










ORIGINAL RESEARCH

# Quantitative Prediction of Right Ventricular Size and Function From the ECG

Son Q. Duong , MD, MS,\* Akhil Vaid , MD,\* Vy Thi Ha My , PhD; Liam R. Butler , BS; Joshua Lampert, MD; Robert H. Pass, MD; Alexander W. Charney, MD, PhD; Jagat Narula , MD, PhD; Rohan Khara , MD, MS; Ankit Sakhuja, MBBS, MS; Hayit Greenspan, PhD; Bruce D. Gelb , MD; Ron Do , PhD; Girish N. Nadkarni , MD, MPH

**BACKGROUND:** Right ventricular ejection fraction (RVEF) and end-diastolic volume (RVEDV) are not readily assessed through traditional modalities. Deep learning-enabled ECG analysis for estimation of right ventricular (RV) size or function is unexplored.

**METHODS AND RESULTS:** We trained a deep learning-ECG model to predict RV dilation (RVEDV >120 mL/m<sup>2</sup>), RV dysfunction (RVEF ≤40%), and numerical RVEDV and RVEF from a 12-lead ECG paired with reference-standard cardiac magnetic resonance imaging volumetric measurements in UK Biobank (UKBB; n=42 938). We fine-tuned in a multicenter health system (MSH<sub>original</sub> [Mount Sinai Hospital]; n=3019) with prospective validation over 4 months (MSH<sub>validation</sub>; n=115). We evaluated performance with area under the receiver operating characteristic curve for categorical and mean absolute error for continuous measures overall and in key subgroups. We assessed the association of RVEF prediction with transplant-free survival with Cox proportional hazards models. The prevalence of RV dysfunction for UKBB/MSH<sub>original</sub>/MSH<sub>validation</sub> cohorts was 1.0%/18.0%/15.7%, respectively. RV dysfunction model area under the receiver operating characteristic curve for UKBB/MSH<sub>original</sub>/MSH<sub>validation</sub> cohorts was 0.86/0.81/0.77, respectively. The prevalence of RV dilation for UKBB/MSH<sub>original</sub>/MSH<sub>validation</sub> cohorts was 1.6%/10.6%/4.3%. RV dilation model area under the receiver operating characteristic curve for UKBB/MSH<sub>original</sub>/MSH<sub>validation</sub> cohorts was 0.91/0.81/0.92, respectively. MSH<sub>original</sub> mean absolute error was RVEF=7.8% and RVEDV=17.6 mL/m<sup>2</sup>. The performance of the RVEF model was similar in key subgroups including with and without left ventricular dysfunction. Over a median follow-up of 2.3 years, predicted RVEF was associated with adjusted transplant-free survival (hazard ratio, 1.40 for each 10% decrease; P=0.031).

**CONCLUSIONS:** Deep learning-ECG analysis can identify significant cardiac magnetic resonance imaging RV dysfunction and dilation with good performance. Predicted RVEF is associated with clinical outcome.

**Key Words:** cardiac MRI ■ deep learning ■ ECG ■ right ventricle

**R**ight ventricular (RV) size and functional metrics have important prognostic implications for many diseases, including cardiomyopathy, pulmonary hypertension, and structural heart disease.<sup>1</sup> Volumetric measurements such as RV ejection fraction (RVEF) and end-diastolic volume (RVEDV) are correlated with major adverse outcomes<sup>2</sup> and have particular

importance in patients with congenital heart disease<sup>3</sup> and pulmonary hypertension.<sup>4</sup>

However, current tools for accurate quantification of RVEF and RVEDV are limited. Estimation of volumes by traditional 2-dimensional echocardiographic RV measurements are not recommended in adults or children<sup>5,6</sup> due to poor reproducibility, especially in the

Correspondence to: Son Q. Duong, MD, MS, Mount Sinai School of Medicine, 1468 Madison Ave, Annenberg 3rd Fl, New York, NY 10029. Email: [son.duong@mssm.edu](mailto:son.duong@mssm.edu)

\*S. Q. Duong and A. Vaid are co-first authors.

Preprint posted on MedRxiv April 26, 2023. doi: <https://doi.org/10.1101/2023.04.25.23289130>.

This manuscript was sent to Barry London, MD, PhD, Senior Guest Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at <https://www.ahajournals.org/doi/suppl/10.1161/JAHA.123.031671>

For Sources of Funding and Disclosures, see page 14.

© 2023 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

JAHA is available at: [www.ahajournals.org/journal/jaha](http://www.ahajournals.org/journal/jaha)

## RESEARCH PERSPECTIVE

### What Is New?

- We developed and fine-tuned a deep learning algorithm on ECGs that estimates right ventricular volume and ejection fraction; the ECG-predicted right ventricular ejection fraction was significantly associated with composite cardiac outcome.

### What Questions Should Be Addressed Next?

- How does this deep learning model perform across a wider range of pathological conditions?
- How does the model perform in external validation across multiple centers?

## Nonstandard Abbreviations and Acronyms

|              |  |
|--------------|--|
| <b>AUPRC</b> | area under the precision-recall curve  |
| <b>CNN</b>   | convolutional neural network           |
| <b>DL</b>    | deep learning                          |
| <b>MSH</b>   | Mount Sinai Hospital                   |
| <b>RVEDV</b> | right ventricular end-diastolic volume |
| <b>UKBB</b>  | UK Biobank                             |

setting of significant pathology.<sup>7</sup> Three-dimensional echocardiography is a promising novel technology but has significant technical and processing limitations and is not broadly available.<sup>8,9</sup> Cardiac magnetic resonance imaging (cMRI) is the clinical reference standard for volumetric quantification, but cMRI is not globally accessible, is time and resource intensive, and cannot be performed in some patients, such as those with implanted devices and those who will not tolerate the examination.

There is an urgent clinical need for novel methods to assess RV size and function that are validated against gold-standard metrics, widely available, and simple to perform. We sought to develop such a method through application of modern machine learning techniques to a ubiquitous clinical tool: the ECG.

Recent developments in deep learning (DL) technology have produced models to interpret ECG for accurate diagnosis of structural heart disease and prognosis,<sup>10</sup> but the use of DL on ECG to quantify RVEDV and RVEF has not been explored. We hypothesized that we could develop a DL algorithm to accurately quantify RVEDV and RVEF from a pathological clinical cohort and a large healthy registry of paired ECG and cMRI and that ECG assessment of RV function was associated with clinical outcome.

## METHODS

This project was approved by the institutional review board of the Icahn School of Medicine at Mount Sinai School of Medicine and waiver of informed consent was obtained. The code pipeline is available at <https://github.com/akhilvaid/RVSizeFunction>. The code provided in the repository contains all functions required to load and analyze data using a choice of models. The ECG data and some specialized code for preprocessing these data cannot be released due to concerns of patient privacy and institutional intellectual property, respectively.

### Overview

Broadly, we trained a convolutional neural network (CNN) on a larger data set of paired ECG and cMRI data from the UK Biobank (UKBB) and then fine-tuned the model on a smaller, retrospectively collected cohort of clinically indicated cMRI–ECG pairs from a single urban referral center. We then performed a prospective validation on the subsequent 4 months following the model training/testing set.

### UKBB Data Set

We accessed paired ECG and cMRI data from the UKBB. Participant enrollment and cMRI acquisition parameters have been described previously.<sup>11</sup> We used a previously validated automated segmentation method to obtain RVEDV and RVEF<sup>12</sup> from the balanced steady-state free precession cine short-axis stack. This method contours the right ventricle to include the papillary muscles within the RV volume. A cMRI-trained cardiologist (S.D.) performed manual segmentation of 10 randomly selected studies for comparison with the automated method, with Bland–Altman analysis results shown in [Figure S1](#). RVEF mean difference was 3.7 (95% upper limit of agreement, 12.8%; lower limit of agreement, –5.4%), and RVEDV indexed to body surface area mean difference was –2.9 mL/m<sup>2</sup> (95% upper limit of agreement, 17.7; lower limit of agreement, –23.5%). We excluded entries with RVEF values <10% or >80%. Per the UKBB study protocol, a standard 12-lead ECG was obtained at the same visit as cMRI. ECG acquisition parameters have been previously published<sup>13</sup> and were accessed in Extensible Markup Language format with preprocessing per below.

### Clinical Data Set

As the UKBB did not contain a high prevalence of dilated or dysfunctional RV measurements, we fine-tuned the model on a clinical cohort of diverse patients in a large urban health system with 5 hospitals (MSH<sub>original</sub> [Mount Sinai Hospital, New York, NY]). We queried our cMRI

reporting software (Precession, Intelrad, Montreal, Canada) for studies performed from April 4, 2012, to November 22, 2022, in patients aged  $\geq 18$  years and collected clinically indicated cMRI reports paired with ECGs within 2 weeks before or 2 weeks after cMRI examination. The MUSE Cardiology Information System (GE, Boston, MA) ECG database was queried for relevant patient medical record numbers, filtered by date of acquisition, and exported as Extensible Markup Language files. We excluded patients with congenital heart disease as an indication for MRI, or without a recorded ECG within the inclusion time frame. Like the UKBB data set, we excluded entries with RVEF values  $< 10\%$  or  $> 80\%$ . We extracted clinically measured RVEDV and RVEF from reports. MRI is typically performed on a 1.5T scanner using balanced steady-state free precession sequences with manual contouring of the cine short-axis stack to include RV papillary muscles within the blood pool for the RV, though all reports were included. Left ventricular volume (contoured to exclude the papillary muscles from the blood pool) and ejection fraction were also collected. Patient age at cMRI, sex, indication for cMRI, cardiac rhythm at cMRI, weight, and height were also collected from the report.

## ECG Preprocessing

ECG waveforms represented as a time series of voltage sampled at 500 Hz in each lead were extracted as Extensible Markup Language files from both the UKBB and MSH data sets. We included 8 channels from the 12-lead ECG: leads I, II, and V1 to V6 as the other leads are linear transformations without additional information. Our ECG preprocessing pipeline was described in detail previously.<sup>14</sup> Briefly, we corrected for baseline wander through the application of a median filter applied over a 2-second window and a subsequent Butterworth bandpass filter applied to a 0.5- to 40-Hz range. We excluded low-quality ECG recordings/ECGs with missing lead information by discarding ECGs with variations in QRS amplitude or frequency greater than  $\pm 3$  SD of the population mean for that lead. A total of 633 (3.4%) ECGs in the MSH data set and 549 (1.3%) ECGs in the UKBB data set were removed (see Figure 1). ECGs were then restricted to the first 5 seconds to ensure all patients had equivalent data and plotted to images with standardization of lead amplitude to the minimum and maximum voltages within each preprocessed lead.

## Model Construction and Training

The model input was the preprocessed cMRI-paired ECG. The model outputs of interest were numerical RVEF (%) and RVEDV/body surface area ( $\text{mL}/\text{m}^2$ ) (ie, supervised learning for regression tasks) as well as these values dichotomized into pathological thresholds

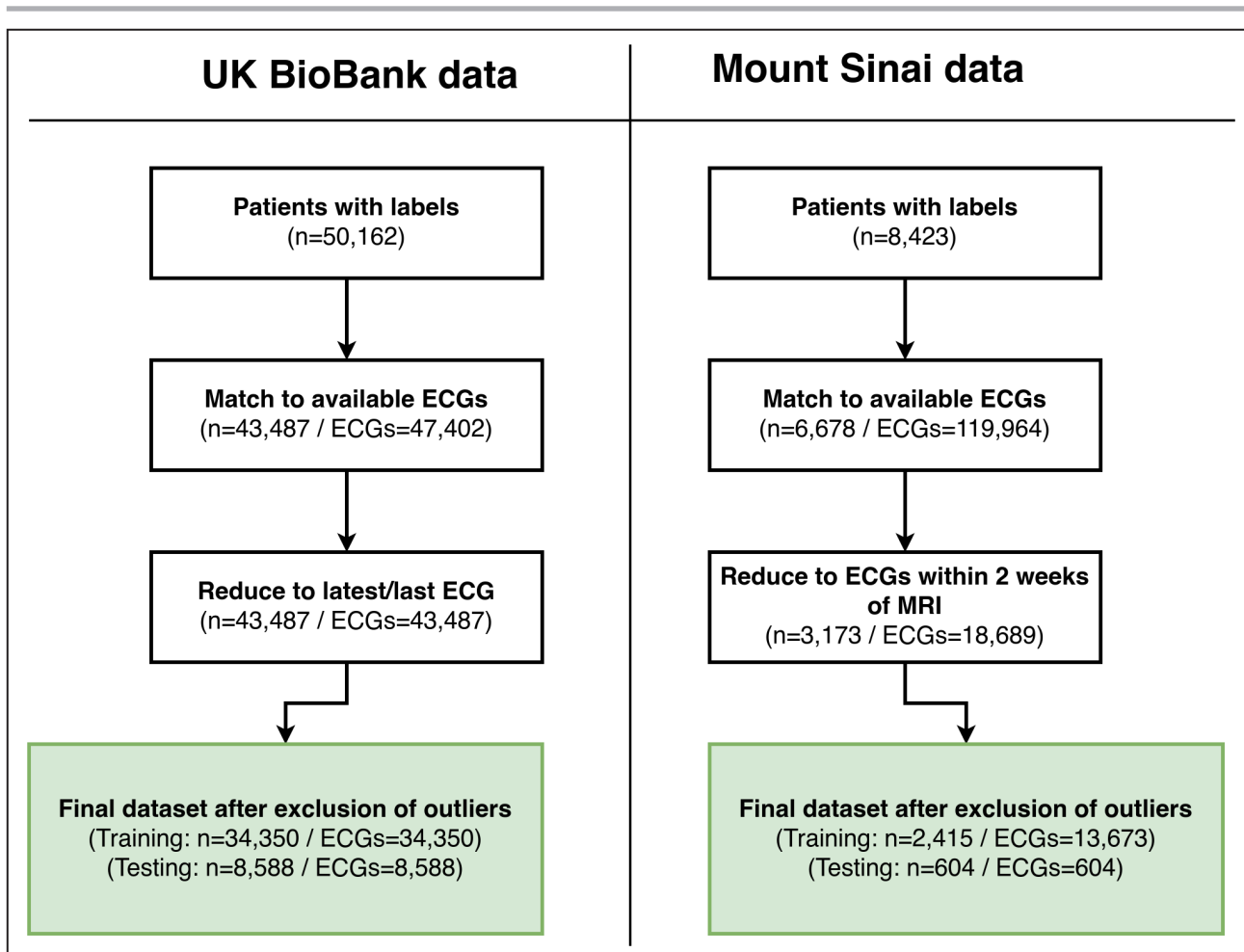
of  $\text{RVEF} < 40\%$  and  $\text{RVEDV} \geq 120 \text{ mL}/\text{m}^2$  (ie, supervised learning for classification tasks). Thus, for each data set, 4 separate models were trained: (1) RVEF regression, (2) RVEF classification, (3) RVEDV regression, and (4) RVEF classification. In the UKBB study protocol, only 1 ECG was obtained per subject and cMRI, which was randomly split into 80% training and 20% testing subsets. For the MSH<sub>original</sub> data set, as multiple ECGs per cMRI could occur within the inclusion time frame, patients were group shuffle split into 80% train and 20% test groups. This method ensured that no patient was represented both in the test and training set but allowed multiple ECGs obtained in the inclusion period for a patient to be considered in the training set. For the MSH<sub>original</sub> test set, only 1 subject with 1 paired ECG closest to the time of cMRI was included.

We selected a DenseNet-201 2-dimensional CNN architecture that had been previously trained on  $> 700\,000$  ECGs<sup>14</sup> as initialization parameters. Neural network architectures are uniquely able to extract information-rich features from complex waveform data and form the basis of several published DL-ECG models.<sup>14–18</sup> We trained the network using the Adam optimizer with cross-entropy as the loss function for classification and mean absolute error for regression tasks, respectively. To minimize overfitting, we monitored train-test loss over training epochs against an internal validation set of 5% of ECGs within the training cohort. Training was stopped at the epoch that performed best with respect to this internal validation set, despite continuing training with an adaptive learning rate for several more epochs. This method is functionally equivalent to early stopping as a form of regularization.

We first trained on the UKBB data set as a way of training features specific to the right ventricle, but the data set has a low prevalence of RV pathology. We then fine-tuned the model to predict outcomes in the MSH cohort, which had a higher outcome prevalence to learn features specific to RV pathology. Fine-tuning a neural network refers to the process of using pre-trained parameters (as opposed to random parameters) from a related task as a starting point for training on a new task or data set. This may potentially improve model performance on smaller data sets because the network may have already learned the basic features needed for the new task. We also tested whether using a pretrained model resulted in better performance compared with random initialization of parameters and found that use of the pretrained model increased model performance in the MSH<sub>original</sub> (Tables S1 and S2) and UKBB test sets.

## Model Evaluation

We evaluated model performance with area under the receiver operating characteristic curve (AUROC)



**Figure 1. Study Inclusion flow diagram.**

Cardiac MRI and ECG inclusion diagram chart for UK Biobank (left) and Mount Sinai Hospital data sets (right). MRI indicates magnetic resonance imaging.

and area under the precision-recall curve (AUPRC) for classification tasks. Precision-recall curves are generated by plotting the precision (ie, the positive predictive value) against the recall (ie, the sensitivity) over the range of predictions generated. AUPRC, by emphasizing the ability of the classifier to predict outcomes, is useful to evaluate the performance of a classifier in the setting of class imbalance where outcomes are more important to detect than non-outcomes.<sup>19</sup> Unlike AUROC, AUPRC depends on the baseline prevalence of disease and therefore can be interpreted as the ability of the classifier to identify a group “enriched” for disease over the baseline. We reported sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio (LR), and negative LR with bootstrapped 95% CIs at the probability threshold defined by the maximum sum of sensitivity and specificity (Youden’s J). We also evaluated classification model performance in key subgroups, including sex, left ventricular (LV) dysfunction

(defined as left ventricular ejection fraction [LVEF] <50%), arrhythmia, obesity (body mass index ≥30), and age ≥60 years. We performed AUROC comparisons and tested for effect modification between subgroups using the method of DeLong.<sup>20</sup> Patients with missing subgroup data (see Table 1) were excluded from subgroup analysis. For regression tasks, we visualized predicted versus expected scatterplots and described the relationship using nonparametric Passing–Bablok regression analysis, which is commonly used to compare 2 measurement methods.<sup>21</sup> The relationship between the predicted (ECG) and expected (cMRI) RV metric was also evaluated with linear  $R^2$  calculation and Lin’s concordance correlation coefficient. Difference plots in which the difference between ECG-predicted and cMRI RV metric were plotted against the cMRI reference standard, and a linear slope with prediction CIs based on  $1.96 \times \text{SD}$  of the residuals were reported as recommended by Bland and Altman.<sup>22</sup> Mean absolute error

**Table 1. Demographic and Clinical Characteristics in Mount Sinai Data Set**

|                                      | MSH <sub>original</sub> training set (n=2415) | MSH <sub>original</sub> test set (n=604) | MSH <sub>validation</sub> temporal validation (n=115) |
|--------------------------------------|---|--|---|
| Age, y, mean (SD)                    | 56.0 (16.4)                                   | 55.9 (17.0)                              | 56.3 (16.1)   |
| Sex, female, n (%)                   | 865 (37.5)*                                   | 209 (36.3)†                              | 45 (39.8)‡  |
| Sex, male, n (%)                     | 1550 (62.5)                                   | 395 (63.7)                               | 70 (60.2)   |
| Race or ethnicity, n (%)             | 995 (41.2)                                    | 258 (42.7)                               | 41 (35.7)   |
| White                                |   |  |   |
| Black                                | 476 (19.7)                                    | 101 (16.7)                               | 28 (24.4)   |
| Other/Unknown                        | 944 (39.1)                                    | 245 (40.6)                               | 46 (40)   |
| BSA, m <sup>2</sup> , mean (SD)      | 1.95 (0.29)                                   | 1.96 (0.27)                              | 1.95 (0.29)   |
| BMI                                  | 27.7 (6.4)*                                   | 27.6 (5.8)†                              | 27.6 (6.2)  |
| BMI ≥30kg/ht <sup>2</sup> , n (%)    | 663 (28.5)*                                   | 152 (26)†                                | 36 (31.3)   |
| Indication for cMRI, n (%)           | 1959 (81.1)                                   | 485 (80.3)                               | 86 (74.8)   |
| Cardiomyopathy                       |   |  |   |
| Abnormal ECG                         | 183 (7.6)                                     | 56 (9.3)                                 | 1 (0.9)   |
| Tumor                                | 107 (4.4)                                     | 22 (3.6)                                 | 8 (7.0)   |
| Valvular function                    | 20 (0.8)                                      | 5 (0.8)                                  | 0 (0)   |
| Other/unknown                        | 146 (6.0)                                     | 36 (6.0)                                 | 20 (17.4)   |
| Rhythm at cMRI, n (%)                | 1957 (91.1)*                                  | 476 (90.5)†                              | 106 (95.5)‡   |
| Normal sinus                         |   |  |   |
| Sinus with ectopic beats             | 97 (4.5)                                      | 25 (4.8)                                 | 2 (1.8)   |
| Atrial fibrillation/flutter          | 78 (3.6)                                      | 21 (4.0)                                 | 1 (0.9)   |
| Other                                | 15 (0.7)                                      | 4 (0.7)                                  | 2 (1.8)   |
| Admitted as inpatient at time of MRI | 863 (35.7)                                    | 217 (35.9)                               | 45 (39.1)   |
| LVEF, median (IQR)                   | 53 (36–61)*                                   | 53 (36–62)                               | 52 (34–63)  |
| LVEF <50%, n (%)                     | 1016 (42)*                                    | 261 (43.4)                               | 44 (38.3)   |
| RVEF, median (IQR)                   | 54 (46–61)                                    | 53 (46–60)                               | 54 (47–59)  |
| RVEF ≤40%, n (%)                     | 411 (17.0)                                    | 109 (18.0)                               | 18 (15.7)   |
| LVEDVi, median (IQR)                 | 84 (69–110)*                                  | 84 (70–108)†                             | 85 (65–107)   |
| RVEDVi, median (IQR)                 | 78 (63–95)*                                   | 80 (64–99)†                              | 70 (58–81)  |
| RVEDVi ≥120, n (%)                   | 202 (8.7)                                     | 62 (10.6)                                | 5 (4.3)   |

BMI indicates body mass index; BSA, body surface area; cMRI, cardiac magnetic resonance imaging; IQR, interquartile range; LVEF, left ventricular ejection fraction; LVEDVi, left ventricular end-diastolic volume indexed to body surface area; MRI, magnetic resonance imaging; MSH, Mount Sinai Hospital; RVEDVi, right ventricular end-diastolic volume indexed to body surface area; and RVEF, right ventricular ejection fraction.

\*Training set missing data: rhythm at cMRI, n=268; patient sex, n=108; LVEDVi, n=91; BSA; BMI; RVEDVi, n=87; LVEF, n=4.

†Test set missing data: rhythm at cMRI, n=78; BSA; BMI; LVEDVi; RVEDVi, n=20 missing; LVEF, n=3 missing; patient sex, n=28 missing.

‡Temporal validation set missing data: patient rhythm, n=4; patient sex, n=2.

of the models with 95% CI was calculated through 500 bootstrap iterations. We also performed saliency mapping using the GradCAM library to highlight the regions of a given ECG input image that were most associated with a prediction.

### Temporal Validation

We performed temporal validation (MSH<sub>validation</sub>) of the final classification models in the period of November 23, 2022 to March 21, 2023, from a cohort of patients with clinically indicated cMRI and paired ECG within +/-14 days of cMRI like the MSH<sub>original</sub> data set. Patients with prior ECG or cMRI included in the MSH<sub>original</sub> data set were excluded. AUROC and AUPRC metrics were reported as above.

### Survival Analysis for Composite Outcome

We reviewed the medical records of patients included within the MSH<sub>original</sub> test set to ascertain date of last known follow-up, death, and date of heart transplant if performed. We performed survival analysis to evaluate the association between predicted numerical RVEF and the combined outcome of death or heart transplantation. This composite outcome is commonly used to represent end-stage heart failure in heart failure prediction models.<sup>23</sup> Kaplan–Meier survival plots stratified by MRI-LVEF and ECG-predicted RVEF were generated. Using Cox proportional hazards for multivariable modeling, we tested several variables associated with heart failure outcome including MRI-quantified LVEF, age, sex, obesity (body mass index



>30), hospitalization at time of MRI (as a proxy for severe heart failure),<sup>23</sup> and self-reported race or ethnicity.<sup>24</sup> Variable selection was determined using the Akaike information criterion for the model excluding RVEF, and then the additive value of ECG-predicted RVEF versus MRI-quantified RVEF was evaluated using Akaike information criterion and Harrell's C-statistic. The linear risk assumption for continuous variables was assessed visually through martingale residual plots with a running mean smoother line plotted to identify nonlinearity. A nonlinear inflection point in risk related to age at around 60 years old was identified, and therefore age was dichotomized at  $\geq 60$  years old. We tested proportional hazards assumptions in the final model by evaluating the relationship between scaled Schoenfeld residuals and time. A *P* value <0.05 was considered statistically significant.

## Software

Model training was performed in Python (version 3.8; Python Software Foundation, Beaverton, OR) with pandas, numpy, scipy, scikit-learn, PyTorch, torchvision, PIL, matplotlib, seaborn, and GradCAM packages. Data visualization and descriptive statistics were performed in R (R Foundation for Statistical Computing, Vienna, Austria) and Stata13 (StataCorp, College Station, TX). Continuous metrics are reported as mean (SD).

## RESULTS

### Demographic and Clinical Information

We included 42 938 patient MRI-ECG pairs from UKBB after quality control. Mean age was 64.7 years, 48% were men, 96.5% were White individuals, and the most common comorbidities were hypertension (22.5%) and hyperlipidemia (17.3%). Heart surgery and history of myocardial infarction were reported in 1.8% and 1.4%, respectively. The mean RVEF was 57% (SD, 6.3), with RV dysfunction (RVEF  $\leq 40\%$ ) present in 425 (1.0%). Mean RVEDV was 82 (15.6) mL/m<sup>2</sup>, with RV dilation (RVEDV  $\geq 120$  mL/m<sup>2</sup>) in 680 (1.6%).

The MSH<sub>original</sub> data set consisted of 3019 patients that were split 80% to 20% into training and test cohorts. The training set consisted of 2415 patients and 13 673 ECG recordings meeting the inclusion criteria. The test set contained 604 patients, each with 1 paired cMRI and ECG. Test set mean age was 56 (17) years, and cardiomyopathy was the predominant indication for cMRI (80%). Prevalence of RV dysfunction was 18%, and prevalence of RV dilation was 11%. Demographic, clinical, and MRI characteristics of the MSH<sub>original</sub> cohorts split by test and training groups are listed in Table 1. A comparison between the available overlapping variables of the MSH and UKBB data sets

are shown in Table S3. A graphical overview of patient and ECG inclusion is shown in Figure 1.

The MSH<sub>validation</sub> group consisted of 115 patients, each with 1 paired ECG and cMRI. Demographic and clinical characteristics were concordant with the MSH<sub>original</sub> group, except for smaller RVEDV measurements and a lower prevalence of RV dilation (*n*=5; 4.3%). However, presence of RV systolic dysfunction was similar (*n*=18; 15.7%). Clinical and demographic information of the MSH training, test, and temporal validation cohorts is shown in Table 1.

### RVEF Classification Models

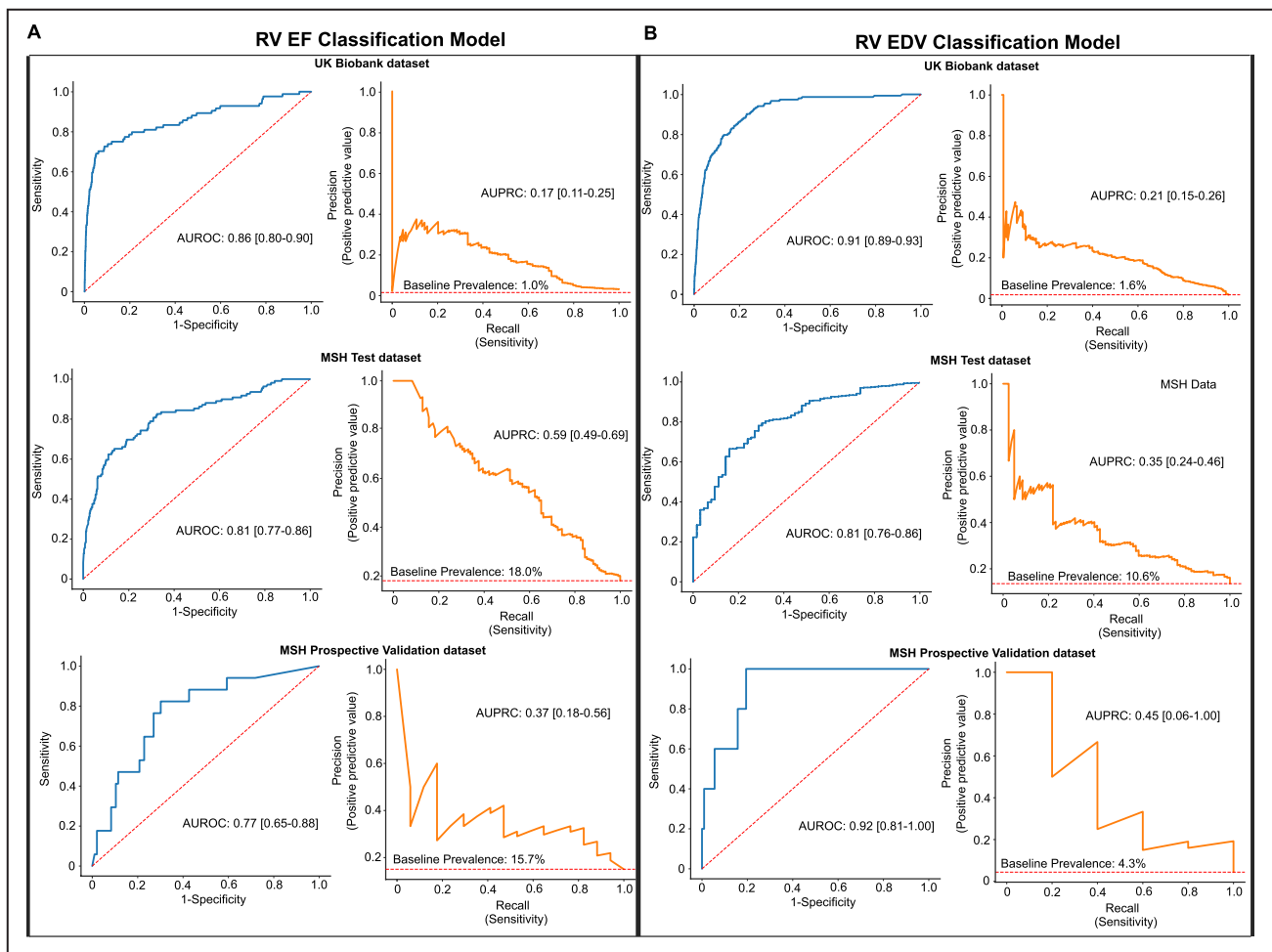
Model performance characteristics to classify RV systolic dysfunction (defined as RVEF  $\leq 40\%$ ) in UKBB and MSH data sets are shown in Figure 2. In UKBB (baseline prevalence 1.0%), AUROC was 0.86 (0.80–0.90), and AUPRC was 0.17 (0.11–0.26). In the MSH<sub>original</sub> data set (baseline prevalence, 18.0%), AUROC was 0.81 (0.77–0.86), and AUPRC was 0.59 (0.49–0.69). At Youden's J threshold, there was a sensitivity of 65% (95% CI, 57–75), specificity of 86% (83–89), positive predictive value of 51% (43–59), negative predictive value of 92% (89–94), positive LR of 4.7 (3.7–6.2), and negative LR of 0.40 (0.29–0.50) in the MSH<sub>original</sub> test data set. Performance between clinically relevant subgroups in the MSH<sub>original</sub> test data set is shown in Table 2. RVEF model AUROC did not differ between any tested subgroup, including those with and without LV dysfunction.

### RVEDV Classification Models

Model performance to classify RV dilation (defined as RVEDV  $\geq 120$  mL/m<sup>2</sup>) is shown in Figure 2B. In UKBB (baseline prevalence, 1.6%), AUROC was 0.91 (0.89–0.93), and AUPRC was 0.21 (0.15–0.26). In the MSH<sub>original</sub> data set (baseline prevalence, 10.6%), AUROC was 0.81 (0.76–0.86), and AUPRC was 0.35 (0.24–0.46). At Youden's J threshold, there was a sensitivity of 84% (95% CI, 75–93), specificity of 66% (63–71), positive predictive value of 23% (18–29), negative predictive value of 97% (95–99), positive LR of 2.5 (2.1–3.0), and negative LR of 0.24 (0.10–0.38) in the MSH<sub>original</sub> test data set. Classification performance between clinically relevant subgroups in the MSH data set is shown in Table 2. RVEDV model performance differed in patients aged  $\geq 60$  versus <60 years (AUROC, 0.67 versus 0.81; *P*=0.015) and those with LV systolic dysfunction versus normal LV function (AUROC, 0.87 versus 0.75; *P*=0.024).

### Prospective Validation

The RVEDV and RVEF classification models were further tested in a temporal validation data set consisting



**Figure 2. Classification of abnormal right ventricular ejection fraction and end-diastolic volume.**

Receiver operating characteristic curves (blue) and precision recall curves (orange) to classify RV systolic dysfunction (A, left) and RV dilation (B, right) in UK Biobank data set (top), MSH data set (middle), and MSH prospective validation data set (bottom). Red dashed lines on receiver operating characteristic curves represent no model skill, and on precision recall curves represent baseline incidence of disease. AUROC indicates area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; MSH, Mount Sinai Hospital; and RV, right ventricular.

of  $\approx 4$  months of data collected after development of the final models, the MSH<sub>validation</sub> group. For RV systolic dysfunction classification, model AUROC was 0.77 (0.65–0.88), and AUPRC was 0.37 (0.18–0.56). For RV dilation classification, model AUROC was 0.92 (0.81–1.00), and AUPRC was 0.45 (0.06–1.00).

### RVEDV and RVEF Regression Models

We developed models to predict the numerical value of RVEDV and RVEF. They were subsequently fine-tuned in the MSH data set, with test set regression results shown in Figure 3. For RVEF prediction, the mean absolute error between predicted and expected RVEF was clinically reasonably low at 7.8%, the  $R^2=0.360$  for the linear fit between the 2 measures, and there was moderate agreement (concordance correlation coefficient, 0.57). Passing–Bablok regression analysis demonstrated systematic

differences between the ECG-predicted and cMRI RVEF evidenced by significant differences in the regression intercept and slope from 0 and 1, respectively (see Figure 3A). Difference plots for RVEF are shown in Figure 3B. The errors are nonnormally distributed, with ECG overestimating cMRI RVEF at low (pathological) values and ECG underestimating cMRI RVEF at higher values, and 95% prediction CI of  $\pm 14.0\%$ . In the RVEDV model, the mean absolute error was 17.6 mL/m<sup>2</sup>,  $R^2=0.250$ , and the concordance correlation coefficient was 0.43. Like the RVEF models, Passing–Bablok regression demonstrated systematic differences between methods (Figure 3C), and difference plots exhibited nonnormally distributed error with a 95% prediction CI of  $\pm 27.5$  mL/m<sup>2</sup>. ECG overestimated cMRI RVEDV at low values, and ECG underestimated cMRI RVEDV at higher (pathological) values (Figure 3D).

**Table 2. Classification Model Performance, MSH Subgroup Analysis**

|                           | n   | Model AUROC (95% CI)     |          | Model AUROC (95% CI)    |          |
|---------------------------|-----|--------------------------|----------|-------------------------|----------|
|                           |     | MSH RVEDV classification | P value* | MSH RVEF classification | P value* |
| Overall                   | 604 | 0.81 (0.76–0.86)         |          | 0.81 (0.77–0.86)        |          |
| Subgroup                  |     |                          |          |                         |          |
| Female                    | 209 | 0.72 (0.62–0.81)         | 0.40     | 0.79 (0.69–0.90)        | 0.46     |
| Male                      | 395 | 0.77 (0.70–0.83)         |          | 0.84 (0.79–0.89)        |          |
| Age ≥60y                  | 272 | 0.67 (0.57–0.76)         | 0.015    | 0.80 (0.72–0.87)        | 0.58     |
| Age <60y                  | 332 | 0.81 (0.74–0.87)         |          | 0.82 (0.76–0.89)        |          |
| Arrhythmia <sup>†</sup>   | 50  | 0.74 (0.53–0.94)         | 0.56     | 0.81 (0.67–0.94)        | 0.69     |
| Nonarrhythmia             | 476 | 0.80 (0.74–0.86)         |          | 0.84 (0.79–0.89)        |          |
| LVEF <50%                 | 261 | 0.87 (0.81–0.94)         | 0.024    | 0.74 (0.68–0.81)        | 0.93     |
| LVEF ≥50%                 | 340 | 0.75 (0.67–0.83)         |          | 0.75 (0.53–0.98)        |          |
| BMI ≥30kg/ht <sup>2</sup> | 152 | 0.80 (0.75–0.88)         | 0.79     | 0.88 (0.73–0.86)        | 0.053    |
| BMI <30kg/ht <sup>2</sup> | 432 | 0.80 (0.68–0.91)         |          | 0.79 (0.73–0.86)        |          |
| Race                      |     |                          |          |                         |          |
| White                     | 258 | 0.83 (0.74–0.92)         | 0.83     | 0.82 (0.74–0.91)        | 0.90     |
| Black                     | 101 | 0.78 (0.64–0.92)         |          | 0.82 (0.74–0.91)        |          |
| Other/Unknown             | 245 | 0.82 (0.75–0.89)         |          | 0.79 (0.72–0.88)        |          |

BMI indicates body mass index; and LVEF, left ventricular ejection fraction; RVEDV, right ventricular end-diastolic volume; and RVEF, right ventricular ejection fraction.

\*P value for comparison of area under the receiver operating characteristic curves between subgroups.

<sup>†</sup>Defined as normal sinus rhythm vs sinus with atrial/ventricular ectopy, atrial fibrillation/flutter, or other.

## Saliency Mapping

RV dilation and RV dysfunction classification model saliency mapping examples from the MSH<sub>original</sub> test set are shown in Figure 4. Qualitative review of several saliency mapping examples suggested that P waves and QRS complexes in leads II, V1, and V5/6 are particularly influential.

## Survival Analysis

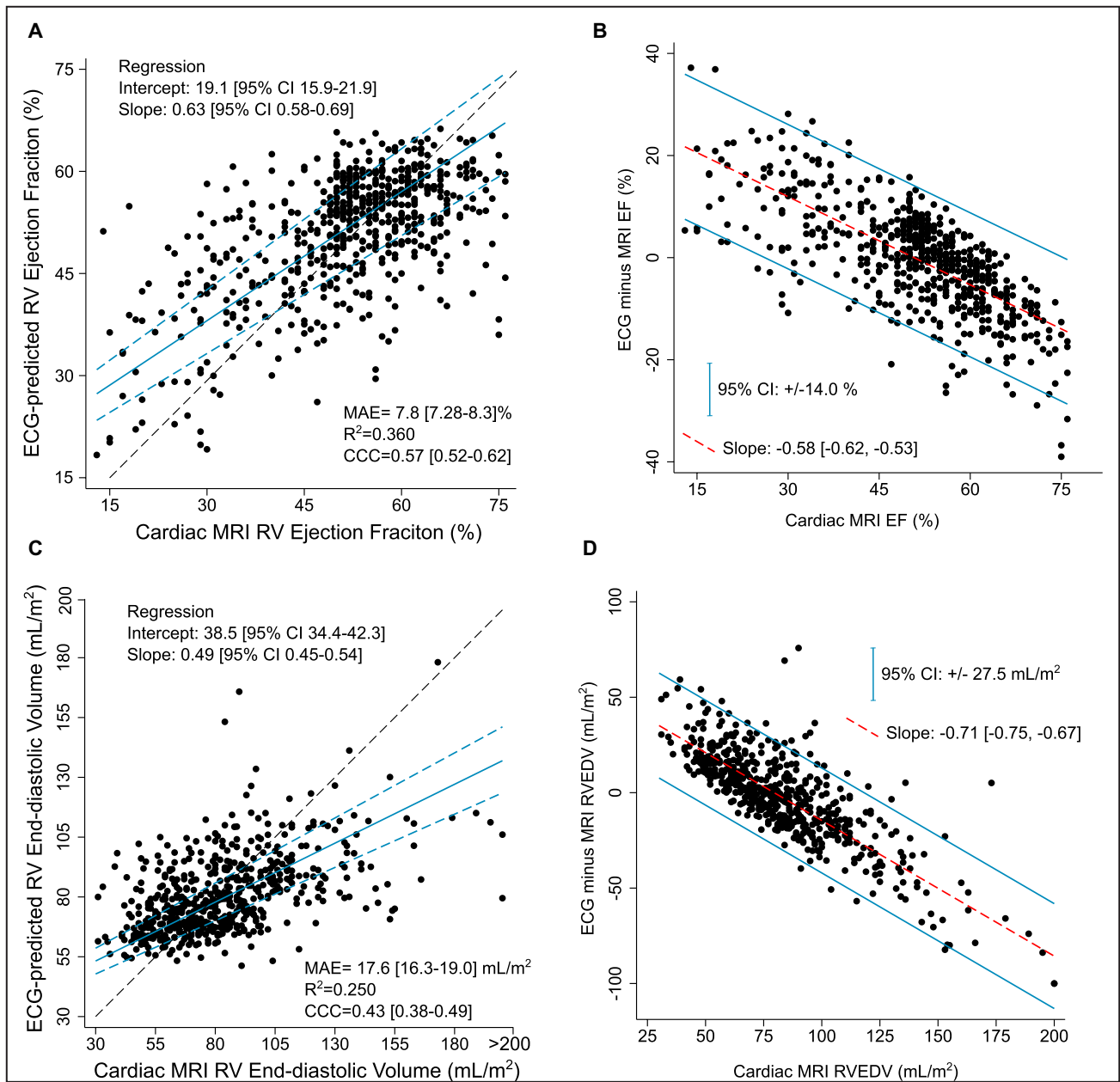
Follow-up information was available for 575 of 604 (95%) patients in the MSH<sub>original</sub> test cohort, with a median follow-up time of 2.27 (interquartile range, 0.67–5.03) years. The combined outcome of death or heart transplantation occurred in 65 of 604 (10.8%; n=59 deaths and n=6 heart transplants) patients. Kaplan–Meier survival curves are illustrated in Figure 5. Multivariable modeling excluding RVEF as a predictor identified the optimal baseline model including LVEF, age (dichotomized at ≥60 years), and hospitalized status to be by Akaike information criterion (Table S4). Addition of either ECG-predicted or MRI-quantified RVEF to the baseline model modestly improved Akaike information criterion and C-statistic, though MRI-quantified RVEF resulted in greater improvement of the baseline model compared with ECG-predicted RVEF (Table S5). The final multivariable survival model including ECG-predicted RVEF is shown in Table 3. In a model including MRI-quantified LVEF and age

and hospitalized status, ECG-predicted RVEF was significantly associated with increased risk of death/transplant (hazard ratio, 1.40 for each 10% decrease in RVEF;  $P=0.031$ ). No significant interaction was identified between LVEF and RVEF. The proportional hazards assumption and linearity assumptions for the remaining continuous variables (LVEF and RVEF) were satisfied (Figure S2). We repeated the analysis using MRI-quantified RVEF and found similar results (Table S6).

## DISCUSSION

In summary, we used DL on ECGs to quantify RVEDV and RVEF on the basis of paired cMRI as reference standard. We first developed this model in a large cohort of mostly healthy participants in the UKBB registry, then further fine-tuned the model in a smaller cohort of clinically indicated paired ECG and cMRI samples from MSH. Finally, we prospectively validated the model in the subsequent ≈4 months after model development. The classification models to detect RV dilation (RVEDV ≥120 mL/m<sup>2</sup>) and RV dysfunction (RVEF ≤40%) have good performance in a clinical cohort. Regression models demonstrated moderate predictive ability of ECG for numerical estimation of cMRI-measured RVEDV and RVEF, and ECG-predicted RVEF was associated with worse short-term freedom from death/heart transplantation. Saliency mapping suggested





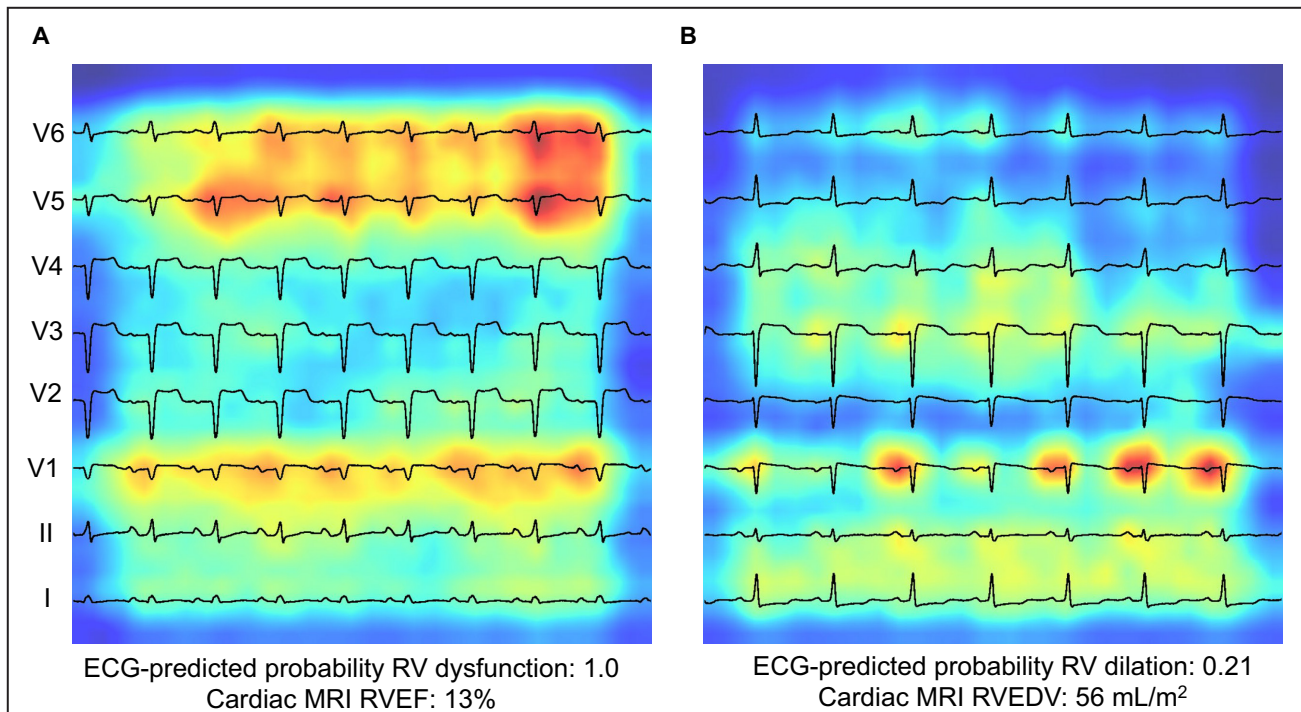
**Figure 3. RVEF and RVEDV regression models, MSH data set.**

MSH<sub>original</sub> regression model with ECG-predicted vs cMRI-expected scatter plots to predict numerical RVEF (A) and numerical RVEDV (C). Scatter plots plotted with line of perfect concordance (dashed black line) and Passing–Bablok regression best-fit line (solid blue) with 95% CIs (dashed blue line). Difference plots analysis of RVEF (B) and RVEDV (D) with regression slope (red line) and 95% CIs of the errors (blue line) displayed. CCC indicates concordance correlation coefficient; EF ejection fraction; MAE, mean absolute error; MRI, magnetic resonance imaging; MSH, Mount Sinai Hospital; RV, right ventricle; and RVEDV, right ventricular end-diastolic volume.

P-wave and QRS complexes of leads II, V1, and V5/V6 as particularly important. This represents a significant advancement in prediction of RV enlargement by 12-lead ECG, which is known to be insensitive by traditional methods (eg, voltage criteria),<sup>25</sup> and this work is the first to show that RV volumetric markers of dilation and dysfunction can be predicted from ECG.

We find the classification models to be more clinically applicable, as they provide good test performance characteristics at clinically important RVEF and RVEDV

values on the basis of AUROC and AUPRC metrics. We provided test statistics at probability thresholds on the basis of statistical criteria (Youden’s J) for the MSH RVEDV and RVEF classification models, but it is important to note that selection of an appropriate threshold is a clinical decision based on the use-case of the model. It is also worth noting that AUROC may be deceptively inflated in the setting of class imbalance, as is the case in the UKBB data set, and therefore AUPRC is an important metric to consider in this case. However,



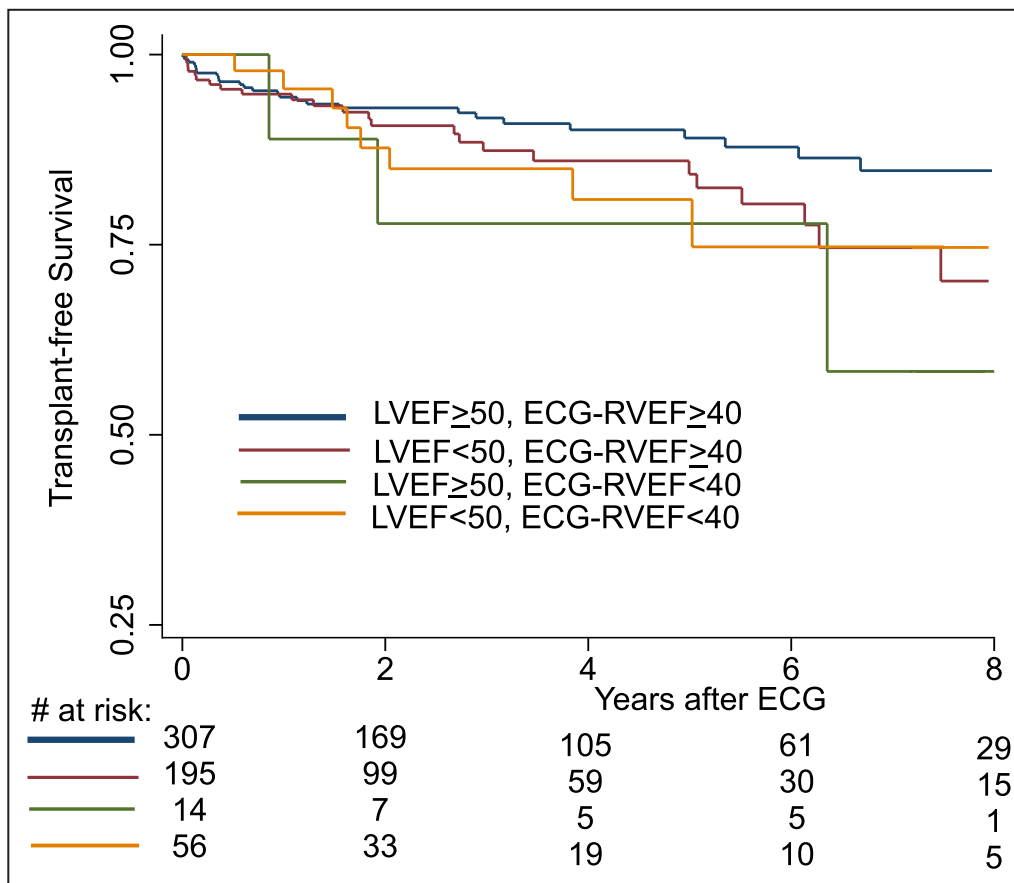
**Figure 4. Saliency mapping.**

Saliency mapping ECG examples from the MSH<sub>original</sub> test set. Leads are arranged from bottom to top: lead I, II, V1-V6. Increasing shades of red indicate increasingly influential pixels in the model. **A** shows a true-positive example of RV dysfunction. The P and QRS portions of V1, V5, and V6 are particularly influential. **B** shows a true-negative example of RV dilation. QRS complex of lead I and P-wave in V1 is particularly influential. MRI indicates magnetic resonance imaging; RVEF, right ventricular ejection fraction; and RVEDV, right ventricular end-diastolic volume.

AUPRC should be interpreted in relation to the baseline prevalence of disease, which explains why AUPRC is lower in the UKBB cohort compared with the MSH cohort. As an example of this, at a baseline prevalence of RV dysfunction of  $\approx 18\%$  in the MSH data set, examination of the AUPRC reveals that at a prediction threshold recall (sensitivity) of  $\approx 50\%$ , the precision (positive predictive value) is  $\approx 60\%$ . In other words, an ECG in this population could alert the practitioner to half of all cases of RV dysfunction in the population, and of cases identified, over half would have true RV dysfunction. Contrast this with the performance of the UKBB model, with only a 1% baseline prevalence of RV dysfunction. At a similar recall (sensitivity) of 50%, the precision (positive predictive value) is  $\approx 20\%$ , demonstrating a higher false-positive rate despite a higher AUROC value, but still a large enrichment in positive cases over random selection from the baseline population. Future clinical implementation should take into consideration the population distribution of disease when determining the clinical utility of this algorithm and in development of appropriate probability thresholds for model prediction.

In contrast to the classification model performance, our regression model performance is less clinically applicable at this time, as difference plot analysis suggested there is high bias at the extremes

of measurement (which are the pathological cases) and wide limits of agreement between ECG and MRI. Passing–Bablok regression confirms that the 2 measurement methods are not equivalent. However, the fact that a correlation is detected at all is itself a novel breakthrough in DL-ECG analysis. Additionally, ECG-predicted RVEF added predictive performance to a multivariable survival model and was associated with combined outcome of death/heart transplant in short-term follow-up. This demonstrates the potential clinical utility of this algorithm for prediction of patient outcome. It is interesting that both ECG-predicted RVEF and MRI-quantified RVEF were associated with poorer short-term survival, whereas LVEF was not in our multivariable model. Low RVEF is a strong risk factor for poor outcome in several conditions, including myocarditis and ischemic cardiomyopathy. We suspect that LVEF was not a significant predictor of outcome for reasons related to cohort and follow-up time. As the most common cause of RV dysfunction is pulmonary hypertension due to left heart dysfunction, RV dysfunction is observed to be a “common final pathway” in heart failure.<sup>1</sup> Thus, it is possible that RVEF may be a more sensitive marker for short-term outcomes than LVEF, and the statistical observation of LVEF nonsignificance, therefore, might be explained as a statistical



**Figure 5. Unadjusted survival curves.**

Kaplan–Meier survival curves by presence of LVEF and RVEF dysfunction. Note that curves are truncated at 8 years due to low sample size after this time point. Risk table is displayed at bottom of figure. LVEF indicates left ventricular ejection fraction; and ECG-RVEF, electrocardiogram-predicted right ventricular ejection fraction.

adjustment for a mediator (ie, RVEF) in the causal pathway of poor short-term outcome. It is possible that in a larger sample size or longer follow-up time LVEF would become a significant predictor, as is well described in established risk prediction models.<sup>23</sup> Further exploration of this is beyond the scope of this report, as our intention was to demonstrate clinical utility of ECG-predicted RV.

**Table 3. Survival Analysis for Freedom From Death/Heart Transplant, Bivariable and Multivariable Analyses**

| Model variable                          | HR (95% CI)      | P value |
|---|------------------|---------|
| ECG-predicted RVEF (every 10% decrease) | 1.40 (1.03–1.91) | 0.031   |
| cMRI LVEF (every 10% decrease)          | 0.98 (0.83–1.17) | 0.86    |
| Age ≥60 y                               | 2.81 (1.65–4.79) | <0.001  |
| Hospitalized at cMRI                    | 1.86 (1.10–3.15) | 0.022   |

cMRI indicates cardiac magnetic resonance imaging; HR, hazard ratio; LVEF, left ventricular ejection fraction; and RVEF, right ventricular ejection fraction.

DL-ECG models have been shown to predict LV systolic dysfunction, elevated LV mass, and primarily left-sided structural heart disease.<sup>14,16,26,27</sup> ECG detection of LVEF has been incorporated into clinical workflow to increase detection of asymptomatic low LVEF in the primary care setting.<sup>28</sup> The right heart is largely uninvestigated in this field. We have previously shown that DL-ECG analysis can detect the qualitative presence of RV systolic dysfunction or RV dilation by 2-dimensional echocardiography.<sup>14</sup> However, the limitations of 2-dimensional echocardiography in quantitative assessment of the right ventricle are well described, and our model represents an advancement over this prior work through training on a diverse group of data sets with paired reference standard cMRI RV measurements, thus allowing us to *quantify* RV size and systolic function by an important clinical metric. Although further work is needed, this algorithm might eventually be incorporated into on-cart analysis to provide a useful screening tool for RV health that is validated against reference standard metrics. This may

allow for better assessments of the RV with a simple test to further inform the need for further diagnostics or therapies. Additional possible benefits of developing non-cMRI-based automated methods for RV quantification include a possible reduction in variability of measurement, as manually contoured RV volumes from cardiac MRI have increased variability compared with the left ventricle with coefficients of variation in the 6% to 9% range for RVEDV, and 8% to 12% for RVEF.<sup>29–31</sup>

In the MSH cohort, LV and RV dysfunction were often found together, but our model performed well in the subgroups with and without concomitant LV dysfunction. This suggests that our model can distinguish RV systolic dysfunction separately from LV dysfunction. Similarly, equivalent performance in those with and without arrhythmia suggests that the model recognizes RV systolic dysfunction independently of cardiomyopathy associated with abnormal rhythm. Our subgroup analysis suggests the classification models have robust performance across subgroups with described differences in ECG morphology including body mass index, sex, and race, which suggests demographic parity in potentially vulnerable groups that may be affected by model unfairness. A more detailed examination of model fairness is an important future step in model deployment.<sup>32</sup> Finally, we did identify differences in the RVEDV (but not RVEF model) in performance across age and LVEF subgroups. This should be addressed in future work through expansion of the training set to include additional patients from these subgroups in a multicenter fashion.

Although the depolarization and repolarization patterns on the 12-lead ECG are dominated by the higher myocardial mass of the left ventricle, prior work has described crude correlations between ECG pattern and RV systolic function,<sup>33–35</sup> which provide a potential scientific rationale for our findings. In our study, saliency mapping demonstrated the importance of lead V1 in evaluating the right ventricle, which may be physiologically explained by its proximity to the right atrium and right ventricle. Further study to differentiate how and why these models can evaluate RV dysfunction independently from LV dysfunction are needed to enhance the explainability of DL-ECG RV models.

We chose a 2-dimensional CNN architecture for deep learning because this theoretically allows us to input images of ECGs (eg, .pdf files, pictures). Not all centers may store ECG data as 1-dimensional vectors, which makes an image-based approach more universally acceptable. Additionally, a 2-dimensional approach allows us to take advantage of availability of 2D CNN models pretrained on millions of images which may enhance performance. The use of 2-dimensional image-based input for DL-ECG analysis may have

superior performance compared with 1-dimensional signal-based ECG input methods,<sup>36</sup> though several groups also report excellent results using a non-image-based approach.<sup>15–17</sup> Further research to determine the generalizability of our models using different sources of ECG images is required.

This work should be viewed in the light of some limitations. First, model training and fine-tuning was performed in only 2 data sets in 2 settings, which may limit the generalizability of these models to other relevant cardiac conditions. The overwhelming majority of patients in the MSH cohort had indication of cardiomyopathy for cMRI, and a higher-than-expected number of patients were admitted to the hospital at the time of MRI. This was likely in part due to the tight inclusion period imposed on paired MRI-ECG samples to ensure that ECGs were valid representations of RV health. However, as MRI is more commonly a scheduled outpatient procedure, it is possible that other important patient subgroups are underrepresented. Additionally, we did not subtype the cardiomyopathy due to limitations in the reporting fields in which the indication for MRI was extracted. Therefore, external validation in other populations where ECG patterns are likely to differ, for instance, in pulmonary hypertension, congenital heart disease, and across different forms of cardiomyopathy, will ensure the generalizability and clinical applicability of these models. The LVEF ECG model used for pretraining was trained on a similar cohort of patients from similar institutions. In experiments performed after our initial model trainings, we identified that  $\approx 30\%$  of cases were present in both the original LVEF model and the MSH RVEF and EDV models (in the MSH<sub>original</sub> test set, 183/604 patients overlapped with the original LVEF pretrained model). We examined the model performance between groups with overlap and without and found no difference in model performance (AUROC for overlapping data was 0.82 [95% CI, 0.72–0.89], AUROC for nonoverlapping samples was 0.81 [95% CI, 0.75–0.87]; data not shown above). We therefore believe the effect of this possible “data leakage” (in which the model makes predictions for data on which it has already been trained) is minimal. Furthermore, the MSH<sub>validation</sub> set, due to temporal inclusion criteria, does not contain patients included in the original LVEF model and exhibited generally stable performance characteristics. The theoretical effect of data leakage is to inflate the performance of the model in the validation set through overfitting to nongeneralizable aspects of the overlapping data. We believe we did not observe this because the labels between the current and pretrained models, RVEF and RVEDV versus LVEF, are biologically different enough phenomena that this overlap does not make a difference. Still, this concern, in addition to other concerns raised about the need to ensure generalizability across health care



systems and patient groups, emphasizes the need for external validation of this model.

Second, we did not compare performance across different CNN architectures because the chosen Densenet-201 architecture had been pretrained to detect LVEF from prior works,<sup>14</sup> which we demonstrated had superior accuracy compared with a nonpretrained model (Table S1). To judge whether alternate architectures are superior, we would need to retrain several model architectures across a large corpus of ECGs for LVEF prediction and subsequently use this model as pretraining for the current task. These experiments are important from a data science perspective, but beyond the scope of the current project goals to demonstrate the feasibility of DL-ECG analysis to quantify metrics of RV size and systolic function. Similarly, our decision to include 8 of 12 available leads in the model was based on a priori knowledge that the augmented leads aVF, aVL, and aVR, and lead III are linear transformations of the other leads and should theoretically not provide additional information to the model. As these models have been built on prior works, we are unable to test model performance across different lead configurations without significant model retraining. Further work to describe the optimal lead configuration for DL-ECG models should be an area of further research.

Third, we trained models on the large UKBB data set, in which RV volumes were automatically contoured through a cMRI segmentation algorithm. This method may be more prone to error and systematically biased compared with clinician labeling, and this method differed from the MSH cohort in which we derived RV measurements from clinician-verified cMRI reports. We performed a comparison between manual and automated contouring methods and found minimal mean differences between methods, though some variability was observed (Figure S1). We fine-tuned the model on MSH data to account for the differences in distribution between data sets, but it is possible that we might have obtained better performance if the UKBB data set were contoured by clinicians. However, this is infeasible due to the large number of studies (>40 000) in UKBB. It is also important to note that we did not conduct experiments to demonstrate the added value of using the large amounts of UKBB data that went into model training (total  $n > 42\,000$  ECGs with paired cMRI). Our group has demonstrated that CNN architectures continue to increase performance with increasing amounts of ECG data even above 500 000 training examples.<sup>18</sup> Given these prior works, we did not seek to demonstrate the additive value of UKBB data, which would require multiple model retrainings at different training sample sizes. This is an interesting direction of research in the data science realm and would be useful to establish minimum sample size estimates for neural network training in the ECG analysis.

Finally, our regression models demonstrate the potential of DL analysis to derive unseen functional data from the ECG; however, currently they exhibit clinically unacceptable limits of agreement with cMRI. The regression models demonstrate the tendency for the networks to predict closer to the population mean. The model architecture we selected is complex enough to fit the training data; however, high variance still exists. Further steps to improve prediction may require experimentation with different model architectures and training on additional pathological cases (increasing sample size), but it may also be that the ECG is too limited to allow for accurate numerical prediction, in which case the classification models may still be useful to rule in or rule out cases. Performance appeared stable in prospective validation, but due to small sample size of the validation group and a shift in baseline prevalence, we may miss subtle differences in performance. As these models improve, research into the usefulness of serial DL-ECG analysis to trend RV size or function should be explored, as trending of RV metrics over time is an important clinical application of this technology that may reduce the need for serial cMRI.

In summary, we developed and validated a DL algorithm to predict RV dilation or systolic dysfunction as defined by reference standard cMRI metrics. This method can provide novel RV quantification from a simple and globally available tool. These prediction tools may be useful in decreasing the need for costly cMRI, with implications for disease screening, risk stratification, and progression of disease. Future work to improve accuracy of prediction to validate findings in heterogeneous clinical populations, and clinical trials to show improved workflow and outcomes are required.

## ARTICLE INFORMATION

Received July 25, 2023; accepted November 20, 2023.

### Affiliations

Division of Pediatric Cardiology, Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY (S.Q.D., L.R.B., R.H.P., B.D.G.); The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY (S.Q.D., A.V., V.T.M., A.W.C., R.D., G.N.N.); Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY (S.Q.D., B.D.G.); Helmsley Center for Electrophysiology at The Mount Sinai Hospital, New York, NY (J.L.); Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY (A.W.C., B.D.G., R.D.); Mount Sinai Heart, Icahn School of Medicine at Mount Sinai, New York, NY (J.N.); Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT (R.K.); Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, CT (R.K.); Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT (R.K.); Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT (R.K.); Division of Cardiovascular Critical Care, Department of Cardiac and Thoracic Surgery, West Virginia University, Morgantown, WV (A.S.); Biomedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, NY (H.G.); and The Division of Data Driven and Digital Medicine (D3M), Department of Medicine (G.N.N.), Icahn School of Medicine at Mount Sinai, New York, NY.



## Sources of Funding

This study was supported by the National Heart, Lung, and Blood Institute, National Institutes of Health (R01HL155915), and National Center for Advancing Translational Sciences, National Institutes of Health (CTSA grant UL1TR004419).

## Disclosures

Dr Nadkarni reports consultancy agreements with AstraZeneca, BioVie, GLG Consulting, Pensieve Health, Reata, Renalytix, Siemens Healthineers, and Variant Bio; reports research funding from Goldfinch Bio and Renalytix; honoraria from AstraZeneca, BioVie, Lexicon, Daiichi Sankyo, Menarini Health, and Reata; reports patents or royalties with Renalytix; owns equity and stock options in Pensieve Health and Renalytix as a scientific cofounder; owns equity in Verici Dx; has received financial compensation as a scientific board member and advisor to Renalytix; serves on the advisory board of Neurona Health; and serves in an advisory or leadership role for Pensieve Health and Renalytix. R. Do reports receiving grants from AstraZeneca; receiving grants and nonfinancial support from Goldfinch Bio; being a scientific cofounder, consultant, and equity holder for Pensieve Health (pending); and being a consultant for Variant Bio, all not related to this work. The remaining authors have no disclosures to report.

## Supplemental Material

Tables S1–S6  
Figures S1–S2

## REFERENCES

- Voelkel NF, Quaife RA, Leinwand LA, Barst RJ, McGoon MD, Meldrum DR, Dupuis J, Long CS, Rubin LJ, Smart FW, et al. Right ventricular function and failure. *Circulation*. 2006;114:1883–1891. doi: [10.1161/CIRCULATIONAHA.106.632208](https://doi.org/10.1161/CIRCULATIONAHA.106.632208)
- Purmah Y, Lei LY, Dykstra S, Mikami Y, Cornhill A, Satriano A, Flewitt J, Rivest S, Sandomato R, Seib M, et al. Right ventricular ejection fraction for the prediction of major adverse cardiovascular and heart failure-related events. *Circ Cardiovasc Imaging*. 2021;14:e011337. doi: [10.1161/CIRCIMAGING.120.011337](https://doi.org/10.1161/CIRCIMAGING.120.011337)
- Geva T. Indications for pulmonary valve replacement in repaired tetralogy of Fallot: the quest continues. *Circulation*. 2013;128:1855–1857. doi: [10.1161/CIRCULATIONAHA.113.005878](https://doi.org/10.1161/CIRCULATIONAHA.113.005878)
- Dong Y, Pan Z, Wang D, Lv J, Fang J, Xu R, Ding J, Cui X, Xie X, Wang X, et al. Prognostic value of cardiac magnetic resonance–derived right ventricular remodeling parameters in pulmonary hypertension. *Circ Cardiovasc Imaging*. 2020;13:e010568. doi: [10.1161/CIRCIMAGING.120.010568](https://doi.org/10.1161/CIRCIMAGING.120.010568)
- Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr*. 2015;28:1–39.e14. doi: [10.1016/j.echo.2014.10.003](https://doi.org/10.1016/j.echo.2014.10.003)
- 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. doi: [10.1016/j.echo.2010.03.019](https://doi.org/10.1016/j.echo.2010.03.019)
- Lai WW, Gauvreau K, Rivera ES, Saleeb S, Powell AJ, Geva T. Accuracy of guideline recommendations for two-dimensional quantification of the right ventricle by echocardiography. *Int J Cardiovasc Imaging*. 2008;24:691–698. doi: [10.1007/s10554-008-9314-4](https://doi.org/10.1007/s10554-008-9314-4)
- Addetia K, Muraru D, Badano LP, Lang RM. New directions in right ventricular assessment using 3-dimensional echocardiography. *JAMA Cardiol*. 2019;4:936–944. doi: [10.1001/jamacardio.2019.2424](https://doi.org/10.1001/jamacardio.2019.2424)
- Dragulescu A, Grosse-Wortmann L, Fackoury C, Mertens L. Echocardiographic assessment of right ventricular volumes: a comparison of different techniques in children after surgical repair of tetralogy of Fallot. *Eur Heart J Cardiovasc Imaging*. 2012;13:596–604. doi: [10.1093/ejechocard/jer278](https://doi.org/10.1093/ejechocard/jer278)
- Petmezas G, Stefanopoulos L, Kilintzis V, Tzavelis A, Rogers JA, Katsaggelos AK, Maglaveras N. State-of-the-art deep learning methods on electrocardiogram data: systematic review. *JMIR Med Inform*. 2022;10:e38454. doi: [10.2196/38454](https://doi.org/10.2196/38454)
- Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, Bell JD, Boulwood C, Collins R, Conroy MC, et al. The UK biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun*. 2020;11:2624. doi: [10.1038/s41467-020-15948-9](https://doi.org/10.1038/s41467-020-15948-9)
- Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson*. 2018;20:65. doi: [10.1186/s12968-018-0471-x](https://doi.org/10.1186/s12968-018-0471-x)
- UK Biobank 12-lead (at rest) ECG. 2015. Accessed April 6, 2023. [https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/12lead\\_ecg\\_explan\\_doc.pdf](https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/12lead_ecg_explan_doc.pdf)
- 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](https://doi.org/10.1016/j.jcmg.2021.08.004)
- 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](https://doi.org/10.1038/s41591-018-0240-2)
- Elias P, Poterucha TJ, Rajaram V, Moller LM, Rodriguez V, Bhave S, Hahn RT, Tison G, Abreau SA, Barrios J, et al. Deep learning electrocardiographic analysis for detection of left-sided valvular heart disease. *J Am Coll Cardiol*. 2022;80:613–626. doi: [10.1016/j.jacc.2022.05.029](https://doi.org/10.1016/j.jacc.2022.05.029)
- Ribeiro AH, Ribeiro MH, Paixão 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](https://doi.org/10.1038/s41467-020-15432-4)
- Vaid A, Jiang J, Sawant A, Lerakis S, Argulian E, Ahuja Y, Lampert J, Charney A, Greenspan H, Narula J, et al. A foundational vision transformer improves diagnostic performance for electrocardiograms. *NPJ Digit Med*. 2023;6:108. doi: [10.1038/s41746-023-00840-9](https://doi.org/10.1038/s41746-023-00840-9)
- Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015;10:e0118432. doi: [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432)
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845. doi: [10.2307/2531595](https://doi.org/10.2307/2531595)
- Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, part I. *J Clin Chem Clin Biochem*. 1983;21:709–720. doi: [10.1515/cclm.1983.21.11.709](https://doi.org/10.1515/cclm.1983.21.11.709)
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160. doi: [10.1177/096228029900800204](https://doi.org/10.1177/096228029900800204)
- Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, et al. The Seattle heart failure model: prediction of survival in heart failure. *Circulation*. 2006;113:1424–1433. doi: [10.1161/circulationaha.105.584102](https://doi.org/10.1161/circulationaha.105.584102)
- Piña IL, Jimenez S, Lewis EF, Morris AA, Onwuanyi A, Tam E, Ventura HO. Race and ethnicity in heart failure: JACC focus seminar 8/9. *J Am Coll Cardiol*. 2021;78:2589–2598. doi: [10.1016/j.jacc.2021.06.058](https://doi.org/10.1016/j.jacc.2021.06.058)
- Hancock EW, Deal BJ, Mirvis DM, Okin P, Kligfield P, Gettes LS. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part V: electrocardiogram changes associated with cardiac chamber hypertrophy: a scientific statement from the American Heart Association electrocardiography and arrhythmias committee, council on clinical cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology. *Circulation*. 2009;119:e251–e261. doi: [10.1161/CIRCULATIONAHA.108.191097](https://doi.org/10.1161/CIRCULATIONAHA.108.191097)
- Ulloa-Cerna AE, Jing L, Pfeifer JM, Raghunath S, Ruhl JA, Rocha DB, Leader JB, Zimmerman N, Lee G, Steinhilb SR, et al. rECHOmment: 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](https://doi.org/10.1161/CIRCULATIONAHA.121.057869)

27. Vaid A, Argulian E, Lerakis S, Beaulieu-Jones BK, Krittanawong C, Klang E, Lampert J, Reddy VY, Narula J, Nadkarni GN, et al. Multi-center retrospective cohort study applying deep learning to electrocardiograms to identify left heart valvular dysfunction. *Commun Med (Lond)*. 2023;3:24. doi: [10.1038/s43856-023-00240-w](https://doi.org/10.1038/s43856-023-00240-w)
28. 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](https://doi.org/10.1038/s41591-021-01335-4)
29. Blalock SE, Banka P, Geva T, Powell AJ, Zhou J, Prakash A. Interstudy variability in cardiac magnetic resonance imaging measurements of ventricular volume, mass, and ejection fraction in repaired tetralogy of Fallot: a prospective observational study. *J Magn Reson Imaging*. 2013;38:829–835. doi: [10.1002/jmri.24050](https://doi.org/10.1002/jmri.24050)
30. Grothues F, Moon JC, Bellenger NG, Smith GS, Klein HU, Pennell DJ. Interstudy reproducibility of right ventricular volumes, function, and mass with cardiovascular magnetic resonance. *Am Heart J*. 2004;147:218–223. doi: [10.1016/j.ahj.2003.10.005](https://doi.org/10.1016/j.ahj.2003.10.005)
31. Mooij CF, de Wit CJ, Graham DA, Powell AJ, Geva T. Reproducibility of MRI measurements of right ventricular size and function in patients with normal and dilated ventricles. *J Magn Reson Imaging*. 2008;28:67–73. doi: [10.1002/jmri.21407](https://doi.org/10.1002/jmri.21407)
32. Corbett-Davies S, Gaebler JD, Nilforoshan H, Shroff R, Goel S. The measure and mismeasure of fairness: a critical review of fair machine learning. *arXiv preprint arXiv*. 2018. doi: [10.48550/arXiv.1808.00023](https://doi.org/10.48550/arXiv.1808.00023)
33. Adams JC, Nelson MR, Chandrasekaran K, Jahangir A, Srivathsan K. Novel ECG criteria for right ventricular systolic dysfunction in patients with right bundle branch block. *Int J Cardiol*. 2013;167:1385–1389. doi: [10.1016/j.ijcard.2012.04.025](https://doi.org/10.1016/j.ijcard.2012.04.025)
34. Alonso P, Andrés A, Rueda J, Buendía F, Igual B, Rodríguez M, Osa A, Arnau MA, Salvador A. Value of the electrocardiogram as a predictor of right ventricular dysfunction in patients with chronic right ventricular volume overload. *Rev Esp Cardiol*. 2015;68:390–397. doi: [10.1016/j.rec.2014.04.019](https://doi.org/10.1016/j.rec.2014.04.019)
35. Daralammouri Y, Qaddumi J, Ayoub K, Abu-Hantash D, Al-sadi MA, Ayaseh RM, Azamtta M, Sawalmeh O, Hamdan Z. Pathological right ventricular changes in synthesized electrocardiogram in end-stage renal disease patients and their association with mortality and cardiac hospitalization: a cohort study. *BMC Nephrol*. 2022;23:79. doi: [10.1186/s12882-022-02707-9](https://doi.org/10.1186/s12882-022-02707-9)
36. Sangha V, Mortazavi BJ, Haimovich AD, Ribeiro AH, Brandt CA, Jacoby DL, Schulz WL, Krumholz HM, Ribeiro ALP, Khera R. Automated multilabel diagnosis on electrocardiographic images and signals. *Nat Commun*. 2022;13:1583. doi: [10.1038/s41467-022-29153-3](https://doi.org/10.1038/s41467-022-29153-3)