

1 **Detection of Left Ventricular Systolic Dysfunction from Electrocardiographic Images**

2 Veer Sangha^{1*}, Arash A Nargesi MD, MPH^{2,3*}, Lovedeep S Dhingra MBBS³, Akshay Khunte¹,
3 Bobak J Mortazavi PhD^{4,5}, Antônio H Ribeiro PhD⁶, Evgeniya Banina MD⁷, Oluwaseun Adeola
4 MD, MPH⁸, Nadish Garg MD⁹, Cynthia A Brandt MD, MPH^{10,11}, Edward J Miller MD, PhD³,
5 Antonio Luiz J Ribeiro MD, PhD^{12,13}, Eric J Velazquez MD, PhD³, Luana Giatti MD, PhD¹⁴,
6 Sandhi M Barreto MD, PhD¹⁵, Murilo Foppa MD, PhD¹⁶, Neal Yuan MD^{17,18}, David Ouyang
7 MD^{19,20}, Harlan M Krumholz MD, SM^{5,3,21}, Rohan Khera MD, MS^{3,5,22}

8

9 ¹Department of Computer Science, Yale University, New Haven, CT, USA

10 ²Heart and Vascular Center, Brigham and Women's Hospital, Harvard Medical School, Boston,
11 MA, USA

12 ³Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University, New
13 Haven, CT, USA

14 ⁴Department of Computer Science & Engineering, Texas A&M University, College Station, TX,
15 USA

16 ⁵Center for Outcomes Research and Evaluation (CORE), Yale New Haven Hospital, New
17 Haven, CT, USA

18 ⁶Department of Information Technology, Uppsala University, Uppsala, Sweden

19 ⁷Internal Medicine Department, Lake Regional Hospital Health, Osage Beach, MO, USA

20 ⁸Methodist Cardiology Clinic of San Antonio, San Antonio, TX, USA

21 ⁹Heart and Vascular Institute, Memorial Hermann Southeast Hospital, Houston, TX, USA

22 ¹⁰Department of Emergency Medicine, Yale University, New Haven, CT, USA

23 ¹¹VA Connecticut Healthcare System, West Haven, CT, USA

24 ¹²Telehealth Center and Cardiology Service, Hospital das Clínicas, Universidade Federal de
25 Minas Gerais, Belo Horizonte, Brazil

26 ¹³Department of Internal Medicine, Faculdade de Medicina, Universidade Federal de Minas
27 Gerais, Belo Horizonte, Brazil

28 ¹⁴Department of Preventive Medicine, School of Medicine and Hospital das Clínicas,
29 Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

30 ¹⁵Department of Preventive Medicine, School of Medicine, and Hospital das Clínicas,
31 Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

32 ¹⁶Postgraduate Studies Program in Cardiology and Division of Cardiology, Hospital de Clinicas
33 de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

34 ¹⁷Department of Medicine, University of California, San Francisco, San Francisco, CA, USA

35 ¹⁸Section of Cardiology, San Francisco Veterans Affairs Medical Center, San Francisco, CA,
36 USA

37 ¹⁹Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles,
38 CA, USA

39 ²⁰Division of Artificial Intelligence in Medicine, Cedars-Sinai Medical Center, Los Angeles, CA,
40 USA

41 ²¹Department of Health Policy and Management, Yale School of Public Health, New Haven, CT,
42 USA

43 ²²Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New
44 Haven, CT, USA

45

46 *Contributed equally

47

48 **Correspondence to:** Rohan Khera, MD, MS

49 195 Church St, 5th Floor, New Haven, CT 06510, 203-764-5885; rohan.khera@yale.edu;

50 @rohan_khera

51

52

53 **ABSTRACT**

54 **Background:** Left ventricular (LV) systolic dysfunction is associated with over 8-fold increased
55 risk of heart failure and a 2-fold risk of premature death. The use of electrocardiogram (ECG)
56 signals in screening for LV systolic dysfunction is limited by their availability to clinicians. We
57 developed a novel deep learning-based approach that can use ECG images for the screening of
58 LV systolic dysfunction.

59 **Methods:** Using 12-lead ECGs plotted in multiple different formats, and corresponding
60 echocardiographic data recorded within 15 days from the Yale-New Haven Hospital (YNHH)
61 during 2015-2021, we developed a convolutional neural network algorithm to detect LV ejection
62 fraction < 40%. The model was validated within clinical settings at YNHH as well as externally
63 on ECG images from Cedars Sinai Medical Center in Los Angeles, CA, Lake Regional Hospital
64 (LRH) in Osage Beach, MO, Memorial Hermann Southeast Hospital in Houston, TX, and
65 Methodist Cardiology Clinic of San Antonio, TX. In addition, it was validated in the prospective
66 Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). Gradient-weighted class activation
67 mapping was used to localize class-discriminating signals in ECG images.

68 **Results:** Overall, 385,601 ECGs with paired echocardiograms were used for model development.
69 The model demonstrated high discrimination power across various ECG image formats and
70 calibrations in internal validation (area under receiving operation characteristics [AUROC] 0.91,
71 area under precision-recall curve [AUPRC] 0.55), and external sets of ECG images from Cedars
72 Sinai (AUROC 90, AUPRC 0.53), outpatient YNHH clinics (AUROC 0.94, AUPRC 0.77), LRH
73 (AUROC 0.90, AUPRC 0.88), Memorial Hermann Southeast Hospital (AUROC 0.91, AUPRC
74 0.88), Methodist Cardiology Clinic (AUROC 0.90, AUPRC 0.74), and ELSA-Brasil cohort
75 (AUROC 0.95, AUPRC 0.45). An ECG suggestive of LV systolic dysfunction portended over
76 27-fold higher odds of LV systolic dysfunction on TTE (OR 27.5, 95% CI, 22.3-33.9 in the held-

77 out set). Class-discriminative patterns localized to the anterior and anteroseptal leads (V2-V3),
78 corresponding to the left ventricle regardless of the ECG layout. A positive ECG screen in
79 individuals with LV ejection fraction $\geq 40\%$ at the time of initial assessment was associated with
80 a 3.9-fold increased risk of developing incident LV systolic dysfunction in the future (HR 3.9,
81 95% CI 3.3-4.7, median follow-up 3.2 years).

82 **Conclusions:** We developed and externally validated a deep learning model that identifies LV
83 systolic dysfunction from ECG images. This approach represents an automated and accessible
84 screening strategy for LV systolic dysfunction, particularly in low-resource settings.

85 **ABBREVIATIONS AND ACRONYMS**

86	LV	Left ventricle
87	ECG	Electrocardiography
88	AI	Artificial Intelligence
89	LVEF	Left ventricular ejection fraction
90	YNHH	Yale New Haven Hospital
91	TTE	Transthoracic echocardiography
92	LRH	Lake Regional Hospital
93	ELSA-Brasil	ELSA-Brasil, Estudo Longitudinal de Saúde do Adulto
94		(The Brazilian Longitudinal Study of Adult Health)
95	Grad-CAM	Gradient-weighted Class Activation Mapping
96	AUROC	Area under receiving operation characteristics
97	AUPRC	Area under precision recall curve

98 **CLINICAL PERSPECTIVE**

99 **What is New?**

- 100 • A convolutional neural network model that accurately identifies LV systolic dysfunction
101 from ECG images across subgroups of age, sex, and race.
- 102 • The model shows robust performance across multiple institutions and health settings,
103 both applied to ECG image databases as well as directly uploaded single ECG images to
104 a web-based application by clinicians.
- 105 • The approach provides information for both screening of LV systolic dysfunction and its
106 risk based on ECG images alone.

107 **What are the clinical implications?**

- 108 • Our model represents an automated screening strategy for LV systolic dysfunction on a
109 variety of ECG layouts.
- 110 • With availability of ECG images in practice, this approach overcomes implementation
111 challenges of deploying an interoperable screening tool for LV systolic dysfunction in
112 resource-limited settings.
- 113 • This model is available in an online format to facilitate real-time screening for LV
114 systolic dysfunction by clinicians.

115 INTRODUCTION

116 Left ventricular (LV) systolic dysfunction is associated with over 8-fold increased risk of
117 subsequent heart failure and nearly 2-fold risk of premature death.¹ While early diagnosis can
118 effectively lower this risk,²⁻⁴ individuals are often diagnosed after developing symptomatic
119 disease due to lack of effective screening strategies.⁵⁻⁷ The diagnosis traditionally relies on
120 echocardiography, a specialized imaging modality that is resource intensive to deploy at scale.^{8,9}
121 Algorithms using raw signals from electrocardiography (ECG) have been developed as a strategy
122 to detect LV systolic dysfunction.¹⁰⁻¹² However, clinicians, particularly in remote settings, do not
123 have access to ECG signals. The lack of interoperability in signal storage formats from ECG
124 devices further limits the broad uptake of such signal-based models.¹³ The use of ECG images is
125 an opportunity to implement interoperable screening strategies for LV systolic dysfunction.

126 We previously developed a deep learning approach of format-independent inference from
127 real-world ECG images.¹⁴ The approach can interpretably diagnose cardiac conduction and
128 rhythm disorders using any layout of real-world 12-lead ECG images and can be accessed on
129 web- or application-based platforms. Extension of this artificial intelligence (AI)-driven
130 approach to ECG images to screen for LV systolic dysfunction could rapidly broaden access to a
131 low cost, easily accessible, and scalable diagnostic approach to underdiagnosed and undertreated
132 at-risk populations. This approach adapts deep learning for end-users, without disruption of data
133 pipelines or clinical workflow. Moreover, the ability to add localization of predictive cues in the
134 ECG images relevant to the LV can improve the uptake of these models in clinical practice.¹⁵

135 In this study, we present a model for accurate identification of LV ejection fraction
136 (LVEF) less than 40%, a threshold with therapeutic implications, based on ECG images. We

137 developed, tested, and externally validated this approach using paired ECG-echocardiographic
138 data from large academic hospitals, rural hospital systems, and a prospective cohort study.

139

140 **METHODS**

141 The Yale Institutional Review Board reviewed the study, which approved the study protocol and
142 waived the need for informed consent as the study represents a secondary analysis of existing
143 data. The data cannot be shared publicly though an online version of the model is publicly
144 available for research use at <https://www.cards-lab.org/ecgvision-lv>.

145 **Data Source for Model Development**

146 We used 12-lead ECG signal waveform data from the Yale New Haven Hospital (YNHH)
147 collected between 2015 and 2021. These ECGs were recorded as standard 12-lead recordings
148 sampled at a frequency of 500 Hz for 10 seconds. These were recorded on multiple different
149 machines and a majority were collected using Philips PageWriter machines and GE MAC
150 machines. Among patients with an ECG, those with a corresponding transthoracic
151 echocardiogram (TTE) within 15 days of obtaining the ECG were identified from the YNHH
152 electronic health records. LVEF values were extracted based on a cardiologist's read of the
153 nearest TTE to each ECG. To augment the evaluation of models built on an image dataset
154 generated from this YNHH signal waveform, six sets of ECG image datasets were used for
155 external validation.

156 **Data Preprocessing**

157 All ECGs were analyzed to determine whether they had 10 seconds of continuous recordings
158 across all 12 leads. The 10 second samples were preprocessed with a one second median filter,
159 which was subtracted from the original waveform to remove baseline drift in each lead,

160 representing processing steps pursued by ECG machines before generating printed output from
161 collected waveform data.

162 ECG signals were transformed into ECG images using the python library `ecg-plot`,¹⁶ and
163 stored at 100 DPI. Images were generated with a calibration of 10 mm/mV, which is standard for
164 printed ECGs in most real-world settings. In sensitivity analyses, we evaluated model
165 performance on images calibrated at 5 and 20 mm/mV. All images, including those in train,
166 validation, and test sets, were converted to greyscale, followed by down-sampling to 300x300
167 pixels regardless of their original resolution using Python Image Library (PIL v9.2.0). To ensure
168 that the model was adaptable to real-world images, which may vary in formats and the layout of
169 leads, we created a dataset with different plotting schemes for each signal waveform recording
170 (**Figure 1**). This strategy has been used to train a format-independent image-based model for
171 detecting conduction and rhythm disorders as well as the hidden label of gender.¹⁴ The model in
172 this study learned ECG lead-specific information based on the label regardless of the location of
173 the lead.

174 Four formats of images were included in the training image dataset (**Figure 1**). The first
175 format was based on the standard printed ECG format in the United States, with four 2.5 second
176 columns printed sequentially on the page. Each column contained 2.5 second intervals from three
177 leads. The full 10-second recording of the lead I signal was included as the rhythm strip. The
178 second format, a two-rhythm format, added lead II as an additional rhythm strip to the standard
179 format. The third layout was the alternate format which consisted of two columns, the first with
180 six simultaneous 5-second recordings from the limb leads, and the second with six simultaneous
181 5-second recordings from the precordial leads, without a corresponding rhythm lead. The fourth
182 format was a shuffled format, which had precordial leads in the first two columns and limb leads

183 in the third and fourth. All images were rotated a random amount between -10 and 10 degrees
184 before being input into the model to mimic variations seen in uploaded ECGs and to aid in
185 prevention of overfitting.

186 The process of converting ECG signals to images was independent of model
187 development, ensuring that the model did not learn any aspects of the processing that generated
188 images from the signals. All ECGs were converted to images in all different formats without
189 conditioning on clinical labels. The validation required uploaded images to be upright, cropped
190 to the waveform region, with no brightness and contrast consideration as long as the waveform is
191 distinguishable from the background and lead labels are discernible.

192 **Experimental Design**

193 Each included ECG had a corresponding LVEF value from its nearest TTE within 15 days of
194 recording. Low LVEF was defined as $LVEF < 40\%$, the cutoff used as an indication for most
195 guideline-directed pharmacotherapy for heart failure.⁴ Patients with at least one ECG within 15
196 days of its nearest TTE were randomly split into training, validation, and held-out test patient
197 level sets (85%, 5%, 10%, **Figure S1**). This sampling was stratified by whether a patient had
198 ever had $LVEF < 40\%$ to ensure cases of preserved and reduced LVEF were split proportionally
199 among the sets. In the training cohort, all ECGs within 15 days of a TTE were included for all
200 patients to maximize the data available. In validation and testing cohorts, only one ECG was
201 included per patient to ensure independence of observations in the assessment of performance
202 metrics. This ECG was randomly chosen amongst all ECGs within 15 days of a TTE.
203 Additionally, to ensure that model learning was not affected by the relatively lower frequency of
204 $LVEF < 40\%$, higher weights were given to these cases at the training stage based on the
205 effective number of samples class sampling scheme.¹⁷

206 **Model Training**

207 We built a convolutional neural network model based on the EfficientNet-B3 architecture,¹⁸
208 which previously demonstrated an ability to learn and identify both rhythm and conduction
209 disorders, as well as the hidden label of gender in real-world ECG images.¹⁴ The EfficientNet-B3
210 model requires images to be sampled at 300 x 300 square pixels, includes 384 layers, and has over
211 10 million trainable parameters (**Figure S2**). We utilized transfer learning by initializing model
212 weights as the pretrained EfficientNet-B3 weights used to predict the six physician-defined
213 clinical labels and gender from Sangha et al.¹⁴ We first only unfroze the last four layers and
214 trained the model with a learning rate of 0.01 for 2 epochs, and then unfroze all layers and trained
215 with a learning rate of 5×10^{-6} for 6 epochs. We used an Adam optimizer, gradient clipping, and a
216 minibatch size of 64 throughout training. The optimizer and learning rates were chosen after
217 hyperparameter optimization. For both stages of training the model, we stopped training when
218 validation loss did not improve in 3 consecutive epochs.

219 We trained and validated our model on a generated image dataset that had equal numbers
220 of standard, two-rhythm, alternate, and standard shuffled images (**Figure 1**). In sensitivity
221 analyses, the model was validated on three novel ECG layouts constructed from the held-out set
222 to assess its performance on ECG formats not encountered in the training process. These novel
223 ECG outlines included three-rhythm (with leads I, II, and V1 as the rhythm strip), no rhythm,
224 and rhythm on top formats (with lead I as the rhythm strip located above the 12-lead, **Figure S3**).
225 Additional sensitivity analyses were performed using ECG images calibrated at 5, 10, and 20
226 mm/mV (**Figure S4**). A custom class-balanced loss function (weighted binary cross-entropy)
227 based on the effective number of samples was used given the lower frequency of the LVEF <
228 40% label relative to those with an LVEF \geq 40%.

229 **External validation**

230 We pursued a series of validation studies. These represented both clinical and population-based
231 cohort studies. Clinical validation represented non-synthetic image datasets from clinical settings
232 spanning (1) consecutive patients undergoing outpatient echocardiography at the Cedars Sinai
233 Medical Center in Los Angeles, CA, and (2) stratified convenience samples of LV systolic
234 dysfunction and non-LV systolic dysfunction ECGs from four different settings (a) outpatient
235 clinics of YNHH, (b) inpatient admissions at Lake Regional Hospital (LRH) in Osage Beach,
236 MO, (c) inpatient admissions at Memorial Hermann Southeast Hospital in Houston, TX, (d)
237 outpatient visits and inpatient admissions at Methodist Cardiology Clinic in San Antonio, TX. In
238 addition, we validated our approach in the prospective cohort from Brazil, the Brazilian
239 Longitudinal Study of Adult Health (ELSA-Brasil),¹⁹ with protocolized ECG and
240 echocardiogram in study participants.

241 Inclusion and exclusion criteria for external validation sets were similar to the internal
242 YNHH dataset. Patients were limited to those having a 12-lead ECG within 15 days of a TTE
243 with reported LVEF. For patients with more than one TTE in this interval, the LVEF from the
244 nearest TTE was used for analysis.

245 At Cedars Sinai, all index ECG images from consecutive patients undergoing outpatient
246 visits between, January through March 2019, representing 879 individuals, including 99 with
247 LVEF < 40%, were included. These analyses were performed in a fully federated and blinded
248 fashion without access to any of the ECG data to the algorithm's developers.

249 For the other clinical validation sites, a stratified convenience sample enriched for low
250 LVEF was drawn. This was done to evaluate the broad use in a clinical setting by practicing
251 clinicians without access to a research dataset. Our preliminary assessment of LV systolic

252 dysfunction prevalence in outpatient and inpatient settings were 10% and 20%, respectively. We
253 sought to achieve twice this prevalence in our external validation data in these sites to ensure our
254 performance was not driven by patients with preserved LVEF and that the model could detect
255 those with LV systolic dysfunction. Specifically, a 1:4 ratio of ECGs corresponding to LVEF <
256 40% and $\geq 40\%$ was sought at three of the four sites (YNHH, Memorial Hermann Southeast
257 Hospital, and Methodist Cardiology Clinic). At the fourth site, LRH, a 1:2 ratio was requested to
258 better measure the model's discriminative ability in an inpatient-only setting.

259 In addition to the clinical validation studies, where concurrent ECG and echocardiogram
260 are always clinically indicated, imposing a selection of the population, we evaluated our model
261 in the ELSA-Brasil study, a community-based prospective cohort in Brazil that obtained ECG
262 and echocardiography from participants on the enrollment visit between 2008-2010. This set
263 included data from 2,577 individuals, including 30 from individuals with LVEF < 40%.

264 Before validation, patient identifiers, ECG measurements, and reported diagnoses were
265 removed from all ECG images. The differences in ECG layouts and the procedures for validation
266 are described in further detail in the **Online Supplement**. Deidentified samples of ECG images
267 are presented in **Figure S5** (Cedars Sinai Medical Center), **Figure S6** (YNHH and LRH),
268 **Figure S7** (Memorial Hermann Southeast Hospital), and **Figure S8** (Methodist Cardiology
269 Clinic), and images are available from the authors upon request.

270 **Localization of Model Predictive Cues**

271 We used Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight which portions
272 of an image were important for the prediction of LVEF < 40%.²⁰ We calculated the gradients on
273 the final stack of filters in our EfficientNet-B3 model for each prediction and performed a global
274 average pooling of the gradients in each filter, emphasizing those that contributed to a prediction.

275 We then multiplied these filters by their importance weights and combined them across filters to
276 generate Grad-CAM heatmaps, which we overlaid on the original ECG images. We averaged
277 class activation maps among 100 positive cases with the most confident model predictions for
278 LVEF < 40% across ECG formats to determine the most important image areas for the prediction
279 of low LVEF. We took an arithmetic mean across the heatmaps for a given image format, and
280 overlaid this average heatmap across a representative ECG to understand it in context. The
281 activation map, a 10x10 array was upsampled to the original image size using the bilinear
282 interpolation built into TensorFlow v 2.8.0. We also evaluated the Grad-CAM for individual
283 ECGs to evaluate the consistency of the information on individual examples.

284 **Statistical Analysis**

285 Categorical variables were presented as frequency and percentages, and continuous variables as
286 means and standard deviations or median and interquartile range, as appropriate. Model
287 performance was evaluated in the held-out test set and external ECG image datasets. We used
288 area under the receiver operator characteristic (AUROC) to measure model discrimination. The
289 cut-off for binary prediction of LV systolic dysfunction was set at 0.10 for all internal and
290 external validations, based on the threshold that achieved a sensitivity of over 90% in the internal
291 validation set. We also assessed area under precision recall curve (AUPRC), sensitivity,
292 specificity, positive predictive value (PPV), negative predictive value (NPV), and diagnostic odds
293 ratio. 95% CIs for AUROC and AUPRC were calculated using DeLong's algorithm and
294 bootstrapping with 1000 variations for each estimate, respectively.^{21,22} Model performance was
295 assessed across demographic subgroups and ECG outlines, as described above. We conducted
296 further sensitivity analyses of model performance across ECG calibrations, PR intervals, and after
297 excluding paced rhythms, conduction disorders, atrial fibrillation, and atrial flutter. Moreover, we

298 assessed the association of the model's predicted probability of LV systolic dysfunction across
299 LVEF categories.

300 Next, we evaluated the future development of LV systolic dysfunction in time-to-event
301 models using a Cox proportional hazards model. In this analysis, we took the first temporal ECG
302 from the patients in the held-out test set, and then modeled the first development of LVEF < 40%
303 across the groups of patients who screened positive but did not have concurrent LV systolic
304 dysfunction (false positives), and those that screened negative (true negative) from this first ECG,
305 with censored at death or end of study period in June 2021. Additionally, we computed an
306 adjusted hazard ratio that accounted for differences in age, sex, and baseline LVEF at the time of
307 index screening for visualization of survival trends. Analytic packages used in model
308 development and statistical analysis are reported in **Table S1**. All model development and
309 statistical analyses were performed using Python 3.9.5 and the level of significance was set at an
310 alpha of 0.05.

311

312 **RESULTS**

313 **Study Population**

314 Out of the 2,135,846 ECGs obtained between 2015 to 2021, 440,072 were from patients who had
315 TTEs within 15 days of obtaining the ECG. Overall, 433,027 had a complete ECG recording,
316 representing 10 seconds of continuous recordings across all 12 leads. These ECGs were drawn
317 from 116,210 unique patients and were split into train, validation, and test sets at a patient level
318 (**Figure S1**).

319 A total of 116,210 individuals with 385,601 ECGs constituted the study population,
320 representing those included in the train, validation, test sets. Individuals whose ECGs were used

321 for model development had a median age of 68 years (IQR 56, 78) at the time of ECG recording,
322 and 59,282 (51.0%) were women. Overall, 75,928 (65.3%) were non-Hispanic white, 14,000
323 (12.0%) non-Hispanic Black, 9,349 (8.0%) Hispanic, and 16,843 (14.5%) were from other races.
324 A total of 56,895 (14.8%) ECGs had a corresponding echocardiogram with an LVEF below
325 40%, 36,669 (9.5%) had an LVEF greater than or equal to 40% but less than 50%, and 292,037
326 (75.7%) had LVEF 50% or greater (**Table S2**).

327 **Detection of LV Systolic Dysfunction**

328 The model's AUROC for detecting LVEF < 40% on the held-out test set composed of standard
329 images was 0.91 and its AUPRC was 0.55 (**Figure 2**). A probability threshold for predicting
330 LVEF < 40% was chosen based on a sensitivity of 0.90 or higher in the validation subset. With
331 this threshold, the model had sensitivity and specificity of 0.89 and 0.77 in the held-out test set,
332 and PPV and NPV of 0.26 and 0.99, respectively. Overall, an ECG suggestive of LV systolic
333 dysfunction portended over 27-fold higher odds (OR 27.5, 95% CI, 22.3 – 33.9) of LV systolic
334 dysfunction on TTE (**Table 1**). The model's performance was comparable across subgroups of
335 age, sex, and race (**Table 1** and **Figure 2**). Moreover, across successive deciles of the model
336 predicted probabilities, the proportion of individuals with LV systolic dysfunction increased,
337 while the mean LVEF decreased (**Figure S9**).

338 **Model Performance Across ECG Formats and Calibrations**

339 The model performance was comparable across the four original layouts of ECG images in the
340 held-out set with AUROC of 0.91 in detecting concurrent LV systolic dysfunction (**Table S3**).
341 The model had a sensitivity of 0.89 and a positive prediction conferred 26- to 27-fold higher
342 odds of LV systolic dysfunction on the standard and the three variations of the data. In sensitivity
343 analyses, the model demonstrated similar performance in detecting LV systolic dysfunction from

344 novel ECG formats that were not encountered before, with AUROC between 0.88-0.91 (**Table**
345 **S4**).

346 The model performance was also consistent across ECG calibrations with an AUROC
347 between 0.88 and 0.91 on ECG calibrations of 5, 10, and 20 mm/mV and AUROC 0.908 (0.899
348 – 0.918) and AUPRC of 0.538 (0.503 – 0.573) with mixed calibrations in the held-out test set.
349 The mixed calibration was generated with a random sample of 5 mm/mV and 20 mm/mV
350 calibrations from the highest and lowest quartiles of voltages, respectively, in lead I (together
351 representing 25% of the sample from the test set), along with 10 mm/mV (remaining 75% of test
352 set) (**Table S5**). Further sensitivity analyses demonstrated consistent model performance on
353 ECGs (a) without prolonged PR interval (AUROC 0.920 and AUPRC 0.537, **Table S6**), (b)
354 without paced rhythms (AUROC 0.908, AUPRC 0.519, **Table S7**), and (c) without atrial
355 fibrillation, atrial flutter, and conduction disorders (AUROC 0.919, AUPRC 0.536, **Table S8**).
356 Model performance was also consistent across subsets on the held-out test set based on the
357 timing of the ECG relative to the echocardiogram (**Table S9**).

358 **LV Systolic Dysfunction in Model-predicted False Positives**

359 Of the 10,666 ECGs in the held-out test set with an associated LVEF \geq 40% on a proximate
360 echocardiogram, the model classified 2,469 (23.1%) as “false positives”, and 8,197 (76.9%) as
361 true negatives. In further evaluation of false positives, 562 (22.8% of false positives) had
362 evidence of mild LV systolic dysfunction with LVEF between 40-50% on concurrent
363 echocardiography.

364 In this group of individuals, 4,046 patients had at least one follow-up TTE, including
365 1,125 (27.8%) false positives and 2,921 (72.2%) true negatives on the initial index screen. There
366 were 2,665 and 6,083 echocardiograms in the false positive and true negative populations during

367 the follow-up, with the longest follow-up of 6.1 years. Overall, 264 (23.5%) patients with model-
368 predicted positive screen and 199 (6.8%) with negative screen developed new LVEF < 40% over
369 the median follow-up of 3.2 years (IQR 1.8-4.4 years, **Figure 3**). This represented a 3.9-fold
370 higher risk of incident low LVEF based on having a positive screening result (HR 3.9, 95% CI
371 3.3-4.7). After adjustment for age, sex, and LVEF at the time of screening, patients with positive
372 screen had a 2.3-fold higher risk of incident low LVEF (Adjusted HR 2.3, 95% CI 1.9-2.8).

373 **Localization of Predictive Cues for LV Systolic Dysfunction**

374 Class activation heatmaps of the 100 positive cases with the most confident model predictions
375 for reduced LVEF prediction across four ECG layouts are presented in **Figure 4**. For all four
376 formats of images, the region corresponding to leads V2 and V3 were the most important areas
377 for prediction of reduced LVEF. Representative images of Grad-CAM analysis in sampled
378 individuals with positive and negative screens in the held-out test set, and non-synthetic ECG
379 images in validation sites are presented in **Figures S10** and **S11**, respectively.

380 **External Validation**

381 The validation performance of the model was consistent and robust across each of the 6
382 validation datasets (**Figure 5**). The first validation set at Cedars Sinai Medical Center included
383 879 ECGs from consecutive patients who underwent outpatient echocardiography, including 99
384 (11%) individuals with LVEF < 40%. The model demonstrated an AUROC of 0.90 and an
385 AUPRC of 0.53 in this set. Second, a total of 147 ECG images drawn from YNHH outpatient
386 clinics were used for validation and included 27 images (18%) from patients with LVEF < 40%.
387 The model had an AUROC of 0.94 and AUPRC of 0.77 in validation on these images. The third
388 image dataset included ECG images from inpatient visits to the LRH. It included 100 ECG
389 images, with 43 images (43%) from patients with LVEF < 40%, with a model AUROC of 0.90

390 and AUPRC of 0.88. The fourth dataset from Memorial Hermann Southeast Hospital included 50
391 ECG images, 11 (22%) from patients with LVEF < 40%, with a model AUROC and AUPRC of
392 0.91 and 0.88 on these images, respectively. The fifth validation set contained 50 ECG images
393 from the Methodist Cardiology Clinic, which included 11 (20%) ECGs from patients with LVEF
394 < 40%, with model AUROC of 0.90 and AUPRC of 0.74.

395 The sixth set included 2,577 ECGs from prospectively enrolled individuals in the ELSA-
396 Brasil study, including 30 with LVEF < 40%. The model demonstrated an AUROC 0.95 and
397 AUPRC 0.45 on this set. The model performance on these 6 validation sets is outlined in **Table 2**
398 and **Table S10**.

399

400 **DISCUSSION**

401 We developed and externally validated an automated deep learning algorithm that accurately
402 identifies LV systolic dysfunction solely from ECG images. The algorithm has high
403 discrimination and sensitivity, representing characteristics ideal for a screening strategy. It is
404 robust to variations in the layouts of ECG waveforms and detects the location of ECG leads
405 across multiple formats with consistent accuracy, making it suitable for implementation in a
406 variety of settings. Moreover, the algorithm was developed and tested in a diverse population
407 with high performance in subgroups of age, sex, and race, and across geographically dispersed
408 academic and community health systems. It performed well in 6 external validation sites,
409 spanning both clinical settings as well as a prospective cohort study where protocolized
410 echocardiograms were performed concurrently with ECGs. An evaluation of the class-
411 discriminating signals localized it to the anteroseptal and anterior leads regardless of the ECG
412 layout, topologically corresponding to the left ventricle. Finally, among individuals who did not

413 have a concurrently recorded low LVEF, a positive ECG screen was associated with a 3.9-fold
414 increased risk of developing LV systolic dysfunction in the future compared with those with
415 negative screen, which was significant after adjustment for age, sex, and baseline LVEF.
416 Therefore, an ECG image-based approach can represent a screening as well as predictive strategy
417 for LV systolic dysfunction, particularly in low-resource settings.

418 Image-based analysis of ECGs through deep learning represents a novel application of AI
419 to improve clinical care. Convolutional neural networks have previously been designed to detect
420 low LVEF from ECG signals.^{10,11} Although reliance of signal-based models on voltage data is
421 not computationally limited, their use in both retrospective and prospective settings requires
422 access to a signal repository where the ECG data architecture varies by ECG device vendors.
423 Moreover, data are often not stored beyond generating printed ECG images, particularly in
424 remote settings.²³ Furthermore, widespread adoption of signal-based models is limited by the
425 implementation barriers requiring health system-wide investments to incorporate them into
426 clinical workflow, something that may not be available or cost-effective in low-resource settings
427 and, to date, is not widely available in higher resource setting such as the US. The algorithm
428 reported in this study overcomes these limitations by making detection of LV systolic
429 dysfunction from ECGs interoperable across acquisition formats and directly available to
430 clinicians who only have access to ECG images. Since scanned ECG images are the most
431 common format of storage and use of electrocardiograms, untrained operators can implement
432 large scale screening through chart review or automated applications to image repositories – a
433 lower resource task than optimizing tools for different machines.

434 The use of ECG images in our model overcomes the implementation challenges arising
435 from black box algorithms. The origin of risk-discriminative signals in precordial leads of ECG

436 images suggests a left ventricular origin of the predictive signals. Moreover, the consistent
437 observation of these predictive signals in the anteroseptal and anterior leads, regardless of the
438 lead location on printed images, also serves as a control for the model predictions. Despite
439 localizing the class-discriminative signals in the image to the left ventricle, heatmap analysis
440 may not necessarily capture all the model predictive features, such as the duration of ECG
441 segments, intervals, or ECG waveform morphologies which might have been used in model
442 predictions. However, visual representations that are consistent with clinical knowledge could
443 explain parts of the model prediction process and address the hesitancy in uptake of these tools in
444 clinical practice.²⁴

445 An important finding was the significantly increased risk of incident LV systolic
446 dysfunction among patients with model-predicted positive screen but LVEF \geq 40% on
447 concurrent echocardiography. These findings demonstrate an electrocardiographic signature that
448 may precede the development of echocardiographic evidence of LV systolic dysfunction. This
449 was previously reported in signal-based models,¹⁰ further suggesting that the detection of LV
450 systolic dysfunction on ECG images represents a similar underlying pathophysiological process.
451 These observations suggest a potential role for AI-based ECG models in risk stratification for
452 future development of cardiovascular disease.²⁵

453 Our study has certain limitations that merit consideration. First, we developed this model
454 among patients with both ECGs and echocardiograms. Therefore, the training population
455 selected likely had a clinical indication for echocardiography, differing from the broader real-
456 world use of the algorithm for screening tests for LV systolic dysfunction among those without
457 any clinical disease. The excellent performance of our algorithm across demographic subgroups
458 and the validation population would suggest robustness and generalizability of the effects though

459 prospective assessments in the intended screening setting are warranted. Second, the model
460 performance may vary by degree of LV systolic dysfunction. Though we chose an LVEF
461 threshold of 40% due to its therapeutic implications, such as an indication for disease-modifying
462 guideline-directed medical therapies,⁴ the model identifies individuals with mild dysfunction.
463 This may highlight a shared signature of LV systolic dysfunction among those with LVEF<40%,
464 and with LVEF of 40-50%, but could also represent the lack of precision of LVEF measurement
465 by echocardiography relative to more precise approaches, such as magnetic resonance
466 imaging.^{26,27} Third, while we incorporated four ECG formats during its development and
467 demonstrated that the model had a consistent performance on a range of commonly used and
468 novel layouts that were not included in the development, we cannot ascertain whether it
469 maintains performance on every novel format. Fourth, while the model development pursues
470 preprocessing the ECG signal for plotting images, these represent standard processes performed
471 before ECG images are generated and/or printed by ECG machines. Therefore, any other
472 processing of images is not required for real-world application, as demonstrated in the
473 application of the model to the external validation sets.

474

475 **CONCLUSIONS**

476 We developed an automated algorithm to detect LV systolic dysfunction from ECG images,
477 demonstrating a robust performance across subgroups of patient demographics, ECG formats and
478 calibrations, and clinical practice settings. Given the ubiquitous availability of ECG images, this
479 approach represents a strategy for automated screening of LV systolic dysfunction, especially in
480 resource-limited settings.

481 **ACKNOWLEDGEMENTS**

482 **Author contributions:** RK conceived the study and accessed the data. VS and RK developed the
483 model. VS, AAN, LD, AK, and RK pursued the statistical analysis. VS and AAN drafted the
484 manuscript. All authors provided feedback regarding the study design and made critical
485 contributions to writing of the manuscript. RK supervised the study, procured funding, and is the
486 guarantor.

487 **Funding:** This study was supported by research funding awarded to Dr. Khera by the Yale School
488 of Medicine and grant support from the National Heart, Lung, and Blood Institute of the National
489 Institutes of Health under the award K23HL153775. The funders had no role in the design and
490 conduct of the study; collection, management, analysis, and interpretation of the data; preparation,
491 review, or approval of the manuscript; and decision to submit the manuscript for publication.

492 **Competing Interests:** Dr. Mortazavi reported receiving grants from the National Institute of
493 Biomedical Imaging and Bioengineering, National Heart, Lung, and Blood Institute, US Food
494 and Drug Administration, and the US Department of Defense Advanced Research Projects
495 Agency outside the submitted work; in addition, Dr. Mortazavi has a pending patent on
496 predictive models using electronic health records (US20180315507A1). Antonio H. Ribeiro is
497 funded by *Kjell och Märta Beijer Foundation*. Dr. Krumholz works under contract with the
498 Centers for Medicare & Medicaid Services to support quality measurement programs, was a
499 recipient of a research grant from Johnson & Johnson, through Yale University, to support
500 clinical trial data sharing; was a recipient of a research agreement, through Yale University, from
501 the Shenzhen Center for Health Information for work to advance intelligent disease prevention
502 and health promotion; collaborates with the National Center for Cardiovascular Diseases in
503 Beijing; receives payment from the Arnold & Porter Law Firm for work related to the Sanofi

504 clopidogrel litigation, from the Martin Baughman Law Firm for work related to the Cook Celect
505 IVC filter litigation, and from the Siegfried and Jensen Law Firm for work related to Vioxx
506 litigation; chairs a Cardiac Scientific Advisory Board for UnitedHealth; was a member of the
507 IBM Watson Health Life Sciences Board; is a member of the Advisory Board for Element
508 Science, the Advisory Board for Facebook, and the Physician Advisory Board for Aetna; and is
509 the co-founder of Hugo Health, a personal health information platform, and co-founder of
510 Refactor Health, a healthcare AI-augmented data management company. Dr Ribeiro is supported
511 in part by CNPq (465518/2014-1, 310790/2021-2 and 409604/2022-4) and by FAPEMIG (PPM-
512 00428-17, RED-00081-16 and PPE-00030-21). Mr. Sangha and Dr. Khera are the coinventors of
513 U.S. Provisional Patent Application No. 63/346,610, “Articles and methods for format
514 independent detection of hidden cardiovascular disease from printed electrocardiographic images
515 using deep learning”. Dr. Khera is also a founder of Evidence2Health, a precision health
516 platform to improve evidence-based care.

517 **Model Availability:** The model is available in an online format for research use at
518 <https://www.cards-lab.org/ecgvision-lv>

519

520 REFERENCES

- 521 1. Wang, T. J. *et al.* Natural history of asymptomatic left ventricular systolic dysfunction in
522 the community. *Circulation* **108**, 977–982 (2003).
- 523 2. Srivastava, P. K. *et al.* Heart Failure Hospitalization and Guideline-Directed Prescribing
524 Patterns Among Heart Failure With Reduced Ejection Fraction Patients. *JACC Heart Fail* **9**, 28–
525 38 (2021).
- 526 3. Wolfe, N. K., Mitchell, J. D. & Brown, D. L. The independent reduction in mortality
527 associated with guideline-directed medical therapy in patients with coronary artery disease and
528 heart failure with reduced ejection fraction. *Eur Heart J Qual Care Clin Outcomes* **7**, 416–421
529 (2021).
- 530 4. Heidenreich, P. A. *et al.* 2022 AHA/ACC/HFSA Guideline for the Management of Heart
531 Failure: A Report of the American College of Cardiology/American Heart Association Joint
532 Committee on Clinical Practice Guidelines. *Circulation* 101161CIR0000000000001063 (2022).

- 533 5. Wang, T. J., Levy, D., Benjamin, E. J. & Vasan, R. S. The epidemiology of
534 “asymptomatic” left ventricular systolic dysfunction: implications for screening. *Ann. Intern.*
535 *Med.* **138**, 907–916 (2003).
- 536 6. Vasan, R. S. *et al.* Plasma natriuretic peptides for community screening for left
537 ventricular hypertrophy and systolic dysfunction: the Framingham heart study. *JAMA* **288**, 1252–
538 1259 (2002).
- 539 7. McDonagh, T. A., McDonald, K. & Maisel, A. S. Screening for asymptomatic left
540 ventricular dysfunction using B-type natriuretic Peptide. *Congest. Heart Fail.* **14**, 5–8 (2008).
- 541 8. Galasko, G. I., Barnes, S. C., Collinson, P., Lahiri, A. & Senior, R. What is the most cost-
542 effective strategy to screen for left ventricular systolic dysfunction: natriuretic peptides, the
543 electrocardiogram, hand-held echocardiography, traditional echocardiography, or their
544 combination? *Eur Heart J* **27**, 193–200 (2006).
- 545 9. Atherton, J. J. Screening for left ventricular systolic dysfunction: is imaging a solution?
546 *JACC Cardiovasc Imaging* **3**, 421–428 (2010).
- 547 10. Attia, Z. I. *et al.* Screening for cardiac contractile dysfunction using an artificial
548 intelligence-enabled electrocardiogram. *Nat Med* **25**, 70–74 (2019).
- 549 11. Vaid, A. *et al.* Using Deep-Learning Algorithms to Simultaneously Identify Right and
550 Left Ventricular Dysfunction From the Electrocardiogram. *JACC Cardiovasc Imaging* **15**, 395–
551 410 (2022).
- 552 12. Yao, X. *et al.* Artificial intelligence-enabled electrocardiograms for identification of
553 patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med* **27**, 815–819
554 (2021).
- 555 13. Stamenov, D., Gusev, M. & Armenski, G. *Interoperability of ECG standards*. (IEEE,
556 2018).
- 557 14. Sangha, V. *et al.* Automated multilabel diagnosis on electrocardiographic images and
558 signals. *Nat Commun* **13**, 1583 (2022).
- 559 15. He, J. *et al.* The practical implementation of artificial intelligence technologies in
560 medicine. *Nat Med* **25**, 30–36 (2019).
- 561 16. ECG Plot Python Library. Accessed at <https://pypi.org/project/ecg-plot/> on May 25,
562 2022.
- 563 17. Cui, Y., Jia, M., Lin, T.-Y., Song, Y. & Belongie, S. Class-Balanced Loss Based on
564 Effective Number of Samples. arXiv:1901.05555 (2019).
- 565 18. Mingxing Tan and Quoc V Le. EfficientNet: Rethinking model scaling for convolutional
566 neural networks. *International Conference on Machine Learning, 2019*.
- 567 19. Aquino, E. M. L. *et al.* Brazilian Longitudinal Study of Adult Health (ELSA-Brasil):
568 objectives and design. *Am. J. Epidemiol.* **175**, 315–324 (2012).
- 569 20. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via
570 Gradient-based Localization. *2017 Ieee International Conference on Computer Vision (Iccv)*
571 618–626 (2017).
- 572 21. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two
573 or more correlated receiver operating characteristic curves: a nonparametric approach.
574 *Biometrics* **44**, 837–845 (1988).
- 575 22. Sun, X. & Xu, W. Fast implementation of DeLong’s algorithm for comparing the areas
576 under correlated receiver operating characteristic curves. *IEEE Signal Process. Lett.* **21**, 1389–
577 1393 (2014).

- 578 23. Siontis, K. C., Noseworthy, P. A., Attia, Z. I. & Friedman, P. A. Artificial intelligence-
579 enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol* **18**, 465–
580 478 (2021).
- 581 24. Quer, G., Arnaout, R., Henne, M. & Arnaout, R. Machine Learning and the Future of
582 Cardiovascular Care: JACC State-of-the-Art Review. *J Am Coll Cardiol* **77**, 300–313 (2021).
- 583 25. Maurovich-Horvat, P. Current trends in the use of machine learning for diagnostics
584 and/or risk stratification in cardiovascular disease. *Cardiovasc Res* **117**, e67–e69 (2021).
- 585 26. Pellikka, P. A. *et al.* Variability in Ejection Fraction Measured By Echocardiography,
586 Gated Single-Photon Emission Computed Tomography, and Cardiac Magnetic Resonance in
587 Patients With Coronary Artery Disease and Left Ventricular Dysfunction. *JAMA Netw Open* **1**,
588 e181456 (2018).
- 589 27. Jenkins, C. *et al.* Left ventricular volume measurement with echocardiography: a
590 comparison of left ventricular opacification, three-dimensional echocardiography, or both with
591 magnetic resonance imaging. *Eur Heart J* **30**, 98–106 (2009).
592

Table 1. Performance of model on test images across demographic subgroups in the held-out test set. Abbreviations: PPV, positive predictive value; NPV, negative predictive value; AUROC, area under receiver operating characteristic curve; AUPRC, area under precision recall curve; OR, odds ratio

Labels	Number	PPV	NPV	Specificity	Sensitivity	AUROC	AUPRC
All	11621 (100%)	0.257	0.988	0.769	0.892	0.910 (0.901-0.919)	0.545 (0.511 – 0.579)
Male	5952 (51.2%)	0.285	0.984	0.735	0.897	0.901 (0.889-0.914)	0.583 (0.539 – 0.621)
Female	5668 (48.8%)	0.215	0.991	0.802	0.884	0.917 (0.903-0.932)	0.470 (0.416 – 0.530)
≥ 65	6550 (56.4%)	0.252	0.985	0.717	0.896	0.892 (0.880-0.905)	0.522 (0.480 – 0.561)
< 65	5068 (43.6%)	0.266	0.991	0.833	0.886	0.931 (0.916-0.945)	0.590 (0.534 – 0.655)
Hispanic	942 (8.1%)	0.253	0.992	0.802	0.908	0.926 (0.892-0.961)	0.576 (0.453 – 0.696)
White	7557 (65.0%)	0.261	0.988	0.770	0.895	0.910 (0.898-0.921)	0.537 (0.498 – 0.580)
Black	1417 (12.2%)	0.263	0.984	0.712	0.897	0.899 (0.872-0.925)	0.590 (0.498 – 0.665)
Other	1705 (14.7%)	0.231	0.987	0.787	0.864	0.912 (0.887-0.937)	0.532 (0.437 – 0.625)

* Gender information was not available for 1 patient and age was not available for 3 patient of the total 11,621 patients in the held-out test set

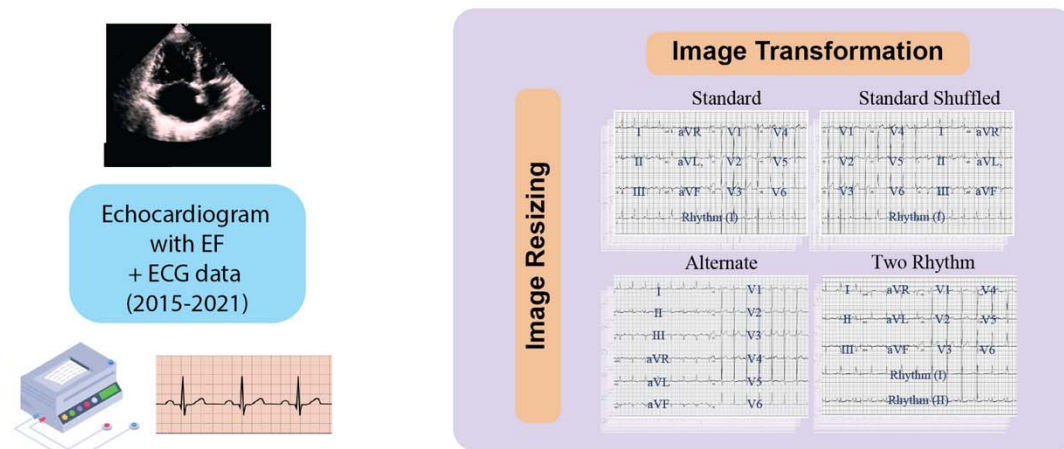
Table 2. Performance of model on external validation datasets. Abbreviations: PPV, positive predictive value; NPV, negative predictive value; AUROC, area under receiver operating characteristic curve; AUPRC, area under precision recall curve; ELSA-Brasil, Estudo Longitudinal de Saúde do Adulto (The Brazilian Longitudinal Study of Adult Health)

Site	PPV	NPV	Specificity	Sensitivity	AUROC	AUPRC
Cedars Sinai Medical Center	0.326	0.979	0.772	0.869	0.902 (0.877 – 0.926)	0.533 (0.432 – 0.640)
Outpatient Clinics of YNHH	0.338	1.000	0.558	1.000	0.946 (0.910 - 0.982)	0.775 (0.605 – 0.916)
LRH	0.538	0.955	0.368	0.977	0.901 (0.843 - 0.959)	0.889 (0.810 – 0.946)
Memorial Hermann Southeast Hospital	0.385	0.958	0.590	0.909	0.918 (0.790 - 1.000)	0.888 (0.699 – 1.000)
Methodist Cardiology Clinic	0.458	1.000	0.667	1.000	0.902 (0.816 - 0.989)	0.738 (0.470 – 0.928)
ELSA-Brasil	0.256	0.996	0.976	0.700	0.949 (0.915 – 0.983)	0.449 (0.290 – 0.651)

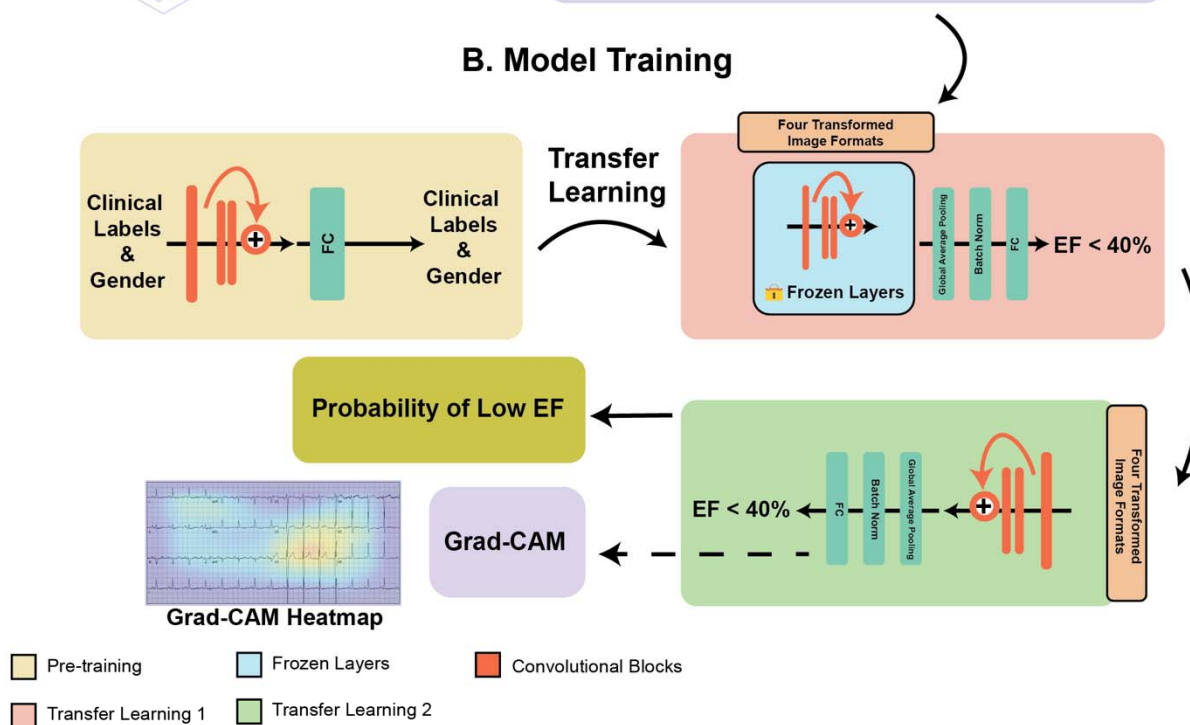
Figure 1. Study Outline A) Data processing, B) Model training, and C) Model validation.

Abbreviations: ECG, electrocardiogram; EF, ejection fraction; FC, fully connected layers; Grad-CAM, gradient-weighted class activation mapping; CT, Connecticut; ELSA-Brasil, Estudo Longitudinal de Saúde do Adulto (The Brazilian Longitudinal Study of Adult Health); MO, Missouri; TX, Texas.

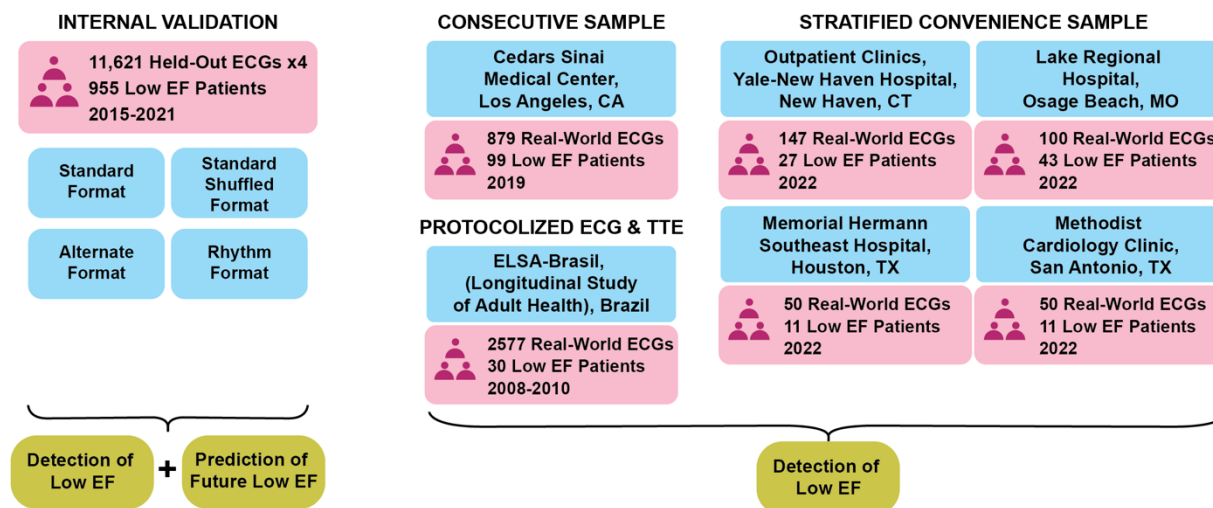
A. Data Processing



B. Model Training



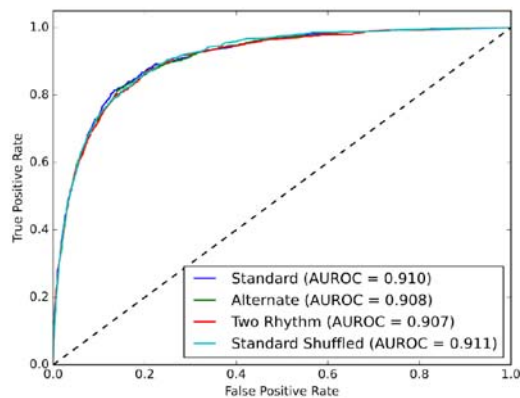
C. Model Validation



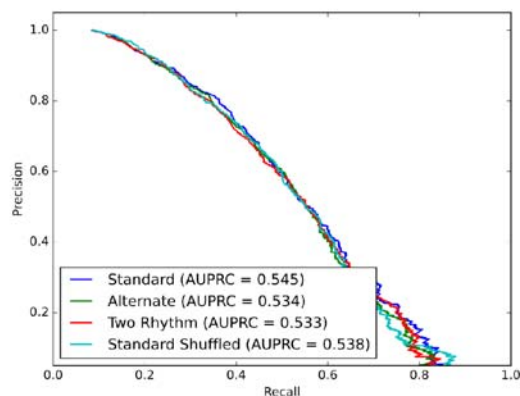
*We pursued a transfer learning strategy in developing the current model from our previous algorithm which was originally trained to detect cardiac rhythm disorders and the hidden label of gender from ECG images. The transfer learning was used as initialization weights for the EfficientNet B3 convolutional neural network being trained to detect LV systolic dysfunction. Other than the weights, clinical and gender labels were not input to the current model.

Figure 2. Model Performance Measures A) Receiver-Operating and B) Precision-Recall Curves on images in held-out test set C) Diagnostic Odds Ratios across age, gender, and race subgroups on standard format images in the held-out test set. Abbreviations: AUROC, area under receiver-operating characteristic curve; AUPRC, area under precision-recall curve.

A. Receiver-Operating Curve



B. Precision-Recall Curve



C. Diagnostic Odds Ratios

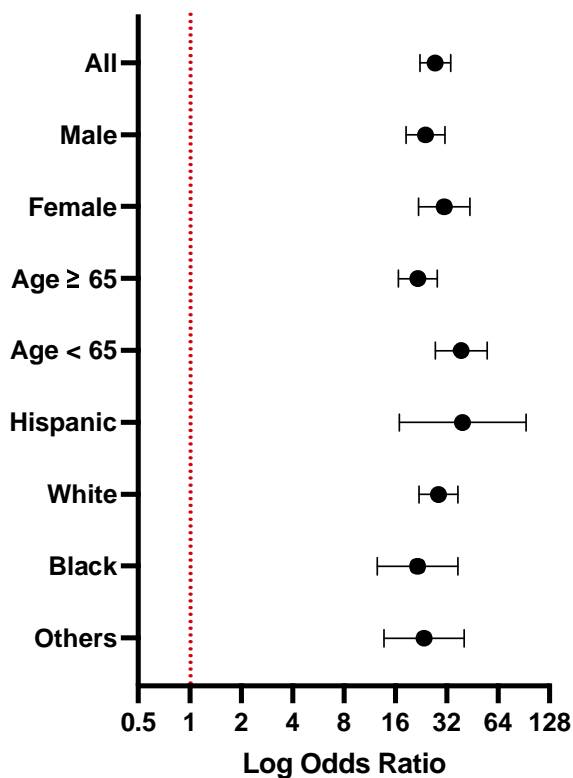
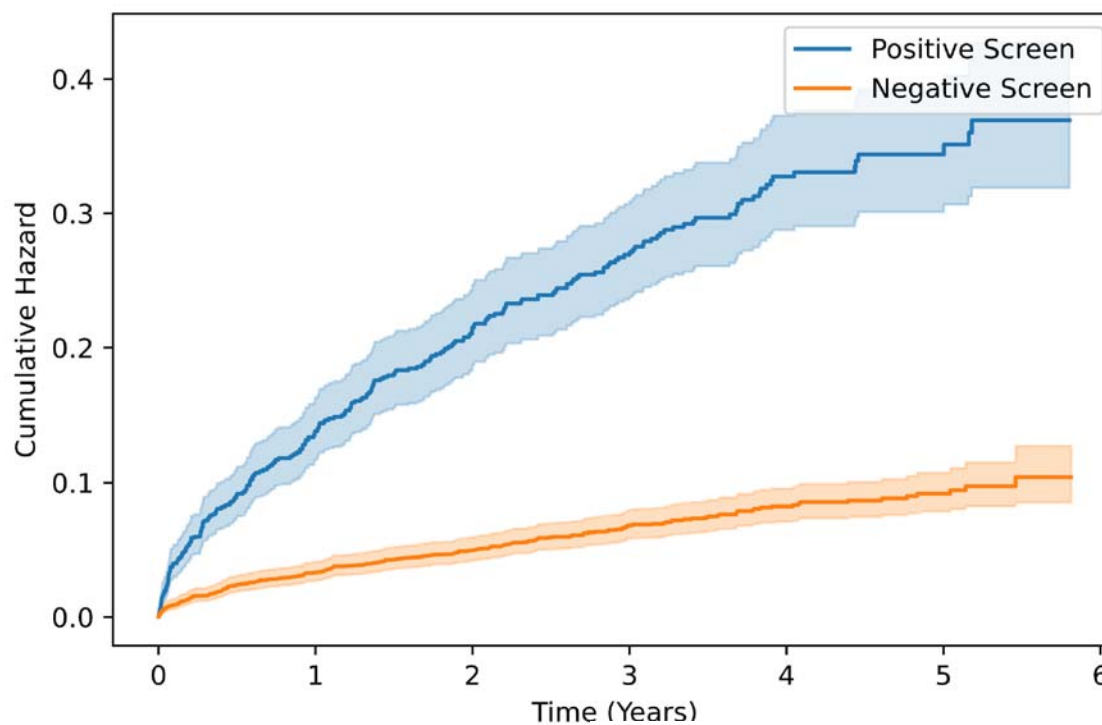


Figure 3. Cumulative hazard curves for incident LV systolic dysfunction in model-predicted positive and negative screens amongst the members of the held-out test set with LVEF $\geq 40\%$ and at least one follow-up measurement.



Positive Screen					
At risk	1125	810	567	297	56
Censored	0	146	333	570	805
Events	0	169	225	258	264
Negative Screen					
At risk	2921	2504	1839	976	138
Censored	0	305	921	1753	2584
Events	0	112	161	192	199

Figure 4. Gradient-weighted Class Activation Mapping (Grad-CAMs) across ECG formats. A) Standard format B) Two rhythm leads C) Standard shuffled format D) Alternate format. The heatmaps represent averages of the 100 positive cases with the most confident model predictions for LVEF < 40%.

A. Standard Format



C. Standard Shuffled Format



B. Two Rhythm Leads



D. Alternate Format



Figure 5. Receiver-Operating Curves for external validation sites. Abbreviations: AUROC, area under receiver-operating characteristic curve; EF, Ejection fraction; LRH, Lake Regional Hospital; YNHH, Yale New Haven Hospital

