Identification of Hypertrophic Cardiomyopathy on Electrocardiographic Images with Deep Learning

Veer Sangha BS[1,2*], Lovedeep Singh Dhingra MBBS[1*], Evangelos Oikonomou MD, DPhil[1], Arya Aminorroaya MD, MPH[1], Nikhil V Sikand MD, FACC[1], Sounok Sen MD[1], Harlan M Krumholz MD, SM[1,3], Rohan Khera MD, MS[1,3,4]

[1]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
[2]Department of Engineering Science, Oxford University, Oxford, UK
[3]Center for Outcomes Research and Evaluation (CORE), Yale New Haven Hospital, New Haven, CT, USA
[4]Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

*Contributed equally as co-first authors

Abstract Word Count: 267 words

Word Count: 3656 words

Figures/Tables: 5 Figures / 2 Tables

Keywords: Artificial Intelligence, Electrocardiogram, Hypertrophic Cardiomyopathy, Sudden Cardiac Death, Machine Learning, Health Technology

Correspondence to:
Rohan Khera, MD, MS
195 Church Street, 6th Floor, New Haven, CT 06510
rohan.khera@yale.edu

ABSTRACT

Background: Hypertrophic cardiomyopathy (HCM) affects 1 in every 200 individuals and is the leading cause of sudden cardiac death in young adults. HCM can be identified using an electrocardiogram (ECG) raw voltage data and deep learning approaches, but their point-of-care application is limited by the inaccessibility of these signal data. We developed a deep learning-based approach that overcomes this limitation and detects HCM from images of 12-lead ECGs across layouts.

Methods: We identified ECGs from patients with HCM features present on cardiac magnetic resonance imaging (CMR) or those within 30 days of an echocardiogram documenting thickened interventricular septum (end-diastolic interventricular septum thickness > 15mm). Patients with CMR-confirmed HCM were considered as cases during the final model evaluation. The model was validated within clinical settings at YNHH and externally on ECG images from the prospective, population-based UK Biobank cohort. We localized class-discriminating signals in ECG images using gradient-weighted class activation mapping.

Results: Overall, 124,553 ECGs from 66,987 individuals (HCM cases and controls) were used for model development. The model demonstrated high discrimination for HCM across various ECG image formats and calibrations in internal validation (area under receiving operation characteristics [AUROC] 0.96) and external sets of ECG images from UK Biobank (AUROC 0.94). A positive screen for HCM was associated with a 100-fold higher odds of CMR-confirmed HCM (OR 102.4, 95% Confidence Interval, 57.4 – 182.6) in the held-out set. Class-discriminative patterns localized to the anterior and lateral leads (V4-V5).

Conclusions: We developed and externally validated a deep learning model that identifies HCM from ECG images with excellent discrimination. This approach represents an automated, efficient, and accessible screening strategy for HCM.

INTRODUCTION

While hypertrophic cardiomyopathy (HCM) is among the leading causes of sudden cardiac death, scalable solutions for screening for the disease have remained elusive.[1,2] HCM is a genetically determined disease that affects up to 1 in every 200 people globally.[3,4] An early diagnosis of HCM can enable regular healthcare follow-up, rigorous cardiovascular risk management, and timely initiation of highly effective risk-reducing therapies.[2,5] The diagnosis of HCM has relied on cardiac imaging, such as echocardiography and cardiac magnetic resonance imaging (CMR).[5,6] However, given the requirement of expensive technology and extensive clinical expertise in deploying and interpreting these modalities, using echocardiography or CMR as a screening strategy is not feasible.[6,7]

Given the inaccessibility of advanced cardiac imaging, deep learning or artificial intelligence (AI)-enhanced interpretation of electrocardiograms (AI-ECG) has been proposed as an alternative for the early detection of HCM.[8,9] While ECG abnormalities, such as prominent Q waves, repolarization changes, left axis deviation, or giant negative T waves, can be apparent in over 90% of patients with the disease, these changes are not specific to HCM.[10,11] Deep learning-based approaches, such as convolutional neural networks (CNNs) can leverage HCM-specific pathological signatures to identify people the disease using clinical ECGs.[8,9,12] Current models, however, use raw ECG voltage data as the inputs, which are often stored in vendor-specific formats and rarely accessible to clinicians at the point-of-care. Moreover, for identifying patients with HCM in development cohorts, a combination of diagnosis codes and other administrative data sources, such as visits to the HCM clinic, have been used.[9,12] Such administrative-code-based phenotyping of diseases in the electronic health record is prone to misclassification and variability due to vast differences in health-system-specific coding practices for HCM.[13,14]

Therefore, there is an unmet need for the development of models that use ubiquitous and interoperable data formats for disease diagnosis and rely on objective imaging-based biological features for defining the presence of disease to enable AI-ECG's use as a practical, generalizable, and scalable screening modality for HCM.

In this study, we report the development and validation of a deep learning-based approach for identifying CMR-confirmed HCM using images of clinical 12-lead ECGs.

METHODS

The Yale Institutional Review Board approved the study protocol and waived the need for informed consent as the study represents a secondary analysis of existing data. Patients who opted out of research studies at the Yale New Haven Hospital (YNHH) were not included. An online version of the model is publicly available for research use at https://www.cards-lab.org/ecgvision-hcm. This web application represents a prototype of the eventual application of the model, with instructions for required image standards and a version that demonstrates an automated image standardization pipeline.

Data Source and Study Population

We used 12-lead ECG signal waveform data collected during the clinical care of patients at the YNHH between 2012 and 2021. These ECGs were recorded as standard 12-lead recordings sampled at a frequency of 500 Hz for 10 seconds. These were recorded on multiple different machines, primarily the Philips PageWriter and GE MAC machines.

We identified the earliest MRI reports for 1,061 patients containing any mention of HCM. Each of these reports was manually reviewed by three cardiologists, and 904 people were

identified as having confirmed or possible HCM. Of these, 779 reports included a confirmed HCM diagnosis and 125 were identified as possible HCM, defined as the inclusion of HCM as one of the reported differential diagnoses, or the presence of features potentially suggestive, but not conclusive for HCM. The data from these patients were split into mutually exclusive training, validation, and test sets in an 85:5:10 ratio. Only MRI-confirmed HCM cases were retained in validation and testing, while confirmed and possible HCM cases were included in the training cohort. For each of these patients, all ECGs recorded up to a year before the MRI, and any time after the MRI were considered, except for those ECGs after a septal reduction procedure, including alcohol septal ablation or ventricular myectomy. To further augment the training cohort, we incorporated ECGs from patients whose transthoracic echocardiograms (TTE) demonstrated severe ventricular hypertrophy (LVH), defined by interventricular septal thickness in diastole (IVSd) of greater than 15mm. We posited that these may represent individuals with possible HCM. ECGs performed within 30 days before or after a TTE demonstrating severe LVH were included as cases in the training set, but were not considered cases in the test set, which only included CMR-confirmed HCM. For patients with more than one recorded ECG filling these criteria, a maximum of five most recent ECGs were used to create cohorts to avoid overrepresenting patients undergoing frequent ECGs.

To derive a control cohort for the training set, we identified ECGs recorded within 30 days of a TTE in a cohort of patients who did not have a diagnosis code recorded for HCM (Table S2) and were not in the cohort of patients with possible positive HCM cases based on MRI reports or IVSd values. Thus, patients with any ICD code for HCM (Table S2), any mention of HCM in cardiac MRI reports, or any TTE with IVSd > 15mm were not included in the control cohort. We

randomly sampled control ECGs so the train set had a 10% prevalence of HCM ECGs to allow the model to learn signatures of HCM on ECG successfully.

To identify a control cohort for the validation and test sets, we used ECGs from any patients who did not have any diagnosis code suggestive of conditions causing cardiomyopathy, along with an inpatient hospitalization for heart failure **(Table S2)**. These patients were also not present in the cohort of patients with confirmed or possible HCM cases based on MRI reports. Since HCM more commonly affects men and is diagnosed in the younger population, it is important to ensure that the model identifies the pathological signature of the HCM and does not rely on age- and sex-based ECG features for the HCM prediction. Thus, the validation and test controls were age- and sex-matched to HCM ECGs at a 10:1 prevalence. We identified 10 control ECGs from patients within 5 years of age, and of the same sex, as each HCM case ECG. We ensured no patient overlap between training, validation, and test sets.

Image Generation

We generated ECG images to recapitulate the variation in ECG layouts in a real-world setting. Our approach to image plotting has been previously described and represents the processing steps of ECG machines to convert acquired waveform data to printed outputs.[15–18] Briefly, 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, subtracted from the original waveform to remove baseline drift in each lead. The process of converting ECG signals to images was independent of model development, ensuring that the model did not learn any aspects of the processing that generated images from the signals.

ECG signals were transformed into ECG images using the Python library ecg-plot.[19] Images were generated with a calibration of 10 mm/mV, which is standard for printed ECGs in most real-world settings. Using the Python Image Library (PIL v9.2.0), we converted all images to greyscale, followed by down-sampling to 300x300 pixels regardless of their original resolution. Given that real-world ECG images may vary in the layout of leads, we created a dataset with four different plotting schemes for each signal waveform recording (**Figure 1**). The first format was based on the standard printed ECG format in the United States. This format consisted of four columns printed sequentially, each containing 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, a two-rhythm format, added lead II as an additional rhythm strip to the standard format. The third layout was the alternate format, which 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, without a corresponding rhythm lead. The fourth format was a shuffled format, which had precordial leads in the first two columns and limb leads in the third and fourth. All images were rotated a random amount between -10 and 10 degrees before being input into the model to mimic variations seen in uploaded ECGs and to aid in the prevention of overfitting.

Model Architecture and Training

We built a convolutional neural network model based on the EfficientNet-B3 architecture.[20] 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. To allow label-efficient model development, we initialized the model using weights from a pretrained EfficientNet-B3 model that leveraged a

novel self-supervised biometric contrastive learning approach, wherein the model was trained to identify individual patient-specific patterns in ECGs regardless of their interpretation.[21] None of the ECGs on the self-supervised pretraining task represented individuals in the model development. For training, we first unfroze the last four layers and trained the model with a learning rate of 0.01 for 2 epochs. Then, we unfroze all layers and trained the model with a learning rate of 5×10^{-6} for 6 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. A custom class-balanced loss function (weighted binary cross-entropy) based on the effective number of samples was used given that the case and control labels were not equally balanced.

Evaluation of Hypertrophy in Patients without Confirmed HCM

Type 1 error, or a high false positive rate, is a major concern for screening strategies for low-prevalence conditions like HCM. Furthermore, given that HCM is often underdiagnosed, it is important to evaluate the phenotypic characteristics of false positive cases. Therefore, among patients without CMR-confirmed HCM, we applied the model to 5,000 randomly selected ECGs recorded within 30 days of a TTE. These ECGs were taken from patients who had previously not been analyzed in development or evaluation sets. We extracted the end-diastolic interventricular septal wall thickness (IVSd), available as a continuous measure from TTEs. To evaluate the instances of model-positive screens in patients without CMR-confirmed HCM, we compared the IVSd measurements in false positive and true negative screens.

Localization of Model Predictive Cues

To obtain a heatmap highlighting the portions of an ECG image that were important for predicting HCM, we used Gradient-weighted Class Activation Mapping (Grad-CAM).[22] 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. Among the 25 positive cases with the most confident model predictions for HCM across ECG formats, we averaged class activation maps to determine the most important image areas for the prediction of HCM. We took an arithmetic mean across the heatmaps for a given image format and overlayed this average heatmap across a representative ECG before the conversion of the image to grayscale. The Grad-CAM intensities were converted from their original scale (0 – 1) to a color range using the jet colormap array in the Python library matplotlib, which was then overlaid on the original ECG image with an alpha of 0.3. The activation map, a 10x10 array, was upsampled to the original image size using the bilinear interpolation built into TensorFlow v2.8.0. We also evaluated the Grad-CAM for individual ECGs in the UK Biobank to evaluate the consistency of the information on individual examples.

External Validation

We used data from the UK Biobank, under research application #71033, to pursue external validation of our model. UK Biobank represents the largest population-based cohort of 502,468 people in the United Kingdom with protocolized imaging and laboratory testing, along with linked electronic health records. Given the mean age at diagnosis for HCM among adults is 51 ±

16 years,[23] we evaluated our model among ECGs from participants where a majority were < 70 years old, representing the population where an AI-ECG model may be used for identifying HCM. We used linked electronic health records for the participants to identify the presence of HCM diagnosis codes. In patients without HCM, we also used CMR-derived left-ventricular mass index (LVMI) to compare the characteristics of participants with a positive and negative AI-ECG screen for HCM. LVH was defined as LVMI > 70 in men and LVMI > 55 in women.[24,25]

Statistical Analysis

Categorical variables were reported as number (percentage, %), and continuous variables as mean (standard deviation [SD]) or median (interquartile range [IQR]), as appropriate. The model's performance was presented as area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC). The 95% confidence intervals (CI) for AUROC and AUPRC were calculated using DeLong's algorithm and bootstrapping with 1000 iterations, respectively.[26,27] Furthermore, we reported sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score of the model at the model threshold for 90% sensitivity in the validation set. The statistical significance level was set at $P < 0.05$. All statistical analyses were executed using Python 3.11.2 and R version 4.2.0.

RESULTS

Study Population

We used data from 126,203 12-lead ECGs obtained from 68,109 patients at YNHH. The data from these patients were split into train, validation, and test sets at a patient level, as described in the methods. Individuals in the model development population (training and validation sets) had

a median age of 63.2 years (IQR 51.2-74.1) at the time of ECG recording, and 33257 (49.6%) were women. Overall, 48475 (72.4%) were non-Hispanic White, 7978 (11.9%) were non-Hispanic Black, 6029 (9.0%) were Hispanic, 1300 (1.9%) were Asian, 376 (0.6%) were from other races, and information was missing for 2829 (4.2%). In the development population, there were 12,680 ECGs from 4745 patients with CMR-confirmation of HCM or echocardiographic parameters consistent with HCM **(Table 1)**.

Detection of HCM

In the age- and sex-matched held-out test set comprising standard format images, the model for detecting HCM achieved an AUROC of 0.96 (**Figure 2**). A probability threshold for predicting HCM was chosen based on a sensitivity of 0.90 or higher in the validation subset. With this threshold, the model had sensitivity and specificity of 0.91 and 0.91 in the held-out test set and PPV and NPV of 0.51 and 0.99, respectively. Overall, an ECG suggestive of HCM portended over 100-fold higher odds (OR 102.4, 95% CI, 57.4 – 182.6) of HCM (**Figure 2**). The model's performance was comparable across subgroups of age, sex, and race (**Table 2** and **Figure 2**). The model performance was also comparable across the four original layouts of ECG images in the held-out set with an AUROC of 0.95 – 0.96 for detecting HCM. Sensitivity analyses demonstrated consistent model performance on ECGs without paced rhythms, atrial fibrillation and flutter, conduction disorders, and in the presence of LVH (**Table 2**).

Evaluation of LVH Phenotype in Model-predicted False Positives

We applied the model to 5,000 randomly selected ECGs recorded within 30 days of a TTE in patients outside our development and held-out test sets. Of these, 647 (12.9%) were classified as

false positives and 4,353 (87.1%) as true negatives. The median IVSd among the false-positive subset was 11mm (IQR 9.7 – 12.4), compared with 9.5mm (IQR 8.4 – 10.8) among false negatives ($p < 0.001$) **(Figure 3A)**.

Localization of Predictive Cues for HCM

Class activation heatmaps of the 25 positive cases with the most confident model predictions for HCM prediction across four ECG layouts are presented in **Figure 4**. For all four formats of images, the region corresponding to leads V4 and V5 were the most important areas for prediction of HCM. Representative images of Grad-CAM analysis in sampled individuals with positive screens in UK Biobank, the external validation site, showed similar patterns (**Figure 5B**).

External Validation

We applied the model to the UK Biobank validation set, consisting of 32,885 ECGs from prospectively enrolled individuals, including 18 (0.05%) with an ICD code for HCM. The model had an AUROC of 0.94 (0.89 – 0.99) on these images (**Figure 5A**), with a sensitivity of 0.61 and specificity of 0.96 at the threshold set in the development population. Of the 32,867 ECGs in this set without an HCM diagnosis, 32,859 were from individuals who had undergone cardiac MRIs. The model classified 1,322 (4.0%) as false positives and 31,537 (96.0%) as true negatives. Of the false positive screens, 153 (11.6%) had LVH, compared with 313 (1.0%) among true negative screens **(Figure 3B)**.

DISCUSSION

We developed and validated an automated deep learning model for identifying HCM solely from ECG images. The model is robust to variations in the layouts of ECG waveforms, making it suitable for implementation in various settings. Moreover, the model has excellent discrimination and sensitivity, representing characteristics ideal for screening. The model was developed and tested in a diverse population with high performance in subgroups of age, sex, and race. The model performance was consistent in the UK Biobank, a population-based cohort, despite a different, diagnosis code-based definition of HCM. An evaluation of the class-discriminating signals localized it to the anterior and lateral leads regardless of the ECG layout, topologically corresponding to the left ventricle. Therefore, an ECG image-based approach can represent a screening strategy for HCM, particularly in low-resource settings.

ECG-image-based deep learning models represent a novel application of AI that has the potential to improve clinical care and public health by offering a feasible modality for the early detection of HCM. Previously, various criteria have been proposed for the identification of HCM based on clinical interpretation of visible ECG features, such as abnormal Q or T waves, repolarization changes, or high QRS complexes.[10] However, the accuracy of these clinical ECG-based criteria is limited, ranging from 55% to 80%.[28,29] Moreover, the visible ECG abnormalities are not specific for HCM and are often present in various clinical conditions causing left ventricular hypertrophy, or rarely even as physiological variants.[10,30] More recently, deep learning models utilizing raw ECG voltage signals have been proposed for the detection of HCM.[8,9,12] While these models have excellent performance in internal validation, their development has often relied on diagnosis codes and visits to a specialty HCM clinic for the identification of HCM cases.[9,12] These practices for recording diagnosis and administrative codes are often health-system-specific, and the use of diagnosis codes has been shown to misclassify

up to one-third of HCM cases.[13,14] Thus, to ensure the generalizability of deep learning models in diverse settings, it is critical to incorporate objective imaging-based biologic features for the identification of disease.[8,18] Our use of CMR for the identification of disease represents a robust definition for identifying HCM, and enables the model to differentiate HCM from LVH-causing HCM mimics such as severe AS or hypertension.[8,10] This is also demonstrated by consistently high discrimination of the model for HCM in the patient subgroup with increased septal wall thickness.

Using images to detect HCM signifies an advance over signal-based models, allowing for accessible implementation of a potential ECG-based screening approach. Digital or printed ECG images represent the most commonly available format, especially in low-resource settings. Moreover, ECG images are an interoperable data stream that is not tied to proprietary formats from specific ECG machine vendors, making them readily available to clinicians at the point-of-care. Currently, the guidelines for universal screening for HCM are equivocal, given the limited affordability of advanced cardiac imaging and the high number of false-positive and negative screens on clinically apparent ECG anomalies.[31–33] However, an accurate and accessible approach to HCM diagnosis using AI can potentially make HCM screening economically viable, especially for people at elevated risk of sudden cardiac death, such as young athletes.[11,34,35] Moreover, using ECG images in our model overcomes the implementation challenges of black box algorithms. The consistent localization of the risk-discriminative signals in anterior and lateral leads of ECG images, regardless of the lead location on printed images, indicates the left ventricular origin of the underlying pathology. Visual representations consistent with clinical knowledge could explain parts of the model prediction process and address the hesitancy in using these tools in clinical practice.[36]

Our study has several limitations. First, our model was developed using retrospective data from a single center, including ECGs from patients with clinical indication for both ECGs and advanced cardiac imaging such as CMR or echocardiography. While the Yale New Haven Health System serves a large and diverse population in Connecticut and Rhode Island, prospective validation of the models is necessary before widespread deployment in a screening setting. Second, despite the model's high discrimination for HCM, its implementation for community-based screening can result in high false positive rates. Since screening for low prevalence conditions inherently limits the yield for diagnostic modalities, a role for universal HCM screening must be established before widespread deployment of the model.[37,38] Third, despite localizing the class-discriminative signals in the ECG image to the left ventricular areas, heatmap analysis may not necessarily capture all the model predictive features, such as the duration of ECG segments, intervals, or ECG waveform morphologies that might have been used in model predictions. Fourth, the model demonstrated a lower sensitivity and higher specificity on the UK-Biobank cohort, which is composed of younger and generally healthier individuals. However, we relied on diagnosis codes to identify HCM cases during testing in the UK Biobank. Our analysis of participants screened as false positives screens showed significantly higher LVMI than true negatives, suggesting that the diagnosis-code-based classification could have missed some cases of HCM. Regardless, depending on the intended result of the screening approach and resource constraints with downstream testing, prediction thresholds for HCM may need to be recalibrated when deployed in such settings.

CONCLUSIONS

We developed and validated a high-performing deep learning-based model that detects HCM from images of clinical 12-lead ECGs. This approach represents an accessible strategy for HCM screening, especially in low-resource settings.

ACKNOWLEDGMENTS

Funding

The study was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (under award K23HL153775 to Dr. Khera) and the Doris Duke Charitable Foundation (under award, 2022060 to Dr. Khera). The funders had no role in the design and conduct of the protocol; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflict of Interest Disclosures

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” and are co-founders of Ensight-AI. Dr. Khera receives support from the National Heart, Lung, and Blood Institute of the National Institutes of Health (under award K23HL153775) and the Doris Duke Charitable Foundation (under award 2022060). He receives support from the Blavatnik Foundation through the Blavatnik fund for Innovation at Yale. He also receives research support, through Yale, from Bristol-Myers Squibb, and Novo Nordisk. He is an Associate Editor at JAMA. In addition to 63/346,610, Dr. Khera is a coinventor of U.S. Provisional Patent Applications 63/177,117, 63/428,569, and 63/484,426. Dr. Khera and Dr. Oikonomou are co-founders of Evidence2Health, a precision health platform to improve evidence-based cardiovascular care. Dr. Oikonomou is a co-inventor of the U.S. Patent Applications 63/508,315 & 63/177,117 and has served as a consultant to Caristo Diagnostics Ltd (all outside the current work).

Data Availability

The dataset cannot be made publicly available because they are electronic health records. Sharing this data externally without proper consent could compromise patient privacy and would violate the Institutional Review Board's approval for the study.

Code Availability

The code for the study is available from the authors upon request.

REFERENCES

1. Maron BJ. Hypertrophic cardiomyopathy. *JAMA*. 2002;287:1308–1320.

2. Maron BJ, Maron MS. Hypertrophic cardiomyopathy. *Lancet*. 2013;381:242–255.

3. Semsarian C, Ingles J, Maron MS, Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2015;65:1249–1254.

4. Sabater-Molina M, Pérez-Sánchez I, Hernández del Rincón JP, Gimeno JR. Genetics of hypertrophic cardiomyopathy: A review of current state. *Clin Genet*. 2018;93:3–14.

5. Geske JB, Ommen SR, Gersh BJ. Hypertrophic cardiomyopathy. *JACC Heart Fail*. 2018;6:364–375.

6. Maron BJ, Desai MY, Nishimura RA, Spirito P, Rakowski H, Towbin JA, Rowin EJ, Maron MS, Sherrid MV. Diagnosis and evaluation of hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2022;79:372–389.

7. Writing Committee Members, Ommen SR, Mital S, Burke MA, Day SM, Deswal A, Elliott P, Evanovich LL, Hung J, Joglar JA, Kantor P, Kimmelstiel C, Kittleson M, Link MS, Maron MS, Martinez MW, Miyake CY, Schaff HV, Semsarian C, Sorajja P. 2020 AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy. *Circulation* [Internet]. 2020;142. Available from: http://dx.doi.org/10.1161/cir.0000000000000937

8. Goto S, Solanki D, John JE, Yagi R, Homilius M, Ichihara G, Katsumata Y, Gaggin HK, Itabashi Y, MacRae CA, Deo RC. Multinational Federated Learning Approach to Train

ECG and Echocardiogram Models for Hypertrophic Cardiomyopathy Detection. *Circulation*. 2022;146:755–769.

9. Ko W-Y, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, Demuth SJ, Ackerman MJ, Gersh BJ, Arruda-Olson AM, Geske JB, Asirvatham SJ, Lopez-Jimenez F, Nishimura RA, Friedman PA, Noseworthy PA. Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram. *J Am Coll Cardiol*. 2020;75:722–733.

10. Finocchiaro G, Sheikh N, Biagini E, Papadakis M, Maurizi N, Sinagra G, Pelliccia A, Rapezzi C, Sharma S, Olivotto I. The electrocardiogram in the diagnosis and management of patients with hypertrophic cardiomyopathy. *Heart Rhythm*. 2020;17:142–151.

11. McLeod CJ, Ackerman MJ, Nishimura RA, Tajik AJ, Gersh BJ, Ommen SR. Outcome of patients with hypertrophic cardiomyopathy and a normal electrocardiogram. *J Am Coll Cardiol*. 2009;54:229–233.

12. Siontis KC, Liu K, Bos JM, Attia ZI, Cohen-Shelly M, Arruda-Olson AM, Zanjirani Farahani N, Friedman PA, Noseworthy PA, Ackerman MJ. Detection of hypertrophic cardiomyopathy by an artificial intelligence electrocardiogram in children and adolescents. *Int J Cardiol*. 2021;340:42–47.

13. Magnusson P, Palm A, Branden E, Mörner S. Misclassification of hypertrophic cardiomyopathy: validation of diagnostic codes. *Clin Epidemiol*. 2017;9:403–410.

14. Farahani NZ, Arunachalam SP, Sundaram DSB, Pasupathy K, Enayati M, Arruda-Olson AM. Explanatory analysis of a machine learning model to identify hypertrophic cardiomyopathy patients from EHR using diagnostic codes. *Proceedings (IEEE Int Conf Bioinformatics Biomed)*. 2020;2020:1932–1937.

15. Daskalov IK, Dotsinsky IA, Christov II. Developments in ECG acquisition, preprocessing, parameter measurement, and recording. *IEEE Eng Med Biol Mag*. 1998;17:50–58.

16. Blanco-Velasco M, Weng B, Barner KE. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Comput Biol Med*. 2008;38:1–13.

17. Sangha V, Mortazavi BJ, Haimovich AD, Ribeiro AH, Brandt CA, Jacoby DL, Schulz WL, Krumholz HM, Ribeiro ALP, Khera R. Automated multilabel diagnosis on electrocardiographic images and signals. *Nat Commun*. 2022;13:1583.

18. Sangha V, Nargesi AA, Dhingra LS, Khunte A, Mortazavi BJ, Ribeiro AH, Banina E, Adeola O, Garg N, Brandt CA, Miller EJ, Ribeiro ALJ, Velazquez EJ, Giatti L, Barreto SM, Foppa M, Yuan N, Ouyang D, Krumholz HM, Khera R. Detection of Left Ventricular Systolic Dysfunction From Electrocardiographic Images. *Circulation* [Internet]. 2023;Available from: http://dx.doi.org/10.1161/CIRCULATIONAHA.122.062646

19. ECG Plot Python Library. Accessed at https://pypi.org/project/ecg-plot/ on May 25, 2022.

20. Mingxing Tan and Quoc V Le. EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning, 2019*.

21. Sangha V, Khunte A, Holste G, Mortazavi BJ, Wang Z, Oikonomou EK, Khera R. Biometric Contrastive Learning for data-efficient deep learning from electrocardiographic images [Internet]. bioRxiv. 2023 [cited 2023 Dec 9];2023.09.13.23295494. Available from: https://www.medrxiv.org/content/10.1101/2023.09.13.23295494v1

22. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *2017 Ieee International Conference on Computer Vision (Iccv)*. 2017;618–626.

23. Canepa M, Fumagalli C, Tini G, Vincent-Tompkins J, Day SM, Ashley EA, Mazzarotto F, Ware JS, Michels M, Jacoby D, Ho CY, Olivotto I, The SHaRe Investigators. Temporal trend of age at diagnosis in hypertrophic cardiomyopathy. *Circ Heart Fail* [Internet]. 2020;13. Available from: http://dx.doi.org/10.1161/circheartfailure.120.007230

24. Petersen SE, Aung N, Sanghvi MM, Zemrak F, Fung K, Paiva JM, Francis JM, Khanji MY, Lukaschuk E, Lee AM, Carapella V, Kim YJ, Leeson P, Piechnik SK, Neubauer S. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *J Cardiovasc Magn Reson* [Internet]. 2017;19. Available from: http://dx.doi.org/10.1186/s12968-017-0327-9

25. Naderi H, Ramírez J, van Duijvenboden S, Pujadas ER, Aung N, Wang L, Anwar Ahmed Chahal C, Lekadir K, Petersen SE, Munroe PB. Predicting left ventricular hypertrophy from the 12-lead electrocardiogram in the UK Biobank imaging study using machine learning. *Eur Heart J Digit Health*. 2023;4:316–324.

26. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett*. 2014;21:1389–1393.

27. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.

28. Delcrè SDL, Di Donna P, Leuzzi S, Miceli S, Bisi M, Scaglione M, Caponi D, Conte MR, Cecchi F, Olivotto I, Gaita F. Relationship of ECG findings to phenotypic expression in patients with hypertrophic cardiomyopathy: A cardiac magnetic resonance study. *Int J Cardiol*. 2013;167:1038–1045.

29. Charron P, Dubourg O, Desnos M, Isnard R, Hagege A, Millaire A, Carrier L, Bonne G, Tesson F, Richard P, Bouhour J-B, Schwartz K, Komajda M. Diagnostic value of electrocardiography and echocardiography for familial hypertrophic cardiomyopathy in a genotyped adult population. *Circulation*. 1997;96:214–219.

30. Sharma S, Drezner JA, Baggish A, Papadakis M, Wilson MG, Prutkin JM, La Gerche A, Ackerman MJ, Borjesson M, Salerno JC, Asif IM, Owens DS, Chung EH, Emery MS, Froelicher VF, Heidbuchel H, Adamuz C, Asplund CA, Cohen G, Harmon KG, Marek JC, Molossi S, Niebauer J, Pelto HF, Perez MV, Riding NR, Saarel T, Schmied CM, Shipon DM, Stein R, Vetter VL, Pelliccia A, Corrado D. International recommendations for electrocardiographic interpretation in athletes. *Eur Heart J*. 2018;39:1466–1480.

31. Norrish G, Jager J, Field E, Quinn E, Fell H, Lord E, Cicerchia MN, Ochoa JP, Cervi E, Elliott PM, Kaski JP. Yield of clinical screening for hypertrophic cardiomyopathy in child first-degree relatives. *Circulation*. 2019;140:184–192.

32. Lafreniere-Roula M, Bolkier Y, Zahavich L, Mathew J, George K, Wilson J, Stephenson EA, Benson LN, Manlhiot C, Mital S. Family screening for hypertrophic cardiomyopathy: Is it time to change practice guidelines? *Eur Heart J*. 2019;40:3672–3681.

33. Corrado D, Basso C, Schiavon M, Thiene G. Screening for hypertrophic cardiomyopathy in young athletes. *N Engl J Med*. 1998;339:364–369.

34. Screening program in the athlete in the suspicion of HCM. *J Integr Cardiol* [Internet]. 2015;1. Available from: http://dx.doi.org/10.15761/jic.1000110

35. Anderson BR, McElligott S, Polsky D, Vetter VL. Electrocardiographic screening for hypertrophic cardiomyopathy and long QT syndrome: The drivers of cost-effectiveness for the prevention of sudden cardiac death. *Pediatr Cardiol*. 2014;35:323–331.

36. Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, Albahri OS, Alamoodi AH, Bai J, Salhi A, Santamaría J, Ouyang C, Gupta A, Gu Y, Deveci M. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Inf Fusion*. 2023;96:156–191.

37. O'Toole BI. Screening for low prevalence disorders. *Aust N Z J Psychiatry*. 2000;34:A39–A46.

38. Basavarajaiah S, Wilson M, Whyte G, Shah A, McKenna W, Sharma S. Prevalence of hypertrophic cardiomyopathy in highly trained athletes. *J Am Coll Cardiol*. 2008;51:1033–1039.

FIGURES

Figure 1. Model Development and Study Design Abbreviations: CMR, Cardiac Magnetic Resonance Imaging; HCM, Hypertrophic Cardiomyopathy; IVSd, End-diastolic Interventricular Septal Thickness

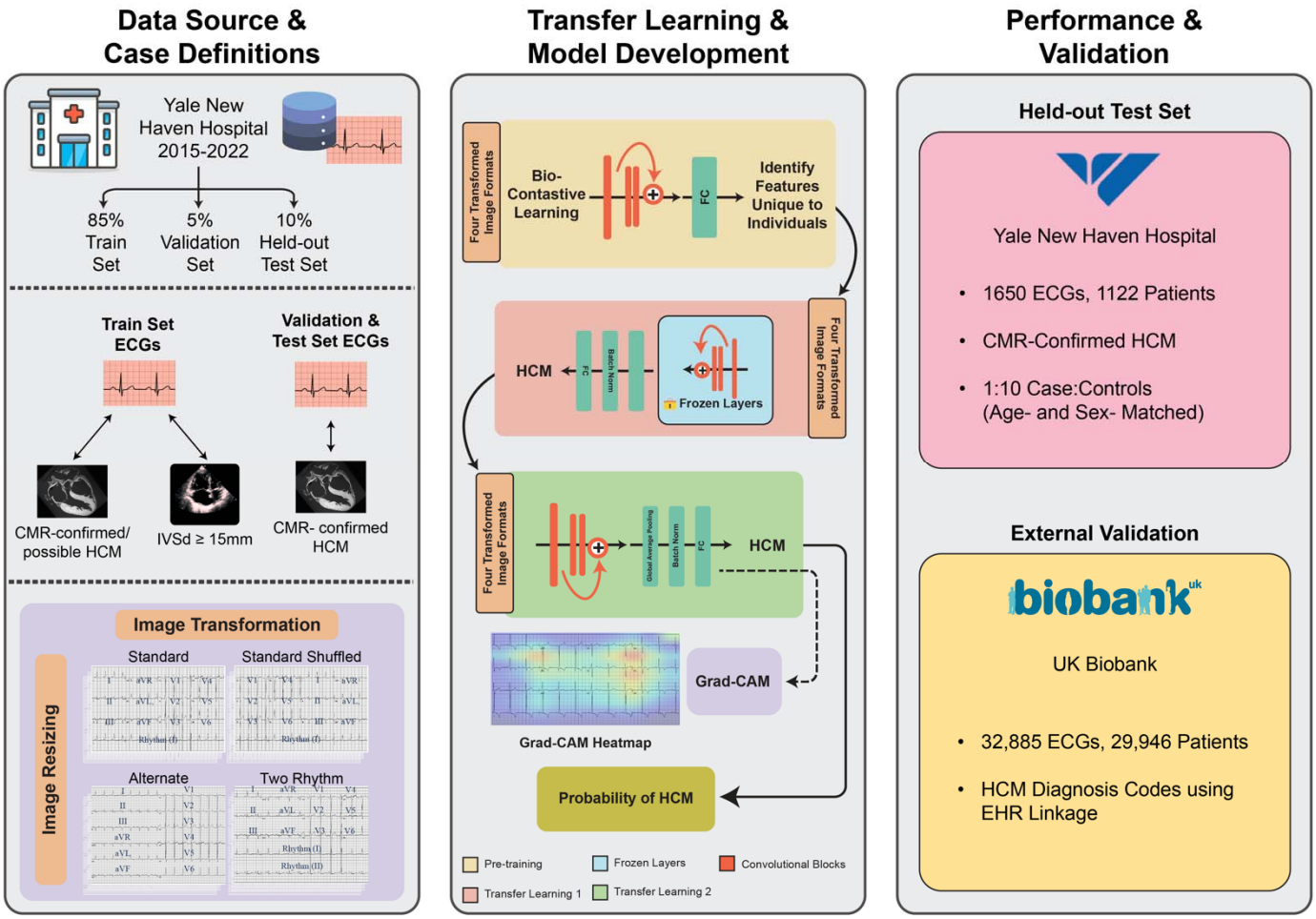
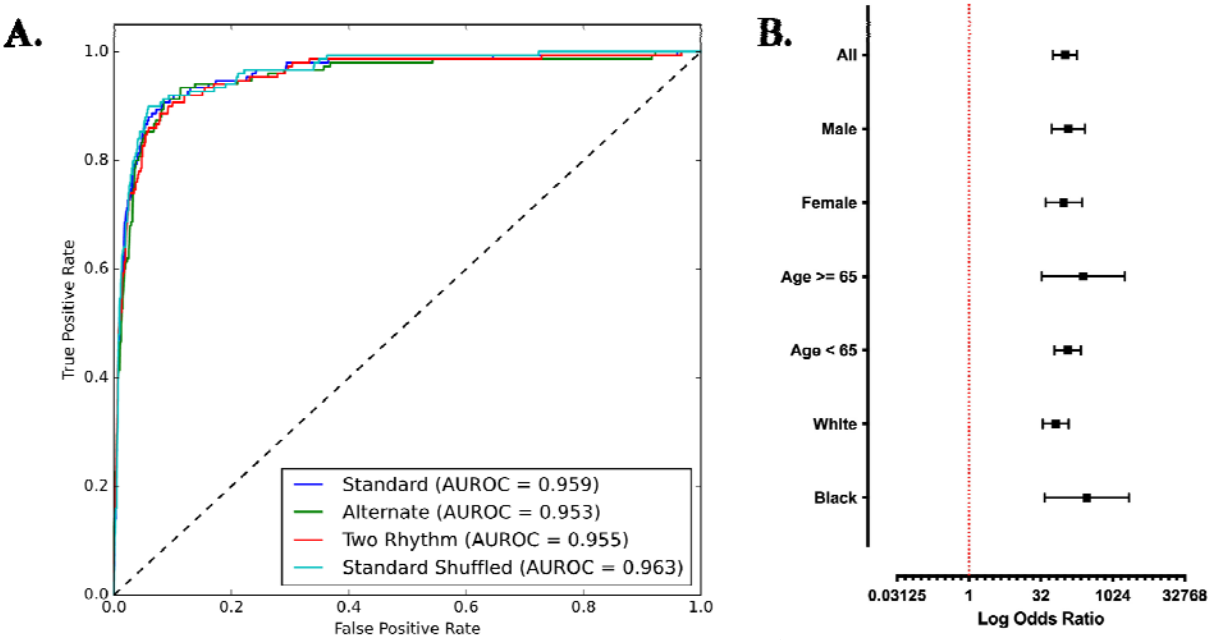


Figure 2. Model performance measures (A) Receiver operating characteristic curves across image formats in the held-out test set. B) Diagnostic odds ratios across age, gender, and race subgroups on standard format images in the held-out test set. Abbreviations: AUROC, area under receiver-operating characteristic curve*



*Note: Diagnostic odds ratios were not available for Hispanic, Asian, or other Races because there were no FN screens in the held-out test set

Figure 3. (A) Distribution of interventricular septal thickness in Yale New Haven Hospital patients without hypertrophic cardiomyopathy. (B). Distribution of left ventricular mass index in UK Biobank without hypertrophic cardiomyopathy.

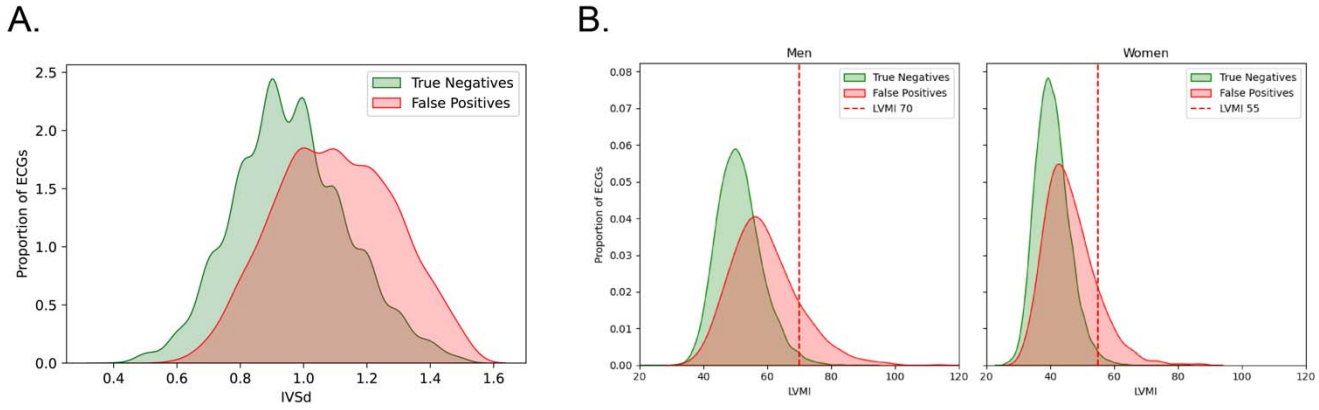


Figure 4. Gradient-weighted Class Activation Mapping (Grad-CAMs) across Electrocardiogram formats. A) Standard format B) Two rhythm leads C) Standard shuffled format D) Alternate format. The heatmaps represent averages of the 25 positive cases with the most confident model predictions for HCM.

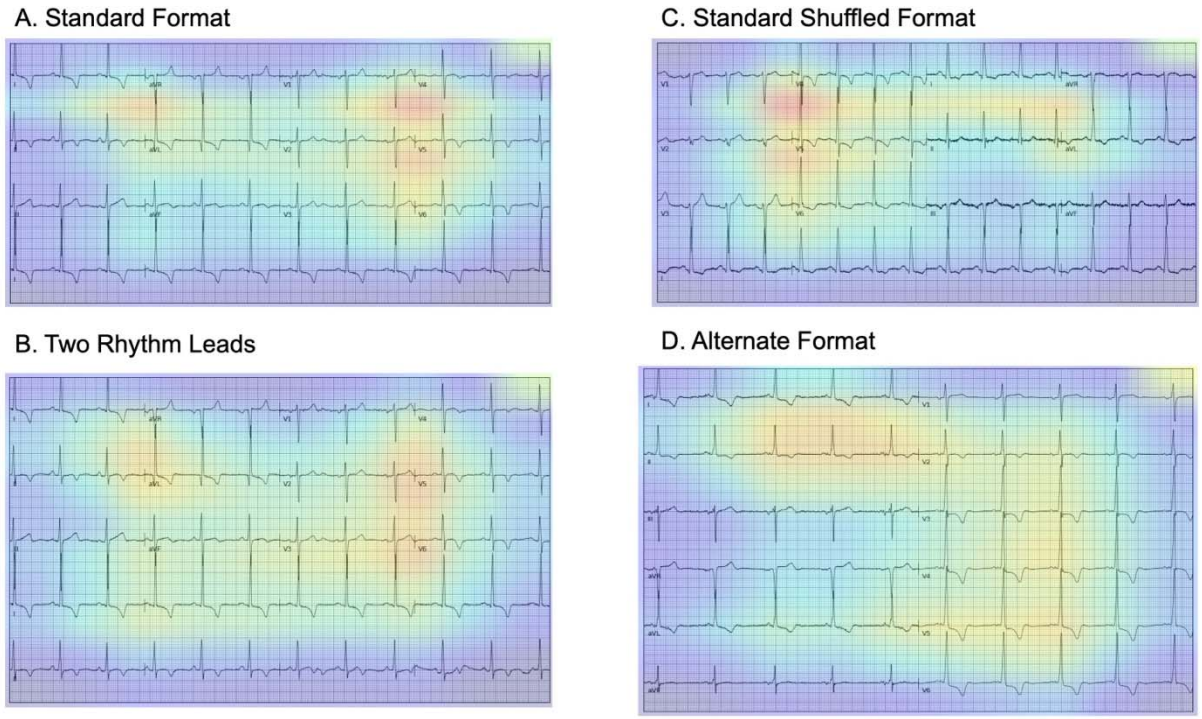


Figure 5. Model Performance in the UK Biobank (A) Receiver Operating Characteristic Curves. B) Examples of Gradient-weighted Class Activation Mapping (Grad-CAM) analysis of electrocardiograms from two individuals with HCM in UK Biobank.
Abbreviations: AUROC, area under receiver-operating characteristic curve

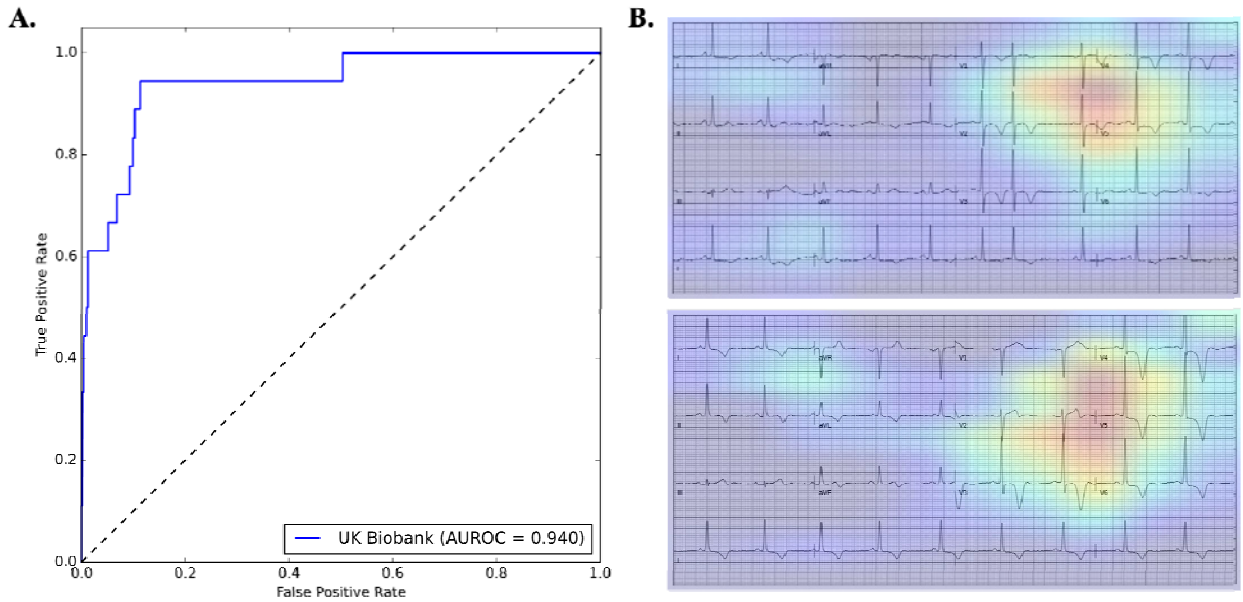
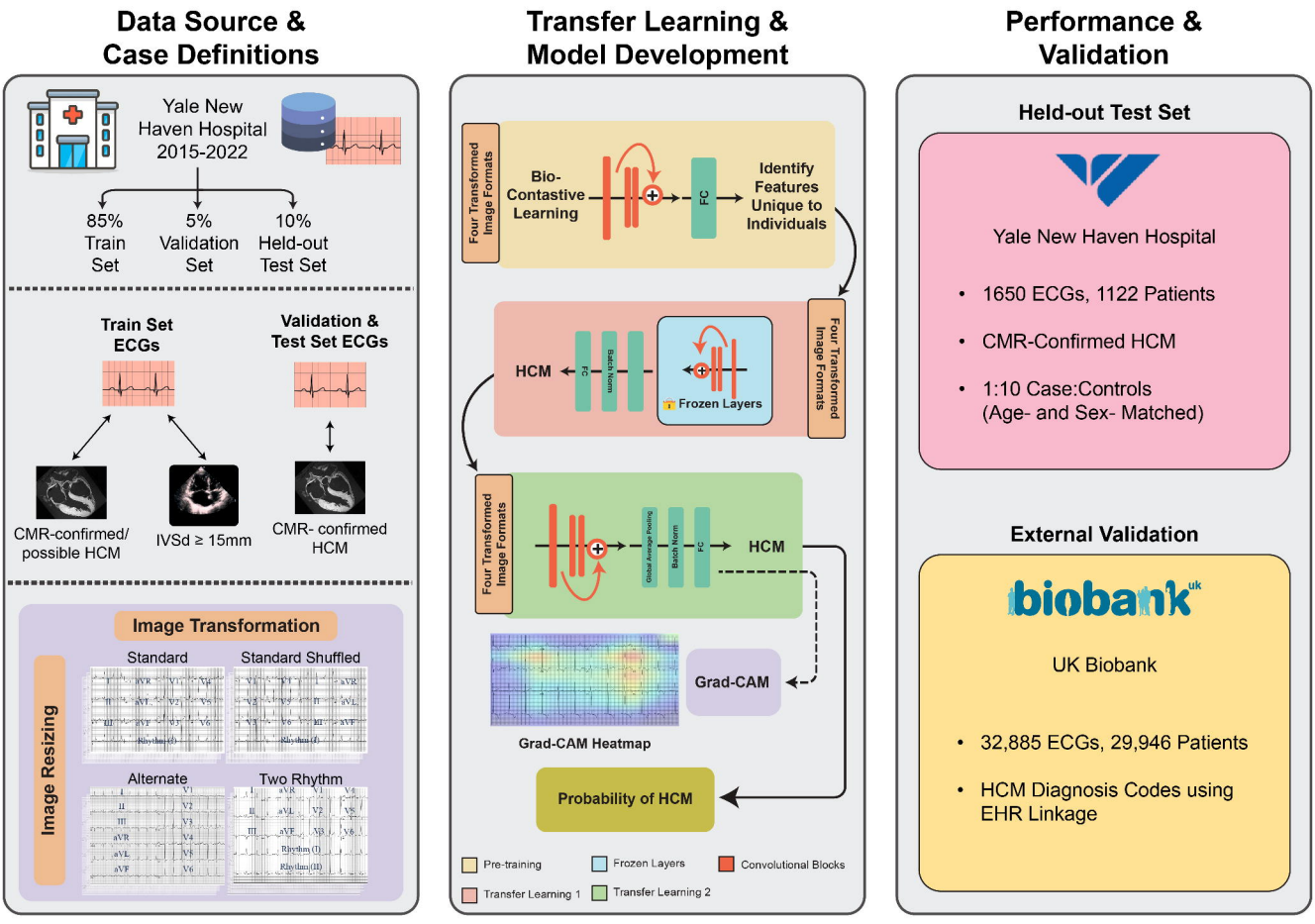


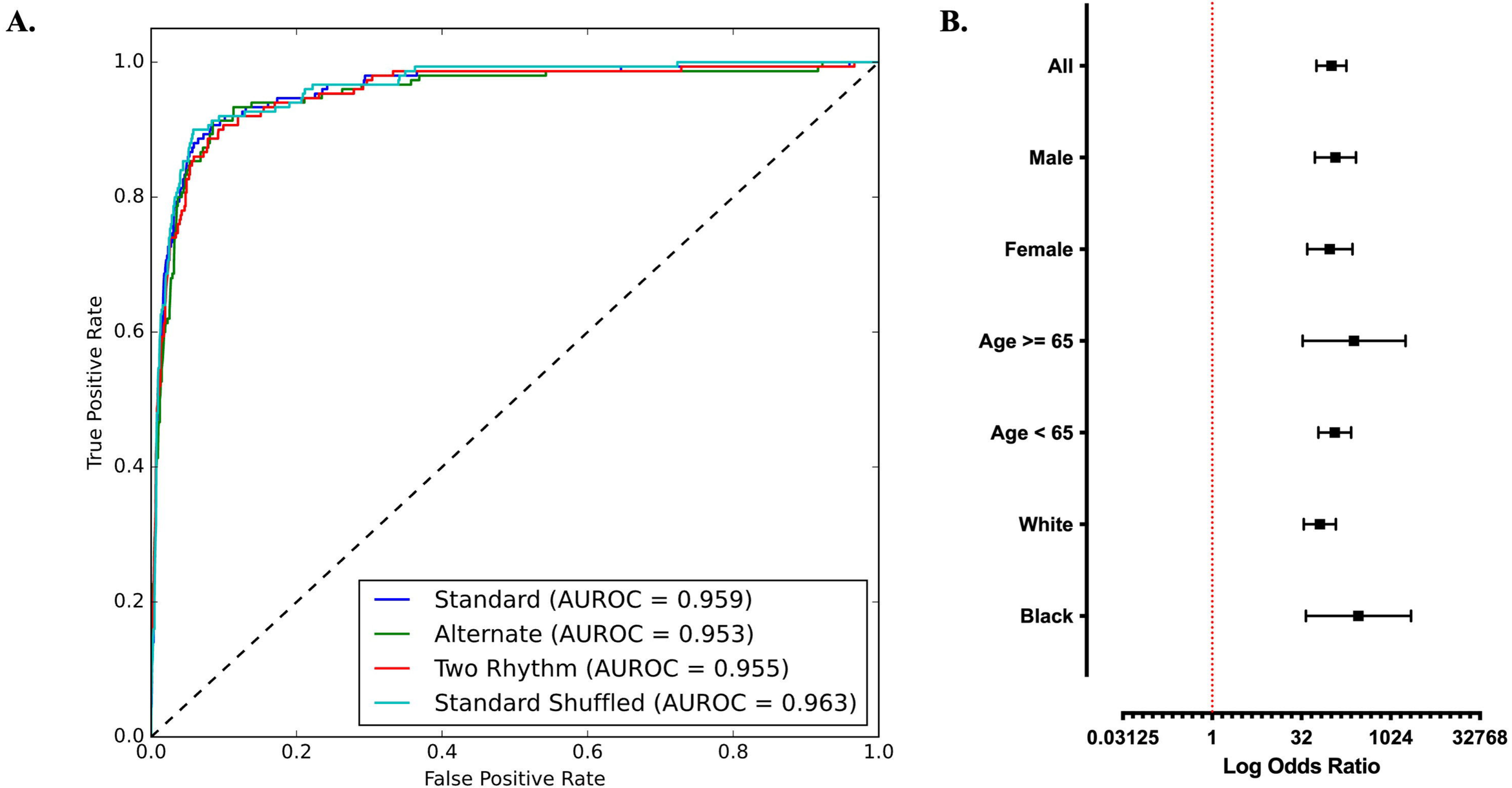
Table 1. Baseline characteristics of study population. Data presented as median [IQR] for age and number (percent) for other variables. Abbreviations: CMR, cardiac magnetic resonance imaging; ECGs, electrocardiograms; HCM, hypertrophic cardiomyopathy; IVSd, interventricular septal thickness in diastole.

Characteristic	**Train**		**Validation**		**Test**	
	Patients	**ECGs**	**Patients**	**ECGs**	**Patients**	**ECGs**
Number	66339	123640	648	913	1122	1650
Sex; N(%)						
Female	33044 (49.8)	58390 (47.2)	213 (32.9)	291 (31.9)	372 (33.2)	542 (32.8)
Male	33013 (49.8)	64675 (52.3)	423 (65.3)	606 (66.4)	729 (65.0)	1076 (65.2)
Missing	282 (0.4)	575 (0.5)	12 (1.9)	16 (1.8)	21 (1.9)	32 (1.9)
Age (years); Median[IQR]	63.3 [51.2,74.2]	64.1 [52.5,75.0]	58.7 [52.8,67.4]	60.5 [53.2,67.6]	56.9 [44.8,67.2]	58.4 [45.6,68.4]
Race; N(%)						
Asian	1289 (1.9)	2286 (1.8)	11 (1.7)	15 (1.6)	27 (2.4)	35 (2.1)
Black	7892 (11.9)	15984 (12.9)	86 (13.3)	122 (13.4)	146 (13.0)	216 (13.1)
Hispanic	5964 (9.0)	11297 (9.1)	65 (10.0)	84 (9.2)	121 (10.8)	170 (10.3)
White	48038 (72.4)	88711 (71.7)	437 (67.4)	628 (68.8)	743 (66.2)	1107 (67.1)
Other	371 (0.6)	702 (0.6)	5 (0.8)	7 (0.8)	9 (0.8)	14 (0.8)
Unknown	2785 (4.2)	4660 (3.8)	44 (6.8)	57 (6.2)	76 (6.8)	108 (6.5)
HCM; N(%)	4715 (7.1)	12597 (10.0)	30 (4.6)	83 (9.1)	59 (5.3)	150 (9.1)
CMR-Confirmed HCM	505 (0.8)	1887 (1.5)	30 (4.6)	83 (9.1)	59 (5.3)	150 (9.1)
CMR-Possible HCM	79 (0.1)	287 (0.2)	-	-	-	-
IVSd > 15mm	4131 (6.2)	10190 (8.3)	-	-	-	-

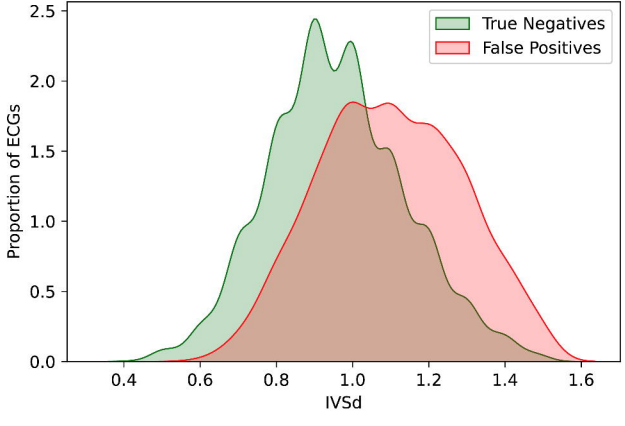
Table 2. Performance of model on test images across demographic subgroups in the age sex matched held-out test set.
Abbreviations: PPV, positive predictive value; NPV, negative predictive value; AUROC, area under receiver operating characteristic curve; AUPRC, area under precision recall curve; A-Fib, atrial fibrillation; ECG, electrocardiogram; LBBB, left bundle branch block; RBBB, right bundle branch block; LVH, left ventricular hypertrophy.

Labels	Number	PPV	NPV	Specificity	Sensitivity	AUROC	AUPRC	F1 Score
All	1650	0.511	0.99	0.913	0.907	0.959 (0.941-0.977)	0.754 (0.678-0.828)	0.654
Male	1076 (65.2%)	0.482	0.992	0.900	0.93	0.958 (0.933-0.983)	0.743 (0.642-0.839)	0.635
Female	542 (32.8%)	0.589	0.985	0.940	0.86	0.966 (0.947-0.986)	0.807 (0.691-0.901)	0.699
>=65	506 (30.7%)	0.388	0.997	0.846	0.978	0.958 (0.938-0.978)	0.633 (0.506-0.793)	0.556
<65	1144 (69.3%)	0.607	0.987	0.943	0.875	0.96 (0.937-0.983)	0.796 (0.705-0.87)	0.717
Hispanic	170 (10.3%)	0.462	1	0.957	1	1 (1-1)	1 (1-1)	0.632
White	1107 (67.1%)	0.48	0.986	0.909	0.867	0.943 (0.917-0.97)	0.673 (0.568-0.77)	0.618
Black	216 (13.1%)	0.661	0.993	0.882	0.975	0.986 (0.974-0.999)	0.941 (0.871-0.986)	0.788
Asian	35 (2.1%)	0.4	1	0.897	1	0.983 (0.935-1)	0.833 (0.333 - 1.0)	0.571
Other	14 (0.8%)	0	1	0.946	-	-	-	-
Unknown	108 (6.5%)	0.444	1	0.928	1	0.971 (0.933-1)	0.560 (0.25-1.0)	0.615
Paced ECGs	28 (1.7%)	0.409	1	0.316	1	0.655 (0.45-0.86)	0.406 (0.234-0.716)	0.581
Not Paced ECGs	1622 (9.8%)	0.52	0.99	0.921	0.901	0.961 (0.943-0.98)	0.803 (0.74-0.863)	0.66
A-Fib or Flutter	75 (4.5%)	0.324	1	0.641	1	0.886 (0.812-0.961)	0.424 (0.237-0.708)	0.489
No A-Fib or Flutter	1575 (95.5%)	0.539	0.99	0.925	0.899	0.962 (0.943-0.981)	0.795 (0.718-0.86)	0.674
LBBB	19 (1.2%)	0.353	1	0.154	1	0.872 (0.706-1)	0.719 (0.365-1.0)	0.522
No LBBB	1600 (97.0%)	0.523	0.992	0.921	0.920	0.963 (0.945-0.982)	0.772 (0.692-0.841)	0.667
RBBB	71 (4.3%)	0.5	1	0.855	1	0.996 (0.988-1)	0.98 (0.904-1)	0.667
No RBBB	1548 (93.8%)	0.512	0.992	0.916	0.919	0.961 (0.943-0.98)	0.758 (0.679-0.836)	0.658
LVH	221 (13.4%)	0.649	0.982	0.735	0.973	0.959 (0.934-0.984)	0.916 (0.855-0.965)	0.778
No LVH	1429 (86.6%)	0.413	0.991	0.933	0.842	0.942 (0.91-0.974)	0.543 (0.442-0.669)	0.554

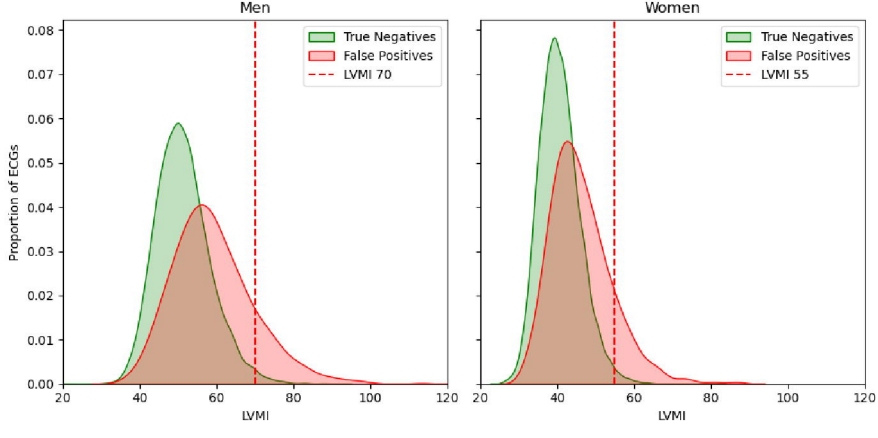

Data Source & Case Definitions
Yale New Haven Hospital 2015-2022
85% Train Set
5% Validation Set
10% Held-out Test Set
Train Set ECGs
Validation & Test Set ECGs
CMR-confirmed/ possible HCM
IVSd ≥ 15mm
CMR- confirmed HCM
Image Transformation
Image Resizing
Standard
Standard Shuffled
Alternate
Two Rhythm
Transfer Learning & Model Development
Four Transformed Image Formats
Bio-Contastive Learning
FC
Identify Features Unique to Individuals
HCM
Batch Norm
Frozen Layers
Global Average Pooling
Grad-CAM
Grad-CAM Heatmap
Probability of HCM
Pre-training
Frozen Layers
Convolutional Blocks
Transfer Learning 1
Transfer Learning 2
Performance & Validation
Held-out Test Set
Yale New Haven Hospital
1650 ECGs, 1122 Patients
CMR-Confirmed HCM
1:10 Case:Controls (Age- and Sex- Matched)
External Validation
biobank uk
UK Biobank
32,885 ECGs, 29,946 Patients
HCM Diagnosis Codes using EHR Linkage


A.
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Standard (AUROC = 0.959)
Alternate (AUROC = 0.953)
Two Rhythm (AUROC = 0.955)
Standard Shuffled (AUROC = 0.963)
B.
All
Male
Female
Age >= 65
Age < 65
White
Black
0.03125
1
32
1024
32768
Log Odds Ratio

A.


True Negatives
False Positives
Proportion of ECGs
IVSd
0.0
0.5
1.0
1.5
2.0
2.5
0.4
0.6
0.8
1.0
1.2
1.4
1.6

B.

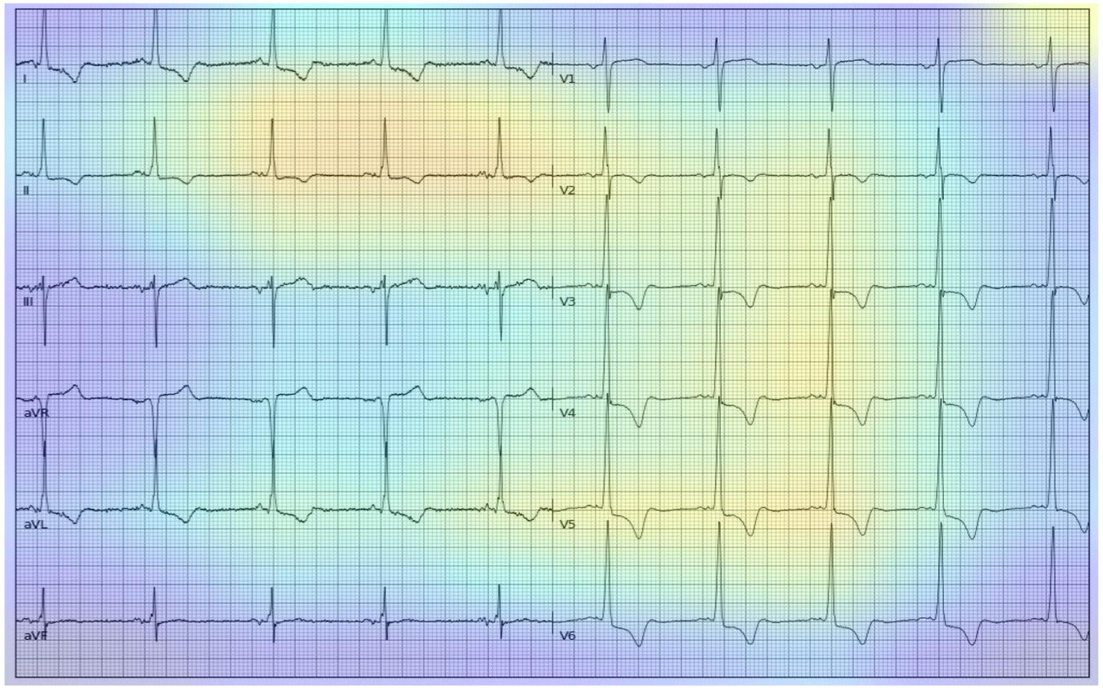

Men
Women
True Negatives
False Positives
LVMI 70
LVMI 55
Proportion of ECGs
LVMI
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
20
40
60
80
100
120

A. Standard Format

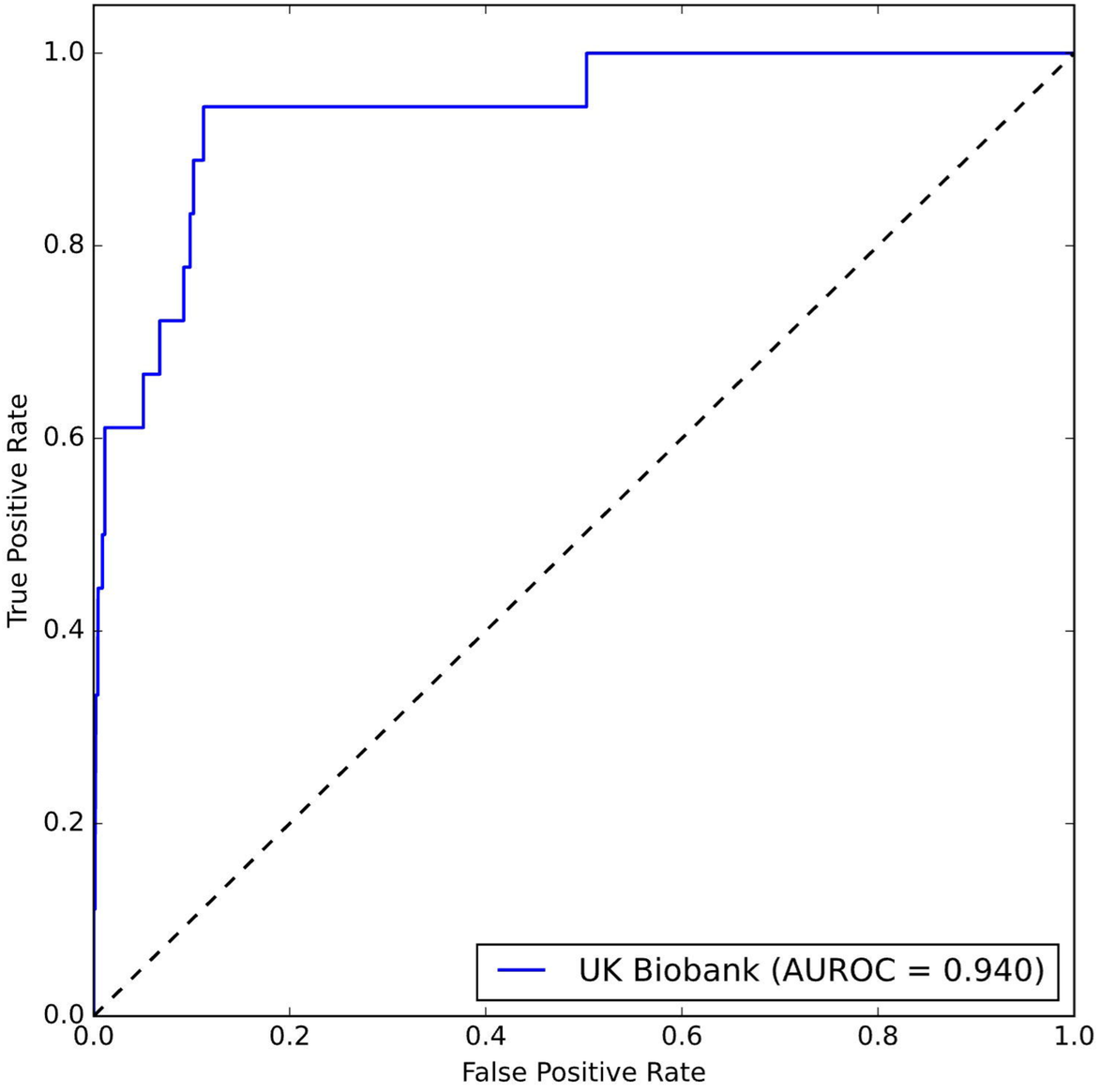
C. Standard Shuffled Format

B. Two Rhythm Leads

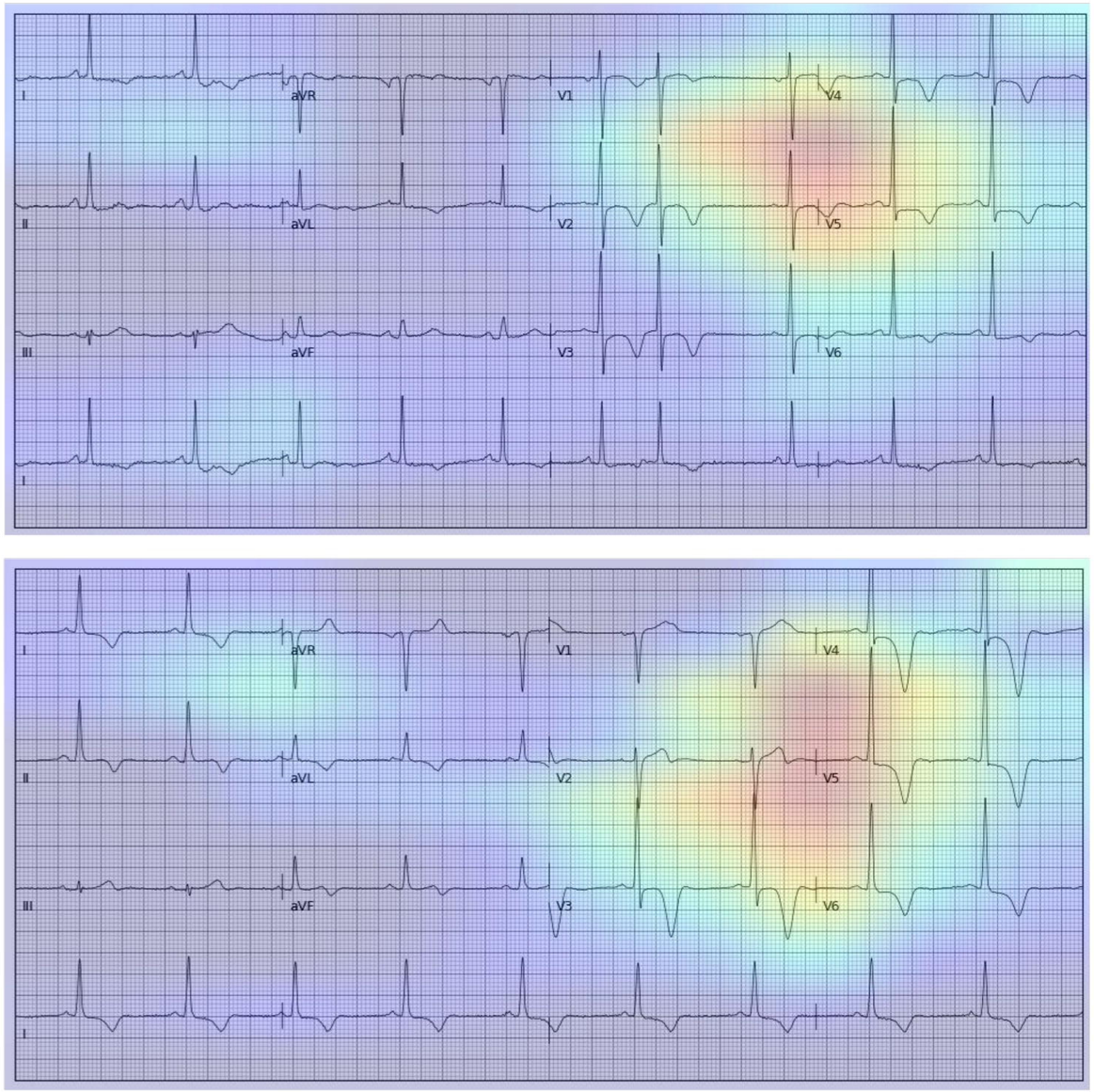
D. Alternate Format



A.


1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
UK Biobank (AUROC = 0.940)

B.


I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6
I
I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6
I