Artificial intelligence-guided detection of under-recognized cardiomyopathies on point-of-care cardiac ultrasound

Evangelos K. Oikonomou MD DPhil[a], Gregory Holste BA[b], Andreas Coppi PhD[c], Robert L. McNamara MD MHS[a], Girish N. Nadkarni MD MPH[d,e], Cristiana Baloescu[f], Harlan M. Krumholz MD SM[a,c], Zhangyang Wang PhD[b], Rohan Khera MD MS[a,c,g,h,i*]

[a] Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
[b] Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA
[c] Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA
[d] The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[e] Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[a] Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, USA
[g] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[h] Section of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT
[i] Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, CT

Manuscript type: Original Research Manuscript
Word count: 3,756 words
Figures: 3
Tables: 2

[*]Corresponding author:
Rohan Khera, MD, MS
195 Church St, 6th Floor, New Haven, CT 06510
203-764-5885; rohan.khera@yale.edu; @rohan_khera

ABSTRACT

Background: Point-of-care ultrasonography (POCUS) enables access to cardiac imaging directly at the bedside but is limited by brief acquisition, variation in acquisition quality, and lack of advanced protocols.

Objective: To develop and validate deep learning models for detecting underdiagnosed cardiomyopathies on cardiac POCUS, leveraging a novel acquisition quality-adapted modeling strategy.

Methods: To develop the models, we identified transthoracic echocardiograms (TTEs) of patients across five hospitals in a large U.S. health system with transthyretin amyloid cardiomyopathy (ATTR-CM, confirmed by Tc^{99m}-pyrophosphate imaging), hypertrophic cardiomyopathy (HCM, confirmed by cardiac magnetic resonance), and controls enriched for the presence of severe AS. In a sample of 290,245 TTE videos, we used novel augmentation approaches and a customized loss function to weigh image and view quality to train a multi-label, view agnostic video-based convolutional neural network (CNN) to discriminate the presence of ATTR-CM, HCM, and/or AS. Models were tested across 3,758 real-world POCUS videos from 1,879 studies in 1,330 independent emergency department (ED) patients from 2011 through 2023.

Results: Our multi-label, view-agnostic classifier demonstrated state-of-the-art performance in discriminating ATTR-CM (AUROC 0.98 [95%CI: 0.96-0.99]) and HCM (AUROC 0.95 [95% CI: 0.94-0.96]) on standard TTE studies. Automated metrics of anatomical view correctness confirmed significantly lower quality in POCUS vs TTE videos (median view classifier confidence of 0.63 [IQR: 0.44-0.88] vs 0.93 [IQR: 0.69-1.00], $p<0.001$). When deployed to POCUS videos, our algorithm effectively discriminated ATTR-CM and HCM with AUROC of

up to 0.94 (parasternal long-axis (PLAX)), and 0.85 (apical 4 chamber), corresponding to positive diagnostic odds ratios of 46.7 and 25.5, respectively. In total, 18/35 (51.4%) of ATTR-CM and 32/57 (41.1%) of HCM patients in the POCUS cohort had an AI-positive screen in the year before their eventual confirmatory imaging.

Conclusions: We define and validate an AI framework that enables scalable, opportunistic screening of under-diagnosed cardiomyopathies using POCUS.

INTRODUCTION

Point-of-care ultrasonography (POCUS) enables cardiac phenotyping at the point-of-care as a direct extension of the physical exam,[1] and is increasingly used as an adjunctive diagnostic tool across outpatient clinics, emergency rooms and inpatient facilities.[2,3] Nonetheless, POCUS studies often adopt an opportunistic approach with incomplete protocols and place less emphasis on image quality or anatomical accuracy compared to standard transthoracic echocardiograms (TTE).[4] Consequently the generated videos and images are rarely used for reasons beyond addressing acute medical questions.

With the expanding use of handheld and portable technologies and algorithms that can assist novice operators in acquiring standard echocardiographic views,[5] there is a growing realization that POCUS imaging represents a missed window for detecting potentially modifiable, chronic cardiac disease. Further to quantifying left ventricular function or ruling our acute, life-threatening findings, the acquired POCUS data may, in theory, enable the opportunistic screening of underdiagnosed cardiomyopathies such as hypertrophic cardiomyopathy (HCM),[6–8] or transthyretin amyloid cardiomyopathy (ATTR-CM),[9,10] effectively discriminating these from controls or other conditions associated with left ventricular hypertrophy, such as aortic stenosis (AS).[11–14] In fact, it is known that only a minority (10-20%) of these cases are identified clinically,[6,7,11,15–17] with disparities in outpatient diagnosis and care disproportionately affecting individuals from lower socioeconomic classes and racial/ethnic minorities.[18–20]

In the recent years, echocardiography has benefited from advances in computer vision and medical artificial intelligence (AI) which not only provide automated summaries of routine TTE studies, but also augment our ability to phenotype cardiovascular pathology,[21–27] with

improved accuracy relative to standard workflows.[28,29] Nevertheless, such AI algorithms are almost invariantly developed and validated using videos acquired by expert technicians and interpreted by board-certified readers and are not adapted for use with POCUS studies and their unique challenges.

Here, we define and implement a framework for POCUS-adapted video-based AI models, demonstrating their ability to efficiently detect under-diagnosed cardiomyopathies from real-world POCUS videos acquired over a decade across the emergency rooms of a large hospital system. Our approach harnesses the availability of large TTE repositories and incorporates a range of natural and synthetic augmentation methods to simulate off-axis acquisitions from variable echocardiographic views, thus enabling the downstream view-agnostic inference from limited POCUS acquisitions.

METHODS

Study population

The study was designed as a case-control analysis of patients seeking care across five distinct hospitals, emergency rooms, and outpatient clinics affiliated with the Yale-New Haven Health System (YNHHS) in Connecticut and Rhode Island. We identified patients with HCM, ATTR-CM, and controls, enriched for conditions with phenotypic overlap (such as, severe AS), all with available TTE and/or ED POCUS videos.

First, we queried the electronic health record (EHR) and linked YNHHS echocardiographic library (n=522,507 unique studies) for TTE exams performed between 5/12/2015 (database inception) and 12/31/2022 in patients with HCM, ATTR-CM, AS, and controls. We chose these conditions as representative examples due to their often indolent course

and delayed diagnosis, overlapping phenotypic features yet distinct underlying pathology, and given that they are not routinely discernable on two-dimensional POCUS. We used the following non-mutually exclusive definitions:

HCM: All individuals with an ICD-9/ICD-10 code for any cardiomyopathy (425, I42.0, I42.1, I42.2, I42.5, I42.8, I42.9, I43.1, I43.8) or heart failure (428, I50*), inclusive of HCM-specific ICD-9/10 codes (425.1, 425.11, 425.18, I42.1, I42.2) were identified. To maximize the specificity of our definition, given the known unreliability of billing/administrative codes in accurately capturing patient phenotypes,[30] we required positive labels for this condition to have undergone cardiac magnetic resonance (CMR) imaging with the final interpretation/conclusion supporting the presence of the diagnosis.[31] Given that HCM is a genetic cardiomyopathy, we included all available echocardiograms regardless of their timing relative to the time of diagnosis.[6]

Amyloid cardiomyopathy (ATTR-CM): We screened for all individuals with an ICD-based diagnosis of any cardiomyopathy or heart failure as above or an amyloidosis-specific code (277.3, 277.30, 277.39, E85.2, E85.82, E85.4, E85.8, E85.9, excluding E85.81 [light chain amyloidosis]). Similar to HCM, to increase the specificity of the label, we required positive labels to have undergone bone scintigraphy (with Tc^{99m}-pyrophosphate [PYP]), which was interpreted as positive for cardiac uptake by the interpreting physician (i.e., a semi-quantitative visual score of 2 or 3 or heart to contralateral lung ratio >1.5).[32] For positive cases, we defined the time of diagnosis as the time of the positive PYP scan, and, to account for the delay between disease onset and diagnosis (median delay of ~13 months as previously reported in the

literature),[33] we included echocardiograms performed up to 12 months before this date (and any time after).

Controls: These were defined by randomly sampling echocardiograms from the same period, after excluding any positive HCM or ATTR-CM cases, and after excluding intermediate phenotypes (i.e., CMR findings suggestive of possible HCM, or equivocal PYP results). The study sample was further enriched for cases of severe AS, including severe low-flow, low-gradient AS, based on the interpretation of a TTE exam by a board-certified reader and in agreement with existing guidelines.[34,35] This was done to ensure the model learned to identify AS, a separate pathology, rather than a confounder of ATTR-CM or other cardiomyopathies.

Development (TTE) and testing (POCUS) sets

Development cohort: We developed our models using TTE videos (all views included) from participants who did not contribute data to the ED POCUS arm. This ensured that there would be no data leakage between TTE and POCUS studies performed in the same individual. After assigning HCM, ATTR-CM, and AS labels as reviewed above, we randomly split our development cohort at the patient level into a training 75%, validation (15%), and *internal* testing (10%) set. This ensured that distinct echocardiograms belonging to the same patient would not be present across more than one of the three (training, validation, testing) sets. To further boost the validity of the observations, we excluded any echocardiographic study with a measured interventricular septal thickness during diastole (IVSd) of 1.3 cm or greater from the controls of our training set. However, this was not done during validation or testing to ensure a reliable assessment of the model's performance.

Testing (POCUS) cohort: We identified 16,487 participants (both cases and controls) who underwent a cardiac POCUS by an emergency provider between 2/2011 and 10/2023. Such studies are performed by emergency department staff using a general-purpose ultrasound system (Sparq Ultrasound system, Philips) and are not read by echocardiographers. We excluded any POCUS studies performed after the onset of end-stage renal disease ('585.6', 'N18.6'), after heart transplantation ('V42.1', 'Z94.1', 'Z94.3), and/or aortic valve replacement (ICD procedure codes '35.21', '35.22', '02RF07Z', '02RF08Z', '02RF0JZ', '02RF0KZ'), as well as POCUS studies that exclusively included non-cardiac imaging (i.e., lung ultrasound) with none of the major cardiac views typically captured in a cardiac POCUS (at least one of parasternal long axis [PLAX], parasternal short axis at the papillary muscle level [PSAX] and/or apical four-chamber [A4C] views). Our approach resulted in 1,879 unique studies in 1,330 unique patients with 3,758 cardiac-focused views performed between 2/23/2011 and 10/30/2023.

Automated view characterization and alignment assessment

We implemented our previously published end-to-end pre-processing pipeline on TTE studies stored in DICOM format,[22,23] that involves loading the pixel data, masking our pixels in the periphery to remove identifying information and converting to Audio Video Interleave (.AVI) format. We then randomly sample ten frames from each video, down-sample to 224x224 pixels, and fed through a previously validated VGG19 convolutional neural network (CNN) that enables video-level classification of 18 echocardiographic views by assigning a probability that a given video corresponds to a standard anatomical view (with probabilities adding up to 1 across all views).[37] A predicted view was then assigned based on the view class that has the highest

probability. The highest probability value (0-1) is then used to define a metric of anatomical alignment with standard echocardiographic views. In other words, greater anatomical correctness and view quality was associated with higher confidence in the CNN model's output. Next, we performed more thorough cleaning and de-identification by binarizing each video frame with a fixed threshold, masking out all pixels outside the convex hull of the largest contour, and down-sampling to 112x112 pixels.

Designing a view-agnostic training pipeline adapted for low-quality acquisitions.

We designed a training framework that incorporated a naïve approach to using multiple views (apical, parasternal long, parasternal short, and subcostal views) without annotations as well as a customized training loss to heavily favor low-quality, off-axis videos of patients with the labels of interest. We first initialized a 3D-ResNet18 CNN architecture by using pre-trained weights from the Kinetics-400 dataset, a large corpus of over 300,000 natural videos for human action classification provided by PyTorch,[38] and further modified the output layer of the label to enable multi-label classification for the three representative labels of interest, namely HCM, ATTR-CM, and AS. To deploy our model in POCUS scans for screening of relatively rare cardiomyopathies, we implemented a range of customizations:

***Natural and synthetic data augmentation methods*:** We trained both separate models for each key views-of-interest, namely PLAX, PSAX, and A4C, followed by all-inclusive, view-naïve models trained in pooled datasets that included all parasternal (long and short) and apical views with the classifier blinded to the input view. This enabled a head-to-head comparison of how view-specific versus view-agnostic approaches generalize to real-world POCUS acquisitions. We

further applied a series of data augmentations to account for variable orientation and off-axis views that included random zero padding by up to 8 pixels in each spatial dimension, random horizontal flipping with (probability 0.5), and a random rotation within -10 and 10 degrees (probability 0.5). After augmentation, each video clip's intensities were normalized to 0-1 and standardized using the channel-wise means and standard deviations from the Kinetics-400 training dataset.

Quality-adjusted weights and loss function: We defined a loss function to prompt the model to learn from lower-quality cases. We took the binary cross entropy (BCE) with logits loss function, that combines a sigmoid layer and the BCE loss function in one class, and incorporated both label-specific weights to account for rare labels,[39] as well as inverse weighting based on the view alignment probabilities (see ***Supplement***).[37] Higher probabilities (i.e., PLAX view probability of 1.00) denote greater anatomical correctness compared with lower ones. The aim was to penalize the model for missing under-represented labels, especially in the context of a challenging view. We also applied label smoothing (α=0.1) to penalize over-confidence in our model's predictions.[40,41]

Model training

All models were trained for a maximum of 30 epochs with early stopping, such that if the mean validation area under the receiver operating characteristic (AUROC) for the three labels of interest did not improve for 5 consecutive epochs, training was terminated and the weights from the epoch with maximum validation AUROC were used for final evaluation. Models were trained on four NVIDIA Tesla T4 GPUs with the Adam optimizer, a learning rate of 10^{-4}, a batch

size of 56 to maximize GPU utilization, and a random dropout of 0.25, using randomly sampled video clips of 16 frames and sampling one out of every five frames to enable a global capture of the cardiac cycle (median number of frames 61 25th-75th percentile: 50-85]). We applied optional padding with empty frames along the temporal axis if either the video was too short or the randomly chosen starting point of the clip was near the end of the video.

Model performance assessment

At the time of inference, we averaged four 16-frame-clip-level predictions to obtain video-level predictions for each label. We analyzed the performance of video- and study-level predictions by averaging class-specific probabilities from all videos acquired during the same study. We evaluated the discriminatory performance for HCM, ATTR-CM, and AS using standard metrics, including AUROC, the area under the precision-recall curve (AUPRC), as well as metrics across the thresholds that maximized Youden's J (sum of sensitivity and specificity minus one) as well as 90% sensitivity, including the F1 score, diagnostic odds ratio (OR), positive (PPV) and negative predictive value (NPV) at 3% prevalence, an estimated average of the prevalence of HCM, ATTR-CM, AS in non-randomly selected individuals with a mean age of >65 years with known or suspected cardiovascular disease.[15,16,42–44]

Model explainability

To assist with explainability, we generated sample saliency maps for the most confident cases using Gradient-weighted Class Activation Mapping (Grad-CAM).[45] We present the pixelwise maximum along the temporal axis to capture the most salient regions as a heatmap overlayed on the videos being leveraged in defining the respective conditinos.

Statistical analysis

Categorical variables are summarized as counts (valid percentages), and continuous variables as mean ± standard deviation, or median [25^{th}-75^{th} percentile], unless specified otherwise. Categorical variables were compared across distinct groups using the χ^2 test. Continuous variables were compared using the non-parametric Mann-Whitney or Kruskal-Wallis tests for two or three or more groups, respectively. Metrics of discrimination and corresponding 95% confidence intervals (CI) are derived from bootstrapping with 200 replications. The probability distribution of continuous variables across distinct groups is visualized using density plots, with values compared by the non-parametric Mann-Whitney test. Analyses were performed using Python 3.9.7, using pytorch 1.8.0, torchvision 0.9.0, and scipy 1.7.3. All statistical tests were 2-sided with a significance level of 0.05, unless specified otherwise.

RESULTS

Study population

The development cohort included a total of 10,702 studies with 290,245 echocardiographic videos among 8,460 unique patients, split into a training (n=8,090 patients), validation (n=1,577 patients) and an (internal) testing set (1,035 patients). In the training set, the mean age at the time of echocardiography was 68.3±15 years, 3,752 (46.4%) were women, 6,087 (75.2%) self-reported White race, 612 (7.6%) Black race and (433) 5.4% Hispanic ethnicity. In total, 1,375 studies corresponded to HCM (17.0%), 228 (2.8%) to ATTR-CM and 1,120 (15.0%) to severe AS, with comparable prevalence among the training, validation, and testing splits (**Table 1**).

The ED POCUS testing population included 1,879 studies with 3,758 key echocardiographic views (mean age at the time of the ED visit was 68.3±16.7 years). Reflecting the differential access to outpatient versus emergency care, the prevalence of non-white race groups or Hispanic ethnicity was significantly higher in the ED POCUS than in the TTE cohort; in summary, 1,053 studies (56.0%) were performed in women, 363 (19.3%) in Black and 198 (10.5%) in Hispanic individuals (p<0.001 for all relative to the TTE cohort*)*. There were 83 (4.4%) scans in patients with HCM, 56 with ATTR-CM (16.1%), and 302 (16.1%) with AS.

Development of a view-agnostic, multi-label model for HCM, ATTR-CM, AS

When tested in a leave-out set of the development (TTE) cohort, our multi-label, view-agnostic classifier successfully discriminated the presence of HCM (AUROC 0.95 [95% CI: 0.94-0.96]), ATTR-CM (0.98 [95% CI: 0.96-0.99]), and AS (AUROC 0.94 [95% CI: 0.93-0.95]). f(**Figure S1**). In a per-video analysis, the highest AUROC for HCM and ATTR-CM was seen in A4C videos (AUROC 0.90 [95% CI 0.89-0.92] and 0.98 [95% CI 0.97-0.99]), respectively) (**Figure S2**). Furthermore, in head-to-head video-level comparisons, the view-agnostic model consistently outperformed an ensemble of three view-specific models for PLAX, PSAX-PAP and A4C views (HCM (0.86 vs 0.81 with Δ[AUC] of 0.05 [95%CI: 0.03-0.07]); ATTR-CM: 0.95 vs 0.92 with Δ[AUC] of 0.03 [95%CI: 0.01-0.06]); and AS (0.84 vs 0.79 with Δ[AUC] of 0.05 [95%CI: 0.03-0.06]) (**Figure S3**).

Cardiac-focused POCUS assessment in the emergency room

As opposed to the TTE cohort where the ratio of PLAX:PSAX:A4C views was almost 1:1:1 (reflecting a standardized protocol) (**Table 1**), PSAX views were seen in the majority of all

(2,442 out of 3,758 [70.0%]) eligible cardiac POCUS videos, followed by the PLAX (836 [22.2%]) and A4C views (480 [12.8%]). Furthermore, the automated view classifier probabilities were significantly lower than in the TTE cohort (median 0.63 [IQR: 0.44-0.88] vs 0.93 [IQR: 0.69-1.00], p <0.001), reflecting lower anatomical correctness and view quality. Among POCUS videos, there was no significant difference in the view classifier probabilities between HCM, ATTR-CM, AS or controls ($p_{\text{Kruskal-Wallis}}$=0.07).

View-agnostic, multi-label AI models for opportunistic screening in the ED via POCUS

When deployed to the external POCUS cohort, our view-agnostic model consistently outperformed an ensemble of three view-specific models for PLAX, PSAX and A4C views for both HCM (AUROC 0.77 vs. 0.71 with Δ[AUROC] of 0.06 [95%CI: 0.02-0.10]), and ATTR-CM (AUROC 0.82 vs. 0.71 with Δ[AUROC] of 0.11 [95%CI: 0.06-0.16]) (**Figure 2**).

Discriminator performance increased further by incorporating a liberal quality control threshold that required input view probabilities to be 0.5 or greater (n=2496 from 3,758 videos) (**Figure 3**), reaching video-level performance consistent with that of standard TTE (**Figure S3**). For instance, at a single video level, our classifier identified ATTR-CM from PLAX views with an AUROC of 0.94, and HCM from A4C views with an AUROC of 0.85. At the threshold level that maximized Youden's J, these thresholds corresponded to a diagnostic odds ratio of 25.5 for HCM and 46.7 for ATTR-CM, respectively (see **Table 2** for threshold-specific summaries of sensitivity, specificity and PPV/NPV metrics).

Explainability and representative cases

Grad-CAM: Representative Grad-CAM maps for the top (true positive predictions) for each label are shown in **Figure 4**. For HCM these generally localized to the left ventricle, for ATTR-CM the signal further localized to the left atrium, whereas for the reference label of severe AS the focus seemed to be on the left ventricle and the aortic valve (when in plane), consistent with our prior work.[22]

Representative cases: In the **Supplement** (**Figures S4-S13**), we provide an illustrative summary of studies corresponding to the top five highest and lowest predictions for both cases and controls across all three key views (PLAX, PSAX, A4C). These examples showcase the key challenges of POCUS imaging, including the variation in acquisition protocols, probe orientation, off-axis views, and significant noise artifacts.

POCUS timing relative to diagnosis

Finally, to understand the relative proportion of patients who would have received a positive screen in the ED before their diagnosis, we estimated that among all patients with HCM (n=35) or ATTR-CM (n=57) who had at least one ED POCUS study in our system (before or after their diagnosis), 51.4% (n=18/35) and 41.1% (n=32/57), respectively, would have had a positive screen on POCUS in the 12 months leading up to their eventual confirmatory imaging (**Figure 5**).

DISCUSSION

In the present study, we demonstrate that AI algorithms trained using large datasets of standard transthoracic echocardiograms (TTE) can be adapted for use with point-of-care ultrasonography

(POCUS), thus enabling opportunistic screening of rare cardiomyopathies across a range of settings without dedicated protocols. Our approach describes a training framework that prompts video-based models to adjust to the unique challenges of handheld echocardiography, such as off-axis views and variable acquisition protocols, ultimately helping the models to learn generalizable signatures detectable across low and high-quality studies. Indeed, we show that view-agnostic models, which are exposed to the whole range of views available during training, consistently outperform view-specific models, a pattern that persists across both standard TTE and POCUS videos. The feasibility and superior performance of this approach are illustrated across more than 3,000 POCUS videos obtained in the emergency rooms of a hospital health system, with POCUS-adaptable CNNs effectively discriminating cases of HCM, ATTR-CM, and AS on single echocardiographic videos. Taken together, these findings provide a guide on how POCUS-adaptable training of AI echocardiographic algorithms can maximize their diagnostic value. Finally, we provide evidence that this approach may expand screening to minority populations receiving outpatient care at significantly lower rates than their counterparts, given their much higher representation in those undergoing POCUS than those undergoing formal echocardiography.

Despite significant progress in AI in cardiovascular imaging, most currently available AI solutions for echocardiography have been developed for standard transthoracic echocardiography.[24–26,29,46,47] Such models benefit, among others, from videos of high quality and anatomical fidelity that are obtained by certified echocardiography technicians, as well as standardized protocols with almost complete data capture that often include multiparametric phenotyping by Doppler and/or strain imaging. Unfortunately, disparities in accessing ambulatory outpatient care do exist and have repeatedly been shown to disproportionately affect

marginalized communities.[48–50] Indeed, as shown in our analysis, the representation of Black participants in the POCUS dataset was nearly 3 times higher than in transthoracic echocardiography. This highlights the untapped potential of diagnostic images already acquired as part of the current clinical workflows, particularly for conditions such as HCM and ATTR-CM that benefit from early detection and risk stratification, yet remain under-diagnosed,[6,7,11,15–17] particularly among under-represented communities.[18–20]

With this unmet need in mind, the present work describes innovative findings on both the methodological and clinical front. On the one hand, our work describes a scheme that can boost the performance of echocardiography-based AI tools when deployed to real-world POCUS studies by addressing POCUS-specific challenges as part of the training process. For instance, we show that natural augmentation of the training set through the naïve inclusion of multiple echocardiographic views and synthetic augmentation methods can help the model to learn a variety of on- and off-axis, high- and low-quality examples. By further introducing a quantitative scheme to assess view quality (or anatomical correctness), we confirm that POCUS videos have a significantly lower degree of anatomical correctness compared to standard TTE studies and provide a quantifiable metric that allows the training algorithm to place more emphasis on examples of low diagnostic quality.

Our work also has direct implications for the scalability of AI-echocardiography tools,[51,52] ensuring equitable access at the first point-of-care and eliminating disparities arising from differential access and referral to outpatient testing. We demonstrate that retrospective screening of large repositories could help detect rare cardiomyopathies even though the respective images and videos were not protocolled for this purpose. This highlights the opportunity to expand opportunistic screening and phenotyping beyond dedicated

echocardiographic studies and paves the path for prospective studies testing screening of high-risk individuals in the community.

Certain limitations merit consideration. First, our study was a retrospective analysis of a large hospital network. Thus, all echocardiograms were performed as part of a clinically indicated work-up, and, by design, POCUS studies were performed in individuals presenting to the emergency department with acute symptoms. Although this may limit our observations' generalizability, it also demonstrates the possibility of real-world AI-POCUS screening in our communities through existing care pathways. It should also be noted that the participants, ultrasound operators, and hardware used differed from the TTE derivation arm, thus enabling independent testing. Second, a drop in performance was noted when transitioning from TTE to POCUS studies. In reviewing representative images, we observed frequent off-axis views; for instance, the aortic valve was often poorly visualized on PLAX views. This is not surprising given that POCUS studies rarely focus on valvular function or LV thickness/remodeling and are most often limited to ruling out acute life-threatening conditions.

CONCLUSIONS

We propose and implement a framework that enables AI-based echocardiographic algorithms to generalize to point-of-care ultrasonography, thus enabling scalable opportunistic screening of under-diagnosed cardiomyopathies.

REFERENCES

1. Narula, J., Chandrashekhar, Y. & Braunwald, E. Time to Add a Fifth Pillar to Bedside Physical Examination: Inspection, Palpation, Percussion, Auscultation, and Insonation. *JAMA Cardiol* **3**, 346–350 (2018).
2. Spencer Kirk T. & Flachskampf Frank A. Focused Cardiac Ultrasonography. *JACC Cardiovasc. Imaging* **12**, 1243–1253 (2019).
3. Williams, J. P. *et al.* Current use, training, and barriers in point-of-care ultrasound in hospital medicine: A national survey of VA hospitals. *J. Hosp. Med.* **17**, 601–608 (2022).
4. Dessie, A. S. *et al.* Development and Validation of a Point-of-Care-Ultrasound Image Quality Assessment Tool: The POCUS IQ Scale. *J. Ultrasound Med.* **42**, 135–145 (2023).
5. Narang, A. *et al.* Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use. *JAMA Cardiol* **6**, 624–632 (2021).
6. Maron Barry J. *et al.* Diagnosis and Evaluation of Hypertrophic Cardiomyopathy. *J. Am. Coll. Cardiol.* **79**, 372–389 (2022).
7. Maron, M. S., Hellawell, J. L., Lucove, J. C., Farzaneh-Far, R. & Olivotto, I. Occurrence of Clinically Diagnosed Hypertrophic Cardiomyopathy in the United States. *Am. J. Cardiol.* **117**, 1651–1654 (2016).
8. Olivotto, I. *et al.* Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **396**, 759–769 (2020).
9. null null *et al.* 2023 ACC Expert Consensus Decision Pathway on Comprehensive Multidisciplinary Care for the Patient With Cardiac Amyloidosis. *J. Am. Coll. Cardiol.* **81**, 1076–1126 (2023).
10. Maurer, M. S. *et al.* Tafamidis Treatment for Patients with Transthyretin Amyloid Cardiomyopathy. *N. Engl. J. Med.* **379**, 1007–1016 (2018).

11. d'Arcy, J. L. *et al.* Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: the OxVALVE Population Cohort Study. *Eur. Heart J.* **37**, 3515–3522 (2016).
12. Mack, M. J. *et al.* Transcatheter Aortic-Valve Replacement with a Balloon-Expandable Valve in Low-Risk Patients. *N. Engl. J. Med.* **380**, 1695–1705 (2019).
13. Popma, J. J. *et al.* Transcatheter Aortic-Valve Replacement with a Self-Expanding Valve in Low-Risk Patients. *N. Engl. J. Med.* **380**, 1706–1715 (2019).
14. Forrest, J. K. *et al.* 2-Year Outcomes After Transcatheter Versus Surgical Aortic Valve Replacement in Low-Risk Patients. *J. Am. Coll. Cardiol.* **79**, 882–896 (2022).
15. Perego, S. *et al.* Prevalence, clinical correlates, and burden of undiagnosed aortic stenosis in older patients: a prospective study in a non-cardiologic acute hospital ward. *Aging Clin. Exp. Res.* **32**, 1533–1540 (2020).
16. Lindroos, M., Kupari, M., Heikkilä, J. & Tilvis, R. Prevalence of aortic valve abnormalities in the elderly: an echocardiographic study of a random population sample. *J. Am. Coll. Cardiol.* **21**, 1220–1225 (1993).
17. Ruberg, F. L., Grogan, M., Hanna, M., Kelly, J. W. & Maurer, M. S. Transthyretin Amyloid Cardiomyopathy: JACC State-of-the-Art Review. *J. Am. Coll. Cardiol.* **73**, 2872–2891 (2019).
18. Eberly, L. A. *et al.* Association of Race With Disease Expression and Clinical Outcomes Among Patients With Hypertrophic Cardiomyopathy. *JAMA Cardiol* **5**, 83–91 (2020).
19. Alexander, K. M. *et al.* Geographic Disparities in Reported US Amyloidosis Mortality From 1979 to 2015: Potential Underdetection of Cardiac Amyloidosis. *JAMA Cardiol* **3**, 865–870 (2018).
20. Crousillat, D. R. *et al.* Racial and Ethnic Differences in the Clinical Diagnosis of Aortic Stenosis. *J. Am. Heart Assoc.* **11**, e025692 (2022).
21. Akerman Ashley P. *et al.* Automated Echocardiographic Detection of Heart Failure With Preserved Ejection Fraction Using Artificial Intelligence. *JACC: Advances* **2**, 100452 (2023).

22. Holste, G. *et al.* Severe aortic stenosis detection by deep learning applied to echocardiography. *Eur. Heart J.* (2023) doi:10.1093/eurheartj/ehad456.

23. Holste, G., Oikonomou, E. K., Mortazavi, B., Wang, Z. & Khera, R. Self-supervised learning of echocardiogram videos enables data-efficient clinical diagnosis. *arXiv [cs.CV]* (2022).

24. Duffy, G. *et al.* High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy With Cardiovascular Deep Learning. *JAMA Cardiol* **7**, 386–395 (2022).

25. Goto, S. *et al.* Multinational Federated Learning Approach to Train ECG and Echocardiogram Models for Hypertrophic Cardiomyopathy Detection. *Circulation* **146**, 755–769 (2022).

26. Goto, S. *et al.* Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. *Nat. Commun.* **12**, 2726 (2021).

27. Krishna, H. *et al.* Fully Automated Artificial Intelligence Assessment of Aortic Stenosis by Echocardiography. *J. Am. Soc. Echocardiogr.* **36**, 769–777 (2023).

28. He, B. *et al.* Blinded, randomized trial of sonographer versus AI cardiac function assessment. *Nature* **616**, 520–524 (2023).

29. Ouyang, D. *et al.* Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252–256 (2020).

30. Stausberg, J., Lehmann, N., Kaczmarek, D. & Stein, M. Reliability of diagnoses coding with ICD-10. *Int. J. Med. Inform.* **77**, 50–57 (2008).

31. Arbelo, E. *et al.* 2023 ESC Guidelines for the management of cardiomyopathies. *Eur. Heart J.* **44**, 3503–3626 (2023).

32. Dorbala, S. *et al.* ASNC/AHA/ASE/EANM/HFSA/ISA/SCMR/SNMMI Expert Consensus Recommendations for Multimodality Imaging in Cardiac Amyloidosis: Part 1 of 2-Evidence Base and Standardized Methods of Imaging. *Circ. Cardiovasc. Imaging* **14**, e000029 (2021).

33. Ladefoged, B. *et al.* Diagnostic delay in wild type transthyretin cardiac amyloidosis - A clinical challenge. *Int. J. Cardiol.* **304**, 138–143 (2020).

34. Baumgartner, H., Chair *et al.* Recommendations on the echocardiographic assessment of aortic valve stenosis: a focused update from the European Association of Cardiovascular Imaging and the American Society of Echocardiography. *Eur. Heart J. Cardiovasc. Imaging* **18**, 254–275 (2017).

35. Baumgartner, H. *et al.* 2017 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur. Heart J.* **38**, 2739–2791 (2017).

36. Oikonomou, E. K. *et al.* A digital biomarker for aortic stenosis development and progression using deep learning for two-dimensional echocardiography. *medRxiv* (2023) doi:10.1101/2023.09.28.23296234.

37. Zhang, J. *et al.* Fully Automated Echocardiogram Interpretation in Clinical Practice. *Circulation* **138**, 1623–1635 (2018).

38. Paszke, Gross, Massa & Lerer. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* (2019).

39. Pedregosa, Varoquaux & Gramfort. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* (2011).

40. Müller, Kornblith & Hinton. When does label smoothing help? *Adv. Neural Inf. Process. Syst.* (2019).

41. Szegedy, Vanhoucke & Ioffe. Rethinking the inception architecture for computer vision. *Proc. Estonian Acad. Sci. Biol. Ecol.* (2016).

42. González-López, E. *et al.* Wild-type transthyretin amyloidosis as a cause of heart failure with preserved ejection fraction. *Eur. Heart J.* **36**, 2585–2594 (2015).

43. Castaño, A. *et al.* Unveiling transthyretin cardiac amyloidosis and its predictors among elderly patients with severe aortic stenosis undergoing transcatheter aortic valve replacement. *Eur. Heart J.* **38**, 2879–2887 (2017).

44. Obi, C. A., Mostertz, W. C., Griffin, J. M. & Judge, D. P. ATTR Epidemiology, Genetics, and Prognostic Factors. *Methodist Debakey Cardiovasc. J.* **18**, 17–26 (2022).

45. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *arXiv [cs.CV]* (2016).

46. Madani, A., Arnaout, R., Mofrad, M. & Arnaout, R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med* **1**, (2018).

47. Arnaout, R. *et al.* An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nat. Med.* **27**, 882–891 (2021).

48. Hyland, P. M. *et al.* Race, sex and age disparities in echocardiography among Medicare beneficiaries in an integrated healthcare system. *Heart* **108**, 956–963 (2022).

49. Ahmed, Y. *et al.* Racial and ethnic disparities in diagnosis, management and outcomes of aortic stenosis in the Medicare population. *PLoS One* **18**, e0281811 (2023).

50. Spencer-Bonilla, G. *et al.* Racial and ethnic disparities in transthyretin cardiac amyloidosis. *Curr. Cardiovasc. Risk Rep.* **15**, (2021).

51. Hswen, Y. & Voelker, R. New AI Tools Must Have Health Equity in Their DNA. *JAMA* **330**, 1604–1607 (2023).

52. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. & Chin, M. H. Ensuring Fairness in Machine Learning to Advance Health Equity. *Ann. Intern. Med.* **169**, 866–872 (2018).

DISCLOSURES

R.K. is an Associate Editor of JAMA and receives research support, through Yale, from the Blavatnik Foundation, Bristol-Myers Squibb, Novo Nordisk, and BridgeBio. He is a coinventor of U.S. Provisional Patent Applications 63/177,117, 63/428,569, 63/346,610, 63/484,426, and 63/508,315, and a co-founder of Ensight-AI, Inc. R.K. and E.K.O. are co-founders of Evidence2Health, a health analytics company. E.K.O. is a co-inventor in patent applications (US17/720,068, 63/619,241, 63/177,117, 63/580,137, 63/606,203, 63/619,241, WO2018078395A1, WO2020058713A1), has been an ad hoc consultant for Caristo Diagnostics,

Ltd, and has received royalty fees from technology licensed through the University of Oxford. H.M.K. 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, and Ensight-AI, Inc. All other authors declare no competing interests.

FUNDING

This study was supported by grants 1F32HL170592-01 (EKO) and K23HL153775 (RK) from the National Heart, Lung, and Blood Institute of the National Institutes of Health and award 2022060 (RK) from the Doris Duke Charitable Foundation. 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.

TABLES

Table 1 | Summary of cohort characteristics.

	Development (TTE) cohort			ED POCUS testing
	Training	Validation	TTE testing	
Available videos and studies				
Number of unique patients	6,376	1,245	839	1,330
Number of unique studies	8,090	1,577	1,035	1,879
Number of unique videos	218,725	43,031	28,489	3,758
PLAX	25,569	4,987	3,283	836
PSAX	26,665	5,114	3,313	2,442
A4C	21,328	4,200	2,649	480
Demographic characteristics (study-level)				
Age, mean (SD)	68.3 (15.0)	68.7 (14.8)	69.2 (15.3)	70.3 (16.7)
Female gender, n (%)	3,752 (46.4)	709 (45.0)	496 (47.9)	1,053 (56.0)
Hispanic Ethnicity, n (%)	433 (5.4)	80 (5.1)	51 (4.9)	198 (10.5)
Race, n (%)				
African American	612 (7.6)	98 (6.2)	73 (7.1)	363 (19.3)
Asian	109 (1.3)	23 (1.5)	12 (1.2)	20 (1.1)
White	6,087 (75.2)	1,231 (78.1)	813 (78.6)	1,247 (66.4)
Other or Unknown	1,282 (15.8)	225 (14.3)	137 (13.2)	249 (13.3)
Cardiomyopathy diagnosis (study-level)				
HCM, n (%)	1,375 (17.0)	243 (15.4)	168 (16.2)	83 (4.4)
ATTR-CM, n (%)	228 (2.8)	39 (2.5)	31 (3.0)	56 (3.0)
Severe AS, n (%)	1,120 (15.0)	238 (16.0)	152 (15.5)	302 (16.1)

Variables summarized as mean (standard deviation [SD]) or counts (percentages), as appropriate. A4C: Apical 4-chamber; AS: aortic stenosis; HCM: hypertrophic cardiomyopathy; PLAX: parasternal long axis view; PSAX: parasternal short axis; SD: standard deviation; TTE: transthoracic echocardiography.

Table 2 | POCUS-based discrimination of HCM, ATTR-CM and AS.

Label	View	Threshold Label	AUROC	Threshold	AUPRC	F-score	Specificity	Sensitivity	PPV	NPV	Diagnostic OR
HCM	A4C	Youden's J	0.849	0.148	0.440	0.274	0.561	0.952	0.063	0.997	25.52
		90% Sensitivity		0.190		0.288	0.640	0.857	0.069	0.993	10.67
	PLAX	Youden's J	0.760	0.167	0.206	0.158	0.595	0.765	0.055	0.988	4.77
		90% Sensitivity		0.123		0.143	0.505	0.824	0.049	0.989	4.75
	PSAX	Youden's J	0.779	0.208	0.253	0.161	0.706	0.688	0.067	0.986	5.29
		90% Sensitivity		0.121		0.127	0.490	0.875	0.050	0.992	6.73
	All 3	Youden's J	0.820	0.180	0.273	0.178	0.646	0.818	0.067	0.991	8.23
		90% Sensitivity		0.148		0.158	0.557	0.879	0.058	0.993	9.12
ATTR-CM	A4C	Youden's J	0.842	0.308	0.162	0.200	0.850	0.714	0.128	0.990	14.14
		90% Sensitivity		0.308		0.200	0.850	0.714	0.128	0.990	14.14
	PLAX	Youden's J	0.936	0.462	0.548	0.271	0.745	0.941	0.102	0.998	46.68
		90% Sensitivity		0.518		0.275	0.787	0.824	0.107	0.993	17.22
	PSAX	Youden's J	0.879	0.497	0.354	0.189	0.777	0.811	0.101	0.993	14.96
		90% Sensitivity		0.401		0.166	0.719	0.865	0.087	0.994	16.35
	All 3	Youden's J	0.908	0.383	0.354	0.171	0.719	0.911	0.091	0.996	26.20
		90% Sensitivity		0.396		0.175	0.735	0.889	0.094	0.995	22.21
AS	A4C	Youden's J	0.617	0.442	0.273	0.371	0.490	0.692	0.040	0.981	2.17
		90% Sensitivity		0.290		0.376	0.293	0.885	0.037	0.988	3.18
	PLAX	Youden's J	0.803	0.598	0.428	0.456	0.755	0.750	0.086	0.990	9.24
		90% Sensitivity		0.425		0.359	0.523	0.875	0.054	0.993	7.68
	PSAX	Youden's J	0.580	0.484	0.182	0.296	0.478	0.690	0.039	0.980	2.03
		90% Sensitivity		0.334		0.270	0.166	0.897	0.032	0.981	1.72
	All 3	Youden's J	0.633	0.512	0.251	0.334	0.579	0.645	0.045	0.981	2.49
		90% Sensitivity		0.314		0.289	0.185	0.892	0.033	0.982	1.87

A4C: apical 4-chamber; AS: aortic stenosis; ATTR-CM: transthyretin amyloid cardiomyopathy; HCM: hypertrophic cardiomyopathy; OR: odds ratio; PLAX: parasternal long axis; POCUS: point-of-care ultrasonography; PSAX: parasternal short axis (papillary muscle level). * PPV and NPV reported at simulated 3% prevalence.

FIGURES

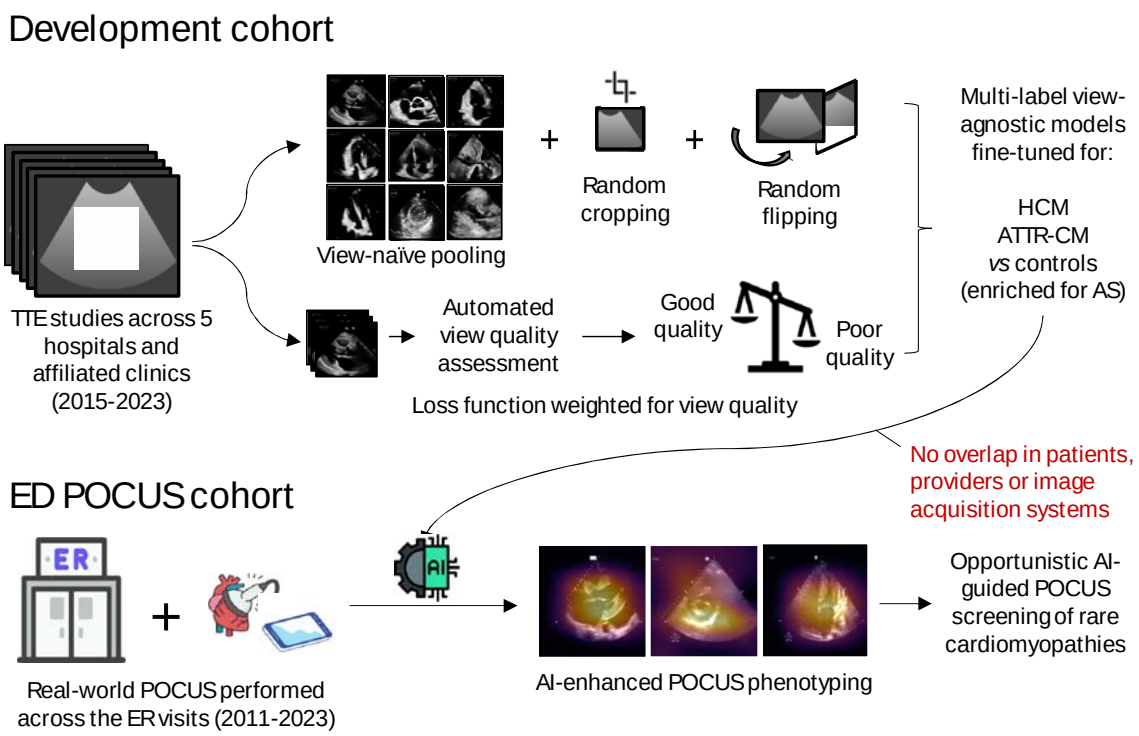


Figure 1 | Study overview. Overview of study design and datasets. AI: artificial intelligence; AS: (severe) aortic stenosis; ATTR-CM: amyloid transthyretin cardiomyopathy; ER: emergency room; HCM: hypertrophic cardiomyopathy; POCUS: point-of-care ultrasound; TTE: (standard) transthoracic echocardiography.

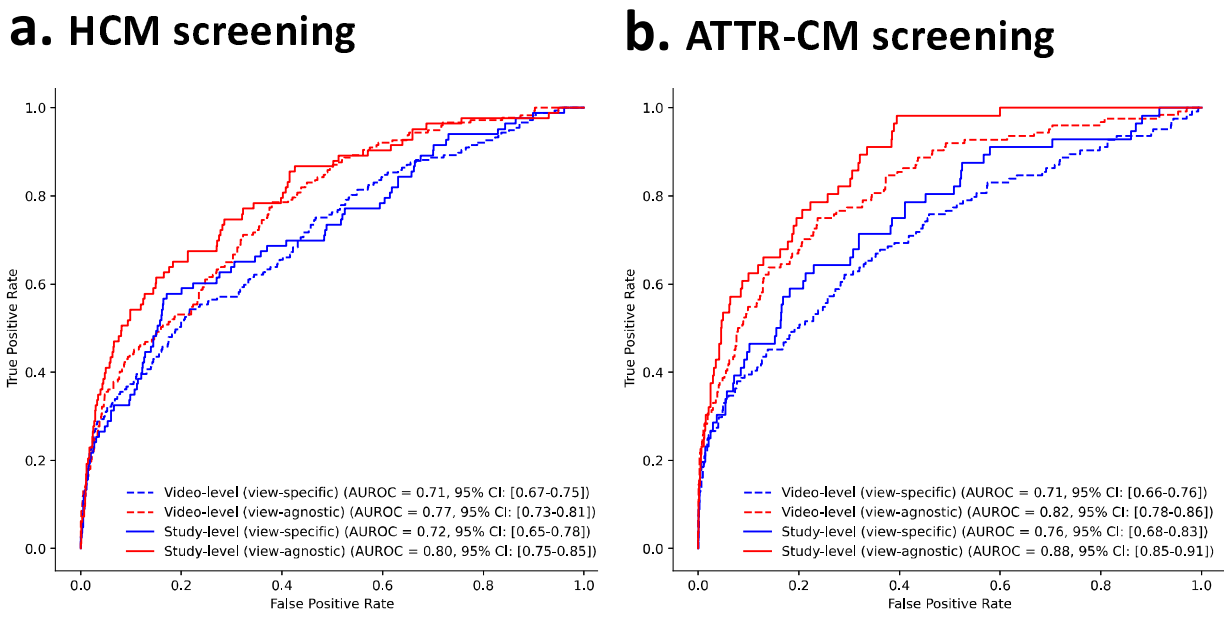


Figure 2 | Head-to-head comparison of view-agnostic vs view-specific multi-label models for discriminating HCM and ATTR-CM at the point-of-care. Head-to-head comparison of view-agnostic (red line) versus an ensemble of view-specific models (blue line) across included parasternal long, parasternal short and apical four-chamber views for discrimination of **(a)** HCM, or **(b)** ATTR-CM using POCUS videos. Continuous lines denote study-level estimates, where dotted lines reflect video-level estimate. ATTR-CM: amyloid transthyretin cardiomyopathy; AUROC: area under the receiver operating characteristic curve; CI: confidence interval; HCM: hypertrophic cardiomyopathy; POCUS: point-of-care ultrasonography.

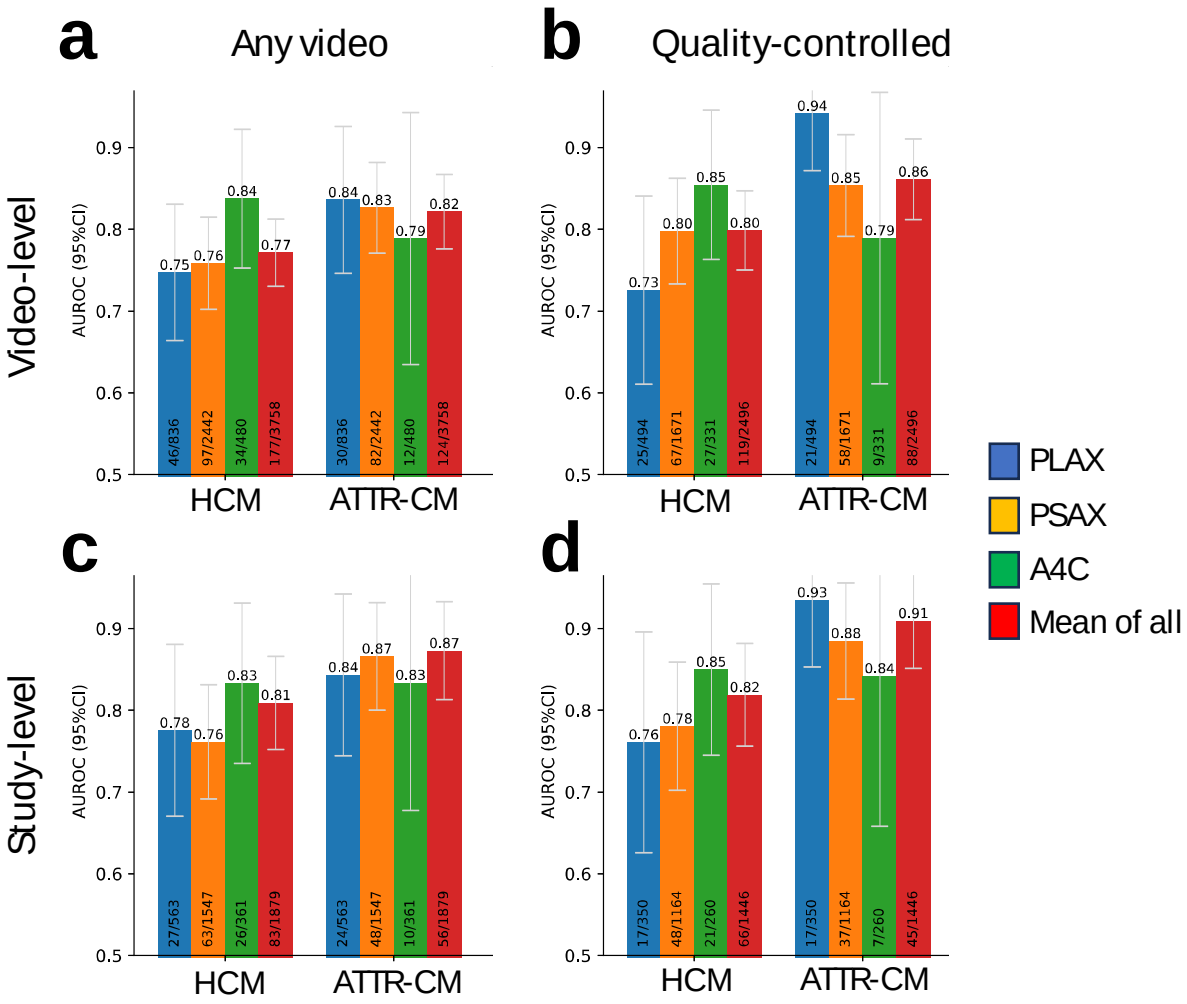
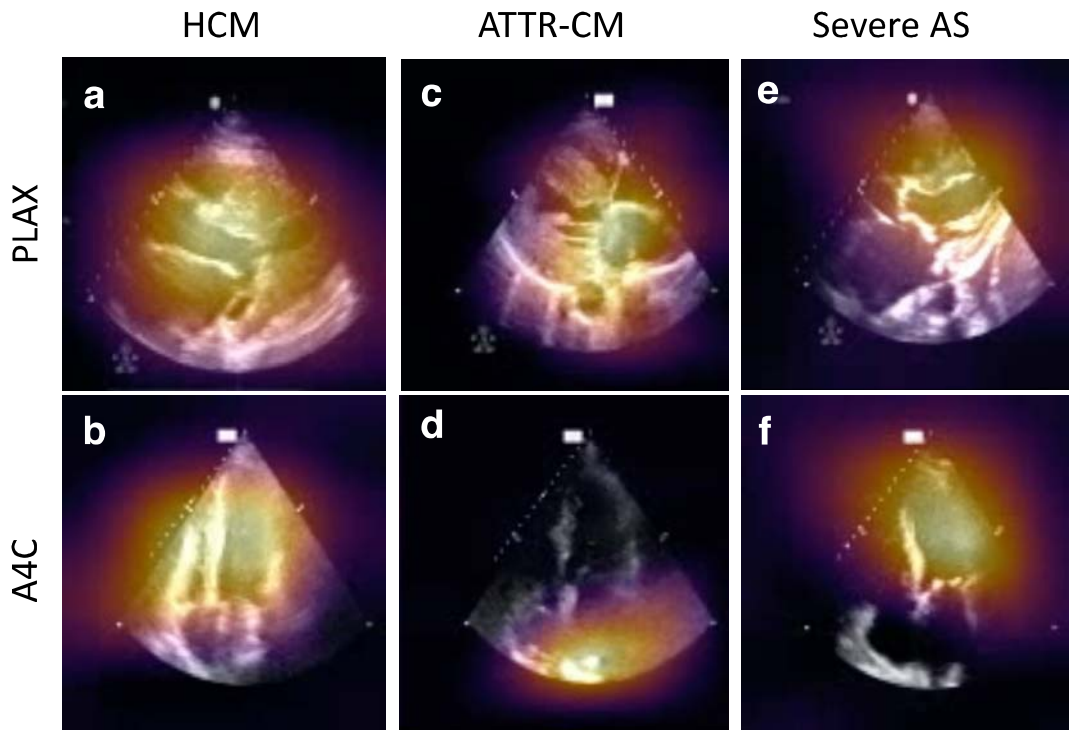


Figure 3 | Video and study-level performance of a view-agnostic multi-label deep learning algorithm applied to POCUS. (a) Video-level performance (AUROC with 95% CI) for discrimination of HCM and ATTR-CM, by deploying a POCUS-adapted, view-agnostic model to different echocardiographic views obtained in the ED (blue = PLAX; orange = PSAX at the papillary muscle level; green = A4C; and red = any of the 3 views combined). **(b)** Results presented for videos with an automated view classifier certainty of 50% of greater. **(c-d)** Study-level results obtained by simple mean averaging of the video-level predictions for each study. A4C: apical-4-chamber view; ATTR-CM: amyloid transthyretin cardiomyopathy; AUROC: area under the receiver operating characteristic curve; CI: confidence interval; HCM: hypertrophic cardiomyopathy; PLAX: parasternal long axis view; POCUS: point-of-care ultrasonography; PSAX: parasternal short axis view.



Figu
PLA
cham
HCM

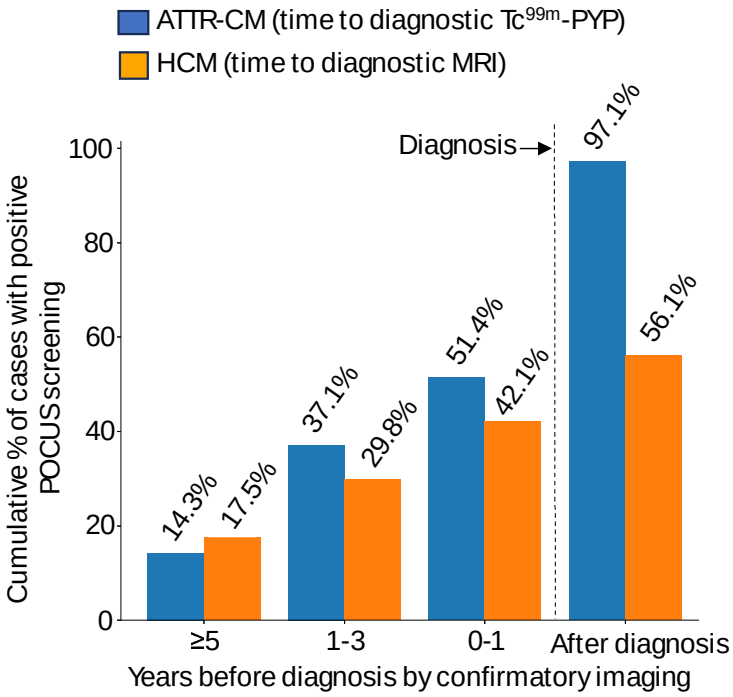


Figure 5 | Positive screening by POCUS relative to eventual confirmatory testing in HCM and ATTR-CM among cases with ED POCUS. Cumulative percentages of patients eventually diagnosed with HCM or ATTR-CM who had positive screening by AI applied to their POCUS studies before their eventual confirmatory imaging test. Positive screens are defined based on probabilities exceeding the value that maximizes Youden's J statistic for each label. ATTR-CM: amyloid transthyretin cardiomyopathy; HCM: hypertrophic cardiomyopathy; POCUS: point-of-care ultrasonography.