Artificial Intelligence Enabled Prediction of Heart Failure Risk from Single-lead Electrocardiograms

Short Title: AI-enabled HF Risk Prediction Using Single-lead ECG

Lovedeep S Dhingra MBBS[1]*, Arya Aminorroaya MD, MPH[1]*, Aline F Pedroso PhD[1], Akshay Khunte BS[2], Veer Sangha BS[1,3], Daniel McIntyre MD, PhD[4], Clara K Chow MBBS, PhD[4,5], Folkert W Asselbergs MD, PhD[6,7,8], Luisa CC Brant MD, PhD[9,10], Sandhi M Barreto MD, PhD[11], Antonio Luiz P Ribeiro MD, PhD[9,10], Harlan M Krumholz MD, SM[1,12,13], Evangelos K Oikonomou MD, DPhil[1], Rohan Khera MD, MS[1,12,14,15]

[1]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
[2]Department of Computer Science, Yale University, New Haven, CT, USA
[3]Department of Engineering Science, University of Oxford, Oxford, UK
[4]Westmead Applied Research Centre, Faculty of Medicine and Health, The University of Sydney, Westmead, Australia
[5]Department of Cardiology, Westmead Hospital, Sydney, Australia
[6]Department of Cardiology, Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, Netherlands
[7]Institute of Health Informatics, University College London, London, UK
[8]The National Institute for Health Research University College London Hospitals Biomedical Research Centre, University College London, London, UK
[9]Department of Internal Medicine, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[10]Telehealth Center and Cardiology Service, Hospital das Clínicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[11]Department of Preventive Medicine, School of Medicine, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[12]Center for Outcomes Research and Evaluation (CORE), Yale New Haven Hospital, New Haven, CT, USA
[13]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA
[14]Section of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA
[15]Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

* Contributed equally as co-first authors

Correspondence to:
Rohan Khera, MD, MS
195 Church Street, 6th Floor, New Haven, CT 06510
(203) 764-5885; rohan.khera@yale.edu; @rohan_khera

Word Count: 3399

Keywords: Heart failure, Deep learning, Wearable Devices, Electrocardiograms, Risk Stratification, Predictive Modelling

KEY POINTS

Question: Can single-lead electrocardiograms (ECG) predict heart failure (HF) risk?

Findings: We evaluated a noise-adapted artificial intelligence (AI) algorithm for single-lead ECGs across multinational cohorts, spanning a diverse US health system and community-based cohorts in the UK and Brazil. A positive AI-ECG screen was associated with 3- to 7-fold higher HF risk, independent of age, sex, and comorbidities. The AI model demonstrated similar or improved performance, compared with two established clinical risk scores for HF prediction.

Meaning: A noise-adapted AI model for single-lead ECG predicted new-onset HF risk, representing a scalable HF risk-stratification strategy for portable and wearable devices.

ABSTRACT

Importance: Despite the availability of disease-modifying therapies, scalable strategies for heart failure (HF) risk stratification remain elusive. Portable devices capable of recording single-lead electrocardiograms (ECGs) can enable large-scale community-based risk assessment.

Objective: To evaluate an artificial intelligence (AI) algorithm to predict HF risk from noisy single-lead ECGs.

Design: Multicohort study.

Setting: Retrospective cohort of individuals with outpatient ECGs in the integrated Yale New Haven Health System (YNHHS) and prospective population-based cohorts of UK Biobank (UKB) and Brazilian Longitudinal Study of Adult Health (ELSA-Brasil).

Participants: Individuals without HF at baseline.

Exposures: AI-ECG-defined risk of left ventricular systolic dysfunction (LVSD).

Main Outcomes and Measures: Among individuals with ECGs, we isolated lead I ECGs and deployed a noise-adapted AI-ECG model trained to identify LVSD. We evaluated the association of the model probability with new-onset HF, defined as the first HF hospitalization. We compared the discrimination of AI-ECG against two risk scores for new-onset HF (PCP-HF and PREVENT equations) using Harrel's C-statistic, integrated discrimination improvement (IDI), and net reclassification improvement (NRI).

Results: There were 192,667 YNHHS patients (age 56 years [IQR, 41-69], 112,082 women [58%]), 42,141 UKB participants (65 years [59-71], 21,795 women [52%]), and 13,454 ELSA-Brasil participants (56 years [41-69], 7,348 women [55%]) with baseline ECGs. A total of 3,697 developed HF in YNHHS over 4.6 years (2.8-6.6), 46 in UKB over 3.1 years (2.1-4.5), and 31 in ELSA-Brasil over 4.2 years (3.7-4.5). A

positive AI-ECG screen was associated with a 3- to 7-fold higher risk for HF, and each 0.1 increment in the model probability portended a 27-65% higher hazard across cohorts, independent of age, sex, comorbidities, and competing risk of death. AI-ECG's discrimination for new-onset HF was 0.723 in YNHHS, 0.736 in UKB, and 0.828 in ELSA-Brasil. Across cohorts, incorporating AI-ECG predictions alongside PCP-HF and PREVENT equations was associated with a higher Harrel's C-statistic ($\Delta_{PCP\text{-}HF}$=0.080-0.107; $\Delta_{PREVENT}$=0.069-0.094). AI-ECG had IDI of 0.091-0.205 and 0.068-0.192, and NRI of 18.2%-47.2% and 11.8%-47.5%, vs. PCP-HF and PREVENT, respectively.

Conclusions and Relevance: Across multinational cohorts, a noise-adapted AI model defined HF risk using lead I ECGs, suggesting a potential portable and wearable device-based HF risk-stratification strategy.

BACKGROUND

Accessible strategies for heart failure (HF) risk stratification remain elusive despite the availability of evidence-based therapies that can effectively modify the disease trajectory.[1,2] Clinical scores to predict HF risk, such as the pooled cohort equations to prevent HF (PCP-HF), the predicting risk of cardiovascular disease events (PREVENT) equations, and the Health ABC score,[3–5] require clinical evaluation, including a detailed history, physical exam, electrocardiogram (ECG), and laboratory testing.[3–9] These complex inputs limit their use, systematically excluding those without healthcare access.[8–10] Similarly, serum-based biomarkers such as N-terminal pro–B-type natriuretic peptide (NT-proBNP) and high-sensitivity cardiac troponin, which portend a higher HF risk when elevated, are limited by the need for blood draws and sample storage, and frequent inaccessibility at the point-of-contact.[11–16] Thus, there is an unmet need for a simple and efficient strategy for HF risk stratification in the community.

Given their increasing utility and ubiquity, portable devices capable of recording single-lead ECG have been proposed as a platform for cardiovascular monitoring and screening.[17–20] Further, artificial intelligence (AI)-enhanced interpretation of ECGs (AI-ECG) has been shown to detect hidden cardiovascular disease signatures from single-lead ECGs.[21–26] However, these portable ECGs are prone to noise introduction during acquisition, which can limit the AI model performance unless specialized measures are taken to ensure they are resilient to such noise.[21,27] Recently, we reported a novel approach for single-lead ECGs that incorporates random noising during model development, enabling consistent diagnostic performance across varying levels of real-world noises.[21] Our initial model development focused on detecting reduced left ventricular ejection fraction (LVEF)

on single-lead ECG based on information from a concurrent echocardiogram, with the potential application of identifying subclinical left ventricular systolic dysfunction (LVSD). Recent studies also suggest that the AI-ECG signature for LVSD identifies other subtle markers of LV dysfunction, including abnormal LV strain and diastolic function, especially among those with a positive screen but preserved LVEF.[28–30]

Given the increasing accessibility of single-lead ECGs, we hypothesized that an AI model developed to detect the cross-sectional signature of LVSD from single-lead ECGs can predict future HF risk. We evaluated our approach in individuals undergoing outpatient ECGs within a diverse US health system and two large population-based cohorts in the UK and Brazil.

METHODS

Data Sources

We included three large cohorts spanning different countries and settings who had undergone an ECG: (i) individuals seeking outpatient care in the Yale New Haven Health System (YNHHS), a large healthcare system in the Northeastern US, including 5 independent hospitals and an outpatient network, (ii) participants in the UK Biobank, a nationwide UK-based cohort study, and (iii) participants in the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil), the largest community-based cohort study from Brazil. While YNHHS included testing and follow-up as a part of routine clinical care in an integrated health system, participants in UKB and ELSA-Brasil had protocolized evaluation at baseline and comprehensive longitudinal follow-up (**eMethods**).

Study Population

In YNHHS, to approximate a screening setting, we identified patients undergoing an outpatient 12-lead ECG during 2014-2023 without HF before the ECG. To account for the ECGs potentially being obtained as a part of HF workup, we included a 1-year blanking period from the first recorded encounter in the electronic health records (EHR) to identify those with prevalent HF (**eMethods; eFigure 1**). A total 255,604 individuals had at least one outpatient ECG after the blanking period. We excluded 47,720 patients in the model development population and 11,954 with prevalent HF. Additionally, we excluded 1,590 patients with LV dysfunction (LVEF under 50% or moderate/severe LV diastolic dysfunction) and 1,673 patients with an NT-proBNP of >300 pg/mL before the index ECG (**eFigure 2**).

To avoid selection bias in UKB and ELSA-Brasil, we identified all participants who received an ECG. In UKB, 42,366 participants underwent a 12-lead ECG during imaging visits (2014-2020). We used the linkage with the UK National Health Service EHR to exclude 225 participants who had been hospitalized with a principal or secondary discharge diagnosis of HF before the ECG. In ELSA-Brasil, we included 13,739 participants who had undergone a 12-lead ECG during 2008-2010, excluding those with HF (N=227) or an LVEF under 50% (N=58) on their baseline echocardiogram (**eFigure 2**).

Study Outcomes

We defined the outcome as new-onset HF, characterized by HF hospitalizations. In YNHHS, this was defined as a hospitalization with an International Classification of Disease Tenth Revision – Clinical Modification (ICD-10-CM) code for HF as the principal discharge diagnosis (**eTable 1**). This approach was guided by the over 95% specificity of diagnosis codes, especially as the principal discharge diagnosis, for HF

diagnosis.[31] Similarly, in UKB, we used the linked EHR to identify hospitalizations with HF as the principal diagnosis code. In ELSA-Brasil, HF was identified by in-person interview or telephonic surveillance for all hospitalizations, followed by independent medical record review and adjudication of HF hospitalizations by two cardiologists (**eMethods**).[32]

We further evaluated the association of AI-ECG probabilities with alternate definitions of HF, and composite outcomes, including (i) any hospitalization with a principal or secondary HF diagnosis code, (ii) a subsequent echocardiogram with LVEF under 50%, and (iii) a composite outcome of HF or all-cause death (**eMethods**). To evaluate the specificity of AI-ECG-defined HF risk, we examined the risk of other cardiovascular conditions, including acute myocardial infarction (AMI), stroke hospitalizations, and all-cause mortality (**eTable 1**). A composite outcome of major adverse cardiovascular events (MACE) was defined as HF, AMI, stroke, or death.

Study Exposure

We defined the exposure as the output of an AI-ECG model trained to detect concurrent LVSD on lead I of a 12-lead ECG, representing the lead commonly captured by portable ECG devices.[21] This was developed at the Yale New Haven Hospital (YNHH) using a novel approach of augmenting training data with random Gaussian noise (**eMethods**). The model achieved excellent discrimination (area under the receiver operating characteristic curve of 0.899 [95% CI, 0.889-0.909]) for detecting concurrent LVSD in the YNHH held-out test set and performed consistently across clinical and population-based external validation cohorts (**eTable 2; eFigure 3**).

We deployed this established model without further development to lead I ECG signals to obtain the LVSD probability, representing a continuous HF risk score. We defined a positive AI-ECG screen as an output probability greater than 0.08, representing the model threshold for 90% sensitivity for detecting LVSD during internal validation.[21]

Study Comparator

We compared the performance of the AI-ECG algorithm with two established risk scores for predicting HF risk, the PCP-HF and PREVENT equations. The PCP-HF score was developed and validated in 7 community-based cohorts.[4] It uses 12 input features, including demographics (age, sex, race), physical exam-based features (smoking status, BMI, systolic blood pressure), laboratory measurements (total cholesterol, high-density lipoprotein cholesterol, fasting blood glucose), medication history (use of antihypertensive and antihyperglycemic medications), and electrocardiographically-defined QRS duration. The PREVENT equations were recently developed using data from over 3.2 million individuals and were validated in 21 datasets.[5,33] The PREVENT equations for HF risk prediction employ 8 inputs, entailing demographics (age, sex), medical history (type 2 diabetes mellitus), physical exam-based features (smoking status, BMI, systolic blood pressure), laboratory measurements (estimated glomerular filtration rate), and medication history (antihypertensive medication use). Across cohorts, these features were determined using the EHR and/or study visits (**eMethods**).[34–37]

Statistical Analysis

We used age-, sex-, and comorbidity-adjusted Cox proportional hazard models with time-to-first HF event as the dependent variable and the AI-ECG-based screen results (positive/negative) or continuous model probability as the independent variable to evaluate the association of the model output with HF risk. Multi-outcome Fine-Gray subdistribution hazard models were used to account for the competing risk of death.[38]

The incremental discrimination of AI-ECG over PCP-HF and PREVENT for predicting time-to-HF hospitalization was reported as the difference in Harrel's C-statistics using a one-shot nonparametric approach.[39] We calculated integrated discrimination improvement (IDI), and categorical and continuous time-to-event net reclassification improvement (NRI).[40] We further compared the net benefit of the AI-ECG model with PCP-HF and PREVENT across probability thresholds (**eMethods)**.[41] Our study follows the TRIPOD + AI reporting guidelines (**eTable 3**).[42] The code for statistical analyses is publicly available at https://github.com/CarDS-Yale/AI-ECG-HF-Pred.

RESULTS

Study Population

From YNHHS, we included 192,667 individuals with a median age of 56 years (IQR, 41-69), comprising 111,181 (57.7%) women, 117,857 (61.2%) non-Hispanic White, 30,623 (15.9%) non-Hispanic Black, and 33,256 (17.3%) Hispanic individuals. Over a median 4.6-year follow-up (IQR, 2.8-6.6), 3,697 (1.9%) had an HF hospitalization, 7,514 (3.9%) had an HF hospitalization or an LVEF below 50% on subsequent echocardiogram, and 10,381 (5.4%) died (**Table 1, eTable 4**).

The 42,141 UKB participants had a median age of 65 years (IQR, 59-71), including 21,795 (51.7%) women, with 40,691 (96.6%) identifying as White and 304 (0.7%) as Black. Over a median follow-up of 3.1 years (IQR, 2.1-4.5), 46 (0.1%) had an HF hospitalization, and 346 (0.8%) died (**Table 1**).

From ELSA-Brasil, the 13,454 participants had a median age of 51 years (IQR, 45-58), comprising 7,348 (54.6%) women, 6,920 (51.4%) adults identifying as White, 2,130 (15.8%) as Black, and 3,767 (28.0%) as “Pardo”. Over a median of 4.2 years (IQR, 3.7-4.5), 31 (0.2%) people developed HF, and 229 (1.7%) died.

Risk Stratification for New-Onset HF

In YNHHS, 42,775 (22.2%) patients screened positive on the AI model applied to the baseline single-lead ECG. A positive screen was associated with over 5-fold higher risk of developing HF (HR 5.05 [95% CI, 4.73-5.39]; **Table 2**). After accounting for differences in age and sex, a positive AI-ECG screen was associated with a 3.3-fold higher risk of HF compared with a negative screen (adjusted HR [aHR], 3.31 [95% CI, 3.10-3.54]). The association remained statistically significant after accounting for differences in HF risk factors of prior ischemic heart disease, hypertension, type 2 diabetes, and obesity (aHR 2.81 [95% CI, 2.63-3.01]) and after additionally accounting for the competing risk of death (aHR of 2.73 [95% CI, 2.55-2.93]). The association of a positive screen with an elevated HF risk was noted across YNHHS sites (**eTable 5**), demographic subgroups (**eTable 6**), and different HF definitions (**eTables 7-8**).

In UKB, 5,513 (13.1%) participants screened positive with the AI-ECG model. A positive AI-ECG screen portended a 7.5-fold higher hazard for developing HF (HR 7.52 [95% CI, 4.21-13.41]). After accounting for age, sex, HF risk factors, and the

competing risk of death, screen-positive participants had a 5-fold higher risk of HF (aHR 5.02 [95% CI, 2.77-9.09]; **Table 2**).

In the ELSA-Brasil cohort, 1,928 (14.3%) participants had a positive AI-ECG screen, with a 9-fold higher HF risk (age- and sex-adjusted HR 8.74 [95% CI, 4.13-18.48]) compared with screen-negative participants. This association was consistent even after accounting for the comorbidities and the competing risk of death (aHR 7.71 [95% CI, 3.62-16.46]).

Risk Across Model Probability Increments

Across the YNHHS network, each 0.1 increment in the model output probability portended a 28% higher hazard of developing HF, adjusted for age, sex, comorbidities, and accounting for the competing risk of death (aHR 1.28 [95% CI, 1.26-1.30]; **Table 2**). Higher model probabilities were progressively associated with higher HF risk across probability bins, with consistent patterns across hospitals and the outpatient network (**eFigure 4**, **eTables 5, 9-10**).

Across UKB and ELSA-Brasil cohorts, a 0.1 increment in model probability was associated with 51% and 66% higher adjusted HF risk (aHR 1.49 [95% CI, 1.36-1.63] and aHR 1.65 [95% CI, 1.46-1.87], respectively, **Table 2**), respectively.

Comparison with PCP-HF and PREVENT

The AI-ECG model had a discrimination based on Harrel's C-statistic of 0.723 (95% CI, 0.694-0.752) in YNHHS, compared with 0.640 (95% CI, 0.612-0.668) for PCP-HF ($p<0.001$), and 0.674 (95% CI, 0.645-0.703) for PREVENT ($p< 0.001$; **Table 3**). In UKB and ELSA-Brasil, the AI-ECG model's discrimination for HF was 0.736 (95% CI, 0.606-0.867) and 0.828 (95% CI, 0.692-0.964), respectively, which was not

significantly different from the clinical risk scores (AI-ECG vs PCP-HF: p_{UKB}=0.96; $p_{ELSA\text{-}Brasil}$=0.80; AI-ECG vs PREVENT: p_{UKB}=0.86; $p_{ELSA\text{-}Brasil}$=0.52). Across cohorts, incorporating AI-ECG predictions in addition to PCP-HF and PREVENT resulted in improved Harrel's C-statistic ($\Delta_{PCP\text{-}HF}$=0.080-0.107; $\Delta_{PREVENT}$=0.069-0.094), compared with the use of the clinical risk equations alone. However, this increase was not statistically significant in the UKB (**Table 3**). Further, in all cohorts, the AI-ECG discrimination for new-onset HF was similar to the base input features for the clinical risk scores (**eTable 11**). Incorporating AI-ECG predictions with the base features resulted in significantly higher Harrel's C-statistics for both PCP-HF and PREVENT input variables in YNHHS and for PREVENT variables in ELSA-Brasil.

Compared with the PCP-HF and PREVENT, the AI-ECG algorithm had a positive IDI across study cohorts. The AI model was associated with a significant improvement in continuous NRI at YNHHS, but not in UKB and ELSA-Brasil (**Table 4**). The AI-ECG model also significantly improved categorical NRI across cohorts, except for PCP-HF in UKB. This improvement in categorical NRI was driven by improved event NRI, while non-event NRI decreased (**eTable 12**). Despite the differential improvement in reclassifying cases and controls, the AI-ECG's PPV was comparable with PCP-HF and PREVENT across sites (**eTable 13**). The AI-ECG model demonstrated consistent superior net benefit over PCP-HF across probability thresholds greater than 0.06 across data sources, where the AI-ECG threshold is 0.08 (**eFigure 5**). A positive AI-ECG screen was an independent predictor of HF risk after accounting for the clinical risk scores, with consistent patterns across racial groups (**eFigures 6-9**).

Non-HF Cardiovascular Outcome Prediction

In YNHHS, a positive AI-ECG screen was associated with a modestly elevated risk of stroke and MACE (age- and sex-adjusted HRs: stroke, 1.17; MACE, 1.77; **eTables 14-15**) compared with a 3-fold increase in HF risk. In UKB and ELSA-Brasil, a positive screen portended a 1.5- to 4-fold hazard of stroke, death, and MACE compared with a 6- to 9-fold increase in HF risk.

DISCUSSION

Across clinically and geographically distinct cohorts, a noise-adapted AI model, trained to detect cross-sectional LVSD from only a lead I ECG, predicted the risk of future HF among individuals seeking outpatient care and community-dwelling adults. Individuals with a positive AI-ECG screen had a 3- to 7-fold higher risk of developing HF compared with those with a negative screen, independent of demographic and clinical characteristics. Higher AI-ECG probabilities were progressively associated with a higher HF risk, with each 10% increment portending a 27-65% higher risk-adjusted hazard for HF across cohorts. Further, the AI-ECG model demonstrated incremental discrimination, improved reclassification, and superior net benefit over PCP-HF and PREVENT. Therefore, our AI-based approach demonstrates promising characteristics for use as a non-invasive digital biomarker for elevated HF risk using a single-lead ECG.

Applications of deep learning for ECGs have demonstrated the ability to identify subtle signatures of structural heart disorders previously considered electrically silent,[29,43–52] with applications extending to detecting LVSD from single-lead tracings.[21,22,53–55] Further, the US Food and Drug Administration recently cleared an AI tool using electronic stethoscope-based single-lead ECGs for cross-sectional LVSD detection.[56] Our study demonstrates that a noise-adapted AI-ECG

model can predict new-onset HF risk using single-lead ECGs. Given the increasing accessibility of portable devices capable of acquiring ECG outside a clinical setting,[20,57,58] this approach can potentially be applied widely to identify individuals at a high risk of HF.[58] While the ECGs acquired with these devices are often distorted by electrode movements or artifacts due to skeletal muscle contraction during acquisition,[27,59] our unique noise-adapted training approach can enable reliable inference from these noisy ECGs.[21]

In this study, we opted for a definition of HF based on the principal discharge diagnosis code, a criterion with high specificity.[31] Nonetheless, the association of a positive screen with elevated HF risk was consistent across several sensitivity analyses defining the condition differently in YNHHS and UKB, and in ELSA-Brasil where the outcomes were explicitly adjudicated. The robust performance across clinically and demographically distinct cohorts indicates that the model captures a predictive HF signature independent of site-specific coding practices.[60–62] Moreover, the dose-dependent association of higher AI-ECG scores with progressively elevated HF risk enables graded risk stratification and risk-informed management. Notably, while a positive screen was also associated with a modestly elevated risk of other cardiovascular outcomes, including MACE, the predictive signature was more specific for HF.

Our study has important implications for defining HF risk. While several clinical risk scores have been proposed to identify those at high risk, these strategies often require clinical evaluation and blood testing.[8,9,16] This limits their scope to patients with established access to healthcare services.[9,16,63–65] In contrast, our AI-based approach using single-lead ECGs may offer a means for HF risk stratification outside clinical settings. Notably, the model demonstrated positive IDI, improved

reclassification, and greater net benefit compared with PCP-HF and PREVENT across sites. While the improvement of discrimination did not reach statistical significance in the UKB and ELSA-Brasil, consistent improvement in categorical NRI–relevant for clinical decision-making–indicates the clinical utility of the AI model.[66] Furthermore, the model's PPV was comparable to PCP-HF and PREVENT, suggesting that an AI-ECG-based strategy for screening will not lead to unnecessary additional testing. Despite these advancements, the AI-ECG model does not eliminate the need for clinical risk scores, but offers a robust and resource-efficient strategy for use in community settings. The risk scores might represent an adjunct in these settings where the focus may be more on the identification of modifiable risk factors.

The ability to use a single portable device to record ECGs for multiple individuals could support the design of efficient community-based screening programs.[67,68] Successful health promotion strategies, such as targeted hypertension management in barbershops and cancer screening in churches across the US,[69,70] can be adapted to promote HF screening, especially among those traditionally less likely to seek preventive care.[63] The ease of use and the brief time for ECG acquisition with these devices can enable a non-laboratory-based strategy, potentially suitable for integration into national-level non-communicable disease screening programs globally, especially in low- and middle-income countries.[67,68,71] This scalability and potential community health benefits necessitate prospective clinical and cost-effectiveness assessments for AI-based HF risk stratification.

Our study has certain limitations. First, waveforms extracted from lead I of clinical ECGs may not be identical to those from portable devices. While our noise-augmentation approach previously demonstrated sustained performance on ECGs

with real-world noises,[21] prospective validation of the model on portable-acquired ECGs is necessary before deployment for community HF screening. This includes evaluating device types, acquisition methods, and handling of ECG segments of longer durations. Second, despite YNHHS's wide geographic coverage, out-of-hospital clinical outcomes may not have been captured, thereby representing a lower HF risk compared with the protocolized follow-up in UKB and ELSA-Brasil. Moreover, while we included only ECGs performed in an outpatient setting, the patients underwent clinically-indicated ECGs, indicating an unmeasured potential risk profile of those who had a negative AI-ECG screen. However, the controls in this setting all underwent ECG screens as well. Third, the number of HF outcome events was low in the UKB and ELSA-Brasil. However, the HF hospitalization rates were comparable to other population-based cohorts,[72,73] and the outcome capture and adjudication in UKB and ELSA-Brasil have been extensively validated.[74–81] Further, in UKB, the smaller subset of participants who underwent an ECG, the shorter follow-up period after the ECGs were performed, and our approach of excluding those with prevalent HF contributed to the lower absolute number of incident HF in our study. Fourth, given the lack of NT-proBNP assessments in UKB and ELSA-Brasil, we could not evaluate NT-proBNP as a comparator in this study. In YNHHS, the use of NT-proBNP could incorporate substantial selection bias, since it is typically ordered for evaluation of cardiopulmonary symptoms and rarely for primary prevention. Nevertheless, a future head-to-head assessment of AI-ECG and NT-proBNP as predictors for HF is warranted. Further, while we performed an analysis that excluded individuals with elevated pre-ECG NT-proBNP levels in YNHHS, the lack of NT-proBNP testing precluded this analysis in UKB and ELSA-Brasil. Fifth, while PCP-HF and PREVENT are utilized to estimate the 10-year risk of HF, we

applied these risk scores to assess HF risk during the available follow-up period. Nevertheless, we factored in varying follow-up durations for each individual for our comparison between risk stratification strategies. Finally, while the AI-ECG approach identifies individuals at elevated HF risk, it is unclear if this risk is modifiable. Nonetheless, a robust screening strategy can enable targeted management of known HF risk factors.

CONCLUSION

Across clinically and geographically distinct cohorts, we used a noise-resilient AI model with a lead I ECG tracing as the sole input to define the risk of future HF, with value over conventional risk scores. With the increasing availability of single-lead ECGs on portable and wearable devices, this AI-ECG-based non-invasive digital biomarker can enable scalable stratification of HF risk across communities.

ACKNOWLEDGMENTS

Author Contributions

Drs Dhingra, Aminorroaya, and Khera had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors approved the final version for submission.

Study concept and design: Dhingra, Lovedeep S; Aminorroaya, Arya; Oikonomou, Evangelos K; Khera, Rohan.

Acquisition, analysis, or interpretation of data: Dhingra, Lovedeep S; Aminorroaya, Arya; Pedroso Camargos, Aline; Khunte, Akshay; Sangha, Veer; McIntyre, Daniel; Chow, Clara K; Asselbergs, Folkert W; Brant, Luisa CC; M Barreto, Sandhi; Ribeiro, Antonio Luiz P; Krumholz, Harlan M; Oikonomou, Evangelos K; Khera, Rohan

Drafting of the manuscript: Dhingra, Lovedeep S; Aminorroaya, Arya; Pedroso Camargos, Aline.

Critical revision of the manuscript for important intellectual content: Khunte, Akshay; Sangha, Veer; McIntyre, Daniel; Chow, Clara K; Asselbergs, Folkert W; Brant, Luisa CC; M Barreto, Sandhi; Ribeiro, Antonio Luiz P; Krumholz, Harlan M; Oikonomou, Evangelos K; Khera, Rohan.

Statistical analysis: Dhingra, Lovedeep S; Aminorroaya, Arya.

Obtained funding: Khera, Rohan.

Administrative, technical, or material support: Aminorroaya, Arya; Dhingra, Lovedeep S; Pedroso Camargos, Aline; Khera, Rohan.

Study supervision: Khera, Rohan.

Conflict of Interest Disclosures

Dr. Khera is an Associate Editor of JAMA. Dr. Khera and Mr. Sangha are the coinventors of U.S. Provisional Patent Application No. 63/346,610, “Articles and methods for format-independent detection of hidden cardiovascular disease from printed electrocardiographic images using deep learning” and are co-founders of Ensight-AI. Dr. Khera receives support the National Institutes of Health (under awards R01AG089981, R01HL167858, and K23HL153775) and the Doris Duke Charitable Foundation (under award 2022060). He receives support from the Blavatnik Foundation through the Blavatnik Fund for Innovation at Yale. He also receives research support, through Yale, from Bristol-Myers Squibb, BridgeBio, and Novo Nordisk. In addition to 63/346,610, Dr. Khera is a coinventor of U.S. Pending Patent Applications WO2023230345A1, US20220336048A1, 63/484,426, 63/508,315, 63/580,137, 63/606,203, 63/619,241, and 63/562,335. Dr. Khera and

Dr. Oikonomou are co-founders of Evidence2Health, a precision health platform to improve evidence-based cardiovascular care. Dr. Oikonomou is a co-inventor of the U.S. Patent Applications 63/508,315 & 63/177,117 and has been a consultant to Caristo Diagnostics Ltd (all outside the current work). Dr. Krumholz works under contract with the Centers for Medicare & Medicaid Services to support quality measurement programs. He is associated with research contracts through Yale University from Janssen, Kenvue, and Pfizer. In the past three years, Dr. Krumholz received options for Element Science and Identifeye and payments from F-Prime for advisory roles. He is a co-founder of and holds equity in Hugo Health, Refactor Health, and Ensight-AI. Dr. Ribeiro is supported in part by the National Council for Scientific and Technological Development - CNPq (grants 465518/2014-1, 310790/2021-2, 409604/2022-4 e 445011/2023-8). Dr. Brant is supported in part by CNPq (307329/2022-4). Dr. Asselbergs is supported by Heart4Data, which received funding from the Dutch Heart Foundation and ZonMw (2021-B015), and UCL Hospitals NIHR Biomedical Research Centre.

Funding/Support

Dr. Khera was supported by the National Institutes of Health (under awards R01AG089981, R01HL167858, and K23HL153775) and the Doris Duke Charitable Foundation (under award 2022060). Dr. Oikonomou was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (under award F32HL170592).

Role of the Funder/Sponsor

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer

The views expressed in this article are those of the authors and not necessarily any funders.

Data Sharing Statement

The data from the Yale New Haven Health System represent protected health information. To protect patient privacy, the Yale Institutional Review Board does not allow the sharing of these data. Data from the UK Biobank and the Brazilian Longitudinal Study of Adult Health are available for research to licensed users. The code for cohort creation and statistical analyses is publicly available at https://github.com/CarDS-Yale/AI-ECG-HF-Pred.

REFERENCES

1. Shahim B, Kapelios CJ, Savarese G, Lund LH. Global public health burden of heart failure: An updated review. *Card Fail Rev*. 2023;9. doi:10.15420/cfr.2023.05

2. Bozkurt B, Ahmad T, Alexander KM, et al. Heart failure epidemiology and outcomes statistics: A report of the heart failure society of America. *J Card Fail*. 2023;29(10):1412-1451.

3. Butler J, Kalogeropoulos A, Georgiopoulou V, et al. Incident Heart Failure Prediction in the Elderly. *Circ Heart Fail*. 2008;1(2):125-133.

4. Khan SS, Ning H, Shah SJ, et al. 10-Year Risk Equations for Incident Heart Failure in the General Population. *J Am Coll Cardiol*. 2019;73(19):2388-2397.

5. Khan SS, Matsushita K, Sang Y, et al. Development and validation of the American heart association's PREVENT equations. *Circulation*. 2024;149(6):430-449.

6. Pandey A, Khan MS, Patel KV, Bhatt DL, Verma S. Predicting and preventing heart failure in type 2 diabetes. *Lancet Diabetes Endocrinol*. 2023;11(8):607-624.

7. Kalogeropoulos A, Psaty BM, Vasan RS, et al. Validation of the Health ABC Heart Failure Model for Incident Heart Failure Risk Prediction. *Circ Heart Fail*. 2010;3(4):495-502.

8. Echouffo-Tcheugui JB, Greene SJ, Papadimitriou L, et al. Population Risk Prediction Models for Incident Heart Failure. *Circ Heart Fail*. 2015;8(3):438-447.

9. Nadarajah R, Younsi T, Romer E, et al. Prediction models for heart failure in the community: A systematic review and meta-analysis. *Eur J Heart Fail*. 2023;25(10):1724-1738.

10. Agarwal SK, Chambless LE, Ballantyne CM, et al. Prediction of Incident Heart Failure in General Practice. *Circ Heart Fail*. 2012;5(4):422-429.

11. Grewal J, McKelvie R, Lonn E, et al. BNP and NT-proBNP predict echocardiographic severity of diastolic dysfunction. *Eur J Heart Fail*. 2008;10(3):252-259.

12. Campbell DJ, Gong FF, Jelinek MV, et al. Prediction of incident heart failure by serum amino-terminal pro-B-type natriuretic peptide level in a community-based cohort. *Eur J Heart Fail*. 2019;21(4):449-459.

13. Yan I, Börschel CS, Neumann JT, et al. High-sensitivity cardiac troponin I levels and prediction of heart failure: Results from the BiomarCaRE consortium. *JACC Heart Fail*. 2020;8(5):401-411.

14. Kelder JC, Cowie MR, McDonagh TA, et al. Quantifying the added value of BNP in suspected heart failure in general practice: an individual patient data meta-analysis. *Heart*. 2011;97(12):959-963.

15. Jia X, Al Rifai M, Hoogeveen R, et al. Association of Long-term Change in N-Terminal Pro-B-Type Natriuretic Peptide With Incident Heart Failure and Death. *JAMA Cardiol*. 2023;8(3):222-230.

16. Watson CJ, Gallagher J, Wilkinson M, et al. Biomarker profiling for risk of future heart failure (HFpEF) development. *J Transl Med*. 2021;19(1):61.

17. Al-Alusi MA, Khurshid S, Wang X, et al. Trends in Consumer Wearable Devices With Cardiac Sensors in a Primary Care Cohort. *Circ Cardiovasc Qual Outcomes*. 2022;15(7):e008833.

18. Gehr S, Russmann C. Shaping the future of cardiovascular medicine in the new era of wearable devices. *Nat Rev Cardiol*. 2022;19(8):501-502.

19. Aminorroaya Arya, Dhingra Lovedeep S., Nargesi Arash A., Oikonomou Evangelos K., Krumholz Harlan M., Khera Rohan. Use of Smart Devices to Track Cardiovascular Health Goals in the United States. *JACC: Advances*. 2023;2(7):100544.

20. Dhingra LS, Aminorroaya A, Oikonomou EK, et al. Use of Wearable Devices in Individuals With or at Risk for Cardiovascular Disease in the US, 2019 to 2020. *JAMA Netw Open*. 2023;6(6):e2316634.

21. Khunte A, Sangha V, Oikonomou EK, et al. Detection of left ventricular systolic dysfunction from single-lead electrocardiography adapted for portable and wearable devices. *NPJ Digit Med*. 2023;6(1):124.

22. Aminorroaya A, Dhingra LS, Pedroso Camargos A, et al. Development and multinational validation of an ensemble deep learning algorithm for detecting and predicting structural heart disease using noisy single-lead electrocardiograms. *medRxiv*. Published online October 8, 2024:2024.10.07.24314974.

23. Shankar SV, Oikonomou EK, Khera R. CarDS-Plus ECG Platform: Development and Feasibility Evaluation of a Multiplatform Artificial Intelligence Toolkit for Portable and Wearable Device Electrocardiograms. *medRxiv*. Published online October 3, 2023. doi:10.1101/2023.10.02.23296404

24. Mathews SM, Kambhamettu C, Barner KE. A novel application of deep learning for single-lead ECG classification. *Comput Biol Med*. 2018;99:53-62.

25. Dupulthys S, Dujardin K, Anné W, et al. Single-lead electrocardiogram Artificial Intelligence model with risk factors detects atrial fibrillation during sinus rhythm. *Europace*. 2024;26(2):euad354.

26. Gibson CM, Mehta S, Ceschim MRS, et al. Evolution of single-lead ECG for STEMI detection using a deep learning approach. *Int J Cardiol*. 2022;346:47-52.

27. Bläsing D, Buder A, Reiser JE, Nisser M, Derlien S, Vollmer M. ECG performance in simultaneous recordings of five wearable devices using a new morphological noise-to-signal index and Smith-Waterman-based RR interval comparisons. *PLoS One*. 2022;17(10):e0274994.

28. Oikonomou EK, Sangha V, Dhingra LS, et al. Artificial intelligence-enhanced risk stratification of cancer therapeutics-related cardiac dysfunction using electrocardiographic images. *Circ Cardiovasc Qual Outcomes*. Published online September 2, 2024. doi:10.1161/CIRCOUTCOMES.124.011504

29. Dhingra LS, Aminorroaya A, Sangha V, et al. Heart failure risk stratification using artificial intelligence applied to electrocardiogram images: a multinational study. *Eur Heart J*. Published online January 13, 2025. doi:10.1093/eurheartj/ehae914

30. Haji K, Huynh Q, Wong C, Stewart S, Carrington M, Marwick TH. Improving the characterization of stage A and B heart failure by adding global longitudinal strain. *JACC Cardiovasc Imaging*. 2022;15(8):1380-1387.

31. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of heart failure diagnoses in administrative databases: A systematic review and meta-analysis. *PLoS One*. 2014;9(8):e104519.

32. Dicionários de variáveis e Data Books. Elsa. October 26, 2021. Accessed March 6, 2024. http://elsabrasil.org/pesquisadores/dicionarios-de-variaveis-e-data-books/

33. Khan SS, Coresh J, Pencina MJ, et al. Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: A scientific statement from the American Heart Association. *Circulation*. 2023;148(24):1982-2004.

34. Littlejohns TJ, Sudlow C, Allen NE, Collins R. UK Biobank: opportunities for cardiovascular research. *Eur Heart J*. 2019;40(14):1158-1166.

35. Elliott P, Peakman TC, UK Biobank. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol*. 2008;37(2):234-244.

36. Mill JG, Pinto K, Griep RH, et al. Medical assessments and measurements in ELSA-Brasil. *Rev Saude Publica*. 2013;47 Suppl 2:54-62.

37. Bensenor IM, Griep RH, Pinto KA, 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.

38. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. 2017;36(27):4391-4400.

39. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med*. 2015;34(4):685-703.

40. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11-21.

41. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:i6.

42. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:q902.

43. Sangha V, Nargesi AA, Dhingra LS, et al. Detection of Left Ventricular Systolic Dysfunction From Electrocardiographic Images. *Circulation*. Published online July 25, 2023. doi:10.1161/CIRCULATIONAHA.122.062646

44. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25(1):70-74.

45. Bjerkén LV, Rønborg SN, Jensen MT, Ørting SN, Nielsen OW. Artificial intelligence enabled ECG screening for left ventricular systolic dysfunction: a systematic review. *Heart Fail Rev*. 2023;28(2):419-430.

46. Dhingra LS, Sangha V, Aminorroaya A, et al. A multicenter evaluation of the impact of therapies on deep learning-based electrocardiographic hypertrophic cardiomyopathy markers. *Am J Cardiol*. 2024;237:35-40.

47. König S, Hohenstein S, Nitsche A, et al. Artificial intelligence-based identification of left ventricular systolic dysfunction from 12-lead electrocardiograms: External validation and advanced application of an existing model. *Eur Heart J Digit Health*. doi:10.1093/ehjdh/ztad081

48. Acharya UR, Fujita H, Oh SL, et al. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. *Applied Intelligence*. 2019;49(1):16-27.

49. Sudarshan VK, Acharya UR, Oh SL, et al. Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2s of ECG signals. *Comput Biol Med*. 2017;83:48-58.

50. Sangha V, Dhingra LS, Oikonomou EK, et al. Identification of hypertrophic cardiomyopathy on electrocardiographic images with deep learning. *medRxiv*. Published online December 28, 2023:2023.12.23.23300490. doi:10.1101/2023.12.23.23300490

51. Aminorroaya A, Dhingra LS, Sangha V, et al. Deep learning-enabled detection of aortic stenosis from noisy single lead electrocardiograms. *medRxiv*. Published online 2023. doi:10.1101/2023.09.29.23296310

52. Dhingra LS, Aminorroaya A, Sangha V, et al. An ensemble deep learning algorithm for structural heart disease screening using electrocardiographic

images: PRESENT SHD. *medRxiv*. Published online October 7, 2024. doi:10.1101/2024.10.06.24314939

53. Attia ZI, Harmon DM, Dugan J, et al. Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. *Nat Med*. 2022;28(12):2497-2503.

54. Attia ZI, Dugan J, Rideout A, et al. Automated detection of low ejection fraction from a one-lead electrocardiogram: application of an AI algorithm to an electrocardiogram-enabled Digital Stethoscope. *Eur Heart J Digit Health*. 2022;3(3):373-379.

55. Bachtiger P, Petri CF, Scott FE, et al. Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during ECG-enabled stethoscope examination in London, UK: a prospective, observational, multicentre study. *Lancet Digit Health*. 2022;4(2):e117-e125.

56. U.S. Food & Drug Administration. 21 CFR 870.2380 Eko Low Ejection Fraction Tool (ELEFT). Published online March 28, 2024. https://www.accessdata.fda.gov/cdrh_docs/pdf23/K233409.pdf

57. Kumar S, Victoria-Castro AM, Melchinger H, et al. Wearables in cardiovascular disease. *J Cardiovasc Transl Res*. Published online September 9, 2022. doi:10.1007/s12265-022-10314-0

58. Singhal A, Cowie MR. The role of wearables in heart failure. *Curr Heart Fail Rep*. 2020;17(4):125-132.

59. Lacirignola F, Pasero E. Hardware design of a wearable ECG-sensor: Strategies implementation for improving CMRR and reducing noise. In: *2017 European Conference on Circuit Theory and Design (ECCTD)*. IEEE; 2017. doi:10.1109/ecctd.2017.8093244

60. 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.

61. Nargesi AA, Adejumo P, Dhingra LS, et al. Automated identification of heart failure with reduced ejection fraction using deep learning-based natural language processing. *JACC Heart Fail*. Published online October 9, 2024. doi:10.1016/j.jchf.2024.08.012

62. Adejumo P, Thangaraj PM, Dhingra LS, et al. Natural language processing of clinical documentation to assess functional status in patients with heart failure. *JAMA Netw Open*. 2024;7(11):e2443925.

63. Sandhu AT, Tisdale RL, Rodriguez F, et al. Disparity in the setting of incident heart failure diagnosis. *Circ Heart Fail*. 2021;14(8). doi:10.1161/circheartfailure.121.008538

64. Lee WC, Serag H, Ohsfeldt RL, et al. Racial disparities in type of heart failure and hospitalization. *J Immigr Minor Health*. 2019;21(1):98-104.

65. Aminorroaya A, Dhingra LS, Oikonomou EK, et al. Development and multinational validation of an algorithmic strategy for high Lp(a) screening. *Nat Cardiovasc Res*. 2024;3(5):558-566.

66. Cook NR. Clinically relevant measures of fit? A note of caution. *Am J Epidemiol*. 2012;176(6):488-491.

67. Shah K, Pandya A, Kotwani P, et al. Cost-effectiveness of portable electrocardiogram for screening cardiovascular diseases at a primary health center in Ahmedabad district, India. *Front Public Health*. 2021;9. doi:10.3389/fpubh.2021.753443

68. Chen W, Khurshid S, Singer DE, et al. Cost-effectiveness of Screening for Atrial Fibrillation Using Wearable Devices. *JAMA Health Forum*. 2022;3(8):e222419-e222419.

69. Victor RG, Lynch K, Li N, et al. A cluster-randomized trial of blood-pressure reduction in black barbershops. *N Engl J Med*. 2018;378(14):1291-1301.

70. McNeill LH, Reitzel LR, Escoto KH, et al. Engaging black churches to address cancer health disparities: Project CHURCH. *Front Public Health*. 2018;6. doi:10.3389/fpubh.2018.00191

71. Proietti M, Farcomeni A, Goethals P, et al. Cost-effectiveness and screening performance of ECG handheld machine in a population screening programme: The Belgian Heart Rhythm Week screening programme. *Eur J Prev Cardiol*. 2019;26(9):964-972.

72. Ciapponi A, Alcaraz A, Calderón M, et al. Burden of heart failure in Latin America: A systematic review and meta-analysis. *Rev Esp Cardiol (Engl Ed)*. 2016;69(11):1051-1060.

73. Conrad N, Judge A, Tran J, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet*. 2018;391(10120):572-580.

74. Banerjee A, Dashtban A, Chen S, et al. Identifying subtypes of heart failure from three electronic health record sources with machine learning: an external, prognostic, and genetic validation study. *Lancet Digit Health*. 2023;5(6):e370-e379.

75. Oguntade AS, Taylor H, Lacey B, Lewington S. Adiposity, fat-free mass and incident heart failure in 500 000 individuals. *Open Heart*. 2024;11(2):e002711.

76. Ho FK, Zhou Z, Petermann-Rocha F, et al. Association between device-measured physical activity and incident heart failure: A prospective cohort study of 94 739 UK Biobank participants. *Circulation*. 2022;146(12):883-891.

77. Aragam KG, Chaffin M, Levinson RT, et al. Phenotypic refinement of heart failure in a national biobank facilitates genetic discovery. *Circulation*. 2019;139(4):489-501.

78. Rastogi T, Ho FK, Rossignol P, et al. Comparing and contrasting risk factors for heart failure in patients with and without history of myocardial infarction: data from HOMAGE and the UK Biobank. *Eur J Heart Fail*. 2022;24(6):976-984.

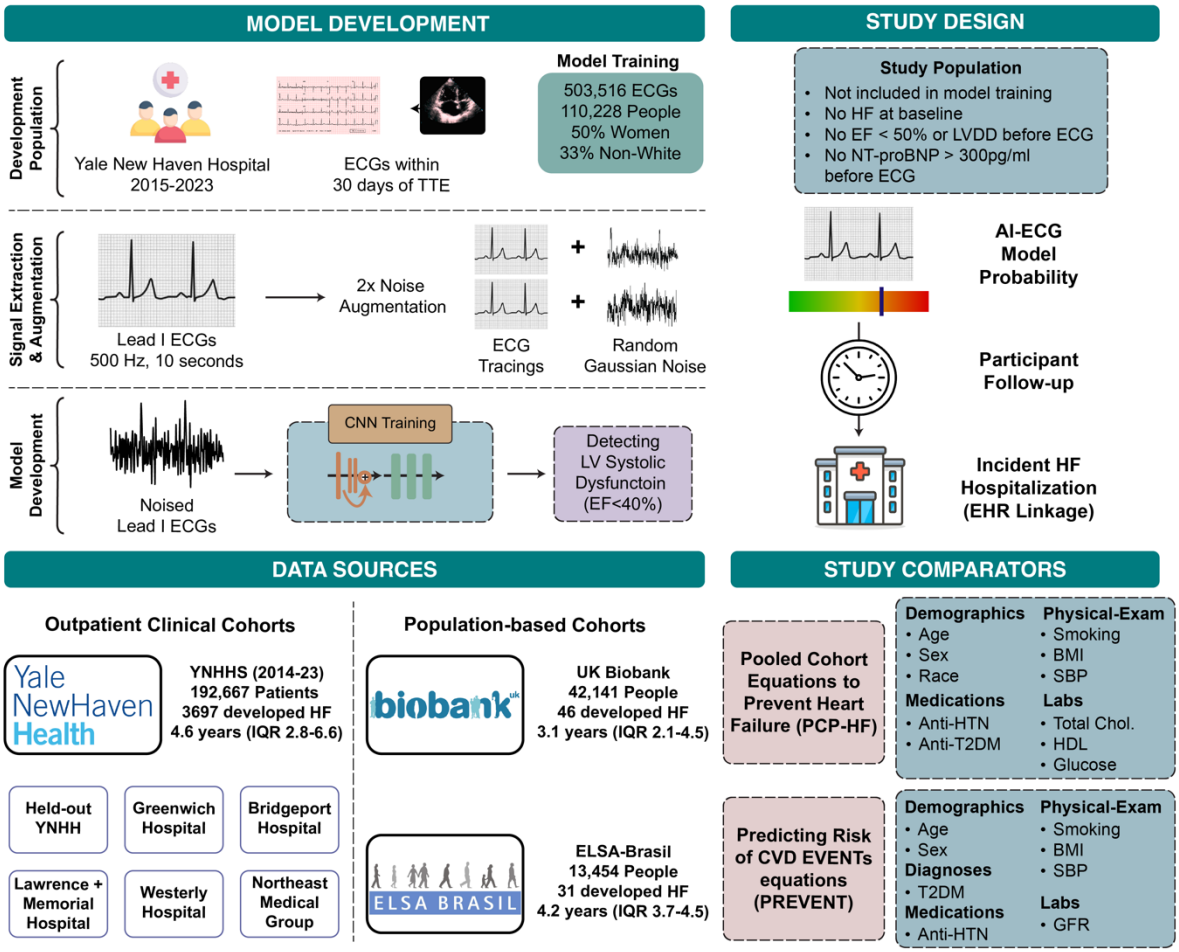
79. Pedroso Camargos A, Barreto S, Brant L, et al. Performance of contemporary cardiovascular risk stratification scores in Brazil: an evaluation in the ELSA-Brasil study. *Open Heart*. 2024;11(1):e002762.

80. Schmidt MI, Duncan BB, Mill JG, et al. Cohort profile: Longitudinal study of adult health (ELSA-brasil). *Int J Epidemiol*. 2015;44(1):68-75.

81. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026-1034.

FIGURES

Figure 1. Study Overview. Abbreviations: BMI, Body Mass Index; BP; Blood Pressure; CNN, Convolutional Neural Network; ECG, Electrocardiogram; EF, Ejection Fraction; EHR, Electronic Health Record; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; HDL, High-density Lipoprotein Cholesterol; HF, Heart Failure; LV, Left Ventricle; YNH, Yale New Haven Hospital; YNHHS, Yale New Haven Health System.



TABLES

Table 1. Population Characteristics of the Study Cohorts. Abbreviations: AMI, acute myocardial infarction; ECG, Electrocardiogram; ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; HF, heart failure; IQR, Interquartile Range; LVEF, Left Ventricular Ejection Fraction; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Characteristic		YNHHS	UKB	ELSA-Brasil
Number		192667	42141	13454
Age at ECG, Median [IQR]		56.1 [41.1,68.7]	65 [59,71]	51 [45,58]
Female Sex, N (%)		111181 (57.7)	21795 (51.7)	7348 (54.6)
Race/Ethnicity, N (%)	**Asian**	3553 (1.8)	600 (1.4)	0 (0.0)
	Black	30623 (15.9)	304 (0.7)	2130 (15.8)
	Hispanic	33256 (17.3)	0 (0.0)	0 (0.0)
	Brazilian "Pardo"	0 (0.0)	0 (0.0)	3767 (28.0)
	Missing	5219 (2.7)	0 (0.0)	0 (0.0)
	Other[a]	2159 (1.1)	546 (1.3)	637 (4.7)
	White	117857 (61.2)	40691 (96.6)	6920 (51.4)
Death, N (%)		10381 (5.4)	346 (0.8)	229 (1.7)
Follow-up Time, Years; Median [IQR]		4.6 [2.8,6.6]	3.1 [2.1,4.5]	4.2 [3.7, 4.5]
Positive Screens, N (%)		42775 (22.2)	5513 (13.1)	1928 (14.3)
Hypertension at baseline, N (%)		88215 (45.8)	6126 (14.5)	4739 (35.3)
Type-2 diabetes mellitus at baseline, N (%)		35522 (18.4)	1258 (3.0)	2105 (15.6)
Obesity at baseline, N (%)		30493 (15.8)	7535 (17.9)	3045 (22.6)
Atrial fibrillation at baseline, N (%)		4746 (2.5)	637 (1.5)	- [b]
Left bundle branch block at baseline, N (%)		2397 (1.2)	383 (0.9)	- [b]
Use of antihypertensive drugs at baseline, N (%)		47611 (24.7)	9936 (23.9)	3640 (27.1)
Use of antihyperglycemic drugs at baseline, N (%)		30520 (15.8)	321 (0.8)	1072 (8.0)
End-stage renal disease, N (%)		547 (0.3)	0 (0.0)	10 (0.1)
Primary HF hospitalization during follow-up, N (%)		3697 (1.9)	46 (0.1)	31 (0.2)
Primary HF hospitalization or an echocardiogram with LVEF < 50% during follow-up, N (%)		7514 (3.9)	- [c]	- [c]
Any HF hospitalization during follow-up, N (%)		13705 (7.1)	231 (0.5)	- [c]
Any HF hospitalization or an echocardiogram with LVEF < 50% during follow-up, N (%)		15705 (8.2)	- [c]	- [c]
Primary AMI hospitalization during follow-up, N (%)		366 (0.2)	208 (0.5)	60 (0.4)
Primary Stroke hospitalization during follow-up, N (%)		3281 (1.7)	210 (0.5)	59 (0.4)
Major Adverse Cardiovascular Events during follow-up, N (%)		16039 (8.3)	768 (1.8)	338 (2.5)

[a] 'Other' races included Native Americans, Pacific Islanders, and mixed races.
[b] ECG-level information about rhythm and conduction disorders not available in the ELSA-Brasil cohort
[c] Follow-up echocardiogram data not available in UKB or ELSA-Brasil.

Table 2. Model Performance for Predicting Heart Failure Risk Based on AI-ECG Probability. Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; IHD, Ischemic Heart Disease; HTN, hypertension; T2DM, type-2 diabetes mellitus; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Model Type	Predictive Model Inputs	YNHHS		UKB		ELSA-Brasil	
		Positive Screen	Per 0.1 Increment	Positive Screen	Per 0.1 Increment	Positive Screen	Per 0.1 Increment
Cox Proportional Hazard Model	AI-ECG Probability	5.05 (4.73-5.39)	1.45 (1.44-1.47)	7.52 (4.21-13.41)	1.55 (1.40-1.71)	11.11 (5.32-23.19)	1.83 (1.64-2.05)
Cox Proportional Hazard Model	AI-ECG Probability + Age + Sex	3.31 (3.10-3.54)	1.32 (1.30-1.34)	5.96 (3.32-10.68)	1.52 (1.37-1.68)	8.74 (4.13-18.48)	1.75 (1.56-1.97)
Cox Proportional Hazard Model	AI-ECG Probability + Age + Sex + IHD + HTN + T2DM + Obesity	2.81 (2.63-3.01)	1.28 (1.26-1.30)	5.02 (2.77-9.09)	1.49 (1.33-1.66)	7.71 (3.62-16.46)	1.72 (1.52-1.93)
Fine-Gray Subdistribution Hazard Model	AI-ECG Probability + Age + Sex and accounting for competing risk of death	3.22 (3.01-3.45)	1.30 (1.29-1.32)	5.91 (3.33-10.50)	1.51 (1.38-1.66)	8.67 (4.02-18.70)	1.74 (1.55-1.96)
Fine-Gray Subdistribution Hazard Model	AI-ECG Probability + Age + Sex + IHD + HTN + T2DM + Obesity and accounting for competing risk of death	2.73 (2.55-2.93)	1.27 (1.25-1.28)	4.99 (2.81-8.87)	1.49 (1.36-1.63)	6.53 (2.91-14.67)	1.65 (1.46-1.87)

Table 3. Comparison of Discrimination for AI-ECG Model Output Probability with Pooled Cohort Equations to Prevent Heart Failure and Predicting Risk of Cardiovascular Disease Events Equations for Predicting Incident Heart Failure.
Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; PCP-HF, Pooled Cohort Equations to Prevent Heart Failure; PREVENT, Predicting Risk of Cardiovascular Disease Events; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Clinical Risk Score	Covariates	YNHHS			UKB			ELSA-Brasil		
		Harrel's C-statistic	Marginal difference over Harrel's C-statistic for clinical risk score	P-value	Harrel's C-statistic	Marginal difference over Harrel's C-statistic for clinical risk score	P-value	Harrel's C-statistic	Marginal difference over Harrel's C-statistic for clinical risk score	P-value
PCP-HF	PCP-HF	0.640 (0.612 - 0.668)	-	-	0.732 (0.620 - 0.844)	-	-	0.850 (0.789 - 0.912)	-	-
	AI-ECG Model Output Probability	0.723 (0.694 - 0.752)	0.083 (0.044 - 0.122)	< 0.001	0.736 (0.606 - 0.867)	0.004 (-0.165 - 0.173)	0.96	0.828 (0.692 - 0.964)	-0.023 (-0.194 - 0.149)	0.80
	AI-ECG Model Output Probability + Age + Sex	0.720 (0.692 - 0.748)	0.081 (0.049 - 0.112)	< 0.001	0.800 (0.707 - 0.894)	0.068 (-0.060 - 0.196)	0.30	0.897 (0.820 - 0.975)	0.047 (-0.064 - 0.157)	0.41
	AI-ECG Model Output Probability + PCP-HF	0.747 (0.721 - 0.773)	0.107 (0.078 - 0.136)	< 0.001	0.812 (0.722 - 0.902)	0.080 (-0.013 - 0.172)	0.09	0.935 (0.898 - 0.971)	0.084 (0.010 - 0.160)	0.03
PREVENT	PREVENT	0.674 (0.645 - 0.703)	-	-	0.753 (0.635 - 0.871)	-	-	0.882 (0.762 - 0.906)	-	-
	AI-ECG Model Output Probability	0.723 (0.694 - 0.752)	0.049 (0.009 - 0.088)	0.02	0.736 (0.606 - 0.867)	-0.017 (-0.197 - 0.164)	0.86	0.828 (0.692 - 0.964)	-0.054 (-0.218 - 0.111)	0.52
	AI-ECG Model Output Probability + Age + Sex	0.720 (0.692 - 0.748)	0.046 (0.012 - 0.080)	0.007	0.800 (0.707 - 0.894)	0.047 (-0.088 - 0.182)	0.49	0.897 (0.820 - 0.975)	0.016 (-0.088 - 0.119)	0.77
	AI-ECG Model Output Probability + PREVENT	0.768 (0.742 - 0.793)	0.094 (0.068 - 0.120)	< 0.001	0.822 (0.730 - 0.913)	0.069 (-0.019 - 0.157)	0.12	0.950 (0.927 - 0.974)	0.069 (0.011 - 0.127)	0.02

Table 4. Integrated Discrimination Improvement and Categorical and Continuous Time-to-Event Net Reclassification Index of AI-ECG Model Output Probability over Pooled Cohort Equations to Prevent Heart Failure and Predicting Risk of Cardiovascular Disease Events Equations for Heart Failure. Abbreviations: ELSA-Brasil, Brazilian Longitudinal Study of Adult Health; IDI, Integrated Discrimination Improvement; NRI, Net Reclassification Index; PCP-HF, Pooled Cohort Equations to Prevent Heart Failure; PREVENT, Predicting Risk of Cardiovascular Disease Events; UKB, UK Biobank; YNHHS, Yale New Haven Health System.

Metric	YNHHS		UKB		ELSA-Brasil	
	PCP-HF	PREVENT	PCP-HF	PREVENT	PCP-HF	PREVENT
IDI	0.091 (0.068 to 0.118)	0.068 (0.044 to 0.098)	0.103 (0.011 to 0.214)	0.113 (0.024 to 0.211)	0.205 (0.075 to 0.347)	0.192 (0.064 to 0.339)
Categorical NRI	0.182 (0.100 to 0.263)	0.118 (0.034 to 0.199)	0.198 (-0.076 to 0.465)	0.289 (0.017 to 0.537)	0.472 (0.131 to 0.749)	0.475 (0.173 to 0.809)
Continuous NRI	0.210 (0.094 to 0.325)	0.207 (0.094 to 0.323)	0.096 (-0.347 to 0.506)	0.309 (-0.140 to 0.724)	0.095 (-0.242 to 0.324)	0.188 (-0.268 to 0.531)