Circulation

ORIGINAL RESEARCH ARTICLE

Pediatric ECG-Based Deep Learning to Predict Left Ventricular Dysfunction and Remodeling

Joshua Mayourian, MD, PhD, ME; William G. La Cava, PhD; Akhil Vaid, MD; Girish N. Nadkarni, MD, MPH; Sunil J. Ghelani, MD; Rebekah Mannix, MD, MPH; Tal Geva, MD; Audrey Dionne, MD; Mark E. Alexander, MD; Son Q. Duong, MD; John K. Triedman, MD

BACKGROUND: Artificial intelligence–enhanced ECG analysis shows promise to detect ventricular dysfunction and remodeling in adult populations. However, its application to pediatric populations remains underexplored.

METHODS: A convolutional neural network was trained on paired ECG–echocardiograms (≤2 days apart) from patients ≤18 years of age without major congenital heart disease to detect human expert–classified greater than mild left ventricular (LV) dysfunction, hypertrophy, and dilation (individually and as a composite outcome). Model performance was evaluated on single ECG–echocardiogram pairs per patient at Boston Children's Hospital and externally at Mount Sinai Hospital using area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC).

RESULTS: The training cohort comprised 92 377 ECG–echocardiogram pairs (46 261 patients; median age, 8.2 years). Test groups included internal testing (12 631 patients; median age, 8.8 years; 4.6% composite outcomes), emergency department (2830 patients; median age, 7.7 years; 10.0% composite outcomes), and external validation (5088 patients; median age, 4.3 years; 6.1% composite outcomes) cohorts. Model performance was similar on internal test and emergency department cohorts, with model predictions of LV hypertrophy outperforming the pediatric cardiologist expert benchmark. Adding age and sex to the model added no benefit to model performance. When using quantitative outcome cutoffs, model performance was similar between internal testing (composite outcome: AUROC, 0.88, AUPRC, 0.43; LV dysfunction: AUROC, 0.92, AUPRC, 0.23; LV hypertrophy: AUROC, 0.88, AUPRC, 0.28; LV dilation: AUROC, 0.91, AUPRC, 0.47) and external validation (composite outcome: AUROC, 0.86, AUPRC, 0.39; LV dysfunction: AUROC, 0.94, AUPRC, 0.32; LV hypertrophy: AUROC, 0.84, AUPRC, 0.25; LV dilation: AUROC, 0.87, AUPRC, 0.33), with composite outcome negative predictive values of 99.0% and 99.2%, respectively. Saliency mapping highlighted ECG components that influenced model predictions (precordial QRS complexes for all outcomes; T waves for LV dysfunction). High-risk ECG features include lateral T-wave inversion (LV dysfunction), deep S waves in V1 and V2 and tall R waves in V6 (LV hypertrophy), and tall R waves in V4 through V6 (LV dilation).

CONCLUSIONS: This externally validated algorithm shows promise to inexpensively screen for LV dysfunction and remodeling in children, which may facilitate improved access to care by democratizing the expertise of pediatric cardiologists.

Key Words: artificial intelligence ■ electrophysiology ■ pediatric cardiology ■ ventricular dysfunction ■ ventricular remodeling

ECG is a rapid, standardized, and cost-effective tool used ubiquitously for initial cardiac screening of adults and children.[1] The use of rule-, feature-, and measurement-based human interpretation of the ECG varies by level of experience and expertise. This has historically motivated the development of computer-generated interpretations on the basis of predefined rules and feature recognition algorithms that may not capture subtleties of an ECG.[2] Recent work has demonstrated that deep learning–based artificial intelligence–enhanced ECG (AI-ECG) algorithms may result in greater diagnostic fidelity; studies of this approach in adult populations

Correspondence to: John K. Triedman, MD, Department of Cardiology, Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115. Email john.triedman@cardio.chboston.org

Supplemental Material, the podcast, and transcript are available with this article at https://www.ahajournals.org/doi/suppl/10.1161/CIRCULATIONAHA.123.067750.
For Sources of Funding and Disclosures, see page 929.

Circulation is available at www.ahajournals.org/journal/circ

ORIGINAL RESEARCH ARTICLE

Clinical Perspective

What Is New?

- An artificial intelligence–enhanced pediatric ECG algorithm is predictive of left ventricular (LV) dysfunction and remodeling in children across multiple health care systems.
- The model outperforms a pediatric cardiologist benchmark for left ventricular hypertrophy; the addition of age and sex does not improve overall model performance.
- Saliency mapping provides insight into ECG components (precordial QRS complexes for all outcomes and T waves for LV dysfunction) influencing model predictions, with high-risk features including lateral T-wave inversion for LV dysfunction, deep S waves in V1 and V2 and tall R waves in V6 for LV hypertrophy, and tall R waves in V4 through V6 for LV dilation.

What Are the Clinical Implications?

- This artificial intelligence–enhanced pediatric ECG algorithm shows promise to inexpensively screen for and diagnose LV dysfunction or remodeling in children, which may facilitate improved access to care and democratize the expertise of pediatric cardiologists.
- Prospective trials may help guide model implementation to support clinical decision making.
- Saliency mapping may promote clinician discovery of novel age-dependent ECG waveform patterns consistent with LV dysfunction and remodeling.

Nonstandard Abbreviations and Acronyms

AI-ECG	artificial intelligence–enhanced ECG
AI-pECG	artificial intelligence–enhanced pediatric ECG
AUPRC	area under the precision-recall curve
AUROC	area under the receiver operating characteristic curve
ED	emergency department
ICD-11	International Classification of Diseases, 11th revision
LV	left ventricular
NPV	negative predictive value
PHN	Pediatric Heart Network
PPV	positive predictive value
SHAP	Shapley Additive Explanations

have reliably predicted a range of adult cardiovascular phenotypes, including ventricular dysfunction,[3–7] ventricular hypertrophy,[8–10] and ventricular dilation.[9,11]

Progressive anatomic and physiological changes occurring from birth to adolescence lead to age-dependent variations in pediatric ECGs. The epidemiology and patterns of normal versus abnormal pediatric ECGs differ significantly from those of adults, which may be expected to limit generalizability of applying adult AI-ECG algorithms to pediatric cohorts.[12] As an example, application of an adult hypertrophic cardiomyopathy AI-ECG model on a pediatric cohort had reduced performance with decreasing age.[13] That work[13] represents one of only a handful[14] of available AI-ECG applications to pediatric cardiology, which highlights the paucity of pediatric AI-ECG models to date that could benefit pediatric populations.

In this work, this technological gap was addressed by development and external validation of an AI-ECG model on a pediatric population (AI-pECG) to predict left ventricular (LV) dysfunction or remodeling on echocardiograms. To do so, a convolutional neural network was trained to predict human expert–classified LV dysfunction, hypertrophy, and dilation (individually and as a composite outcome) using nearly 100 000 ECG–echocardiogram pairs obtained ≤2 days apart. Model performance was then tested in >10 000 patients from an independent internal test cohort, nearly 3000 patients from a separate clinical setting (emergency department [ED]), as well as >5000 patients from an outside health care system. Saliency mapping was performed to provide model explainability and identify regions of the ECG waveform that influence model predictions.

METHODS

Internal Study Population and Patient Assignment

Patient data from Boston Children's Hospital between January 1, 2002, and December 31, 2021, were used. Inclusion criteria consisted of children ≤18 years of age with at least one echocardiogram. Echocardiograms performed in the operating room, medical intensive care unit, or cardiac intensive care unit were excluded. Patients with known major congenital heart disease[15,16] or implantable cardioverter defibrillators or pacemakers were excluded. Patients with known congenital heart disease were identified on the basis of the institutional Fyler coding system.[16] This coding system has been mapped into the International Pediatric and Congenital Cardiac Code *ICD-11* (*International Classification of Diseases*, 11th revision) nomenclature.[15] Fyler codes used to exclude major congenital heart disease in this study are shown in Table S1.

Each qualifying echocardiogram event was paired to an ECG; only ECG–echocardiogram pairs ≤2 days apart were included. For patients with multiple ECGs within this timeframe, only the ECG closest in time to the echocardiogram was included. ECG–echocardiogram pairs with ECGs failing to pass quality control (see Quality Control and Data Processing for details) were removed. The remaining ECG–echocardiogram pairs were included in the main cohort.

Similar to other studies,[7] a group-stratified design was implemented for partitioning of the main cohort. Each patient was treated as a separate group, which restricts ECG–echocardiogram pairs for a given patient to either training or testing data sets to minimize leakage of ECG–echocardiogram pair data. If an ECG or echocardiogram within an ECG–echocardiogram pair was performed in the ED, then the ECG–echocardiogram pair was placed in the ED group. These same patients with other ECG–echocardiogram pairs were forced into the internal testing group to ensure no data leakage occurred between training and testing. The remaining patients were then randomly partitioned 80:20 into training and internal testing data sets.

External Study Population

For external validation, patient data from Mount Sinai Hospital between January 1, 2002, and December 31, 2018, were used (Fyler codes were no longer available for this institution starting in 2019). Inclusion criteria consisted of ≤18 years of age and at least one ECG–echocardiogram pair ≤7 days apart. The same Fyler codes (Table S1) were used to exclude patients.

Data Retrieval

All raw ECG signals were exported from the MUSE ECG data management system (GE Healthcare). Waveform data were obtained from XML files, in which each 1-dimensional vector of data sampled at a rate of 250 Hz for a 10-second duration (2500 samples) corresponds to a lead (I, II, or V1–V6). As performed by others,[17] linear transformations of the vectors were performed on the basis of the Einthoven law[18] and Goldberger equation[19] to obtain leads III, aVF, aVL, and aVR. Age, sex, and physician-reviewed ECG measurements (QRS interval, QRS axis, T axis, P axis, PR interval, QT interval, QTc, and heart rate) are archived in an internal database at Boston Children's Hospital, which was also retrieved. In addition, ECG-based diagnosis of LV hypertrophy by expert pediatric cardiologists using conventional scoring systems[20,21] was retrieved using Fyler codes at Boston Children's Hospital for benchmarking purposes.

Echocardiogram reports written by pediatric cardiologists are archived in an internal database at Boston Children's Hospital; extracted records contained the human expert classification of the degree of LV systolic dysfunction, hypertrophy, and dilation (if any). Potential grades were trivial, mild, mild to moderate, moderate, moderate to severe, and severe. When available, quantitative measures were also obtained of LV ejection fraction (percentage and *z* score), LV mass (raw and *z* score), and LV end-diastolic volume (raw and *z* score). In an effort to make the model generalizable across multiple institutions, the Pediatric Heart Network (PHN) *z* scores were used (on the basis of healthy children with normal echocardiogram results) for LV mass and LV end-diastolic volume.[22] Because PHN *z* scores were not available for LV ejection fraction, institutional *z* scores were used (given that normative values are related to age[23]), which are publicly available online.

Quality Control and Data Preprocessing

In the case of multiple ECG recording attempts for a given ECG event, the final recorded ECG is retrieved. This ECG is then discarded if any lead is not 2500 samples long or if any lead recording has no lead information (ie, flat line). Given that ECGs are prone to recording errors (eg, baseline wander or electrical interference), a high pass filter was used[24] with a cutoff frequency of 0.8 Hz, rejection band of 0.2 Hz, ripple in passband of 0.5 dB, and attenuation in rejection band of 40 dB. The ECG was then trimmed to 2048 samples (≈8 seconds) to facilitate convenient working with convolution neural networks.

Definition of Primary Outcomes

Individual outcomes included LV systolic dysfunction, LV hypertrophy, and LV dilation. Human expert knowledge was considered as the ground truth, whereby LV systolic dysfunction was considered positive if the echocardiogram report was coded by a pediatric cardiologist for qualitatively greater than mild LV systolic dysfunction; LV hypertrophy was considered positive if the echocardiogram report was coded by a pediatric cardiologist for qualitatively greater than mild LV hypertrophy, or LV hypertrophic cardiomyopathy; and LV dilation was considered positive if the echocardiogram report was coded by a pediatric cardiologist for qualitatively greater than mild LV dilation, or LV dilated cardiomyopathy. The composite outcome was defined as having positive LV systolic dysfunction, hypertrophy, or dilation. The primary qualitative outcomes were used to train and internally test the models used herein. A similar coding structure was not available at the external site, restricting the use of external validation with qualitative cutoffs.

To further evaluate the performance of the human expert–trained model both internally (Boston Children's Hospital) and externally (Mount Sinai Hospital), quantitative cutoffs were also implemented for the outcomes, whereby LV ejection fraction, LV mass, and LV end-diastolic volume *z* scores of ≤ −4, ≥ +4, and ≥ +4 (corresponding to quantitative moderate cutoffs), respectively, were considered positive. The quantitative composite outcome was defined as having positive quantitative LV systolic dysfunction, hypertrophy, or dilation. Note that an LV ejection fraction *z* score ≤ −4 corresponds to an ejection fraction of 42% in the newborn, and linearly increases to 47% at 18 years of age.

Model Selection, Architecture, and Training

The model was developed solely on the training set, which was further partitioned 95% for training and 5% for validation to allow for hyperparameter tuning. A total of 12×2048 ECG samples were used as inputs to a convolutional neural network similar to the residual network described previously[17] that is adapted for unidimensional signals. This architecture allows neural networks to be efficiently trained with skip connections.[17] A diagram of the architecture used in this study is shown in Figure S1.

The AI-pECG network consisted of a convolutional layer followed by 4 residual blocks with 2 convolutional layers per block.[17] The convolutional layers start with 64 filters for the first layer and residual block, with a filter increase and subsampling as shown in Table S2. The output of each convolutional layer is rescaled using batch normalization and fed into a rectified linear activation unit, with subsequent dropout at a rate of 0.2. Max pooling and convolutional layers with filter length 1 are included in the skip connections to match main branch signal dimensions.[17] The output of the last block is fed into a fully

connected layer with a sigmoid activation function given that outcomes are not mutually exclusive.

Similar to other studies,[24] demographic characteristics (ie, age and sex) were also incorporated as inputs along with ECG waveforms in a separate deep learning model (AI-pECG+age+sex). The architecture (Figure S1) was modified by adding a separate part of the model, in which demographic characteristics (ie, age and sex) were concatenated and passed through a fully connected layer. The outputs for demographic and ECG model parts are individually flattened and then concatenated to obtain one feature vector. The resulting feature vector was fed into the final fully connected layer with a sigmoid activation function. For details of this model architecture, see Table S3.

For each model, the final hyperparameters were obtained by a grid search on the training set among the following options: kernel size (3, 9, or 17), batch size (8, 32, or 64), and initial learning rate (0.01, 0.001, 0.0001, or 0.00001). The average cross-entropy was minimized using the Adam optimizer. A maximum of 150 epochs was used with early stopping on the basis of validation loss. The model with the lowest validation loss during hyperparameter tuning was selected as the final model. For the AI-pECG model, final hyperparameters were kernel size 17, batch size 32, and learning rate 0.001. For the AI-pECG+age+sex model, final hyperparameters were kernel size 17, batch size 64, and learning rate 0.001.

Performance Evaluation and Statistical Analyses

Consistent with previous works,[25–27] multiple ECG–echocardiogram pairs per patient were allowed in the training cohort, as progressive anatomic and physiological changes lead to age-dependent variations in pediatric ECGs that are important to capture over time, and a patient may have a normal echocardiogram at one point in time but not another. In contrast, model performance was evaluated on the internal and external test groups using one ECG–echocardiogram pair per patient. To minimize confounding variables, the ECG–echocardiogram pair with the smallest time difference was selected for each patient.

Given the nature of an imbalanced data set, the area under the receiver operating characteristic curve (AUROC) and area under the precision-recall (ie, positive predictive value [PPV]—sensitivity) curve (AUPRC) were computed. To benchmark the LV hypertrophy model, pediatric cardiologist ECG-based diagnoses of LV hypertrophy were used. Other performance metrics evaluated included PPV, negative predictive value (NPV), sensitivity, and specificity. These metrics were calculated on the basis of thresholds achieving 90% sensitivity in the training set. For all metrics, a higher value is indicative of better performance. Resampling with 1000 bootstraps was implemented to obtain performance metric CIs.

Subgroup Analyses

Subgroup analyses were performed on the internal test set when considering all available ECG–echocardiogram pairs ≤2 days apart. Age and sex are known to influence ECG characteristics in a healthy pediatric population,[1] and were therefore explored in subgroup analyses. Age partitioning was adapted from previous reports,[1,28] with groupings by age: <1, 1 to <3, 3 to <8, 8 to <12, and 12 to 18 years. AUROCs and AUPRCs were calculated for each subgroup.

Model Explainability

In an effort to provide model interpretability, median waveform analysis and saliency mapping were performed.

As described in previous studies,[26] median waveform analysis is a technique to visualize aggregated ECG samples into a single beat. In doing so, examples of high-risk and low-risk ECGs can be visualized. Herein, the 100 highest predicted ECGs for a given outcome were used in the internal test set to create high-risk median waveforms, and the 100 lowest predicted ECGs were used in the internal test set to create low-risk median waveforms. Median waveforms were generated in each lead using the NeuroKit Python toolbox[29] by QRS complex detection, interpolating all ECGs to the same heart rate, computing the median voltage across beats for each patient, and computing the median voltage across patients for each time bin in the cardiac cycle.[26]

Saliency mapping helps identify which features of the ECG input contribute to model prediction. Saliency maps highlight components of the ECG during which a change in input (ie, ECG voltage) leads to a change in prediction.[26] Saliency maps were created using a SHAP (Shapley Additive Explanations) framework.[30] To highlight the most influential components of the ECG waveform, SHAP values for the high-risk ECGs were obtained. The previously described steps to generate median waveforms were implemented on SHAP values over time. The resultant darker regions in saliency maps correspond to greater contribution to the prediction.

Data Availability and Software

Requests for Boston Children's Hospital data and related materials will be reviewed internally to clarify whether the request is subject to intellectual property or confidentiality constraints. Sharable data and materials will be released under a material transfer agreement for noncommercial research purposes. Use of Boston Children's Hospital and Mount Sinai Hospital data were approved by their respective institutional review boards.

Programming codes used to perform the analyses are available upon reasonable request. The convolutional neural network used the Keras framework with a Tensorflow (Google) backend using Python 3.9.[31] Deep learning was executed on institutional graphics processing units. All other preprocessing and postprocessing code was written in Python 3.9[31] and R 4.0,[32] which was executed locally.

RESULTS

Training Cohort: Patient Population Characteristics

Of the 272 221 echocardiograms performed at Boston Children's Hospital on 104 508 children ≤18 years of age without major congenital heart disease, there were 122 757 ECG–echocardiogram pairs ≤2 days apart. Of these ECG–echocardiogram pairs, 119 787 ECGs

(61 722 patients) passed quality control, thus forming the main internal study cohort (Figure 1).

The training cohort comprised 92 377 ECG–echocardiogram pairs (46 261 patients; median age, 8.2 years [interquartile range, 2.9–13.8]; 54% male; 56% White), 8.2% with composite LV outcomes, 2.4% with LV dysfunction, 3.5% with LV hypertrophy, and 3.8% with LV dilation (Table 1). Median LV mass and volume *z* scores derived from the PHN network were 0.6, consistent with previous reports.[33] Training cohort patient characteristics and outcomes stratified by age group are shown in Table S4. ECG characteristics stratified by age are within range of previously reported values for healthy children.[1] ECG findings by age include a more rightward QRS, T, and P axis for age <1 year, an increasing PR and QT interval with age, and decreasing heart rate with age (Table S4). Tables S5 through S8 highlight the numerous significant differences in ECG–echocardiogram pair demographics, ECG characteristics, and echocardiogram data when stratifying by each outcome in the training cohort. Of note, ≈40% of patients with LV dysfunction had concomitant LV dilation (Table S6).

Testing Cohort: Patient Population Characteristics

The testing cohorts for evaluating model performance comprised one ECG–echocardiogram pair per patient, with 12 631 for the internal testing, 2830 for the ED, and 5088 for the external validation cohort (Figure 1; Table 2).

As shown in Table 2, age at ECG was similar between the training and internal test cohorts. In contrast, the ED, and more notably, the external validation cohort, had younger ages at ECG. Echocardiogram characteristics were similar between the training and internal test cohorts. In contrast, the ED cohort had lower LV ejection

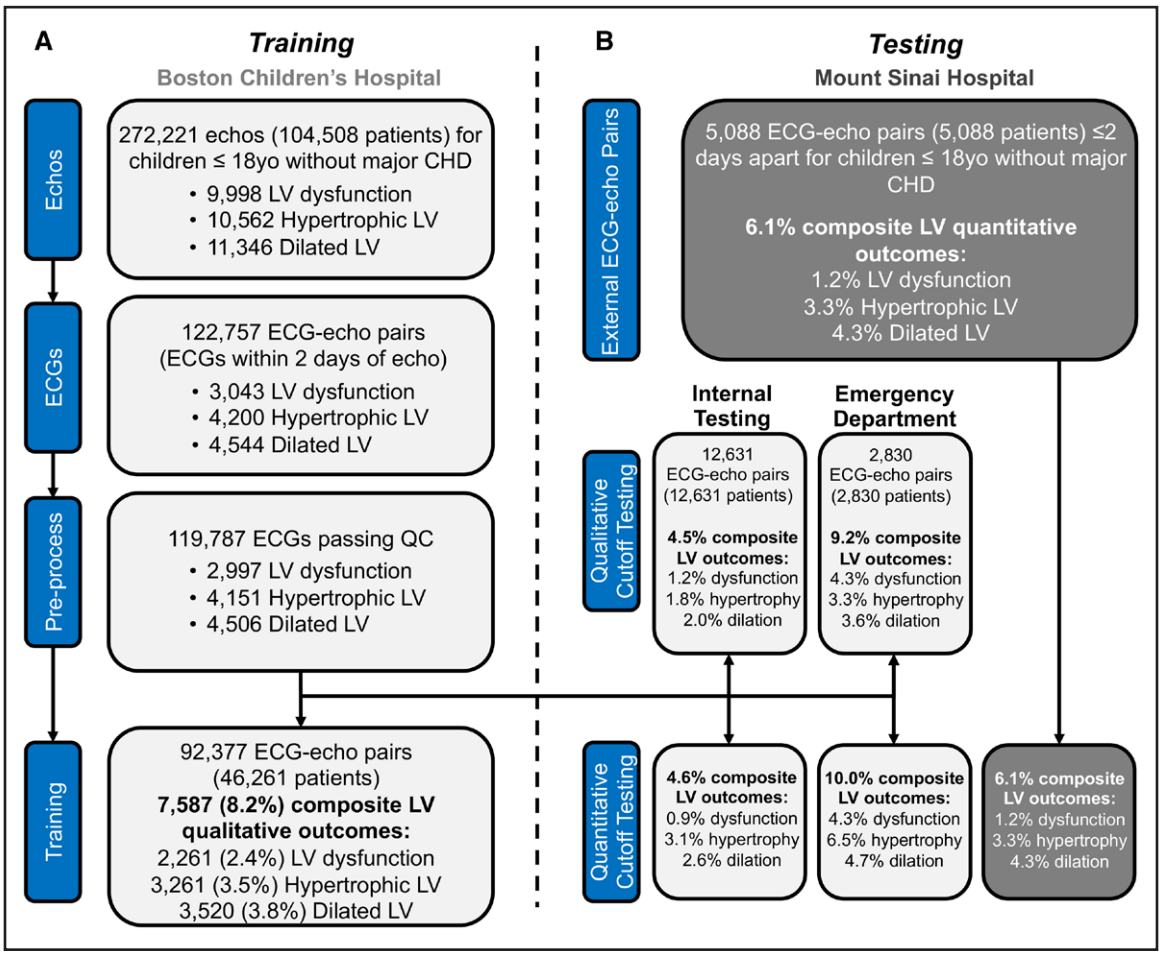

Figure 1. Schematic of study design.
A, Schematic of training design. STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) diagram showing initial patient selection and filtering at each data processing stage (with primary outcome rates shown) at Boston Children's Hospital (light gray), leading to the final training cohort. **B**, Schematic of testing design. Model performance was tested on one ECG–echocardiogram pair per patient using qualitative and quantitative outcome cutoffs across different Boston Children's Hospital settings. External validation (Mount Sinai; dark gray) was performed on 1 ECG–echocardiogram pair per patient using quantitative cutoffs. CHD indicates congenital heart disease; LV, left ventricular; and QC, quality control.

Table 1. Training Cohort Baseline Characteristics

Characteristics	Values
ECG–echocardiogram pairs	92 377
Patients	46 261
Sex	
Female	21 300 (46)
Male	24 950 (54)
Missing	11 (<0.01)
Age at ECG, y	8.2 (2.9, 13.8)
Race and ethnicity	
White	25 856 (56.0)
Black	2509 (5.4)
Hispanic	3894 (8.4)
Asian	1408 (3.0)
Other	2877 (6.2)
Missing	9717 (21.0)
ECG characteristics	
QRS interval, ms	80.0 (70.0, 90.0)
QRS axis	72.0 (52.0, 86.0)
T axis	48.0 (33.0, 61.0)
P axis	46.0 (32.0, 58.0)
PR interval, ms	126.0 (112.0, 142.0)
QT interval, ms	350.0 (312.0, 382.0)
QTc, ms	421.0 (407.0, 438.0)
Heart rate, BPM	89.0 (73.0, 113.0)
Echocardiogram characteristics	
LVEF, %	63.0 (59.0, 66.0)
LVEF, *z* score	−0.1 (−1.0, 0.6)
LV mass, g	60.7 (32.9, 98.8)
LV mass, z score	0.6 (−0.2, 1.7)
LV EDV, mL	72.6 (39.4, 115.6)
LV EDV, z score	0.6 (−0.2, 1.5)
Qualitative outcomes	
Composite LV outcome	7587 (8.2)
LV dysfunction	2261 (2.4)
LV hypertrophy	3261 (3.5)
LV dilation	3520 (3.8)

Data presented as n (%) or median (interquartile range). BPM indicates beats per minute; EDV, end-diastolic volume; EF, ejection fraction; and LV, left ventricle.

fraction and higher LV mass, with an increased prevalence in all outcomes. The external validation site had median echocardiogram *z* scores closer to 0, but nonetheless had outcome rates similar to the internal test cohort (Table 2).

AI-pECG Model Performance on Qualitative Cutoffs

After training the AI-pECG model on nearly 100 000 ECG–echocardiogram pairs with corresponding human expert–classified greater than mild LV dysfunction, hypertrophy, and dilation, model performance was evaluated.

During internal testing (Figure 2; left), the AI-pECG model achieved AUROCs of 0.86, 0.86, 0.86, and 0.88 for LV composite outcome, LV dysfunction, LV hypertrophy, and LV dilation, respectively. AUPRCs of 0.37, 0.18, 0.24, and 0.31 were achieved, respectively. During testing on the ED cohort (Figure 2; right), the AI-pECG model achieved AUROCs of 0.79, 0.81, 0.84, and 0.85 for the LV composite outcome, LV dysfunction, LV hypertrophy, and LV dilation, respectively. AUPRCs of 0.39, 0.33, 0.29, and 0.35 were achieved, respectively.

In both cases, the AI-pECG model outperformed the pediatric cardiologist expert ECG-based diagnosis of LV hypertrophy (Figure 2C; gray dot). Adding age and sex to the AI-pECG model led to similar performance (Figure 2). Model performance improved when considering all available ECG–echocardiogram pairs ≤2 days apart (Table S9) for patients in each cohort (Figure S2).

AI-pECG Model External Validation

Performance of the AI-pECG model to discriminate between quantitative cutoffs was subsequently explored internally (Figure 3; left) and externally using an outside health care system (Figure 3; right).

In general, performance when using quantitative cutoffs (Figure 3; left) was higher than qualitative cutoffs (Figure 2). During internal testing (Figure 3; left), the AI-pECG model achieved AUROCs of 0.88, 0.92, 0.88, and 0.91 for LV composite outcome, LV dysfunction, LV hypertrophy, and LV dilation, respectively; AUPRCs of 0.43, 0.23, 0.28, and 0.47 were achieved, respectively. For the ED cohort (Figure 3; left), the AI-pECG model achieved AUROCs of 0.81, 0.84, 0.82, and 0.84 for LV composite outcome, LV dysfunction, LV hypertrophy, and LV dilation, respectively; AUPRCs of 0.47, 0.43, 0.35, and 0.38 were achieved, respectively. LV dysfunction model performance was similar when using a quantitative cutoff of ejection fraction ≤40% and ≤50% (Figure S3). Again, model performance improved when considering all available ECG–echocardiogram pairs ≤2 days apart (Table S9) for patients in the internal cohorts (Figure S4).

During external validation (Figure 3; right), the AI-pECG model achieved AUROCs of 0.86, 0.94, 0.84, and 0.87 for LV composite outcome, LV dysfunction, LV hypertrophy, and LV dilation, respectively; AUPRCs of 0.39, 0.32, 0.25, and 0.33 were achieved, respectively.

Model sensitivity, specificity, NPV, PPV, and percentage predicted negative were subsequently evaluated when setting the following thresholds to achieve 90% sensitivity in the training set: 0.015 (LV dysfunction), 0.019 (LV hypertrophy), 0.04 (LV dilation), and 0.05 (LV composite outcome). Within the Boston Children's Hospital cohorts, sensitivities were slightly lower for each

Table 2. Comparison of Demographic Characteristics, Echocardiogram Characteristics, and Outcomes Stratified by Test Cohort

Characteristics	Boston Children's Hospital		Mount Sinai Hospital, external validation
	Internal testing	Emergency department	
Demographic			
Patients	12631	2830	5088
ECG–echocardiogram pairs	12631	2830	5088
Sex			
Female	5859 (46)	1217 (43)	2696 (53)
Male	6772 (54)	1613 (57)	2376 (47)
Missing	–	–	16 (0.3)
Age at ECG, y	8.8 (2.8, 14.4)	7.7 (1.2, 14.5)	4.3 (0.3, 12.2)
Echocardiogram			
LVEF, %	63.0 (59.0, 66.0)	62.0 (58.0, 66.0)	62.1 (58.6, 65.8)
LVEF, *z* score	−0.1 (−0.9, 0.6)	−0.3 (−1.2, 0.6)	−0.3 (−1.0, 0.5)
LV mass, g	62.7 (32.5, 102.4)	59.9 (25.3, 105.0)	36.9 (13.4, 78.3)
LV mass, *z* score	0.5 (−0.2, 1.4)	0.7 (−0.1, 1.8)	0.0 (−0.8, 1.1)
LV EDV, mL	77.3 (39.9, 121.8)	70.6 (29.5, 118.2)	47.8 (15.8, 97.8)
LV EDV, *z* score	0.6 (−0.2, 1.3)	0.6 (−0.3, 1.6)	0.4 (−0.4, 1.4)
Qualitative outcomes			
Composite LV outcome	567 (4.5)	260 (9.2)	–
LV dysfunction	150 (1.2)	123 (4.3)	–
LV hypertrophy	232 (1.8)	94 (3.3)	–
LV dilation	254 (2.0)	102 (3.6)	–
Quantitative outcomes*			
Composite LV outcome	437/9476 (4.6)	203/1974 (10.0)	280/4602 (6.1)
LV dysfunction	83/9565 (0.9)	87/2036 (4.3)	61/5080 (1.2)
LV hypertrophy	294/9484 (3.1)	128/1980 (6.5)	153/4602 (3.3)
LV dilation	253/9576 (2.6)	95/2036 (4.7)	206/4747 (4.3)

Data presented as n (%) or median (interquartile range). Qualitative outcomes are not available for the external validation site. BPM indicates beats per minute; EDV, end-diastolic volume; EF, ejection fraction; and LV, left ventricle.

*Outcome rates presented as number of events/number of eligible echocardiograms on the basis of *z* score availability.

outcome, ranging from 0.80 (LV hypertrophy) to 0.87 (LV dysfunction and dilation). In contrast, for the Mount Sinai cohort, sensitivities were markedly higher, ranging from 0.92 to 1.00. In correspondence, the NPVs were higher, the percentages predicted negative were lower, and the PPVs were lower in the external validation cohort compared with the Boston Children's cohort (Table 3).

Subgroup Analysis

In a subgroup analysis (Figure 4), the AI-pECG model performance for <1 year of age was lower for predicting LV hypertrophy and higher for predicting LV dilation. For ≥12 years of age, performance was slightly lower for predicting LV dilation. There was slightly better performance for predicting LV dysfunction and dilation in females and for predicting LV hypertrophy in males (Figure 4).

To investigate the influence of age-dependent ECG changes on overall model performance, a sensitivity analysis was performed by comparing overall model performance with model performance when excluding each age group of interest (Figure S5). Composite outcome model performance was insensitive to specific age groups (Figure S5). LV hypertrophy model performance increased slightly when excluding <1 year of age. LV dilation model performance decreased marginally when excluding <1 year of age and increased marginally when excluding ≥12 years of age.

Model Explainability

In an attempt to gain model interpretability, saliency mapping and median waveform analysis were performed.

As shown in Figure 5, the most salient features of an ECG to predict LV dysfunction include the lateral precordial (V4–V6) QRS complexes, as well as the lateral precordial (V4–V6) T waves. High-risk features to predict LV dysfunction include inverted T waves in the

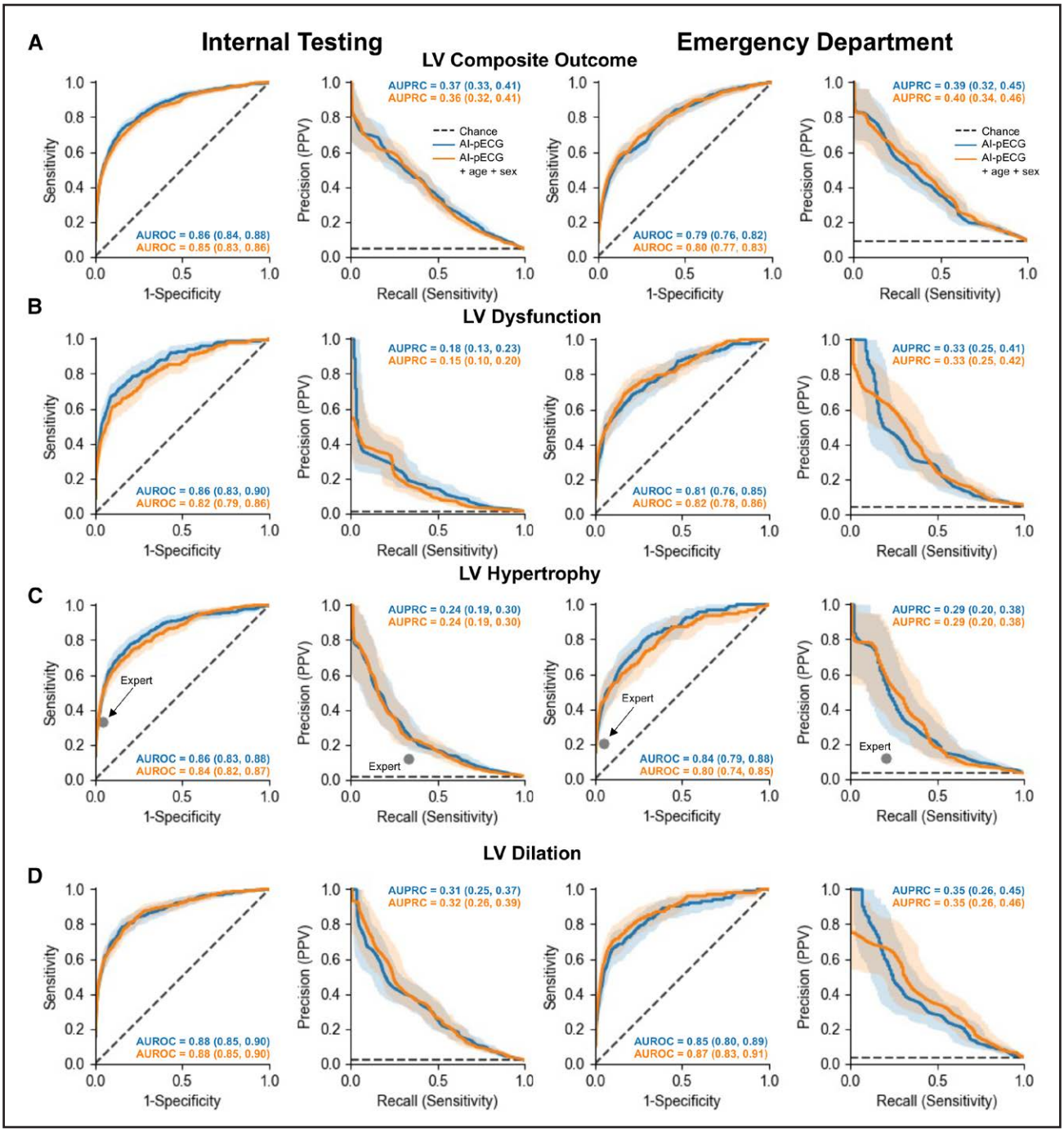

Figure 2. Pediatric ECG-based deep learning model performance at qualitative outcome cutoffs.
Performance of the artificial intelligence–enhanced pediatric ECG (blue) and artificial intelligence–enhanced pediatric ECG with age and sex (orange) models evaluated using the internal test (**left**) and emergency department (**right**) cohorts with receiver operating characteristic and precision-recall curves for the left ventricular (LV) composite (**A**), LV dysfunction (**B**), LV hypertrophy (**C**), and LV dilation (**D**) qualitative outcomes. In **C**, the gray dot represents the benchmark of pediatric cardiologist expert ECG-based diagnosis of LV hypertrophy. Area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) metric values for each model and outcome are inset. The dotted line represents chance; 95% CIs are shown using bootstrapping. PPV indicates positive predictive value.

lateral precordial leads (V4–V6). In V1 and V2, S waves are also found as salient, with high-risk features including deep S waves. To predict LV hypertrophy, the most salient features include precordial QRS complexes. High-risk features to predict LV hypertrophy include deep S waves in V1 and V2. In limb lead I, the QRS complex was also salient, with high-risk features including a high-amplitude R wave (Figure 5). For LV dilation, the most salient features include lateral precordial (V4 through V6) QRS complexes. High-risk features to

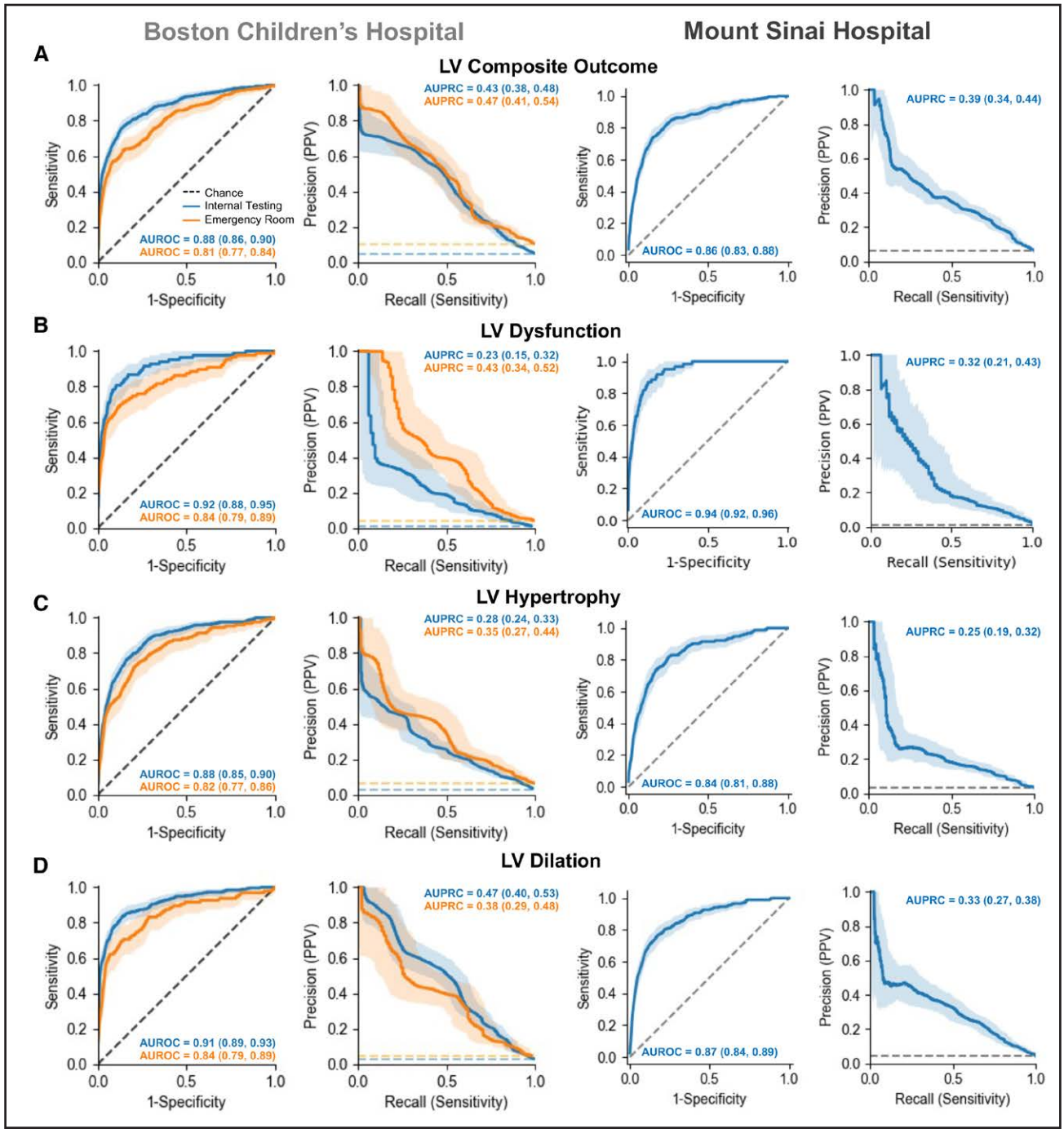

Figure 3. External validation of the model to predict quantitative cutoffs of left ventricular function and remodeling. Performance of the artificial intelligence–enhanced pediatric ECG algorithm evaluated in the internal (**left**) and external (**right**) cohorts using receiver operating characteristic and precision-recall curves. **A**, Left ventricular (LV) composite outcome (LV ejection fraction *z* score ≤ −4 or LV mass *z* score ≥ +4 or LV end-diastolic volume *z* score ≥ +4). **B**, LV dysfunction (LV ejection fraction *z* score ≤ −4); **C**, LV hypertrophy (LV mass *z* score ≥ +4). **D**, LV dilation (LV end-diastolic volume *z* score ≥ +4). Within the internal group, internal testing (blue) and emergency department (orange) performance is shown. Area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) metric values for each model and outcome are inset. The dotted line represents chance (the blue and orange lines represent the prevalence of each respective group); 95% CIs are shown using bootstrapping.

predict LV dilation include high-amplitude R waves in V4 through V6 (Figure 5).

When stratifying by age (Figures S6 through S8), distinct age-dependent high-risk features were identified. For LV dysfunction (Figure S6), tall R waves in lateral precordial leads were high-risk features for <1 year but not age ≥12 years of age. Conventional scoring system components to predict LV hypertrophy[20,21] were detected as high-risk features in ≥12 years of age, and less so in <1 year of age (Figure S7). For LV dilation (Figure S8),

Table 3. Summary of Internal and External Validation Model Performance at a Select Threshold

Outcomes	Boston Children's Hospital		Mount Sinai Hospital, external validation
	Internal testing	Emergency department	
LV composite outcome			
Prevalence, %	4.6	10.0	6.1
Sensitivity	0.84 (0.80–0.87)	0.84 (0.79–0.89)	0.96 (0.93–0.98)
Specificity	0.73 (0.73–0.74)	0.57 (0.55–0.60)	0.34 (0.33–0.36)
NPV, %	99.0 (98.7–99.2)	96.8 (95.9–97.8)	99.2 (98.7–99.6)
PPV, %	13.2 (12.6–13.9)	18.4 (17.2–19.7)	8.6 (8.4–8.9)
Predicted negative, %	70.8 (70.0–71.6)	53.3 (51.1–55.4)	32.6 (31.2–33.7)
LV dysfunction			
Prevalence, %	0.9	4.3	1.2
Sensitivity	0.87 (0.80–0.93)	0.85 (0.78–0.92)	1.00 (1.00–1.00)
Specificity	0.79 (0.79–0.80)	0.58 (0.55–0.60)	0.58 (0.57–0.59)
NPV, %	99.9 (99.8–99.9)	98.9 (98.3–99.4)	100 (100–100)
PPV, %	3.4 (3.1–3.7)	8.2 (7.5–9.0)	2.8 (2.7–2.9)
Predicted negative, %	78.2 (77.3–79.0)	55.8 (53.7–57.9)	57.4 (56.0–58.7)
LV hypertrophy			
Prevalence, %	3.1	6.5	3.3
Sensitivity	0.80 (0.76–0.85)	0.81 (0.74–0.88)	0.92 (0.88–0.96)
Specificity	0.78 (0.77–0.79)	0.67 (0.65–0.69)	0.40 (0.39–0.42)
NPV, %	99.2 (99.0–99.4)	98.1 (97.4–98.7)	99.3 (98.9–99.7)
PPV, %	10.6 (9.9–11.3)	14.6 (13.2–15.9)	5.0 (4.8–5.3)
Predicted negative, %	76.5 (75.6–77.3)	64.1 (62.0–66.1)	39.2 (37.8–40.6)
LV dilation			
Prevalence, %	2.6	4.7	4.3
Sensitivity	0.87 (0.83–0.91)	0.85 (0.78–0.92)	0.95 (0.91–0.97)
Specificity	0.77 (0.76–0.78)	0.63 (0.61–0.65)	0.40 (0.38–0.41)
NPV, %	99.5 (99.4–99.7)	98.9 (98.3–99.4)	99.4 (99.0–99.7)
PPV, %	9.3 (8.8–9.8)	10.2 (9.2–11.1)	6.7 (6.4–6.9)
Predicted negative, %	75.4 (74.6–76.2)	60.7 (58.5–62.9)	38.3 (37.0–39.5)

Data presented as median (95% CI). Predicted negative indicates the fraction of ECGs predicting negative echocardiogram findings at the given threshold. LV indicates left ventricular; NPV, negative predictive value; and PPV, positive predictive value.

tall R waves in lateral precordial leads were high-risk features for <1 year and less so in ≥12 years of age.

The saliency map for the composite outcome (Figure S9) appears to merge features from each of the individual outcomes of interest.

DISCUSSION

In this work, a technological gap in the application of ECG-based deep learning to a pediatric cohort for prediction of LV dysfunction or remodeling was addressed. The convolutional neural network trained on nearly 100 000 ECG–echocardiogram pairs with human expert–classified greater than mild LV dysfunction, hypertrophy, and dilation performed well on an independent internal test set of >10 000 patients, on nearly 3000 patients in the ED, as well as during external validation on an outside health care system with >5000 patients. During internal and external testing, a high NPV was achieved. Saliency mapping and median waveform analysis provide physiologically relevant insights into ECG waveforms influencing model prediction of pediatric LV dysfunction or remodeling, allowing for visual comparison with existing algorithms for ECG interpretation. Altogether, these findings demonstrate the promise of AI-pECG to inexpensively screen for or diagnose LV dysfunction or remodeling in children, which may facilitate improved access to care and help prioritize patients for further studies or interventions.

Clinical Significance and Implications

Artificial intelligence has had a transformative effect on adult cardiovascular medicine,[2] but its potential for implementation in pediatric cardiology is only beginning to be

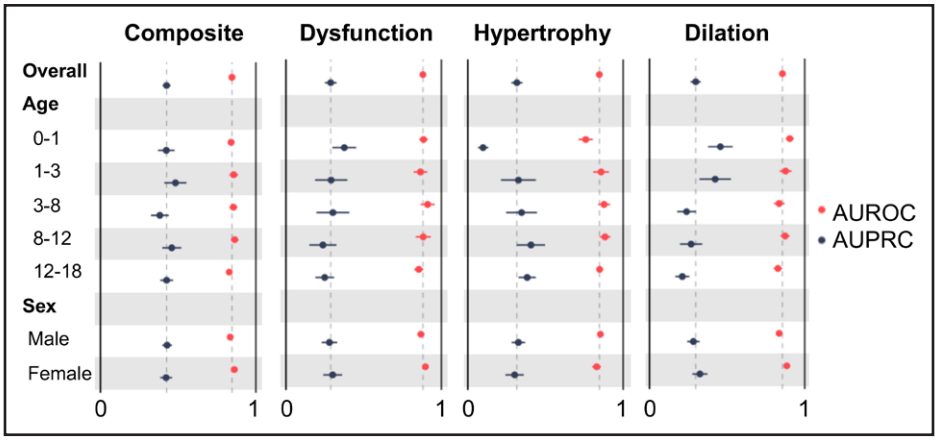

Figure 4. Model performance in age and sex subgroups.
Forest plot showing artificial intelligence–enhanced pediatric ECG area under the receiver operating characteristic curve (AUROC; red) and area under the precision-recall curve (AUPRC; black) performance when stratifying by age (<1, 1 ≤ age <3, 3 ≤ age <8, 8 ≤ age <12, age ≥12) and sex for the following outcomes: left ventricular composite, left ventricular dysfunction, left ventricular hypertrophy, and left ventricular dilation; 95% CIs are shown using bootstrapping.

appreciated.[34,35] To date, AI has been used in pediatric cardiology primarily for image-based deep learning applications.[36–39] Analysis of ECG waveforms provides a rapid, easy-to-implement, and cost-effective application for artificial intelligence. Its use in adults has been wide-ranging, including prediction of ventricular dysfunction,[3–7] ventricular hypertrophy,[8–10] ventricular dilation,[9,11] atrial fibrillation and other arrhythmias,[17,26,40,41] and age,[42,43] sex,[42] and time to death.[8,43] Our findings provide proof-of-concept evidence that similar ECG applications can be explored in children and suggest that deep learning may also be applicable to other data streams (eg, wearable biosensor data) that could aid in predicting outcomes for children[44] similar to what has been performed in adults.[45]

As a direct case example of clinical usefulness, ECG–echocardiogram pairs from the ED were considered, a setting in which AI-pECG could be of substantial clinical and economic value. The economic burden includes misdiagnosis leading to unnecessary referrals and associated costs of echocardiograms, as well as missed diagnoses resulting in adverse clinical outcomes. AI-pECG predictions could help guide an emergency physician's need to consult a pediatric cardiologist. The benchmark herein of pediatric cardiologists applying conventional LV hypertrophy criteria is on par with previous literature[46]; given that model performance is superior, it may also help guide pediatric cardiologists on whether to obtain an echocardiogram for a child without major congenital heart disease. This democratization of specialty expertise is likely to be particularly valuable for hospitals with low pediatric volumes or limited pediatric cardiology experience.[47] As a thought example, using the composite LV outcome, the model has the capacity to achieve an NPV of 99% during internal testing, 97% in the ED, and 99% during external validation, with a potential to reduce echocardiograms obtained by 71%, 53%, and 33%, respectively (Table 3). NPVs and clinical effectiveness to reduce echocardiograms improve further when assessing individual outcomes (Table 3).

In addition to its clinical usefulness, the AI-pECG model only requires the inexpensive and rapidly generated ECG waveform data. The robust performance of the model suggests it is at least partially resistant to noise generated from obtaining data, as well as age dependence on overall model performance. The model uses only a single modality (ie, independent of age or sex) rather than a complex clinical scoring system that would require user interaction and is susceptible to input error.

The saliency mapping also provides insight for clinicians to detect LV dysfunction or remodeling. It is reassuring that features identified by this model are relevant to findings currently used by clinicians to identify LV hypertrophy and dysfunction. First, saliency mapping and median waveform analysis identified lateral precordial leads as influential for LV dysfunction, with high-risk features of inverted T waves in V5 and V6, which has been previously considered pathological.[48] Second, LV hypertrophy explainability was focused on precordial QRS complexes (eg, deep S waves in V1 and V2 and tall R waves in V6), in keeping with previously established scoring systems.[20,21] Third, tall R waves in lateral precordial leads (V4 through V6) were high-risk features for LV dilation. Fourth, there were distinct age-dependent characteristics in saliency maps and median waveforms; conventional scoring system components to predict LV hypertrophy[20,21] were detected as high-risk features in ≥12 years of age, and less so in <1 year of age (Figure S7).

Use of Both Qualitative and Quantitative Cutoffs

The primary objective was to create an ECG-based deep learning model on the basis of human expert interpretation

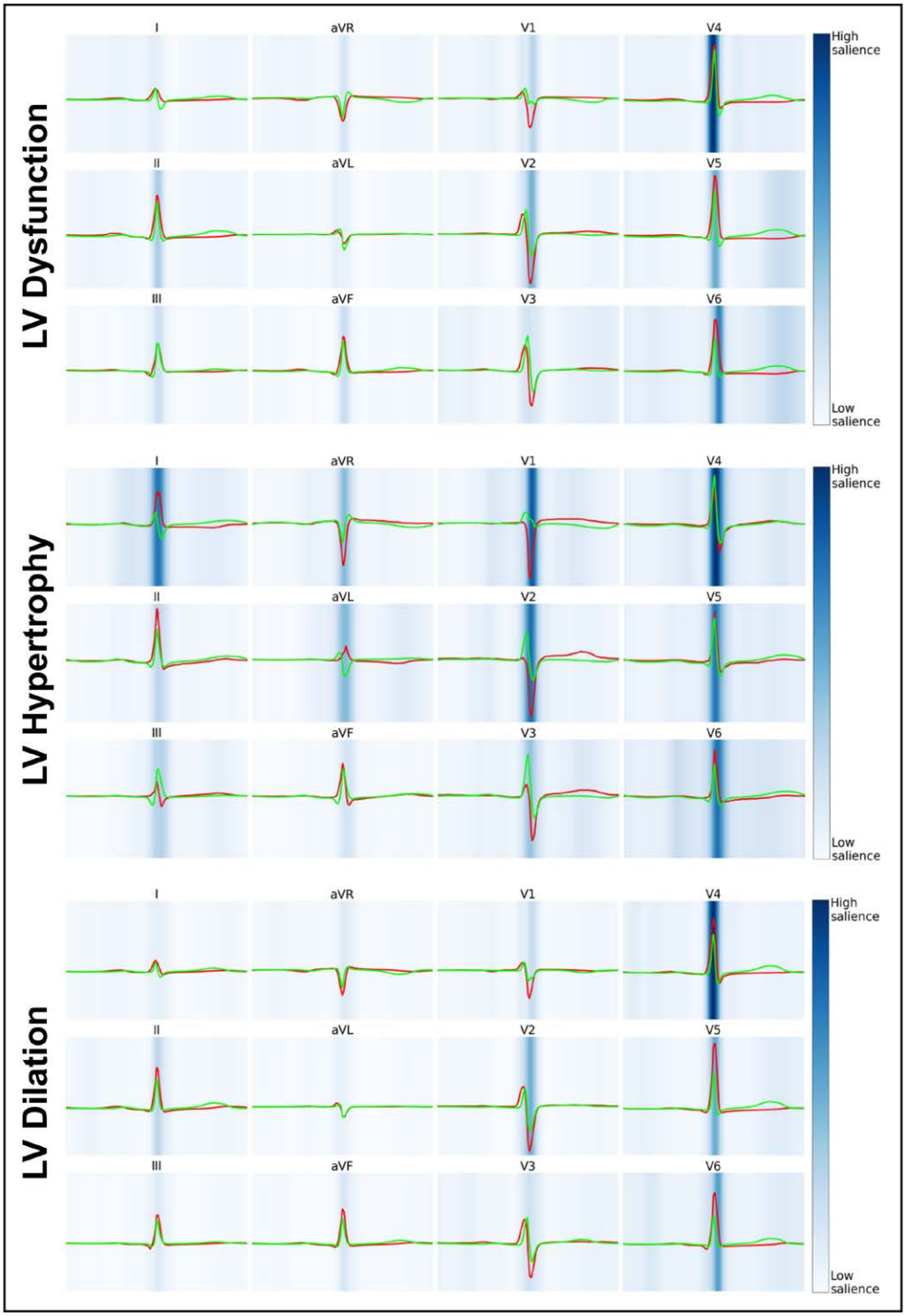

Figure 5. Explainability of artificial intelligence–enhanced pediatric ECG predictions.
Visualization of median waveforms generated in each lead using ECGs from the 100 highest (red) and 100 lowest (green) artificial intelligence–enhanced pediatric ECG predictions of left ventricular dysfunction, hypertrophy, and dilation. Saliency mapping demarcates regions of the ECG waveform having the greatest (dark blue) and least (light blue) influence on each outcome. Saliency was averaged across the 100 highest predicted ECGs for each outcome.

of echocardiograms of LV function and remodeling. This approach was taken given that ≈25% of patients were given only qualitative measures of LV function, which would have substantially reduced the training data set. Human expert opinion incorporates multiple clinical data points that aid in decision making, which was also of interest when training a model. Furthermore, most emergency medicine physicians will attempt cardiac point of care ultrasound and report qualitative outcomes on function. In contrast, from a pediatric cardiology perspective,

z scores are commonly reported for LV ejection fraction, LV mass, and LV volume. However, both qualitative[49] and quantitative[39,50] cutoffs have limitations. Given the goal to externally validate the model for eventual multicenter use and validation, PHN[22] *z* scores were incorporated. In practical clinical use, it would be of value to have a model that is agnostic to cutoff method used.

Herein, PHN *z* scores led to higher internal median LV mass and volume *z* scores. Whereas these findings are consistent with previous work,[33] they, nonetheless, may contribute to the variation in sensitivity and percentage predicted negative across institutions (Table 3), underscoring the need for improved standardization of echocardiogram measurements, multicenter collaboration (eg, federated learning) for model training and testing, or consideration of institution-specific AI-pECG thresholds.

Limitations and Future Directions

There are several limitations of this work. First, human expert qualitative classification of outcomes on the basis of echocardiogram is subject to interrater and intrarater variability, as highlighted previously. This was addressed by demonstrating effective model performance using quantitative cutoffs across multiple institutions (Figure 3). Second, whereas AUROC and AUPRC performance is similar between the internal and external cohorts (demonstrating generalizable discrimination) and the main objective was to screen for (ie, rule out) pathology, model specificity and PPV were limited (demonstrating limited calibration, which may be attributed to the aforementioned measurement variability). Only 1 example of thresholding was used in evaluation of model performance, because further consideration is required to weigh the effect of resultant false-positives and false-negatives, as well as optimally set thresholds across institutions. To this end, multicenter external validation to further refine thresholds for clinical implementation is warranted. In addition, multicenter collaboration may help improve training and testing sample sizes, which may further improve performance (eg, AUPRC) and generalizability. Third, it is conceivable that echocardiogram (or ECG) findings may change within the time frame of the paired ECG (or echocardiogram), which was minimized by selecting the closest ECG–echocardiogram pair during testing. Fourth, other quantitative cutoffs could have been implemented, but *z* score–based cutoffs were used given their ubiquitous use in pediatrics. Fifth, only one model architecture was attempted; others conceivably may lead to improved model performance. Sixth, these findings are limited to patients without major congenital heart disease.

Future work includes application to other pediatric cardiology outcomes of interest (eg, pediatric arrhythmia detection, mortality), other pediatric populations (eg, major congenital heart disease), multicenter collaboration (for further model and threshold selection refinement), and prospective trials (to determine how to properly implement such tools to support clinical decision making). Methodologies need to be developed further to clarify the mechanistic insights afforded by models in relationship to ECG principles and current scoring systems.

Conclusions

These findings demonstrate the promise of AI-pECG to inexpensively screen for or diagnose LV dysfunction or remodeling in children. This tool may facilitate prioritization of patients for future interventions or studies, provide meaningful insight into novel ECG waveforms suggestive of LV dysfunction or remodeling, and potentially reduce disparities by improving access to care. Future multicenter collaboration, prospective trials, and application to congenital heart disease and pediatric arrhythmias are warranted.

ARTICLE INFORMATION

Received October 26, 2023; accepted December 20, 2023.

Affiliations

Department of Cardiology (J.M., S.J.G., T.G., A.D., M.E.A., J.K.T.), Computational Health Informatics Program (W.G.L.C.), and Department of Medicine, Division of Emergency Medicine (R.M.), Boston Children's Hospital, MA. Department of Pediatrics (J.M., W.G.L.C., S.J.G., T.G., A.D., M.E.A., J.K.T.), Harvard Medical School (R.M.), Boston, MA. The Charles Bronfman Institute of Personalized Medicine (A.V., G.N.N., S.Q.D.) and Division of Pediatric Cardiology, Department of Pediatrics (S.Q.D.), Icahn School of Medicine at Mount Sinai, New York, NY.

Acknowledgments

The High-Performance Computing Resources Cluster Enkefalos 2 (E2), Boston Children's Hospital, was used in conducting the research reported in this article.

Sources of Funding

Funding support was received from the Thrasher Research Fund Early Career Award (Dr Mayourian), the Boston Children's Hospital Electrophysiology Research Education Fund (Drs Mayourian and Triedman), and National Institutes of Health grant R00-LM012926 from the National Library of Medicine (Dr La Cava).

Disclosures

Dr Nadkarni reports consultancy agreements with AstraZeneca, BioVie, GLG Consulting, Pensieve Health, Reata, Renalytix, Siemens Healthineers, and Variant Bio; research funding from Goldfinch Bio and Renalytix; honoraria from AstraZeneca, BioVie, Lexicon, Daiichi Sankyo, Menarini Health, and Reata; patents or royalties with Renalytix; equity and stock options in Pensieve Health and Renalytix as a scientific cofounder; equity in Verici Dx; financial compensation as a scientific board member and advisor to Renalytix; advisory board of Neurona Health; and advisory or leadership role for Pensieve Health and Renalytix, none of which played a role in the design or conduct of this study. The other authors report no disclosures.

Supplemental Material

Tables S1–S9

Figures S1–S9

REFERENCES

1. Saarel EV, Granger S, Kaltman JR, Minich LL, Tristani-Firouzi M, Kim JJ, Ash K, Tsao SS, Berul CI, Stephenson EA, et al; Pediatric Heart Network Investigators. Electrocardiograms in healthy North American children in the digital age. *Circ Arrhythm Electrophysiol.* 2018;11:e005808. doi: 10.1161/CIRCEP.117.005808

2. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. 2021;18:465–478. doi: 10.1038/s41569-020-00503-2
3. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25:70–74. doi: 10.1038/s41591-018-0240-2
4. Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA, Carter RE, Shah ND, Friedman PA, Noseworthy PA. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J Cardiovasc Electrophysiol*. 2019;30:668–674. doi: 10.1111/jce.13889
5. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM, Bernard ME, Rosas SL, Akfaly A, Misra A, 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. doi: 10.1038/s41591-021-01335-4
6. Adedinsewo D, Carter RE, Attia Z, Johnson P, Kashou AH, Dugan JL, Albus M, Sheele JM, Bellolio F, Friedman PA, et al. Artificial intelligence-enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea. *Circ Arrhythm Electrophysiol*. 2020;13:e008437. doi: 10.1161/CIRCEP.120.008437
7. Vaid A, Johnson KW, Badgeley MA, Somani SS, Bicak M, Landi I, Russak A, Zhao S, Levin MA, Freeman RS, et al. Using deep-learning algorithms to simultaneously identify right and left ventricular dysfunction from the electrocardiogram. *JACC Cardiovasc Imaging*. 2022;15:395–410. doi: 10.1016/j.jcmg.2021.08.004
8. Liu CM, Hsieh ME, Hu YF, Wei TY, Wu IC, Chen PF, Lin YJ, Higa S, Yagi N, Chen SA, et al. Artificial intelligence-enabled model for early detection of left ventricular hypertrophy and mortality prediction in young to middle-aged adults. *Circ Cardiovasc Qual Outcomes*. 2022;15:e008360. doi: 10.1161/CIRCOUTCOMES.121.008360
9. Kokubo T, Kodera S, Sawano S, Katsushika S, Nakamoto M, Takeuchi H, Kimura N, Shinohara H, Matsuoka R, Nakanishi K, et al. Automatic detection of left ventricular dilatation and hypertrophy from electrocardiograms using deep learning. *Int Heart J*. 2022;63:939–947. doi: 10.1536/ihj.22-132
10. Ko WY, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, Demuth SJ, Ackerman MJ, Gersh BJ, Arruda-Olson AM, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol*. 2020;75:722–733. doi: 10.1016/j.jacc.2019.12.030
11. Shrivastava S, Cohen-Shelly M, Attia ZI, Rosenbaum AN, Wang L, Giudicessi JR, Redfield M, Bailey K, Lopez-Jimenez F, Lin G, et al. Artificial intelligence-enabled electrocardiography to screen patients with dilated cardiomyopathy. *Am J Cardiol*. 2021;155:121–127. doi: 10.1016/j.amjcard.2021.06.021
12. Dickinson DF. The normal ECG in childhood and adolescence. *Heart*. 2005;91:1626–1630. doi: 10.1136/hrt.2004.057307
13. Siontis KC, Liu K, Bos JM, Attia ZI, Cohen-Shelly M, Arruda-Olson AM, Zanjirani Farahani N, Friedman PA, Noseworthy PA, Ackerman MJ. Detection of hypertrophic cardiomyopathy by an artificial intelligence electrocardiogram in children and adolescents. *Int J Cardiol*. 2021;340:42–47. doi: 10.1016/j.ijcard.2021.08.026
14. Mori H, Inai K, Sugiyama H, Muragaki Y. Diagnosing atrial septal defect from electrocardiogram with deep learning. *Pediatr Cardiol*. 2021;42:1379–1387. doi: 10.1007/s00246-021-02622-0
15. Jacobs JP, Franklin RCG, Beland MJ, Spicer DE, Colan SD, Walters HL 3rd, Bailliard F, Houyel L, St Louis JD, Lopez L, et al. Nomenclature for pediatric and congenital cardiac care: unification of clinical and administrative nomenclature: the 2021 International Paediatric and Congenital Cardiac Code (IPCCC) and the Eleventh Revision of the International Classification of Diseases (ICD-11). *World J Pediatr Congenit Heart Surg*. 2021;12:E1–E18. doi: 10.1177/21501351211032919
16. 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*. Springer London; 2015:163–169.
17. Ribeiro AH, Ribeiro MH, Paixao GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira W Jr, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11:1760. doi: 10.1038/s41467-020-15432-4
18. Einthoven W. Weiteres über das Elektrokardiogramm. *Pflüger Arch Gesammte Physiol Menschen Thiere*. 1908;122:517–584. doi: 10.1007/bf01677829
19. Goldberger E. A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads. *Am Heart J*. 1942;23:483–492. doi: 10.1016/s0002-8703(42)90293-x
20. Sokolow M, Lyon TP. The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads. *Am Heart J*. 1949;37:161–186. doi: 10.1016/0002-8703(49)90562-1
21. Okin PM, Roman MJ, Devereux RB, Kligfield P. Electrocardiographic identification of increased left ventricular mass by simple voltage-duration products. *J Am Coll Cardiol*. 1995;25:417–423. doi: 10.1016/0735-1097(94)00371-v
22. Lopez L, Colan S, Stylianou M, Granger S, Trachtenberg F, Frommelt P, Pearson G, Camarda J, Cnota J, Cohen M, et al; Pediatric Heart Network Investigators. Relationship of echocardiographic *Z* scores adjusted for body surface area to age, sex, race, and ethnicity: the Pediatric Heart Network Normal Echocardiogram Database. *Circ Cardiovasc Imaging*. 2017;10:e006979. doi: 10.1161/CIRCIMAGING.117.006979
23. Lai WW, Mertens L, Cohen M, Geva T. *Echocardiography in Pediatric and Congenital Heart Disease: From Fetus to Adult*. Wiley-Blackwell; 2021.
24. Gustafsson S, Gedon D, Lampa E, Ribeiro AH, Holzmann MJ, Schon TB, Sundstrom J. Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients. *Sci Rep*. 2022;12:19615. doi: 10.1038/s41598-022-24254-x
25. Ulloa-Cerna AE, Jing L, Pfeifer JM, Raghunath S, Ruhl JA, Rocha DB, Leader JB, Zimmerman N, Lee G, Steinhubl SR, et al. rECHOmmend: an ECG-based machine learning approach for identifying patients at increased risk of undiagnosed structural heart disease detectable by echocardiography. *Circulation*. 2022;146:36–47. doi: 10.1161/CIRCULATIONAHA.121.057869
26. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, Harrington LX, Wang X, Al-Alusi MA, Sarma G, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation*. 2022;145:122–133. doi: 10.1161/CIRCULATIONAHA.121.057480
27. Sangha V, Nargesi AA, Dhingra LS, Khunte A, Mortazavi BJ, Ribeiro AH, Banina E, Adeola O, Garg N, Brandt CA, et al. Detection of left ventricular systolic dysfunction from electrocardiographic images. *Circulation*. 2023;148:765–777. doi: 10.1161/CIRCULATIONAHA.122.062646
28. Rijnbeek PR, Witsenburg M, Schrama E, Hess J, Kors JA. New normal limits for the paediatric electrocardiogram. *Eur Heart J*. 2001;22:702–711. doi: 10.1053/euhj.2000.2399
29. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA. NeuroKit2: a Python toolbox for neurophysiological signal processing. *Behav Res Methods*. 2021;53:1689–1696. doi: 10.3758/s13428-020-01516-y
30. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global understanding with explainable ai for trees. *Nat Mach Intell*. 2020;2:56–67. doi: 10.1038/s42256-019-0138-9
31. Python Software Foundation. Python: a dynamic, open source programming language. Python Software Foundation; 2015.
32. R Foundation for Statistical Computing. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing; 2019.
33. Lopez L, Frommelt PC, Colan SD, Trachtenberg FL, Gongwer R, Stylianou M, Bhat A, Burns KM, Cohen MS, Dragulescu A, et al; Pediatric Heart Network Investigators. Pediatric Heart Network echocardiographic *Z* scores: comparison with other published models. *J Am Soc Echocardiogr*. 2021;34:185–192. doi: 10.1016/j.echocardiogram.2020.09.019
34. Gaffar S, Gearhart AS, Chang AC. The next frontier in pediatric cardiology: artificial intelligence. *Pediatr Clin North Am*. 2020;67:995–1009. doi: 10.1016/j.pcl.2020.06.010
35. Jone P-N, Gearhart A, Lei H, Xing F, Nahar J, Lopez-Jimenez F, Diller G-P, Marelli A, Wilson L, Saidi A, et al. Artificial intelligence in congenital heart disease. *JACC Adv*. 2022;1:100153. doi: 10.1016/j.jacadv.2022.100153
36. Sethi Y, Patel N, Kaka N, Desai A, Kaiwan O, Sheth M, Sharma R, Huang H, Chopra H, Khandaker MU, et al. Artificial intelligence in pediatric cardiology: a scoping review. *J Clin Med*. 2022;11. doi: 10.3390/jcm11237072
37. Arnaout R, Curran L, Zhao Y, Levine JC, Chinn E, Moon-Grady AJ. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nat Med*. 2021;27:882–891. doi: 10.1038/s41591-021-01342-5
38. Gearhart A, Goto S, Deo RC, Powell AJ. An automated view classification model for pediatric echocardiography using artificial intelligence. *J Am Soc Echocardiogr*. 2022;35:1238–1246. doi: 10.1016/j.echocardiogram.2022.08.009
39. Reddy CD, Lopez L, Ouyang D, Zou JY, He B. Video-based deep learning for automated assessment of left ventricular ejection fraction in pediatric patients. *J Am Soc Echocardiogr*. 2023;36:482–489. doi: 10.1016/j.echocardiogram.2023.01.015
40. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, et al. An artificial

intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. 2019;394:861–867. doi: 10.1016/S0140-6736(19)31721-0
41. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25:65–69. doi: 10.1038/s41591-018-0268-3
42. Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, Satam G, Pellikka PA, Munger TM, Asirvatham SJ, Scott CG, et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythm Electrophysiol*. 2019;12:e007284. doi: 10.1161/CIRCEP.119.007284
43. Lima EM, Ribeiro AH, Paixao GMM, Ribeiro MH, Pinto-Filho MM, Gomes PR, Oliveira DM, Sabino EC, Duncan BB, Giatti L, et al. Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nat Commun*. 2021;12:5117. doi: 10.1038/s41467-021-25351-7
44. Tandon A, Nguyen HH, Avula S, Seshadri DR, Patel A, Fares M, Baloglu O, Amdani S, Jafari R, Inan OT, et al. Wearable biosensors in congenital heart disease: needs to advance the field. *JACC Adv*. 2023;2:100267. doi: 10.1016/j.jacadv.2023.100267
45. Attia ZI, Harmon DM, Dugan J, Manka L, Lopez-Jimenez F, Lerman A, Siontis KC, Noseworthy PA, Yao X, Klavetter EW, et al. Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. *Nat Med*. 2022;28:2497–2503. doi: 10.1038/s41591-022-02053-1
46. Rivenes SM, Colan SD, Easley KA, Kaplan S, Jenkins KJ, Khan MN, Lai WW, Lipshultz SE, Moodie DS, Starc TJ, et al; Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted HIV Infection Study Group. Usefulness of the pediatric electrocardiogram in detecting left ventricular hypertrophy: results from the Prospective Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted HIV Infection (P2C2 HIV) multicenter study. *Am Heart J*. 2003;145:716–723. doi: 10.1067/mhj.2003.15
47. Morris SA, Lopez KN. Deep learning for detecting congenital heart disease in the fetus. *Nat Med*. 2021;27:764–765. doi: 10.1038/s41591-021-01354-1
48. D'Ascenzi F, Anselmi F, Berti B, Capitani E, Chiti C, Franchini A, Graziano F, Nistri S, Focardi M, Capitani M, et al. Prevalence and significance of T-wave inversion in children practicing sport: a prospective, 4-year follow-up study. *Int J Cardiol*. 2019;279:100–104. doi: 10.1016/j.ijcard.2018.09.069
49. Lopez L, Colan SD, Frommelt PC, Ensing GJ, Kendall K, Younoszai AK, Lai WW, Geva T. 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. doi: 10.1016/j.echocardiogram.2010.03.019
50. Frommelt PC, Minich LL, Trachtenberg FL, Altmann K, Camarda J, Cohen MS, Colan SD, Dragulescu A, Frommelt MA, Johnson TR, et al; Pediatric Heart Network Investigators. Challenges with left ventricular functional parameters: the pediatric heart network normal echocardiogram database. *J Am Soc Echocardiogr*. 2019;32:1331–1338.e1. doi: 10.1016/j.echocardiogram.2019.05.025