Electrocardiogram-based deep learning to predict left ventricular systolic dysfunction in paediatric and adult congenital heart disease in the USA: a multicentre modelling study

Joshua Mayourian, Ivor B Asztalos, Amr El-Bokl, Platon Lukyanenko, Ryan L Kobayashi, William G La Cava, Sunil J Ghelani, Victoria L Vetter, John K Triedman

Summary

Background Left ventricular systolic dysfunction (LVSD) is independently associated with cardiovascular events in patients with congenital heart disease. Although artificial intelligence-enhanced electrocardiogram (AI-ECG) analysis is predictive of LVSD in the general adult population, it has yet to be applied comprehensively across congenital heart disease lesions.

Methods We trained a convolutional neural network on paired ECG–echocardiograms (≤2 days apart) across the lifespan of a wide range of congenital heart disease lesions to detect left ventricular ejection fraction (LVEF) of 40% or less. Model performance was evaluated on single ECG–echocardiogram pairs per patient at Boston Children's Hospital (Boston, MA, USA) and externally at the Children's Hospital of Philadelphia (Philadelphia, PA, USA) using area under the receiver operating (AUROC) and precision-recall (AUPRC) curves.

Findings The training cohort comprised 124 265 ECG–echocardiogram pairs (49 158 patients; median age 10·5 years [IQR 3·5–16·8]; 3381 [2·7%] of 124 265 ECG–echocardiogram pairs with LVEF ≤40%). Test groups included internal testing (21 068 patients; median age 10·9 years [IQR 3·7–17·0]; 3381 [2·7%] of 124 265 ECG–echocardiogram pairs with LVEF ≤40%) and external validation (42 984 patients; median age 10·8 years [IQR 4·9–15·0]; 1313 [1·7%] of 76 400 ECG–echocardiogram pairs with LVEF ≤40%) cohorts. High model performance was achieved during internal testing (AUROC 0·95, AUPRC 0·33) and external validation (AUROC 0·96, AUPRC 0·25) for a wide range of congenital heart disease lesions. Patients with LVEF greater than 40% by echocardiogram who were deemed high risk by AI-ECG were more likely to have future dysfunction compared with low-risk patients (hazard ratio 12·1 [95% CI 8·4–17·3]; p<0·0001). High-risk patients by AI-ECG were at increased risk of mortality in the overall cohort and lesion-specific subgroups. Common salient features highlighted across congenital heart disaese lesions include precordial QRS complexes and T waves, with common high-risk ECG features including deep V2 S waves and lateral precordial T wave inversion. A case study on patients with ventricular pacing showed similar findings.

Interpretation Our externally validated algorithm shows promise in prediction of current and future LVSD in patients with congenital heart disease, providing a clinically impactful, inexpensive, and convenient cardiovascular health tool in this population.

Funding Kostin Innovation Fund, Thrasher Research Fund Early Career Award, Boston Children's Hospital Electrophysiology Research Education Fund, National Institutes of Health, National Institute of Childhood Diseases and Human Development, and National Library of Medicine.

Lancet Digit Health 2025; 7: e264–74

Department of Cardiology (J Mayourian MD, A El-Bokl MD, R L Kobayashi MD, S J Ghelani MD, Prof J K Triedman MD) **and Computational Health Informatics Program** (P Lukyanenko PhD, W G La Cava MD), **Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, MA, USA; Division of Pediatric Cardiology, Perelman School of Medicine at the University of Pennsylvania, Children's Hospital of Philadelphia, Philadelphia, PA, USA** (I B Asztalos MD, Prof V L Vetter MD)

Correspondence to: Prof John K Triedman, Department of Cardiology, Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
john.triedman@cardio.chboston.org

Introduction

Medical and surgical advancements have led to improved survival in children with congenital heart disease, with more than 90% of children with congenital heart disease reaching adulthood and more than 1 million adults living with congenital heart disease in the USA and Europe.[1] This growing population remains at increased risk of heart failure, a leading cause of death in people with congenital heart disease, with higher rates in more complex forms of disease.[2] Pharmacological interventions targeting neurohormonal pathways and device implantation play integral roles in improving heart failure symptoms and survival in the general adult population.[3–5] However, there remains a paucity of evidence-based therapies specific to heart failure in congenital heart disease,[6,7] making it of interest to improve preventive strategies by conveniently and inexpensively detecting early markers, such as left ventricular systolic dysfunction (LVSD). LVSD is independently associated with cardiovascular events in congenital heart disease, with guideline-directed medical

Research in context

Evidence before this study

Artificial intelligence-enhanced electrocardiogram (AI-ECG) has shown promise as an inexpensive, ubiquitous, and non-invasive screening tool to detect left ventricular systolic dysfunction (LVSD) in the general adult population. However, at present there have been few available AI-ECG applications to congenital cardiology. We searched PubMed for English-language articles using the search terms "artificial intelligence" AND "electrocardiogram" AND "congenital heart disease", from database inception to July 16, 2024. This search identified three studies, none of which involved prediction of LVSD.

Added value of this study

An externally validated AI-ECG algorithm is predictive of current and future LVSD across a range of congenital heart disease lesions across multiple health-care systems. To our knowledge, this study represents the most comprehensive application of ECG-based deep learning to the heterogeneous paediatric and adult populations with congenital heart disease to predict LVSD. AI-ECG screening provides prognostic value for future LVSD or all-cause mortality in congenital heart disease. Saliency mapping and median waveform analysis provide insight into common (deep S waves in V2 and inverted T waves in V6) and unique salient features across congenital heart disease lesions to predict LVSD.

Implications of all the available evidence

Our AI-ECG algorithm shows promise to inexpensively screen for current and predict future left ventricular dysfunction across the lifespan of patients with congenital heart disease, which might influence clinical decision making and facilitate improved access to care. Prospective trials are needed to help guide model implementation to support clinical decision making.

therapy (GDMT) and cardiac resynchronisation therapy associated with improvement in left ventricular ejection fraction (LVEF).[6]

Artificial intelligence-enhanced electrocardiogram (AI-ECG) has shown promise as an inexpensive, ubiquitous, and non-invasive screening tool to detect LVSD in the general adult population.[8] However, there is a paucity of AI-ECG applications to predict LVSD in paediatric cardiology, limited to patients without major congenital heart disease or select patients undergoing cardiac MRI.[9,10] There remains a large unmet need to leverage AI-ECG to predict LVSD across the spectrum of paediatric congenital heart disease lesions, which have substantially different epidemiology, anatomic structure, and ECG patterns that limit generalisability of applying adult AI-ECG algorithms.[11,12] In this study, we aim to address this gap by developing and externally validating an AI-ECG model on a comprehensive paediatric and adult population with congenital heart disease to predict imaging-defined LVSD.

Methods

Internal study population and patient assignment

Our study adheres to the TRIPOD+AI guidelines.[13] Patient data from Boston Children's Hospital (Boston, MA, USA) up to January, 2023 were used. Inclusion criteria comprised any patient with at least one echocardiogram with a recorded LVEF. Patients with cardiomyopathy and patients without congenital heart disease were also included to enrich the training set with overlapping pathophysiology that ultimately leads to paediatric heart failure,[14] and to broaden application to the diverse cohort encountered in the paediatric cardiology clinic.

Each qualifying echocardiogram was paired to the closest ECG, with only ECG–echocardiogram pairs 2 days or less apart included. ECGs that did not pass quality control were removed. The remaining ECG–echocardiogram pairs comprised the main study cohort. A group-stratified design was implemented to minimise data leakage by partitioning the main cohort at the patient level into training (70%) and test (30%) sets.

Restricting qualifying echocardiograms only to those with available LVEF would lead to selection bias, especially for specific lesions where LVEF is less often reported (hypoplastic left heart syndrome [HLHS], tricuspid atresia, and L-loop transposition of the great arteries [TGA]). To obtain LVEF in these cases, cardiac MRI is at times required. To address this limitation, we also assessed model performance on ECG–cardiac MRI pairs 30 days or less apart without an intermediate intervention.[10] Only patients outside the main echocardiogram cohort were included to properly assess this systematic difference.

External study population

For external validation, patient data from the Children's Hospital of Philadelphia (Philadelphia, PA, USA) were obtained. Inclusion criteria comprised echocardiograms with a recorded LVEF, and at least one ECG–echocardiogram pair 2 days or less apart. Each qualifying echocardiogram event was paired with the closest ECG. ECGs that did not pass quality control were removed. The remaining ECG–echocardiogram pairs comprised the external cohort.

Data retrieval

ECG lead placement is consistent, even in the case of dextrocardia. At both Boston Children's Hospital and the Children's Hospital of Philadelphia, all raw ECG signals were obtained from the MUSE ECG data management system (GE Healthcare; Chicago, IL, USA).

For both institutions, LVEF was obtained from echocardiogram reports, where the left ventricle always

corresponds to the morphological left ventricle. At Boston Children's Hospital, LVEF was calculated via the bullet method. At the Children's Hospital of Philadelphia, LVEF was calculated via the biplane Simpson method.[15]

At Boston Children's Hospital, patient congenital heart disease lesions were identified based on the institutional Fyler coding system.[16] At the Children's Hospital of Philadelphia, congenital heart disease lesions were identified via ICD-9 and ICD-10 codes. At both institutions, paced patients were identified based on ECG diagnoses of dual chamber pacing or ventricular pacing.

Quality control and data preprocessing

ECGs less than 10 s long or missing lead information were discarded. Less than 2% of ECGs did not pass quality control, which we deemed to occur at random (eg, accidentally unconnected ECG leads). The passing ECGs then were resampled to 250 Hz and underwent a high pass filter and trimming to 2048 samples (approximately 8 s) to facilitate conveniently working with convolutional neural networks. Details of quality control and preprocessing have been published previously.[9]

Outcomes

The primary outcome was LVEF of 40% or less (quantitatively at least moderate dysfunction). Secondary outcomes included LVEF of 50% or less (quantitatively at least mild dysfunction) and LVEF of 30% or less (quantitatively severe dysfunction). As secondary analyses, we also evaluated time to mortality and time to LVSD onset.

Model selection, architecture, and training

The model was developed on the training set, which, in accordance with our previous work,[9] was further partitioned into 95% for training and 5% for validation and hyperparameter tuning. 12-lead ECG samples of length 20148 were used as inputs to a convolutional neural network with residual block architecture, which has been adapted for unidimensional signals as previously described.[9] Further details are given in the appendix (pp 3, 8).

The final hyperparameters were obtained via a grid search, as follows: 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 minimised using the Adam optimiser.[17] A maximum of 150 epochs were used with early stopping based on validation loss with a patience of five epochs. This choice was empirically based on our previous work[9] to limit computational expense and target generalisability. The model with the lowest validation loss during hyperparameter tuning was selected as the final model (kernel size 9, batch size 64, learning rate 0·001).

Performance evaluation and statistical analyses

As justified in the appendix (p 3), multiple ECG–echocardiogram pairs per patient were allowed in the training cohort. By contrast, model performance was evaluated on test groups using one randomly selected ECG–echocardiogram pair per patient. As an ancillary approach, model performance was assessed on the first or last available ECG–echocardiogram pair.

Given the imbalanced dataset, both the area under the receiver operating curve (AUROC_ and area under the precision-recall (ie, positive predictive value [PPV]-sensitivity) curve (AUPRC) were computed. Other performance metrics evaluated included PPV, negative predictive value (NPV), sensitivity, and specificity. These metrics were calculated based on a low-risk threshold (achieving 95% sensitivity in the training set) and high-risk threshold (achieving 95% specificity in the training set). Confidence intervals were obtained via resampling with 1000 bootstraps.

Subgroup analyses

Subgroup analyses were done across a range of congenital heart disease lesions and age subgroups on test sets using all available ECG–echocardiogram pairs 2 days or less apart.

Benchmarking model performance

To benchmark our model, we compared the performance of our current model with our previous AI-ECG model for children without major congenital heart disease.[9] The benchmarking test cohort comprised all known patients with congenital heart disease who were independent from both model training sets. A subgroup analysis was done to benchmark across individual congenital heart disease lesions. Model performance was compared using the DeLong test.[18]

Time-to-event analysis

Time-to-event analysis was done for LVSD onset and all-cause mortality. Time-to-LVSD onset analysis (where the event of interest was LVEF ≤40%) was done by: including only patients with multiple ECG–echocardiogram pairs and a first echo LVEF >40%; stratifying patients into three groups based on the AI-ECG classification of the first ECG–echocardiogram pair (low risk [AI-ECG probability of at least low-risk cutoff], intermediate risk [low-risk cutoff less than AI-ECG probability of at least high-risk cutoff], and high risk [AI-ECG cutoff greater than high-risk cutoff]); and assessing time-to-event after ECG within each group. Patients who did not have LVSD onset were censored at the time of last echocardiogram.

See Online for appendix

Time-to-mortality was done by stratifying patients into the same groups based on a single random ECG per patient and assessing time-to-event after ECG within each group. Patients who did not have all-cause mortality were censored at time last known alive. Patients with unknown follow-up time after ECG were excluded from survival analysis.

Cox proportional hazards regression was used to evaluate AI-ECG classification association with time

from ECG until the event of interest (LVSD onset or all-cause mortality). Hazard ratios (HRs) were adjusted for age, with the low-risk group as reference. Statistical comparison between groups was based on log-rank testing.

	Boston Children's Hospital		Children's Hospital of Philadelphia
	Training	Internal testing	External validation
Demographics			
Patient total	49158	21068	42984
Sex			
Male	26311 (53·5%)	11251 (53·4%)	23815 (55·4%)
Female	22835 (46·5%)	9813 (46·6%)	19163 (44·6%)
Missing	12 (<0·1%)	4 (<0·1%)	6 (<0·1%)
Lesion			
Tetralogy of Fallot	1911 (3·9%)	813 (3·9%)	740 (1·7%)
Cardiomyopathy	2428 (4·9%)	1060 (5·0%)	1179 (2·7%)
Atrial septal defect	3516 (7·2%)	1523 (7·2%)	1386 (3·2%)
Complete atrioventricular canal	590 (1·2%)	261 (1·2%)	560 (1·3%)
Coarctation of the aorta	2406 (4·9%)	999 (4·7%)	701 (1·6%)
Double outlet right ventricular	529 (1·1%)	205 (1·0%)	121 (0·3%)
D-loop TGA	1003 (2·0%)	445 (2·1%)	410 (1·0%)
Ebstein	341 (0·7%)	140 (0·7%)	57 (0·1%)
Hypoplastic left heart syndrome	450 (0·9%)	191 (0·9%)	23 (<0·1%)
L-loop TGA	251 (0·5%)	109 (0·5%)	14 (<0·1%)
Pulmonary atresia	689 (1·4%)	289 (1·4%)	NA
Total anomalous pulmonary venous return	396 (0·8%)	165 (0·8%)	NA
Tricuspid atresia	242 (0·5%)	97 (0·5%)	102 (0·2%)
Truncus arteriosus	239 (0·5%)	105 (0·5%)	105 (0·2%)
Ventricular septal defect	5178 (10·5%)	2225 (10·6%)	2371 (5·5%)
Dextrocardia	285 (0·6%)	88 (0·4%)	19 (<0·1%)
Pacemaker	465 (0·9%)	212 (1·0%)	58 (0·1%)
Heart failure medications			
Spironolactone	2482 (5·0%)	1082 (5·1%)	..
Metoprolol	749 (1·5%)	324 (1·5%)	..
Enalapril	1791 (3·6%)	785 (3·7%)	..
Entresto	49 (<0·1%)	22 (0·1%)	..
Dapagliflozin	66 (0·1%)	20 (<0·1%)	..
Any of the above	3989 (8·1%)	1734 (8·2%)	..
Mortality	1478 (3·0%)	633 (3·0%)	..
Age at death, years	10·8 (1·7–20·3)	12·8 (2·3–20·6)	..
Last known alive, years	13·3 (5·2–18·1)	13·4 (5·3–18·1)	..
ECG–echocardiogram pairs			
Totals	124265	54230	76400
Age at ECG, years	10·5 (3·5–16·8)	10·9 (3·7–17·0)	10·8 (4·9–15·0)
LVEF	62·0 (57·4–66·0)	62·0 (57·6–66·0)	64·2 (60·4–67·2)
Follow-up after echocardiogram, years	3·1 (0·3–7·5)	3·2 (0·3–7·7)	..
Outcomes			
LVEF ≤50%	8525 (6·9%)	3674 (6·8%)	2888 (3·8%)
LVEF ≤40%	3381 (2·7%)	1473 (2·7%)	1313 (1·7%)
LVEF ≤30%	1490 (1·2%)	598 (1·1%)	605 (0·8%)

Data are n, n (%), or median (IQR). ECG=electrocardiogram. LVEF=left ventricular ejection fraction. NA=not available. TGA=transposition of the great arteries.

Table: **Baseline characteristics**

Coding language

The convolutional neural network used the Keras framework with a Tensorflow (Google) backend using Python 3.9. Deep learning was executed on institutional graphics processing units. All other pre-processing and post-processing code was written in Python 3.9 and R 4.0, which was executed locally.

Model explainability

Model interpretability was explored via median waveform analysis and saliency mapping, as described previously.[9,19,20] The number of samples used to generate representative median waveforms and saliency maps (n=100 for the overall test cohort and n=25 for individual congenital heart disease lesions) was selected empirically based on our previous work.[9,20] For full details see the appendix (pp 3–4).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The training cohort comprised 124265 ECG–echocardiogram pairs (49158 patients; median age 10·5 years [IQR 3·5–16·8]; 22835 [46·5%] of 49158 patients were female and 26311 [53·5%] were male; table). The median echocardiogram LVEF was 62·0% (IQR 57·4–66·0), where 3381 (2·7%) of 124265 ECG–echocardiogram pairs had an LVEF of 40% or less. The most common lesions included tetralogy of Fallot (1911 [3·9%] of 49158 patients), cardiomyopathy (2428 [4·9%] patients), atrial septal defects (3516 [7·2%] patients), coarctation of the aorta (2406 [4·9%] patients), and ventricular septal defects (5178 [10·5%] patients; table). Complex lesions with lower prevalence included HLHS (450 [0·9%] patients), L-loop TGA (251 [0·5%] patients), and tricuspid atresia (242 [0·5%] patients; table). 465 (0·9%) patients were ventricularly paced. 1478 (3·0%) patients died. Similar lesion and outcome breakdowns were found in the internal testing cohort, which comprised of 54230 ECG–echocardiogram pairs (21068 patients; median age 10·9 years [IQR 3·7–17·0]; 9813 [46·6%] of 21068 patients were female and 11251 [53·4%] were male; table). 16930 (24·1%) of 70226 patients had known congenital heart disease in the overall internal cohort.

The external validation cohort comprised 76400 ECG–echocardiogram pairs (42984 patients; median age 10·8 years [IQR 4·9–15·0]; 19163 [44·6%] of 42984 patients were female and 23815 [55·4%] were male; table). The median echo LVEF was slightly higher at 64·2% (IQR 60·4–67·2) in the external validation cohort than in the training and internal testing cohorts, with a lower prevalence of outcomes (1313 [1·7%] of 76400

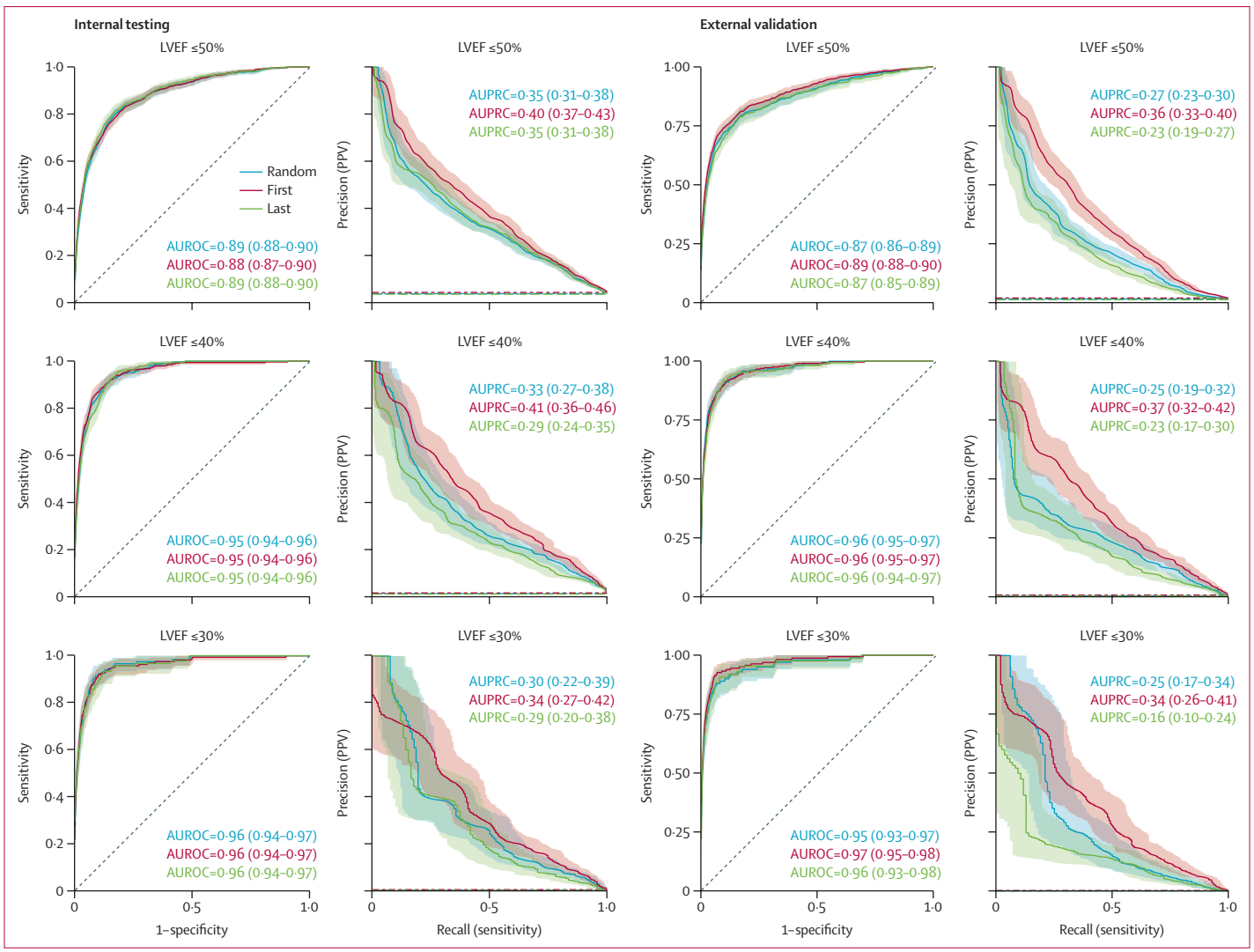

Figure 1: **Internal testing and external validation of the AI-ECG model to predict left ventricular systolic dysfunction**
Performance of the AI-ECG algorithm evaluated in the internal (left) and external (right) cohorts using receiver operating (AUROC) and precision-recall (AUPRC) curves for the following outcomes: LVEF ≤50%, LVEF ≤40%, and LVEF ≤30%. Model performance was assessed on a single ECG-echocardiogram pair per patient: random (blue), first (red), and last (green). AUROC and AUPRC metric values for each model and outcome are inset; values in parentheses are 95% CIs. Dotted line represents chance. Shaded areas represent 95% CIs, which were generated using boostrapping. AI-ECG=artificial intelligence-enhanced electrocardiogram. AUPRC=area under the precision-recall curve. AUROC=area under the receiver operating characteristic curve. LVEF=left ventricular ejection fraction. PPV=positive predictive value.

ECG–echocardiogram pairs with LVEF ≤40%; table). In general, there was a lower prevalence of each lesion and ventricularly paced patients in the external cohort compared with the internal cohort.

AI-ECG achieved high performance to discriminate mild (LVEF ≤50%), moderate (LVEF ≤40%), and severe dysfunction (LVEF ≤30%) during internal (AUROC 0·95, AUPRC 0·33) and external (AUROC 0·96, AUPRC 0·25) testing (figure 1). The model is well calibrated for each outcome of interest (appendix p 9).

Performance metrics for low-risk and high-risk cutoffs across institutions are shown in the appendix (pp 5–6). At the low-risk cutoff for LVEF of 40% or lower, sensitivity is around 90% across institutions with NPV at least 99·8% and approximately 90% of ECGs were predicted negative. At the high-risk cutoff for LVEF of 40% or lower, specificity was 97–98% across institutions with NPV at least 99·5% and PPV 13–20%.

In general, AI-ECG model performance was lower for more complex lesions and for lower prevalence lesions, with the lowest performance in dextrocardia. Modest performance was achieved across institutions for L-loop TGA and functionally single ventricle lesions such as HLHS and tricuspid atresia (figure 2A). Given that it is more difficult to ascertain LVEF on echocardiogram for these lesions, we also assessed model performance on ECG–cardiac MRI pairs. In this set of patients independent from the main cohort, model performance was again

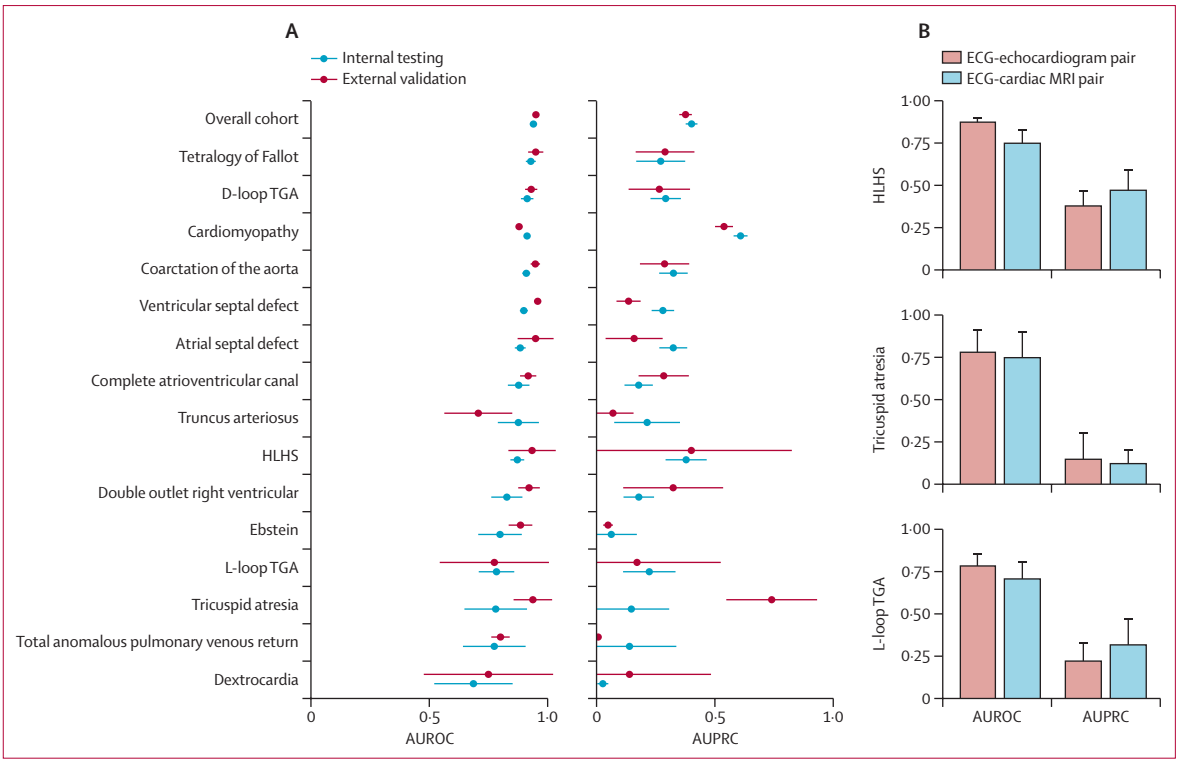

Figure 2: **Model performance across congenital heart disease lesion subgroups**
(A) Internal testing (blue) and external validation (red) AUROC (left column) and AUPRC (right column) performance when stratifying by a range of congenital heart disease lesions (listed from highest to lowest prevalence) for the outcome of LVEF ≤40%. 95% CIs are shown using bootstrapping and indicated by error bars. (B) Model performance (AUROC and AUPRC) to predict LVEF ≤40% using ECG-echocardiogram pairs (red) *vs* ECG-cardiac MRI pairs (blue) for select lesions. Error bars indicate 95% CIs. AUPRC=area under the precision-recall curve. AUROC=area under the receiver operating characteristic curve. ECG=electrocardiogram. HLHS=hypoplastic left heart syndrome. LVEF=left ventricular ejection fraction. TGA=transposition of the great arteries.

modest for HLHS, L-loop TGA, and tricuspid atresia (figure 2B). AI-ECG performance remained high across all age groups (appendix p 10), with the highest discrimination achieved for same-day ECG–echocardiogram pairs (appendix p 7). During benchmarking, this model outperformed a previously established model[9] to predict LVSD in patients with any known congenital heart disease (p<0·0001; appendix p 11), as well as across a range of individual congenital heart disease lesions (appendix p 12).

The prognostic value of the initial ECG–echocardiogram pair to predict future LVSD was assessed (figure 3). Relative to low-risk patients, high-risk patients were more likely to have a future LVEF of 40% or less in the overall cohort (HR 12·1 [95% CI 8·4–17·3]; p<0·0001), as well as in cardiomyopathy (5·8 [95% CI 3·4–10·0]; p<0·0001) and tetralogy of Fallot subgroups (8·2 [95% CI 1·7–39·8]; p=0·0094). AI-ECG predictions were also predictive of future all-cause mortality. Using a single random ECG per patient, patients with ECGs deemed high-risk were more likely to have all-cause mortality compared with those deemed low-risk. Similar prognostic trends were noted in patients independent from the training cohort who presented to the cardiology clinic and did not have an echocardiogram within 2 days of the visit (appendix p 13).

Saliency mapping and median waveform analysis were done to generate hypotheses of ECG features driving LVSD model predictions (figure 4). In the overall test cohort, salient features included the QRS complexes for V2 and V5–V6, as well as the V6 T wave. Predicted high-risk features of LVSD include deep S waves in V2 and tall R waves in V5-V6, with inverted T waves in the lateral precordial leads. In cardiomyopathy, nearly identical saliency and high-risk features were identified. In HLHS, similar saliency and high-risk features were identified, with the exception of lower amplitude high-risk R waves in V5–V6. Tricuspid atresia had similar high-risk features to HLHS, with QRS complexes being salient in precordial leads V2–V6. For tetralogy of Fallot and L-loop TGA, QRS complexes across most precordial leads were salient. Deep S waves in V2 were also high-risk for tetralogy of Fallot and L-loop TGA, with other high-risk features including inverted T waves in lateral precordial leads and wide QRS complexes.

Given the unique considerations and adverse impact of ventricular pacing in congenital heart disease, we assessed model performance and prognostic value in this specific subgroup. Modest model performance was achieved across internal and external cohorts (appendix p 14). Compared with low-risk paced patients, high-risk

patients had an increased risk of all-cause mortality (p=0·018). Salient features similarly included QRS complexes in precordial leads (V2–V4, V6), as well as V6 T waves. High-risk features included deep and wide QRS complexes in V2–V3 and inverted T waves in V6.

Discussion

To our knowledge, this study represents the most comprehensive application of ECG-based deep learning to heterogeneous paediatric and adult congenital heart disease to predict LVSD. After training on more than 100 000 ECG–echocardiogram pairs from around 50 000 patients, model performance was high for a range of congenital heart disease lesions across two large children's hospital health-care systems. AI-ECG provided prognostic value, predicting future LVSD and all-cause mortality in populations with congenital heart disease. Our findings show the promise of AI-ECG to inexpensively screen for or predict LVSD in paediatric and adult congenital heart disease, which might facilitate improved access to care, help prioritise patients for further studies or interventions, and inform ventricular pacing strategies.

Heart failure accounts for approximately 20% of hospital admissions for adults with congenital heart disease[21] and is a substantial (20–40%) cause of all-cause mortality in adults with congenital heart disease.[22] Factors contributing to myocardial dysfunction and heart failure in congenital heart disease are complex and include a multitude of haemodynamic and electrophysiological derangements (eg, volume or pressure loading, sequelae of previous surgeries, arrhythmia, chronic ventricular pacing, and ventricular scarring or fibrosis). Although

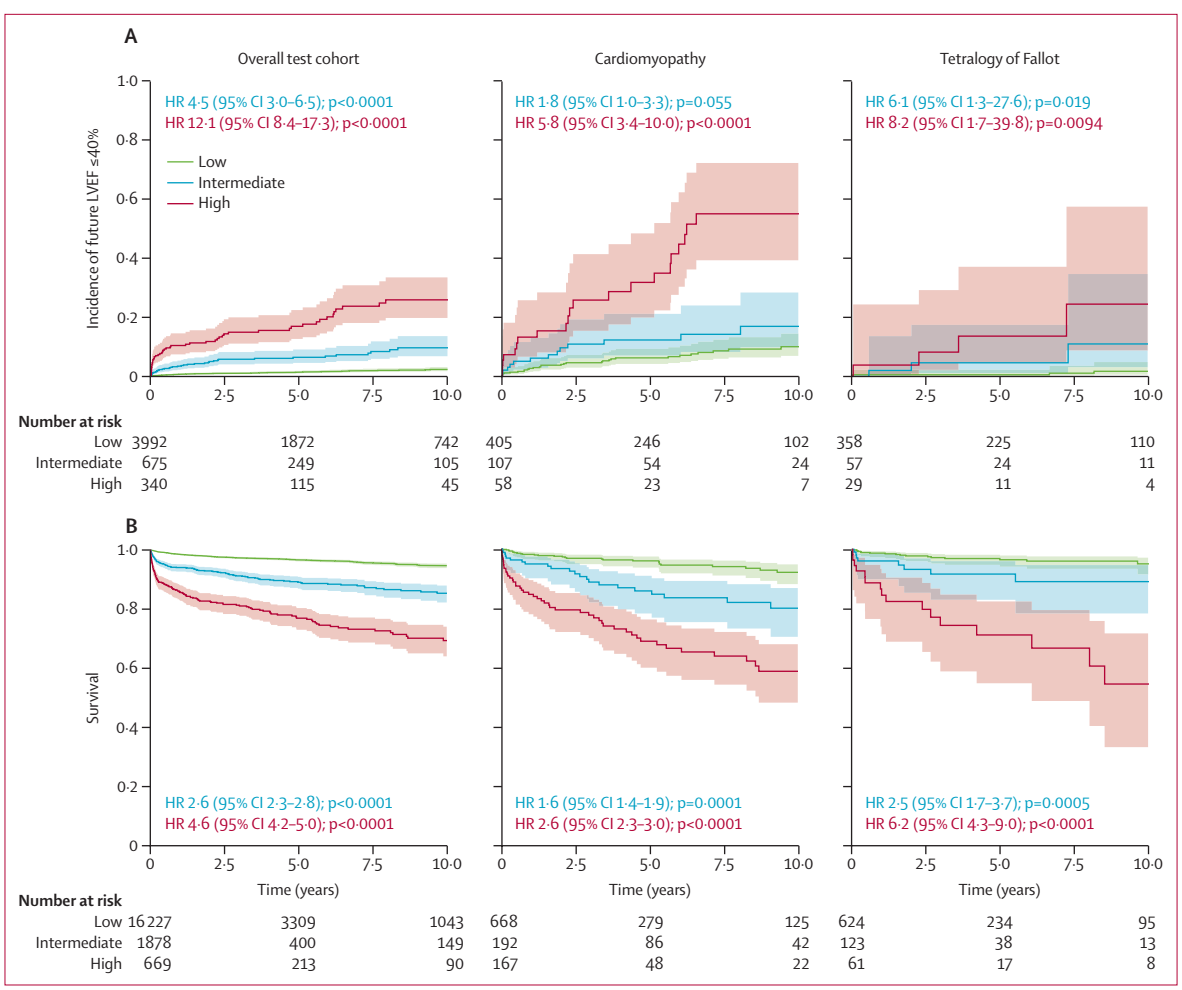

Figure 3: **Future left ventricular systolic dysfunction or mortality based on AI-ECG classification**
(A) Incidence of future LVEF ≤40% for the overall test cohort (left), patients with cardiomyopathy (middle), and patients with tetralogy of Fallot (right) initially with LVEF >40%, stratified by initial network classification (low-risk in green, intermediate-risk in blue, and high-risk in red). Number of patients at risk over the 10-year period given below the graphs. The shaded areas surrounding the curves represent 95% CIs. (B) Survival analysis when stratifying patients as low-risk, intermediate-risk, or high-risk based on AI-ECG left ventricular systolic dysfunction probabilities for the overall cohort (left), cohort with cardiomyopathy (middle), or cohort with tetralogy of Fallot (right). Colour-coded HR with 95% CI is given below the graphs (obtained via Cox regression analysis when adjusting for age, with the low-risk group as a reference). The shaded areas surrounding the curves represent 95% CIs. AI-ECG=artificial intelligence-enhanced electrocardiogram. HR=hazard ratio. LVEF=left ventricular ejection fraction.

recent data suggest sodium-glucose cotransporter-2 inhibitors reduce heart failure hospitalisation rates in adults with congenital heart disease,[7] there remains limited data showing GDMT for heart failure in adults is similarly effective in children or adults with congenital heart disease.[23,24]

The prospect of developing novel approaches for inexpensive and convenient early screening for closely

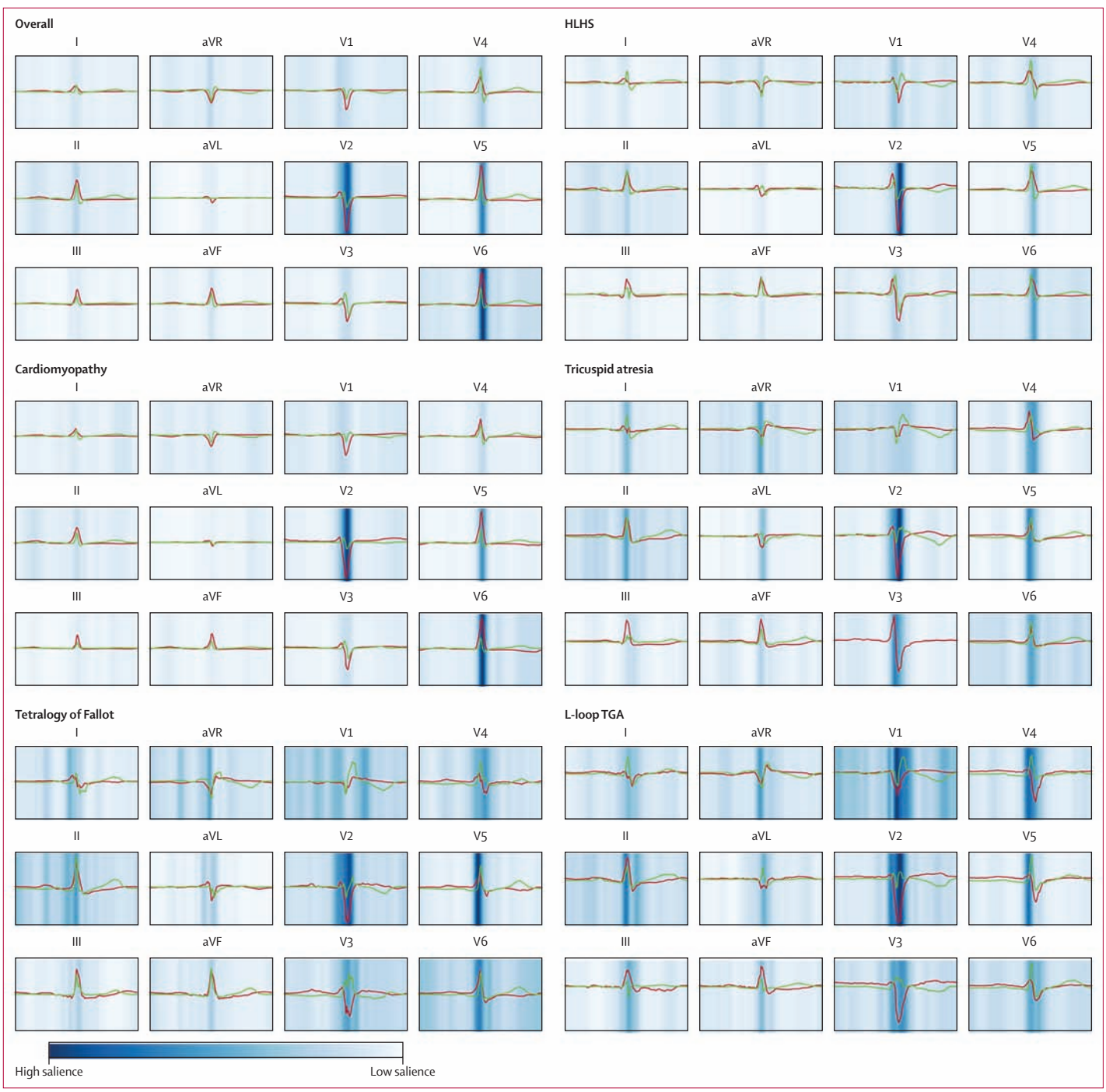

Figure 4: **Explainability of AI-ECG predictions**
Median waveforms generated in each lead using ECGs from the highest (red) and lowest (green) AI-ECG predictions of the overall test cohort, as well as cardiomyopathy, tetralogy of Fallot, HLHS, tricuspid atresia, and L-loop TGA subgroups. Saliency mapping demarcates regions of the ECG waveform having greatest (dark blue) and least (light blue) effect on each outcome. Saliency was averaged over the highest predicted ECGs for left ventricular ejection fraction ≤40%. AI-ECG=artificial intelligence-enhanced electrocardiogram. HLHS=hypoplastic left heart syndrome. TGA=transposition of the great arteries.

linked markers of heart failure, such as LVSD, motivated our efforts to develop an AI-ECG tool to predict LVSD across a range of congenital heart disease lesions. Prediction of LVSD in congenital heart disease is advantageous given that it provides a snapshot of ventricular health, carries prognostic value, and has the potential to be modified with GDMT and cardiac resynchronisation therapy.[6] Additionally, earlier identification of patients with congenital heart disease at risk for LVSD—in particular, more severe degrees of dysfunction—might help tailor treatment strategies and intensities for this group. We envision this tool has potential for multiple clinical applications including screening for LVSD in congenital heart disease (which might reduce health-care costs associated with unnecessary echocardiograms), identifying patients at greater risk of LVSD (which might lead to closer monitoring or earlier initiation of GDMT), and conveniently tracking risk of LVSD across the lifespan.

As a screening tool, our low-risk cutoff achieved around 90% sensitivity across institutions, with the potential to decrease obtaining around 60% of echocardiograms (for detecting mild dysfunction) to around 90% of echocardiograms (for detecting moderate or severe dysfunction) at NPVs of at least 99·5%. Low-risk patients also had significantly lower future LVSD and all-cause mortality, which might facilitate reduced follow-up frequency. By contrast, our high-risk cutoff was highly specific (around 95%) across institutions, with a PPV of 13–20% to detect LVEF of 40% or less. Furthermore, high-risk patients had significantly higher future LVSD and all-cause mortality, which might trigger more thorough follow-up and potentially earlier initiation of GDMT. Finally, for intermediate-risk patients, we anticipate continuing the standard of care.

The model is well calibrated across all risk groups and could theoretically serve as a tracking tool across the lifespan to provide convenient assessment of ventricular health and a response to interventions shown to improve ventricular function (eg, GDMT in congenital heart disease[6]). We envision this tool could be especially valuable for low-resource settings and children with stage B heart failure (given current recommendations for echocardiograms approximately every 6 months).[25]

Ventricular pacing in congenital heart disease—and more specifically in patients with functionally single ventricles—is associated with increased risk of heart transplant and mortality.[26] Additionally, cardiac resynchronisation therapy is associated with improvement in left ventricular systolic function.[6] Taken together with the distinct ECG waveforms of ventricular pacing, it was of interest to investigate this subgroup separately. Similar to the cohort with congenital heart disease, we envision this early case study on patients with ventricular pacing might inspire clinical translation to screen for LVSD (which might reduce health-care echocardiogram use), predict future LVSD or mortality (which might help inform on the highest-risk candidates for pacemakers), and identify salient and high-risk features of LVSD (which might inform lead placement to optimise ECG waveforms beyond using narrow QRS duration).

Congenital heart disease-specific models are desirable to account for congenital heart disease-specific anatomical structures and ECG patterns. Similar reasoning might explain why adult AI-ECG algorithms have limited generalisability to paediatric populations.[11,12]

Model explainability might provide insight into the underlying pathology that can subtly manifest within ECGs for patients with congenital heart disease. For example, common salient features across lesions include V2 QRS complexes, with high-risk features of deep S waves. These findings were similarly highlighted by Sangha and colleagues[27] in a deep learning-based ECG model for the general adult population. Mechanistically, one could speculate that the model might be focusing electrocardiographically on the anteroseptal left ventricle, recognising delayed myocardial activation. This interpretation is further reinforced by similar patterns noted in patients with pacemakers, suggesting a mechanism independent of the native conduction system. Delayed myocardial activation can lead to cardiac dyssynchrony, which is linked to LVSD and heart failure.[28] Future work is needed to rigorously investigate this hypothesis in relation to underlying pathophysiology.

There are several limitations of this work. First, inclusion criteria consisted of echocardiograms with recorded LVEF, which inserts selection bias for certain lesions that are less likely to have echocardiograms with recorded LVEF (eg, HLHS, L-loop TGA, and tricuspid atresia). This limitation was addressed by effective model performance using ECG–cardiac MRI pairs in patients outside the main cohort. Selection bias also exists when training on contemporaneous ECG–echocardiogram pairs, although reassuringly the model still holds predictive value in patients presenting to the clinic without an echocardiogram. Second, although AUROC and AUPRC performance remain high between internal and external cohorts (showing generalisability), performance was lower for more complex and less prevalent lesions. To address this limitation, future studies will aim to include multicentre collaboration via federated learning[29] to compile a larger set of heterogeneous data across institutions, incorporation of clinical variables, and use of various training techniques (eg, ensemble models, data augmentation, and regularisation). Additionally, although performance was high in simple lesions across institutions, performance was only evaluated in large referral centres, thus warranting future multicentre external validation across the spectrum of care levels (eg, secondary, tertiary, and quaternary). Third, model inputs require access to digital waveform data, which impedes translation to low-resource settings where such data are unavailable. Future efforts to address this limitation include training a model using ECG image inputs.[27] Our pacemaker cohort provided

proof-of-concept for AI-ECG applications, which requires further comprehensive investigation. Pragmatic randomised clinical trials[30] to prospectively determine effectiveness of AI-ECG as an ECG screening tool, to guide clinical implementation, and inform cost-effectiveness are warranted. Although heart failure is a leading cause of death in adults with congenital heart disease,[31] all-cause mortality rather than cardiac mortality was used for our survival analysis. Although saliency mapping provides insight into model behaviour, limitations must be noted.[32] Although our model does show prognostic value, other approaches to directly predict future heart failure or mortality should be considered.[33,34] Finally, patient subgroups were identified via the available coding infrastructures within each institution, which are limited by potential miscoding errors.

In conclusion, these findings show the promise of AI-ECG to inexpensively screen for and predict future LVSD in paediatric and adult congenital heart disease. This tool might facilitate improved access to care, help prioritise patients for future interventions or studies, and potentially inform ventricular pacing strategies. Future multicentre collaboration and pragmatic randomised clinical trials are warranted.

Contributors

JM and JKT conceptualised the project, obtained funding and resources, and prepared the first draft of the manuscript. JM curated the data, wrote the software, did the internal formal analysis, and created figures and tables. IBA had access to the software and did the external formal analysis. IBA, AE-B, PL, RLK, SJG, WGLC, and VLV assisted with investigation and writing (reviewing and editing). VLV, WGLC, and JKT supervised the study. JM, IBA, and JKT had direct access to and verified the underlying data reported in the manuscript. All authors had full access to all the data and had final responsibility for the decision to submit for publication.

Declaration of interests

JM and JKT are co-inventors of pending patent applications. All other authors declare no competing interests.

Data sharing

Requests for Boston Children's Hospital data, programming code, 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 non-commercial research purposes. Use of Boston Children's Hospital and the Children's Hospital of Philadelphia data were approved by their respective institutional review boards.

Acknowledgments

We acknowledge Boston Children's Hospital's High-Performance Computing Resources Clusters Enkefalos 2, which was made available for conducting the research reported in this publication. We acknowledge support from the Kostin Innovation Fund (JM and JKT), Thrasher Research Fund Early Career Award (JM), Boston Children's Hospital Electrophysiology Research Education Fund (JM), National Institutes of Health (NIH) grant T32HD040128 (National Institute of Childhood Diseases and Human Development; PL), and National Library of Medicine NIH grant R00-LM012926 (WGLC).

References

1 Dellborg M, Giang KW, Eriksson P, et al. Adults with congenital heart disease: trends in event-free survival past middle age. *Circulation* 2023; **147:** 930–38.

2 Menachem JN, Schlendorf KH, Mazurek JA, et al. Advanced heart failure in adults with congenital heart disease. *JACC Heart Fail* 2020; **8:** 87–99.

3 Dargie HJ. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet* 2001; **357:** 1385–90.

4 Pfeffer MA, Braunwald E, Moyé LA, et al. Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction. Results of the survival and ventricular enlargement trial. *N Engl J Med* 1992; **327:** 669–77.

5 Poole JE, Olshansky B, Mark DB, et al. Long-term outcomes of implantable cardioverter-defibrillator therapy in the SCD-HeFT. *J Am Coll Cardiol* 2020; **76:** 405–15.

6 Egbe AC, Miranda WR, Pellikka PA, DeSimone CV, Connolly HM. Prevalence and prognostic implications of left ventricular systolic dysfunction in adults with congenital heart disease. *J Am Coll Cardiol* 2022; **79:** 1356–65.

7 Neijenhuis RML, MacDonald ST, Zemrak F, et al. Effect of sodium-glucose cotransporter 2 inhibitors in adults with congenital heart disease. *J Am Coll Cardiol* 2024; **83:** 1403–14.

8 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.

9 Mayourian J, La Cava WG, Vaid A, et al. Pediatric ECG-based deep learning to predict left ventricular dysfunction and remodeling. *Circulation* 2024; **149:** 917–31.

10 Mayourian J, Gearhart A, La Cava WG, et al. Deep learning-based electrocardiogram analysis predicts biventricular dysfunction and dilation in congenital heart disease. *J Am Coll Cardiol* 2024; **84:** 815–28.

11 Dickinson DF. The normal ECG in childhood and adolescence. *Heart* 2005; **91:** 1626–30.

12 Anjewierden S, O'Sullivan D, Mangold KE, et al. Detection of right and left ventricular dysfunction in pediatric patients using artificial intelligence-enabled ECGs. *J Am Heart Assoc* 2024; **13:** e035201.

13 Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024; **385:** e078378.

14 Hinton RB, Ware SM. Heart failure in pediatric patients with congenital heart disease. *Circ Res* 2017; **120:** 978–94.

15 Tretter JT, Chakravarti S, Bhatla P. Use of echocardiographic subxiphoid five-sixth area length (bullet) method in evaluation of adequacy of borderline left ventricle in hypoplastic left heart complex. *Ann Pediatr Cardiol* 2015; **8:** 243–45.

16 Colan SD. Early database initiatives: the Fyler codes. In: Barach PR, Jacobs JP, Lipshultz SE, Laussen PC, eds. Pediatric and congenital cancer care. Springer Nature, 2014: 163–69.

17 Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv* 2014; published online Dec 22. https://doi.org/10.48550/arXiv.1412.6980.

18 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44:** 837–45.

19 Khurshid S, Friedman S, Reeder C, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation* 2022; **145:** 122–33.

20 Mayourian J, El-Bokl A, Lukyanenko P, et al. Electrocardiogram-based deep learning to predict mortality in paediatric and adult congenital heart disease. *Eur Heart J* 2024; published online Oct 10. https://doi.org/10.1093/eurheartj/ehae651.

21 Burchill LJ, Jain CC, Miranda WR. Advancing new solutions for adult congenital heart disease-related heart failure. *J Am Coll Cardiol* 2024; **83:** 1415–17.

22 Agasthi P, Van Houten HK, Yao X, et al. Mortality and morbidity of heart failure hospitalization in adult patients with congenital heart disease. *J Am Heart Assoc* 2023; **12:** e030649.

23 Hsu DT, Zak V, Mahony L, et al. Enalapril in infants with single ventricle: results of a multicenter randomized trial. *Circulation* 2010; **122:** 333–40.

24 Shaddy RE, Boucek MM, Hsu DT, et al. Carvedilol for children and adolescents with heart failure: a randomized controlled trial. *JAMA* 2007; **298:** 1171–79.

25 Amdani S, Conway J, George K, et al. Evaluation and management of chronic heart failure in children and adolescents with congenital heart disease: a scientific statement from the American Heart Association. *Circulation* 2024; **150:** e33–50.

26 Chubb H, Bulic A, Mah D, et al. Impact and modifiers of ventricular pacing in patients with single ventricle circulation. *J Am Coll Cardiol* 2022; **80:** 902–14.

27 Sangha V, Nargesi AA, Dhingra LS, et al. Detection of left ventricular systolic dysfunction from electrocardiographic images. *Circulation* 2023; **148:** 765–77.

28 Cleland JG, Daubert JC, Erdmann E, et al. The effect of cardiac resynchronisation on morbidity and mortality in heart failure. *N Engl J Med* 2005; **352:** 1539–49.

29 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:** 755–69.

30 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:** 815–19.

31 Diller GP, Kempny A, Alonso-Gonzalez R, et al. Survival prospects and circumstances of death in contemporary adult congenital heart disease patients under follow-up at a large tertiary centre. *Circulation* 2015; **132:** 2118–25.

32 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:** e745–50.

33 Akbilgic O, Butler L, Karabayir I, et al. ECG-AI: electrocardiographic artificial intelligence model for prediction of heart failure. *Eur Heart J Digit Health* 2021; **2:** 626–34.

34 Raghunath S, Ulloa Cerna AE, Jing L, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med* 2020; **26:** 886–91.