

High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy With Cardiovascular Deep Learning

Grant Duffy, BS; Paul P. Cheng, MD, PhD; Neal Yuan, MD; Bryan He, BS; Alan C. Kwan, MD; Matthew J. Shun-Shin, PhD; Kevin M. Alexander, MD; Joseph Ebinger, MD; Matthew P. Lungren, MD; Florian Rader, MD, MSc; David H. Liang, MD, PhD; Ingela Schnittger, MD; Euan A. Ashley, MBChB, DPhil; James Y. Zou, PhD; Jignesh Patel, MD, PhD; Ronald Witteles, MD; Susan Cheng, MD, MPH; David Ouyang, MD

IMPORTANCE Early detection and characterization of increased left ventricular (LV) wall thickness can markedly impact patient care but is limited by under-recognition of hypertrophy, measurement error and variability, and difficulty differentiating causes of increased wall thickness, such as hypertrophy, cardiomyopathy, and cardiac amyloidosis.

OBJECTIVE To assess the accuracy of a deep learning workflow in quantifying ventricular hypertrophy and predicting the cause of increased LV wall thickness.

DESIGN, SETTINGS, AND PARTICIPANTS This cohort study included physician-curated cohorts from the Stanford Amyloid Center and Cedars-Sinai Medical Center (CSMC) Advanced Heart Disease Clinic for cardiac amyloidosis and the Stanford Center for Inherited Cardiovascular Disease and the CSMC Hypertrophic Cardiomyopathy Clinic for hypertrophic cardiomyopathy from January 1, 2008, to December 31, 2020. The deep learning algorithm was trained and tested on retrospectively obtained independent echocardiogram videos from Stanford Healthcare, CSMC, and the Unity Imaging Collaborative.

MAIN OUTCOMES AND MEASURES The main outcome was the accuracy of the deep learning algorithm in measuring left ventricular dimensions and identifying patients with increased LV wall thickness diagnosed with hypertrophic cardiomyopathy and cardiac amyloidosis.

RESULTS The study included 23 745 patients: 12 001 from Stanford Health Care (6509 [54.2%] female; mean [SD] age, 61.6 [17.4] years) and 1309 from CSMC (808 [61.7%] female; mean [SD] age, 62.8 [17.2] years) with parasternal long-axis videos and 8084 from Stanford Health Care (4201 [54.0%] female; mean [SD] age, 69.1 [16.8] years) and 2351 from CSMS (6509 [54.2%] female; mean [SD] age, 69.6 [14.7] years) with apical 4-chamber videos. The deep learning algorithm accurately measured intraventricular wall thickness (mean absolute error [MAE], 1.2 mm; 95% CI, 1.1-1.3 mm), LV diameter (MAE, 2.4 mm; 95% CI, 2.2-2.6 mm), and posterior wall thickness (MAE, 1.4 mm; 95% CI, 1.2-1.5 mm) and classified cardiac amyloidosis (area under the curve [AUC], 0.83) and hypertrophic cardiomyopathy (AUC, 0.98) separately from other causes of LV hypertrophy. In external data sets from independent domestic and international health care systems, the deep learning algorithm accurately quantified ventricular parameters (domestic: R^2 , 0.96; international: R^2 , 0.90). For the domestic data set, the MAE was 1.7 mm (95% CI, 1.6-1.8 mm) for intraventricular septum thickness, 3.8 mm (95% CI, 3.5-4.0 mm) for LV internal dimension, and 1.8 mm (95% CI, 1.7-2.0 mm) for LV posterior wall thickness. For the international data set, the MAE was 1.7 mm (95% CI, 1.5-2.0 mm) for intraventricular septum thickness, 2.9 mm (95% CI, 2.4-3.3 mm) for LV internal dimension, and 2.3 mm (95% CI, 1.9-2.7 mm) for LV posterior wall thickness. The deep learning algorithm accurately detected cardiac amyloidosis (AUC, 0.79) and hypertrophic cardiomyopathy (AUC, 0.89) in the domestic external validation site.

CONCLUSIONS AND RELEVANCE In this cohort study, the deep learning model accurately identified subtle changes in LV wall geometric measurements and the causes of hypertrophy. Unlike with human experts, the deep learning workflow is fully automated, allowing for reproducible, precise measurements, and may provide a foundation for precision diagnosis of cardiac hypertrophy.

JAMA Cardiol. 2022;7(4):386-395. doi:10.1001/jamacardio.2021.6059
Published online February 23, 2022.

 Multimedia

 Supplemental content

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

Corresponding Authors: Susan Cheng, MD, MPH (susan.cheng@cshs.org), and David Ouyang, MD (david.ouyang@cshs.org), Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, 127 S San Vicente Blvd, Ste A3600, Los Angeles, CA 90048.

Despite rapidly advancing developments in targeted therapeutics and genetic sequencing,^{1,2} persistent limits in the accuracy and throughput of clinical phenotyping has led to a widening gap between the potential and the actual benefits realized by precision medicine. This conundrum is exemplified by current approaches to assessing morphologic alterations of the heart.^{3,4} If reliably identified, misdiagnoses of certain cardiac diseases (eg, cardiac amyloidosis and hypertrophic cardiomyopathy [HCM]) could be avoided, and specific targeted therapies could be initiated efficiently. Systematic screening paradigms, including through imaging and automated medical record feature review, have shown the opportunity to identify patients with underdiagnosed diseases that are increasingly recognized as more prevalent than was previously thought.⁵⁻⁹ The ability to reliably distinguish between cardiac disease types with similar morphologic features but different causes would also enhance specificity for linking genetic risk variants and determining mechanisms.

The heart is a dynamic organ capable of remodeling and adapting to physiological stress and extracardiac perturbation. Both intrinsic cardiac disease and systemic insults can result in similar presentations of increased left ventricular (LV) wall thickness and LV hypertrophy (LVH), which are difficult to distinguish on routine imaging by human observation. Pressure overload from long-standing hypertension and aortic stenosis can cause cardiac remodeling to compensate for additional physiological work, and HCM and cardiac amyloidosis can similarly manifest with an increase in LV mass in the absence of physiological stress.^{6,10}

In addition to the presence of LVH, the degree of ventricular thickness also has substantial prognostic value in many diseases.¹⁰⁻¹² Ventricular wall thickness is used to risk-stratify patients for risk of sudden cardiac death and help determine which patients should undergo defibrillator implantation.¹⁰ Nevertheless, quantification of ventricular wall thickness remains subject to substantial intraprovider and interprovider variability across imaging modalities.^{13,14} Even with the high image resolution and signal-to-noise ratio of cardiac magnetic resonance imaging, there is marked test-retest variability owing to the laborious, manual nature of wall thickness measurement.^{15,16} Although abundant,^{17,18} low cost, and without ionizing radiation, echocardiography relies on expert interpretation, and its accuracy is dependent on careful application of measurement techniques.^{19,20}

Recent work²¹⁻²⁴ has shown that deep learning applied to medical imaging can identify clinical phenotypes beyond conventional image interpretation and with higher accuracy than interpretation by human experts. We hypothesized that echocardiography, the most common form of cardiovascular imaging on the basis of cardiology society guidelines for diagnosing hypertrophy including those of the European Society of Cardiology,¹⁰ when enhanced with artificial intelligence (AI) models, could provide additional value in understanding disease states by predicting both the presence and the potential cause of LVH in a screening population. To address current limitations in the assessment of ventricular hypertrophy and disease diagnosis, we developed an end-to-

Key Points

Question Can deep learning be used to automate measurements of left ventricular dimensions and identify patients who could benefit from screening for underdiagnosed diseases?

Findings In this cohort study of 23 745 patients that used a deep learning algorithm to automatically measure left ventricular dimensions and to identify patients with increased wall thickness who may benefit from additional screening for hypertrophic cardiomyopathy and cardiac amyloidosis, the algorithm performed consistently across multiple cohorts while also delivering results in a shorter time than required for human assessment.

Meaning In this study, a deep learning workflow was able to automate wall thickness evaluation while facilitating identification of hypertrophic cardiomyopathy and cardiac amyloidosis.

end deep learning approach for labeling the LV dimensions, quantifying ventricular wall thickness, and predicting the cause of LVH. We first conducted frame-level semantic segmentation of the left ventricular wall thickness from parasternal long-axis echocardiogram videos and then performed beat-to-beat evaluation of ventricular hypertrophy. After identifying LVH, we used a 3-dimensional convolutional neural network with residual connections to predict the cause of the LVH, including predictions for cardiac amyloidosis and aortic stenosis among a background of other hypertrophic diseases.

Methods

Data Curation

A standard full resting echocardiogram study consists of a series of 50 to 100 videos and still images visualizing the heart from different angles, locations, and image acquisition techniques (eg, 2-dimensional images, tissue Doppler images, and color Doppler images). In this cohort study, patients were identified in physician-curated cohorts from the Stanford Amyloid Center and Cedars-Sinai Medical Center (CSMC) Advanced Heart Disease Clinic for cardiac amyloidosis and the Stanford Center for Inherited Cardiovascular Disease and the CMSC Hypertrophic Cardiomyopathy Clinic for hypertrophic cardiomyopathy from January 1, 2008, to December 31, 2020. This research was approved by the Stanford University and CMSC institutional review boards. Written informed consent was obtained from patients for inclusion in the cohorts, but the need for consent for echocardiographic imaging analysis was exempted by the participating institutional review boards because of the use of deidentified images.

Relevant parasternal long-axis (PLAX) and apical 4-chamber 2-dimensional videos were extracted from each echocardiogram study (Table 1 and Table 2). Human clinician annotations of intraventricular septum (IVS), LV internal dimension (LVID), and LV posterior wall (LVPW) measurements were used as training labels to assess ventricular hypertrophy. Parasternal long-axis videos obtained from Stanford Health Care (SHC) were split and used as follows: 9600 for the training set, 1200

Table 1. Baseline Characteristics of Patients With Parasternal Long-Axis Videos of Echocardiograms From SHC and CSMC^a

Characteristic	SHC				CSMC test set (n = 1309)
	Training set (n = 9601)	Validation set (n = 1200)	Test set (n = 1200)	Total (N = 12 001)	
Age, mean (SD), y	61.6 (17.4)	61.3 (17.8)	61.7 (17.5)	61.6 (17.4)	62.8 (17.2)
Sex					
Female	5268 (54.9)	618 (51.5)	623 (51.9)	6509 (54.2)	501 (38.3)
Male	4333 (45.1)	582 (48.5)	577 (48.1)	5492 (45.8)	808 (61.7)
Race and ethnicity					
American Indian	24 (0.2)	2 (0.2)	6 (0.5)	32 (0.3)	4 (0.3)
Asian	1377 (14.3)	176 (14.7)	180 (15.0)	1733 (14.4)	85 (6.5)
Black	373 (3.9)	53 (4.4)	50 (4.2)	476 (4.0)	239 (18.3)
Hispanic	1076 (11.2)	135 (11.2)	117 (9.8)	1328 (11.1)	178 (13.6)
Non-Hispanic White	4043 (42.1)	518 (43.2)	516 (43.0)	5077 (42.3)	697 (53.2)
Pacific Islander	140 (1.5)	14 (1.2)	23 (1.9)	177 (1.5)	4 (0.3)
Other ^b	742 (7.7)	98 (8.2)	96 (8.0)	936 (7.8)	83 (6.3)
Unknown	1826 (19.0)	204 (17.0)	212 (17.7)	2242 (18.7)	19 (1.5)
Atrial fibrillation	2067 (39.3)	281 (40.7)	258 (39.6)	2606 (39.5)	315 (24.1)
Congestive heart failure	3162 (60.1)	417 (60.3)	398 (61.1)	3977 (60.2)	451 (34.5)
Hypertension	3783 (71.9)	502 (72.6)	466 (71.6)	4751 (71.9)	543 (41.5)
Diabetes	1852 (35.2)	265 (38.4)	249 (38.2)	2366 (35.8)	248 (18.9)
Coronary artery disease	2517 (47.8)	343 (49.6)	316 (48.5)	3176 (48.1)	369 (28.2)
Chronic kidney disease	2066 (39.3)	265 (38.4)	252 (38.7)	2583 (39.1)	257 (19.6)
LVPWd thickness, mean (SD), cm	1.00 (0.21)	1.01 (0.21)	1.01 (0.21)	1.01 (0.21)	1.09 (0.25)
LVIDd, mean (SD), cm	4.70 (0.83)	4.69 (0.85)	4.71 (0.82)	4.70 (0.83)	4.70 (0.90)
LVIDs, mean (SD), cm	3.28 (0.90)	3.26 (0.90)	3.28 (0.88)	3.28 (0.90)	3.29 (1.05)
IVSd, mean (SD), cm	1.03 (0.24)	1.03 (0.24)	1.03 (0.25)	1.03 (0.24)	1.12 (0.29)
LVEF, mean (SD), %	55.71 (12.31)	55.62 (12.28)	56.08 (12.01)	55.74 (12.28)	55.92 (15.67)
LV mass, mean (SD), g	173.29 (68.71)	173.57 (69.97)	174.92 (68.84)	173.48 (68.84)	195.19 (83.16)

Abbreviations: CSMC, Cedars-Sinai Medical Center; IVSd, intraventricular septal thickness (diastole); LV, left ventricular; LVEF, left ventricular ejection fraction; LVIDd, left ventricular internal dimension (diastole); LVIDs, left ventricular internal dimension (systole); LVPWd, left ventricular posterior wall (diastole); SHC, Stanford Health Care.

^a Data are presented as number (percentage) of patients unless otherwise indicated.

^b Other racial and ethnic groups were not available because they were not included in the electronic health records.

for the validation set, and 1200 for the test set. An additional 7767 SHC echocardiogram studies were from patients with defined disease characteristics, including cardiac amyloidosis, HCM, and severe aortic stenosis. From these studies, the apical 4-chamber videos were extracted and used as input data for the hypertrophic disease classification task. Videos were processed in a previously described automated preprocessing workflow in which identifying information and human jlabels had been removed.^{21,25}

Domestic and International External Health Care System Test Data Sets

Transthoracic echocardiogram studies from CSMC and the Unity Imaging Collaborative¹⁶ were used to evaluate the deep learning algorithm's performance in identifying key points in PLAX videos and measuring ventricular dimensions. Previously described methods were used to identify PLAX and apical 4-chamber-view videos and to convert Digital Imaging and Communications in Medicine files to AVI files.²² We extracted a total of 3660 videos from CSMC as a domestic held-out test data set. Labeled images from the Unity Imaging jCollaborative were used as an additional held-out international test data set not seen during model training. These echocar-

diogram videos were obtained from the British echocardiography laboratories and were retrospectively annotated by echocardiography-certified cardiologists.

Deep Learning Algorithm Development and Training

Model design and training were done in Python, version 3.8.5 (Python Software Foundation) using the PyTorch deep learning library. A modified DeepLabv3²⁶ architecture trained on parasternal long-axis images to minimize a weighted mean square error loss was used to identify key points used for measuring ventricular dimensions. Three-dimensional implementations of a segmentation model took substantially more computational resources without marked improvement in performance. An Adam optimizer with a learning rate of 0.001 was used, and the model was trained for 50 epochs, with early stopping based on the validation loss. We evaluated different video lengths, resolutions, and temporal resolutions as hyperparameters to optimize model performance. Computational cost was evaluated using 1 NVIDIA GeForce GTX 3090.

For video-based disease classification, an 18-layer ResNet3D²⁷ architecture was used to classify videos. Given the potential for patients who had diagnoses of both aortic steno-

Table 2. Baseline Characteristics of Patients With Apical 4-Chamber Videos of Echocardiograms From SHC and CSMC^a

Characteristic	SHC				CSMC test set (n = 2351)
	Training set (n = 6461)	Validation set (n = 814)	Test set (n = 809)	Total (N = 8084)	
Cardiac amyloidosis	950 (14.7)	117 (14.4)	120 (14.8)	1187 (14.7)	358 (15.2)
Hypertrophic cardiomyopathy	2344 (36.3)	298 (36.6)	294 (36.3)	2936 (36.3)	146 (6.2)
Aortic stenosis	1061 (16.4)	133 (16.3)	132 (16.3)	1326 (16.4)	468 (19.9)
Other LVH	2106 (32.6)	266 (32.7)	263 (32.5)	2635 (32.6)	1379 (58.7)
Age, mean (SD), y	69.0 (16.9)	70.0 (16.6)	69.4 (16.0)	69.1 (16.8)	69.6 (14.7)
Sex					
Female	3352 (51.9)	414 (50.9)	435 (53.8)	4201 (52.0)	731 (31.1)
Male	3109 (48.1)	400 (49.1)	374 (46.2)	3883 (48.0)	1620 (68.9)
Race and ethnicity					
American Indian	11 (0.2)	1 (0.1)	1 (0.1)	13 (0.2)	4 (0.2)
Asian	586 (9.1)	77 (9.5)	53 (6.6)	716 (8.9)	105 (4.5)
Black	252 (3.9)	24 (2.9)	37 (4.6)	313 (3.9)	353 (15.0)
Hispanic	551 (8.5)	59 (7.2)	56 (6.9)	666 (8.2)	266 (11.3)
Non-Hispanic White	2895 (44.8)	387 (47.5)	346 (42.8)	3628 (44.9)	1468 (62.4)
Pacific Islander	58 (0.9)	7 (0.9)	2 (0.2)	67 (0.8)	3 (0.1)
Other ^b	666 (10.3)	86 (10.6)	96 (11.9)	848 (10.5)	146 (6.2)
Unknown	1442 (22.3)	173 (21.3)	218 (26.9)	1833 (22.7)	6 (0.3)
Atrial fibrillation	2209 (49.5)	293 (51.3)	262 (50.0)	2764 (49.8)	630 (26.8)
Congestive heart failure	3632 (81.5)	467 (81.8)	422 (80.5)	4521 (81.4)	876 (37.3)
Hypertension	3657 (82.0)	473 (82.8)	423 (80.7)	4553 (82.0)	1290 (54.9)
Diabetes	1462 (32.8)	178 (31.2)	180 (34.4)	1820 (32.8)	478 (20.3)
Coronary artery disease	2858 (64.1)	359 (62.9)	324 (61.8)	3541 (63.8)	840 (35.7)
Chronic kidney disease	2238 (50.2)	297 (52.0)	292 (55.7)	2827 (50.9)	513 (21.8)
LVPWd thickness, mean (SD), cm	1.18 (0.25)	1.19 (0.24)	1.17 (0.24)	1.18 (0.25)	1.44 (0.26)
LVIDd, mean (SD), cm	4.58 (0.75)	4.61 (0.75)	4.58 (0.72)	4.58 (0.75)	4.52 (0.86)
LVIDs, mean (SD), cm	3.15 (0.78)	3.19 (0.81)	3.17 (0.77)	3.15 (0.78)	3.07 (0.81)
IVSd, mean (SD), cm	1.27 (0.31)	1.28 (0.31)	1.26 (0.32)	1.27 (0.31)	1.52 (0.37)
LVEF, mean (SD), %	56.92 (11.70)	56.50 (11.61)	56.73 (12.03)	56.86 (11.72)	59.06 (13.68)
LV mass, mean (SD), g	215.73 (80.61)	219.52 (76.97)	213.66 (78.67)	215.91 (80.06)	278.64 (96.53)

Abbreviations: CSMC, Cedars-Sinai Medical Center; IVSd, intraventricular septal thickness (diastole); LV, left ventricular; LVEF, left ventricular ejection fraction; LVIDd, left ventricular internal dimension (diastole); LVIDs, left ventricular internal dimension (systole); LVPWd, left ventricular posterior wall (diastole); SHC, Stanford Health Care.

^a Data are presented as number (percentage) of patients unless otherwise indicated.

^b Other racial and ethnic groups were not available because they were not included in the electronic health records.

sis and cardiac amyloidosis with multiple causes of LVH,²⁸ parallel binary classification deep learning models were trained to predict probability of cardiac amyloidosis, HCM, and aortic stenosis secondary to uncontrolled hypertension and in the context of end-stage kidney disease independently. Distinct from previous literature,^{7,8} for each classification task, the negative controls were images from patients with other causes of LVH to mimic the clinical workflow. For example, during amyloid classification, the negative training examples included videos from patients with diagnosed HCM, aortic stenosis, hypertension, and end-stage kidney disease as other causes of LVH. This model was trained to minimize binary cross-entropy loss using an Adam optimizer with a learning rate of 0.01. The model was trained for 100 epochs with a batch size of 14, with early stopping based on area under the curve (AUC) on the validation set (eMethods in the Supplement).

Comparison With Variation in Human Measurement

Using the reporting database of the Stanford Echocardiography Laboratory, we identified paired studies of the same

patient for which the reviewing cardiologist determined there was no substantial change from the current study to the previous study by a structured reporting element. Of these studies with clinical stability, we analyzed the subset of 23 874 studies for which LVID, IVS, and LVPW at diastole were measured for both the current and the subsequent studies. The variance in measurement between the previous and subsequent studies was used as a surrogate of clinician variation and was compared with the variation of the deep learning algorithm. In the CSMC data set, we identified 99 random studies; blinded relabeling was performed by 2 level 3 echocardiography-certified cardiologists, and their performance was compared with the performance of the deep learning algorithm on the consensus label.

Statistical Analysis

The 95% CIs were computed using 10 000 bootstrapped samples and by obtaining 95th percentile ranges for each prediction. The performance of the semantic segmentation task was evaluated by comparing the length of LVID, LVPW, and IVS with human labels in the hold-out test data set. The centroid

of each predicted key point was used to calculate measurements with Python software, version 3.8.5.

Results

The study included 23 745 patients: 12 001 from SHC (6509 [54.2%] female; mean [SD] age, 61.6 [17.4] years) and 1309 from CSMC (808 [61.7%] female; mean [SD] age, 62.8 [17.2] years) with parasternal long-axis videos and 8084 from SHC (4201 [54.0%] female; mean [SD] age, 69.1 [16.8] years) and 2351 from CSMC (6509 [54.2%] female; mean [SD] age, 69.59 [14.7] years) with apical 4-chamber videos. The deep learning workflow for screening of HCM and cardiac amyloidosis had 2 components (Figure 1). First, we designed a deep learning model with atrous convolutions for semantic segmentation of PLAX echocardiogram videos and identification of the IVS, LV internal dimension, and LV posterior wall. With atrous convolutions to capture longer-range features, full-resolution PLAX frames were used as input images for higher-resolution assessment of LVH. Given the tedious nature of annotation, in the standard clinical workflow, only 1 or 2 frames of a video are often labeled, but each video records multiple heartbeats that can be used for clinical measurements (Figure 2). Therefore, we generalized a neural network trained on these sparse annotations into measurement predictions for every frame of the entire video to allow for beat-to-beat estimation of ventricular wall thickness and dimensions. Representative example parasternal long axis videos with model annotations are shown in the Video.

After detection of LVH, identifying the specific cause (eg, infiltrative disease, inherited cardiomyopathies, or chronic elevated afterload) can help guide therapy. We trained a video-based convolutional neural network model with spatiotemporal convolutions to predict the cause of LVH (Figure 3). Integrating spatial and temporal information, the model expanded on previous work²¹ with video-based model interpretation of echocardiograms and classified videos based on probability of hypertension, aortic stenosis, HCM, or cardiac amyloidosis as causes of ventricular hypertrophy. In addition, we performed a video-based model architecture and hyperparameter search to identify the optimal base architecture for the deep learning algorithm (eFigure 1 in the Supplement). The deep learning algorithm was trained on a data set of 17 802 echocardiogram videos from SHC and then evaluated on held-out test cohorts from SHC, CSMC, and Unity Imaging Collaborative.

Evaluation of Hypertrophy Detection

From the held-out test data set from SHC not seen during model training ($n = 1200$), the deep learning algorithm predicted ventricular dimensions with an R^2 of 0.97 compared with annotations by human experts (eFigure 2 in the Supplement). The deep learning algorithm had a mean absolute error (MAE) of 1.2 mm (95% CI, 1.1-1.3 mm) for IVS thickness, 2.4 mm (95% CI, 2.2-2.6 mm) for LVID, and 1.4 mm (95% CI, 1.2-1.5 mm) for LVPW thickness. This compared favorably with clinical inter-provider variation, which had an MAE of 1.3 mm (95% CI, 1.3-1.3 mm) for IVS thickness, 3.7 mm (95% CI, 3.6-3.7 mm) for LVID,

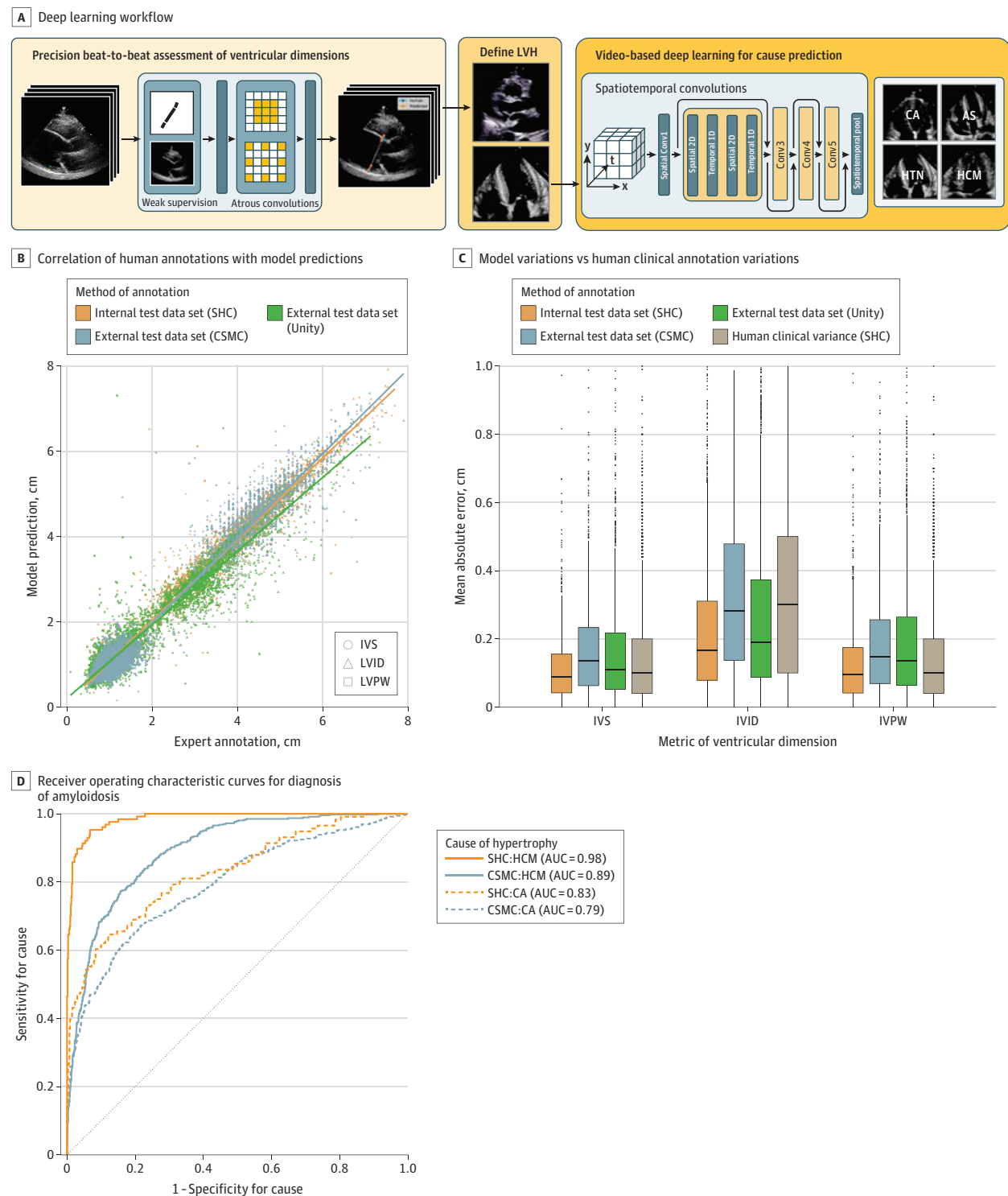
and 1.3 mm (95% CI, 1.3-1.3 mm) for LVPW thickness. The deep learning algorithm also performed well compared with the prospective consensus annotation of 2 level-3 echocardiography-certified cardiologists in 99 random studies from CSMC (eFigure 3 in the Supplement). To assess the reliability of the model across health care systems internationally, the deep learning algorithm was also tested without any tuning on an external test data set of 1791 videos from Unity Imaging Collaborative and 3660 videos from CSMC. On the Unity Imaging Collaborative external test data set, the deep learning algorithm showed a robust prediction accuracy, with an overall R^2 of 0.90 and MAEs of 2.2 mm (95% CI, 1.7-2.6 mm) for IVS thickness, 4.5 mm (95% CI, 3.7-5.3 mm) for LVID, and 2.4 mm (95% CI, 2.0-2.7 mm) for LVPW thickness. When fine-tuned using the training split of the Unity Imaging Collaborative data set, the deep learning algorithm showed an improved performance with an overall R^2 of 0.92 and MAEs of 1.7 mm (95% CI, 1.5-2.0 mm) for IVS thickness, 2.9 mm (95% CI, 2.4-3.3 mm) for LVID, and 2.3 mm (95% CI, 1.9-2.7 mm) for LVPW thickness on the Unity Imaging Collaborative validation data split, indicating data shift and potential variations in practice across institutions and continents (eTable 1 in the Supplement).

A rapid, high-throughput automated approach allowed for measurement of every individual frame, which would be tedious for manual tracing (Figure 2). Differences in filling time and irregularity in the heart rate can cause variation in measurement, but beat-to-beat model assessment can provide higher-fidelity overall assessments. Although the SHC and Unity Imaging Collaborative data sets were directly compared on annotated individual frames, we evaluated the deep learning algorithm's beat-to-beat evaluation on the CSMC data set compared with study-level annotations of ventricular dimensions. In this data set, human measurements were not associated with specific frames of the echocardiogram video, and beat-to-beat analysis was used to predict diastole and mean measurements from each heartbeat throughout the entire video. On the CSMC external test data set, the deep learning algorithm showed a robust prediction accuracy with an overall R^2 of 0.96 and MAEs of 1.7 mm (95% CI, 1.6-1.8 mm) for IVS thickness, 3.8 mm (95% CI, 3.5-4.0 mm) for LVID, and 1.8 mm (95% CI, 1.7-2.0 mm) for LVPW thickness with beat-to-beat evaluation.

Prediction of Cause of LVH

The cause derivation, validation, and test cohorts from SHC had 6215, 787, and 765 videos, respectively. In the held-out test cohort, the deep learning algorithm distinguished cardiac amyloidosis with an AUC of 0.83, HCM with an AUC of 0.98, and aortic stenosis with an AUC of 0.89 from other causes of LVH. On the held-out test cohort, the area under the precision-recall curve of the deep learning algorithm was 0.77 for cardiac amyloidosis, 0.95 for HCM, and 0.79 for aortic stenosis. The proposed ensemble of binary classification video-based deep learning classifiers in the deep learning algorithm was similar in performance to a multilabel, multiclass deep learning model for disease detection but had the flexibility of being able to identify patients who had diagnoses of both aortic stenosis and cardiac amyloidosis. In an external test data set of

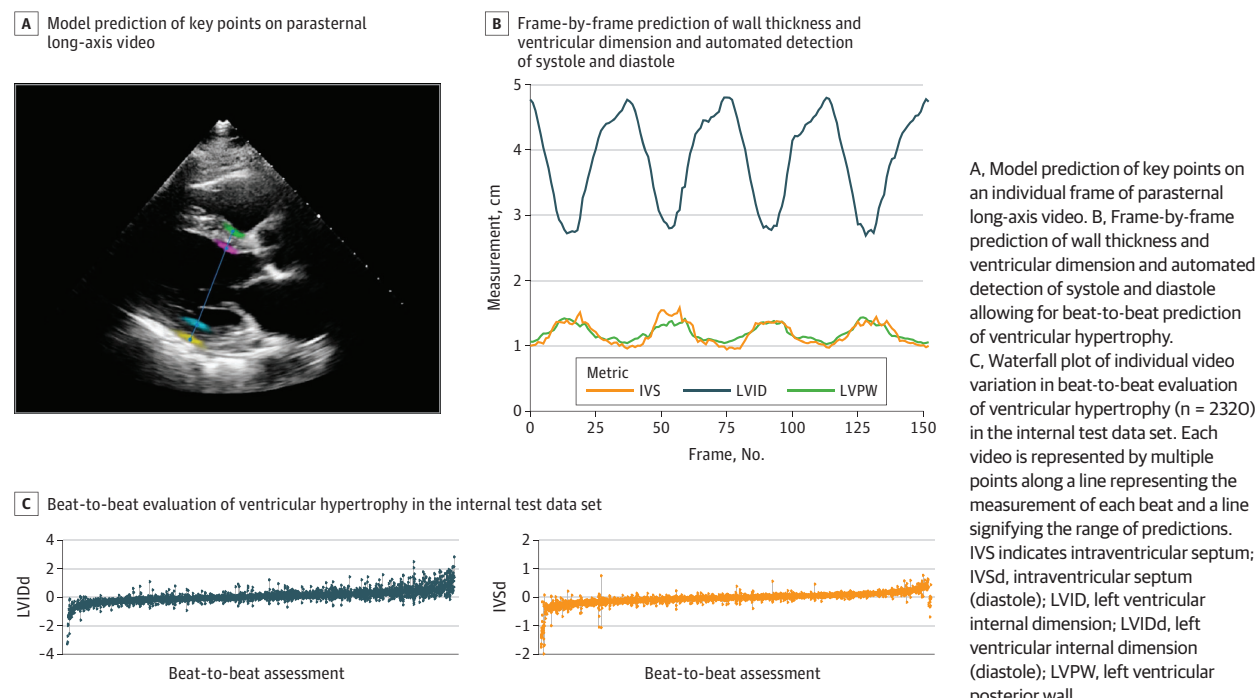
Figure 1. Deep Learning Workflow Combining Evaluation of Ventricular Dimensions and Suspicion for Underdiagnosed Diseases



A, The deep learning algorithm used parasternal long-axis echocardiogram video as input to derive key points and estimate ventricular dimensions. After identifying patients with left ventricular hypertrophy (LVH), the deep learning workflow used a video-based architecture to distinguish common causes of LVH. B, Correlation of human annotations vs model predictions for ventricular dimensions in data sets from Stanford Health Care (SHC; $n = 1200$), Cedars-Sinai Medical Center (CSMC; $n = 1309$), and Unity Imaging Collaborative ($n = 1791$). C, Model variation on the 3 data sets vs

human clinical annotation variation. Middle lines represent means; upper and lower bounds of the boxes, 25th and 75th percentiles; and points, values greater than 1.5 times the IQR. D, Receiver operating characteristic curves for diagnosis of amyloidosis in the SHC validation ($n = 813$) and test ($n = 812$) sets. AS indicates aortic stenosis; AUC, area under the curve; CA, cardiac amyloid; HCM, hypertrophic cardiomyopathy; HTN, hypertension; IVS, intraventricular septum; LVID, LV internal dimension; LVPW, LV posterior wall.

Figure 2. Beat-to-Beat Evaluation of Ventricular Dimensions



2351 apical 4-chamber-view videos from CSMC with 358 videos of cardiac amyloidosis, 146 videos of aortic stenosis, 468 videos of HCM, and 1379 videos of other causes of LVH, the deep learning algorithm had an AUC of 0.79 for predicting cardiac amyloidosis and an AUC of 0.89 for predicting HCM. For the CSMC cohort, the area under the precision-recall curve of the deep learning algorithm was 0.54 for cardiac amyloidosis, 0.69 for HCM, and 0.08 for aortic stenosis. The model performance was consistent across body mass index and image quality (eTable 2 in the [Supplement](#)).

Phenotypic Mimics and Disease-Specific Training Pipeline

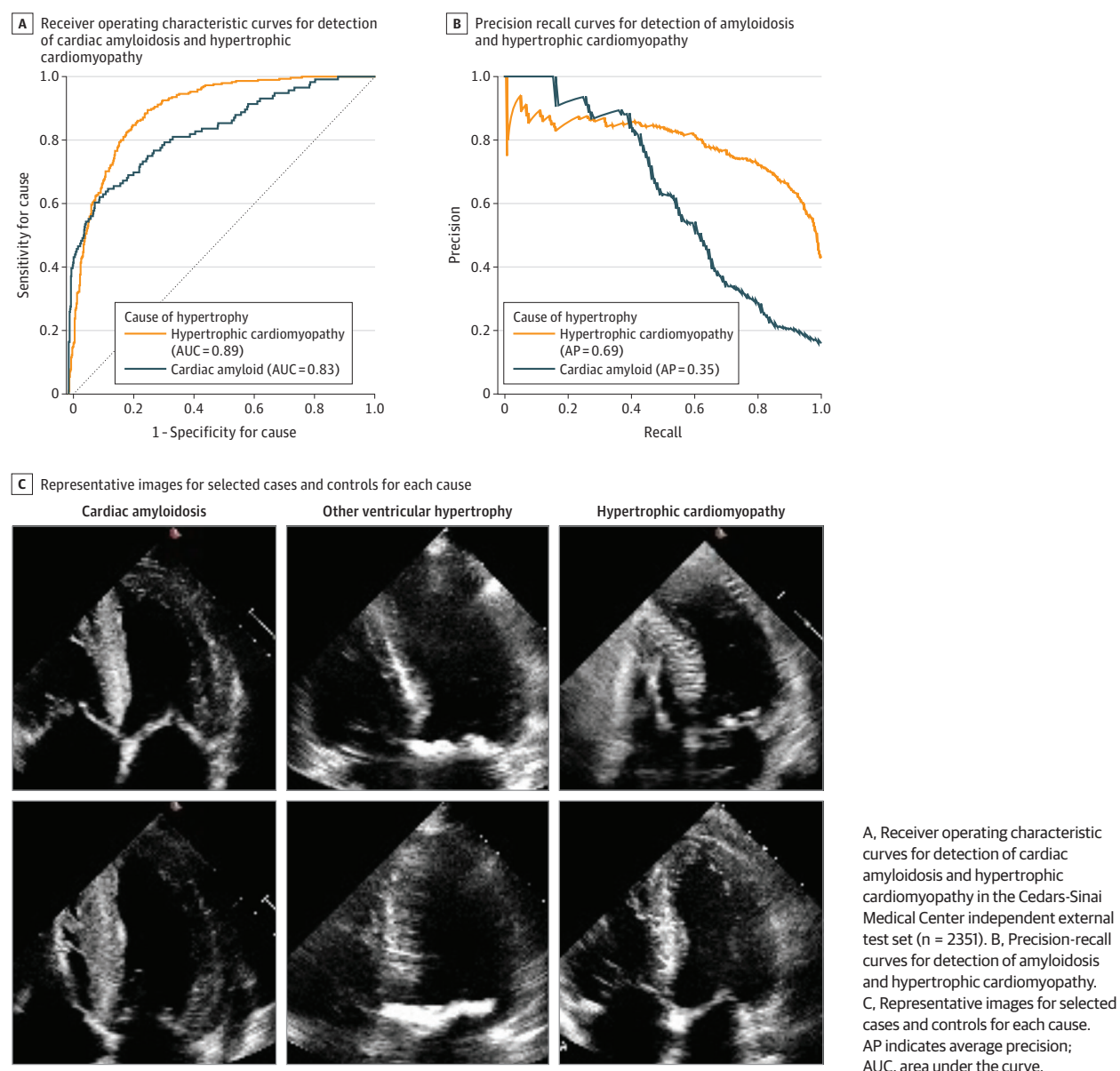
To highlight the benefit of training a model with negative controls derived from other causes of LVH instead of normal controls, we performed a series of experiments to see how a model that was trained without seeing other phenotypic mimics would perform when encountering phenotypic mimics. A confusion matrix was generated in the 2 experimental settings (eTable 3 in the [Supplement](#)) in which a higher AUC outside the diagonal showed the model confusion and a lower AUC suggested improved discrimination between phenotypic mimics. In this experiment, although the model produced a higher AUC (0.96 for cardiac amyloidosis, 0.98 for aortic stenosis, and 0.97 for HCM), there was substantial confusion when other causes were introduced, suggesting that a model trained only on age- and sex-matched controls would primarily identify LVH.

Discussion

The AI-guided workflow used in this cohort study was a deep learning system that automatically quantified LV wall thick-

ness on echocardiography while also predicting the cause of LVH as either HCM or cardiac amyloidosis. The deep learning algorithm performed measurements of ventricular thickness and diameter well within the variance of human clinical test-retest assessment while aiding the detection of subtle ventricular phenotypes that tend to be challenging for human readers. This integration of LV measurement and prediction of cause offers an automated workflow for disease screening from echocardiography, the most frequently used form of cardiac imaging. As such, echocardiography-based screening can provide a high index of suspicion that can facilitate more efficient clinical evaluation, diagnosis, and care. Assimilation of automated diagnostic algorithms with widely available clinical imaging can reduce physician burden while streamlining opportunities for more targeted cardiovascular care.

Studies²⁸⁻³⁰ have suggested that diseases such as cardiac amyloidosis are underdiagnosed rather than rare. Given the large heterogeneous population of patients with heart failure with preserved ejection fraction,³¹ methods that might appropriately and efficiently increase suspicion for under-recognized causes, such as subtypes of amyloidosis with newly available targeted therapies, may help address a persistent unmet need. Accordingly, an opportunity exists in the application of efficient AI algorithms to increase recognition of historically underdiagnosed diseases in stored images in databases of large echocardiography laboratories. Notwithstanding that all patient data should be interpreted in the clinical context, our findings suggest that automated image analysis workflows could be feasibly implemented to rapidly identify patients who could benefit from follow-up screening in large populations. As such, more prospective work is needed to evaluate the potential of such algorithms to expedite appro-

Figure 3. Performance of Disease Cause Classification in the Independent External Validation Cohort

appropriate clinical evaluation, targeted testing, and confirmation before eventual diagnosis and initiation of disease-modifying therapy.

The results of the present study represent a step toward the automated assessment of cardiac structures in echocardiogram videos through deep learning. Although individual linear measurements take only seconds to measure, there is inherent variation in frame and video selection that sets a floor to the precision of manual measurements derived from echocardiography. Future work should augment echocardiographic labels with annotations and information from cardiac magnetic resonance imaging and other imaging modalities to have more precision automation. By using an automated method, potentially more precise measurements can be obtained in busy clinical and research settings. Combined with

previous work⁸ assessing cardiac function, the present study showed that deep learning models on echocardiogram images can automate an increasingly larger proportion of tasks for assessing cardiac function and structure to provide more holistic evaluation of cardiovascular disease. With improved precision to detect ventricular remodeling and cardiac dysfunction, AI systems offer the potential for earlier detection and treatment of subclinical cardiovascular disease, including less common or underdiagnosed conditions.

Strengths and Limitations

This study has several strengths. A key challenge in the use of AI in health care has been the lack of benchmark data sets for direct comparison of models and engineering workflows across institutions. Data set inclusion criteria, differences in annotations and

disease definitions, and protocols of how to annotate images are sources of data set shift that limit the direct comparison of model performance.^{32,33} With fine-tuning on site-specific data, our model compares favorably with prior state-of-the-art approaches to assessing ventricular wall thickness and hypertrophy on open benchmarks.¹⁶ Expanding on previous work,²¹ we collaborated with stakeholders across Stanford Medicine to release our data set of 12 000 deidentified PLAX echocardiogram videos as a resource for the medical machine learning community for future comparison and validation of deep learning models. This expands the prior data set of 10 030 apical 4-chamber videos²¹ to a total of 22 030 echocardiogram videos made publicly available, which to our knowledge, is the largest data set release of labeled medical videos with matched clinician annotations. We hope this data set will facilitate new echocardiogram- and medical video-based machine learning approaches. We also released the full code for our algorithm and data-processing workflow.

This study also has limitations. First, because the training images for this study were obtained from curated cohorts of patients from tertiary care specialty clinics, biases in patient selection for these clinics can be present in deployment of such algorithms. For example, although hereditary cardiac amyloidosis is known to disproportionately affect Black individuals in the US, they are underrepresented in study cohorts, and care must be taken to extrapolate performance of deep learning algorithms in populations with different demographic characteristics.^{34,35} Second, our model was trained on videos

obtained by expert sonographers at an academic medical center. With expansion in the use of point-of-care ultrasonography for evaluation of cardiac function by noncardiologists, further work is needed to understand model performance with input videos of more variable quality and acquisition expertise as well as in comparison with other imaging modalities. Although our analyses across health systems suggest that our deep learning algorithm is robust to variation in practice patterns across continents, prospective deployment and testing of AI systems in diverse clinical environments remain to be done. A key limitation of research in this field has been a dearth of prospective trials and evaluation of model performance during clinical deployments.³⁴ As such, further work and prospective validation are needed to more fully understand the effect of AI-guided screening workflows on clinical care.

Conclusions

In this cohort study, using measurements across multiple heartbeats and validated against 3 international cohorts, the deep learning model more accurately identified subtle changes in LV wall geometric measurements than did human assessment and accurately identified the cause of LVH. A rapid, fully automated workflow, the deep learning algorithm allows for reproducible, precise measurements and may provide the foundation for precision diagnosis of cardiac hypertrophy.

ARTICLE INFORMATION

Accepted for Publication: December 21, 2021.

Published Online: February 23, 2022.

doi:10.1001/jamacardio.2021.6059

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2022 Duffy G et al. *JAMA Cardiology*.

Author Affiliations: Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California (Duffy, Yuan, Kwan, Ebinger, Rader, Patel, S. Cheng, Ouyang); Department of Medicine, Division of Cardiology, Stanford University, Stanford, California (P. P. Cheng, Alexander, Liang, Schnittger, Ashley, Witteles); Department of Computer Science, Stanford University, Stanford, California (He, Zou); National Heart and Lung Institute, Imperial College London, London, United Kingdom (Shun-Shin); Department of Radiology, Stanford University, Stanford, California (Lungren); Department of Biomedical Data Science, Stanford University, Stanford, California (Zou); Division of Artificial Intelligence in Medicine, Cedars-Sinai Medical Center, Los Angeles, California (Ouyang).

Author Contributions: Mr Duffy, Dr P. P. Cheng, and Dr Yuan served as co-first authors and contributed equally to this work. Dr S. Cheng and Dr Ouyang served as co-senior authors and contributed equally to this work. Mr Duffy and Dr Ouyang had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Concept and design:* Duffy, P. P. Cheng, Kwan, Alexander, Lungren, Zou, Witteles, S. Cheng, Ouyang.

Acquisition, analysis, or interpretation of data:

Duffy, P. P. Cheng, Yuan, He, Kwan, Shun-Shin, Alexander, Ebinger, Rader, Liang, Schnittger, Ashley, Patel, Ouyang.

Drafting of the manuscript: Duffy, P. P. Cheng, Ouyang.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Duffy, P. P. Cheng, Zou, Ouyang.

Obtained funding: P. P. Cheng, Ouyang.

Administrative, technical, or material support: Duffy, P. P. Cheng, Kwan, Shun-Shin, Alexander, Ebinger, Patel, S. Cheng, Ouyang.

Supervision: Kwan, Lungren, Ashley, Witteles, Ouyang.

Conflict of Interest Disclosures: Mr Duffy reported having a patent pending for EchoNet-LVH through Cedars-Sinai Medical Center. Dr P. P. Cheng reported having a patent pending for EchoNet-LVH. Dr Kwan reported receiving grants from the Doris Duke Charitable Foundation during the conduct of the study. Dr Alexander reported receiving consulting fees from Alnylam, Eidos, and Pfizer outside the submitted work. Dr Lungren reported receiving personal fees from Nines Radiology, Philips Healthcare, and SegMed; consulting fees from Centaur Equity; and grants from the National Institutes of Health outside the submitted work. Dr Rader reported receiving consulting fees from Medtronic, Recor, and Bristol Myers Squibb outside the submitted work. Dr Ashley reported being a founder of Personalis Inc, DeepCell Inc, and Silicon Valley Sports Analytics; serving on the advisory board for AstraZeneca; serving as an advisor for Nuevoco Founding, Cathay, Third Rock Ventures, Medical Excellence, Foresite, Novartis, Genome Medical, Disney, and Sequence Bio; receiving grants

from Bristol Myers Squibb, Takeda, Google, and Verily; receiving hardware support from Samsung Collaborative, Analog Devices Collaborative, Illumina Collaborative, PacBio Collaborative, and Nanopore Collaborative; and consulting fees from Apple Advisor outside the submitted work. Dr Witteles reported receiving personal fees from Pfizer, Alnylam, Eidos, Ionis, and Intelia outside the submitted work. Dr Ouyang reported having a patent pending for EchoNet-LVH. No other disclosures were reported.

Funding/Support: This study was supported by grants K99-HL157421 (Dr Ouyang) and 5T32HL116273-07 (Dr Yuan) from the National Institutes of Health.

Role of the Funder/Sponsor: The NIH had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Information: The EchoNet-LVH dataset, a publicly available dataset of deidentified echocardiogram videos, is available at <https://echonet.github.io/lvh/>. The code for EchoNet-LVH is available at <https://github.com/echonet/lvh>.

REFERENCES

1. Mardis ER. A decade's perspective on DNA sequencing technology. *Nature*. 2011;470(7333):198-203. doi:10.1038/nature09796
2. National Human Genome Research Institute, National Institutes of Health. DNA sequencing costs: data. Accessed June 1, 2021. <https://www.>

genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data

3. Lindpaintner K, Lee M, Larson MG, et al. Absence of association or genetic linkage between the angiotensin-converting-enzyme gene and left ventricular mass. *N Engl J Med*. 1996;334(16):1023-1028. doi:10.1056/NEJM199604183341604
4. Vasan RS, Glazer NL, Felix JF, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA*. 2009;302(2):168-178. doi:10.1001/jama.2009.978-a
5. Huda A, Castaño A, Niyogi A, et al. A machine learning model for identifying patients at risk for wild-type transthyretin amyloid cardiomyopathy. *Nat Commun*. 2021;12(1):2725. doi:10.1038/s41467-021-22876-9
6. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015;131(3):269-279. doi:10.1161/CIRCULATIONAHA.114.010637
7. 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
8. 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
9. Davies D, Minamisawa M, Scott C, et al. A simple score to predict transthyretin cardiac amyloidosis in heart failure with preserved ejection fraction. *J Am Col Cardiol*. 2021;77(18)(suppl 1):521.
10. Elliott PM, Anastakis A, Borger MA, et al; Authors/Task Force members. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur Heart J*. 2014;35(39):2733-2779. doi:10.1093/eurheartj/ehu284
11. Lee GY, Kim K, Choi JO, et al. Cardiac amyloidosis without increased left ventricular wall thickness. *Mayo Clin Proc*. 2014;89(6):781-789. doi:10.1016/j.mayocp.2014.01.013
12. Pagourelis ED, Mirea O, Duchenne J, et al. Echo Parameters for differential diagnosis in cardiac amyloidosis: a head-to-head comparison of deformation and nondeformation parameters. *Circ Cardiovasc Imaging*. 2017;10(3):e005588. doi:10.1161/CIRCIMAGING.116.005588
13. Phelan D, Sperry BW, Thavendiranathan P, et al. Comparison of ventricular septal measurements in hypertrophic cardiomyopathy patients who underwent surgical myectomy using multimodality imaging and implications for diagnosis and management. *Am J Cardiol*. 2017;119(10):1656-1662. doi:10.1016/j.amjcard.2017.02.009
14. Angeli F, Verdecchia P, Angeli E, et al. Day-to-day variability of electrocardiographic diagnosis of left ventricular hypertrophy in hypertensive patients: influence of electrode placement. *J Cardiovasc Med (Hagerstown)*. 2006;7(11):812-816. doi:10.2459/01.JCM.0000250869.78777.09
15. Augusto JB, Davies RH, Bhuvana 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. doi:10.1016/S2589-7500(20)30267-3
16. 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. doi:10.1161/CIRCIMAGING.120.011951
17. Douglas PS, Garcia MJ, Haines DE, et al; American College of Cardiology Foundation Appropriate Use Criteria Task Force; American Society of Echocardiography; American Heart Association; American Society of Nuclear Cardiology; Heart Failure Society of America; Heart Rhythm Society; Society for Cardiovascular Angiography and Interventions; Society of Critical Care Medicine; Society of Cardiovascular Computed Tomography; Society for Cardiovascular Magnetic Resonance; American College of Chest Physicians. ACCF/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 appropriate use criteria for echocardiography: a report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Association, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance American College of Chest Physicians. *J Am Soc Echocardiogr*. 2011;24(3):229-267. doi:10.1016/j.echo.2010.12.008
18. Popescu BA, Stefanidis A, Nihoyannopoulos P, et al. Updated standards and processes for accreditation of echocardiographic laboratories from the European Association of Cardiovascular Imaging: an executive summary. *Eur Heart J Cardiovasc Imaging*. 2014;15(11):1188-1193. doi:10.1093/ehjci/jeu057
19. Leibundgut G, Rohner A, Grize L, et al. Dynamic assessment of right ventricular volumes and function by real-time three-dimensional echocardiography: a comparison study with magnetic resonance imaging in 100 adult patients. *J Am Soc Echocardiogr*. 2010;23(2):116-126. doi:10.1016/j.echo.2009.11.016
20. Farsalinos KE, Daraban AM, Ünü S, Thomas JD, Badano LP, Voigt JU. Head-to-head comparison of global longitudinal strain measurements among nine different vendors: the EACVI/ASE Inter-Vendor Comparison Study. *J Am Soc Echocardiogr*. 2015;28(10):1171-1181, e2.
21. 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
22. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2(3):158-164. doi:10.1038/s41551-018-0195-0
23. 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
24. Raghunath S, Ulloa Cerna AE, Jing L, et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med*. 2020;26(6):886-891. doi:10.1038/s41591-020-0870-z
25. Github. ConvertDICOMToAVI.ipynb at master echonet/dynamic. Accessed November 1, 2021. <https://github.com/echonet/dynamic/blob/master/scripts/ConvertDICOMToAVI.ipynb>
26. Yurtkulu SC, Şahin YH, Unal G. Semantic segmentation with extended DeepLabv3 architecture. Paper presented at: 27th Signal Processing and Communications Applications Conference; April 24-26, 2019; Sivas, Turkey.
27. Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. A closer look at spatiotemporal convolutions for action recognition. *arXiv*. Preprint posted online November 30, 2017.
28. Castaño A, Narotsky DL, Hamid N, et al. Unveiling transthyretin cardiac amyloidosis and its predictors among elderly patients with severe aortic stenosis undergoing transcatheter aortic valve replacement. *Eur Heart J*. 2017;38(38):2879-2887. doi:10.1093/eurheartj/ehx350
29. Tanskanen M, Peuralinna T, Polvikoski T, et al. Senile systemic amyloidosis affects 25% of the very aged and associates with genetic variation in alpha2-macroglobulin and tau: a population-based autopsy study. *Ann Med*. 2008;40(3):232-239. doi:10.1080/07853890701842988
30. Gillmore JD, Maurer MS, Falk RH, et al. Nonbiopsy diagnosis of cardiac transthyretin amyloidosis. *Circulation*. 2016;133(24):2404-2412. doi:10.1161/CIRCULATIONAHA.116.021612
31. Shah SJ, Borlaug BA, Kitzman DW, et al. Research priorities for heart failure with preserved ejection fraction: National Heart, Lung, and Blood Institute Working Group summary. *Circulation*. 2020;141(12):1001-1026. doi:10.1161/CIRCULATIONAHA.119.041886
32. Quiñero-Candela J, Sugiyama M, Schwaighofer A, Lawrence ND, eds. *Dataset Shift in Machine Learning*. The MIT Press; 2008. doi:10.7551/mitpress/9780262170055.001.0001
33. Biondetti GP, Gauriau R, Bridge CP, Lu C, Andriole KP. "Name that manufacturer". Relating image acquisition bias with task complexity when training deep learning models: experiments on head CT. *arXiv*. Preprint posted online August 19, 2020.
34. Wu E, Wu K, Daneshjouri R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med*. 2021;27(4):582-584. doi:10.1038/s41591-021-01312-x
35. Banerjee I, Bhimireddy AR, Burns JL, et al. Reading race: AI recognises patient's racial identity in medical images. *arXiv*. Preprint posted online July 21, 2021.