JACC VOL. 86, NO. 13, 2025

Artificial Intelligence Automation of Echocardiographic Measurements

Yuki Sahashi, MD, MSc, PhD,[a] Hirotaka Ieki, MD, PhD,[b] Victoria Yuan, BS,[a,c] Matthew Christensen,[a] Milos Vukadinovic, BS,[a,d,e] Christina Binder-Rodriguez, MD,[a] Justin Rhee, BS,[f] James Y. Zou, PhD,[g,h,i] Bryan He, PhD,[g] Paul Cheng, MD, PhD,[b] David Ouyang, MD[a,j]

ABSTRACT

BACKGROUND Accurate measurement of echocardiographic parameters is crucial for the diagnosis of cardiovascular disease and tracking of change over time; however, manual assessment requires time-consuming effort and can be imprecise. Artificial intelligence has the potential to reduce clinician burden by automating the time-intensive task of comprehensive measurement of echocardiographic parameters.

OBJECTIVES The purpose of this study was to develop and validate open-sourced deep learning semantic segmentation models for the automated measurement of 18 anatomic and Doppler measurements in echocardiography.

METHODS We trained models for the automated measurement of echocardiography parameters using data sets between 2011 and 2023 from Cedars-Sinai Medical Center (CSMC). The outputs of segmentation models were compared with sonographer measurements from temporal split data from CSMC and an external data set from Stanford Healthcare (SHC) to access accuracy and precision.

RESULTS We used 877,983 echocardiographic measurements from 155,215 studies from CSMC to develop EchoNet-Measurements, an open-source deep learning model for echocardiographic annotation. The models demonstrated high accuracy when compared with sonographer measurements from held-out data from CSMC and an independent external validation data set from SHC. Measurements across all 9 B-mode and 9 Doppler measurements had high accuracy (mean coverage probability of 0.796 and 0.839 and mean relative difference of 0.120 and 0.096 on held-out test set from CSMC and external data set from SHC, respectively). When evaluated end-to-end on 2,103 temporally distinct studies at CSMC, EchoNet-Measurements had similar reasonable performance (mean coverage probability 0.803 and mean relative difference of 0.108). Performance was consistent across patient characteristics including age, sex, and atrial fibrillation, obesity status, and machine vendors.

CONCLUSIONS EchoNet-Measurements achieves high accuracy in automated echocardiographic quantification and potential for assisting the clinicians in the echocardiography workflow. This open-source model provides the foundation for future developments in artificial intelligence applied to echocardiography. (JACC. 2025;86:964–978)

Listen to this manuscript's audio summary by Editor-in-Chief Dr Harlan M. Krumholz on www.jacc.org/journal/jacc.

From the [a]Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA; [b]Division of Cardiology, Department of Medicine, Stanford University, Palo Alto, California, USA; [c]School of Medicine, University of California, Los Angeles, California, USA; [d]Department of Bioengineering, University of California Los Angeles, Los Angeles, California, USA; [e]Division of Research, Kaiser Permanente Northern California, Pleasanton, California, USA; [f]School of Medicine, Brown University, Providence, Rhode Island, USA; [g]Department of Computer Science, Stanford University, Stanford, California, USA; [h]Department of Biomedical Data Science, Stanford University, Stanford, California, USA; [i]Department of Electrical Engineering, Stanford University, Stanford, California, USA; and the [j]Department of Cardiology, Kaiser Permanente Santa Clara Medical Center, Santa Clara, California, USA.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received April 24, 2025; revised manuscript received July 6, 2025, accepted July 15, 2025.

ISSN 0735-1097 https://doi.org/10.1016/j.jacc.2025.07.053

Accurate assessment and diagnosis of cardiovascular disease requires detailed and reproducible assessment of cardiac dimensions and function. Echocardiography, the most common and readily available modality of cardiac imaging, enables frequent and rapid evaluation and surveillance of disease. However, echocardiographic measurements require detailed manual measurements and even expert evaluation has considerable variability.[1,2] Given the large number of relevant and important echocardiographic parameters, accurate echocardiographic measurement requires a highly time-consuming effort and places a significant burden on the clinical practice of echocardiography.[3]

SEE PAGE 979

The advent of deep learning models with computer vision has enabled fully automated image analysis and processing across a wide range of medical imaging.[4-11] Isolated task-specific annotations have already been proposed in echocardiography,[12,13] including for both wall thickness[14] and left ventricular ejection fraction (LVEF)[15,16]; however, there are many additional measurements that have been tackled with deep learning models. Furthermore, commercially available models are not open source, thus limiting the ability to build on prior developments.

Automation through artificial intelligence (AI) has the potential for more precision as well as more rapid turnaround.[8,11,12,14,17-19] In this study, we aimed to develop and validate a deep learning package (EchoNet-Measurements) for automated measurement of a wide range of echocardiographic parameters and measurements. Using a pipeline of view classification and automated segmentation, EchoNet-Measurements was trained on a large data set of annotated echocardiogram

ABBREVIATIONS AND ACRONYMS

AI = artificial intelligence

CP = coverage probability

CSMC = Cedars-Sinai Medical Center

DICOM = Digital Imaging and Communications in Medicine

LVEF = left ventricular ejection fraction

LVOT Vmax = left ventricular outflow tract maximum velocity

MAE = mean absolute error

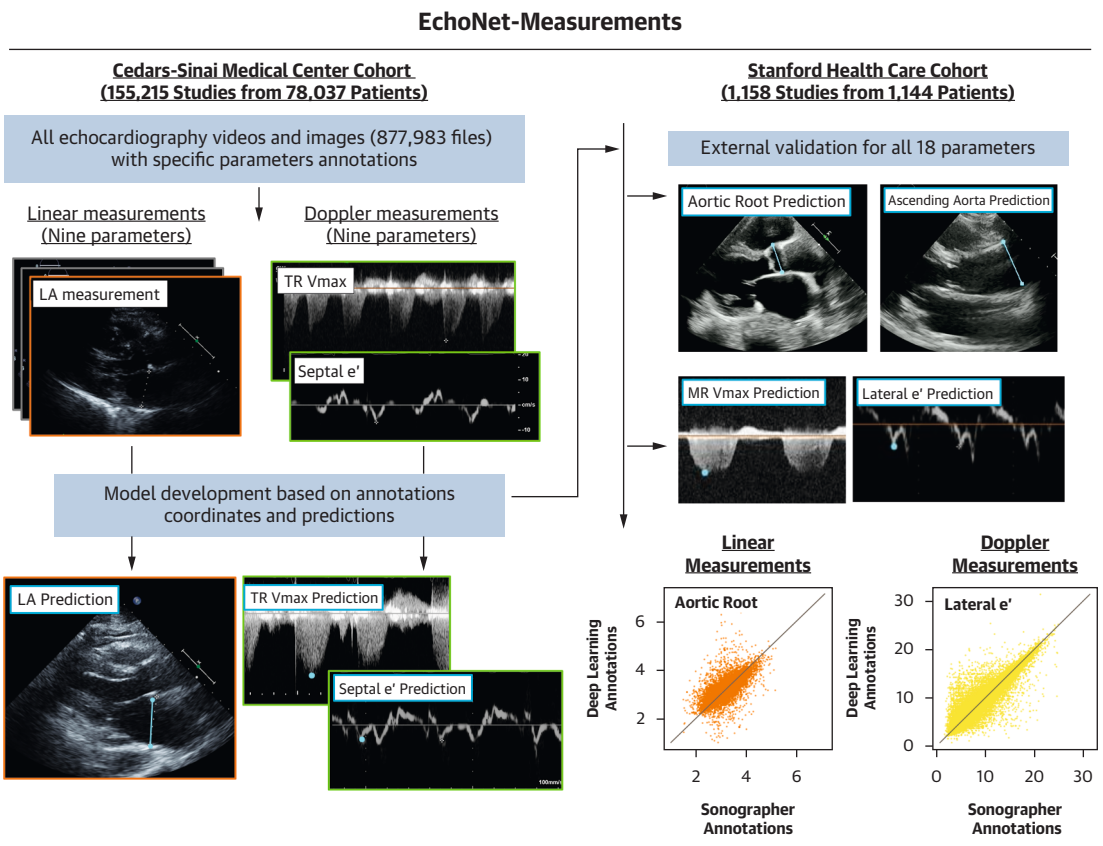
SHC = Stanford Healthcare

TAPSE = tricuspid annular plane systolic excursion

TR Vmax = tricuspid regurgitation maximum velocity

TTE = transthoracic echocardiography

FIGURE 1 Overview of the Study Pipeline



The pipeline of the developed automated echocardiographic parameters measurement model includes 2 main groups: linear measurements (eg, left atrium diameter and intraventricular septum) and Doppler measurements (eg, tricuspid regurgitation peak velocity and septal e' velocity). Evaluation of EchoNet-Measurements was performed on held-out test cohorts (Cedars-Sinai Medical Center) and external data set (Stanford Healthcare), and the model demonstrated accuracy comparable to sonographer annotations. LA = left atrium; MR Vmax = mitral regurgitation maximum velocity; TR Vmax = tricuspid regurgitation maximum velocity.

TABLE 1 Participant Characteristics

	Derivation Cohort (CSMC) (n = 78,037)	Temporal Split Cohort (CSMC) (n = 2,039)	External Validation Cohort (SHC) (n = 1,158)
Age, y	65.0 ± 17.4	66.2 ± 17.7	63.4 ± 18.8
BMI, kg/m^2	27.2 ± 6.4	26.4 ± 5.5	24.9 ± 9.0
Female	35,829 (46.2)	921 (45.2)	567 (49.0)
Unknown	555	1	15
Race/ethnicity			
Asian	5,638 (7.2)	201 (9.9)	193 (16.7)
Black or African American	10,593 (13.6)	279 (13.7)	65 (5.6)
Other	8,427 (10.8)	260 (12.7)	264 (22.8)
White	53,379 (68.4)	1,299 (63.7)	636 (54.9)
Hypertension	38,122 (48.9)	794 (38.9)	581 (50.2)
Dyslipidemia	36,779 (47.1)	883 (43.3)	575 (49.7)
Diabetes	32,583 (41.8)	605 (29.7)	328 (28.3)
Atrial fibrillation	29,324 (37.6)	480 (23.5)	352 (30.4)
Heart failure	29,114 (37.3)	526 (25.8)	379 (32.7)
LVEF, %	58.0 ± 14.6	57.2 ± 13.0	58.1 ± 12.7
Unknown	4,585		

Values are mean ± SD, n, or n (%).
BMI = body mass index; CSMC = Cedars-Sinai Medical Center; LVEF = left ventricular ejection fraction; SHC = Stanford Healthcare.

studies to enable automated comprehensive AI measurements.

METHODS

DATA CURATION. For training and a held-out test data set, we identified all resting transthoracic echocardiography (TTE) at Cedars-Sinai Medical Center (CSMC) between April 2009 and June 2022 and identified sonographer and cardiologist annotations. We included 155,215 studies from 78,037 patients who received a clinical transthoracic echocardiogram (Figure 1). There are no exclusion criteria based on patient characteristics or machine vendors. A total of 18 key TTE parameters were used for model training. Imaging data are saved in Digital Imaging and Communications in Medicine (DICOM) format and annotated by expert sonographers and cardiologists as part of routine practice.

Important linear measurements from B-mode echocardiographic images and videos included intraventricular septum, left ventricular internal diameter, left ventricular posterior wall diameter, left atrium diameter, right ventricular basal diameter, ascending aorta diameter, aortic root diameter, pulmonary artery diameter, and inferior vena cava. Important Doppler measurements included tricuspid regurgitation maximal velocity (TR Vmax), aortic valve maximal velocity, mitral valve regurgitation maximal velocity (MR Vmax), left ventricular outflow tract maximum velocity (LVOT Vmax), septal e', lateral e', mitral valve peak E velocity, E/A, and tricuspid annular plane systolic excursion (TAPSE).

In addition, EchoNet-Measurements was evaluated on 1,158 studies from 1,144 patients from Stanford Healthcare (SHC) who underwent echocardiography between January 2013 and August 2018 for external validation. Furthermore, to assess the end-to-end processing for full TTE studies, 2,103 studies from 2,039 patients from June 2022 to November 2024 at CSMC were used as a temporal split validation data set. In the end-to-end approach, automated parameter measurements corresponding to each view were performed after the view classification. If no images of a relevant view are present, that particular measurement is not performed. All echocardiography studies in the CSMC development cohort data set and the SHC cohort were performed using Philips EPIQ 7 or iE33 ultrasound machines. For both analyses, low-quality images were removed using the quality-control model. For each image-level and study-level analysis, image-level analysis refers to a comparison of the AI model directly with the sonographer measurement in the same image. In addition, we performed end-to-end processing where the AI model chose the view, video, and timing in a study and compared performance with the study-level clinician evaluation. As there are multiple videos of the same view, study-level analysis has more variation but more closely represents comparison with clinician intraobserver variability.[20] All echocardiography studies in the temporal split data set at CSMC were conducted using Philips EPIQ 5C, 7C, iE 33, or EPIQ CVx ultrasound systems. To confirm the robustness of the model across other vendors and different media, we performed inference using a subsample of publicly available MIMIC-Echo data sets[21] (GE HealthCare, DICOM), video screenshots of echocardiography machine monitors (Konica Minolta, AVI), stress echocardiography data (Vivid E95, GE HealthCare), and educational YouTube videos (GE HealthCare, AVI). Approval for this study was obtained from the CSMC and SHC Institutional Review Boards, and the requirement for informed consent was waived for retrospective data analysis without patient contact.

MODEL DEVELOPMENT. In preparing the data set for linear measurements, we used images that contained sonographer-annotated coordinates for measurements. These images then underwent automated

TABLE 2 Imaging Characteristics

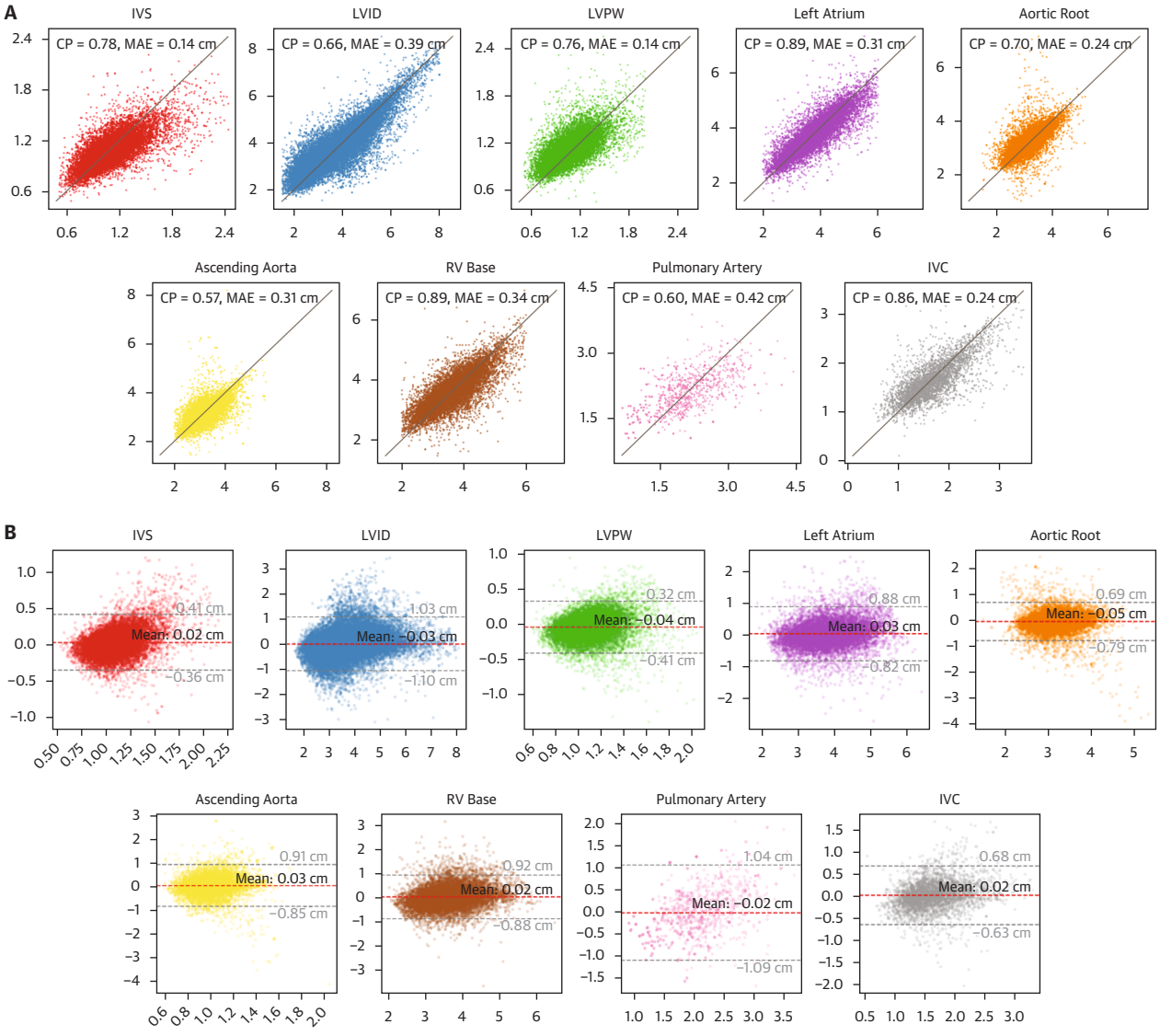
	CSMC Training and Validation Cohort			SHC External Validation Cohort		
	Number of Images	Number of Studies	Number of Patients	Number of Images	Number of Studies	Number of Patients
Linear measurements						
IVS	97,685	94,311	55,284	96	96	94
LVID	224,042	131,873	71,321	129	129	129
LVPW	120,093	116,276	65,394	55	55	54
Left atrium	89,779	89,170	56,805	35	35	35
Ascending aorta	67,034	62,018	37,952	77	77	77
Aortic root	97,944	93,872	57,739	32	32	32
RV base	92,729	90,389	53,566	60	60	60
Pulmonary artery	8,960	6,875	5,723	13	13	13
IVC	32,288	31,815	21,171	67	67	67
Doppler measurements						
TR Vmax	199,243	79,897	47,584	425	425	423
AV Vmax	34,526	30,403	22,067	320	320	320
MR Vmax	16,819	12,417	10,301	89	89	89
LVOT Vmax	14,314	14,035	9,552	304	304	304
Lateral e'	96,051	95,578	58,495	559	559	559
Septal e'	43,462	43,343	33,327	640	640	639
Peak E velocity	62,698	62,331	43,060	71	71	71
E/A	62,698	62,331	43,060	54	54	54
TAPSE	54,344	53,802	36,189	254	254	254

AV Vmax = aortic valve maximum velocity; E/A = Ratio of early diastolic to late diastolic mitral inflow velocities; IVC = inferior vena cava; IVS = intraventricular septum; Lateral e' = lateral mitral annulus e' velocity; LVID = left ventricular internal diameter; LVOT Vmax = left ventricular outflow tract maximum velocity; LVPW = left ventricular posterior wall; MR Vmax = mitral regurgitation maximum velocity; peak E velocity = early diastolic mitral inflow velocity; RV Base = right ventricular basal diameter; Septal e' = septal mitral annulus e' velocity; TAPSE = tricuspid annular plane systolic excursion; TR Vmax = tricuspid regurgitation maximum velocity; other abbreviations as in Table 1.

preprocessing steps, which included removing patient identification details, text information, and electrocardiogram/respirometer tracings. Images were resized to 480 (height) by 640 (width) before training for the linear measurement group and cropped to 426 (height) by 1,024 (width) for the Doppler measurement group based on the DICOM Tag that indicated the region of Doppler analysis. For EchoNet-Measurements training and testing, we used a DeepLabv3 architecture[22] and trained with a binary cross-entropy with logits loss. The CSMC data were split by patient in a ratio of 8:1:1 for model training, validation, and held-out testing to avoid data leakage across data split. Task-specific models were trained for each measurement. For all models, an Adam optimizer with a learning rate of 0.001 was used, and the model was trained for up to 100 epochs with early stopping after 10 epochs based on the validation loss and a batch size of 24. The weights from the epoch with the minimum loss were used for the held-out test data set and external validation. All performance metrics were analyzed using data from held-out data that were not involved in model training. An image quality classifier was also developed to identify low-quality images and videos (examples in Supplemental Figure 1) for post hoc analysis in the held-out data set in CSMC and these were excluded from downstream analysis in the end-to-end analysis. The image quality-control model was developed with a video-based convolutional neural network (R2+1D)[15] to classify low-quality videos and an image-based model (DenseNet)[23] to classify low-quality Doppler images[24] or exclude inaccurate tissue Doppler annotations (remove a' annotation when e' is desired). (Supplemental Methods).

ANALYSIS OF CASES WITH HIGH DISCREPANCY BETWEEN AI AND CLINICIANS. From the 90th percentile or higher of mean absolute error (MAE) between deep learning measurements and sonographer measurements, a total of 180 subsample data (randomly picked 10 images from each measurement variable) from the CSMC data set were manually reviewed by 2 board-certified cardiologists (Y.S. and C.B.-R). This analysis was conducted to identify the causes of significant differences between the deep learning model measurements and those of the sonographers. We identified if any of the following criteria applied: 1) the sonographer measurement (ground truth) was preferred over automatic measurements by the deep learning model; 2) the

FIGURE 2 Model Performance Between Deep Learning Model and Sonographer Annotations for Echocardiographic Measurements in the CSMC Test Data Set (Image-Level Analysis)

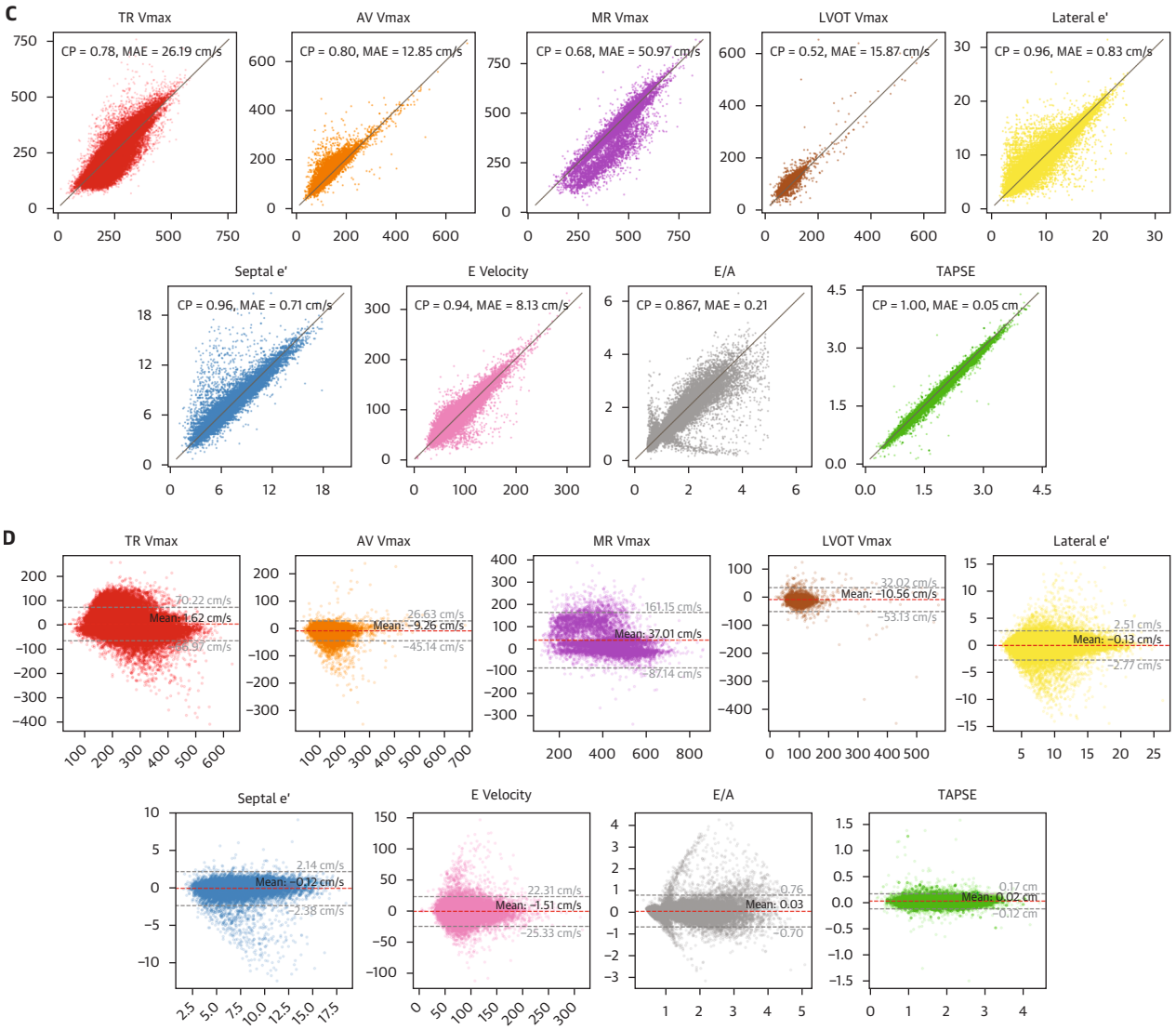


(A and C) Scatter plot comparing deep learning model predictions with sonographer annotations for 9 linear echocardiographic parameters (A) and Doppler echocardiography parameters (C). (B and D) Bland-Altman plots for each parameter in the linear measurement group (B) and Doppler echocardiography parameters (D), displaying the difference between model predictions and sonographer measurements (y-axis) against the mean of the 2 measurements (x-axis). AV Vmax = aortic valve maximum velocity; CP = coverage probability; E/A = ratio of early diastolic to late diastolic mitral inflow velocities; E velocity = Early diastolic mitral inflow velocity; IVC = inferior vena cava; IVS = Intraventricular septum; LA = left atrium; Lateral e' = Lateral mitral annulus e' velocity; LVID = left ventricular internal diameter; LVOT Vmax = left ventricular outflow tract maximum velocity; LVPW = left ventricular posterior wall; MAE = Mean absolute error; PA = pulmonary artery; RV Base: = right ventricular basal diameter; Septal e' = septal mitral annulus e' velocity; TAPSE = tricuspid annular plane systolic excursion; other abbreviations as in Figure 1.

Continued on the next page

automatic measurements by the deep learning model were preferred over sonographer measurements; 3) images that both sonographer measurement and deep learning model predictions were within a clinically acceptable range but with a measurement variability; and 4) predictions on severely low-quality images. For comparison, we also performed a similar analysis on cases in which the discrepancy between AI and sonographer measurements was moderate (from the 50th to 90th percentile of

FIGURE 2 **Continued**



MAE between AI measurements and sonographer measurements).

STATISTICAL ANALYSIS. Coverage probability (CP), mean error, relative difference, MAE, coefficient of determination (R^2), and intraclass correlation coefficients (ICCs 2,1) between clinically measured parameters and parameters predicted by the developed deep learning model were calculated in the held-out data set and the external data set. CP, the chance that 2 measurements of the same parameter differed by less than a predefined acceptable difference, was calculated by the proportion of absolute paired differences that fall within a predefined delta.[25,26] The predefined delta (acceptable difference) was defined based on previous literature on intraobserver variability in echocardiographic measurements[20] (Supplemental Table 1). Supplemental Figures 2 and 3 show the coverage probability curve in linear measurements and Doppler measurements, and Supplemental Figure 4 demonstrates the upper limit of acceptable variation for coverage probability. Bland-Altman plots in which the average of the 2 measurements was plotted against the difference were used to check the agreement between the actual and predicted measurements. Stratified analyses

TABLE 3 EchoNet-Measurements Performance in Internal and External Test Data Sets (Image-Level Comparison)

	CSMC				SHC			
	Sonographer Measurement	DL Measurement	Coverage Probability	MAE	Sonographer Measurement	DL Measurement	Coverage Probability	MAE
Linear measurements								
IVS, cm	1.13 ± 0.27	1.1 ± 0.2	0.777 (0.77-0.79)	0.143 (0.14-0.15)	1.02 ± 0.21 c	0.96 ± 0.17	0.906 (0.84-0.96)	0.106 (0.09-0.12)
LVID, cm	3.71 ± 1.12	3.74 ± 0.99	0.663 (0.66-0.67)	0.386 (0.38-0.39)	4.69 ± 0.78	4.77 ± 0.73	0.853 (0.79-0.91)	0.210 (0.17-0.25)
LVPW, cm	1.09 ± 0.23	1.13 ± 0.19	0.762 (0.75-0.77)	0.142 (0.14-0.14)	1.01 ± 0.21	1.02 ± 0.17	0.927 (0.85-0.98)	0.093 (0.08-0.11)
Left atrium, cm	3.84 ± 0.78	3.81 ± 0.71	0.894 (0.89-0.9)	0.314 (0.31-0.32)	4.06 ± 0.72	3.79 ± 0.76 c	0.771 (0.63-0.91)	0.410 (0.30-0.53)
Aorta, cm	3.24 ± 0.53	3.21 ± 0.5	0.573 (0.56-0.58)	0.308 (0.30-0.32)	3.5 ± 0.47	3.3 ± 0.48	0.532 (0.42-0.65)	0.281 (0.24-0.33)
Aortic root, cm	3.18 ± 0.46	3.23 ± 0.47	0.697 (0.69-0.71)	0.245 (0.24-0.25)	3.25 ± 0.47 m	3.02 ± 0.44	0.5 (0.31-0.66)	0.286 (0.22-0.36)
RV base, cm	3.65 ± 0.71	3.64 ± 0.62	0.894 (0.89-0.9)	0.340 (0.33-0.35)	3.5 ± 0.6	3.54 ± 0.57	0.95 (0.88-1.0)	0.184 (0.14-0.23)
Pulmonary artery, cm	2.05 ± 0.68	2.08 ± 0.47	0.6 (0.57-0.63)	0.425 (0.40-0.45)	1.76 ± 0.5	1.9 ± 0.4	0.896 (0.82-0.97)	0.185 m (0.15-0.23)
IVC, cm	1.68 ± 0.49	1.66 ± 0.41	0.857 (0.84-0.87)	0.244 (0.24-0.25)	3.31 ± 0.78	2.92 ± 0.64	0.615 (0.31-0.85)	0.489 (0.33-0.66)
Doppler measurements								
TR Vmax, cm/s	249.74 ± 58.04	248.16 ± 67.97	0.775 (0.77-0.78)	26.18 (26.05-26.32)	258.41 ± 46.07	252.92 ± 48.74	0.88 (0.85-0.91)	15.39 (14.00-16.80)
AV Vmax, cm/s	130.59 ± 37.94	139.84 ± 40.34	0.805 (0.8-0.81)	12.85 (12.65-13.06)	170.62 ± 67.04 cm/s	166.31 ± 63.52 cm/s	0.862 (0.82-0.9)	10.41 (9.14-11.84)
MR Vmax, cm/s	433.65 ± 109.06	396.65 ± 131.57	0.679 (0.67-0.69)	50.97 (49.85-52.13)	492.54 ± 50.08 cm/s	477.58 ± 64.62 cm/s	0.809 (0.72-0.89)	29.92 (24.45-36.16)
LVOT Vmax, cm/s	99.35 ± 40.76	109.74 ± 44.11	0.52 (0.5-0.54)	15.84 (15.23-16.49)	101.22 ± 25.56	104.91 ± 27.67	0.855 (0.82-0.89)	7.89 (7.01-8.82)
Lateral e', cm/s	9.21 ± 3.33	9.34 ± 3.22	0.956 (0.95-0.96)	0.834 (0.82-0.84)	9.3 ± 3.32	9.57 ± 3.26	0.891 (0.86-0.92)	1.17 (1.05-1.29)
Septal e', cm/s	7.24 ± 2.6	7.36 ± 2.56	0.958 (0.95-0.96)	0.709 (0.70-0.72)	6.94 ± 2.51	7.75 ± 2.41	0.805 (0.77-0.83)	1.23 (1.12-1.33)
E velocity, cm/s	86.33 ± 29.53	87.83 ± 29.72	0.944 (0.94-0.95)	8.13 (8.02-8.24)	84.35 ± 26.73	77.71 ± 27.39	0.944 (0.89-0.99)	6.80 (4.89-9.24)
E/A	1.44 ± 0.79	1.41 ± 0.75	0.867 (0.86-0.87)	0.212 (0.21-0.22)	1.19 ± 0.79	1.12 ± 0.5	0.963 (0.91-1.0)	0.22 (0.09-0.39)
TAPSE, cm	1.91 ± 0.55	1.89 ± 0.54	0.999 (0.999-0.999)	0.051 (0.05-0.05)	2.28 ± 0.5	2.2 ± 0.55	0.949 (0.92-0.97)	0.148 (0.11-0.19)

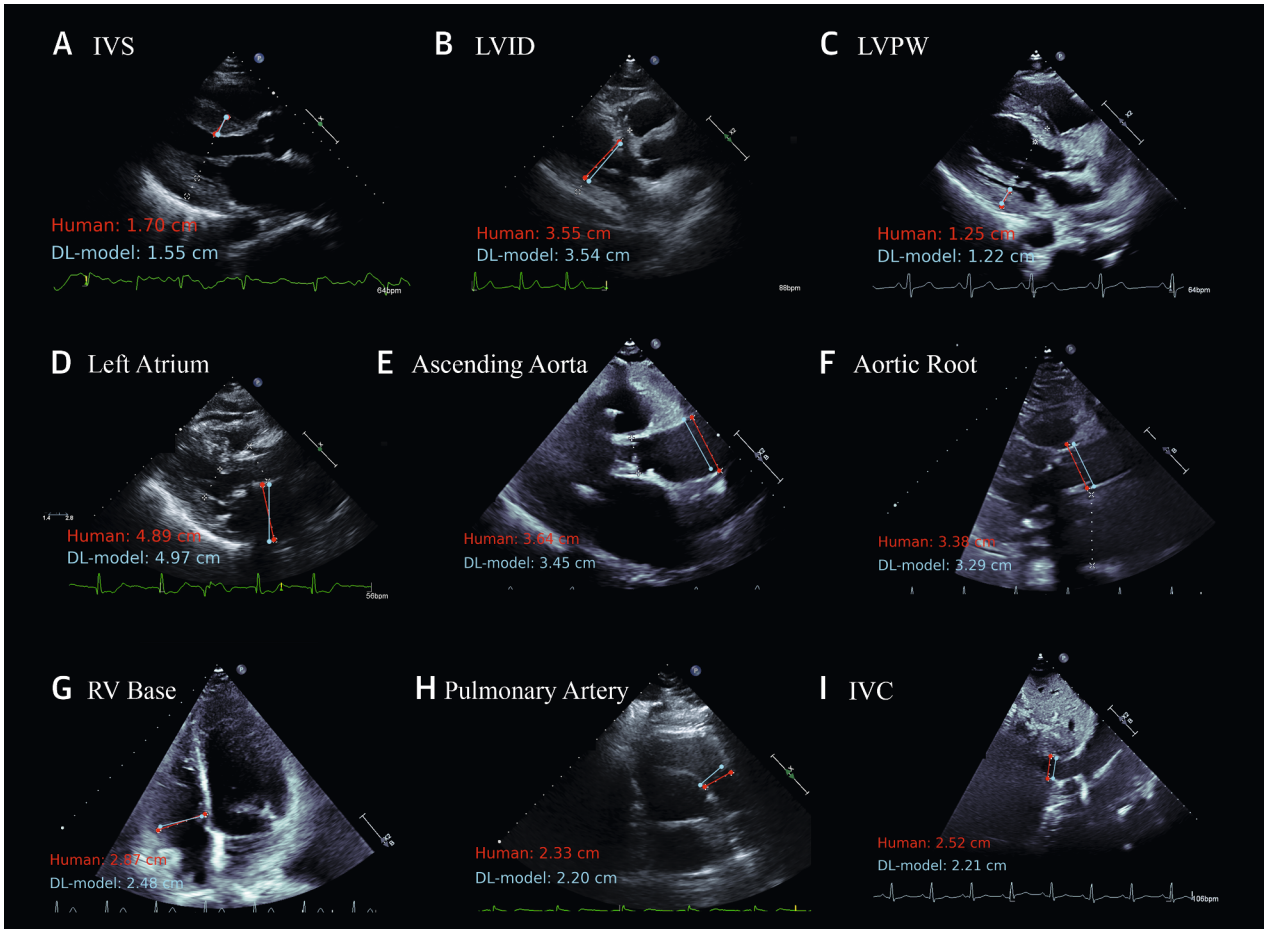
All metrics values are described with 95% CI.
ICC = intraclass correlation coefficient; MAE = mean absolute error; other abbreviations as in Tables 1 and 2.

were performed based on image quality defined by the developed quality-control model, machine vendor, and the following patient characteristics: age (<65 or ≥65 years), sex, history of atrial fibrillation, body mass index (<35 or ≥35 kg/m^2), and LVEF subgroup (<40% or ≥40%). To account for sicker patients potentially undergoing more frequent examinations and detailed measurements, we randomly selected only one measurement per patient if they had multiple exams or measurement for the CSMC held-out data set. Additionally, if a specific measurement was taken multiple times within that chosen exam, we randomly selected only one measurement for the additional analysis. All 95% CIs were calculated with 10,000 bootstrapping samples. Data analysis was performed using both Python version 3.10.12 and R version 4.2.2 programming languages. This study was carried out following the CONSORT-AI (Consolidated Standards of Reporting Trials-Artificial Intelligence) guideline.[27]

RESULTS

MODEL DEVELOPMENT AND VALIDATION COHORTS. A total of 877,983 annotated measurements were identified from 155,215 CSMC echocardiography studies from 78,037 patients. The patient characteristics in the CSMC derivation cohort exhibited similar patient age, female proportion, race, mean LVEF, and comorbidities across data sets (**Table 1**, Supplemental Tables 2

FIGURE 3 **Representative Figures of Comparison Between Sonographer and Deep Learning Model Measurements for Echocardiographic Linear Measurement Parameters**



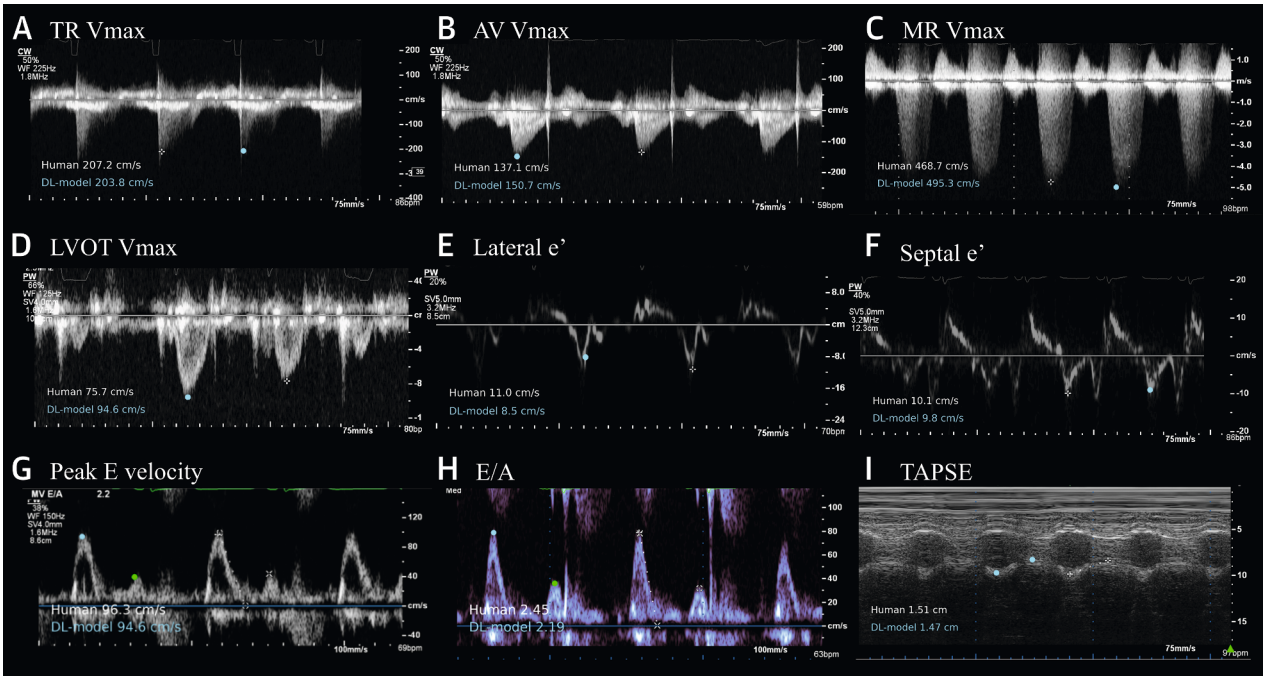
The figure shows a comparison of measurements made by a sonographer (in red) and predictions from the deep learning (DL) model (in light blue) across 9 echocardiographic parameters (A to I). Abbreviations as in Figure 2.

to 4). The number of training and validation examples for each parameter are summarized in Table 2.

ASSESSMENT OF EchoNet-MEASUREMENTS PERFORMANCE. In the CSMC test data set, the EchoNet-Measurements model demonstrated a strong agreement between automated deep learning measurements and sonographer measurements in the held-out data set (metrics ranging from a CP = 0.573 [0.56-0.58], MAE: 0.425 cm [0.404-0.446]) for ascending aorta diameter to CP = 0.894 [0.89-0.9], MAE: 0.314 [0.31-0.32]) for left atrium diameter (Figure 2). Similarly for the Doppler measurement images, the model demonstrated a high agreement between automated deep learning measurements and sonographer measurement, ranging from CP = 0.52 (0.5-0.54), MAE: 15.84 cm/2 (15.2-16.5) for LVOT Vmax to CP = 0.999 (0.999-0.999), MAE: 0.051 cm (0.050-0.052) for TAPSE (Figure 2). Detailed information including comparison between mean value of sonographer measurements and deep learning-based measurements, CP, MAE, R^2, ICC, mean relative difference, and mean error in the held-out data set are described in Table 3 and Supplemental Table 5.

Representative images comparing sonographer annotations and deep learning annotations are shown in Figures 3 and 4 for linear and Doppler

FIGURE 4 Representative Figures of Comparison Between Sonographer and Deep Learning Model Measurements for Echocardiographic Doppler Measurement Parameters



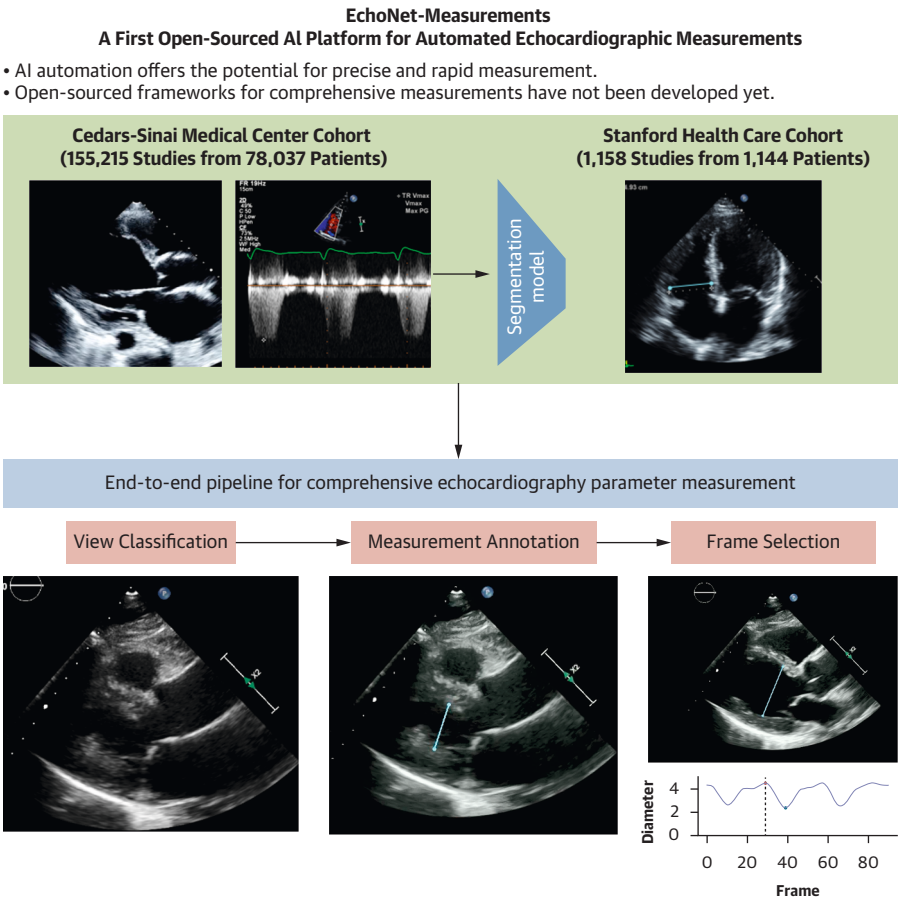
Comparison of measurements made by a sonographer (in white) and predictions from the DL model (in light blue) across 9 echocardiographic Doppler and M-mode parameters (TAPSE) (A to I). For peak E velocity and E/A, deep learning–based annotation on peak E velocity is shown in light blue dot and peak A velocity is shown in green dot. Abbreviations as in **Figures 1 to 3**.

measurements, respectively. Representative generated videos in the linear measurement group are shown in Video 1. A demonstration video for the high-throughput analysis of TR Vmax using the input of multiple DICOM files is shown in Video 2. Image quality greatly affected the performance of automated measurements, with the high-quality image group consistently showing a significant improvement in CP, R^2, and MAE compared with low-quality echocardiography images (Supplemental Table 6, Supplemental Figures 5 and 6). The model performance was consistent across patient characteristics including sex, age, body mass index, LVEF subgroup, machine vendor, and atrial fibrillation status in all measurement parameters (Supplemental Tables 7 to 12). Similar findings were observed in a subgroup analysis in which only 1 measurement from a single study per single patient was performed (Supplemental Table 13).

On the temporal split data set from CSMC (June 2022-November 2024), the model showed robust predictive accuracy (eg, CP of 0.866, R^2 of 0.725, and MAE of 16.09 cm/s for TR Vmax, and CP of 0.871, R^2 of 0.74, and a MAE of 0.204 cm for inferior vena cava diameter) (**Figure 5**, Supplemental Table 2). In the external SHC data set with image-level ground truth measurement value, the performance of EchoNet-Measurements was robust (**Figure 5**, Supplemental Table 3) in the SHC data set. MAE for linear measurements ranged from 0.106 cm and CP of 0.906 for intraventricular septum diameter to MAE of 0.489 cm and CP of 0.615 for pulmonary artery diameter. The model's performance was also consistent and robust across all parameters for Doppler measurements (eg, MAE of 0.148 cm and CP of 0.949 in TAPSE and MAE of 7.89 cm/s and CP of 0.855 in LVOT Vmax). The model performance was similarly robust in an external data set with study-level ground truth label for both linear measurement and all Doppler parameters (Supplemental Table 4, Supplemental Figure 7).

ANALYSIS OF HIGH MEASUREMENT DISCREPANCIES CASES. Among echocardiography images with absolute differences between sonographer measurement and deep learning measurement in the 90th percentile or higher, a subset of 180 images (10 randomly

CENTRAL ILLUSTRATION Overview of EchoNet-Measurements



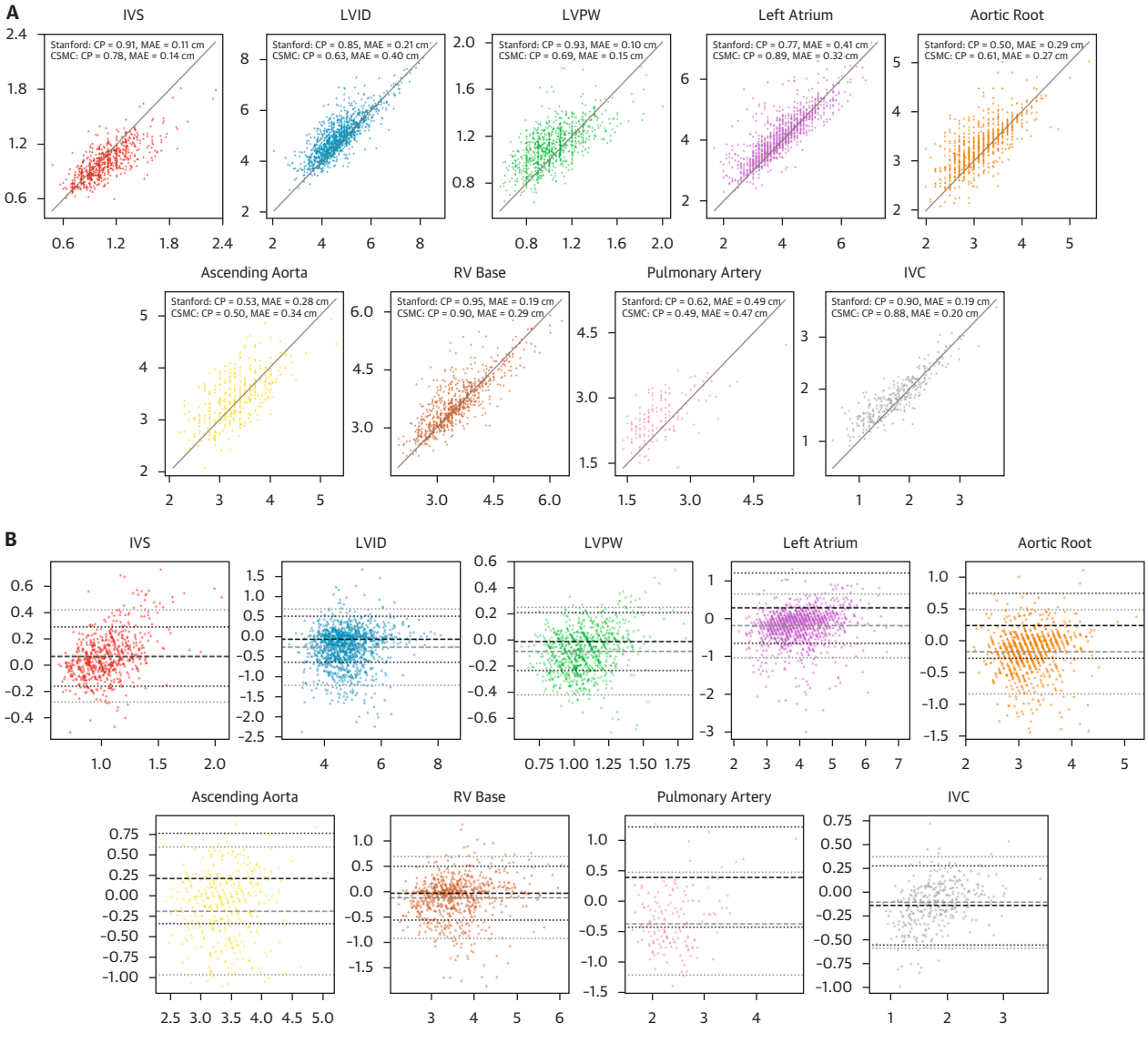
High accuracy in automated echocardiographic measurements compared with clinicians.

Sahashi Y, et al. JACC. 2025;86(13):964–978.

EchoNet-Measurements, an automated echocardiographic parameter measurement model, includes 2 main groups: linear measurements for 9 key parameters (eg, left atrium diameter and interventricular septum) and Doppler measurements for 9 key parameters (eg, tricuspid regurgitation peak velocity and septal e' velocity). The model demonstrated accuracy comparable to sonographer annotations (top). An end-to-end pipeline for comprehensive echocardiography parameter measurement was designed for the automated measurement of echocardiographic parameters, which includes: 1) classifying views using artificial intelligence (AI) models; 2) executing corresponding measurements for each identified view using EchoNet-Measurements; and 3) extracting optimal values from videos based on cardiac cycle estimation, when applicable (bottom).

selected images from each measurement variable) was manually reviewed. Of a total of 180 images, 91 (50.6%) were categorized as the cases in which sonographer annotations were preferred over automatic measurements by the deep learning model, 20 images (11.1%) were classified as images in which the deep learning model was preferred over sonographer measurements, and 54 images (30.0%) were categorized as images that both sonographer annotations and deep learning model predictions were clinically acceptable but demonstrated measurement variabilities. In the remaining 15 images (8.3%), echocardiography images were severely noisy or blurred, seen as inappropriate images for deep learning-based measurements. Representative images are described in Supplemental Figure 8 and the number of images in each category is described in Supplemental Table 14. In contrast, in 180 randomly selected images in which the MAE was between the 50th and 90th percentiles, both measurements by

FIGURE 5 End-to-End Image-level Model Performance Between Deep Learning Model and Sonographer Annotations for Echocardiographic Measurements in the SHC External Test Data Set



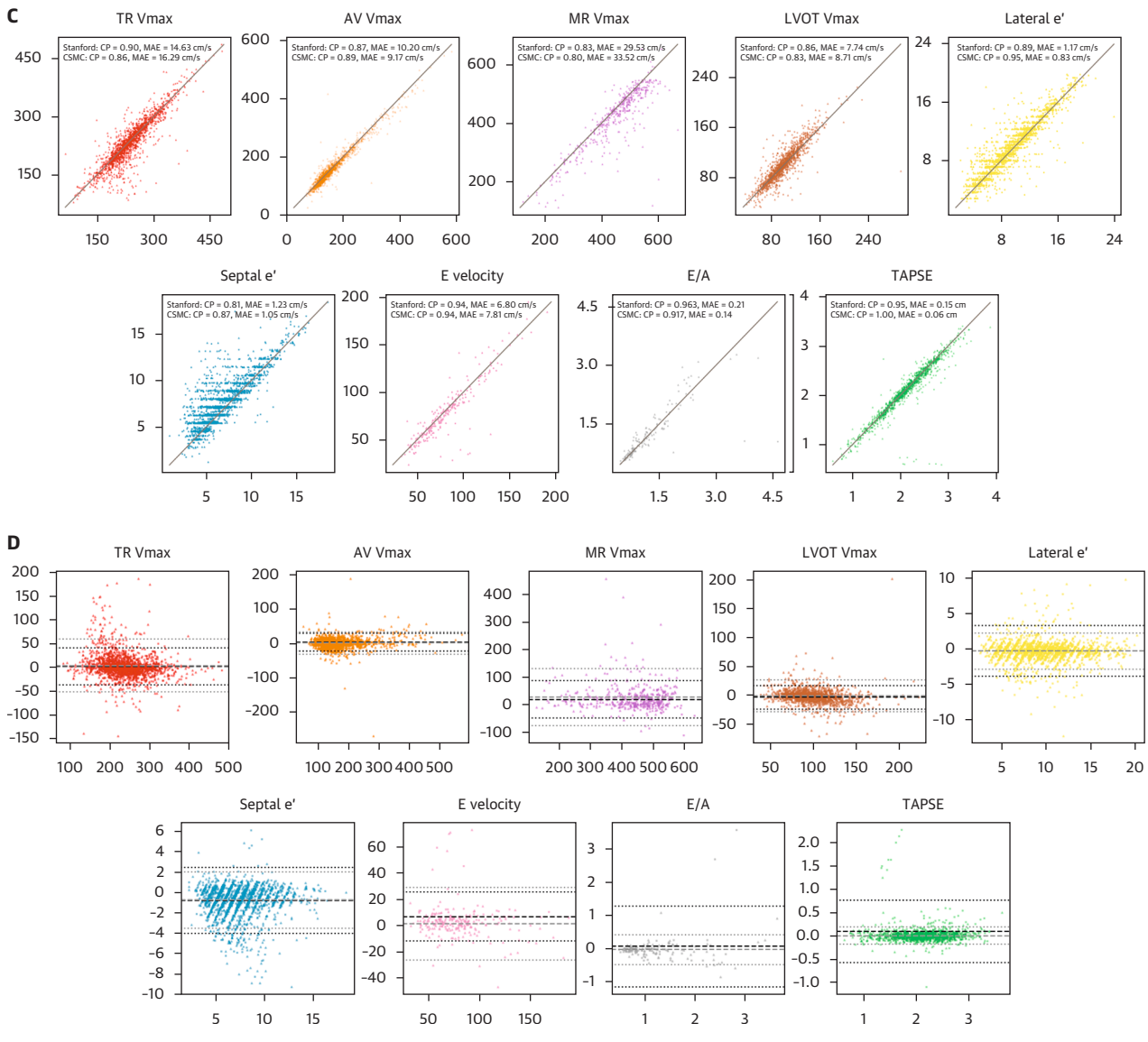
(A and C) Scatter plots comparing deep learning model predictions with sonographer annotations for (A) linear parameters and (C) Doppler echocardiography parameters. (B and D) Bland-Altman plots for each parameter in the linear measurement group (B) and Doppler echocardiography parameters (D). Data from Stanford Healthcare (SHC) are described as square dots and in the Cedars-Sinai Medical Center (CSMC) temporal split data are shown in triangle dots in all figures. Details of bias and limits of agreement are described in Supplemental Tables 2 and 3. Abbreviations as in **Figures 1 and 2**.

Continued on the next page

sonographers and AI model measurements were judged to be clinically reasonable in 123 of 180 cases (68.3%) (Supplemental Table 15).

EchoNet-Measurements performance was robust and generalizable on a wide range of echocardiography vendor formats, including on the MIMIC-Echo data set (GE Healthcare, DICOM), screenshots from Konica Minolta echocardiography screens (AVI), and screenshots from educational YouTube echocardiography videos (GE Healthcare, AVI) in Supplemental Figures 9 and 10. Because these data sets do not have released clinician measurements, statistical analysis was not conducted, but images demonstrate reasonable automated measurements.

FIGURE 5 **Continued**



DISCUSSION

In this study, we developed EchoNet-Measurements, a deep learning model for the automatic measurement of 18 echocardiographic parameters, and showed the model performed well in multiple geographically and temporally distinct cohorts. Using the largest yet real-world data set of sonographer clinical annotations, this open-source model demonstrated favorable measurement accuracy comparable to that of expert sonographers. Further, we have released the code and weights of these models publicly, along with demo user interface, to facilitate further research and clinical deployment in the field.

Comprehensive measurement is needed because clinical echocardiography involves the measurement of many parameters that holistically assess patient status. Complementary to approaches that directly predict measurements or disease classes,[8,11] EchoNet-Measurements is a series of segmentation models that directly measure key clinical findings

that result in downstream clinical interpretations. For example, diagnosis for heart failure with preserved ejection fraction requires a combination of multiple echocardiographic parameters[28,29]—including left atrial volume index, lateral e', septal e', E/A ratio, right ventricular systolic pressure, and TR velocity. In addition, by annotating relevant images with specific regions of interest rather than only showing the final numerical measurements, EchoNet-Measurements also has the added benefit of being more interpretable and adjustable by clinicians. One of the primary benefits of measurement automation is the reduction of examination time and variability between sonographers, leading to a more consistent result.[30,31] Echocardiography includes multiple steps, including image acquisition, measurement, assessment, and report generation. The present study and prior research[8,11,14,15] have demonstrated that some of these steps can be supported by deep learning models. However, the utility and accuracy of deep learning models that performed the entire echocardiography process have not yet been tested in clinical practice or prospective trials, which will be necessary for the clinical deployment of comprehensive echocardiography AI models.

A few considerations merit discussion in differences between sonographer and AI annotations. In cases of poor image quality (Supplemental Figure 8D), sonographers are more able to overcome noise and artifacts. Sonographers might also consider clinical context and surrounding images when annotating low-quality images, leading to over-annotation when evaluated on a single image, whereas the AI contextualizes and generates annotations from a single image at this time. In addition, there was higher agreement in Doppler measurements than linear measurements, likely attributable to fewer degrees of freedom (more frames to choose from and more subjectivity in annotation in linear measurements).

STUDY STRENGTHS AND LIMITATIONS. Previous studies have already focused on automating the measurement of limited echocardiographic parameters using deep learning segmentation models,[12,17,18] but we note that many prior models do not share code or model weights, limiting utility for reproducible research purposes and future development. Automated measurements derived from the EchoNet-Measurements model have indeed already been used for automated diastolic function evaluation showing similar accuracy across a wide range of systolic and diastolic functions.[32] As demonstrated in this study, EchoNet-Measurements has strong accuracy both compared directly with the manually selected images sonographers annotated as well as end-to-end study-level assessment with automated choice of videos, frames, and views (**Central Illustration**). EchoNet-Measurements is trained on data from only 1 medical center, which might capture biases and quirks associated with the data set; however, we demonstrate favorable results compared with previous deep learning models.[14,18,19,33] Furthermore, one of the strengths of our model is the size of the training data, which is the largest data set yet of sonographer annotations. In multiple prior works both within medicine[34] and outside,[35] the scale of training data has a strong impact on downstream model performance. Another feature of our work is the public availability of our model for the end-to-end approach and demo interface that enables the creation of clinical data sets, supports reproducible clinical research and clinical deployments, and allows for the integration of EchoNet-Measurements models into multitasking AI systems for echocardiographic interpretation.[8,11]

One of the limitations is that our current work does not yet include automated measurements for all parameters assessed in clinical echocardiography (eg, velocity time integral, proximal isovelocity surface area, and left atrium volume). Automated measurement of these parameters would enable application to a wide range of echocardiography tasks, including aortic valve area calculation for aortic stenosis severity assessment. Future studies are warranted to develop these important clinical measurements. Other limitations include that the model could not be compared with other AI-based measurement models due to limited availability and the lack of a benchmark data set, and that the cardiologists who performed the evaluations in the analysis of measurement discrepancies cases were not blinded. Benchmark data set and prospective randomized controlled trials, similar to a previous study,[34] would be necessary.

CONCLUSIONS

In this study, we developed and validated EchoNet-Measurements, an open-source deep learning framework for the automated assessment of a wide range of echocardiographic parameters. Trained on the largest data set to date of clinical annotations, the

model demonstrated robust performance across both linear and Doppler measurements.

DATA SHARING STATEMENT

The code, model weights, and demonstration video are available at https://github.com/echonet/measurements/. The patient data are not publicly available due to their potentially identifiable nature.

ACKNOWLEDGMENTS The authors appreciate Dr Kazuharu Kawano for providing the screenshot of Konica Minolta echocardiography videos.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This work is funded by National Institutes of Health, National Heart, Lung, and Blood Institute grants R00HL157421, R01HL173526, and R01HL173487 to Dr Ouyang. No funders had a role in the design/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. Dr Ouyang has received support from Alexion; and has received consulting or honoraria for lectures from EchoIQ, Ultromics, Pfizer, InVision, the Korean Society of Echo, and the Japanese Society of Echo. Dr Sahashi has received honoraria for consulting from m3.com Inc and InVision. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr David Ouyang, Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, 4480 Hacienda Drive, Pleasanton, California 94588, USA. E-mail: David.ouyang@kp.org. X handle: @David_Ouyang.

REFERENCES

1. Anderson DR, Blissett S, O'Sullivan P, Qasim A. Differences in echocardiography interpretation techniques among trainees and expert readers. *J Echocardiogr*. 2021;19(4):222–231.

2. Nagata Y, Kado Y, Onoue T, et al. Impact of image quality on reliability of the measurements of left ventricular systolic function and global longitudinal strain in 2D echocardiography. *Echo Res Pract*. 2018;5(1):27–39.

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 (US); 2014.

4. Christensen M, Vukadinovic M, Yuan N, Ouyang D. Vision-language foundation model for echocardiogram interpretation. *Nat Med*. 2024;30(5):1481–1488.

5. Xu H, Usuyama N, Bagga J, et al. A whole-slide foundation model for digital pathology from real-world data. *Nature*. 2024;630(8015):181–188.

6. Zhang S, Xu Y, Usuyama N, et al. A multimodal biomedical foundation model trained from fifteen million image-text pairs. *NEJM AI*. Published online December 20, 2024. https://doi.org/10.1056/aioa2400640

7. Chia MA, Zhou Y, Keane PA. A new foundation model for multimodal ophthalmic images: Advancing disease detection and prediction. *NEJM AI*. 2024;1(12). https://doi.org/10.1056/aie2401024

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

9. Goto S, Solanki D, John JE, et al. Multinational federated learning approach to train ECG and echocardiogram models for hypertrophic cardiomyopathy detection. *Circulation*. 2022;146(10):755–769.

10. Oikonomou EK, Vaid A, Holste G, et al. Artificial intelligence-guided detection of under-recognised cardiomyopathies on point-of-care cardiac ultrasonography: a multicentre study. *Lancet Digit Health*. 2025;7(2):e113–e123.

11. Holste G, Oikonomou EK, Tokodi M, Kovács A, Wang Z, Khera R. Complete AI-enabled echocardiography interpretation with multitask deep learning. *JAMA*. 2025;34(4):306–318. https://doi.org/10.1001/jama.2025.8731

12. 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.

13. Kwan AC, Chang EW, Jain I, et al. Deep learning-derived myocardial strain. *JACC Cardiovasc Imaging*. 2024;17(7):715–725. https://doi.org/10.1016/j.jcmg.2024.01.011

14. 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.

15. 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.

16. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation*. 2018;138(16):1623–1635.

17. Howard JP, Stowell CC, Cole GD, et al. Automated left ventricular dimension assessment using artificial intelligence developed and validated by a UK-wide collaborative. *Circ Cardiovasc Imaging*. 2021;14(5):e011951.

18. Jevsikov J, Ng T, Lane ES, et al. Automated mitral inflow Doppler peak velocity measurement using deep learning. *Comput Biol Med*. 2024;171:108192.

19. 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.

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

21. Gow B, Pollard T, Greenbaum N, et al. MIMIC-IV-ECHO: echocardiogram matched subset. Published online July 21, 2023. https://doi.org/10.13026/EF48-V217.

22. Yurtkulu SC, Şahin YH, Unal G. Semantic segmentation with extended DeepLabv3 architecture. In: *2019 27th Signal Processing and Communications Applications Conference (SIU); April 24-26, 2019; Sivas, Turkey*. IEEE; 2019:1–4. https://doi.org/10.1109/siu.2019.8806244

23. Huang G, Liu Z, Weinberger KQ. *Densely connected convolutional networks*. Honolulu, HI: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit; July 21-26, 2017; Published online August 25, 2017:2261–2269. https://doi.org/10.1109/CVPR.2017.243

24. Sahashi Y, Vukadinovic M, Amrollahi F, et al. Opportunistic screening of chronic liver disease with deep-learning-enhanced echocardiography. *NEJM AI*. 2025;2(3). https://doi.org/10.1056/aioa2400948

25. Crowley AL, Yow E, Barnhart HX, et al. Critical review of current approaches for echocardiographic reproducibility and reliability assessment in clinical research. *J Am Soc Echocardiogr*. 2016;29(12):1144–1154.e7.

26. Barnhart HX, Yow E, Crowley AL, et al. Choice of agreement indices for assessing and improving measurement reproducibility in a core laboratory setting. *Stat Methods Med Res*. 2016;25(6):2939–2958.

27. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385. https://doi.org/10.1136/bmj-2023-078378

28. Amanai S, Harada T, Kagami K, et al. The H2FPEF and HFA-PEFF algorithms for predicting exercise intolerance and abnormal hemodynamics

in heart failure with preserved ejection fraction. *Sci Rep*. 2022;12(1):13.

29. Pieske B, Tschöpe C, de Boer RA, et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). *Eur Heart J*. 2019;40(40):3297–3317.

30. Lin A, Manral N, McElhinney P, et al. Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. *Lancet Digit Health*. 2022;4(4):e256–e265.

31. Augusto JB, Davies RH, Bhuva AN, et al. Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *Lancet Digit Health*. 2021;3(1):e20–e28.

32. Yuan V, Sahashi Y, Ieki H, et al. Automated deep learning pipeline for characterizing left ventricular diastolic function. *medRxiv*. Published online April 30, 2025. https://doi.org/10.1101/2025.04.29.25326683

33. Lau ES, Di Achille P, Kopparapu K, et al. Deep learning-enabled assessment of left heart structure and function predicts cardiovascular outcomes. *J Am Coll Cardiol*. 2023;82(20):1936–1948.

34. He B, Kwan AC, Cho JH, et al. Blinded, randomized trial of sonographer versus AI cardiac function assessment. *Nature*. 2023;616(7957):520–524.

35. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36(4):1234–1240.

KEY WORDS automated measurement, convolutional neural network, deep learning, Doppler wave, echocardiography

APPENDIX For supplemental tables, figures, videos, and methods, please see the online version of this paper.