

Estimation of Left Ventricular Systolic Function in Pediatric and Congenital Heart Disease from Serial Electrocardiograms

Platon Lukyanenko, PhD^{a,c}, Sunil J. Ghelani, MD^{b,c}, John K. Triedman, MD^{b,c},
Joshua Mayourian, MD PhD^{b,c,+}, William G. La Cava, PhD^{a,c,+*}

^aComputational Health Informatics Program, Boston Children's Hospital

^bDepartment of Cardiology, Boston Children's Hospital

^cDepartment of Pediatrics, Harvard Medical School

+ These authors contributed equally

* Correspondence: william.lacava@childrens.harvard.edu. 617-355-6000; Lab page: <https://cavalab.org>

Abstract

Estimating left ventricular ejection fraction (LVEF) from electrocardiograms (ECGs) is a useful task enabled by artificial intelligence (AI-ECG). Prior work focuses on predicting LVEF from single ECGs. Here, we investigate whether a patient's history of ECGs can improve LVEF prediction. We study a longitudinal cohort from Boston Children's Hospital, deriving LVEF from echocardiograms conducted within 2 days of ECGs ($n=178,495$, n patients: 70,226, median age: 10.6). We propose a sequential AI-ECG approach using convolutional layers to represent single ECGs and a sequential neural network to reason over longitudinal ECGs. We build and test several sequential architectures. For predictions with at least 5 previous ECGs, sequential AI-ECG improved median AUROC (IQR) by 3.4 points (1.4, 4.4). When predicting the LVEF value, sequential AI-ECG improves Pearson R by 0.08 (0.02, 0.13). Results suggest that patients' longitudinal ECG history contains valuable information for improving AI-ECG risk stratification beyond current snapshot-based models.

Introduction

Electrocardiography measures the electric activity of the heart and is a staple of cardiac diagnostics. AI-ECG refers to the application of artificial intelligence (AI) to electrocardiograms (ECG). Large public ECG datasets are now available (Code-15¹, MIMIC-IV²), and recent AI-ECG models train on millions of ECGs^{3,4,5} to predict targets such as physiological measures or patient diagnoses^{6,7,8}. While most work and datasets have focused on adult cohorts, here we focus on a primarily pediatric and congenital heart disease-enriched dataset from Boston Children's Hospital (BCH), many of whom require life-long care management.

One promising application of AI-ECG is predicting measures that generally require more expensive and resource-intensive modalities, such as echocardiograms⁹. By marking potentially abnormal measures, patients can be prioritized for further diagnostic steps. In this work, we predict left ventricular ejection fraction (LVEF), a central echocardiography measure of LV systolic function and target of prior AI-ECG work^{10,11}. Low LVEF is a strong prognostic factor in adults and children alike, providing clues toward heart failure and the need for guideline-directed medical therapy. LVEF trajectories are therefore central to longitudinal patient health monitoring. While past work has estimated LVEF from ECG in adult¹⁰ and pediatric¹¹ cohorts using Convolutional Neural Networks (CNNs), predicting LVEF from serial ECGs remains unexplored. More broadly, AI methodology using sequential ECGs is limited to recent work demonstrating serial ECGs can improve prediction of future atrial fibrillation onset.¹²

The serial ECG approach is particularly compelling in pediatrics and congenital heart disease for several reasons. First, ECG morphology rapidly evolves as a child grows from neonatal to adolescent stages, and after catheter/surgical intervention. Serial ECGs may therefore allow the model to distinguish normal developmental changes from early disease signals. Second, pediatric ECG datasets are far smaller and more heterogeneous than adult counterparts due to the rarity and wide variety of congenital heart lesions. Incorporating longitudinal data in each patient’s personal trajectory reframes the learning problem: instead of asking the model to detect left ventricular systolic dysfunction across a very heterogeneous range of cardiac lesions, we allow it to focus on a simpler, more physiologically grounded question: “What changes in this patient’s ECG over time reflect ventricular dysfunction?” This shift may reduce the burden of inter-patient heterogeneity and improve signal-to-noise ratio. Third, sequential modeling may help differentiate transient perturbations (e.g., tachycardia from fever, dehydration, or medications) from persistent pathologic changes that reflect true myocardial dysfunction. Finally, sequential ECG models offer the potential to generate personalized risk trajectories with higher model performance, which may enable earlier identification of deterioration and inform timing of follow-up imaging/surveillance.

The goal of this study is to investigate whether considering sequences of ECGs can improve AI-ECG prediction of LVEF. We address this by first building non-sequential models classifying LVEF at several clinically relevant thresholds. Sequential models are then built upon the resultant snapshot ECG representations, allowing them to consider all prior patient ECGs when estimating LVEF. We report a detailed sweep over a large set of sequential modeling approaches, including models trained end-to-end on entire ECG histories. Finally, after model selection, we report measures on a held-out test set for LVEF classification and regression, comparing sequential and non-sequential approaches to robustly estimate performance differences.

Methods

Data processing.

ECGs are first adjusted to a 7-second, 12-channel signal sampled at 400Hz, padded with 48 zeros at the start and end, standardizing to the format used by Code-15¹³. ECGs, as in past work¹¹, are split randomly 56/14/30 into train/validation/test sets by patient, fixing the test set across random seeds, without considering LVEF distribution. ECG signals are z-score normalized per channel based on the training set. BCH LVEF measures were acquired from an institutional data archiving system and are calculated using the bullet (5/6 area-length) method.

Modeling Approach

ECG sequences are modeled as in (Figure 1). There are two main pieces: a ‘non-sequential model’ which handles ECG at a single time point, and a ‘sequential model’ that uses the non-sequential model representations across time points. Models are built to predict two sets of outcomes: first classifying LVEF below five clinical thresholds of 50%, 45%, 40%, 35%, and 30%, and second to regress the LVEF value. We first build the non-sequential models to predict the outcomes, then remove their final layer, wrap them with the sequential component, and train the resulting system to predict the outcome, sweeping over training parameters and sequential-modeling approaches.

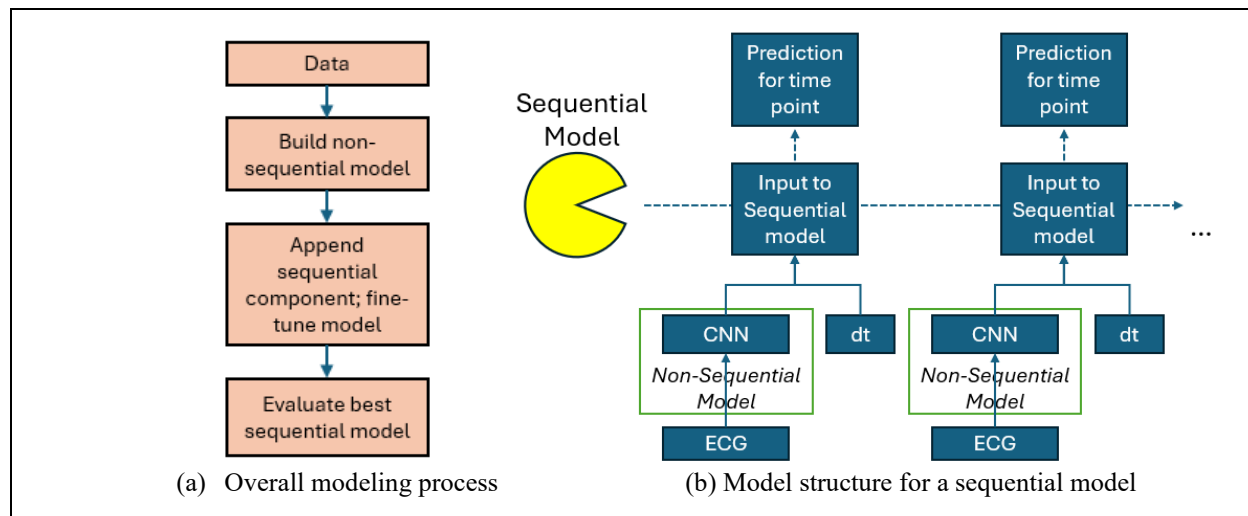


Figure 1. Modeling process. (a) Data is first used to build non-sequential models that treat each ECG as if from a different patient; these models then have their final layer removed and are wrapped with a sequential component (such as an LSTM) and trained on each patient’s forward-masked ECG history. The model that performs best on the validation-set LVEF <50% AUROC task is chosen for reporting test set metrics. (b) Illustration of how the sequential and non-sequential components interact in the LSTM case. ECG is first processed by non-sequential models, whose features then act as inputs to the sequential components.

Non-sequential models

The non-sequential model processes an individual ECG record. It follows an adapted ResNet¹³ ECG-processing CNN module with a feed-forward layer adapting CNN outputs to dimension 128 or 2048. Variants of this architecture have previously been used to predict patient age¹³, mortality¹⁴, and LVEF¹¹ from ECG.

Sequential models

The sequential model operates on the features of the non-sequential model and the timing between successive ECGs. Our data is an irregularly sampled time-series: the timing between successive ECGs is non-uniform and since small differences may be informative (e.g. preceding or following a surgical intervention), they are provided with day-accuracy to each model. We explore three sequential approaches: LSTM¹⁵, T-LSTM¹⁶, and a transformer-encoder¹⁷:

Long Short-Term Memory, “LSTM”, is a common approach to processing sequential information. To encode timing, we concatenate the time between the previous ECG and the current ECG (‘0’ for the first ECG in the sequence) with the non-sequential model’s features and feed the resulting vector to the LSTM.

T-LSTM is an LSTM variant that explicitly includes a separate input pathway that marks input timing, adapting the LSTM architecture to irregular timing inputs.

Transformers encode information using attention mechanisms and have become incredibly popular with the advent of large language models. We use the encoder portion of the transformer architecture, which uses masked self-attention to encode a sequence of inputs causally. We use 8 attention heads and encode time-since-first-ECG as an additive sin/cos signal¹⁸ rather than providing the time between patient ECGs as an additional channel.

Model training

Models are trained for up to 200 epochs or until validation-set performance fails to improve for 20 epochs. Models are trained with an AdamW optimizer using learning rate scheduler that scales the rate by 0.1x when validation performance fails to improve over 10 epochs. Models train in batches of 512 ECGs. Classification error is calculated with binary cross-entropy loss, while regression error is calculated with mean square error. To generate batches that adapt to variable-length patient records, patient IDs are shuffled, and training batches append ECGs per patient until the batch limit of 512 ECGs would be exceeded. One epoch goes through each patient’s information once. As noted earlier, non-sequential models treat each ECG as if sourced from a unique patient.

Non-sequential models trained with a learning rate of $1e-3$ and used an established ECG-processing architecture¹³. Two non-sequential models were trained per random seed, using a linear layer to compress the models’ 5120-dimension embedding to either 128 or 2048 dimensions prior to predicting outputs. For sequential models, we freeze non-sequential weights and sweep architecture depths of $\{1\dots3\}$, layer dimension of $2^{\{5\dots11\}}$, and learning rates of $\{1e-5, 5e-5, 1e-4, 1e-3\}$. We also explore the same depth and dimension sweep, with learning rate $1e-3$, with non-sequential weights unfrozen (i.e. “end-to-end” training). This sweep was first run for LSTM, T-LSTM, and transformer approaches on one random seed. After no transformer models outperformed the median performance of non-sequential models, further seeds were not explored for transformers. Otherwise, the experiment was repeated across five random seeds. Additionally, as most of our cluster GPUs could not handle the depth-3 seq-dimension-2048 configuration, this combination was ultimately excluded from analysis.

We also explored training models to classify discretized LVEF and regress its value simultaneously, rather than training separate models for both tasks. We then used the top sequential approach from the previous classification task to fine-tune these models across each patient’s ECGs. Mean Square Error was used for regression loss, while binary cross-entropy was used for classification loss, without weighting.

Results

Data

The BCH dataset¹¹ used consists of 178,495 ECGs recorded between 10/1987 and 12/2022 from a hospital system in Boston, Massachusetts. These include both emergency and routine visits and are enriched for pediatric and congenital heart disease cases. Cohort details are in (Table 1). LVEF measures were taken from echocardiogram studies recorded within two days of the ECG, preferring the temporally closest echo if multiple were available. Here, 25% of patients are followed for at least two years, and 10% for at least eight years.

Table 1. Details of the patient cohort from Boston Children’s Hospital, collected between October 1987 and December 2022

		Overall	Fold Training	Testing
ECG - Echo pairs, n		178495	124265	54230
Patients, n		70226	49158	21068
Patients with at least 5 ECGs		8707	5995	2712
LVEF, mean (SD)		0.61 (0.08)	0.61 (0.08)	0.61 (0.08)
LVEF Bin, n (%)	< 30%	2242 (1.3)	1587 (1.3)	655 (1.2)
	30 - 35%	1216 (0.7)	828 (0.7)	388 (0.7)
	35 - 40%	1760 (1.0)	1229 (1.0)	531 (1.0)
	40 - 45%	2910 (1.6)	2028 (1.6)	882 (1.6)
	45 - 50%	5217 (2.9)	3640 (2.9)	1577 (2.9)
	>= 50%	165150 (92.5)	114953 (92.5)	50197 (92.6)
Age at Echo, yrs., mean (SD)		11.8 (10.5)	11.7 (10.5)	12.0 (10.6)

LVEF Classification Results

The top model’s classification performance is compared to the non-sequential baseline on the test sets in (Figure 2). Predicting LVEF<50% median AUROC (IQR) increased from 0.873 (0.870, 0.882) to 0.892 (0.890, 0.894) with paired differences of 0.024 (0.010, 0.025) when considering all ECGs, and 0.865 (0.850, 0.869) to 0.891 (0.885, 0.894) with paired differences of 0.034 (0.014, 0.044) when only considering ECGs with at least 5 preceding patient ECGs.

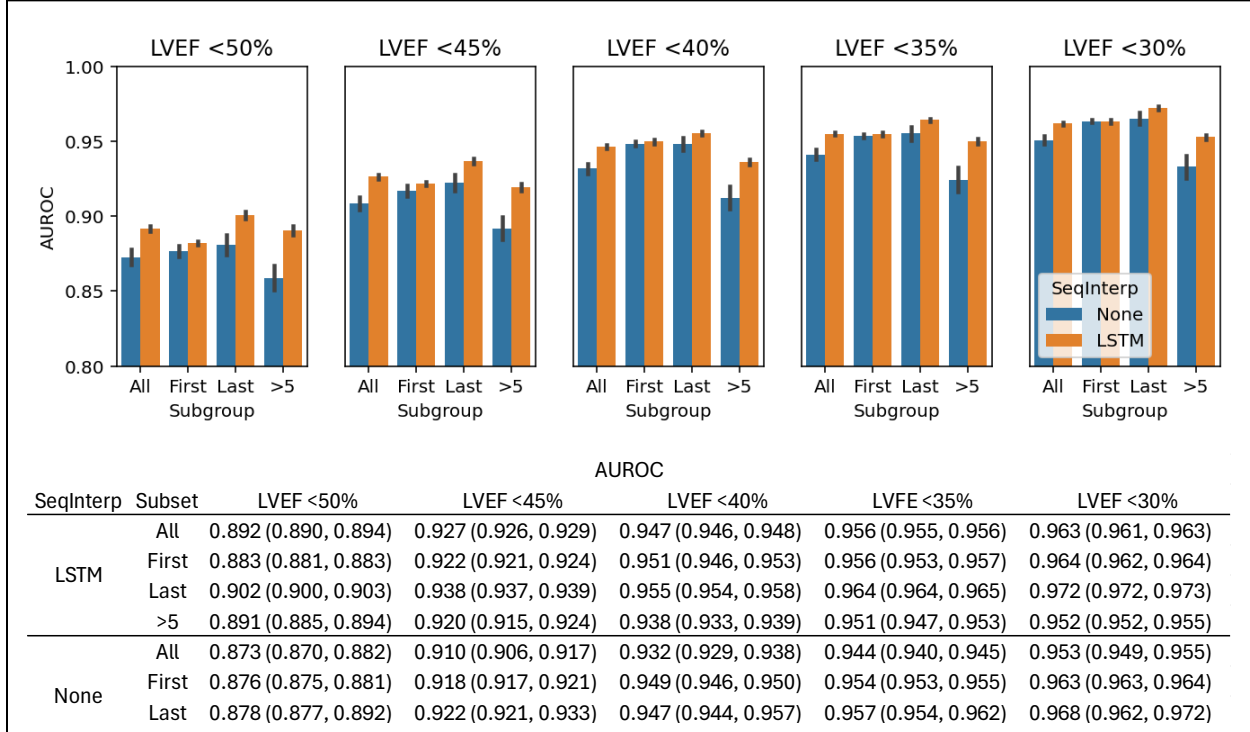


Figure 2. Bar plots and table show Median (IQR) for top model AUROC when predicting LVEF bin class using the “top” sequential approach. The subgroup label indicates which ECGs were considered when calculating the performance shown in the corresponding bar: ‘All’ considers all ECGs, ‘First’ and ‘Last’ consider only the first/last ECG per patient, and ‘>5’ considers only ECGs preceded by at least 5 ECGs.

Regression and Classification of LVEF

Modeling results when predicting both binned LVEF and LVEF value for the test sets with and without previous ECGs is shown in (Figure 3). Using the same best sequential architecture as before, median Pearson R (IQR) improved from 0.46 (0.38, 0.46) to 0.50 (0.47, 0.50) when considering all ECGs with a paired difference of 0.04 (0.01, 0.12), and 0.44 (0.37, 0.45) to 0.50 (0.46, 0.52) with a paired difference of 0.08 (0.02, 0.13) when considering patient ECGs with at least 5 preceding recordings.

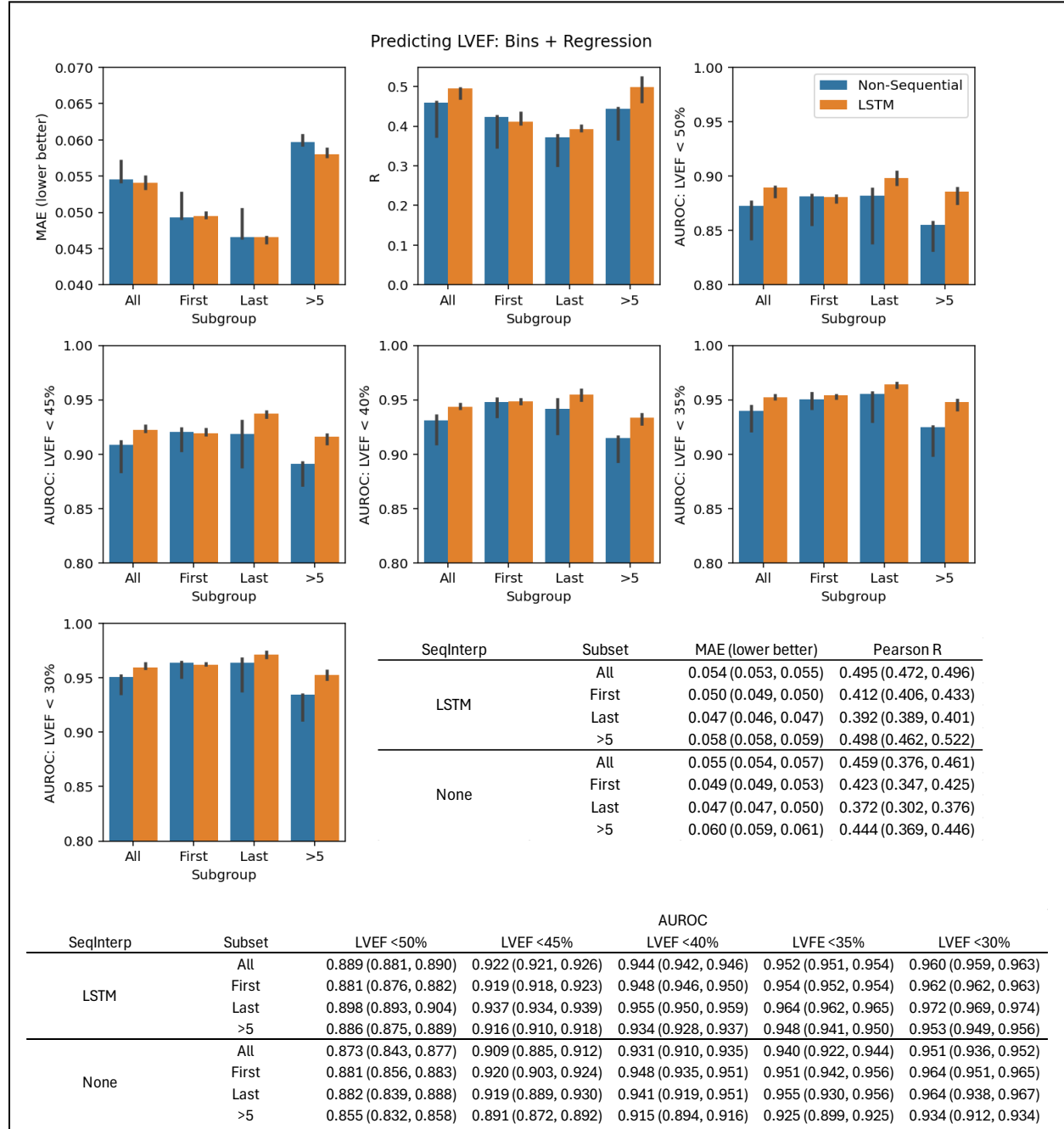


Figure 3. Bar plots and tables show Median (IQR) for model AUROC, R, and MAE when predicting LVEF value as well as bin class inclusion using “top” sequential approach. The subgroup label indicates which ECGs were considered when calculating the performance shown in the corresponding bar: ‘All’ considers all ECGs, ‘First’ and ‘Last’ consider only the first/last ECG per patient, and ‘>5’ considers only ECGs preceded by at least 5 ECGs.

Discussion

Adult AI-ECG models have rapidly accelerated¹⁹ since the landmark work by the Mayo Clinic in 2019²⁰, with countless single-snapshot ECG models now existing (including an FDA-cleared product). However, nearly every approach to date has relied on a single snapshot ECG as the sole input, and prospective evaluations of these models in clinical practice have shown modest positive predictive value. In this study, we hypothesized that leveraging a patient's full ECG history, rather than a single tracing, could capture meaningful longitudinal patterns and thereby improve predictive performance, especially in the tailored pediatric and congenital heart disease population.

Sequential Time series in ECG-LVEF

Overall, our results confirm that considering ECG sequences can improve AUROC when predicting abnormal LVEF. On the LVEF <50% task, median AUROC (IQR) improves by 2.4 (1.0, 2.5) points when considering all ECGs, and 3.4 (1.4, 4.4) points when considering patient ECGs with at least 5 preceding records. LVEF regression also improves with, for example, median Pearson R (IQR) rising by 0.04 (0.01, 0.12) when considering all ECGs, and 0.08 (0.02, 0.13) when only considering ECGs with at least 5 preceding patient ECGs. Several examples are shown in (Figure 3).

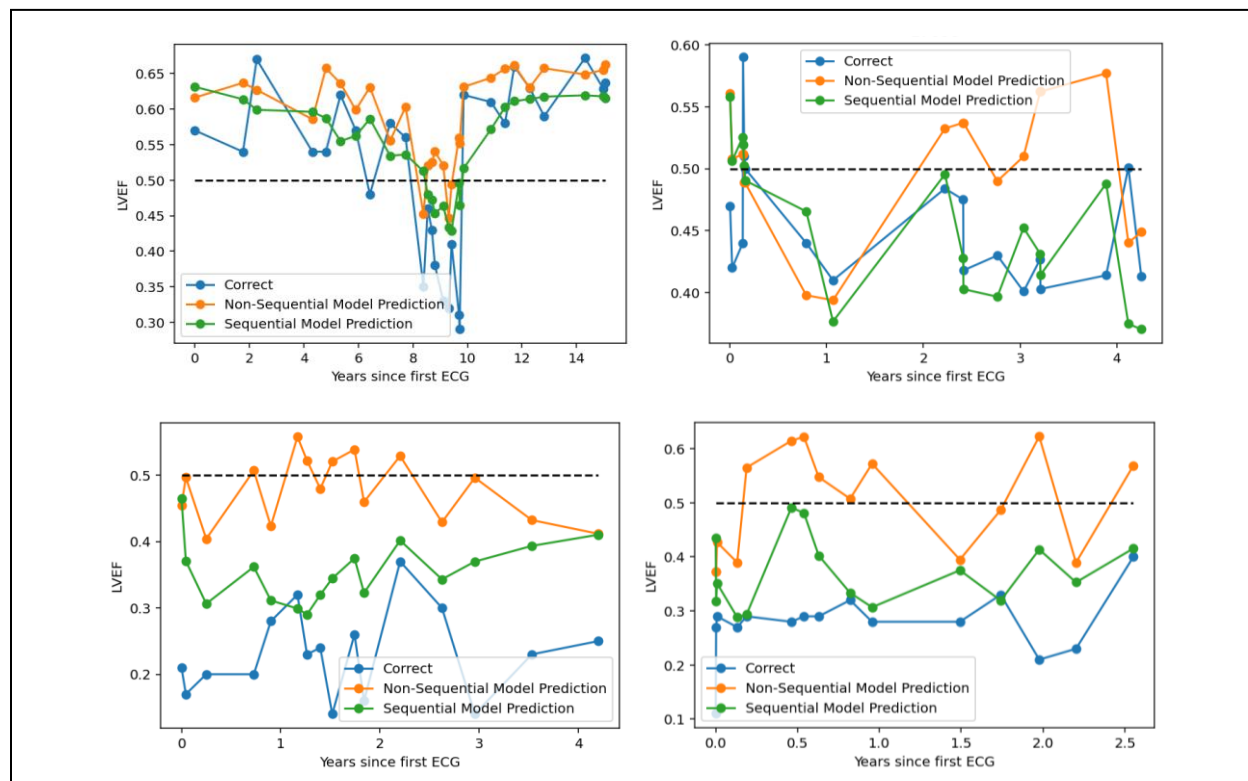


Figure 3. Examples of patient LVEF predictions from non-sequential and sequential models for one random seed.

Novelty and Significance

The first contribution of our work is that it provides new model architectures for AI-ECG that predict measures of heart function from the full history ECGs in a patient's medical record, rather than at a single snapshot. This allows AI-ECG models to leverage much richer longitudinal data about a patient than was previously possible. The architectures allow for variable-length histories, which is a key consideration in modeling diverse patient populations that use the hospital at non-uniform intervals and with varying cadences. Our work's second contribution is providing a rich empirical benchmark of state-of-the-art AI-ECG architectures that fulfill this task in a large patient cohort with rich longitudinal data. The model designs, experimental analysis and findings should advance our understanding of sequential AI-ECG and provide a promising first measure of performance in applications to longitudinal patient data.

Clinical Value of this Serial AI-ECG Model

Our prior single ECG work¹¹ established the potential for AI-ECG to support: 1) screening for LV systolic dysfunction in congenital heart disease (which may reduce unnecessary echos); 2) identifying patients at greater risk of LV systolic dysfunction (which may facilitate earlier initiation of medical therapy); and 3) tracking LV systolic dysfunction risk across the lifespan. Here, we demonstrate that AUROC and MAE improve specifically for high-utilization patients, which largely corresponds to our high-risk and lifelong congenital heart disease care cohort. Improving model performance for this subset may improve our ability to detect LV systolic dysfunction early and initiate life-saving guideline-directed medical therapy in this vulnerable group. Using such models for clinical decision support requires formalized trials establishing how models affect clinical outcomes and workflow.

Future Applications

Our results substantially expand past sequential-ECG work, which focused on atrial fibrillation in adult patients¹², to also include LVEF class and value prediction in a primarily pediatric context. The approach used here can be readily extended to other ECG modeling tasks, such as pathology identification or tracking pathology progression. It aligns well with the recent trend in generating ECG foundation models²¹. While here we use the features from a classifier to embed the ECG signals prior to sequential modeling, other embeddings such as those from foundation models could be used.

Limitations and future directions

This study has the following limitations. First, the current work is limited to a single center, and so it is crucial that future work validates performance on external sites. Although nascent and publicly available ECG-echo datasets exist for adult cardiology,^{9,22} to our knowledge they do not exist in the tailored pediatric and congenital heart disease population. Second, we do not examine the impact of additional covariates such as age or sex; while work has shown that these do not substantially alter AI-ECG LVEF prediction¹¹, their effect has not been examined in a sequential context. Additionally, although we sweep a large set of models, our analysis is not exhaustive. For example, we do not explore more complex training paradigms such as unfreezing the non-sequential component after several epochs, nor do we consider alternative time encodings. Lastly, this work is based on a pediatric dataset with a high prevalence of congenital heart disease and works better for patients with many ECGs, which may skew the model performance towards sicker subpopulations.

Conclusion

Overall, results indicate that AI-ECG that considers serial patient ECGs can improve AUROC when predicting LVEF.

Supplemental

Ethics approval and consent to participate: The BCH data study used here was approved for a waiver of informed consent by Boston Children's Hospital Institutional Review Board (IRB-P00044967).

Funding: Funding support included the Thrasher Research Fund Early Career Award (J.M.), Boston Children's Hospital Electrophysiology Research Education Fund (J.M., J.K.T.), and the Kostin Innovation Fund (P.L., W.G.L.).

Acknowledgements: The authors would like to acknowledge Boston Children's Hospital's High-Performance Computing Resources Clusters Enkefalos 3 which were used to conduct this research.

References

- ¹ Lima EM, Ribeiro AH, Paixão GM, Ribeiro MH, Pinto-Filho MM, Gomes PR, Oliveira DM, Sabino EC, Duncan BB, Giatti L, Barreto SM. Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nature communications*. 2021 Aug 25;12(1):5117.
- ² Gow B, Pollard T, Nathanson LA, Johnson A, Moody B, Fernandes C, Greenbaum N, Waks JW, Eslami P, Carbonati T, Chaudhari A. Mimic-iv-ecg: Diagnostic electrocardiogram matched subset. Type: dataset. 2023 Sep;6:13-4.
- ³ Zhu H, Jiang Y, Cheng C, Wang J, Zhu L, Chen X, Feng K, Liu Y, Zhang L, Luo Q, He X. Four-channel ECG as a single source for early diagnosis of cardiac hypertrophy and dilation—a deep learning approach. *NEJM AI*. 2024 Sep 26;1(10):AIoa2300297.
- ⁴ Raghunath S, Cerna AE, Jing L, vanMaanen DP, Stough J, Hartzel DN, Leader JB, Kirchner HL, Good CW, Patel AA, Delisle BP. Deep neural networks can predict mortality from 12-lead electrocardiogram voltage data. *arXiv preprint arXiv:1904.07032*. 2019 Apr 15.
- ⁵ Sau A, Pastika L, Sieliweczyk E, Patlatzoglou K, Ribeiro AH, MCGurk KA, Zeidaabadi B, Zhang H, Macierzanka K, Mandic D, Sabino E. Artificial intelligence-enabled electrocardiogram for mortality and cardiovascular risk estimation: a model development and validation study. *The Lancet Digital Health*. 2024 Nov 1;6(11):e791-802.
- ⁶ Mayourian J, Gearhart A, La Cava WG, Vaid A, Nadkarni GN, Triedman JK, Powell AJ, Wald RM, Valente AM, Geva T, Duong SQ. Deep learning-based electrocardiogram analysis predicts biventricular dysfunction and dilation in congenital heart disease. *Journal of the American College of Cardiology*. 2024 Aug 27;84(9):815-28.
- ⁷ Takahiro Kokubo et al. Automatic detection of left ventricular dilatation and hypertrophy from electrocardiograms using deep learning. *International Heart Journal*. 63(5):939–947, 2022. DOI 10.1536/ihj.22-132
- ⁸ Adedinsewo D, Carter RE, Attia Z, Johnson P, Kashou AH, Dugan JL, Albus M, Sheele JM, Bellolio F, Friedman PA, Lopez-Jimenez F. Artificial intelligence-enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea. *Circulation: Arrhythmia and Electrophysiology*. 2020 Aug;13(8):e008437.
- ⁹ Elias P, Finer J. Echonext: A dataset for detecting echocardiogram-confirmed structural heart disease from ecgs.
- ¹⁰ Vaid A, Johnson KW, Badgeley MA, Somani SS, Bick M, Landi I, Russak A, Zhao S, Levin MA, Freeman RS, Charney AW. Using deep-learning algorithms to simultaneously identify right and left ventricular dysfunction from the electrocardiogram. *Cardiovascular Imaging*. 2022 Mar 1;15(3):395-410.
- ¹¹ Mayourian J, Asztalos IB, El-Bokl A, Lukyanenko P, Kobayashi RL, La Cava WG, Ghelani SJ, Vetter VL, Triedman JK. Electrocardiogram-based deep learning to predict left ventricular systolic dysfunction in paediatric and adult congenital heart disease in the USA: a multicentre modelling study. *Lancet Digital Health* 2025 Apr 1;7(4):e264-74
- ¹² Choi JH, Song SH, Kim J, Jeon J, Woo K, Cho SJ, Park SJ, On YK, Kim JY, Park KM. Sequential AI-ECG Diagnostic Protocol for Opportunistic Atrial Fibrillation Screening: A Retrospective Single-Center Study. *Journal of Clinical Medicine*. 2025 Sep 22;14(18):6675.
- ¹³ Ribeiro AH, Ribeiro MH, Paixão GM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MP, Andersson CR, Macfarlane PW, Meira Jr W, Schön TB. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature communications*. 2020 Apr 9;11(1):1760.
- ¹⁴ Lukyanenko P, Mayourian J, Liu M, Triedman JK, Ghelani SJ, La Cava WG. Deep Survival Analysis from Adult and Pediatric Electrocardiograms: A Multi-center Benchmark Study. *arXiv e-prints*. 2024 Jun:arXiv-2406.
- ¹⁵ Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation*. 1997 Nov 15;9(8):1735-80.
- ¹⁶ Baytas IM, Xiao C, Zhang X, Wang F, Jain AK, Zhou J. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD conference on knowledge discovery and data mining* 2017 Aug 4 (pp. 65-74).
- ¹⁷ Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Advances in neural information processing systems*. 2017;30.
- ¹⁸ Chen Y, Ren K, Wang Y, Fang Y, Sun W, Li D. Contiformer: Continuous-time transformer for irregular time series modeling. *Advances in Neural Information Processing Systems*. 2023 Dec 15;36:47143-75.
- ¹⁹ Anumana. Anumana receives U.S. FDA 510(k) clearance for ECG-AI algorithm to detect low ejection fraction. October 2, 2023 (<https://anumana.ai/newsroom/ZRo9DhAAACYAxfK>).
- ²⁰ Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac con- tractile dysfunction using an artificial intelligence-enabled electro- cardiogram. *Nat Med* 2019;25:70-74. DOI: 10.1038/s41591-018- 0240-2.
- ²¹ Li J, Aguirre AD, Junior VM, Jin J, Liu C, Zhong L, Sun C, Clifford G, Brandon Westover M, Hong S. An Electrocardiogram Foundation Model Built on over 10 Million Recordings. *NEJM AI* 2025 Jun 26;2(7):AIoa2401033.
- ²² Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*. 2000 Jun 13;101(23):e215-20.