JAMA | Original Investigation | AI IN MEDICINE

Complete AI-Enabled Echocardiography Interpretation With Multitask Deep Learning

Gregory Holste, MSE; Evangelos K. Oikonomou, MD, DPhil; Márton Tokodi, MD, PhD; Attila Kovács, MD, PhD; Zhangyang Wang, PhD; Rohan Khera, MD, MS

IMPORTANCE Echocardiography is a cornerstone of cardiovascular care, but relies on expert interpretation and manual reporting from a series of videos. An artificial intelligence (AI) system, PanEcho, has been proposed to automate echocardiogram interpretation with multitask deep learning.

OBJECTIVE To develop and evaluate the accuracy of an AI system on a comprehensive set of 39 labels and measurements on transthoracic echocardiography (TTE).

DESIGN, SETTING, AND PARTICIPANTS This study represents the development and retrospective, multisite validation of an AI system. PanEcho was developed using TTE studies conducted at Yale New Haven Health System (YNHHS) hospitals and clinics from January 2016 to June 2022 during routine care. The model was internally validated in a temporally distinct YNHHS cohort from July to December 2022, externally validated across 4 diverse external cohorts, and publicly released.

MAIN OUTCOMES AND MEASURES The primary outcome was the area under the receiver operating characteristic curve (AUC) for diagnostic classification tasks and mean absolute error for parameter estimation tasks, comparing AI predictions with the assessment of the interpreting cardiologist.

RESULTS This study included 1.2 million echocardiographic videos from 32 265 TTE studies of 24 405 patients across YNHHS hospitals and clinics. The AI system performed 18 diagnostic classification tasks with a median (IQR) AUC of 0.91 (0.88-0.93) and estimated 21 echocardiographic parameters with a median (IQR) normalized mean absolute error of 0.13 (0.10-0.18) in internal validation. For instance, the model accurately estimated left ventricular ejection fraction (mean absolute error: 4.2% internal; 4.5% external) and detected moderate or worse left ventricular systolic dysfunction (AUC: 0.98 internal; 0.99 external), right ventricular systolic dysfunction (AUC: 0.93 internal; 0.94 external), and severe aortic stenosis (AUC: 0.98 internal; 1.00 external). The AI system maintained excellent performance in limited imaging protocols, performing 15 diagnosis tasks with a median (IQR) AUC of 0.91 (0.87-0.94) in an abbreviated TTE cohort and 14 tasks with a median (IQR) AUC of 0.85 (0.77-0.87) on real-world point-of-care ultrasonography acquisitions from YNHHS emergency departments.

CONCLUSIONS AND RELEVANCE In this study, an AI system that automatically interprets echocardiograms maintained high accuracy across geography and time from complete and limited studies. This AI system may be used as an adjunct reader in echocardiography laboratories or AI-enabled screening tool in point-of-care settings following prospective evaluation in the respective clinical workflows.

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Rohan Khera, MD, MS, Yale School of Medicine, 195 Church St, 6th Floor, New Haven, CT 06510 (rohan.khera@yale.edu).

JAMA. 2025;334(4):306-318. doi:10.1001/jama.2025.8731
Published online June 23, 2025.

Echocardiography is a pillar of cardiovascular diagnostics, providing in-depth phenotyping of cardiac, valvular, and vascular structure and function.[1] More than 7.5 million echocardiographic studies are performed annually in the US alone, and increasing referrals contribute to rising health care expenditures across most nations.[2,3] Accurate reporting of echocardiography requires time, skilled acquisition, and expert readers, but remains limited by the availability of sufficient experts and substantial interrater variability.[4,5] Reliance on scarcely available expert interpretation poses a barrier to access to this important diagnostic modality, especially given the increasing accessibility of handheld ultrasonography devices at the point of care.[6,7]

Artificial intelligence (AI) algorithms have shown promise in automating various aspects of echocardiography reporting, from detecting valvular abnormalities[8-11] to quantifying key measurements such as left ventricular ejection fraction (LVEF)[12-17] and others.[18-23] However, existing solutions typically rely on single-view inputs and are limited to single tasks.[8,11,13,20,22-26] This process differs from clinical practice, in which multiple views and imaging modes, such as color Doppler imaging, are integrated to form a comprehensive evaluation, spanning functional and structural metrics of all major chambers, valves, and vessels. Versatile AI systems that handle this multiview, multitask workflow accurately and robustly would enable efficient, reader-independent phenotyping of echocardiography, but are currently lacking.

To bridge this gap and provide a scalable solution for fully automated echocardiographic interpretation, an end-to-end, view-agnostic deep learning model capable of simultaneously performing the full range of key echocardiographic reporting tasks was developed and validated. This study presents the development of PanEcho, a novel AI model that uses multitask deep learning on more than 1 million standard 2-dimensional (2D) B-mode and color Doppler echocardiogram videos to perform 39 diverse interpretation tasks and report its internal and external validation across a range of acquisition protocols. In addition to this study report, the model and source code have been publicly released to accelerate research on AI-enabled echocardiographic interpretation.

Key Points

Question Can artificial intelligence (AI) fully automate echocardiogram interpretation?

Findings This study reports the development and validation of an automated AI system for echocardiogram analysis, PanEcho, that performed 18 diagnostic classification tasks with a median area under the receiver operating characteristic curve of 0.91 and 21 echocardiographic parameter estimation tasks with a median normalized mean absolute error of 0.13.

Meaning An AI system can automate complete echocardiogram interpretation with high accuracy, potentially accelerating workflows and enabling rapid cardiovascular health screening in point-of-care settings with limited access to trained experts.

Methods

Data Source and Patient Population

This study was approved by the Yale University and local institutional review boards, and the need for informed consent was waived because this research represents secondary analysis of existing data. This study adhered to all applicable elements of the TRIPOD+AI reporting guideline.

Model Development and Internal Validation

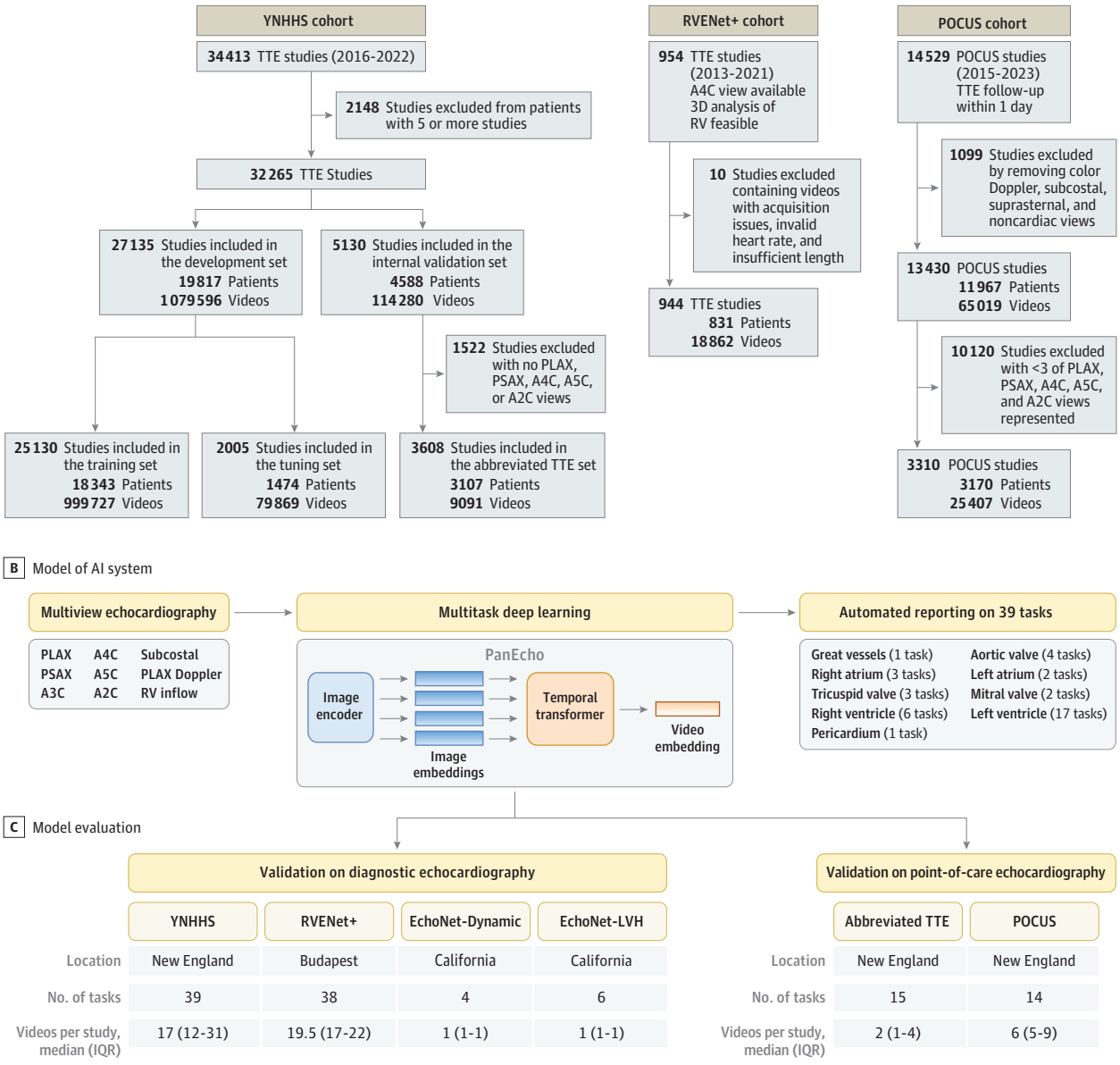
Data for model development and internal validation were derived from transthoracic echocardiography (TTE) studies performed at Yale New Haven Health System (YNHHS) hospitals from 2016 to 2022 during routine clinical care. We used a previously published approach[8] to extract and deidentify echocardiographic videos by masking out pixels containing protected health information. This study included 2D grayscale and color Doppler videos from all echocardiographic views. The YNHHS cohort was split into development and validation sets, with studies from July 2022 to December 2022 set aside as a temporally distinct internal validation set. The remaining studies from January 2016 to June 2022 were used for model development after removing studies from all validation set patients to prevent data leakage. The development set was randomly partitioned into training (92.5%) and tuning (7.5%) sets at the patient level for model training. See the eMethods in Supplement 1 for full details on YNHHS dataset curation and preprocessing.

External Validation

This study featured 4 cohorts for external validation: RVENet+, POCUS (point-of-care ultrasonography), EchoNet-Dynamic, and EchoNet-LVH, representing a broad range of phenotypes with standard and point-of-care echocardiographic acquisitions across time and geography. The RVENet+ cohort consisted of complete echocardiographic studies conducted at the Heart and Vascular Center of Semmelweis University in Budapest, Hungary, from 2013 to 2021 and included all available 2D views. As previously described,[27,28] this cohort comprised 7 distinct subpopulations including elite athletes and heart transplant recipients, representing a diverse population. This cohort underwent the same deidentification and preprocessing steps as the YNHHS cohort.

The POCUS cohort consisted of consecutive patients who underwent cardiac-focused POCUS imaging as part of care in YNHHS-affiliated emergency departments (EDs) from 2016 to 2023. In contrast to all other cohorts that included studies performed by echocardiography technicians, these abbreviated scans were performed by ED clinicians. We present analysis in a subset of patients with POCUS imaging and a temporally linked complete echocardiographic examination performed no more than 1 day apart, essential to provide reliable ground truth labels. This cohort underwent the same deidentification and preprocessing steps as the YNHHS and RVENet+ cohorts, ensuring no patient overlap with the YNHHS development cohort. Additional quality control was needed to ensure minimal diagnostic quality in these noisy acquisitions by nonexperts, removing noncardiac views and filtering for studies with at least 3 of these 5 key views: parasternal long axis (PLAX), parasternal short axis at the papillary level, apical 4-chamber (A4C),

Figure 1. Artificial Intelligence (AI) System Model Development and Study Design



PanEcho is an AI model that analyzes all 2-dimensional views acquired during a transthoracic echocardiogram and automatically performs 39 key echocardiographic interpretation tasks (B). The model is externally validated on complete diagnostic-quality echocardiograms as well as limited acquisitions acquired at the point of care (C).

3D indicates 3-dimensional; A2C, apical 2-chamber; A4C, apical 4-chamber; A5C, apical 5-chamber; AV, aortic valve; LVH, left ventricle hypertrophy; PLAX, parasternal long axis; POCUS, point-of-care ultrasonography; PSAX, parasternal short axis; TTE, transthoracic echocardiography; YNHHS, Yale New Haven Health System.

apical 2-chamber (A2C), and apical 5-chamber (A5C). Full POCUS cohort curation details are described in previous work[29] and the eMethods in Supplement 1. See **Figure 1**A for an exclusion flowchart of all key internal and external cohorts.

EchoNet-Dynamic[13] and EchoNet-LVH[24] are publicly available datasets for assessment of LV function and structure, respectively, that were included as additional external validation cohorts in this study. EchoNet-Dynamic consists of 10 030 individual A4C videos from unique echocardiograms performed at Stanford University Hospital from 2016 to 2018 with measurements of LV function and volumes; EchoNet-LVH consists of 12 000 PLAX videos from echocardiograms performed at Stanford Health Care from 2008 to 2020 with measurements of LV dimensions.

Echocardiographic Reporting Labels

For each study in the YNHHS, RVENet+, and POCUS cohorts, the team of investigators jointly reviewed and defined a list

of 39 key reporting tasks, representing a variety of categorical findings and continuous measurements covering all major axes of echocardiographic reporting. This included 18 classification tasks, such as capturing the size, structure, and function of all 4 heart chambers and valvular disease, and 21 regression tasks quantifying key dimensions of each chamber, blood flow velocities, and more. Labels were extracted from the local electronic echocardiography reporting systems and reflected the final measurements and reporting confirmed by a certified echocardiographer. All interpretations were performed in line with standard guidelines of the American Society of Echocardiography[1] and to the discretion of the echocardiographer; descriptions of all diagnostic labels are detailed in eTable 1 in Supplement 1. We also extracted a total of 10 labels from EchoNet-Dynamic and EchoNet-LVH based on the LV measurements provided, such as LVEF and LV interventricular septum thickness. Because categorical labels were not explicitly provided, we determined these via thresholds described in the eMethods in Supplement 1.

Model Development

We designed a multitask, video-based deep learning model with a 2D image encoder, temporal transformer, and output heads dedicated to individual tasks (Figure 1B). In its design, the AI system mimics a human reader, integrating all available echocardiographic views and assessing all 39 echocardiographic reporting tasks described above. Each video frame is first processed by the 2D image encoder, a convolutional neural network that produces a learned embedding of each frame. These framewise embeddings are then interpreted as an ordered sequence and modeled using self attention[30] to learn time-varying associations over the frames. Finally, a video-level embedding is formed and used as input to the task-specific output heads. Rather than develop separate view- and task-specific models, this design efficiently shares computation across views with a unified view-agnostic encoder and parallelizes task-specific modeling with lightweight output heads specialized for each task. The model was trained end to end to minimize the mean of all task-specific losses. See the eMethods in Supplement 1 for full implementation details.

Multitask Evaluation on Diagnostic Echocardiography

The model was first evaluated across all 39 labels in the internal YNHHS validation set as well as the 3 external cohorts RVENet+ (38 labels), EchoNet-Dynamic (4 labels), and EchoNet-LVH (6 labels) with diagnostic-quality videos from complete echocardiographic studies (Figure 1C). Because interpretations are unique to each echocardiographic study, evaluation was performed at the study level using all available videos and tasks. For multiview datasets like the YNHHS and RVENet+ cohorts, the AI system automatically aggregated its predictions across all videos acquired during a study at inference time to form study-level interpretations. In contrast, EchoNet-Dynamic and EchoNet-LVH feature 1 video per study, evaluating the model's ability to provide interpretations from a single view. Beyond predictive performance, the model was also evaluated along dimensions of interpretability, fairness, and robustness to image quality in the internal validation set. Details on these auxiliary analyses can be found in the eMethods in Supplement 1.

Multitask Evaluation on Point-of-Care Echocardiography

To illustrate the versatility of our model across imaging protocols, we also evaluated its performance in both a simulated abbreviated TTE protocol and a real-world point-of-care setting. To simulate an abbreviated acquisition, inference was performed on the internal YNHHS validation set, but the model was only given access to a single video from each of the following key views per study if available: PLAX, parasternal short axis at the papillary level, A4C, A5C, and A2C. Validation was also performed across actual point-of-care acquisitions from YNHHS EDs in the POCUS cohort described above. Because POCUS imaging is focused on ruling out key abnormalities and lacks the protocolized approach for reliable estimation of linear or volumetric measurements, evaluation was restricted to all 15 classification tasks evaluable on grayscale 2D echocardiography; this excluded estimation tasks and diagnostic findings, such as valvular regurgitation, that require Doppler imaging for proper assessment. See the eMethods in Supplement 1 for full details on point-of-care validation.

Statistical Analysis

Categorical values are summarized as counts and percentages, whereas continuous variables are summarized as median (IQR) to account for skewness. Classification tasks were primarily evaluated by area under the receiver operating characteristic curve (AUC) in addition to sensitivity, specificity, average precision, and Brier score; class-specific thresholds were determined by maximizing Youden index on the tuning set. Regression tasks were primarily evaluated by mean absolute error (MAE) in addition to Pearson correlation coefficient (r), median absolute deviation, root mean-squared error, R^2, and normalized MAE (divided by the mean ground truth measurement). For multiclass classification tasks, we present AUC results on the most severe class in the main text to simplify the presentation. When summarizing performance across regression tasks with different units, we report median normalized MAE to account for variations in scale across measurements. We computed 95% CIs for all metrics with 1000 bootstrap samples at the study level using the percentile method. Statistical analysis was performed using Python version 3.10.8 (Python Foundation).

Results

Study Cohorts and Model Development

The YNHHS cohort included 1 193 876 echocardiographic videos comprising all 2D views from 32 265 transthoracic echocardiography studies of 24 405 unique patients across YNHHS hospitals (median [IQR] age, 69.0 [58.0-79.0] years; 16 819 [52.1%] men). The model was developed using 999 727 videos from 25 130 studies of 18 343 YNHHS patients from January 2016 to June 2022, while 5130 studies from 4588 distinct patients from July to December 2022 were used as a temporally distinct internal validation set. This set featured a broad range of echocardiographic phenotypes with a median (IQR) LVEF of 61.4%

(57.0%-65.0%) and 284 studies (5.8%) including individuals with moderate or worse LV systolic dysfunction, 86 (1.8%) with severe aortic stenosis, and 612 (12.2%) with moderate mitral regurgitation; see additional details in the **Table**.

The external RVENet+ cohort included 18 862 echocardiographic videos from 944 complete echocardiographic studies of 831 patients from the Heart and Vascular Center of Semmelweis University in Budapest, Hungary, from November 2013 to March 2021 (median [IQR] age, 50.0 [25.0-66.0] years; 586 [62.1%] male). This cohort featured a unique composition of 7 subpopulations, with median (IQR) LVEF of 57.5% (52.2%-61.4%) and 122 (12.9%) studies with moderate or worse LV systolic dysfunction, 85 (9.0%) with severe aortic stenosis, and 141 (17.0%) with moderate or worse mitral regurgitation. The POCUS cohort consisted of 25 407 videos from 3310 studies of 3170 patients undergoing cardiac-focused point-of-care echocardiography in YNHHS-affiliated EDs from September 2015 to November 2023 (median [IQR] age: 66.0 [53.0-79.0] years, 1811 [54.8%] male) (Table). Full description and statistics of all cohorts can be found in eTable 2 in Supplement 1.

Validation on Diagnostic Echocardiography

The AI system performed 18 diagnostic classification tasks with a median (IQR) AUC of 0.91 (0.88-0.93) in internal YNHHS validation and 17 diagnostic tasks with a median (IQR) AUC of 0.91 (0.85-0.94) in international RVENet+ validation (**Figure 2**; eTable 3 in Supplement 1). The model accurately assessed ventricular structure and function, with internal AUCs of 0.94 (95% CI, 0.93-0.95) (external AUC, 0.98 [95% CI, 0.98-0.99]) for moderate or higher increased LV size, 0.98 (95% CI, 0.98-0.99) (external AUC, 0.99 [95% CI, 0.98-0.99]) for moderate or worse LV systolic dysfunction, 0.92 (95% CI, 0.91-0.93) (external AUC, 0.92 [95% CI, 0.90-0.93]) for moderate or worse LV diastolic dysfunction, 0.88 (95% CI, 0.86-0.90) (external AUC, 0.98 [95% CI, 0.97-0.99]) for LV wall motion abnormalities, and 0.94 (95% CI, 0.91-0.94) (external AUC, 0.94 [95% CI, 0.92-0.96]) for right ventricular systolic dysfunction. The model also achieved excellent discrimination of valvular disease, with internal AUCs of 0.98 (95% CI, 0.98-0.99) (external AUC, 1.00 [95% CI, 0.99-1.00]) for severe aortic stenosis, 0.96 (95% CI, 0.94-0.98) (external AUC, 1.00 [95% CI, 1.00-1.00]) for mitral stenosis, 0.91 (95% CI, 0.90-0.91) (external AUC, 0.92 [95% CI, 0.91-0.94]) for moderate or worse mitral regurgitation, and 0.90 (95% CI, 0.89-0.91) (external AUC, 0.91 [95% CI, 0.90-0.93]) for moderate or worse tricuspid regurgitation. Additional phenotypes of pericardial effusion and LV outflow tract obstruction were classified with internal AUCs of 0.91 (95% CI, 0.83-0.97) (external AUC, 0.94 [95% CI, 0.89-0.98]) and 0.94 (95% CI, 0.88-0.89), respectively.

Beyond categorical classification, the model estimated 21 echocardiographic parameters with median (IQR) normalized MAEs of 0.13 (0.10-0.18) in internal validation and 0.16 (0.11-0.23) in the external RVENet+ cohort (**Figure 3**; eTable 4 in Supplement). The model accurately quantified LV dimensions and function, with MAEs of 4.2% (95% CI, 4.1%-4.3%) (external AUC, 4.5% [95% CI, 4.3%-4.7%]) for LVEF, 1.3 mm (95% CI, 1.3-1.3) (external AUC, 1.3 mm [95% CI, 1.3-1.4]) for LV interventricular septum thickness, 1.2 mm (95% CI, 1.1-1.2) (external AUC, 1.1 mm [95% CI, 1.0-1.1]) for LV posterior wall thickness, and 3.8 mm (95% CI, 3.7-3.9) (external AUC, 3.7 mm [95% CI, 3.6-3.7]) for LV internal diameter. For the right ventricle, the AI system estimated MAEs of 4.0 mm (95% CI, 3.9-4.1) (external AUC, 3.9 mm [95% CI, 3.7-4.0]) for right ventricular internal diameter diastole, 3.4 mm (95% CI, 3.3-3.4) (external AUC, 3.8 mm [95% CI, 3.6-3.9]) for tricuspid plane excursion velocity, and 1.9 cm/s (95% CI, 1.9-2.0) (external AUC, 2.0 cm/s [95% CI, 1.9-2.1]) for RV systolic excursion velocity. Atrial dimensions of left atrial internal diameter and left atrial volume were estimated with internal MAEs of 4.0 mm (95% CI, 3.9-4.1) (external AUC, 4.0 mm [95% CI, 3.8-4.3]) and 9.4 cm^3 (95% CI, 9.2-9.6) (external AUC, 13.4 cm^3 [95% CI, 12.9-13.9]), respectively. The model quantified Doppler-derived measurements with MAEs of 0.3 m/s (95% CI, 0.3-0.3) (external AUC, 0.4 m/s [95% CI, 0.3-0.4]) for peak aortic velocity, 5.6 mm Hg (95% CI, 5.4-5.7) (external AUC, 6.5 mm Hg [95% CI, 6.1-6.9]) for tricuspid peak gradient, and 2.0 (95% CI, 1.9-2.0) (external AUC, 2.2 [95% CI, 2.1-2.3]) for mean E/e' ratio. See eFigure 1 in Supplement 1 for Bland-Altman plots comparing key AI vs ground truth measurements.

In addition to accuracy, the model demonstrated group fairness by exhibiting demographic parity across sex and race (eFigure 2 in Supplement 1), robustness by reliably estimating LVEF under varying image quality (eFigure 3 in Supplement 1) and interpretability by correctly identifying the most relevant views for each task in accordance with guideline-recommended best practices[1] (**Figure 4**). For instance, in line with standard echocardiographic interpretation, the PLAX view was most informative for LV dimension measurements as well as aortic valve and aortic root characterization. Similarly, A4C was most informative for estimating LVEF and classifying LV dysfunction, Additionally, color Doppler videos were the most informative for all valvular regurgitation tasks, which typically involve assessment with color Doppler echocardiography. Detailed task-specific view relevance scores are depicted in eFigure 4 in Supplement 1.

Validation on Point-of-Care Echocardiography

The model performed 15 diagnostic classification tasks with a median (IQR) AUC of 0.91 (0.87-0.94) in an abbreviated TTE cohort, as well as 14 tasks with a median (IQR) AUC of 0.85 (95% CI, 0.77-0.87) in a POCUS cohort of bedside ED acquisitions (**Figure 5**; eTable 5 in Supplement 1). The AI system accurately performed ventricular assessment, with abbreviated TTE AUCs of 0.98 (95% CI, 0.97-0.98) (POCUS: 0.93 [95% CI, 0.92-0.94]) for moderate or worse LV systolic dysfunction in the simulated point-of-care cohort , 0.94 (95% CI, 0.92-0.95) (POCUS: 0.89 [95% CI, 0.87-0.90]) for moderate or higher increased LV size, 0.91 (95% CI, 0.89-0.92) (POCUS: 0.83 [95% CI, 0.82-0.85]) for moderate or worse LV diastolic dysfunction, 0.93 (95% CI, 0.91-0.94) (POCUS: 0.85 [95% CI, 0.83-0.87]) for moderate or greater increased LV wall thickness, 0.92 (95% CI, 0.90-0.93) (POCUS: 0.85 [95% CI, 0.84-0.86]) for RV systolic dysfunction, and 0.88 (95% CI, 0.86-0.90) (POCUS: 0.87 [95% CI, 0.86-0.89]) for moderate or greater increased RV size. The model also accurately detected severe aortic stenosis with an AUC of 0.96 (95% CI, 0.94-0.97) and mitral stenosis

Table. Description of Internal and External Validation Cohorts

	YNHHS			
	Development	Validation	RVENet+	POCUS
Dates sampled	1/2016-6/2022	7/2022-12/2022	11/2013-3/2021	9/2015-11/2023
Unique patients, No.	19 817	4588	831	3170
Unique studies, No.	27 135	5130	944	3310
Unique videos, No.	1 079 596	114 280	18 862	25 407
Unique videos per study, median (IQR)	39.0 (30.0-48.0)	17.0 (12.0-31.0)	19.5 (17.0-22.0)	6.0 (5.0-9.0)
Demographics				
Age, median (IQR), y	69.0 (58.0-79.0)	68.0 (57.0-78.0)	50.0 (25.0-66.0)	66.0 (53.0-79.0)
Sex, No. (%)				
Male	14 234 (52.5)	2585 (50.4)	586 (62.1)	1817 (54.9)
Female	12 901 (47.5)	2545 (49.6)	358 (37.9)	1493 (45.1)
Race, No. (%)[a]				
American Indian	68 (0.3)	15 (0.3)	0	115 (3.7)
Asian	436 (1.7)	112 (2.3)	0	68 (2.2)
Black	3711 (14.4)	679 (14.0)	0	871 (27.7)
Pacific Islander	36 (0.1)	10 (0.2)	0	12 (0.4)
White	20874 (80.9)	3893 (80.6)	944 (100)	2073 (65.8)
Other	693 (2.7)	124 (2.6)	0 (0)	10 (0.3)
Ethnicity, No. (%)[a]				
Hispanic	2094 (7.9)	472 (9.5)	0	2884 (88.6)
Non-Hispanic	24316 (92.1)	4486 (90.5)	944 (100)	372 (11.4)
Patient status, No. (%)				
Outpatient	16842 (62.5)	3511 (68.7)	524 (55.5)	0
Inpatient	8398 (31.2)	1296 (25.4)	420 (45.5)	0
Observation	1638 (6.1)	291 (5.7)	0	0
Emergency	63 (0.2)	11 (0.2)	0	3310 (100)
Hypertension, No. (%)	19982 (73.6)	3577 (69.7)	372 (39.4)	2323 (70.2)
Diabetes, No. (%)	7266 (26.8)	1243 (24.2)	119 (12.6)	1001 (30.2)
Dyslipidemia, No. (%)	5754 (21.2)	1164 (22.7)	239 (25.3)	493 (14.9)
Chronic kidney disease, No. (%)	2576 (9.5)	350 (6.8)	108 (11.4)	442 (13.4)
Coronary artery disease, No. (%)	5520 (20.3)	824 (16.1)	91 (9.6)	1103 (33.3)
Left ventricle				
LV systolic dysfunction, No. (%)				
Mild	1906 (7.5)	319 (6.5)	130 (13.8)	430 (13.8)
≥Moderate	1989 (7.9)	284 (5.8)	122 (12.9)	755 (24.2)
LV ejection fraction, median (IQR), %	62.0 (56.2-66.9)	61.4 (57.0-65.0)	57.5 (52.2-61.4)	57.0 (40.0-63.0)
LV diastolic dysfunction, No. (%)				
Mild	6872 (35.9)	1214 (29.8)	182 (22.0)	473 (26.7)
≥ Moderate	2965 (15.5)	333 (8.2)	159 (19.2)	416 (23.5)
Increased LV size, No. (%)				
Mild	1489 (7.3)	338 (7.9)	81 (8.6)	351 (13.3)
≥Moderate	1283 (6.3)	270 (6.3)	89 (9.4)	547 (20.8)
LVIDd, median (IQR), mm	45.5 (40.9-50.1)	46.0 (41.8-50.7)	49.0 (45.0-55.0)	47.2 (42.1-52.8)
LVIDs, median (IQR), mm	29.5 (25.7-34.0)	30.0 (26.2-34.0)	33.0 (28.0-38.0)	32.2 (27.6-39.0)
Increased LV wall thickness, No. (%)				
Any	13 499 (54.8)	1971 (40.2)	232 (24.6)	1206 (44.1)
≥ Moderate	3260 (13.2)	315 (6.4)	46 (4.9)	258 (9.4)
IVSd, median (IQR), mm	10.9 (9.1-12.6)	10.0 (8.8-11.5)	11.0 (9.0-12.0)	10.0 (8.9-11.7)
LVPWd, median (IQR), mm	10.4 (9.0-12.0)	10.0 (8.7-11.2)	9.0 (8.0-10.0)	10.0 (8.9-11.4)
LV wall motion abnormalities, No. (%)	1847 (20.7)	348 (22.6)	116 (12.3)	630 (44.5)

(continued)

Table. Description of Internal and External Validation Cohorts (continued)

	YNHHS			
	Development	Validation	RVENet+	POCUS
Right ventricle				
RV systolic dysfunction, No. (%)	2785 (11.3)	371 (7.8)	76 (8.1)	819 (29.1)
TAPSE, median (IQR), mm	21.0 (18.0-24.9)	21.2 (18.0-24.8)	21.0 (17.0-26.0)	19.0 (16.0-23.0)
RV S′, median (IQR), cm/s	12.3 (10.4-14.4)	12.6 (10.9-14.7)	13.0 (11.0-15.0)	11.9 (9.5-14.0)
Increased RV size, No. (%)				
Mild	2483 (10.0)	457 (9.6)	94 (10.0)	427 (15.3)
≥Moderate	1515 (6.1)	215 (4.5)	112 (11.9)	411 (14.7)
RVIDd, median (IQR), mm	31.0 (27.1-35.1)	30.9 (27.0-34.9)	31.0 (28.0-35.0)	32.0 (28.2-36.3)
Left atrium				
Increased LA size, No. (%)				
Mild	4897 (18.0)	805 (15.7)	125 (14.8)	403 (12.2)
≥Moderate	5392 (19.9)	798 (15.6)	273 (32.3)	741 (22.4)
LA volume, median (IQR), cm^3	58.5 (45.0-74.3)	55.2 (43.0-70.0)	64.5 (46.2-91.3)	61.9 (47.0-79.0)
Right atrium				
RA transverse dimension, median (IQR), mm	37.3 (33.0-43.0)	37.0 (33.0-42.0)	40.5 (35.0-45.0)	40.0 (34.0-46.0)
Increased RA size, No. (%)				
Any	3219 (29.7)	507 (29.3)	220 (23.3)	541 (41.3)
Valvular assessment				
AV stenosis, No. (%)				
Mild to moderate	2808 (12.1)	401 (8.5)	9 (1.0)	164 (7.1)
Severe	1934 (8.3)	86 (1.8)	85 (9.0)	38 (1.6)
AV regurgitation, No. (%)				
Mild	4309 (17.7)	773 (16.4)	140 (14.8)	398 (15.3)
≥Moderate	1645 (6.8)	272 (5.8)	31 (3.3)	164 (6.3)
AV structure, No. (%)				
Bicuspid	212 (0.9)	52 (1.1)	8 (0.8)	10 (0.4)
MV stenosis, No. (%)				
Any	812 (4.1)	89 (2.1)	11 (1.3)	71 (3.4)
MV regurgitation, No. (%)				
Mild	8594 (33.5)	1419 (28.3)	209 (25.2)	815 (29.1)
≥Moderate	4000 (15.6)	612 (12.2)	141 (17.0)	544 (19.4)
TV regurgitation, No. (%)				
Mild	8829 (33.8)	1580 (31.1)	310 (32.8)	939 (30.7)
≥ Moderate	3960 (15.2)	610 (12.0)	102 (10.8)	692 (22.6)
RV systolic pressure, median (IQR), mm Hg	29.0 (23.0-38.0)	27.5 (22.5-35.0)	28.0 (22.0-38.0)	37.0 (28.0-49.0)
Other				
Aortic root dim., median (IQR), mm	33.0 (30.0-36.0)	33.0 (30.0-36.6)	32.0 (30.0-34.2)	33.0 (30.0-36.0)
Any pericardial effusion, No. (%)	1006 (4.0)	21 (3.4)	47 (5.0)	305 (10.9)
Elevated RA pressure (≥ 8 mm Hg), No. (%)	4940 (20.8)	722 (15.5)	531 (56.2)	1200 (43.2)

Abbreviations: AV, aortic valve; IVSd, interventricular septum thickness at diastole; LA, left atrium; LV, left ventricle; LVIDd, left ventricular internal diameter at diastole; LVIDs, left ventricular internal diameter at systole; LVPWd, left ventricular posterior wall thickness at diastole; MV, mitral valve; RA, right atrium; RV, right ventricle; RVIDd, right ventricular internal diameter at diastole; RV S′, right ventricular systolic excursion velocity; TAPSE, tricuspid annular plane systolic excursion; TV, tricuspid valve; YNHHS, Yale New Haven Health System.

[a] Race was included in this specific analysis of fairness as per the TRIPOD+AI guidelines. Race was self-reported by patients based on fixed categories presented.

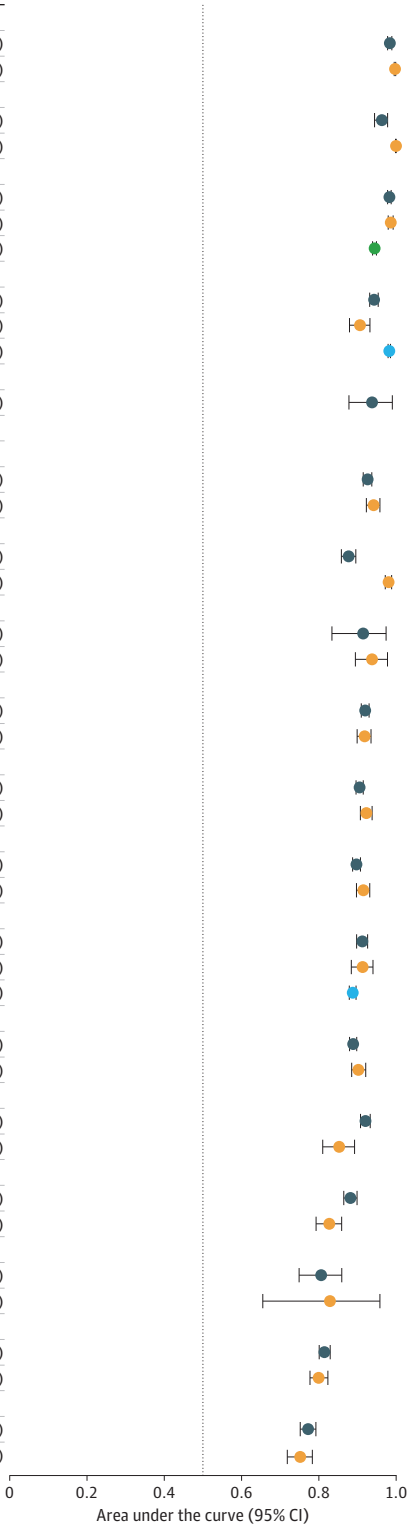
with an AUC of 0.94 (95% CI, 0.92-0.96) (POCUS: 0.92 [95% CI, 0.88-0.95]) from simplified acquisitions. Additional diagnostic findings of pericardial effusion and elevated LV outflow tract pressure were detected with AUCs of 0.91 (95% CI, 0.82-0.97) (POCUS: 0.86 [95% CI, 0.84-0.88]) and 0.94 (95% CI, 0.87-0.99), respectively.

Discussion

This study presents a view-agnostic deep learning model for automated echocardiography interpretation developed and validated on more than 1 million videos spanning a broad range

Figure 2. Multitask Classification Performance Evaluation

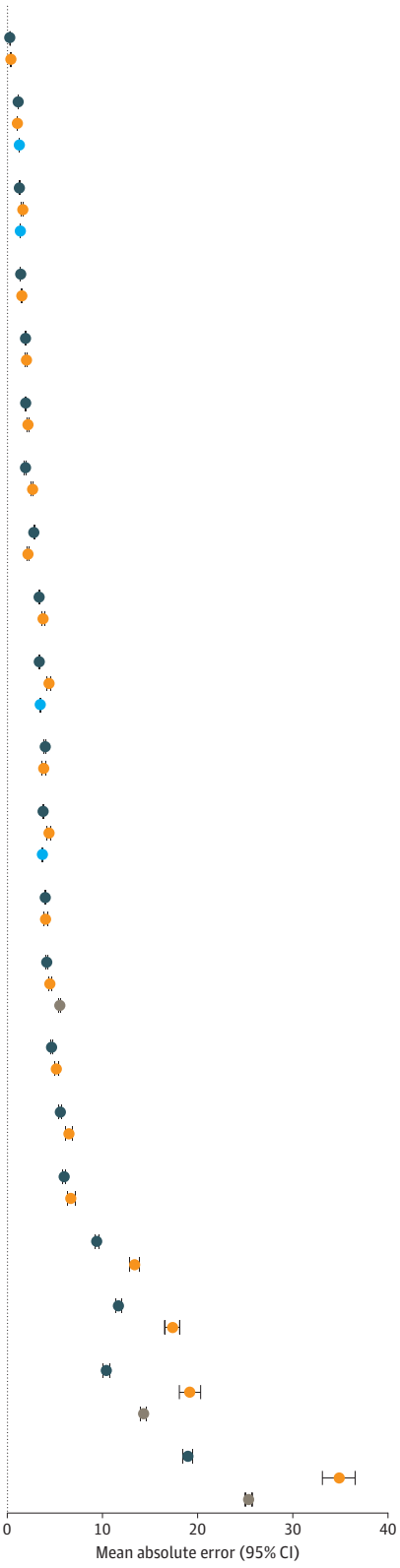
Diagnostic task	No. of studies meeting label definition/ No. with labels	Sensitivity (95% CI)	Specificity (95% CI)	Area under the curve (95% CI)
AV stenosis[a]				
YNHHS (US)	86/4701	0.94 (0.89-0.98)	0.95 (0.94-0.95)	0.98 (0.98-0.99)
RVENet+ (Hungary)	85/944	0.96 (0.93-0.99)	0.98 (0.97-0.99)	1.00 (0.99-1.00)
Mitral valve stenosis				
YNHHS (US)	89/4202	0.96 (0.96-0.97)	0.82 (0.75-0.88)	0.96 (0.94-0.98)
RVENet+ (Hungary)	11/829	1.00 (1.00-1.00)	0.95 (0.94-0.97)	1.00 (1.00-1.00)
LV systolic dysfunction[b]				
YNHHS (US)	284/4878	0.92 (0.90-0.95)	0.94 (0.94-0.95)	0.98 (0.98-0.99)
RVENet+ (Hungary)	122/944	0.97 (0.94-0.99)	0.89 (0.87-0.91)	0.99 (0.98-0.99)
EchoNet-Dynamic (US)	1264/10032	0.80 (0.78-0.82)	0.91 (0.91-0.92)	0.94 (0.94-0.95)
Increased LV size[b]				
YNHHS (US)	270/4268	0.81 (0.77-0.85)	0.92 (0.91-0.93)	0.94 (0.93-0.95)
RVENet+ (Hungary)	89/944	0.71 (0.63-0.78)	0.90 (0.88-0.92)	0.91 (0.88-0.93)
EchoNet-LVH (US)	418/11713	0.95 (0.93-0.97)	0.93 (0.92-0.93)	0.98 (0.98-0.99)
Elevated LVOT pressure gradient				
YNHHS (US)	24/4354	0.88 (0.75-1.00)	0.95 (0.95-0.96)	0.94 (0.88-0.99)
RVENet+ (Hungary)	0/944	NA	NA	NA
RV systolic dysfunction				
YNHHS (US)	371/4743	0.89 (0.88-0.90)	0.80 (0.77-0.84)	0.93 (0.91-0.94)
RVENet+ (Hungary)	76/944	0.92 (0.87-0.97)	0.82 (0.80-0.84)	0.94 (0.92-0.96)
LV wall motion abnormalities				
YNHHS (US)	348/1541	0.76 (0.73-0.80)	0.85 (0.83-0.86)	0.88 (0.86-0.90)
RVENet+ (Hungary)	116/944	0.98 (0.96-1.00)	0.71 (0.69-0.74)	0.98 (0.97-0.99)
Pericardial effusion				
YNHHS (US)	21/623	0.91 (0.89-0.93)	0.86 (0.73-0.96)	0.91 (0.83-0.97)
RVENet+ (Hungary)	47/944	0.68 (0.56-0.79)	1.00 (1.00-1.00)	0.94 (0.89-0.98)
LV diastolic dysfunction[b]				
YNHHS (US)	333/4071	0.88 (0.85-0.91)	0.81 (0.80-0.82)	0.92 (0.91-0.93)
RVENet+ (Hungary)	159/826	0.95 (0.92-0.98)	0.78 (0.76-0.81)	0.92 (0.90-0.93)
Mitral valve regurgitation[b]				
YNHHS (US)	612/5007	0.74 (0.71-0.77)	0.87 (0.86-0.88)	0.91 (0.90-0.91)
RVENet+ (Hungary)	141/829	0.88 (0.83-0.92)	0.83 (0.80-0.85)	0.92 (0.91-0.94)
Tricuspid valve regurgitation[b]				
YNHHS (US)	610/5076	0.76 (0.73-0.78)	0.86 (0.86-0.87)	0.90 (0.89-0.91)
RVENet+ (Hungary)	102/944	0.86 (0.81-0.92)	0.80 (0.77-0.82)	0.91 (0.90-0.93)
Increased LV wall thickness[b]				
YNHHS (US)	315/4815	0.90 (0.89-0.90)	0.75 (0.71-0.79)	0.91 (0.90-0.93)
RVENet+ (Hungary)	46/944	0.50 (0.38-0.62)	0.96 (0.95-0.97)	0.91 (0.88-0.94)
EchoNet-LVH (US)	703/11651	0.92 (0.92-0.93)	0.57 (0.54-0.60)	0.89 (0.88-0.90)
Increased LA size[b]				
YNHHS (US)	798/5130	0.79 (0.76-0.81)	0.83 (0.82-0.84)	0.89 (0.88-0.90)
RVENet+ (Hungary)	273/845	0.78 (0.74-0.82)	0.88 (0.86-0.91)	0.90 (0.88-0.92)
Atrial valve regurgitation[b]				
YNHHS (US)	272/4721	0.85 (0.82-0.89)	0.83 (0.82-0.84)	0.92 (0.91-0.93)
RVENet+ (Hungary)	31/944	0.77 (0.64-0.90)	0.76 (0.74-0.78)	0.85 (0.81-0.89)
Increased RV size[b]				
YNHHS (US)	215/4760	0.72 (0.66-0.77)	0.85 (0.84-0.86)	0.88 (0.86-0.90)
RVENet+ (Hungary)	112/944	0.64 (0.56-0.72)	0.81 (0.79-0.84)	0.83 (0.79-0.86)
Bicuspid aortic valve				
YNHHS (US)	52/4559	0.70 (0.69-0.71)	0.69 (0.58-0.80)	0.81 (0.75-0.86)
RVENet+ (Hungary)	8/944	0.88 (0.67-1.00)	0.60 (0.58-0.63)	0.83 (0.65-0.96)
Elevated RA pressure				
YNHHS (US)	722/4646	0.62 (0.59-0.65)	0.82 (0.81-0.83)	0.81 (0.80-0.83)
RVENet+ (Hungary)	531/944	0.49 (0.46-0.53)	0.96 (0.94-0.98)	0.80 (0.78-0.82)
Increased RA size				
YNHHS (US)	507/1732	0.71 (0.69-0.73)	0.69 (0.66-0.72)	0.77 (0.75-0.79)
RVENet+ (Hungary)	220/944	0.70 (0.65-0.76)	0.68 (0.65-0.71)	0.75 (0.72-0.78)



Multitask international validation of PanEcho on diagnostic classification tasks. Results are presented on the internal YNHHS validation cohort and the external RVENet+, EchoNet-Dynamic, and EchoNet-LVH cohorts. See eTable 1 for diagnostic definitions and guidelines. [a]Severe. [b]Moderate or worse.

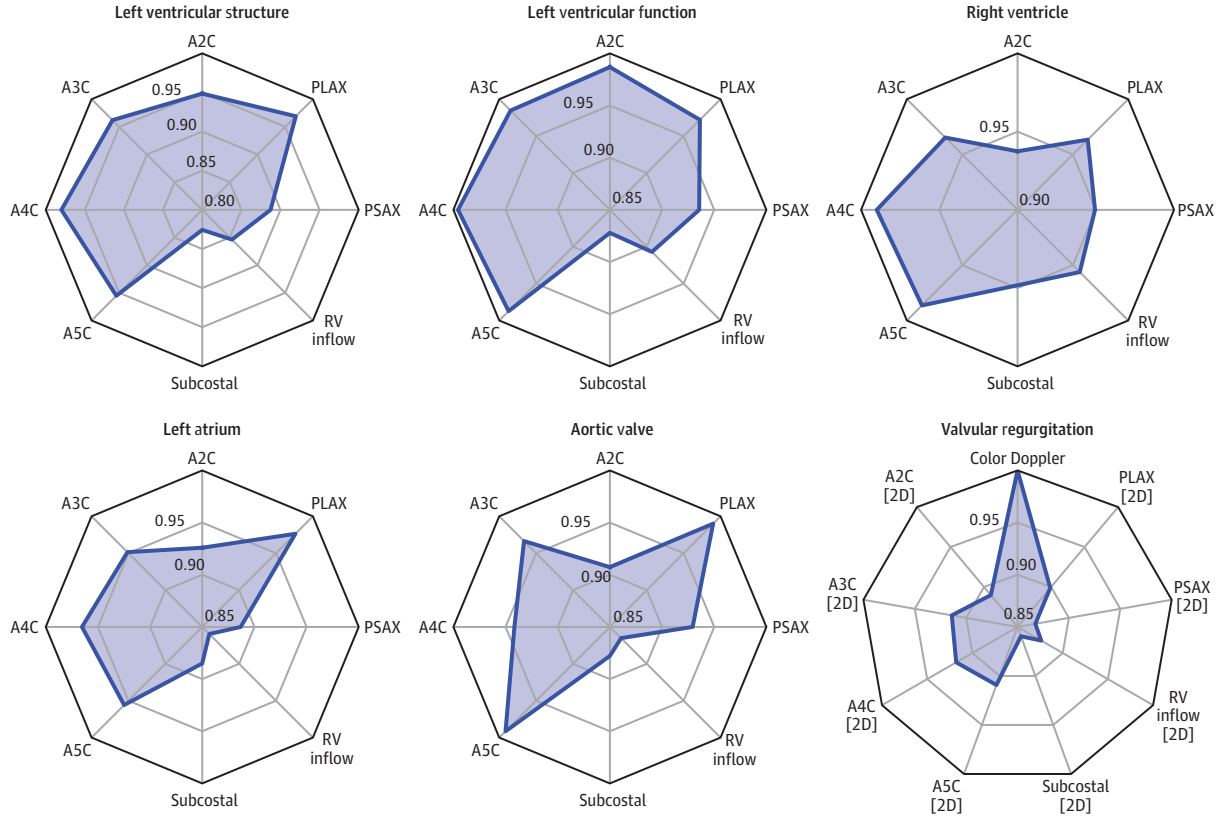
Figure 3. Multitask Regression Performance Evaluation

	No.	Median (IQR)	*r* (95% CI)	Mean absolute error (95% CI)
Aortic valve peak velocity, mis				
YNHHS (US)	4499	1.4 (1.2-1.8)	0.79 (0.78-0.81)	0.31 (0.30-0.32)
RVENet+ (Hungary)	934	1.3 (1.2-1.6)	0.84 (0.81-0.86)	0.36 (0.34-0.38)
LV posterior wall thickness at diastole				
YNHHS (US)	5040	10.0 (8.7-11.2)	0.68 (0.66-0.69)	1.15 (1.13-1.17)
RVENet+ (Hungary)	944	9.0 (8.0-10.0)	0.68 (0.65-0.71)	1.05 (1.01-1.09)
EchoNet-LVH (US)	11672	9.9 (8.6-11.3)	0.64 (0.63-0.65)	1.28 (1.26-1.29)
Interventricular septum thickness at diastole, mm				
YNHHS (US)	5045	10.0 (8.8-11.5)	0.70 (0.68-0.72)	1.30 (1.27-1.33)
RVENet+ (Hungary)	944	11.0 (9.0-12.0)	0.73 (0.70-0.75)	1.58 (1.52-1.65)
EchoNet-LVH (US)	11681	10.0 (8.6-11.7)	0.69 (0.68-0.70)	1.34 (1.32-1.36)
LV outflow tract diameter, mm				
YNHHS (US)	3719	20.0 (19.0-22.0)	0.59 (0.58-0.61)	1.35 (1.32-1.38)
RVENet+ (Hungary)	614	20.2 (18.9-22.0)	0.53 (0.47-0.58)	1.54 (1.47-1.61)
RV systolic excursion velocity, cm/s				
YNHHS (US)	3631	12.6 (10.9-14.7)	0.60 (0.58-0.63)	1.91 (1.86-1.96)
RVENet+ (Hungary)	831	13.0 (11.0-15.0)	0.62 (0.59-0.66)	1.99 (1.90-2.08)
E/e' ratio				
YNHHS (US)	4349	9.2 (7.2-12.1)	0.80 (0.79-0.81)	1.97 (1.92-2.01)
RVENet+ (Hungary)	728	7.0 (5.5-10.6)	0.84 (0.82-0.86)	2.18 (2.07-2.31)
Global longitudinal strain, %				
YNHHS (US)	461	19.0 (18.0-21.0)[a]	0.65 (0.59-0.71)	1.89 (1.78-2.01)
RVENet+ (Hungary)	866	17.6 (13.3-20.7)[a]	0.82 (0.80-0.84)	2.65 (2.54-2.75)
Aortic root dimension, mm				
YNHHS (US)	4968	33.0 (30.0-36.6)	0.69 (0.68-0.70)	2.82 (2.77-2.87)
RVENet+ (Hungary)	943	32.0 (30.0-34.2)	0.51 (0.47-0.55)	2.16 (2.08-2.26)
Tricuspid annular plane, mm				
YNHHS (US)	3734	21.2 (18.0-24.8)	0.59 (0.58-0.61)	3.36 (3.29-3.43)
RVENet+ (Hungary)	944	21.0 (17.0-26.0)	0.72 (0.69-0.74)	3.78 (3.64-3.92)
LVIDs, mm				
YNHHS (US)	5028	30.0 (26.2-34.0)	0.80 (0.78-0.81)	3.38 (3.32-3.45)
RVENet+ (Hungary)	814	33.0 (28.0-38.0)	0.86 (0.84-0.87)	4.40 (4.20-4.62)
EchoNet-LVH (US)	11268	31.1 (27.2-36.0)	0.86 (0.85-0.86)	3.45 (3.40-3.50)
RVIDd, mm				
YNHHS (US)	3430	30.9 (27.0-34.9)	0.57 (0.54-0.59)	3.97 (3.87-4.06)
RVENet+ (Hungary)	944	31.0 (28.0-35.0)	0.58 (0.53-0.62)	3.85 (3.67-4.03)
LVIDd, mm				
YNHHS (US)	5039	46.0 (41.8-50.7)	0.74 (0.73-0.75)	3.78 (3.70-3.85)
RVENet+ (Hungary)	944	49.0 (45.0-55.0)	0.80 (0.78-0.82)	4.37 (4.19-4.56)
EchoNet-LVH (US)	11713	46.1 (41.7-51.2)	0.81 (0.81-0.82)	3.69 (3.64-3.73)
LAIDs, mm				
YNHHS (US)	4832	39.6 (35.0-45.0)	0.79 (0.78-0.80)	3.98 (3.90-4.05)
RVENet+ (Hungary)	601	40.0 (34.0-46.0)	0.79 (0.77-0.81)	4.03 (3.81-4.25)
LV ejection fraction, %				
YNHHS (US)	5045	61.7 (57.0-65.1)	0.85 (0.84-0.86)	4.17 (4.08-4.25)
RVENet+ (Hungary)	944	57.5 (52.2-61.4)	0.87 (0.86-0.89)	4.53 (4.33-4.71)
EchoNet-Dynamic (US)	10032	59.2 (51.6-64.0)	0.79 (0.78-0.80)	5.53 (5.45-5.62)
RA transverse dimension, mm				
YNHHS (US)	2897	37.0 (33.0-42.0)	0.62 (0.60-0.64)	4.67 (4.57-4.78)
RVENet+ (Hungary)	944	40.5 (35.0-45.0)	0.57 (0.53-0.61)	5.15 (4.95-5.35)
Tricuspid valve peak gradient, mm Hg				
YNHHS (US)	3515	24.0 (19.0-30.0)	0.64 (0.62-0.66)	5.57 (5.44-5.72)
RVENet+ (Hungary)	689	22.0 (17.0-30.0)	0.68 (0.65-0.71)	6.48 (6.08-6.85)
RV systolic pressure, mm Hg				
YNHHS (US)	3298	27.5 (22.5-35.0)	0.68 (0.66-0.70)	5.98 (5.84-6.14)
RVENet+ (Hungary)	689	28.0 (22.0-38.0)	0.73 (0.70-0.75)	6.72 (6.30-7.12)
LA volume, cm^3				
YNHHS (US)	4107	55.2 (43.0-70.0)	0.78 (0.77-0.79)	9.41 (9.22-9.60)
RVENet+ (Hungary)	845	64.5 (46.2-91.3)	0.80 (0.78-0.82)	13.4 (12.9-13.9)
LV stroke volume, cm^3				
YNHHS (US)	3395	62.0 (48.3-79.0)	0.77 (0.76-0.78)	11.7 (11.4-12.0)
RVENet+ (Hungary)	944	69.5 (56.4-84.6)	0.62 (0.58-0.66)	17.3 (16.5-18.1)
LV end systolic volume, cm^3				
YNHHS (US)	3435	40.1 (28.9-56.4)	0.90 (0.90-0.91)	10.4 (10.1-10.8)
RVENet+ (Hungary)	944	55.0 (41.9-78.0)	0.90 (0.88-0.91)	19.2 (18.1-20.3)
EchoNet-Dynamic (US)	10032	33.6 (23.7-49.1)	0.82 (0.81-0.83)	14.3 (14.0-14.6)
LV end diastolic volume, cm^3				
YNHHS (US)	3449	105.1 (80.6-136.9)	0.86 (0.85-0.87)	19.0 (18.5-19.5)
RVENet+ (Hungary)	944	128.9 (104.0-166.1)	0.81 (0.78-0.83)	34.8 (33.1-36.6)
EchoNet-Dynamic (US)	10032	82.1 (62.2-108.3)	0.70 (0.69-0.71)	25.4 (25.0-25.8)



Results are presented on the internal Yale New Haven Health System (YNHHS) validation cohort and the external RVENet+, EchoNet-Dynamic, and EchoNet-LVH cohorts. Median (IQR) is reported for the ground truth measurements in each cohort.

Figure 4. Task-Specific View Relevance



Radar plots depicting the relative importance of each echocardiographic view for a given aspect of cardiovascular diagnosis. For each task, a normalized view relevance score was computed, where 1 indicates the most relevant view; this score reflects the fraction of the maximum performance metric (area under the receiver operator characteristic curve for classification tasks and mean absolute error [MAE] for regression tasks) on a given task when only using videos from an individual view. Presented view relevance scores were then averaged over all tasks falling under a given category in each radar plot. Analysis was performed on the internal YNHHS validation set using up to 3 videos from a given echocardiographic view per study. A2C indicates apical 2-chamber; A3C, apical 3-chamber; A4C, apical 4-chamber; A5C, apical 5-chamber; AUC, area under the receiver operator characteristic curve; MAE, mean absolute error; PLAX, parasternal long axis; PSAX, parasternal short axis; RV, right ventricle; YNHHS, Yale New Haven Health System.

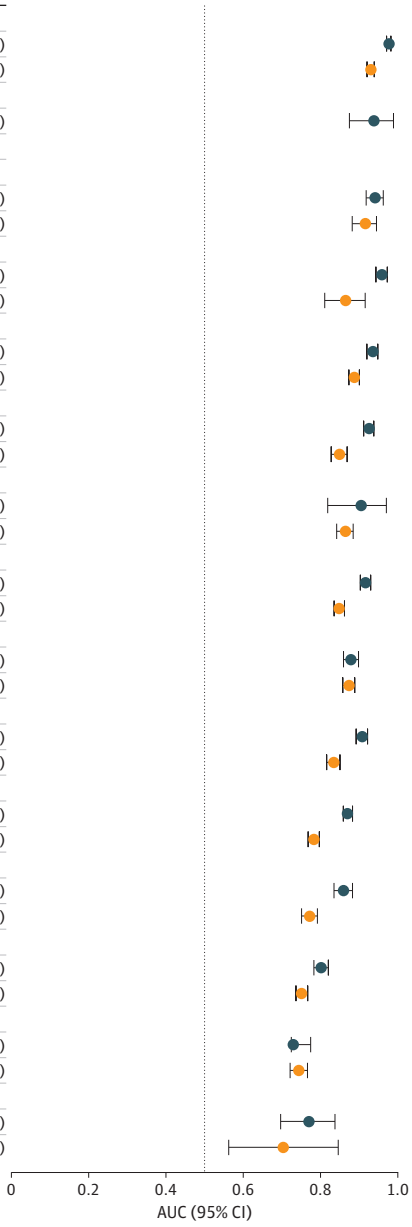
of acquisitions and phenotypes. PanEcho, the AI system, advances the current state of AI-enabled echocardiography, enabling simultaneous estimation of all key parameters of cardiac structure and function from any combination of 2D views. The model integrates multiview information from complete echocardiographic studies while maintaining accurate reporting on simplified imaging protocols and technically limited POCUS acquisitions from the ED. The model is open-source, extensively validated internationally, and at the point of care and may signal a shift from specialized single-task models toward generalized multitask AI models poised for clinical deployment.

This AI system was developed to address a critical gap in AI for echocardiography: the predominance of single-view, single-task models. Unlike prior approaches that require a particular echocardiographic view or sequence,[12,14,16,24,31] the model flexibly integrates videos from all available 2D views. Further, although previous work has primarily been limited to specialized single-task models,[8,13,14,20,22-24] the current model performs end-to-end estimation of all variables forming the core of a complete echocardiographic report. This unified approach is closer to clinical practice, in which a cardiologist synthesizes multiple views and imaging modes to assess all aspects of cardiovascular health, and more efficiently scales to clinical deployment. In the previous paradigm, single-task models would pose considerable implementation challenges, especially in computationally constrained environments such as on-device deployment on a handheld device for point-of-care use.

To understand the contributions of this study, the AI system should be evaluated in the context of state-of-the-art commercial systems[16,31] and academic efforts.[25,32] Although all report video-based AI models that are capable of automating many aspects of echocardiographic reporting, key differences exist in accessibility and extent of validation. First, unlike commercial products, the current model is fully open-source, accelerating research on AI for echocardiography. Second, the model is evaluated across 39 diverse labels spanning key diagnostic findings in a complete echocardiographic report, while comparable studies validate only on subsets of these labels. Third, the model has been internationally validated across multiple sites on both complete diagnostic quality studies and technically limited acquisitions by noncardiovascular operators, in contrast to prior work. This extensive

Figure 5. Performance Evaluation at the Point of Care

	No. of studies meeting label definition/ No. with labels	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
LV systolic dysfunction[a]				
Abbreviated TTE	212/3440	0.81 (0.77-0.85)	0.86 (0.82-0.90)	0.98 (0.97-0.98)
POCUS	755/3114	0.89 (0.87-0.91)	0.80 (0.79-0.82)	0.93 (0.92-0.94)
Elevated LVOT PG				
Abbreviated TTE	22/3096	0.54 (0.37-0.71)	0.73 (0.56-0.88)	0.94 (0.87-0.99)
POCUS	2/2433	NA	NA	NA
MV stenosis[b]				
Abbreviated TTE	67/2915	1.00 (1.00-1.00)	0.97 (0.96-0.97)	0.94 (0.92-0.96)
POCUS	71/2062	0.95 (0.95-0.96)	0.63 (0.54-0.72)	0.92 (0.88-0.95)
AV stenosis				
Abbreviated TTE	69/3285	0.42 (0.33-0.53)	0.83 (0.75-0.90)	0.96 (0.94-0.97)
POCUS	38/2311	0.55 (0.41-0.68)	0.96 (0.95-0.96)	0.86 (0.81-0.92)
Increased LV size[a]				
Abbreviated TTE	201/2981	0.63 (0.58-0.69)	0.76 (0.71-0.81)	0.94 (0.92-0.95)
POCUS	547/2632	0.72 (0.69-0.75)	0.89 (0.88-0.90)	0.89 (0.87-0.90)
Increased LV wall thickness[a]				
Abbreviated TTE	250/3355	0.99 (0.99-0.99)	0.90 (0.89-0.91)	0.93 (0.91-0.94)
POCUS	258/2710	0.70 (0.68-0.71)	0.83 (0.79-0.86)	0.85 (0.83-0.87)
Pericardial effusion				
Abbreviated TTE	21/619	0.99 (0.98-1.00)	0.91 (0.89-0.93)	0.91 (0.82-0.97)
POCUS	305/2796	0.48 (0.46-0.50)	0.93 (0.90-0.95)	0.86 (0.84-0.88)
RV systolic dysfunction				
Abbreviated TTE	263/3347	0.99 (0.99-0.99)	0.90 (0.89-0.91)	0.92 (0.90-0.93)
POCUS	819/2813	0.71 (0.69-0.72)	0.81 (0.79-0.83)	0.85 (0.84-0.86)
Increased RV size[a]				
Abbreviated TTE	158/3360	0.27 (0.22-0.33)	0.69 (0.63-0.75)	0.88 (0.86-0.90)
POCUS	411/2793	0.97 (0.96-0.98)	0.38 (0.37-0.40)	0.87 (0.86-0.89)
LV diastolic dysfunction[a]				
Abbreviated TTE	247/2825	0.51 (0.46-0.57)	0.85 (0.81-0.88)	0.91 (0.89-0.92)
POCUS	416/1769	0.79 (0.75-0.82)	0.74 (0.72-0.76)	0.83 (0.82-0.85)
Increased LA size[a]				
Abbreviated TTE	598/3608	0.56 (0.53-0.60)	0.74 (0.72-0.77)	0.87 (0.86-0.88)
POCUS	741/3310	0.67 (0.64-0.69)	0.74 (0.72-0.75)	0.78 (0.77-0.80)
LV wall motion abnormalities				
Abbreviated TTE	249/1096	0.72 (0.67-0.76)	0.73 (0.68-0.78)	0.86 (0.84-0.88)
POCUS	630/1415	0.81 (0.78-0.83)	0.53 (0.50-0.56)	0.77 (0.75-0.79)
Elevated RA pressure				
Abbreviated TTE	485/3289	0.48 (0.44-0.53)	0.62 (0.58-0.66)	0.80 (0.78-0.82)
POCUS	1200/2775	0.96 (0.95-0.97)	0.22 (0.20-0.23)	0.75 (0.74-0.77)
Increased RA size				
Abbreviated TTE	374/1267	0.87 (0.84-0.89)	0.70 (0.68-0.73)	0.75 (0.72-0.77)
POCUS	541/1310	0.36 (0.34-0.39)	0.89 (0.87-0.91)	0.74 (0.72-0.77)
Bicuspid AV				
Abbreviated TTE	37/3219	1.00 (0.99-1.00)	0.75 (0.74-0.76)	0.77 (0.70-0.84)
POCUS	10/2567	0.95 (0.94-0.95)	0.30 (0.00-0.56)	0.70 (0.56-0.85)



Multitask validation of PanEcho on diagnostic classification tasks in a simulated and real-world POCUS dataset. Yale New Haven Health System (YNHHS) 5-view refers to evaluation on the internal YNHHS validation set, except only using up to 1 video from the following key views per study: PLAX, PSAX, A4C, A5C, or A2C. Error bars and values in parentheses represent bootstrapped 95% CIs. The dotted line represents the performance of random guessing. A2C indicates apical 2-chamber; A3C, apical 3-chamber; A4C, apical 4-chamber; A5C, apical 5-chamber; AUC, area under the receiver operating characteristic curve; AV, aortic valve; LA, left atrium; LV, left ventricle; LVOT, left ventricular outflow tract; MV, mitral valve; NA, not applicable; PG, pressure gradient; PLAX, parasternal long axis; POCUS, point-of-care ultrasonography; PSAX, parasternal short axis; RA, right atrium; RV, right ventricle; TTE, transthoracic echocardiography; TV, tricuspid valve; YNHHS, Yale New Haven Health System.

[a]Moderate or worse.

[b]Severe.

validation enables unique opportunities for advancing diagnostic care in point-of-care settings and community clinics, where access to equipment and personnel needed for complete diagnostic echocardiography is limited.

There are 3 primary clinical applications for PanEcho: as a preliminary reader for assisted interpretation in echocardiography laboratories, as a tool to identify potentially missed abnormalities in existing echocardiography databases, and as

an efficient screening system for accelerated protocols and point-of-care examinations performed by nonexperts. Given its excellent performance on complete echocardiograms, the model can support expert readers by identifying key abnormalities for further interrogation. Prospective integration into a real-world echocardiography laboratory workflow is needed to evaluate factors such as interface design and the effect of human-AI collaboration on interpretation speed and accuracy. The model may also be used to efficiently parse existing repositories of echocardiograms for potential clinical abnormalities that were missed at the time of examination to be flagged for review. Also, the AI system is uniquely suited for point-of-care settings and community clinics given its extensive validation on POCUS acquisitions. Given the scarcity of echocardiographic expertise[33,34] and increasing accessibility of portable ultrasound devices,[6,7,35] the model could provide cardiovascular health care that might otherwise be inaccessible to many communities. Future work to validate this approach includes prospective human vs AI evaluation on point-of-care scans from nonsonographers.

Limitations

Certain limitations merit consideration. First, this model was limited to 2D grayscale and color Doppler echocardiographic videos, excluding other acquisitions such as still frames, spectral Doppler, strain imaging, and 3-dimensional (3D) echocardiography, which could improve the predictive performance and versatility of the model. Second, unlike other approaches,[13,15-17,31] the current method does not incorporate a segmentation step for echocardiographic measurements. Although the direct estimation approach yields strong numerical results, it may limit the model's interpretability; future iterations of the AI system may supplement the current approach with a segmentation overlay on select views for improved transparency and ease of use in clinical deployment. Systematic differences in measurement methods[36] limit performance on tasks such as LV volumes, where RVENet+ labels were derived from 3D echocardiography vs the mixture of 2D and 3D methods used in routine practice across other cohorts. Similarly, performance on certain labels— such as bicuspid aortic valve and right atrial abnormalities—is limited by low prevalence[37] and difficulty of interpretation for human experts,[38-40] causing class imbalance and noisy ground truth labels, respectively. Third, this model was limited to the specific suite of 39 tasks defined in this study and, for certain tasks, may fail to provide sufficiently granular interpretations to independently guide patient care. For this reason, a PanEcho-assisted workflow would still ultimately rely on expert supervision to ensure reliable diagnosis and treatment based on the broader clinical context of each patient beyond echocardiography.

Conclusions

Results of this study describe the development and validation of an AI system that automatically interprets key aspects of echocardiograms—spanning ventricular structure and function to valvular disease and more—maintaining high accuracy across geography and time from complete and technically limited studies. These findings support further research, including prospective evaluation in real-world clinical workflows, which is needed to assess the efficacy of PanEcho and similar tools as adjunct readers in echocardiography laboratories or AI-enabled screening tools in point-of-care settings.

ARTICLE INFORMATION

Accepted for Publication: May 10, 2025.

Published Online: June 23, 2025. doi:10.1001/jama.2025.8731

Author Affiliations: Department of Electrical and Computer Engineering, The University of Texas at Austin (Holste, Wang); Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut (Holste, Oikonomou, Khera); Cardiovascular Data Science (CarDS) Lab, Yale School of Medicine, New Haven, Connecticut (Holste, Oikonomou, Khera); Department of Experimental Cardiology and Surgical Techniques, Heart and Vascular Center, Semmelweis University, Budapest, Hungary (Tokodi, Kovács); Institute for Clinical Data Management, Semmelweis University, Budapest, Hungary (Kovács); Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut (Khera); Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, Connecticut (Khera); Section of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, Connecticut (Khera); Associate Editor, *JAMA* (Khera).

Author Contributions: Dr Khera had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Mr Holste and Dr Oikonomou contributed equally as co–first authors.
Concept and design: Holste, Oikonomou, Wang, Khera.
Acquisition, analysis, or interpretation of data: Holste, Oikonomou, Tokodi, Kovács, Khera.
Drafting of the manuscript: Holste, Oikonomou.
Critical review of the manuscript for important intellectual content: All authors.
Statistical analysis: Holste, Oikonomou, Tokodi, Khera.
Obtained funding: Khera.
Administrative, technical, or material support: Kovács, Khera.
Supervision: Kovács, Wang, Khera.
Other - AI model development and validation: Holste.

Conflict of Interest Disclosures: Dr Oikonomou reported receiving support from the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number F32HL170592 and being cofounder of Evidence2Health, LLC, providing ad hoc consultancy and having stock options in Caristo Diagnostics Ltd, and consulting for Ensight-AI, Inc outside the submitted work. Dr Tokodi reported receiving grants from the Hungarian Academy of Sciences during the conduct of the study. Dr Kovács reported receiving grants from European Union–MILAB, National Research, Development, and Innovation Office of Hungary (FK 142573), and the Hungarian Academy of Sciences during the conduct of the study and personal fees from Argus Cognitive, Inc and equity in CardioSight, Inc outside the submitted work. Dr Khera reported receiving grants from National Institutes of Health (1R01HL167858, 1R01AG089981, and 5K23HL153775), Doris Duke Charitable Foundation (award No. 2022060), Bristol Myers Squibb (through Yale), Novo Nordisk (through Yale), and BridgeBio (through Yale) and serving as an academic cofounder of Evidence2Health and Ensight-AI outside the submitted work. No other disclosures were reported.

Disclaimer: Dr Khera is an associate editor of *JAMA* but was not involved in any of the decisions regarding review of the manuscript or its acceptance

Data Sharing Statement: See Supplement 2.

Additional Contributions: We thank Andreas Coppi, PhD (Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine), for instrumental work facilitating the external validation of PanEcho in the RVENet+ cohort. We also gratefully acknowledge Béla Merkely, MD, PhD (Heart and Vascular Center, Semmelweis University, Budapest, Hungary), for his leadership in the establishment and sustained development of the RVENet+ dataset.

REFERENCES

1. Mitchell C, Rahko PS, Blauwet LA, et al. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the American Society of Echocardiography. *J Am Soc Echocardiogr*. 2019;32(1):1-64. doi:10.1016/j.echo.2018.06.004

2. Wei C, Milligan M, Lam M, Heidenreich PA, Sandhu A. Variation in cost of echocardiography within and across United States hospitals. *J Am Soc Echocardiogr*. 2023;36(6):569-577.e4. doi:10.1016/j.echo.2023.01.002

3. Virnig BA, Shippee ND, O'Donnell B, Zeglin J, Parashuram S. *Trends in the Use of Echocardiography, 2007 to 2011*. Agency for Healthcare Research and Quality; 2014.

4. Pillai B, Salerno M, Schnittger I, Cheng S, Ouyang D. Precision of echocardiographic measurements. *J Am Soc Echocardiogr*. 2024;37(5):562-563. doi:10.1016/j.echo.2024.01.001

5. He B, Kwan AC, Cho JH, et al. Blinded, randomized trial of sonographer versus AI cardiac function assessment. *Nature*. 2023;616(7957):520-524. doi:10.1038/s41586-023-05947-3

6. Díaz-Gómez JL, Mayo PH, Koenig SJ. Point-of-care ultrasonography. *N Engl J Med*. 2021;385(17):1593-1602. doi:10.1056/NEJMra1916062

7. Narang A, Bae R, Hong H, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. *JAMA Cardiol*. 2021;6(6):624-632. doi:10.1001/jamacardio.2021.0185

8. Holste G, Oikonomou EK, Mortazavi BJ, et al. Severe aortic stenosis detection by deep learning applied to echocardiography. *Eur Heart J*. 2023;44(43):4592-4604. doi:10.1093/eurheartj/ehad456

9. Krishna H, Desai K, Slostad B, et al. Fully automated artificial intelligence assessment of aortic stenosis by echocardiography. *J Am Soc Echocardiogr*. 2023;36(7):769-777. doi:10.1016/j.echo.2023.03.008

10. Huang Z, Long G, Wessler B, Hughes MC. A New Semi-supervised Learning Benchmark for Classifying View and Diagnosing Aortic Stenosis from Echocardiograms. In: Jung K, Yeung S, Sendak M, Sjoding M, Ranganath R, eds. Proceedings of the 6th Machine Learning for Healthcare Conference. Vol 149. Proceedings of Machine Learning Research. 2021:614-647.

11. Oikonomou EK, Holste G, Yuan N, et al. A multimodal video-based AI biomarker for aortic stenosis development and progression. *JAMA Cardiol*. 2024;9(6):534-544. doi:10.1001/jamacardio.2024.0595

12. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation*. 2018;138(16):1623-1635. doi:10.1161/CIRCULATIONAHA.118.034338

13. Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*. 2020;580(7802):252-256. doi:10.1038/s41586-020-2145-8

14. Ghorbani A, Ouyang D, Abid A, et al. Deep learning interpretation of echocardiograms. *NPJ Digit Med*. 2020;3:10. doi:10.1038/s41746-019-0216-8

15. Reddy CD, Lopez L, Ouyang D, Zou JY, He B. Video-based deep learning for automated assessment of left ventricular ejection fraction in pediatric patients. *J Am Soc Echocardiogr*. 2023;36(5):482-489. doi:10.1016/j.echo.2023.01.015

16. Tromp J, Bauer D, Claggett BL, et al. A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *Nat Commun*. 2022;13(1):6776. doi:10.1038/s41467-022-34245-1

17. Zeng Y, Tsui PH, Pang K, et al. MAEF-Net: multi-attention efficient feature fusion network for left ventricular segmentation and quantitative analysis in two-dimensional echocardiography. *Ultrasonics*. 2023;127:106855. doi:10.1016/j.ultras.2022.106855

18. Khera R, Oikonomou EK, Nadkarni GN, et al. Transforming cardiovascular care with artificial intelligence: from discovery to practice: JACC state-of-the-art review. *J Am Coll Cardiol*. 2024;84(1):97-114. doi:10.1016/j.jacc.2024.05.003

19. Goto S, Mahara K, Beussink-Nelson L, et al. Artificial intelligence-enabled fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. *Nat Commun*. 2021;12(1):2726. doi:10.1038/s41467-021-22877-8

20. Long A, Haggerty CM, Finer J, et al. deep learning for echo analysis, tracking, and evaluation of mitral regurgitation (DELINEATE-MR). *Circulation*. 2024;150(12):911-922. doi:10.1161/CIRCULATIONAHA.124.068996

21. Ferreira DL, Salaymang Z, Arnaout R. Label-free segmentation from cardiac ultrasound using self-supervised learning. arXiv. Preprint posted online October 10, 2022. http://arxiv.org/abs/2210.04979

22. Vrudhula A, Duffy G, Vukadinovic M, et al. High throughput deep learning detection of mitral regurgitation. *medRxiv*. Preprint posted online February 14, 2024. doi:10.1101/2024.02.08.24302547

23. Vrudhula A, Vukadinovic M, Haeffle C, et al. Deep learning phenotyping of tricuspid regurgitation for automated high throughput assessment of transthoracic echocardiography. *medRxiv*. Preprint posted June 24, 2024. doi:10.1101/2024.06.22.24309332

24. Duffy G, Cheng PP, Yuan N, et al. High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning. *JAMA Cardiol*. 2022;7(4):386-395. doi:10.1001/jamacardio.2021.6059

25. Christensen M, Vukadinovic M, Yuan N, Ouyang D. Vision-language foundation model for echocardiogram interpretation. *Nat Med*. 2024;30(5):1481-1488. doi:10.1038/s41591-024-02959-y

26. Holste G, Oikonomou EK, Mortazavi BJ, Wang Z, Khera R. Efficient deep learning-based automated diagnosis from echocardiography with contrastive self-supervised learning. *Commun Med (Lond)*. 2024;4(1):133. doi:10.1038/s43856-024-00538-3

27. Magyar B, Tokodi M, Soós A, et al. RVENet: A large echocardiographic dataset for the deep learning-based assessment of right ventricular function. In: *Lecture Notes in Computer Science*. Springer Nature; 2023:569-583.

28. Tokodi M, Magyar B, Soós A, et al. Deep learning-based prediction of right ventricular ejection fraction using 2D echocardiograms. *JACC Cardiovasc Imaging*. 2023;16(8):1005-1018. doi:10.1016/j.jcmg.2023.02.017

29. Oikonomou EK, Vaid A, Holste G, et al. Artificial intelligence-guided detection of under-recognised cardiomyopathies on point-of-care cardiac ultrasonography. *Lancet Digit Health*. 2025;7(2):e113-e123. doi:10.1016/S2589-7500(24)00249-8

30. Vaswani A, Shazeer NM, Parmar N, et al. Attention is all you need. In: Guyon I, ed. *Advances in Neural Information Processing Systems 30*. NIPS Foundation; 2017: 5998-6008.

31. Tromp J, Seekings PJ, Hung CL, et al. Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *Lancet Digit Health*. 2022;4(1):e46-e54. doi:10.1016/S2589-7500(21)00235-1

32. Vukadinovic M, Tang X, Yuan N, et al. EchoPrime: a multi-video view-informed vision-language model for comprehensive echocardiography interpretation. arXiv. Preprint posted online October 12, 2024. http://arxiv.org/abs/2410.09704

33. Narang A, Sinha SS, Rajagopalan B, et al. The supply and demand of the cardiovascular workforce. *J Am Coll Cardiol*. 2016;68(15):1680-1689. doi:10.1016/j.jacc.2016.06.070

34. Colebourn CL. Future-proofing UK echocardiography. *Br J Cardiol*. 2023;30(4):36.

35. Ginsburg AS, Liddy Z, Khazaneh PT, May S, Pervaiz F. A survey of barriers and facilitators to ultrasound use in low- and middle-income countries. *Sci Rep*. 2023;13(1):3322. doi:10.1038/s41598-023-30454-w

36. Dorosz JL, Lezotte DC, Weitzenkamp DA, Allen LA, Salcedo EE. Performance of 3-dimensional echocardiography in measuring left ventricular volumes and ejection fraction: a systematic review and meta-analysis. *J Am Coll Cardiol*. 2012;59(20):1799-1808. doi:10.1016/j.jacc.2012.01.037

37. Sillesen AS, Vøgg O, Pihl C, et al. Prevalence of bicuspid aortic valve and associated aortopathy in newborns in Copenhagen, Denmark. *JAMA*. 2021;325(6):561-567. doi:10.1001/jama.2020.27205

38. Jain R, Ammar KA, Kalvin L, et al. Diagnostic accuracy of bicuspid aortic valve by echocardiography. *Echocardiography*. 2018;35(12):1932-1938. doi:10.1111/echo.14167

39. Sun ZY, Li Q, Li J, Zhang MW, Zhu L, Geng J. Echocardiographic evaluation of the right atrial size and function: Relevance for clinical practice. *Am Heart J Plus*. 2023;27(100274):100274. doi:10.1016/j.ahjo.2023.100274

40. Grünig E, Henn P, D'Andrea A, et al. Reference values for and determinants of right atrial area in healthy adults by 2-dimensional echocardiography. *Circ Cardiovasc Imaging*. 2013;6(1):117-124. doi:10.1161/CIRCIMAGING.112.978031