JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
VOL. 84, NO. 9, 2024

PUBLISHED BY ELSEVIER

Deep Learning-Based Electrocardiogram Analysis Predicts Biventricular Dysfunction and Dilation in Congenital Heart Disease

Joshua Mayourian, MD, PhD,[a,*] Addison Gearhart, MD,[a,*] William G. La Cava, PhD,[b] Akhil Vaid, MD,[c] Girish N. Nadkarni, MD,[c] John K. Triedman, MD,[a] Andrew J. Powell, MD,[a] Rachel M. Wald, MD,[d] Anne Marie Valente, MD,[a] Tal Geva, MD,[a] Son Q. Duong, MD,[c,e] Sunil J. Ghelani, MD[a]

ABSTRACT

BACKGROUND Artificial intelligence–enhanced electrocardiogram (AI-ECG) analysis shows promise to detect biventricular pathophysiology. However, AI-ECG analysis remains underexplored in congenital heart disease (CHD).

OBJECTIVES The purpose of this study was to develop and externally validate an AI-ECG model to predict cardiovascular magnetic resonance (CMR)-defined biventricular dysfunction/dilation in patients with CHD.

METHODS We trained (80%) and tested (20%) a convolutional neural network on paired ECG-CMRs (≤30 days apart) from patients with and without CHD to detect left ventricular (LV) dysfunction (ejection fraction ≤40%), RV dysfunction (ejection fraction ≤35%), and LV and RV dilation (end-diastolic volume z-score ≥4). Performance was assessed during internal testing and external validation on an outside health care system using area under receiver-operating curve (AUROC) and area under precision recall curve.

RESULTS The internal and external cohorts comprised 8,584 ECG-CMR pairs (n = 4,941; median CMR age 20.7 years) and 909 ECG-CMR pairs (n = 746; median CMR age 25.4 years), respectively. Model performance was similar for internal testing (AUROC: LV dysfunction 0.87; LV dilation 0.86; RV dysfunction 0.88; RV dilation 0.81) and external validation (AUROC: LV dysfunction 0.89; LV dilation 0.83; RV dysfunction 0.82; RV dilation 0.80). Model performance was lowest in functionally single ventricle patients. Tetralogy of Fallot patients predicted to be at high risk of ventricular dysfunction had lower survival ($P < 0.001$). Model explainability via saliency mapping revealed that lateral precordial leads influence all outcome predictions, with high-risk features including QRS widening and T-wave inversions for RV dysfunction/dilation.

CONCLUSIONS AI-ECG shows promise to predict biventricular dysfunction/dilation, which may help inform CMR timing in CHD. (J Am Coll Cardiol 2024;84:815-828) © 2024 by the American College of Cardiology Foundation.

Listen to this manuscript's audio summary by Editor Emeritus Dr Valentin Fuster on www.jacc.org/journal/jacc.

From the [a]Department of Cardiology, Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA; [b]Computational Health Informatics Program, Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA; [c]The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [d]Division of Cardiology, University of Toronto, Peter Munk Cardiac Centre, Toronto, Ontario, Canada; and the [e]Division of Pediatric Cardiology, Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York, USA. *Drs Mayourian and Gearhart contributed equally to this work.

Ambarish Pandey, MD, served as Guest Associate Editor for this paper. Javed Butler, MD, MPH, MBA, served as Guest Editor-in-Chief for this paper.

Review and acceptance occurred under Dr Valentin Fuster's term as Editor-in-Chief.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received February 16, 2024; revised manuscript received April 23, 2024, accepted May 20, 2024.

ISSN 0735-1097/$36.00
https://doi.org/10.1016/j.jacc.2024.05.062

ABBREVIATIONS AND ACRONYMS

AI-ECG = artificial intelligence-enhanced electrocardiogram

AUPRC = area under the precision-recall curve

AUROC = area under the receiver operating curve

CHD = congenital heart disease

CMR = cardiovascular magnetic resonance

ECG = electrocardiogram

EDV = end-diastolic volume

EF = ejection fraction

LV = left ventricle

RV = right ventricle

ToF = Tetralogy of Fallot

Patients with congenital heart disease (CHD) are a heterogeneous population with complex anatomy and physiology. In the current era, the majority survive to adulthood.[1] Noninvasive imaging plays a central role in risk stratifying this growing population and informing medical/surgical interventions, with biventricular size and function among the strongest predictors of long-term mortality in a multitude of CHD lesions.[2-4] For example, in adults with tetralogy of Fallot (ToF), left ventricular (LV) and right ventricular (RV) dysfunction are both predictive of death and sustained ventricular tachycardia.[5]

SEE PAGE 829

Cardiac magnetic resonance (CMR) imaging has an established role in the lifelong management of patients with CHD to accurately assess LV, RV, and functional single ventricle size and ejection fraction (EF), which are particularly more challenging to measure by echocardiography. However, CMR has practical limitations prohibiting its widespread use including being time-, resource-, and cost-intensive, making it of interest to develop a cheap and effective tool to help inform timing of CMR.

Electrocardiograms (ECGs) are a quick, ubiquitous, and cost-effective tool used for cardiac screening of adults and children. Artificial intelligence-enhanced electrocardiogram (AI-ECG) algorithms reliably predict a range of cardiovascular phenotypes in the general adult population, including biventricular dilation and dysfunction.[6-9] However, there remains a paucity of AI-ECG applications for the CHD population with distinct ECG characteristics attributed to age, CHD lesion, and prior interventions.[10,11] To our knowledge, AI-ECG technology has yet to be applied to predict the gold-standard CMR measurements that congenital cardiologists rely on to risk stratify and inform management of CHD patients.

In this study, our primary objective was to address this gap by developing, internally testing, and externally validating an AI-ECG model to predict LV and RV dysfunction and dilation in patients with and without CHD (**Central Illustration**).

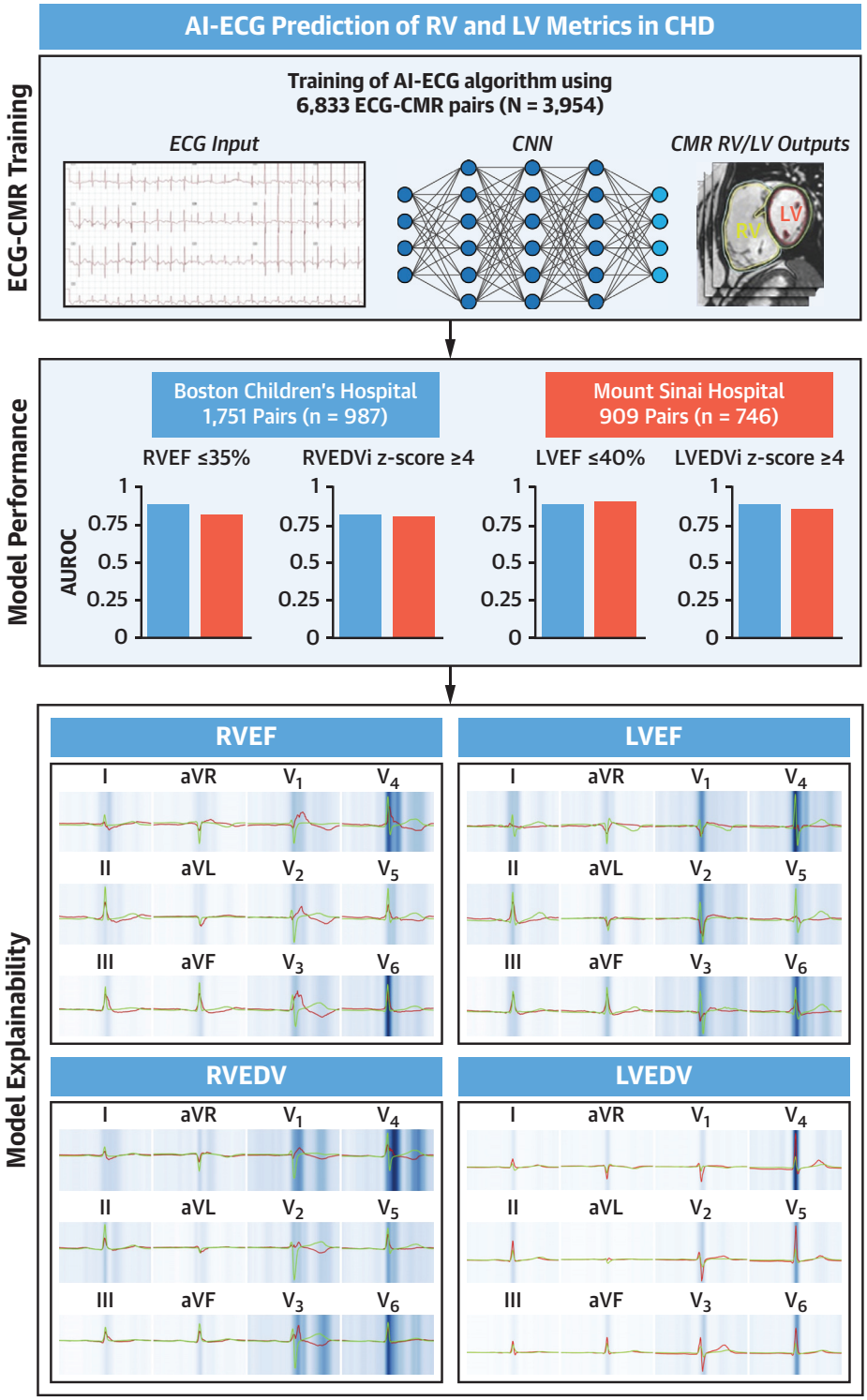
METHODS

INTERNAL STUDY POPULATION AND PATIENT ASSIGNMENT. We utilized patient data from Boston Children's Hospital between 2002 and 2021. All CMR studies with RV and LV EF percentage and end-diastolic volume (EDV) z-scores were considered eligible. LVEDV and RVEDV z-scores were calculated using published equations by Alfakih et al[12]; patients with body surface area <1.0 m^2 were excluded given the paucity of normative CMR data for this group, as well as the phenomenon of heteroscedasticity preventing extrapolation. Each qualifying CMR event was paired with an ECG; only ECG-CMR pairs ≤30 days apart without an intermediate catheterization or surgery were included. In cases of multiple ECGs within this timeframe, only the ECG closest in time to the CMR was included. ECG-CMR pairs with ECGs failing to pass quality control (see the Supplemental Methods for details) were removed. The remaining ECG-CMR pairs were included as the main internal cohort (**Figure 1**).

To assign patients into CHD subgroups, we utilized our institutional Fyler coding system.[13] The coding system allows for identification of patients with specific structural diagnoses (eg, coarctation of the aorta, ventricular septal defect, and so on) as well as clinical diagnoses (eg, myocarditis, right heart failure, and so on). Based on the primary underling cardiac diagnosis, patients were grouped into 4 categories: 1) functionally single ventricles at any stage of palliation with reportable LV and RV function and size (inclusive of hypoplastic left heart syndrome, tricuspid atresia, double outlet LV, double inlet LV, double outlet RV, double inlet RV; ie, functionally single-ventricle patients without a secondary ventricle were excluded); 2) RV at-risk including ToF, right heart failure, right-dominant atrioventricular canal defect, atrial septal defect, pulmonary atresia, total anomalous pulmonary venous return, Ebstein anomaly, and truncus arteriosus; 3) LV at-risk including coarctation of the aorta, left heart failure, myocardial infarction, left-dominant atrioventricular canal defect, L- or D-loop transposition of the great arteries, anomalous left coronary artery from the pulmonary artery, cardiomyopathy, heart transplant, ventricular septal defect, myocarditis, anemia, and iron overload; and 4) other. Grouping was tiered, such that if a patient met functionally single ventricle criteria, the patient was excluded from the LV or RV at-risk group. Next, if a patient met RV at-risk criteria, the patient was excluded from the LV at-risk group. Finally, if a patient did not meet any group criteria, the patient was placed in the other group. Note the other group includes non-CHD to reflect all indications of CMR at our institution and diversify the training set. Patient diagnoses within each subgroup are shown in Supplemental Table 1.

Similar to prior work,[14] a group stratified design was implemented for partitioning of the main cohort. ECG-CMR pairs for a given patient were restricted to

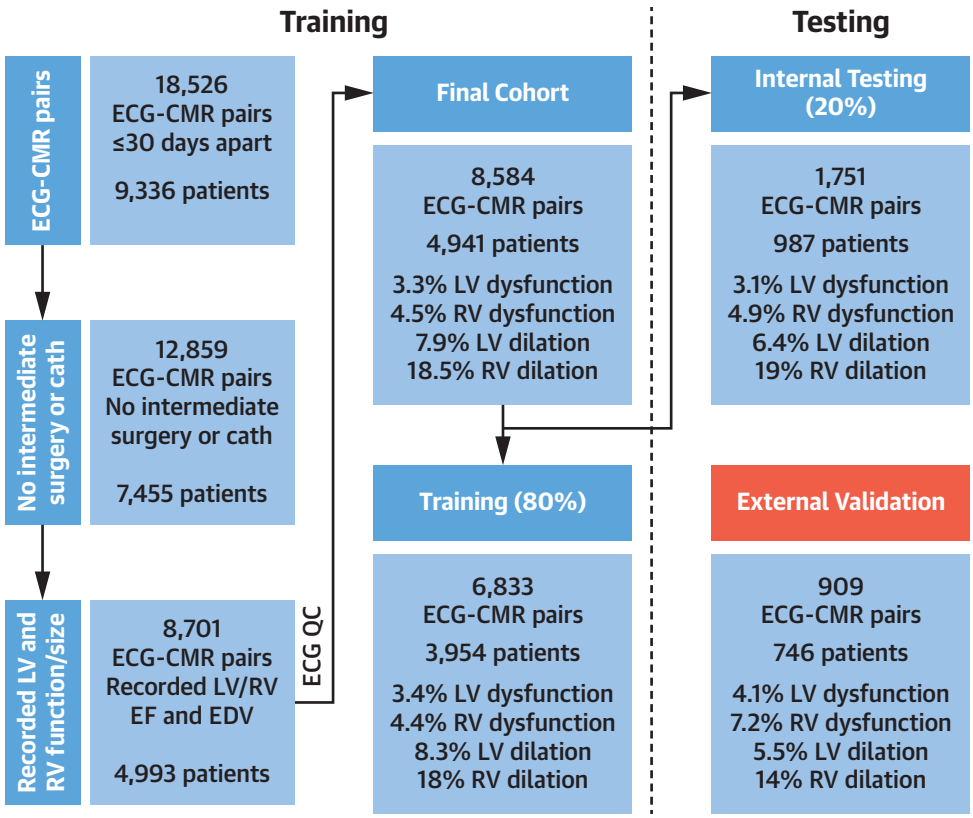
CENTRAL ILLUSTRATION **Artificial Intelligence-Enhanced Electrocardiography to Predict Biventricular Size and Function**



Mayourian J, et al. J Am Coll Cardiol. 2024;84(9):815–828.

An artificial intelligence (AI)-enhanced electrocardiography (ECG) algorithm trained on ECG- cardiovascular magnetic resonance (CMR) pairs at Boston Children's Hospital was predictive of right ventricular (RV) and left ventricular (LV) dysfunction and dilation in a congenital heart disease cohort, with external validation and model explainability. EDV = end-diastolic volume; EF = ejection fraction.

FIGURE 1 STROBE Diagram Showing Initial Patient Selection and the Final Cohort



Filtering at each data processing stage is shown. Patient partitioning for training (80%) and testing (20%) is shown, with external validation at Mount Sinai. CMR = cardiovascular magnetic resonance; EDV = end-diastolic volume; EF = ejection fraction; LV = left ventricle; QC = quality control; RV = right ventricle.

either training or testing data sets to minimize leakage of ECG-CMR pair data. The patients were randomly partitioned 80:20 into training and testing data sets.

EXTERNAL STUDY POPULATION. The external validation cohort (Mount Sinai Hospital, New York, New York, USA) had similar inclusion criteria such that only ECG-CMR pairs ≤30 days apart without an intermediate procedure were included.

DATA RETRIEVAL, QUALITY CONTROL, AND DATA PREPROCESSING. Data retrieval, quality control, and data preprocessing is analogous to our previous work[14]; for details, see the Supplemental Methods.

DEFINITION OF OUTCOMES. The individual outcomes included greater than mild LV dysfunction (LVEF ≤40%), RV dysfunction (RVEF ≤35%), LV dilation (LVEDV z-score ≥4, corresponding to 121 mL/m^2 in women and 141 mL/m^2 in men), and RV dilation (RVEDV z-score ≥4, corresponding to 130 mL/m^2 in women and 143 mL/m^2 in men). The composite biventricular dysfunction outcome was defined as LVEF ≤40% and RVEF ≤35%. The primary outcomes were used to train and test the model used herein. In a secondary RV volume-specific model, individual outcomes included RVEDV z-score ≥4, indexed RVEDV ≥160 mL/m^2, and indexed RVEDV ≥180 mL/m^2.

As a secondary outcome analysis, we evaluated time to all-cause mortality after CMR. When multiple ECG-CMR pairs were available in a patient, a randomly selected ECG-CMR pair was included for survival analysis.

MODEL SELECTION, ARCHITECTURE, AND TRAINING. A transfer learning approach was utilized for model development. We started with model weights from our recently developed AI-ECG model[14] to predict left ventricular dysfunction or remodeling in patients ≤18 years of age without major CHD, and then trained our model on the aforementioned training set.

The training set was further partitioned 95% for training and 5% for validation to allow for hyperparameter tuning. Our convolutional neural network used 12 × 2,048 ECG inputs with an architecture inspired by the residual network (ie, including skip connections) adapted for unidimensional signals.[14,15] Network architecture is identical to our previous work.[14]

Model hyperparameters were tuned by performing a grid search on the training set over the following values: kernel size [3, 9, 17], batch size [8, 32, 64], and initial learning rate [0.01, 0.001, 0.0001]. The average cross-entropy was minimized using the Adam optimizer. We used maximum 150 epochs with early stopping based on validation loss. The model with the lowest validation loss during hyperparameter tuning was selected as the final model (kernel size 17, batch size 32, learning rate 0.001).

PERFORMANCE EVALUATION. Model performance was evaluated only on the test group. To account for class imbalance and capture model performance at various thresholds, we evaluated the area under the receiver operating curve (AUROC) and area under the precision-recall (ie, positive predictive value-sensitivity) curve (AUPRC). In addition, positive predictive value, negative predictive value, sensitivity, and specificity were evaluated at the Youden index threshold (ie, maximizing sensitivity and specificity) in the training set. Metric CIs were computing using 1,000 bootstrap resamples.

SURVIVAL ANALYSIS. For mortality analysis, Kaplan-Meier curves were constructed to visualize survival probabilities of patients based on their AI-ECG predictions. Patients were deemed high- or low-risk using the Youden Index in the training cohort as a cutoff threshold. Kaplan-Meier curves were generated using a single random ECG-CMR pair per patient. Patients were censored at the time last known alive. Statistical comparison between Kaplan-Meier curves was performed using the log-rank test.

MODEL EXPLAINABILITY. To explain model behavior across all outcomes, the following analyses were performed: 1) median waveform analysis; and 2) saliency mapping. For details, see the Supplemental Methods.

DATA AVAILABILITY AND SOFTWARE. Requests for Boston Children's Hospital data and related materials will be internally reviewed to clarify if the request is subject to intellectual property or confidentiality constraints. Shareable data and materials will be released under a material transfer agreement for noncommercial research purposes. Institutional Review Board approval was obtained by each respective institution in this study. Programming code used (Supplemental Methods) to perform the analyses are available upon reasonable request.

TABLE 1 Baseline Characteristics of Internal Training and Testing Cohorts

	Training (80%)	Testing (20%)
Patient-level characteristics		
Patients	3,954	987
Male	2,209 (56)	564 (57)
Death	175 (4.4)	43 (4.4)
Age of death, y	29.6 (20.7-45.2)	24.1 (18.3-43.2)
ECG-CMR pair grouping		
Pairs	6,833	1,751
LV at-risk	2,120 (31)	519 (30)
RV at-risk	2,341 (34)	655 (37)
Functionally SV	666 (9.7)	156 (8.9)
Other	1,706 (25)	421 (24)
ECG-CMR pair characteristic		
Pairs	6,833	1,751
Age at CMR, y	20.7 (15.5-30.4)	20.7 (15.6-29.6)
Heart rate, beats/min	71.0 (62.0-81.0)	70.0 (61.0-81.0)
QRS axis	76.0 (55.0-95.0)	74.0 (54.0-94.0)
T axis	57.0 (40.0-74.0)	59.0 (40.0-76.0)
P axis	47.0 (31.0-61.0)	47.0 (32.0-61.0)
PR interval, ms	152.0 (136.0-172.0)	154.0 (136.0-174.0)
QRS interval, ms	102.0 (90.0-132.0)	104.0 (90.0-134.0)
QT interval, ms	402.0 (376.0-430.0)	404.0 (378.0-434.0)
QTc interval, ms	435.0 (414.0-458.0)	435.0 (415.0-459.0)
LVEF, %	58.8 (54.5-63.6)	58.8 (54.5-63.2)
RVEF, %	54.1 (48.5-59.4)	54.2 (48.2-59.6)
LVEDV, mL	152.2 (121.2-193.4)	148.3 (118.8-189.1)
LVEDV z-score	0.6 (−0.4 to 1.9)	0.6 (−0.4 to 1.7)
RVEDV, mL	179.4 (137.0-229.7)	185.2 (139.3-230.5)
RVEDV z-score	1.5 (0.3-3.2)	1.7 (0.4-3.4)
ECG-CMR pair outcomes		
Pairs	6,833	1,751
LVEF ≤40% (yes)	231 (3.4)	54 (3.1)
RVEF ≤35% (yes)	300 (4.4)	85 (4.9)
LVEDV z-score ≥4	566 (8.3)	112 (6.4)
RVEDV z-score ≥4	1,246 (18)	341 (19)

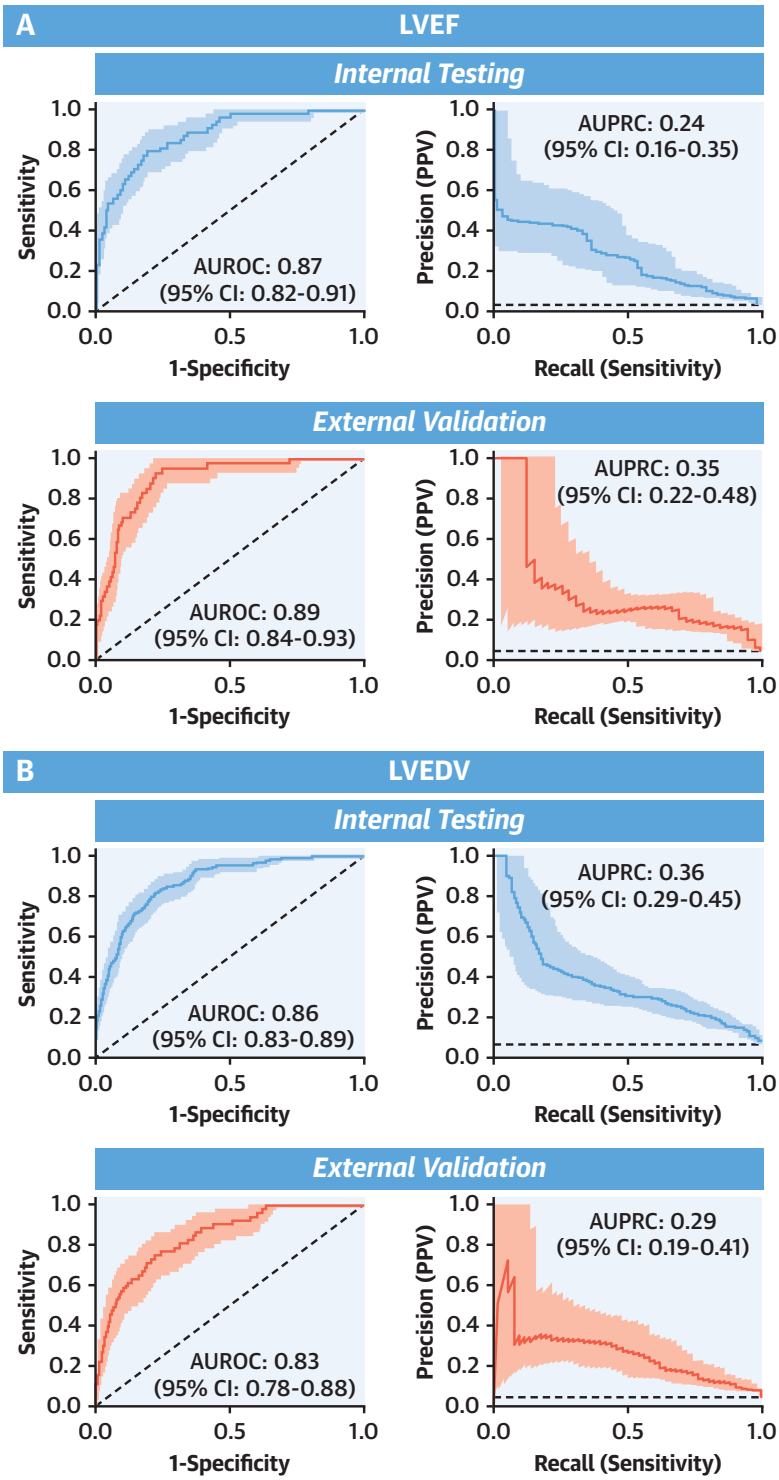
Values are n, n (%), or median (Q1-Q3).
CMR = cardiovascular magnetic resonance; ECG = electrocardiogram; EDV = end-diastolic volume; EF = ejection fraction; LV = left ventricular; RV = right ventricular; SV = single ventricle.

RESULTS

PATIENT POPULATION BASELINE CHARACTERISTICS AND OUTCOMES. Of the 18,526 ECG-CMR pairs ≤30 days apart, 12,859 were without an intermediate surgery or catheterization, with 8,701 ECGs (n = 4,993) meeting entry criteria and 8,584 ECGs (n = 4,941) passing quality control, thus forming the main study cohort (**Figure 1**).

The training cohort comprised of 6,833 ECG-CMR pairs (n = 3,954; median age at CMR 20.7 years [Q1-Q3: 15.5-30.4 years]; 56% men; 19% with ToF), 3.4% with LV dysfunction, 4.4% with RV dysfunction, 8.3% with LV dilation, and 18% with RV dilation

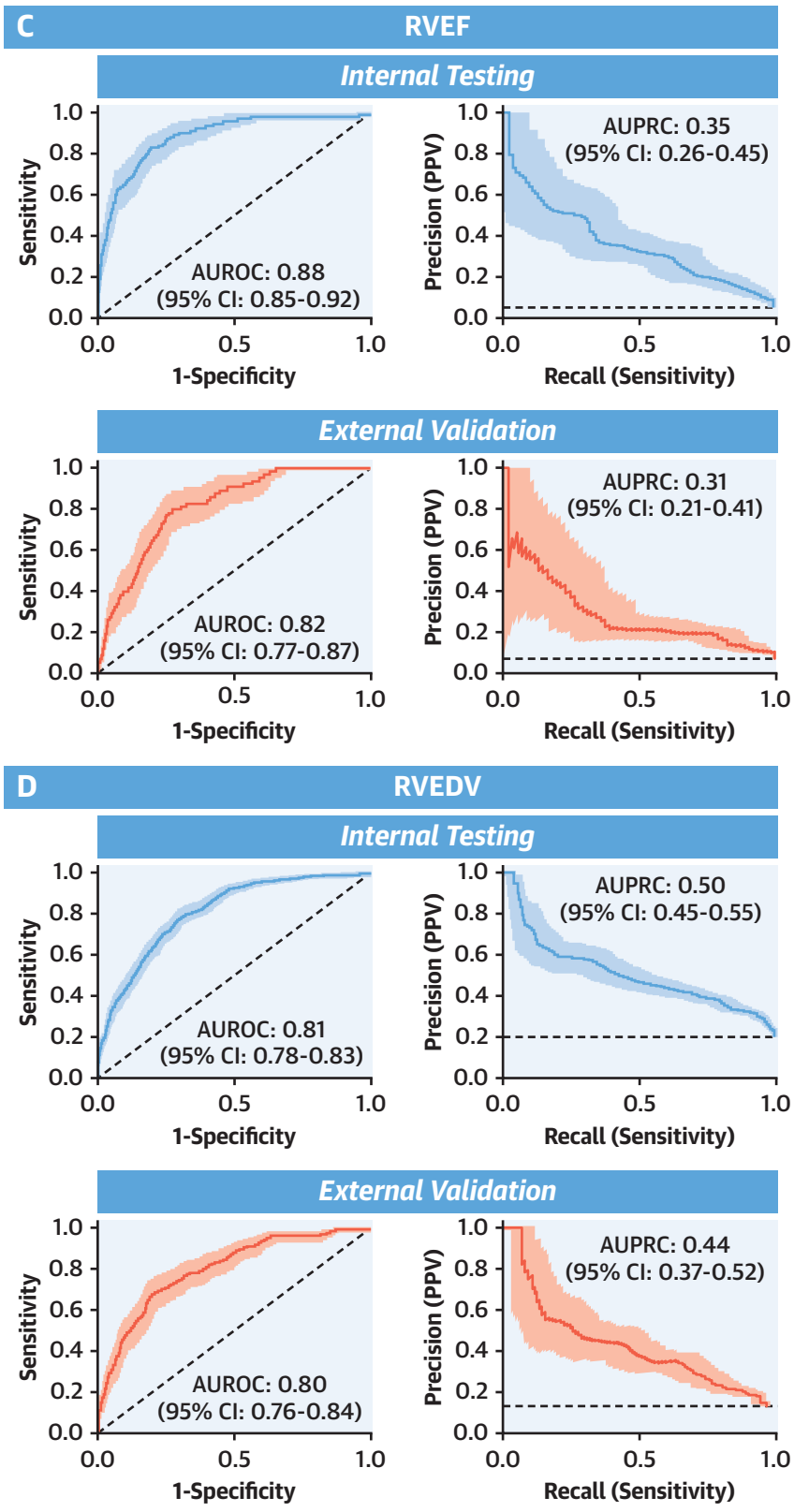
FIGURE 2 External Validation of Electrocardiogram-Based Deep Learning Model Performance



Model performance evaluated with receiver operating and precision-recall curves on internal (left) and external (right) cohorts for: (A) left ventricular ejection fraction (LVEF) ≤40%; (B) left ventricular end-diastolic volume (LVEDV) z-score ≥4; (C) right ventricular ejection fraction (RVEF) ≤35%; and (D) right ventricular end-diastolic volume (RVEDV) z-score ≥4. Area under receiver-operating curve (AUROC) and area under precision recall curve (AUPRC) metric values for each model and outcome are inset. Dotted line represents chance. 95% CIs are shown using bootstrapping. PPV = positive predictive value.

Continued on the next page

FIGURE 2 Continued



(**Table 1**). A majority of the patients were in the LV (31%) and RV at-risk groups (34%), followed by the other (25%) and single-ventricle (10%) groups. When stratifying by group, there were notable differences in ECG, CMR, and outcome characteristics (Supplemental Table 2). Further details on demographics, ECG, and CMR characteristics are summarized in **Table 1**, with similar baseline characteristics in the internal test group. Within the training and test cohorts, 4.4% of patients died at a median age of 29.6 years (Q1-Q3: 20.7-45.2 years) and 24.1 years (Q1-Q3: 18.3-43.2 years), respectively.

The external cohort comprised of 909 ECG-CMR pairs (746 patients; median age at CMR 25.4 years [Q1-Q3: 15.8-32.3 years]; 56% men; 15% with ToF), 4.1% with LV dysfunction, 7.2% with RV dysfunction, 5.5% with LV dilation, and 14% with RV dilation.

AI-ECG MODEL PERFORMANCE. Model performance (**Figure 2**, **Table 2**) was similar between internal testing (LV dysfunction: AUROC: 0.87, AUPRC: 0.24; LV dilation: AUROC: 0.86, AUPRC: 0.36; RV dysfunction: AUROC: 0.88, AUPRC: 0.35; RV dilation: AUROC: 0.81, AUPRC: 0.50) and external cohort (LV dysfunction: AUROC: 0.89, AUPRC: 0.35; LV dilation: AUROC: 0.83, AUPRC: 0.29; RV dysfunction: AUROC: 0.82, AUPRC: 0.31; RV dilation: AUROC: 0.80, AUPRC: 0.44). When looking at single random ECG-CMR pairs per patient, model performance remained similar (Supplemental Figure 1). Model performance was comparable in detecting biventricular dysfunction (Supplemental Figure 2).

Model performance metrics were subsequently evaluated (**Table 2**) at the Youden index threshold in the training set. LVEF, LV dilation, and RV dilation had similar sensitivities (ranging from 0.78-0.94) across internal test and external cohorts. In contrast, for RVEF, the sensitivity was markedly higher in the internal test cohort compared with the external validation cohort (0.83 vs 0.53, respectively).

SUBGROUP ANALYSIS. We next examined AI-ECG performance when stratifying by age and sex (**Figure 3**) (for full model performance metric details, see Supplemental Tables 3 and 4). There was no clear relationship between age and model performance. Performance was higher for men than women for RVEF and RVEDV.

We similarly performed subgroup analysis by CMR grouping (**Figure 3**, Supplemental Table 5). For LVEF, the LV at-risk group performed better than the RV at-risk group, whereas for LVEDV, the RV at-risk group performed better than the LV at-risk group. Across all outcomes, the single-ventricle group had the lowest performance. For RVEDV, performance was worse for the RV at-risk group than the LV at-risk group. Overall model performance was broadly insensitive to exclusion of common CHD lesions in this cohort (Supplemental Figure 3).

Finally, we assessed a specific CHD lesion with a common CMR indication—ToF. ToF performance was highly representative of the encompassing RV at-risk group (Supplemental Figure 4), with notably poor RV dilation performance. As shown in Supplemental Figure 5, the RV at-risk group RV dilation performance improved when excluding ToF (but not other RV at-risk CHD lesions), highlighting the challenge in predicting RVEDV z-score ≥4 for ToF specifically. When using a secondary RV volume-specific model, performance in ToF remained poor for RVEDV z-score ≥4 (AUROC: 0.65), but was improved for higher and more clinically relevant cutoffs such as indexed RVEDV ≥160 mL/m^2 (AUROC: 0.76) and ≥180 mL/m^2 (AUROC 0.78) (Supplemental Figure 6).

SURVIVAL ANALYSIS. In the testing cohort, median follow-up after CMR was 6.4 years (Q1-Q3: 3.2-10.7 years), with 4.4% of patients experiencing all-cause mortality at a median age of 24.1 years (Q1-Q3: 18.3-43.2 years). Given the established use of CMR to risk-stratify patients with ToF, survival after CMR was assessed in this cohort when stratified as high- or low-risk based on AI-ECG predictions of LV or RV dysfunction. There was significantly lower survival in ToF patients (**Figure 4**) with LV or RV dysfunction based on AI-ECG predictions. Similar trends were identified for the overall cohort, LV at-risk patients, and RV at-risk patients (Supplemental Figure 7).

SALIENCY MAPPING. In an attempt to interpret the model, we performed saliency mapping and median waveform analysis. As shown in **Figure 5**, ECGs at high risk of LV dysfunction had lower amplitude and widened QRS complexes with inverted T waves in V_4 to V_6. Saliency mapping demonstrated that the most influential segments were QRS complexes and T waves in V_4 and V_6. For LV dilation, saliency mapping demonstrated that the most influential segments were QRS complexes of lateral precordial leads, with high-risk ECGs having higher amplitude QRS complexes in these regions.

ECGs at high-risk of RV dysfunction had widened QRS complexes in V_1 to V_4 with inverted T waves in all precordial leads. Saliency mapping suggested the most influential segments were QRS complexes in V_4 and V_6, and T waves in V_4. Finally, ECGs at high risk of RV dilation had similar patterns and saliency maps compared to RV dysfunction. More specifically, they had widened QRS complexes with inverted T-wave in V_1 to V_4. In addition, saliency mapping demonstrates

the most influential segments were QRS complexes in V_4 and V_6, and T waves in nearly all precordial leads (ie, except V_5).

DISCUSSION

In this work, we developed and externally validated (to our knowledge) the first AI-ECG algorithm to predict LV and RV dysfunction and dilation using gold standard CMR metrics obtained from a heterogeneous cohort of pediatric and adult patients predominantly with CHD (**Central Illustration**). The model achieved similar performance during internal testing and external validation on diverse cohorts despite different baseline characteristics, suggesting model robustness and generalizability. Performance was lesion-dependent and consistently lowest in functionally single ventricles, highlighting the inherent complexities of CHD not addressed in AI-ECG algorithms to date. Importantly, we showed the model can be refined to predict outcomes that are integral to clinical decision-making in this at-risk population (Supplemental Figure 6). Patients identified at high-risk of ventricular dysfunction had lower survival in the ToF cohort, suggesting potential prognostic value. Finally, saliency mapping and median waveform analysis provided a framework to develop new insights into clinically relevant ECG characteristics predictive of LV and RV dysfunction/dilation. Altogether, these findings demonstrate the promise of AI-ECG to inexpensively screen for biventricular dysfunction/dilation in CHD, which may facilitate improved access to care and help prioritize patients for further imaging studies and/or interventions.

TABLE 2 Model Performance Metrics Across Internal and External Cohorts

	Internal Testing	Mount Sinai Hospital
LVEF ≤40%		
AUROC	0.87 (0.82-0.91)	0.89 (0.84-0.93)
AUPRC	0.24 (0.16-0.35)	0.35 (0.22-0.48)
Sensitivity	0.78 (0.67-0.87)	0.92 (0.82-1.0)
Specificity	0.81 (0.79-0.83)	0.78 (0.75-0.81)
Negative predictive value, %	99.1 (98.7-99.5)	99.5 (99.0-100)
Positive predictive value, %	11.6 (9.8-13.4)	15.9 (13.9-18.0)
Predicted negative, %	79.4 (77.8-81.1)	74.6 (71.8-77.4)
LVEDV z-score ≥4		
AUROC	0.86 (0.83-0.89)	0.83 (0.78-0.88)
AUPRC	0.36 (0.29-0.45)	0.29 (0.19-0.41)
Sensitivity	0.82 (0.75-0.88)	0.94 (0.86-1.0)
Specificity	0.76 (0.74-0.78)	0.39 (0.36-0.43)
Negative predictive value, %	98.4 (97.8-99.0)	99.1 (97.9-100)
Positive predictive value, %	19.2 (17.3-21.1)	8.6 (7.8-9.3)
Predicted negative, %	72.6 (70.6-74.5)	37.5 (34.4-40.8)
RVEF ≤35%		
AUROC	0.88 (0.85-0.92)	0.82 (0.77-0.87)
AUPRC	0.35 (0.26-0.45)	0.31 (0.21-0.41)
Sensitivity	0.83 (0.75-0.92)	0.53 (0.42-0.66)
Specificity	0.79 (0.77-0.81)	0.85 (0.83-0.88)
Negative predictive value, %	98.9 (98.4-99.5)	96.0 (95.1-97.1)
Positive predictive value, %	17.2 (15.3-19.2)	21.4 (16.8-26.5)
Predicted negative, %	76.4 (74.5-78.2)	82.6 (80.2-84.8)
RVEDV z-score ≥4		
AUROC	0.81 (0.78-0.83)	0.80 (0.76-0.84)
AUPRC	0.50 (0.45-0.55)	0.44 (0.37-0.52)
Sensitivity	0.78 (0.74-0.82)	0.91 (0.85-0.95)
Specificity	0.69 (0.67-0.72)	0.46 (0.42-0.49)
Negative predictive value, %	92.9 (91.5-94.1)	97.0 (95.1-98.4)
Positive predictive value, %	38.1 (35.9-40.5)	20.8 (19.2-22.1)
Predicted negative, %	60.2 (58.0-62.1)	40.6 (37.3-43.5)

Values are median (95% CI).

AUPRC = area under the precision-recall curve; AUROC = area under the receiver operating curve; other abbreviations as in Table 1.

CMR IN CHD. The indications for CMR in CHD are broad and include evaluating anatomy, physiology/hemodynamics (eg, pulmonary/systemic flow, collateral flow evaluation), myocardial scarring, valve function, and biventricular size/function.

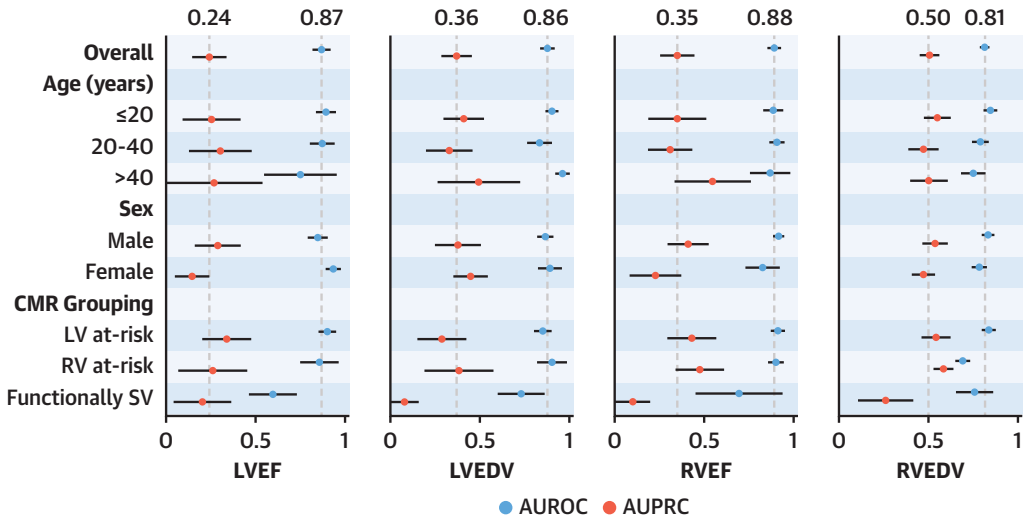
The known limitations of echocardiography for pediatric and adult patients to assess the RV[8,16,17] makes referral for right-sided lesions a common indication for CMR. Common right-sided lesions include ToF, pulmonary atresia, atrial septal defects, and Ebstein anomaly. As these patients age, they are at increased risk of RV myopathy, which can lead to significant morbidity and mortality.[5]

Lesion-specific guidelines provide recommended CMR surveillance frequency to help guide the timing of interventions and heart failure management.[18] For example, in adults with ToF at risk of pulmonary regurgitation and RV or LV dilation/dysfunction, guidelines recommend CMR surveillance every 12 to 36 months.[18] In ToF, biventricular function and size inform the need for pulmonary valve replacement, with consensus criteria for pulmonary valve replacement dependent on RVEDV, RVEF, and LVEF.[19]

Although CMR carries these tremendous benefits specific to the pediatric cardiology population, it is limited by being time-, resource-, and cost-intensive, making it of interest to develop a convenient, standardized, and inexpensive tool to help inform timing of CMR.

AI-ECG CLINICAL SIGNIFICANCE AND IMPLICATIONS. Compared with CMR, AI-ECG has several advantages: 1) it is rapid, conveniently obtained, and cost-effective; 2) it is standardized, without being subject to inter-rater and intrarater variability; and 3) it is safe and there are no practical limitations hindering use. In this study, we utilized AI-ECG to infer only a portion of measurements obtained from CMR (ie, biventricular size/function). We envision the clinical

FIGURE 3 Subgroup Model Performance



Forest plot showing AUROC (blue) and AUPRC (red) performance when stratifying by age and CMR subgroups for the following outcomes: LVEF ≤40%, LVEDV z-score ≥4, RVEF ≤35%, and RVEDV z-score ≥4. Dotted lines represent overall cohort performance, with metrics inset above each forest plot. 95% CIs are shown using bootstrapping. SV = single ventricle; other abbreviations as in Figures 1 and 2.

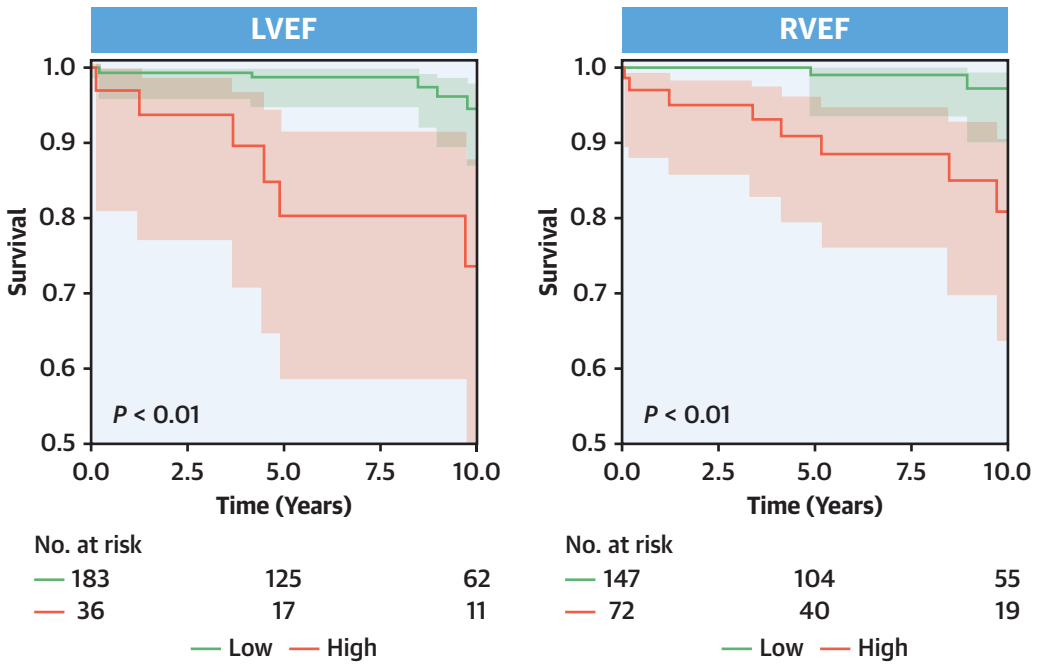
implementation of this AI-ECG algorithm may serve as a screening tool to inform timing of CMR, as well as to improve access to care.

As a screening tool, this algorithm may have economic value and lower the costs associated with care of individuals with CHD[20] by reducing the frequency of noninvasive imaging (eg, echocardiography, CMR), which may even reduce the frequency of diagnostic and/or interventional catheterizations. As a thought example, the model has the capacity to achieve a negative predictive value of 99.1% for LV dysfunction in the overall internal test cohort and 99.5% externally, with the potential to reduce CMRs for LV size/function indications by 75% to 79%. For RV dysfunction, negative predictive values of 98.9% and 96.0% were achieved, respectively, with the potential to reduce CMRs for RV size/function indications by 76% to 83%. Finally, in the case of the ToF subgroup, the secondary RV dilation-specific model has the potential to reduce CMRs for RV size indications by 28% at a 90% sensitivity to predict RVEDV index of 160 mL/m^2 (ie, one of the proactive criteria for pulmonary valve replacement in ToF).[19] On the other hand, the AI-ECG predictions may also help identify high-risk patients at an earlier age who will require closer monitoring and/or earlier CMR studies.

Finally, this algorithm may help improve access to care, especially in centers/areas without reliable access to CMR (and thus limited measurements of RV size/function). Currently, approximately two-thirds of the world population does not have access to specialized CHD services, with the majority of CHD patients now adults.[21] This democratization of specialty expertise may similarly be valuable for hospitals with low pediatric volumes and/or limited pediatric cardiology experience.

MODEL EXPLAINABILITY AND AI-ECG INSIGHTS. Model explainability increases the transparency/interpretability of models for clinicians and may aid clinicians in identifying ECG signatures resembling myopathy. High-risk features for RV dysfunction and dilation include widened QRS complexes and inverted T waves in precordial leads. These findings differ from recent adult AI-ECG saliency maps of RV pathology,[8,9] highlighting the unique considerations in CHD. Notably, widened QRS complexes—especially in ToF—are associated with morbidity, mortality, and RV pathology.[22-24] Salient features identified for LVEF are similar to our previous work[14] in children without major CHD and adults with structurally normal hearts; however, unique high-risk features were identified including reduced QRS amplitude and

FIGURE 4 Survival Analysis in Tetralogy of Fallot Based on AI-ECG Predictions



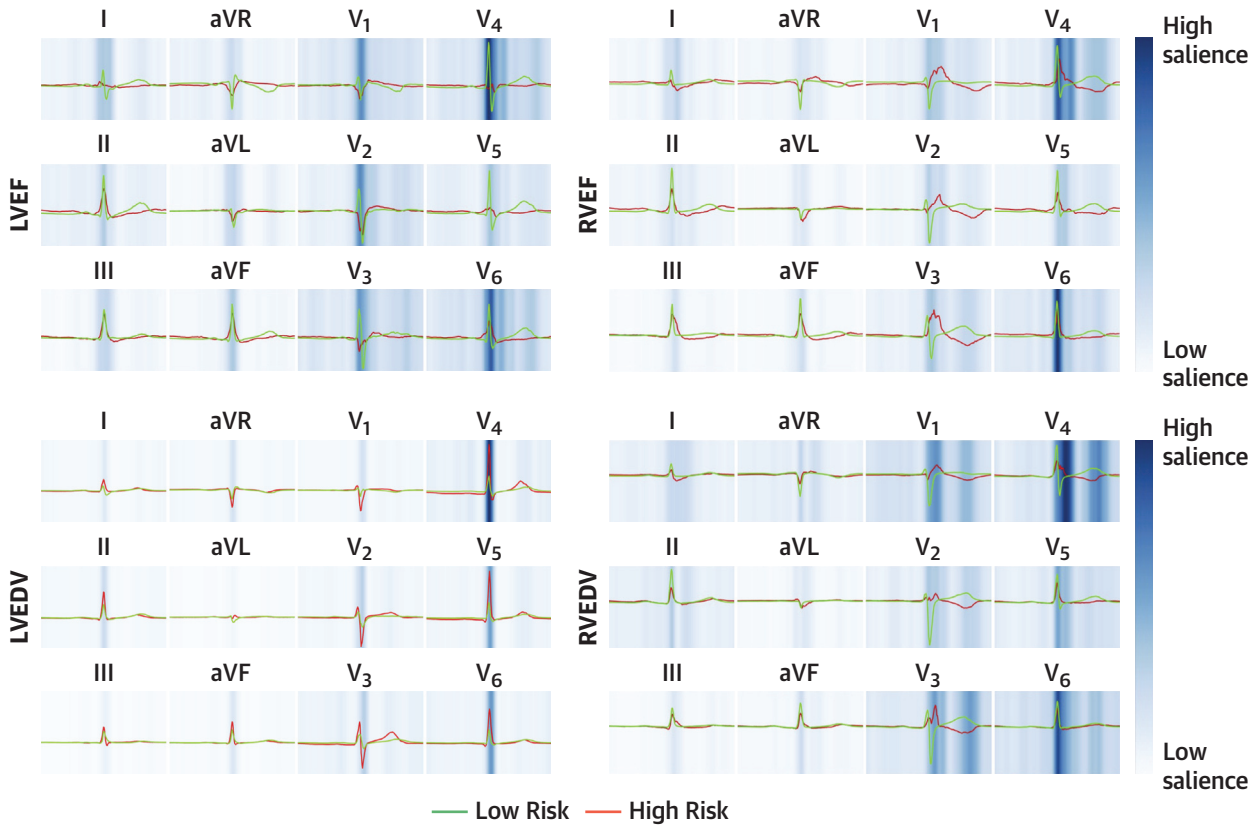
Kaplan-Meier survival curves on the Tetralogy of Fallot subgroup for patients deemed high-risk (red) vs low-risk (green) based on artificial intelligence-enhanced electrocardiogram predictions of left ventricular ejection fraction (LVEF) and right ventricular ejection fraction (RVEF) outcomes, using the Youden Index as a cutoff. Statistics based on log-rank testing. Number at risk inset below plots. 95% CIs are shown in shaded regions.

widened QRS intervals, similarly demonstrating that the distinct ECG features in a CHD cohort require tailored AI-ECG models. Interestingly, LVEDV saliency maps and high-risk features are more similar to our previous work.[14]

Several insights into CHD-specific AI-ECG challenges were also gained by performing subgroup analysis. Most notably, there was poorer performance for RVEDV z-score ≥4 for ToF (Supplemental Figure 4), and all outcomes for functionally single ventricles (**Figure 3**). In the case of ToF, we hypothesize the high prevalence of outcomes with frequently wide QRS duration and right bundle branch block at baseline led to subtle ECG changes that were not easily identified by the model; in contrast, at a higher RVEDV cutoff (eg, RVEDV index ≥160 or ≥180 mL/m^2), performance was similar to the overall model in predicting RVEDV z-score ≥4, possibly explained by more obvious ECG changes at this extreme. In the case of single ventricles, we hypothesize there are several attributable factors for poorer performance, including the following: 1) wide range of distinct anatomic/physiologic considerations; 2) unique postoperative physiology such as septating 1 morphologic ventricle (eg, double inlet left ventricle) into 2; and 3) relatively smaller sample sizes in this extremely heterogeneous subset of patients.

STUDY LIMITATIONS AND FUTURE DIRECTIONS. First, these findings are limited to patients routinely referred for CMR with a measurable secondary contributing chamber. In addition, the requirement of CMR data excludes patients with pacemakers and defibrillators. Future avenues to mitigate this limitation include obtaining biventricular function/size data from cardiac computed tomography or 3-dimensional echocardiography. Patients with body surface area <1 m^2 were excluded, thereby excluding young children. Second, a referral bias for CMR testing may lead to over-representation of certain lesions and sicker patients in the cohort, although performance was consistent across multiple care centers with different patient populations. Third, although external validation was achieved, it is of great interest to obtain external validation for each CHD lesion, and across multiple institutions globally

FIGURE 5 Explainability of Artificial Intelligence-Enhanced Electrocardiogram Outcome Predictions



Averaged median electrocardiogram waveform from the 25 highest (red) and 25 lowest (green) predictions for LVEF, LVEDV, RVEF, and RVEDV. Saliency mapping shows more (dark blue) and less (light blue) contributory regions of the electrocardiography in the background of each lead waveform. Abbreviations as in Figure 2.

to capture more diversity. Fourth, only 1 example of thresholding (ie, maximal sensitivity and specificity) was used in evaluation of model performance, as further consideration (eg, weighted loss function) is required to weigh the impact of resultant false positives (which may lead to unnecessary referrals to CMR) and false negatives (which may lead to clinical consequences of missed ventricular pathophysiology), as well as optimally set thresholds across institutions. Further multicenter external validation is warranted to refine thresholds for clinical implementation. Similarly, multicenter collaboration via federated learning[25] may help improve training/testing sample sizes and enhance diversity, which may further improve performance within each specific lesion (eg, single-ventricle patients) as well as outcome (eg, RV dilation in ToF). In addition, only ECG inputs were utilized; multimodal inputs may lead to improved model performance, especially for the complex/heterogeneous single ventricle subgroup.[26] Until then, our model is likely best served for patients without functionally single ventricles. Fifth, discrete cutoffs corresponding to ≥ moderate dysfunction or dilation were selected, although other cutoffs could have been considered as truth labels. Although all-cause mortality is routinely coded within our database, it is possible that positive cases are undocumented for our survival analyses. Further assessment of prognostic value is needed across a range of other individual CHD lesions. Although saliency mapping provides insight into model behavior, its limitations must be noted[27]; other methodologies such as layer-wise relevance propagation[28] may be considered for ongoing efforts to enhance transparency and build clinician acceptance. Last, the diagnostic categories in this study are quite heterogeneous, with grouping choices performed using clinical experience.

CONCLUSIONS

Our findings demonstrate the promise of AI-ECG to inexpensively screen for and/or predict biventricular dysfunction and dilation in patients with and without CHD as defined by CMR metrics. This tool may facilitate prioritization of patients for future interventions/imaging studies, decrease costs by reducing the frequency of echocardiograms and CMRs, provide meaningful insight into novel ECG waveforms suggestive of biventricular dysfunction/dilation, and potentially reduce disparities by improving access to care. Future multicenter collaboration and prospective trials are warranted.

ACKNOWLEDGMENT The authors thank Kai-Ou Tang for assisting with artwork.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This work was supported in part by the Thrasher Research Fund Early Career Award (Dr Mayourian), National Institutes of Health T32 Research Methods in Pediatric Cardiovascular Disease Award Number: 5T32HL007572-38 (Dr Gearhart), and National Institutes of Health grant R00-LM012926 (Prof La Cava). Dr Nadkarni has consultancy agreements with AstraZeneca, BioVie, GLG Consulting, Pensieve Health, Reata, Renalytix, Siemens Healthineers, and Variant Bio; has received research funding from Goldfinch Bio and Renalytix; has received honoraria from AstraZeneca, BioVie, Lexicon, Daiichi-Sankyo, Menarini Health, and Reata; has patents or royalties with Renalytix; owns equity and stock options in Pensieve Health and Renalytix as a scientific cofounder; owns equity in Verici Dx; has received financial compensation as a scientific board member and advisor to Renalytix; serves on the advisory board of Neurona Health; and serves in an advisory or leadership role for Pensieve Health and Renalytix; none of these relationships played a role in the design or conduct of this study. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Sunil J. Ghelani, Department of Cardiology, Boston Children's Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. E-mail: sunil.ghelani@cardio.chboston.org.

PERSPECTIVES

COMPETENCY IN PATIENT CARE AND PROCEDURAL SKILLS: AI-ECG shows promise to predict LV and RV dysfunction and dilation and mortality in patients with congenital and acquired heart disease.

TRANSLATIONAL OUTLOOK: Further studies are needed to evaluate the utility of AI-ECG to guide timing of advanced imaging modalities like CMR in patients with acquired and congenital heart disease.

REFERENCES

1. Candelino M, Tagi VM, Chiarelli F. Cardiovascular risk in children: a burden for future generations. *Ital J Pediatr*. 2022;48:57.

2. Meyer SL, St Clair N, Powell AJ, Geva T, Rathod RH. Integrated clinical and magnetic resonance imaging assessments late after Fontan operation. *J Am Coll Cardiol*. 2021;77:2480–2489.

3. Knauth AL, Gauvreau K, Powell AJ, et al. Ventricular size and function assessed by cardiac MRI predict major adverse clinical outcomes late after tetralogy of Fallot repair. *Heart*. 2008;94:211–216.

4. Ghai A, Silversides C, Harris L, Webb GD, Siu SC, Therrien J. Left ventricular dysfunction is a risk factor for sudden cardiac death in adults late after repair of tetralogy of Fallot. *J Am Coll Cardiol*. 2002;40:1675–1680.

5. Valente AM, Gauvreau K, Assenza GE, et al. Contemporary predictors of death and sustained ventricular tachycardia in patients with repaired tetralogy of Fallot enrolled in the INDICATOR cohort. *Heart*. 2014;100:247–253.

6. Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med*. 2021;27:815–819.

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

8. Duong SQ, Vaid A, My VTH, et al. Quantitative prediction of right ventricular size and function from the ECG. *J Am Heart Assoc*. 2024;13:e031671.

9. Vaid A, Johnson KW, Badgeley MA, et al. Using deep-learning algorithms to simultaneously identify right and left ventricular dysfunction from the electrocardiogram. *JACC Cardiovasc Imaging*. 2022;15:395–410.

10. O'Connor M, McDaniel N, Brady WJ. The pediatric electrocardiogram. Part I: age-related interpretation. *Am J Emerg Med*. 2008;26:221–228.

11. O'Connor M, McDaniel N, Brady WJ. The pediatric electrocardiogram part III: Congenital heart disease and other cardiac syndromes. *Am J Emerg Med*. 2008;26:497–503.

12. Alfakih K, Plein S, Thiele H, Jones T, Ridgway JP, Sivananthan MU. Normal human left and right ventricular dimensions for MRI as assessed by turbo gradient echo and steady-state free precession imaging sequences. *J Magn Reson Imaging*. 2003;17:323–329.

13. Colan SD. Early database initiatives: the Fyler codes. In: Barach PR, Jacobs JP, Lipshultz SE, Laussen PC, eds. *Pediatric and Congenital Cardiac Care: Volume 1: Outcomes Analysis*. London: Springer; 2015:163–169.

14. Mayourian J, La Cava WG, Vaid A, et al. Pediatric ECG-based deep learning to predict left ventricular dysfunction and remodeling. *Circulation*. 2024;149:917–931.

15. Ribeiro AH, Ribeiro MH, Paixao GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11:1760.

16. Lai WW, Gauvreau K, Rivera ES, Saleeb S, Powell AJ, Geva T. Accuracy of guideline recommendations for two-dimensional quantification of the right ventricle by echocardiography. *Int J Cardiovasc Imaging*. 2008;24:691–698.

17. Lopez L, Colan SD, Frommelt PC, et al. Recommendations for quantification methods during the performance of a pediatric echocardiogram: a report from the Pediatric Measurements Writing Group of the American Society of Echocardiography Pediatric and Congenital Heart Disease Council. *J Am Soc Echocardiogr*. 2010;23:465–495. quiz 576-577.

18. Stout KK, Daniels CJ, Aboulhosn JA, et al. 2018 AHA/ACC guideline for the management of adults with congenital heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2019;73(12):e81–e192.

19. Bokma JP, Geva T, Sleeper LA, et al. Improved outcomes after pulmonary valve replacement in repaired Tetralogy of Fallot. *J Am Coll Cardiol*. 2023;81:2075–2085.

20. Chowdhury D, Johnson JN, Baker-Smith CM, et al. Health care policy and congenital heart disease: 2020 focus on our 2030 future. *J Am Heart Assoc*. 2021;10:e020605.

21. Williams WG. Surgical outcomes in congenital heart disease: expectations and realities. *Eur J Cardiothorac Surg*. 2005;27:937–944.

22. Gatzoulis MA, Balaji S, Webber SA, et al. Risk factors for arrhythmia and sudden cardiac death late after repair of tetralogy of Fallot: a multicentre study. *Lancet*. 2000;356:975–981.

23. Abd El Rahman MY, Abdul-Khaliq H, Vogel M, Alexi-Meskishvili V, Gutberlet M, Lange PE. Relation between right ventricular enlargement, QRS duration, and right ventricular function in patients with tetralogy of Fallot and pulmonary regurgitation after surgical repair. *Heart*. 2000;84:416–420.

24. Cochran CD, Yu S, Gakenheimer-Smith L, et al. Identifying risk factors for massive right ventricular dilation in patients with repaired tetralogy of Fallot. *Am J Cardiol*. 2020;125:970–976.

25. Goto S, Solanki D, John JE, et al. Multinational federated learning approach to train ECG and echocardiogram models for hypertrophic cardiomyopathy detection. *Circulation*. 2022;146:755–769.

26. Jain S, Elias P, Poterucha T, et al. Artificial intelligence in cardiovascular care - part 2: applications: JACC review topic of the week. *J Am Coll Cardiol*. 2024;83(24):2487–2496.

27. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;3:e745–e750.

28. Montavon G, Binder A, Lapuschkin S, Samek W, Müller K-R. Layer-wise relevance propagation: An Overview. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller K-R, Explainable AI, eds. *Interpreting, Explaining and Visualizing Deep Learning*. Springer International Publishing; 2019:193–209.

KEY WORDS artificial intelligence, cardiovascular magnetic resonance, congenital heart disease, tetralogy of Fallot, ventricular function

APPENDIX For an expanded Methods section as well as supplemental tables and figures, please see the online version of this paper.