



# Identification of cardiac wall motion abnormalities in diverse populations by deep learning of the electrocardiogram



Albert J. Rogers <sup>1,2</sup>✉, Neal K. Bhatia<sup>3</sup>, Sabyasachi Bandyopadhyay<sup>1,2</sup>, James Tooley<sup>1,2</sup>, Rayan Ansari<sup>1,2</sup>, Vyom Thakkar<sup>1</sup>, Justin Xu<sup>4</sup>, Jessica Torres Soto <sup>1,2</sup>, Jagteshwar S. Tung<sup>1,2</sup>, Mahmood I. Alhusseini<sup>1,2</sup>, Paul Clopton<sup>1</sup>, Reza Sameni <sup>5,6</sup>, Gari D. Clifford<sup>5,6</sup>, J. Weston Hughes <sup>1,2</sup>, Euan A. Ashley <sup>1,2</sup>, Marco V. Perez <sup>1,2</sup>, Matei Zaharia<sup>7</sup> & Sanjiv M. Narayan <sup>1,2</sup>

Cardiac wall motion abnormalities (WMA) are strong predictors of mortality, but current screening methods using Q waves from electrocardiograms (ECGs) have limited accuracy and vary across racial and ethnic groups. This study aimed to identify novel ECG features using deep learning to enhance WMA detection, referencing echocardiography as the gold standard. We collected ECG and echocardiogram data from 35,210 patients in California and labeled WMA using unstructured language parsing of echocardiographic reports. A deep neural network (ECG-WMA-Net) was trained and outperformed both expert ECG interpretation and Q-wave indices, achieving an AUROC of 0.781 (CI: 0.762–0.799). The model was externally validated in a diverse cohort from Georgia ( $n = 2338$ ), with an AUC of 0.723 (CI: 0.685–0.757). Explainability analysis revealed significant contributions from QRS and T-wave regions. This deep learning approach improves WMA screening accuracy, potentially addressing physiological differences not captured by standard ECG-based methods.

Normal heart function requires coordinated contraction. Wall motion abnormalities (WMAs) of the heart substantially increase adverse outcomes<sup>1–3</sup> including sudden death and all-cause mortality in patients with ischemic<sup>4</sup> or non-ischemic<sup>5</sup> heart disease and even in those without an apparent cardiac history<sup>6</sup>. WMA can result from myocardial infarction but also a multitude of non-ischemic conditions, including cardiac sarcoidosis, myocarditis, takotsubo syndrome, and hypertrophic cardiomyopathy and, notably, predicts events independent of reduced systolic function of the heart<sup>7–9</sup>.

Current screening approaches for WMA have low sensitivity and specificity<sup>10–12</sup>, and primarily assess the electrocardiogram (ECG) to detect Q waves, T-waves, ST segment alterations or indices such as the Cornell product<sup>13,14</sup>. Accordingly, patients with a clinical suspicion of WMA must subsequently undergo confirmatory echocardiography or other imaging studies<sup>15</sup>, which may be unavailable at primary care facilities and in underserved regions, and introduce delays<sup>16</sup>. Studies with readily labeled endpoints have shown that AI-enabled algorithms can detect several abnormalities from the ECG, including reduced left ventricular ejection fraction (LVEF), propensity for atrial fibrillation, and predict all-cause mortality<sup>17–21</sup>.

We hypothesized that a deep neural network (ECG-WMA-Net) model trained on the 12-lead ECG could identify WMA with higher accuracy than analysis of standard ECG indices reported by existing ECG machines or based on qualitative ECG interpretations by physicians. We set out to develop and test our models in California in a large population, addressing the challenges of obtaining labels for WMA from complex imaging data using natural language processing (NLP) of unstructured full-text echocardiography reports in the electronic health record system. We tested the generalizability of the trained model (ECG-WMA-Net) in an external population of patients in Georgia who differed substantially in ethnic and demographic features and comorbidities.

## Results

The development cohort included 35,210 patients who underwent ECG and echocardiography at Stanford University, California, and the external validation cohort comprised 2338 unique patients at Emory Healthcare, Georgia, USA. Demographics of the two groups are contrasted in Table 1, which shows their important differences in race, ethnicity, and the presence of comorbidities known to influence cardiac disease. The overall prevalence

<sup>1</sup>Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. <sup>2</sup>Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, USA. <sup>3</sup>Department of Medicine, Emory University, Atlanta, GA, USA. <sup>4</sup>Department of Computer Science, Stanford University, Stanford, CA, USA. <sup>5</sup>Department of Biomedical Informatics, Emory University, Atlanta, GA, USA. <sup>6</sup>Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

<sup>7</sup>Department of Computer Science, UC Berkeley, Berkeley, CA, USA. ✉e-mail: [rogersaj@stanford.edu](mailto:rogersaj@stanford.edu)

**Table 1 | Demographic and clinical characteristics of patients from Stanford and Emory University cohorts**

	Stanford (N = 35,210)	Emory (N = 2338)	p value
Age (years)	63.4 ± 16.9	61.9 ± 16.0	<0.001
Male	52.6%	51.3%	0.230
Race	-	-	<0.001
Asian	16.6%	3.2%	-
Black	5.6%	41.5%	-
Native American	0.4%	0.2%	-
Other	16.4%	0.6%	-
Pacific Islander	1.2%	0.2%	-
Unknown	4.8%	5.7%	-
White	55.0%	48.7%	-
Ethnicity	-	-	<0.001
Hispanic/ Latino	11.1%	2.8%	-
Non-Hispanic	82.6%	86.0%	-
Unknown	6.3%	11.2%	-
AT/AF	30.6%	33.9%	0.679
CAD	4.1%	10.2%	0.044
HTN	62.9%	81.4%	0.005
HF	27.3%	39.0%	0.063
CVA	6.0%	15.3%	0.007
Echocardiography			
WMA	10.7%	8.9%	0.006
EF (%)	58.9 ± 11.2	56.2 ± 10.4	<0.001

AT/AF atrial tachycardia/atrial fibrillation, CAD coronary artery disease, CVA cardiovascular accident, EF ejection fraction, HF heart failure, HTN hypertension, WMA wall motion abnormality

of WMA in the Stanford dataset was significantly greater than in the Emory dataset (10.7 vs 8.9%,  $p = 0.006$ ). They were further stratified into training, validation, and testing cohorts (Supplementary Table 1). Patients at both centers underwent ECG and echocardiographic assessment within 60 days. Figure 1 shows data flow in the study. Stanford University and Emory University Institutional Review Board approvals were obtained for this study.

#### Natural language processing of electronic health records for data labeling

The clinical records for electrocardiogram interpretations and echocardiogram reports were obtained as unstructured text. We developed a novel scalable approach using NLP to objectively assign ground truth labels of WMA from echocardiography reports, and qualitative myocardial dysfunction patterns from ECG interpretations. This approach enabled us to address the challenge of obtaining expert labels in our large echocardiographic and ECG databases, which are larger than others in the literature<sup>22–24</sup>. Briefly, the NLP used customized regular expression scripts in Python v3.7, and provided an accuracy of 100% for extracting the WMA label when it was documented in the clinical interpretation of the echocardiogram. The accuracy of the NLP engine was analyzed by three expert reviewers in a random subpopulation of 100 studies selected from the test set and not used for engine development. As proof of concept, Fig. 2 illustrates NLP classification of echocardiography reports (left column), associated with raw ECGs (center), and echocardiograms (right) for three patients of the cohort. In Fig. 2a, a Black man in his 60s is shown with a history of cerebrovascular accident/stroke (CVA), heart failure (HF), and hypertension (HTN). An expert echocardiography report (ground truth) showed

lateral (white arrow) and anterior hypokinesis (Supplementary Movie 1). No Q waves were visible on ECG or stated in its interpretation. ECG-WMA-Net (described below) correctly labeled WMA. Figure 2b shows a Hispanic male in his forties with type-2 diabetes mellitus undergoing chemotherapy for acute myelogenous leukemia had Q waves (black arrows) without echocardiographic abnormalities. ECG-WMA-Net classified normal wall motion. Finally, Fig. 2c shows a White female in her sixties with anterior myocardial infarction (MI) had an echocardiogram read as akinesis of the mid-distal anterior wall, anteroseptum, and inferoapex. In this case, the ECG report read sinus rhythm with probable old anteroseptal infarct (black arrows), and ECG-WMA-Net classified abnormal wall motion.

#### ECG-based ECG-WMA-Net and traditional ECG analysis for WMA classification

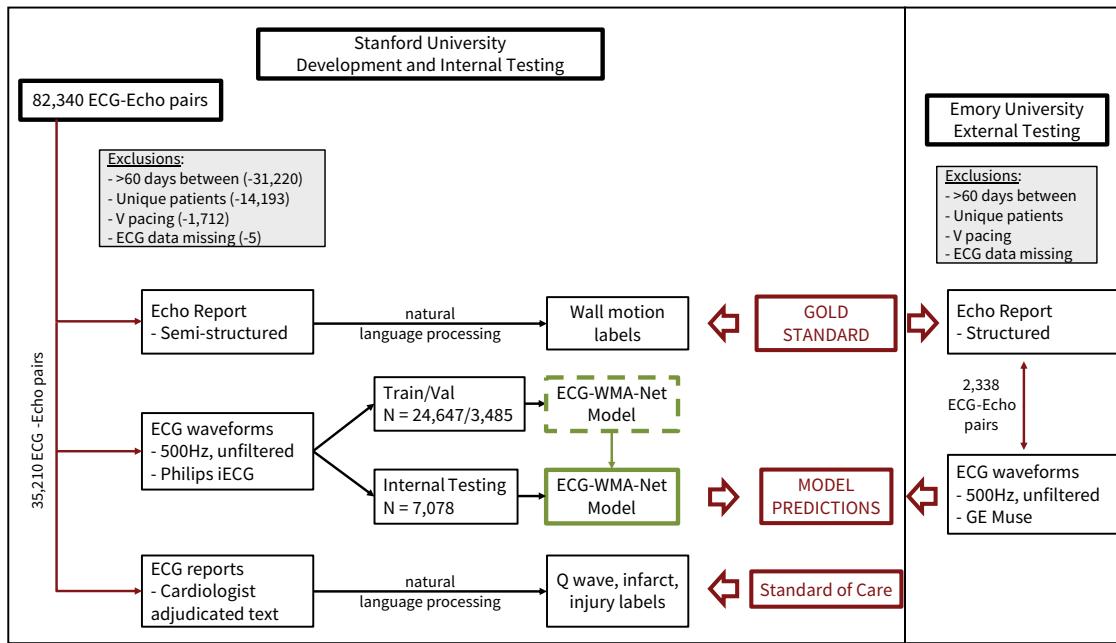
The electrocardiographic data for each patient in the test set was analyzed by three methods to classify the endpoint of echocardiographic WMA. We compared: (1) ECG-WMA-Net analysis of the 12-lead ECG for classification of the presence of WMA, (2) qualitative assessment of the 12-lead ECG by cardiologists interpreting the ECG during routine clinical care, and (3) an automated ECG index model based on logistic regression of quantitative Q-wave, T-wave, and ST-segment measurements.

ECG-WMA-Net was trained and tuned using the TensorFlow 2.0 machine learning library with the Keras API in Python. Input to the model was a matrix of the eight unique surface ECG waveforms recorded at 2500 500 Hz for 5 seconds (2500 samples). Iterative model architectures with permutations from the parameter sets were performed using keras-tuner (<https://github.com/keras-team/keras-tuner>). The architecture that provided the highest AUROC was stored and used for internal and external testing of the model and is shown in Fig. 3a. Overall, ECG-WMA-Net provided an AUROC of 0.781 (CI: 0.762–0.799) for WMAs. The Youden index was used to identify an optimal cut point, providing a sensitivity of 65.2% and specificity of 76.8%, negative predictive value of 94.9%, and positive predictive value of 25.0%. Separately trained models that included demographic variables did not improve these metrics ( $p = 0.558$ , see Supplementary Fig. 1). Model performance did not significantly differ across deciles of LVEF (Supplementary Fig. 2).

Conversely, the AUROCs for traditional qualitative and quantitative ECG analysis were 0.571 (CI: 0.552–0.590), and 0.681 (CI: 0.658–0.705), respectively were lower than the machine learning model ( $p < 0.0001$  in both cases). The three methods are summarized in Fig. 3b with details in Table 2. A net reclassification index analysis was performed, to compare individuals classified by ECG-WMA Net with the physician interpretation and the quantitative linear model, ECG-WMA Net had an NRI of 0.25 (95% CI: 0.20 to 0.29), with an NRI for events of 0.36 (95% CI: 0.32 to 0.40) and a NRI for non-events of -0.11 (95% CI: -0.13 to -0.10) compared with physician interpretation. Compared with the quantitative model, ECG-WMA Net had a NRI of 0.12 (95% CI: 0.08 to 0.17), (see Supplementary Table 2).

#### Generalizability in external test cohort with distinct demographic makeup

In a distinct external population from Emory Healthcare (Table 1), ECG-WMA-Net provided an AUC of 0.723 (CI: 0.685–0.757) (Fig. 3c), which was not significantly different from the California population ( $p = 0.069$ ). The accuracy was 0.710 (CI: 0.695–0.726), F1 score was 0.261 (0.230–0.292), sensitivity and specificity were 58 and 72%, respectively. The negative predictive value was 95% and the positive predictive value was 17%. These results demonstrate the ability of ECG-WMA-Net applied to raw ECGs to classify echocardiographic WMA in ethnically and physiologically distinct populations. To evaluate the possibility of the disparate impact of changes in racial makeup between cohorts on the accuracy of ECG-WMA-Net, we evaluated its performance in White and non-White patients. The ROC AUC was similar at 0.74 for White patients and 0.69 for non-White patients ( $p = 0.312$ ). For patients with low (<3 comorbidities) and high (>3 comorbidities) comorbidity burdens, the AUC was 0.71 and 0.69, respectively ( $p = 0.685$ ).



**Fig. 1 | Process for development and external testing of ECG-WMA-Net using ECG and echocardiogram pairs from Stanford and Emory University cohorts.** Consecutive ECG-echocardiogram pairs were curated from the clinical ECG and echocardiogram databases at Stanford (left) and Emory (right) Universities. Pairs were excluded if the two studies were >60 days apart, if the ECG reported ventricular pacing, or if ECG data were missing. Only the most recent valid pair was used such

that all patients in the database were unique. At Stanford, semi-structured echocardiogram reports and ECG reports were parsed by NLP for WMA labels and ECG labels, respectively. Raw ECG data were used with WMA labels to train ECG-WMA-Net for comparison against the standard of care. Consecutive ECG-echocardiogram pairs from Emory were then used to externally test the model.

### ECG regions used by ECG-WMA-Net to identify WMAs

We probed the trained ECG-WMA-Net to reveal ECG regions that contributed most to identifying the presence or absence of cardiac WMA, first, using *SHapley Additive exPlanations* (SHAP) values<sup>25</sup>. Figure 4a shows the regions that contributed to the output in the three patients from Fig. 2. Figure 4a also shows the summary for the most discriminatory ECG regions for all patients (aggregate). ECG regions arose throughout the QRS and T-waves and not just at early regions corresponding to early ventricular activation (Q waves).

Secondly, a data ablation experiment to create a stepwise analysis of limited windows within aligned ECG beats confirmed this analysis (Fig. 4b). By training models using only 120 ms windows, we found that the highest-performing window was between 80 and 200 ms (AUC 0.78 CI: 0.751–0.799) which encompassed the early QRS complex, with lower performances in the window of 40–160 ms (AUC 0.64 CI: 0.612–0.675). We found that ECG preprocessing, such as peak detection and transformation, did not improve model accuracy and may decrease generalizability.

### Discussion

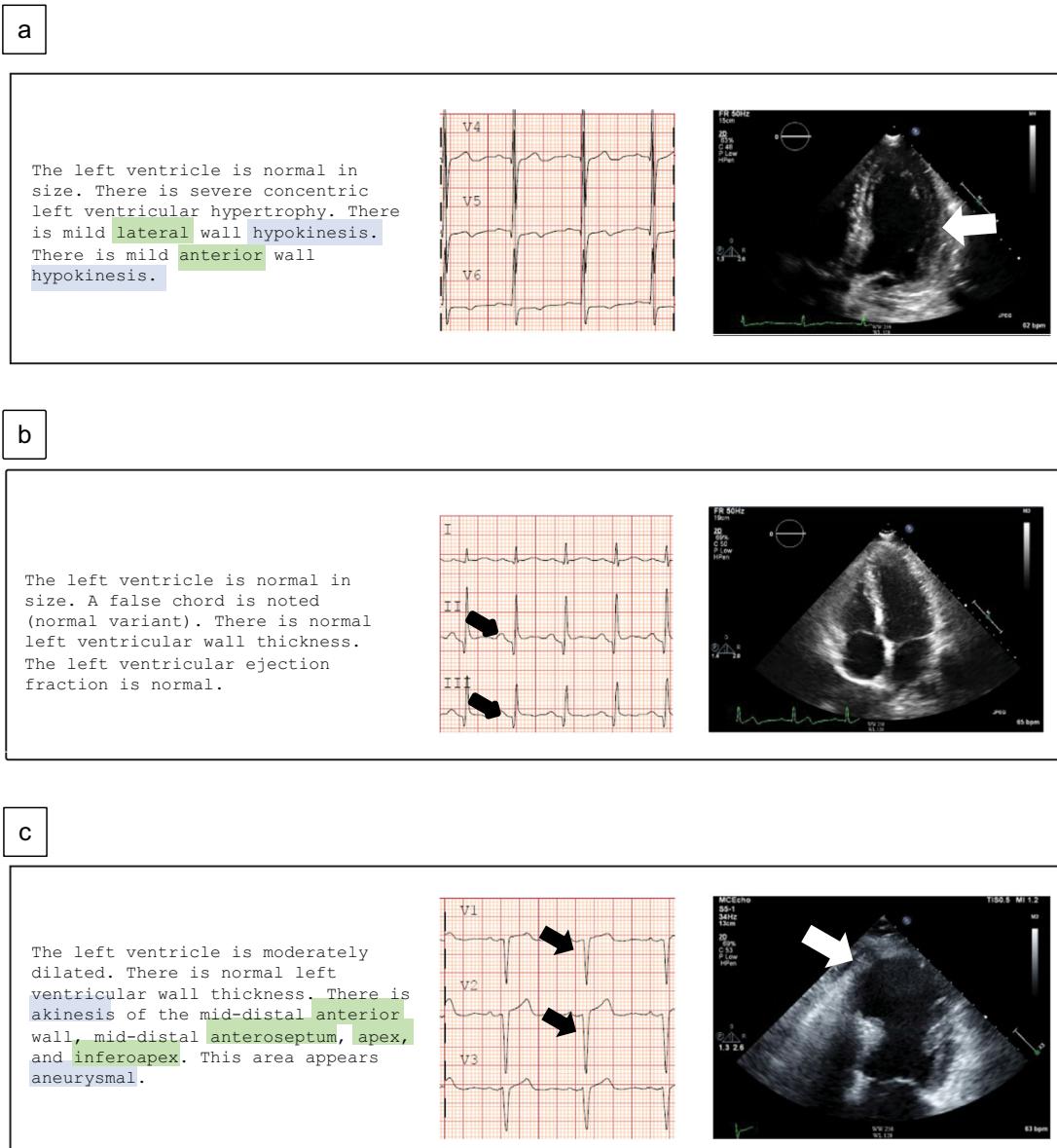
In this work, we present ECG-WMA-Net, a deep learning model that detects cardiac WMAs from the widely available surface ECG and outperforms the current clinical standard of clinician ECG interpretation and other published models focused on this task<sup>14</sup>. The performance of the model was externally validated on a demographically distinct population. This approach may enable broad screening for WMA in diverse populations using the affordable and ubiquitous ECG, which could reduce healthcare disparities and divert resources toward individuals who require additional evaluation, and reduce inconvenience to those incorrectly identified by current screening.

The development of a generalizable model was realized via one of the most extensive datasets in its domain. This dataset was curated by employing NLP techniques on the unstructured text of clinical records, facilitating the generation of labels from qualitative expert clinician

echocardiogram assessments without manual structured documentation. Therefore, it comprises a versatile approach that could be extended for model development across institutions and clinical contexts. Finally, using ablative and attention methods, we highlight ECG regions outside of the traditional ECG Q-wave and ST segment findings that identify WMAs from the surface ECG.

Detecting regional WMA is an important prognostic feature across populations, independent of reduced LVEF and the presence of active ischemia. While WMA may co-migrate with reduced LVEF in some patients<sup>26,27</sup>, compensatory hyperkinesia in others may normalize LVEF (Supplementary Fig. 3 and Supplementary Table 3)<sup>28</sup>. Accordingly, in patients with prior myocardial infarction, regional WMA may outperform low LVEF in predicting cardiovascular events<sup>29</sup>. The approach described in this study could enable more accessible screening for WMA across diverse populations. Using conventional approaches, the prevalence of WMA in the Framingham Heart Study Offspring Cohort was 6.5% in patients with congenital heart defects and heart failure and 4.2% in patients without these conditions<sup>30</sup>. In that study, clinical associations with WMAs were found with male gender and hypertension on multivariate analysis, and they are often seen in asymptomatic individuals. ECG-WMA-Net's use as a screening tool is supported by greatly improved sensitivity over the current practice. Implementing ECG-WMA-Net could identify more patients with WMAs and, despite requiring additional echocardiograms, the low number of studies required per true positive suggests that the benefits of early detection outweigh the resources required. Our simple approach to monitor WMA could be applied to multi-lead ECG devices for ambulatory assessment. This could also be applied to vulnerable populations including after acute coronary syndrome and revascularization. Continuous ECG during ambulatory exercise may allow assessment of ischemia-induced WMAs. Trials designed to assess the impact of artificial intelligence (AI)-enabled decision support tools are forthcoming<sup>31</sup>.

It is noteworthy to mention that ECG-based automatic diagnosis of WMA has been previously attempted in patients presenting acute



**Fig. 2 | Representative cases of ECG and echocardiographic studies with and without WMAs detected by ECG-WMA-Net. a** A black man in his sixties without Q waves on ECG but with mild anterior and lateral WMA (white arrow) (ECG-WMA-net true positive). **b** A Hispanic man in his forties with Q waves on ECG but without WMAs

on echocardiogram (ECG-WMA-Net true negative). **c** A female in her sixties with anterior and lateral Q waves (black arrows) on ECG and akinesis of the apex and mid-distal anterior and inferior walls (white arrow). Green text highlights show language tagged for location, and blue text highlights show language tagged for abnormalities.

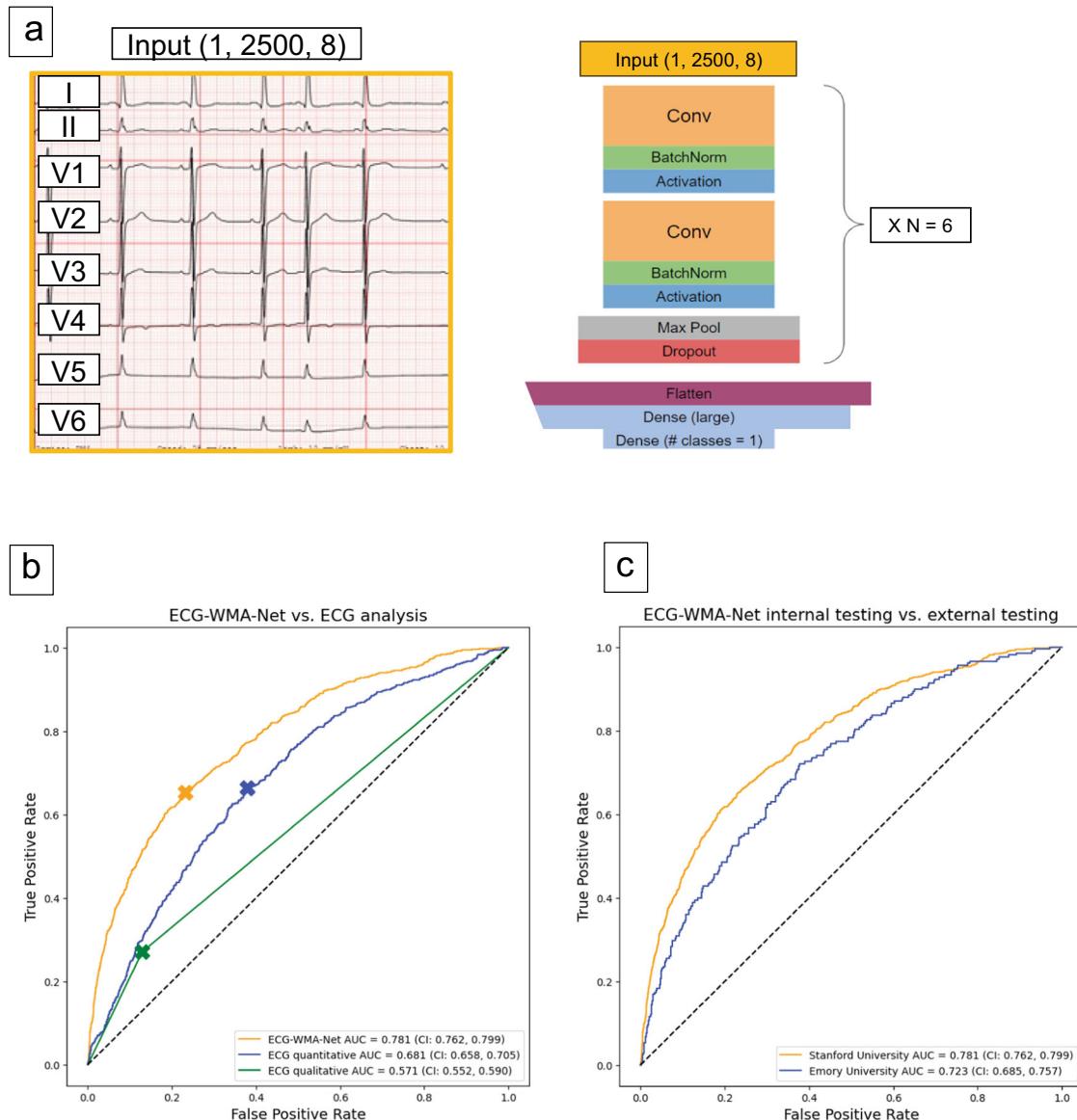
conditions<sup>14</sup>. The ECG possesses higher predictive power in the acute setting and WMA generally has a higher prevalence in these datasets. In contrast, our study is the first to automatically diagnose WMA from ECG data of a broad range of ambulatory and inpatient individuals and validate it in an external population.

In exploratory analysis, we applied ECG-WMA Net to forecast future WMA in patients who had normal function at a previous time point, which showed some ability to identify patients, but has declining performance with longer lead time, as expected (Supplementary Figure 4).

Our results may help reconcile pathophysiological inconsistencies in the literature. Although pathologic Q waves were classically described as sequelae of myocardial infarction scar, they are found in a vascular distribution in only 29% of patients with WMAs<sup>30</sup>. In cardiac magnetic resonance imaging studies to detect late gadolinium enhancement, pathologic Q-waves are actually rare in patients with unrecognized scar (6.7%)<sup>32</sup> although they are more common in selected patients with proven coronary disease<sup>10</sup>. The classical pathophysiological explanation of early loss of

subendocardial electrical forces and Q-waves is thus likely only a subset of patients with WMA. Instead, WMAs represent a collection of cardiomyopathic processes, including myocardial dysfunction, scar, pericardial adhesion, and stiffness. These changes in contractility may manifest across the duration of the electrocardiogram, including the T wave, where isovolumetric contraction stops, and ejection begins.

Several limitations apply to this study. Echocardiography reports were used in this study, rather than direct assessment of echocardiograms. While this provides expert-level adjudications, variations in practice between centers may affect generalizability. Nevertheless, this approach enables rapid deployment from full-text records, at the level of experts at each center. Electronic medical records and echocardiogram reports are readily available and portable. The use of the model as a diagnostic test is limited due to its modest positive predictive value in a broad population, although it outperforms other widely-used models and standard ECG interpretation<sup>33</sup>. In the external test set, approximately six patients identified by the model would need to undergo an echocardiogram to diagnose WMA in one



**Fig. 3 | Structure and performance of ECG-WMA-Net for detecting WMAs from the ECG.** **a** Structure of ECG-WMA-Net – Input to the model was the first 5 s of the eight unique leads from the ECG recording. At 500 Hz, this resulted in an input matrix of (1, 2500, 8). The parameter searching tool optimally found six sets of Conv, BatchNorm, Max Pool, and Dropout prior to Flatten, and Dense layers. **b** Receiver

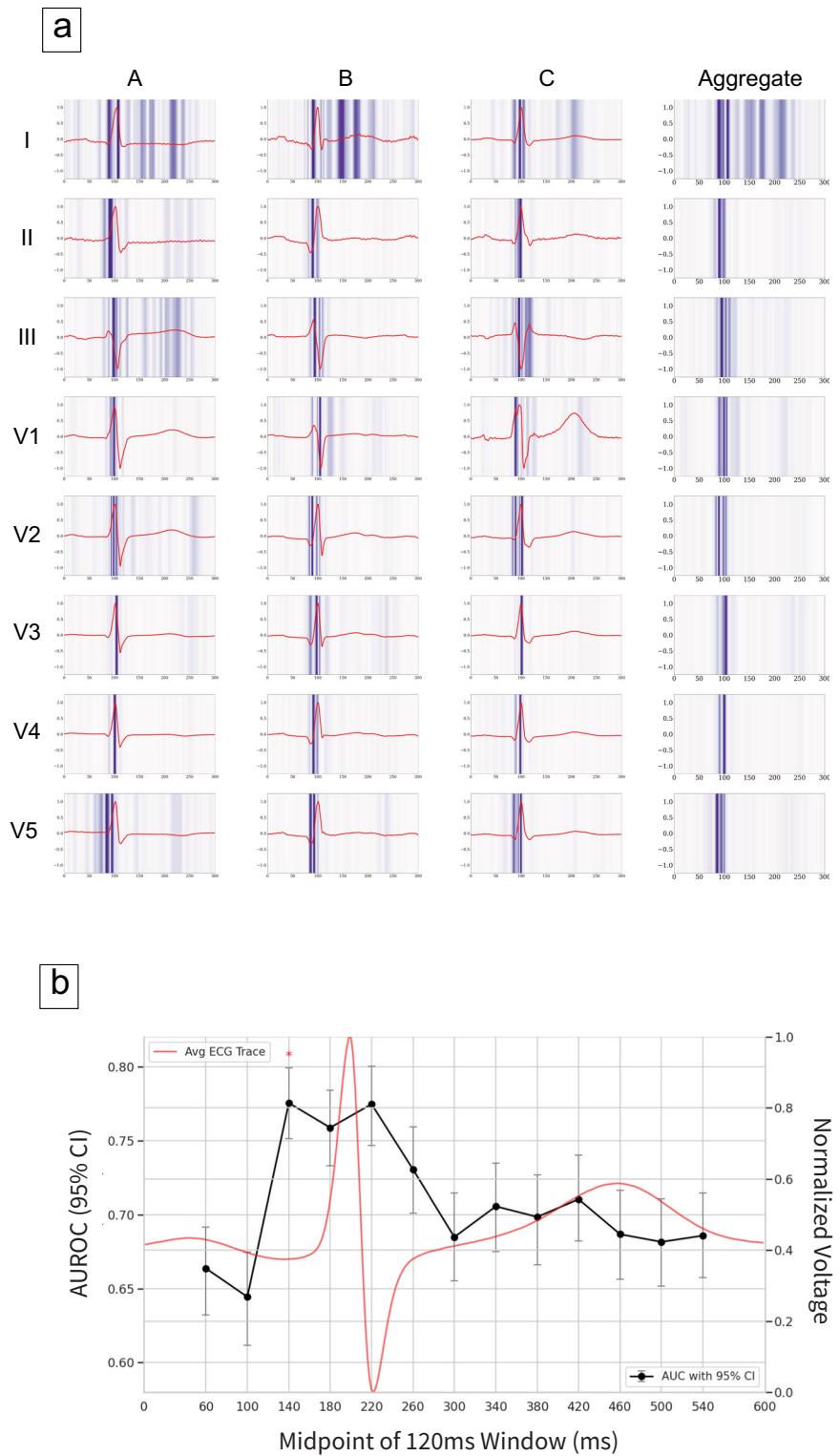
operating characteristics of ECG analysis by ECG-WMA-Net (yellow), quantitative ECG analysis (blue), and qualitative ECG analysis (green) for detection of WMA on echocardiography. The 'x' for each curve corresponds to the Youden Index optimal cut point. **c** Receiver operating characteristics of ECG analysis by ECG-WMA-Net in the internal test cohort (yellow) and external test cohort (blue).

**Table 2 | Model performance: ECG-WMA-Net versus qualitative and quantitative ECG metrics**

	ECG-WMA-Net	ECG qualitative	ECG quantitative
AUC	0.781 (0.762, 0.799)	0.571 (0.552, 0.590)	0.681 (0.658, 0.705)
Accuracy	0.756 (0.747, 0.764)	0.808 (0.799, 0.817)	0.622 (0.611, 0.632)
F1 Score	0.362 (0.343, 0.381)	0.228 (0.203, 0.253)	0.269 (0.251, 0.287)
Sensitivity	0.652 (0.624, 0.680)	0.271 (0.240, 0.301)	0.665 (0.632, 0.698)
Specificity	0.768 (0.759, 0.777)	0.871 (0.863, 0.879)	0.617 (0.605, 0.628)
NPV	0.949 (0.944, 0.954)	0.911 (0.904, 0.918)	0.940 (0.933, 0.947)
PPV	0.250 (0.235, 0.267)	0.197 (0.174, 0.221)	0.168 (0.155, 0.182)

AUC area under the curve, ECG electrocardiogram, NPV negative predictive value, PPV positive predictive value, WMA wall motion abnormality

**Fig. 4 | Temporal localization of ECG features associated with WMAs identified by ECG-WMA-Net.** **a** Times of cardiac activation on the ECG that identify abnormal wall motion. Columns A, B, and C are the output of SHAP analysis of trained machine learning models to each of the patient cases from Fig. 2, in order. The cropped and aligned ECG signal from each patient is shown in red trace. The column labeled aggregate displays the summated ECG SHAP values output map for all patients. Both the example cases and the aggregate shadings support that ECG features throughout the QRS and T-waves were used to identify WMA. **b** A sliding window of 120 ms was used to build successive models for each 40 ms from 60 to 540 ms, while ablating the remainder of waveform data. The black line shows the ROC AUC values with confidence intervals for each model trained on windows centered at each time point. The red tracing indicates a standardized average ECG with normalized voltage.



patient. Given the importance of this diagnosis and the availability of effective treatments this may be a tenable screening paradigm, however, further work is needed. Studies for implementation of the model will have to be carefully designed given the heterogeneous treatment strategies and success rates for different etiologies of WMA and may benefit from pilot studies in certain high-risk subgroups. The model performance may be further improved by training on federated data across institutions, enabled by the NLP labeling engine. AI models may be technically less effective when there is an imbalance between groups, such as prior studies of hypertrophic

cardiomyopathy or hypoglycemia<sup>34–36</sup>, yet our models retained performance in a distinct population. It would be important to identify additional test populations for further validation.

In summary, a novel deep learning classifier applied to the ECG provided improved performance for echocardiographic WMA over standard of care ECG interpretation. This could be used for screening large populations, even on reduced lead sets which are amenable to portable wearable devices. Finally, linking non-q-wave regions of the QRS to abnormal cardiac structure alters our physiological understanding of ECG genesis.

## Methods

### Development cohort

We enrolled data from patients who received an ECG at Stanford University between 05/01/2014 and 08/17/2018 who underwent echocardiography with final reports generated by readers with level III American Society of Echocardiography training within 60 days. Figure 1 shows data flow in the study. The search resulted in 82,340 echo-ECG pairs. We excluded 31,220 pairs when the finalized ECG was not within 60 days of the index echocardiogram and 1712 where ventricular pacing was involved. Ventricular pacing was excluded since it obscures native conduction patterns. Five echo-ECG pairs were excluded due to corrupt ECG files. Finally, we included only the most recent valid echo-ECG pair for each patient, thus excluding 14,193 multiple examinations. This resulted in 35,210 unique patient cases. An external test dataset was developed in a similar fashion at Emory Healthcare. Consecutive ECGs from 5/15/2018 to 6/1/2020 were collected with the same criteria as above. 23,658 cases (with paired ECG and echo) met the inclusion criteria, of which ~10% were randomly selected ( $N = 2338$ ), and 90% were reserved to avoid contamination of the dataset for future model development at Emory University. Table 1 includes the demographics of the two study populations. Demographics are further stratified between the training, validation, and testing partitions (Supplementary Table 1). Stanford University and Emory University Institutional Review Board approvals were obtained for this study (IRB-41045 and STUDY00000160, respectively). Waiver of consent forms were approved by both IRBs due to the retrospective nature of the study.

### Natural language processing for wall motion assessment

Wall motion was assessed by experts in echocardiography with Level III American Society of Echocardiography (ASE) training and coded for all heart walls in six segments using planimetry, transverse and longitudinal strain. Abnormal wall motion was noted as present for reduced wall thickening (hypokinesis), absence of thickening (akinesis), or a thinned wall with paradoxical motion (aneurysm). The text was analyzed using custom regular expression algorithms developed in Python 3.7 using the *re* library. First, the section of the echocardiogram report relating to the left ventricle was extracted. Then, statements were parsed according to terms related to wall motion (normal, hypokinesis, akinesis, and aneurysm) in six clinical regions of the heart (apical, septal, inferior, posterior, lateral, and anterior)<sup>37</sup>. Negative space terms were applied to appropriately assign the sentiment of negating statements. The regular expression terms are reported in Supplementary Table 4 and shared publicly via GitHub<sup>38</sup>. The accuracy of this approach to identify labels from the clinical reports was verified by three reviewers in a random subset of 100 echocardiography reports from the test set and found to have an accuracy of 100%.

### Electrocardiography acquisition and labeling

ECGs for development and internal testing were recorded at Stanford University and stored unfiltered in XML format in a clinical ECG system (iECG, Philips, Amsterdam, NL). A total of eight ECG waveforms (leads I, II, V1-6) were stored for each case. Leads aVR, aVL, aVF, and III can be calculated from leads I and II, so these were not stored to avoid redundancy. Each ECG was reviewed by a cardiologist, and a semi-structured interpretation was stored as plain text during standard clinical care. Quantitative measurements of Q waves, T waves and ST segments in each lead were derived from the clinical ECG software and stored as metadata with each ECG.

For model training, 8-lead ECG waveforms were used. All ECG pre-processing was algorithmic with no manual steps required. ECG voltage time series were extracted from XML files in raw format. They were trimmed to 5 s duration (2500 samples), and baseline wander was removed with the *signal.detrend* method from the SciPy Python library<sup>39</sup>. The eight signals were then scaled by the standard deviation on a per-ECG basis.

For this study, we applied NLP regular expression analysis to “qualitatively” identify the presence/absence of abnormalities suggesting myocardial dysfunction from ECG codes. Statements across all the ECG physician interpretation statements were tokenized and counted by frequency. Analysis of the tokenized statements by expert reviewers (AJR, SN,

and NKB) resulted in keywords (and keyword variants) indicating possible myocardial dysfunction. These included “infarct”, “injury”, “ST elev”, and “Q wave” with segmental notation in one of the distributions “posterior”, “inferior”, “lateral”, “septal”, or “anterior”<sup>40</sup>. The term “ischemia” was not used due to its lack of regional distinction unless another term was present. ECG interpretation statements regarding the apex, including “apex” or “apic” were analyzed individually ( $N < 100$ ). Negation was included to ensure accurate sentiment handling for tokens. The total unique statement phrases were 17,990. Using this NLP method, we parsed the EHR for statements suggesting myocardial dysfunction in the unstructured final ECG record. The presence/absence of these terms (Boolean) is referred to as the Qualitative ECG assessment of WMA.

Separately, quantitative measurements made by the ECG algorithm (iECG, Philips Healthcare, Amsterdam, NL) pertaining to myocardial abnormalities were extracted from ECG metadata, including the Q wave duration (in milliseconds; ms), Q wave amplitude (in millivolts; mV), T wave amplitude (mV), and the mid-point of ST segment (mV). A logistic regression model was trained using these features for the eight leads and tested on the holdout test set of the development cohort. This model is referred to as the quantitative ECG assessment of WMA.

### Machine learning model development

The ECG data from each patient were analyzed in three ways, as follows: (a) deep learning-based classification of 12-lead ECGs (ECG-WMA-Net), (b) qualitative evaluation of 12-lead ECG signals by trained cardiologists, and (c) an ECG index model developed using logistic regression on an assortment of lead-wise features (Q wave, ST segment, and T wave features). ECG-WMA-Net used the TensorFlow 2.0 machine learning system with the Keras API. Model architectures were iterated using keras-tuner<sup>41</sup> over 100 trials to maximize the AUC metric by training with a range of parameter sets, including filters, convolution widths, strides, learning rates, and decay. The best-performing model consisted of six stacks of the following network: two consecutive blocks of 2D convolution, batch normalization and exponential linear unit activation followed by one block of max pooling and dropout. This network was augmented with a flattening layer, and two dense layers with a hidden dimension of 500. The network was optimized using the Adam optimizer over a weighted binary cross-entropy loss, which maximized the average between AUC, Precision, and Recall. The training was performed on the development cohort using a batch size of 8 and over 75 epochs. The optimal learning rate was learned using hyper-parameter tuning. The overall network architecture is shown in Fig. 3b and made available publicly via GitHub<sup>38</sup>. The loss function was binary cross entropy. The parameter set yielding the optimum model was saved for full model evaluation and testing.

Models used ECG data as inputs and were trained to the presence or absence of WMA ascertained from NLP of echocardiography reports. Patients were randomly split into separate cohorts of 70% for training, 10% for validation and 20% for testing so that each patient appeared in only one of these groups. Supplementary Table 1 shows the characteristics of each split. There were no differences between the groups. The Youden index was used to assign the cut point for classification of the continuous output probabilities from the model.

We assessed the impact of demographic variables on model accuracy using a separate multimodal model in which age, gender and ethnicity and race were input along with ECG voltage time series, which did not improve performance (Supplementary Fig. 1).

### External test cohort

The collection of ECG and echocardiogram pairs was performed at Emory University using the same criteria described above (studies within 60 days, excluding ventricular pacing on ECG). The demographics of this cohort, showing a different ethnic makeup are shown in Table 1. ECG-WMA-Net and the optimal threshold cut point was shared with Emory University via GitHub. The classifier was then applied, and results were compiled to assess model performance.

## Identifying ECG regions critical to the identification of WMA

To identify regions of the ECG input used by ECG-WMA-Net to detect WMA, referenced to specific time points within the ECG, we performed a post hoc analysis in which we aligned single ECG beats by their R-wave. Two analyses were performed to assess important segments of the aligned ECG signal: one attention method and one ablation method. R wave peaks were first detected using a wavelet-based algorithm for QRS complex detection<sup>42</sup>. Each ECG beat window was defined as 600 ms (300 samples) total, with 200 ms prior to the R wave peak and 400 ms after the peak were extracted for input to the model. The resultant input shape is a matrix of size  $300 \times 8$ .

To assess attention of the neural network to specific regions of the ECG, We used SHAP via the python library<sup>25</sup>. SHAP values link regions of the trained model most critical to model classifications to values from game theory to identify inputs that most influence classifications. We utilized a python library with built-in neural network methods, called DeepExplainer<sup>43</sup>. DeepExplainer was run on our raw eight-lead ECG inputs to identify hot-spots.

To further dissect ECG time points identified by the SHAP values to classify WMA, we performed an ablation experiment using windows of the aligned ECG waveforms. Classifiers with a similar structure to ECG-WMA-Net were trained using successive 120 ms windows of the aligned ECG input signal. This process was repeated for each sliding window across the entire tracing. Other waveform data not included in each window was ablated. The AUC and confidence intervals for each window-based classifier were recorded.

## Retrospective analysis for prediction of future WMAs

We conducted an additional analysis within our internal testing cohort ( $N = 7078$ ). This analysis aimed to determine the model's ability to predict the presence of wall motion abnormalities (WMA) on future echocardiograms based on prior ECGs. We included patients with available ECG and echocardiogram pairs recorded up to 1 year prior to the echocardiogram included in the overall test set. We excluded those who already had abnormal wall motion at the time of the prior ECG. The analysis was stratified by the time interval between the ECG and the echocardiogram. The results of the retrospective analysis are given in Supplementary Fig. 4.

## Statistical methods

For clinical demographics, continuous variables between development and external datasets were compared by *t*-tests and categorical variables were compared with a Chi-squared test. Normality was assessed using the Shapiro-Wilk test. Non-normal distributions were compared using the Mann-Whitney *U*-test. Continuous variables are reported as mean  $\pm$  SD or median  $\pm$  IQR for non-normal distributions, and categorical variables are reported as percentages. Comparisons between partitions of the development cohort were performed using ANOVA. Confidence intervals on ECG-WMA-Net test characteristics were obtained via bootstrapping with 10,000 trials and an alpha of 0.05. To compare ROC AUCs in paired sets of data, DeLong's test was performed. To compare ROC AUCs in unpaired data, ROC AUCs and standard errors were bootstrapped and a two-tailed *z* test was calculated to assess differences. The mean total, event, and non-event net reclassification indices were calculated with 95% confidence intervals by bootstrapping with  $N = 1000$  iterations.

## Data availability

The raw patient data are not publicly available due to institutional policy and human subjects' approval to protect patient privacy.

## Code availability

The source code for this project, including ECG-WMA-Net model architecture and language processing code is available on GitHub at <https://github.com/sanarayan-code/ECG-WMA-Net.git><sup>38</sup>

Received: 22 January 2024; Accepted: 22 December 2024;

Published online: 11 January 2025

## References

- Tracy, C. M. et al. Determinants of ventricular arrhythmias in mildly symptomatic patients with coronary artery disease and influence of inducible left ventricular dysfunction on arrhythmia frequency. *J. Am. Coll. Cardiol.* **9**, 483–488 (1987).
- Trappe, H. J., Lichtlen, P. R., Klein, H., Wenzlaff, P. & Hartwig, C. A. Natural history of single vessel disease. Risk of sudden coronary death in relation to coronary anatomy and arrhythmia profile. *Eur. Heart J.* **10**, 514–524 (1989).
- Nath, S. et al. Use of a regional wall motion score to enhance risk stratification of patients receiving an implantable cardioverter-defibrillator. *J. Am. Coll. Cardiol.* **22**, 1093–1099 (1993).
- Mahenthiran, J. et al. Prognostic importance of wall motion abnormalities in patients with ischemic cardiomyopathy and an implantable cardioverter-defibrillator. *Am. J. Cardiol.* **98**, 1301–1306 (2006).
- Gaitonde, R. S. et al. Segmental wall-motion abnormalities of the left ventricle predict arrhythmic events in patients with nonischemic cardiomyopathy. *Heart Rhythm* **7**, 1390–1395 (2010).
- Cicala, S. et al. Prevalence and prognostic significance of wall-motion abnormalities in adults without clinically recognized cardiovascular disease: the Strong Heart Study. *Circulation* **116**, 143–150 (2007).
- Schlant, R. C. et al. Guidelines for electrocardiography. a report of the American college of cardiology/American heart association task force on assessment of diagnostic and therapeutic cardiovascular procedures (committee on electrocardiography). *Circulation* **85**, 1221–1228 (1992).
- Ommen, S. R. et al. 2020 AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy: A report of the American college of cardiology/American heart association joint committee on clinical practice guidelines. *Circulation* **142**, e558–e631 (2020).
- Birnie, D. H. et al. HRS expert consensus statement on the diagnosis and management of arrhythmias associated with cardiac sarcoidosis. *Heart Rhythm* **11**, 1305–1323 (2014).
- Nadour, W. et al. Does the presence of Q waves on the EKG accurately predict prior myocardial infarction when compared to cardiac magnetic resonance using late gadolinium enhancement? A cross-population study of noninfarct vs infarct patients. *Heart Rhythm* **11**, 2018–2026 (2014).
- Hurd, H. P. 2nd, Starling, M. R., Crawford, M. H., Dlabal, P. W. & O'Rourke, R. A. Comparative accuracy of electrocardiographic and vectorcardiographic criteria for inferior myocardial infarction. *Circulation* **63**, 1025–1029 (1981).
- Abedin, Z. et al. Predictive value of Q waves in inferior leads for the diagnosis of old inferior wall myocardial infarction. *Clin. Investig.* **08**, 155–160 (2018).
- Ishikawa, J., Yamanaka, Y., Watanabe, S., Toba, A. & Harada, K. Cornell product in an electrocardiogram is related to reduced LV regional wall motion. *Hypertens. Res.* **42**, 541–548 (2019).
- Sugimoto, K. et al. Electrocardiographic scoring helps predict left ventricular wall motion abnormality commonly observed after subarachnoid hemorrhage. *J. Stroke Cerebrovasc. Dis.* <https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.07.008> (2018).
- Cheitlin, M. D. et al. ACC/AHA guidelines for the clinical application of echocardiography. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Clinical Application of Echocardiography). Developed in collaboration with the American Society of Echocardiography. *Circulation* **95**, 1686–1744 (1997).
- Malhotra, C. & Do, Y. K. Socio-economic disparities in health system responsiveness in India. *Health Policy Plan* **28**, 197–205 (2013).
- Attia, Z. I. et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* **394**, 861–867 (2019).

18. Tison, G. H., Zhang, J., Delling, F. N. & Deo, R. C. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. *Circ. Cardiovasc. Qual. Outcomes* **12**, e005289 (2019).
19. Sparapani, R. et al. Detection of left ventricular hypertrophy using Bayesian additive regression trees: the MESA. *J. Am. Heart Assoc.* **8**, e009959 (2019).
20. Hughes, J. W. et al. A deep learning-based electrocardiogram risk score for long term cardiovascular death and disease. *NPJ Digit. Med.* **6**, 169 (2023).
21. Torres-Soto, J. & Ashley, E. A. Multi-task deep learning for cardiac rhythm detection in wearable devices. *NPJ Digit. Med.* **3**, 116 (2020).
22. Kim, Y.-G. et al. ECG-ViEW II, a freely accessible electrocardiogram database. *PLoS ONE* **12**, e0176222 (2017).
23. Zheng, J., Guo, H. & Chu, H. A large scale 12-lead electrocardiogram database for arrhythmia study. *PhysioNet* <https://doi.org/10.13026/WGEX-ER52> (2022).
24. Gow, B. et al. MIMIC-IV-ECHO: echocardiogram matched subset. *PhysioNet* <https://doi.org/10.13026/EF48-V217> (2023).
25. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Neural Inf Process Syst.* 4765–4774 (2017).
26. St John Sutton, M. et al. Quantitative two-dimensional echocardiographic measurements are major predictors of adverse cardiovascular events after acute myocardial infarction. The protective effects of captopril. *Circulation* **89**, 68–75 (1994).
27. Picciotto, S. et al. Determinants of plasma interleukin-6 levels among survivors of myocardial infarction. *Eur. J. Cardiovasc. Prev. Rehabil.* **15**, 631–638 (2008).
28. Kjøller, E., Køber, L., Jørgensen, S. & Torp-Pedersen, C. Long-term prognostic importance of hyperkinesia following acute myocardial infarction. TRACE Study Group. TRAndolapril Cardiac Evaluation. *Am. J. Cardiol.* **83**, 655–659 (1999).
29. Jurado-Román, A. et al. Superiority of wall motion score index over left ventricle ejection fraction in predicting cardiovascular events after an acute myocardial infarction. *Eur. Heart J. Acute Cardiovasc Care* **8**, 78–85 (2019).
30. Tsao, C. W. et al. Subclinical and clinical correlates of left ventricular wall motion abnormalities in the community. *Am. J. Cardiol.* **107**, 949–955 (2011).
31. Yao, X. et al. ECG AI-guided screening for low ejection fraction (EAGLE): rationale and design of a pragmatic cluster randomized trial. *Am. Heart J.* **219**, 31–36 (2020).
32. Ebeling Barbier, C., Bjerner, T., Johansson, L., Lind, L. & Ahlström, H. Myocardial scars more frequent than expected. *J. Am. Coll. Cardiol.* **48**, 765–771 (2006).
33. Brindle, P. M. et al. The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study. *Br. J. Gen. Pr.* **55**, 838–845 (2005).
34. Porumb, M., Stranges, S., Pescapè, A. & Pecchia, L. Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG. *Sci. Rep.* **10**, 170 (2020).
35. Ko, W.-Y. et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J. Am. Coll. Cardiol.* **75**, 722–733 (2020).
36. Lyon, A. et al. Distinct ECG phenotypes identified in hypertrophic cardiomyopathy using machine learning associate with arrhythmic risk markers. *Front. Physiol.* **9**, 213 (2018).
37. IAC. Echocardiography accreditation standards & guidelines. [https://www.intersocietal.org/echo/seeking/echo\\_standards.htm](https://www.intersocietal.org/echo/seeking/echo_standards.htm) (2021).
38. Rogers, A. J. et al. ECG-WMA-Net. *GitHub* <https://github.com/sanarayan-code/ECG-WMA-Net.git> (2023).
39. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
40. Wagner, G. S. et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram. *Circulation* **119** (2009).
41. O’Malley, T. et al. *KerasTuner*. GitHub. (2019).
42. Lin, C.-C., Chang, H.-Y., Huang, Y.-H. & Yeh, C.-Y. A novel wavelet-based algorithm for detection of QRS complex. *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* **9**, 2142 (2019).
43. Shrikumar, A., Greenside, P. & Kundaje, A. Learning important features through propagating activation differences. *Proc. Mach. Learn. Res.* **70**, 3145–3153 (2017).

## Acknowledgements

A.J.R. is supported by the National Institutes of Health (K23 HL166977) and the American Heart Association Career Development Award (23CDA933663). S.B. is supported by the American Heart Association (25POST1361932). J.X. is supported by the National Science Foundation CAREER 1942926 and grants from the Chan-Zuckerberg Initiative. G.D.C. and R.S. are supported by the National Institutes of Health (R01 EB030362). J.W.H. is supported by the National Science Foundation (DGE-1656518). M.V.P. is supported by funding from the National Institutes of Health and Apple Inc. S.M.N. reports grant funding from the National Institutes of Health (R01 HL83359, R01 HL149134, R01 HL1662260, and T32 HL166155).

## Author contributions

A.J.R., M.V.P., and S.M.N. conceived of the study. A.J.R., S.B., J.T., J.T.S., J.S.T., J.W.H., E.A.A., M.V.P., and S.M.N. were involved in acquiring the Stanford data in a format appropriate for this study and preprocessing. N.K.B., R.S., and G.D.C. performed data acquisition, preparation, and analysis at Emory University Medical Center. A.J.R., S.B., R.A., V.T., J.X., M.I.A., P.C., J.W.H., E.A.A., M.V.P., M.Z., and S.M.N. performed data analysis at Stanford. A.J.R., S.B., R.A., V.T., J.X., M.I.A., M.Z., and S.M.N. performed model development and optimization. The work was led by A.J.R. and overseen by E.A.A., M.V.P., and S.M.N. All authors provided comments and approved the manuscript.

## Competing interests

S.N. reports consulting compensation from Abbott Inc., Up to Date, and LifeSignals.ai, and intellectual property rights from the University of California Regents and Stanford University. E.A. reports consulting fees from Apple Inc. M.V.P. reports consulting fees from Apple Inc., Boston Scientific, Biotronik Inc., Bristol Myers Squibb, QALY, Johnson & Johnson, and has an equity interest in QALY. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41746-024-01407-y>.

**Correspondence** and requests for materials should be addressed to Albert J. Rogers.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025