Detection of Left Ventricular Systolic Dysfunction from Electrocardiographic Images

Veer Sangha[1*], Arash A Nargesi MD MPH[2*], Lovedeep S Dhingra MBBS[2], Bobak J Mortazavi PhD[3,4], Antônio H Ribeiro PhD[5], Cynthia A Brandt MD MPH[6,7], Edward J Miller MD PhD[2], Antonio Luiz P Ribeiro MD PhD[8,9], Eric J Velazquez MD[2], Harlan M Krumholz MD SM[2,4,10], Rohan Khera MD MS[2,4]

[1]Department of Computer Science, Yale University, New Haven, CT

[2]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT

[3]Department of Computer Science & Engineering, Texas A&M University, College Station, TX

[4]Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT

[5]Department of Information Technology, Uppsala University, Sweden

[6]Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT

[7]VA Connecticut Healthcare System, West Haven, CT

[8]Telehealth Center and Cardiology Service, Hospital das Clínicas

[9]Department of Internal Medicine, Faculdade de Medicina, Universidade Federal de Minas

[10]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT

*Contributed equally

Correspondence to:

Rohan Khera, MD, MS

195 Church St, 5th Floor, New Haven, CT 06510

203-764-5885; rohan.khera@yale.edu; @rohan_khera

ABSTRACT

Left ventricular (LV) systolic dysfunction is associated with over 8-fold increased risk of heart failure and 2-fold risk of premature death. To allow early detection by clinicians, we developed a layout-independent deep learning model that identifies LV systolic dysfunction directly from images of electrocardiograms (ECGs). Overall, 393,910 ECGs with paired echocardiograms were used for model development. The model performed well across multiple image formats (AUROC 0.91, AUPRC 0.69), and two external sets of real-world ECG images from an outpatient academic center (AUROC 0.93, AUPRC 0.69) and a rural hospital system (AUROC 0.91, AUPRC 0.89). Class-discriminative patterns localized to the anterior and anteroseptal leads corresponding to the left ventricle, regardless of ECG layout. A positive ECG screen in individuals with preserved LV ejection fraction was associated with 4-fold increased risk of developing LV systolic dysfunction in future. This approach represents an automated and accessible screening strategy for LV systolic dysfunction.

INTRODUCTION

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

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

We developed a method for accurate identification of LV ejection fraction (LVEF) less than 40%, a threshold with therapeutic implications, based on real-world ECG images. We developed, tested, and validated this approach using paired ECG-echocardiographic data from a large academic center and a rural hospital system.

RESULTS

Study Population

We used consecutive 12-lead ECGs from the Yale New Haven Hospital (YNHH) collected between 2015 and 2021. Out of the 2,135,846 ECGs obtained during this period, 440,072 were from patients who also had transthoracic echocardiograms (TTEs) within 15 days of obtaining the ECG. Overall, 393,910 had a complete ECG recording, representing 10 seconds of continuous recordings across all 12 leads. All selected ECGs were recorded as standard 12-lead recordings sampled at 500 Hz for 10 seconds.

The median age of patients at the time of ECG recording was 68 years (IQR 57, 78) and 45.3% of the ECGs were obtained from women. There were 34,577 (8.8%) readings from Hispanic individuals, 279,028 (70.8%) from non-Hispanic white individuals, 60,920 (15.5%) from non-Hispanic black individuals, and 19,385 (4.9%) from others. Of these ECGs, 60,293 (15.3%) had LV systolic dysfunction, defined as an associated echocardiographic recording of LVEF below 40%. An additional 38,727 (9.8%) had a LVEF greater than or equal to 40% but lower than 50%, and 294,890 (74.9%) had LVEF 50% or greater (**Supplementary Table 1)**.

Detection of LV Systolic Dysfunction

To develop a model that can adapt to real-world images across ECG formats, we created a dataset with different plotting schemes for each signal waveform recording (**Figure 1**). This strategy has been used to train a format-independent image-based model for the detection of conduction and rhythm disorders as well as the hidden label of gender.[14] The model was able to learn ECG lead-specific information based on the label regardless of the location of the lead.

We built a convolutional neural network similar to the EfficientNet B3 architecture (see Methods for details).[16] To increase the efficiency of learning and model performance, we pursued transfer learning from our previously developed model focused on detecting clinical diagnoses from ECGs.[14] We trained and validated our model on the constructed image dataset, with equal proportions of four different image formats. To mimic real-world variation in images and prevent model overfitting, all images were randomly rotated prior to being used as input in the model for training and validation. A custom class-balanced loss function based on the effective number of samples was used so that training was not affected by the lower frequency of the LVEF < 40% label relative to those with a LVEF ≥ 40%.

The model's area under receiving operation curve (AUROC) for LVEF < 40% on the held-out test set composed of standard images was 0.91 and its area under precision recall curve (AUPRC) was 0.69 **(Figure 2**). A threshold was chosen based on sensitivity of 0.90 or higher in the validation subset. With this threshold the model had sensitivity and specificity of 0.90 and 0.73 respectively in the held-out test set, and positive predictive value (PPV) and negative predictive value (NPV) of 0.38 and 0.98 respectively. Overall, an ECG suggestive of LV systolic dysfunction portended over 24-fold higher odds (OR 24.4, 95% CI, 22.4 – 26.6) of LV systolic dysfunction on echocardiogram (**Table 1**). The model's performance was comparable across subgroups of age, sex, and race (**Table 1** and **Figure 2**).

Model Performance Across ECG Formats

In the held-out set, 4 sets of image layouts were compared for model performance, including (i) the standard format with lead I as the rhythm strip (ii) two rhythm format with leads I and II as the rhythm strips, (iii) standard shuffled format with precordial leads located at the site of limb leads and vice versa, and (iii) alternate format with leads arranged in 6 rows across 2 columns. The model performance was comparable across format with AUROC between 0.90 – 0.91 on all four layouts of ECG images for detecting concurrent LV systolic dysfunction from echocardiogram (**Supplementary Table 2**). The model had a sensitivity between 0.89 and 0.90 and conferred 24- to 25-fold higher odds of LV systolic dysfunction on the standard and the 3 variations of the data.

LV Systolic Dysfunction in Model-predicted False Positives

Of the 33,362 ECGs in the held-out test set with an associated LVEF>40% on a proximate echocardiogram, the model classified 8,973 (26.9%) as "false positives", and 24,389 (73.1%) as true negatives. In further evaluation of false positives, 2,474 (27.6% of false positives) had evidence of mild LV systolic dysfunction with LVEF between 40-50% on concurrent echocardiography.

Among the 15,319 patients with LVEF>40% at the time of initial screening and at least one follow-up screening, the model classified 4,814 (31.4%) as false positives, and 10,505 (68.6%) as true negatives. 1,205 (13.4%) patients with model-predicted positive screen and 796 (3.3%) patients with model-predicted negative screen developed new LVEF<40% over the median follow-up period of 2.3 years (IQR 1.0 – 3.7, **Figure 3**). This represented a four-fold

higher risk of incident low LVEF based on having a positive screening result (HR 4.3, 95% CI 3.9 – 4.7).

Interpretability of Predicted LV Systolic Dysfunction

We used gradient-weighted class activation mapping (Grad-CAM) to highlight regions in each image that contributed most to predicting LVEF<40%. **Figure 4** shows the average class activation heatmaps for reduced LVEF prediction on all four formats of images that the model was trained and tested on. For all four formats of images, the region corresponding to leads V2 and V3 were the most important for prediction of reduced LVEF.

External Validation on Real-World ECG Images

We pursued validation on two real-world image datasets distinct from the data in the derivation cohort and the held-out test set. These cohorts were sampled from patients undergoing echocardiography.

The first were ECG images drawn from YNHH outpatient clinics during January through March 2022 and included 147 ECGs including 27 with LVEF < 40%, manually captured from the electronic health record. These images had a similar layout to the standard ECG format that had been used for model training but had the lead II rather than lead I as the rhythm strip. Moreover, there were several real-world noise artifacts in these images, including the shade of the page, vertical lines demarcating the leads, and differences in the location of lead labels. The model had an AUROC of 0.93 and AUPRC of 0.69 on the images (**Figure 5).**

The second real-world image dataset was randomly drawn from inpatient visits to the Lake Regional Hospital System (LRH) in 2022, a rural US hospital system in Osage Beach, MO,

with oversampling for patients with LV systolic dysfunction. These data included 100 ECG images, with 34 ECG images from patients with LVEF < 40%. These images also had a similar layout to the standard ECG format but had the lead II as the rhythm strip. There were unique noise real-world artifacts present on these images, including a different background color, the layout of the grid over which the waveform data are displayed as well as location and font of lead label. The model demonstrated an AUROC of 0.91 and AUPRC of 0.89 on these images (**Figure 5**).

DISCUSSION

We developed and externally validated an automated deep learning algorithm that accurately identifies LV systolic dysfunction solely from ECG images. The algorithm has high discrimination and sensitivity, representing characteristics ideal for a screening strategy. It is robust to variations in the layouts of ECG waveforms and detects the location of ECG leads across formats with consistent accuracy, making it suitable for broad implementation in varied settings. Moreover, the algorithm was developed and validated in a diverse population with high performance in subgroups of age, sex, and race across geographically dispersed academic and community health systems. The algorithm is also interpretable with localization of class-discriminative signals in anteroseptal and anterior leads regardless of the ECG layout, topologically corresponding to the left ventricle. Finally, among individuals who did not have a concurrently recorded low LVEF, a positive ECG screen was associated with a 4-fold increased risk of developing LV systolic dysfunction in the future compared with those with negative screen. This ECG image-based approach can represent a screening strategy for LV systolic dysfunction, particularly in low resource settings.

Image-based analysis of ECGs through deep learning represents a novel application of AI to improve clinical care. Convolutional neural networks have previously been designed to detect low LVEF from ECG signals[10,11], though the application is limited to harmonized data streams where ECG signal data are available in real time to clinicians. This is seldom the case in clinical practice where the ECG data architecture varies by ECG machine vendors, and data are not stored beyond generating the printed ECG images, particularly in remote settings.[17] The algorithm reported in this study overcomes these limitations by making detection of LV systolic dysfunction from ECGs interoperable across acquisition formats and directly available to clinicians who only have access to ECG images. Since scanned ECG images are the most common format of storage and use of electrocardiograms, untrained operators can implement large scale screening through chart review or automated applications to image repositories, requiring a lower resource task than optimizing tools for different machines.

The use of ECG images in our model overcomes the implementation challenges arising from black box algorithms. The origin of risk-discriminative signals in precordial leads of ECG images makes the model interpretable to clinicians as it localizes over the left ventricle. Moreover, the consistent observation of these predictive signals in the anteroseptal and anterior leads regardless of the lead location on printed image also serves as a control for the model predictions. Such visual representations that are consistent with clinical knowledge could address the hesitancy in uptake of these tools in clinical practice.[18]

An important finding was the significantly increased risk of incident LV systolic dysfunction among patients with model-predicted positive screen but LVEF>40% on concurrent echocardiography. These findings demonstrate an electrocardiographic signature which may precede the development of echocardiographic evidence of LV systolic dysfunction. This was

previously reported in signal-based models,[10] further suggesting that the detection of LV systolic dysfunction on ECG images represents a similar underlying pathophysiological process. These observations suggest a potential role for AI-based ECG models in risk stratification for future development of cardiovascular disease.[19]

Our study has certain limitations that merit consideration. First, we developed this model among patients who had both electrocardiograms and echocardiograms. Therefore, the training population selected likely had a clinical indication for echocardiography, differing from the broader real-world use of the algorithm for screening test for LV systolic dysfunction among those without any clinical disease. The excellent performance of our algorithm across demographic subgroups and the validation population would suggest robustness and generalizability of the effects though prospective assessments in the intended screening setting are warranted. Second, the model performance may vary by degree of LV systolic dysfunction. Though we chose an LVEF threshold of 40% due to its therapeutic implications, such as an indication for disease modifying guideline directed medical therapies,[4] the model identifies individuals with mild dysfunction. This may highlight a shared signature of LV systolic dysfunction among those with LVEF<40%, and with LVEF of 40-50%, but could also represent the lack of precision of LVEF measurement by echocardiography relative to more precise approaches, such as magnetic resonance imaging.[20,21] Third, while we incorporated four ECG formats during its development, the model is not restricted to these formats, with a different model focusing on clinical diagnoses with ECG images using a similar process has a robust performance on novel layouts.[14] Finally, while the model development pursues preprocessing of the ECG signal for plotting images, these represent standard processes performed before ECG images are generated and/or printed by ECG machines. Therefore, any other processing of

images is not required for the real-world application, as demonstrated in the application of the model to the real-world external validation set.

CONCLUSIONS

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

METHODS

The study was reviewed by the Yale Institutional Review Board, which approved the study protocol and waived the need for informed consent as the study represents secondary analysis of existing data.

Data Source and Study Population

We used 12-lead ECGs from the YNHH system collected between 2015 and 2021. These data were collected from the YNHH York Street Campus and the Saint Raphael Campus. These ECGs were recorded as standard 12-lead recordings sampled at a frequency of 500 Hz for 10 seconds and were collected from different machines. The majority of ECGs were collected using Philips PageWriter machines and GE MAC machines. Patient identifiers were used to identify ECGs with a corresponding transthoracic echocardiogram (TTE) within 15 days of obtaining the ECG. LVEF values were extracted based on a cardiologist read of the nearest TTE to each ECG.

To augment the evaluation of models built on an image dataset generated from this YNHH signal waveform, two sets of real-world image datasets were used for external validation.

Data Preprocessing for Model Development

All ECGs were analyzed to determine whether they had 10 seconds of continuous recordings across all 12 leads. The 10 second samples were preprocessed with a one second median filter, which was subtracted from the original waveform to remove baseline drift in each lead, a process that mimics what ECG machines do before printing the readings.

ECG signals were transformed into images using the python library ecg-plot.[22] Four formats of images were included in the training image dataset to ensure that the model learned information independent of the format and was generalizable to real-world images (**Figure 1**). The first format was based on the standard printed ECG format in the United States, with four 2.5 second columns printed sequentially on the page. Each column contained 2.5 second intervals from three leads. The full 10-second recording of the lead I signal was included as the rhythm strip. The second format, or the two-rhythm format, added lead II as an additional rhythm strip to the standard format. The third format consisted of two columns, the first with six simultaneous 5-second recordings from the limb leads, and the second with six simultaneous 5-second recordings from the precordial leads. The fourth format was the standard shuffled format, which had precordial leads in the first two columns and limb leads in the third and fourth. All images were converted to greyscale before being used as the input in the model.

Experimental Design

Each included ECG had a corresponding LVEF value from its nearest TTE within 15 days of recording, and low LVEF was defined as LVEF < 40%, the cutoff used an indication for most guideline-directed pharmacotherapy for heart failure.[4] All ECGs were randomly subset into training, validation, and held-out test sets (85%, 5%, 10%). Because of the low prevalence of LVEF < 40%, cases of normal and low LVEF were split proportionally among the sets. Additionally, to ensure that model learning was not affected by the relatively lower frequency of LVEF < 40%, higher weights were given to these cases based on the effective number of samples class sampling scheme.[23]

Model Training

We built a convolutional neural network model based on the EfficientNet-B3 architecture, which previously demonstrated an ability to learn and identify both rhythm and conduction disorders, as well as the hidden label of gender in real-world ECG images.[14] The EfficientNet-B3 model requires images to be sampled at 300 x 300 square pixels, includes 384 layers, and has over 10 million trainable parameters (**Supplementary Figure 1**). We utilized transfer learning by initializing model weights as the pretrained EfficientNet-B3 weights used to predict the six physician-defined clinical labels and gender from Sangha et al.[14] We first only unfroze the last four layers and trained the model with a learning rate of 0.01 for 2 epochs, and then unfroze all layers and trained with a learning rate of 5×10^{-6} for 4 epochs. We used an Adam optimizer, gradient clipping, and a minibatch size of 64 throughout training. The optimizer and learning rates were chosen after hyperparameter optimization. For both stages of training the model, we stopped training when validation loss did not improve in 3 consecutive epochs.

We trained and validated our model on a generated image dataset that had equal numbers of standard, two-rhythm, alternate, and standard shuffled images (**Figure 1**). All training and validation images were rotated a random amount between -10 and 10 degrees before being input into the model to mimic variation seen in uploaded ECGs and to aid in prevention of overfitting.

External validation: Real-World ECG Images

We pursued validation on two real-world image datasets. Both cohorts were sampled from patients undergoing echocardiography, with patients with LVEF < 40% oversampled. The first are ECG images drawn from YNHH during January through March 2022 and included 147 ECGs including 27 with LVEF < 40%, manually captured from the electronic health record. These images had a similar layout to the standard ECG format that had been created for model training but had the lead II rather than lead I as the rhythm strip. Moreover, there were several real-world noise artifacts in these images, including the shade of the page, vertical lines demarcating the leads, and differences in the location of the lead labels. The second real-world image dataset was drawn from the Lake Regional Hospital System (LRH), a rural US hospital system in Osage Beach, MO. These data included 100 ECG images, and 34 ECG images for patients with LVEF < 40%. These images also had a similar layout as the standard ECG format but had the lead II rather than lead I as the rhythm strip. There were unique noise real-world artifacts present on these images, including a different background color, the layout of the grid over which the waveform data are displayed as well as the location and the font of the lead label. Samples of electrocardiogram images from YNHH and LRH are presented in **Supplemental Figure 2** and images are available from the authors upon request.

Model Interpretability

We used Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight which portions of an image were important for the prediction of LVEF < 40%.[24] We calculated the gradients on the final stack of filters in our EfficientNet-B3 model for each prediction and performed a global average pooling of the gradients in each filter, emphasizing those that contributed to a prediction. We then multiplied these filters by their importance weights and combined them across filters to generate Grad-CAM heatmaps, which we overlayed on the original ECG images. We averaged class activation maps for predictions made on a certain format of image to see which leads were the most important for prediction of low LVEF. We took an arithmetic mean across the heatmaps for a given image format, and overlayed this average heatmap across a representative ECG to understand it in context.

Statistical Analysis

Model performance was evaluated in the held-out test set and real-world ECG image datasets. We used AUROC to measure model discrimination. We also assessed AUPRC, sensitivity, specificity, positive predictive value, negative predictive value, and diagnostic odds ratio, and chose threshold values based on cutoffs that achieved sensitivity of 0.90 in validation data. We evaluated future development of LV systolic dysfunction in time-to-event models using a Cox proportional hazards model. All analyses were performed using Python 3.9 and level of significance was set at an alpha of 0.05.

ACKNOWLEDGEMENTS

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

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

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

from the Siegfried and Jensen Law Firm for work related to Vioxx litigation; chairs a Cardiac Scientific Advisory Board for UnitedHealth; was a member of the IBM Watson Health Life Sciences Board; is a member of the Advisory Board for Element Science, the Advisory Board for Facebook, and the Physician Advisory Board for Aetna; and is the co-founder of Hugo Health, a personal health information platform, and co-founder of Refactor Health, a healthcare AI-augmented data management company. Dr Ribeiro is supported in part by CNPq (310679/2016-8 and 465518/2014-1) and by FAPEMIG (PPM-00428-17 and RED-00081-16). Mr. Sangha and Dr. Khera are the coinventors of U.S. Provisional Patent Application No. 63/346,610, “Articles and methods for format independent detection of hidden cardiovascular disease from printed electrocardiographic images using deep learning”. Dr. Khera is also a founder of Evidence2Health, a precision health platform to improve evidence-based care.

REFERENCES

1 Wang, T. J. *et al.* Natural history of asymptomatic left ventricular systolic dysfunction in the community. *Circulation* **108**, 977-982, doi:10.1161/01.CIR.0000085166.44904.79 (2003).

2 Srivastava, P. K. *et al.* Heart Failure Hospitalization and Guideline-Directed Prescribing Patterns Among Heart Failure With Reduced Ejection Fraction Patients. *JACC Heart Fail* **9**, 28-38, doi:10.1016/j.jchf.2020.08.017 (2021).

3 Wolfe, N. K., Mitchell, J. D. & Brown, D. L. The independent reduction in mortality associated with guideline-directed medical therapy in patients with coronary artery disease and heart failure with reduced ejection fraction. *Eur Heart J Qual Care Clin Outcomes* **7**, 416-421, doi:10.1093/ehjqcco/qcaa032 (2021).

4 Heidenreich, P. A. *et al.* 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*, 101161CIR0000000000001063, doi:10.1161/CIR.0000000000001063 (2022).

5 Wang, T. J., Levy, D., Benjamin, E. J. & Vasan, R. S. The epidemiology of "asymptomatic" left ventricular systolic dysfunction: implications for screening. *Ann Intern Med* **138**, 907-916, doi:10.7326/0003-4819-138-11-200306030-00012 (2003).

6 Vasan, R. S. *et al.* Plasma natriuretic peptides for community screening for left ventricular hypertrophy and systolic dysfunction: the Framingham heart study. *JAMA* **288**, 1252-1259, doi:10.1001/jama.288.10.1252 (2002).

7 McDonagh, T. A., McDonald, K. & Maisel, A. S. Screening for asymptomatic left ventricular dysfunction using B-type natriuretic Peptide. *Congest Heart Fail* **14**, 5-8, doi:10.1111/j.1751-7133.2008.tb00002.x (2008).

8 Galasko, G. I., Barnes, S. C., Collinson, P., Lahiri, A. & Senior, R. What is the most cost-effective strategy to screen for left ventricular systolic dysfunction: natriuretic peptides, the electrocardiogram, hand-held echocardiography, traditional echocardiography, or their combination? *Eur Heart J* **27**, 193-200, doi:10.1093/eurheartj/ehi559 (2006).

9 Atherton, J. J. Screening for left ventricular systolic dysfunction: is imaging a solution? *JACC Cardiovasc Imaging* **3**, 421-428, doi:10.1016/j.jcmg.2009.11.014 (2010).

10 Attia, Z. I. *et al.* Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* **25**, 70-74, doi:10.1038/s41591-018-0240-2 (2019).

11 Vaid, A. *et al.* Using Deep-Learning Algorithms to Simultaneously Identify Right and Left Ventricular Dysfunction From the Electrocardiogram. *JACC Cardiovasc Imaging* **15**, 395-410, doi:10.1016/j.jcmg.2021.08.004 (2022).

12 Yao, X. *et al.* Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med* **27**, 815-819, doi:10.1038/s41591-021-01335-4 (2021).

13 Stamenov, D., Gusev, M. & Armenski, G. in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO).* 0319-0323 (IEEE).

14 Sangha, V. *et al.* Automated multilabel diagnosis on electrocardiographic images and signals. *Nat Commun* **13**, 1583, doi:10.1038/s41467-022-29153-3 (2022).

15 He, J. *et al.* The practical implementation of artificial intelligence technologies in medicine. *Nat Med* **25**, 30-36, doi:10.1038/s41591-018-0307-0 (2019).
16 Mingxing Tan and Quoc V Le. EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning, 2019*.
17 Siontis, K. C., Noseworthy, P. A., Attia, Z. I. & Friedman, P. A. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol* **18**, 465-478, doi:10.1038/s41569-020-00503-2 (2021).
18 Quer, G., Arnaout, R., Henne, M. & Arnaout, R. Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review. *J Am Coll Cardiol* **77**, 300-313, doi:10.1016/j.jacc.2020.11.030 (2021).
19 Maurovich-Horvat, P. Current trends in the use of machine learning for diagnostics and/or risk stratification in cardiovascular disease. *Cardiovasc Res* **117**, e67-e69, doi:10.1093/cvr/cvab059 (2021).
20 Pellikka, P. A. *et al.* Variability in Ejection Fraction Measured By Echocardiography, Gated Single-Photon Emission Computed Tomography, and Cardiac Magnetic Resonance in Patients With Coronary Artery Disease and Left Ventricular Dysfunction. *JAMA Netw Open* **1**, e181456, doi:10.1001/jamanetworkopen.2018.1456 (2018).
21 Jenkins, C. *et al.* Left ventricular volume measurement with echocardiography: a comparison of left ventricular opacification, three-dimensional echocardiography, or both with magnetic resonance imaging. *Eur Heart J* **30**, 98-106, doi:10.1093/eurheartj/ehn484 (2009).
22 ECG Plot Python Library. Accessed at https://pypi.org/project/ecg-plot/ on May 25, 2022.
23 Cui, Y., Jia, M., Lin, T.-Y., Song, Y. & Belongie, S. Class-Balanced Loss Based on Effective Number of Samples. arXiv:1901.05555 (2019). <https://ui.adsabs.harvard.edu/abs/2019arXiv190105555C>.
24 Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Ieee I Conf Comp Vis*, 618-626, doi:10.1109/Iccv.2017.74 (2017).

Table 1. Performance of model on test images across demographic subgroups. 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	PPV	NPV	Specificity	Sensitivity	AUROC	AUPRC	OR
All	0.377	0.976	0.731	0.9	0.907 (0.903 - 0.911)	0.688	24.41
Male	0.411	0.973	0.7	0.914	0.905 (0.900 - 0.910)	0.719	24.86
Female	0.321	0.979	0.765	0.871	0.904 (0.898 - 0.911)	0.624	22.06
>=65	0.366	0.975	0.691	0.911	0.899 (0.894 - 0.904)	0.673	22.91
< 65	0.398	0.976	0.787	0.881	0.918 (0.912 - 0.924)	0.711	27.28
Hispanic	0.407	0.971	0.772	0.87	0.909 (0.895 - 0.923)	0.707	22.76
White	0.368	0.977	0.73	0.9	0.906 (0.902 - 0.911)	0.669	24.31
Black	0.401	0.973	0.704	0.909	0.907 (0.897 - 0.917)	0.749	23.83
Other	0.362	0.984	0.756	0.919	0.923 (0.907 - 0.939)	0.665	35.28

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; YNHH, Yale New Haven Hospital.

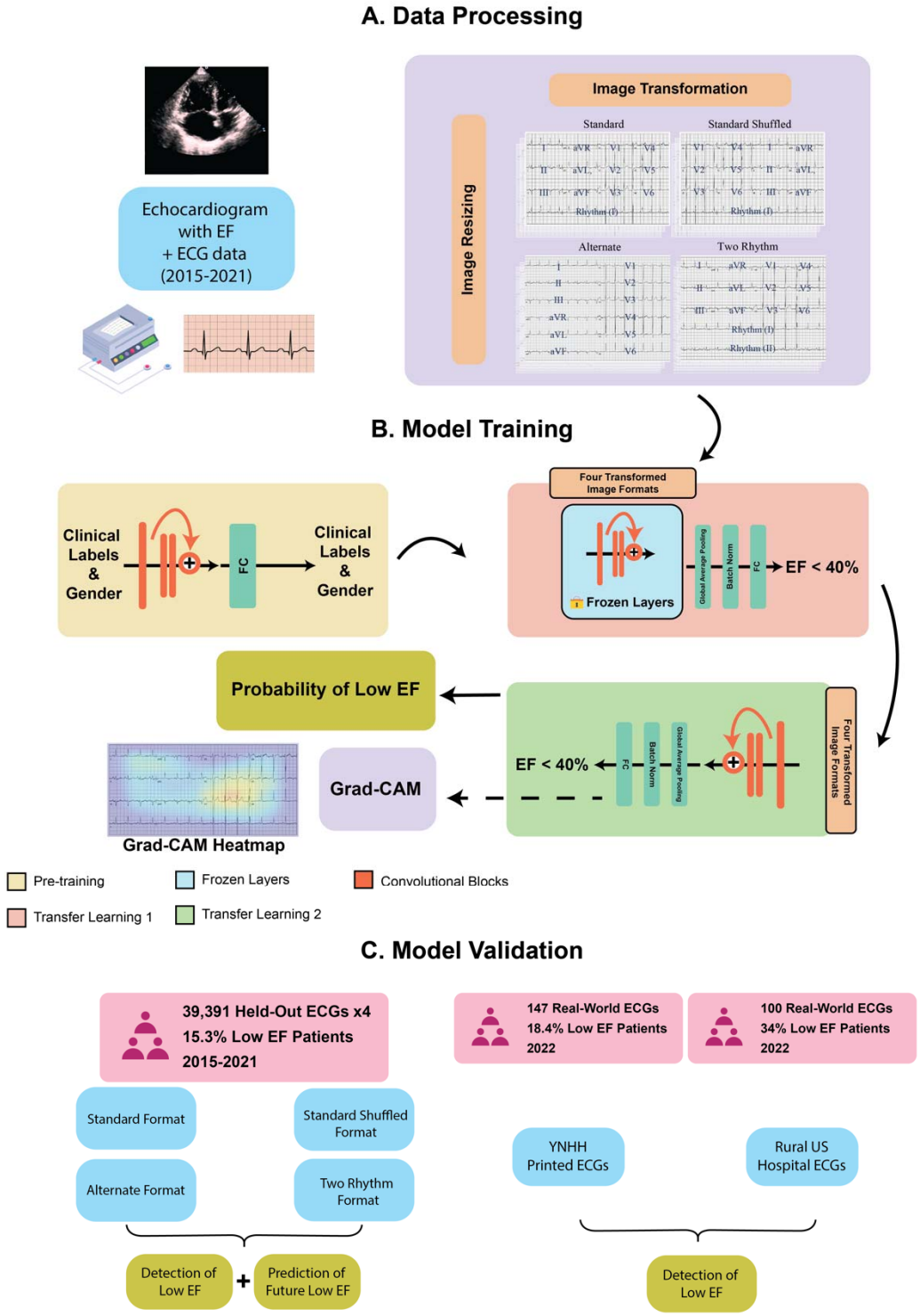


Figure 2. Model Performance Measures A) Receiver-Operating and B) Precision-Recall Curves on images in held-out test set C) Diagnostic Odds Ratios across age, gender, and race subgroups on standard format images in the held-out test set. Abbreviations: AUROC, area under receiver-operating characteristic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

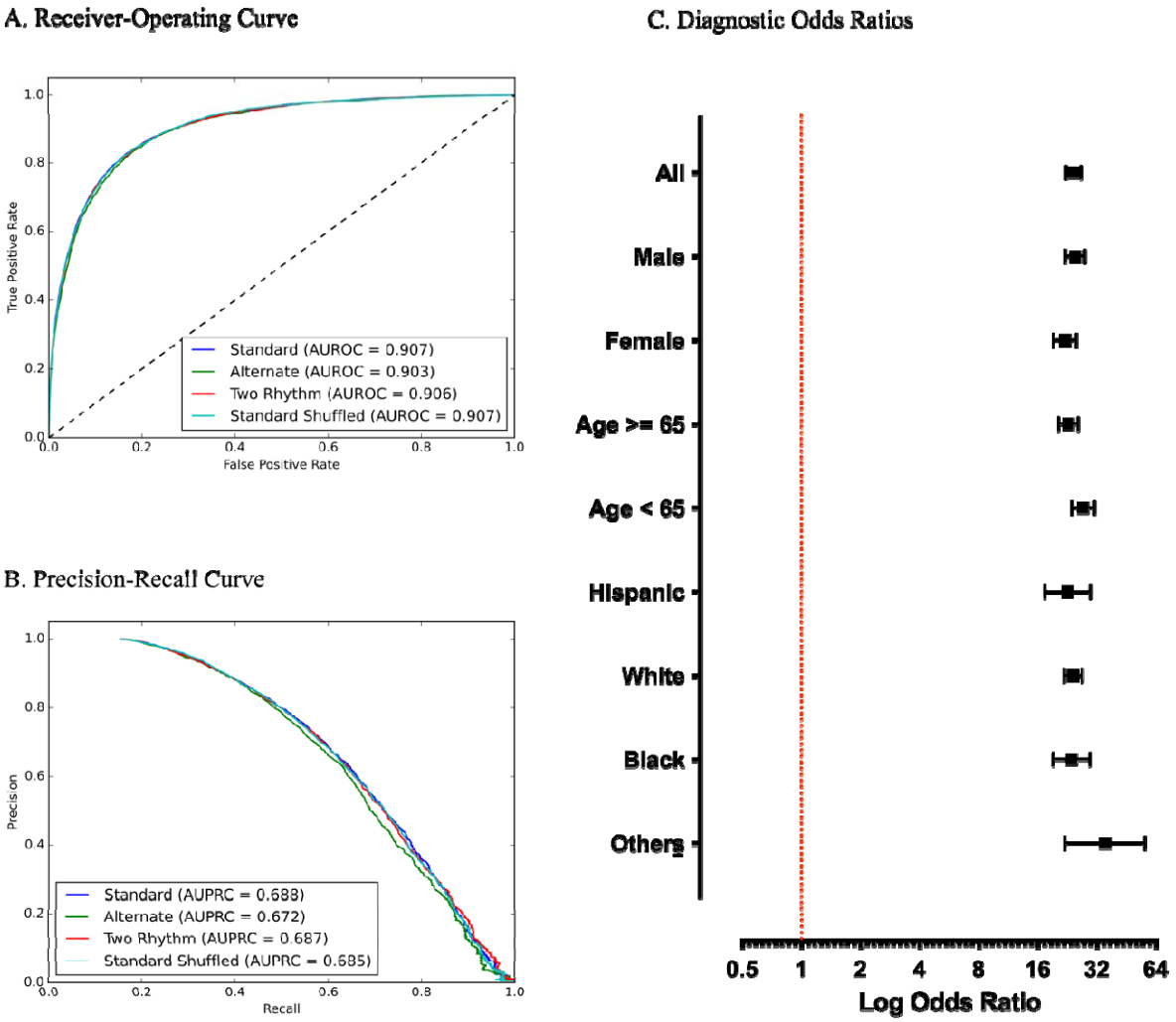
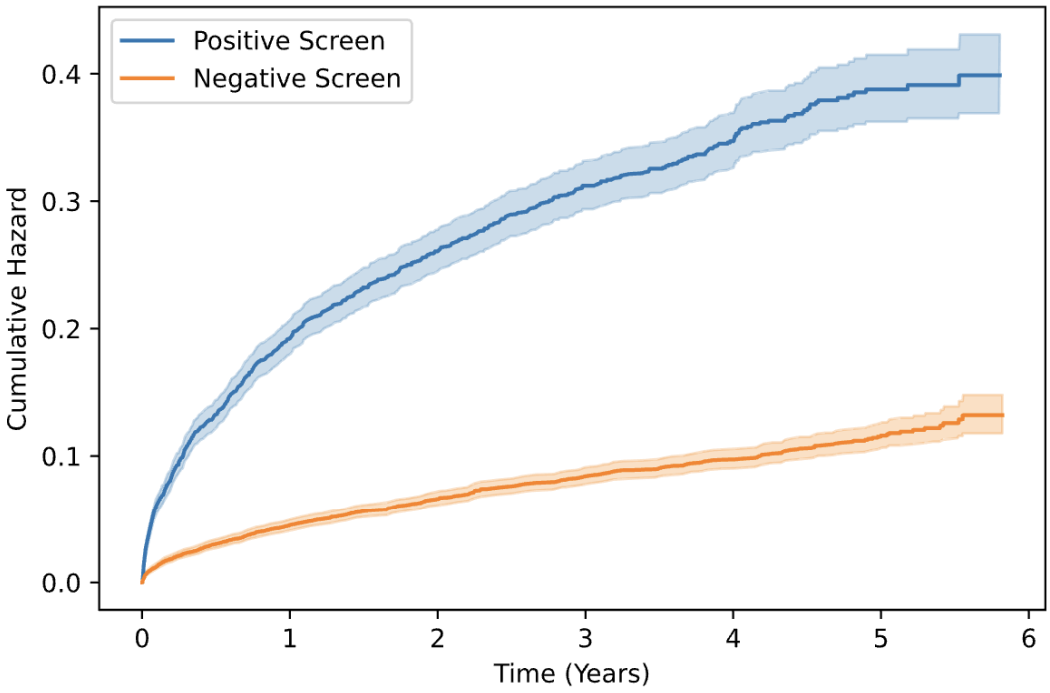


Figure 3. Cumulative hazard curves for incident LV systolic dysfunction in model-predicted positive and negative screens amongst the members of the held-out test set with LVEF > 40% and at least one follow-up measurement.



Positive Screen					
At risk	4814	3116	1959	907	152
Censored	0	792	1750	2722	3457
Events	0	906	1105	1185	1205
Negative Screen					
At risk	10505	8079	5136	2528	404
Censored	0	1918	4685	7219	9305
Events	0	508	684	758	796

Figure 4. Gradient-weighted Class Activation Mapping (Grad-CAMs) A) Standard format B) Two rhythm leads C) Standard shuffled format D) Alternate format

A. Standard Format

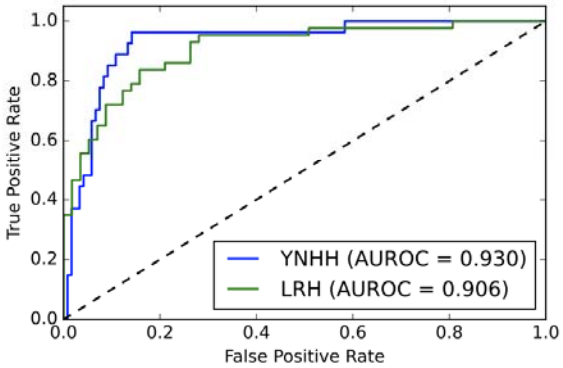
C. Standard Shuffled Format

B. Two Rhythm Leads

D. Alternate Format

Figure 5. Model Performance on real-world validation datasets. A) Receiver-Operating Curve, B) Confusion Matrices. Abbreviations: EF, Ejection fraction; LRH, Lake Regional Hospital; YNHH, Yale New Haven Hospital

A. Receiver-Operating Curve



B. Confusion Matrices

YNHH		Prediction Score			
		0-.1	.1-.2	.2-.5	.5-1
EF	< 40	1	0	0	26
	40 - 50	1	2	1	10
	> 50%	64	18	16	8

LRH		Prediction Score			
		0-.1	.1-.2	.2-.5	.5-1
EF	< 40	2	6	11	24
	40 - 50	9	3	3	1
	> 50%	32	4	4	1