-Brasil Hazard Ratio (95% CI)
Per 0.1 increase in model output probability		1.36 (1.35-1.38)	1.81 (1.58-2.07)	1.93 (1.68-2.21)
Grouped by ranges of model output probabilities	**0-0.1**	Reference	Reference	Reference
	0.1-0.5	3.29 (3.06-3.54)	11.30 (5.75-22.18)	16.80 (7.02-40.25)
	0.5-1	7.16 (6.44-7.95)	29.17 (8.82-96.48)	80.11 (26.96-237.97)
Grouped by ranges of model output probabilities	**0-0.1**	Reference	Reference	Reference
	0.1-0.3	3.00 (2.77-3.26)	10.27 (4.86-21.71)	18.66 (7.44-46.85)
	0.3-0.5	4.38 (3.87-4.95)	16.29 (4.93-53.83)	10.68 (1.41-80.79)
	0.5-0.7	6.09 (5.30-7.00)	22.44 (5.32-94.62)	31.24 (4.16-234.78)
	0.7-0.9	7.63 (6.46-9.00)	74.59 (10.05-553.42)	170.11 (48.09-601.77)
	0.9-1	17.58 (13.28-23.27)	-	-

Table S10. Age- and Sex-adjusted Cox Proportional Hazard Models Across Different Model Output Probability Threshold for Defining a Positive Screen.
Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Model Output Probability Threshold for Positive Screen	Age- and Sex- Adjusted Cox Proportional Hazard Models		
	YNHHS Hazard Ratio (95% CI)	UKB Hazard Ratio (95% CI)	ELSA-Brasil Hazard Ratio (95% CI)
>0.05	3.25 (3.06-3.46)	6.93 (3.83-12.53)	11.71 (5.64-24.36)
>0.1	3.88 (3.63-4.14)	12.85 (6.87-24.02)	23.50 (11.09-49.81)
>0.2	4.32 (4.01-4.66)	14.46 (6.67-31.34)	22.68 (9.12-56.40)
>0.5	5.57 (5.03-6.17)	20.88 (6.42-67.95)	58.16 (20.07-168.48)

Table S11. Comparison of Discrimination for AI-ECG Model Output Probability and Components of the Pooled Cohort equations to Prevent Heart Failure for Incident Heart Failure in the Yale New Haven Health System. Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; PCP-HF, Pooled Cohort equations to Prevent Heart Failure; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Covariates	Harrell's C-statistic	Marginal difference over Harrell's C-statistic for PCP-HF Components	P-value
PCP-HF Components	0.688 (0.670-0.707)	-	-
AI-ECG Model Output Probability	0.718 (0.697-0.738)	0.028 (0.002-0.055)	0.03
AI-ECG Model Output Probability + Age + Sex	0.724 (0.705-0.743)	0.035 (0.016-0.053)	< 0.001
AI-ECG Model Output Probability + PCP-HF Components	0.763 (0.746-0.780)	0.075 (0.061-0.089)	< 0.001

Table S12. Discrimination (Harrell's C-statistic [95% Confidence Intervals]) for New-onset Heart Failure Across Novel Electrocardiogram Image Formats.
Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; CI, confidence intervals; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Novel image format	AI-ECG Model Discrimination for New-onset Heart Failure		
	YNHHS Harrell's C-statistic (95% CI)	UKB Harrell's C-statistic (95% CI)	ELSA-Brasil Harrell's C-statistic (95% CI)
Three-rhythm	0.720 (0.701-0.740)	0.752 (0.647-0.858)	0.879 (0.810-0.947)
No-rhythm	0.713 (0.695-0.733)	0.834 (0.767-0.902)	0.886 (0.808-0.964)
Rhythm-on-top	0.717 (0.697-0.738)	0.764 (0.662-0.867)	0.870 (0.795-0.945)

Table S13. Net Reclassification Index of AI-ECG Model Output Probability over Pooled Cohort Equations to Prevent Heart Failure. Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; NRI, Net Reclassification Index; PCP-HF, Pooled Cohort equations to Prevent Heart Failure; YNHHS, Yale New Haven Health System.

Metric	YNHHS	ELSA-Brasil
Net Reclassification Index (Overall)	0.219 (0.180 to 0.270)	0.356 (0.079 to 0.621)
Net Reclassification Index (Events)	0.145 (0.106 to 0.195)	0.358 (0.081 to 0.622)
Net Reclassification Index (Non-events)	0.073 (0.068 to 0.083)	-0.002 (-0.005 to 0.002)