Important Instructions for Approving Your Edited Manuscript

- This PDF document contains your edited manuscript (text and tables).
- **Please respond to my queries within 24 hours.**
- **Do not generate a new electronic version of your manuscript.**
- If you have Adobe Reader (http://www.adobe.com/products/reader.html) or Adobe Acrobat, you may indicate your changes in the file, but you must use Adobe's Comment tool. **Do not make any changes directly in the file using strikethroughs and underscores**. Upload the annotated PDF file to the JAMA Network production system using the encrypted link sent to you in the email.
- If you are unable to use Adobe, indicate clearly all changes by page number and line number. Please include the entire sentence, with your changes clearly marked (in bold, italics, or underline).
- Please verify that the email address we are displaying is accurate and is the one you wish to use for correspondence.
- Please answer all questions and review the manuscript thoroughly.
- If your manuscript contains tables, they will be more readable when they have been formatted; review them for content only, not for bad word breaks or the shapes of the cells.
- If your manuscript has figures, they are being formatted and will be sent to you separately and/or included in the proof that we will send to you for final review.
- The information in your manuscript is confidential. You should not distribute copies of the edited manuscript to anyone, except coauthors, without our approval. The news media should not release any information about your article until the date of publication. Please contact me if you need to confirm your publication date.
- If you have questions or need help coordinating news releases or media attention, contact JAMA Network Media Relations at mediarelations@jamanetwork.org or 312-464-5262.
- Please verify and confirm that all Conflict of Interest disclosure information for you and all coauthors is accurate, complete, up-to-date, and reported accurately and in its entirety.
- Our policy requires that all authors disclose all potential conflicts of interest, including specific financial interests and relationships and affiliations (other than those listed in the author affiliations) as listed are disclosed. Definitions and terms of such disclosures can be found at the journal's Instructions for Authors. Authors should err on the side of full disclosure and should contact the editorial office if they have questions or concerns.
- You may order author reprints or eprints online at www.ama-authorreprints.com.

I hope you will be pleased with our editing and the published version of your manuscript.

Sincerely yours,

Ruth A. Kaufman, MA
Copy Editor
JAMA Network
rakma1985@sbcglobal.net

Original Investigation | Health informatics

«e1: »High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy With Cardiovascular Deep Learning

Grant Duffy, BS[1]

Paul P. Cheng, MD, PhD[2]

Neal Yuan, MD[1]

Bryan He, BS[3]

Alan C. Kwan, MD[1]

Matthew J. Shun-Shin, PhD[4]

Kevin M. Alexander, MD[2]

Joseph Ebinger, MD[1]

Matthew P. Lungren, MD[5]

Florian Rader, MD, MSc[1]

David H. Liang, MD, PhD[2]

Ingela Schnittger, MD[2]

Euan A. Ashley, MRCP, DPhil[2]

Commented [e1]: ATTENTION: Some spelling, formatting, and style preferences are automatically corrected without track changes.

James Y. Zou, PhD[3,6]

Jignesh Patel, MD, PhD[1]

Ronald Witteles, MD[2]

Susan Cheng, MD, MPH[1]

David Ouyang, MD[1,7]

[1]Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California

[2]Department of Medicine, Division of Cardiology, Stanford University, Stanford, California

[3]Department of Computer Science, Stanford University, Stanford, California

[4]National Heart and Lung Institute, Imperial College London, London, United Kingdom

[5]Department of Radiology, Stanford University, Stanford, California

[6]Department of Biomedical Data Science, Stanford University, Stanford, California

[7]Division of Artificial Intelligence in Medicine, Cedars-Sinai Medical Center, Los Angeles, California

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

Findings: **«SB2:** In this cohort study**»** **«SB3:** of 13 310 patients **»**~~We developed~~ that used a deep learning algorithm to automatically measure left ventricular dimensions and ~~also~~ to identify patients with increased wall thickness who ~~could~~ may benefit from additional screening for hypertrophic cardiomyopathy and cardiac amyloidosis~~.~~, ~~We found~~ the algorithm ~~performs~~ performed consistently across multiple cohorts while also delivering results in a ~~a fraction of the~~shorter time than required for human assessment.

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

Importance: Early detection and characterization of increased left ventricular (LV) wall thickness can ~~significantly~~ markedly impact patient care but is limited by under-recognition of hypertrophy, measurement error and variability, and difficulty differentiating **«SB4:** ~~etiologies~~ causes**»** of increased wall thickness, such ~~h~~as hypertrophy, cardiomyopathy, and cardiac amyloidosis.

Commented [SB(2]: AU: The study type chosen during peer review was cohort study, but this appears to be a diagnostic study. Should the study type be changed to diagnostic study throughout?

Commented [SB(3]: Is this the correct total sample size?

Commented [SB(4]: Per the AMA Manual of Style, “etiology” is used only when describing the study of etiology.

Objective: «RK5: To assess~~To overcome this challenge, we presen~~ the accuracy of a deep learning workflow ~~that automatically~~ in ~~quantifies~~ quantifying ventricular hypertrophy ~~with precision equal to human experts~~ and predicting~~s~~ ~~etiology~~ the cause of increased LV wall thickness.**»**

Design, Settings, and Participants: «SB6: 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 **«SB7:** month day, year, to month day, year. **»**The **«SB8:** deep learning algorithm **»**was ~~Trained~~ trained and tested on retrospectively obtained independent echocardiogram videos from Stanford Healthcare, ~~Cedars-Sinai Medical Center (~~CSMC~~)~~, and the Unity Imaging Collaborative.

«SB9: **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 who

Commented [RK5]: AU: Editing of objective OK?

Commented [SB(6]: Added from Methods.

Commented [SB(7]: Please give the study dates.

Commented [SB(8]: Correct wording? Or should this be “model” or “workflow”?

Commented [SB(9]: Please verify the accuracy of this section or suggest revisions.

may benefit from additional screening for hypertrophic cardiomyopathy and cardiac amyloidosis.

Results: ~~Our~~ **«SB10:** The study included 13 310 patients (X [X%] male; mean [SD] age, X [X] years).**»** ~~-~~The deep learning algorithm accurately ~~measures~~ measured intraventricular wall thickness (mean absolute error [MAE], 1.4 mm; 95% CI, 1.2-1.5 mm), ~~left ventricular~~LV diameter (MAE, 2.4 mm; 95% CI, 2.2-2.6 mm), and posterior wall thickness (MAE, 1.2 mm; 95% CI, 1.1-1.3 mm) and ~~classifies~~ classified cardiac amyloidosis (area under the curve [AUC], ~~of~~ 0.83) and hypertrophic cardiomyopathy (AUC, 0.98) separately from other ~~etiologies~~ causes of ~~left ventricular~~LV hypertrophy. In external data sets from independent domestic and international health care systems, ~~our~~ the deep learning algorithm accurately quantified ventricular parameters (domestic: R^2, ~~of~~ 0.96; ~~-~~international: ~~and~~R^2, 0.90)~~., respectively, with~~ For the domestic data set, the MAE ~~of~~ **«SB11:** was 1.7 mm (95% CI, 1.6-1.8 mm) for intraventricular septum ~~[IVS]~~ thickness, 3.8 mm (95% CI, 3.5-4.0 mm) for ~~left ventricular~~LV internal dimension ~~[LVID]~~, and 1.8 mm (95% CI, 1.7-2.0 mm) **»**for ~~left ventricular~~LV posterior wall **«SB12:** thickness**»**. ~~[LVPW] for the~~

Commented [SB(10]: Is this the correct total sample size? Please provide demographic information.

Commented [SB(11]: 95% CIs added from main text. Please confirm that the correct data have been added.

Commented [SB(12]: Correct?

~~domestic health care system and~~ For the international data set, «SB13: the MAE ~~of~~ was 1.6 mm (95% CI, X-X mm) for intraventricular septum thickness~~IVS~~, 3.6 mm (95% CI, X-X mm) for ~~LVID~~LV internal dimension, and 2.1 mm (95% CI, X-X mm) for LV posterior wall thickness~~L~~»~~VPW for the international health care system)~~. The deep learning algorithm accurately~~and~~ detected cardiac amyloidosis (AUC, 0.79) and hypertrophic cardiomyopathy (AUC, 0.89) ~~on~~ in the domestic external validation site.

Conclusions and Relevance: In this cohort study, «SB14: ~~Leveraging measurements across multiple heart beat~~»~~s, our~~ the deep learning model ~~can more~~ accurately ~~identify~~ identified subtle changes in L«SB15: V wall geometric measurements~~y~~ and ~~its~~ the ~~causal etiologies~~causes» of these changes. ~~Compared~~ Unlike ~~to~~ with human experts, ~~our~~ the deep learning workflow is fully automated, allowing for reproducible, precise measurements, and ~~lays~~ may provide ~~the~~ a foundation for precision diagnosis of cardiac hypertrophy. ~~As a resource to promote further innovation, we also make publicly available a large data set of 12 000 annotated echocardiogram videos.~~

Commented [SB(13)]: Please give 95% CIs.

Commented [SB(14)]: This is not mentioned previously in the abstract and thus was deleted here.

Commented [SB(15)]: Correct as edited?

Introduction

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, ~~misdiagnosis~~ and ~~receive~~ specific targeted therapies~~efficient treatment~~ could be ~~initiation~~ initiated efficiently ~~with specific targeted therapies~~. Systematic screening paradigms, including **«RK16:** throughput **»** imaging and automated ~~chart~~ medical record feature review, have shown the opportunity to identify patients ~~for~~ 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 ~~of~~ with similar ~~morphology~~ morphologic features but different ~~etiology~~ causes would also enhance specificity for linking genetic risk variants and determining mechanisms.

Commented [RK16]: AU: As intended?

The heart is a dynamic organ capable of remodeling and adapting to physiologic stress and extracardiac perturbation. Both intrinsic cardiac disease ~~as well as~~and systemic insults can result in similar presentations of increased left ventricular (LV) wall thickness ~~-~~and ~~left ventricular~~LV hypertrophy (LVH), ~~that~~ which are difficult to distinguish on routine imaging by human observation. **«SB17:** Pressure overload from long-standing hypertension and aortic stenosis can cause cardiac remodeling to compensate for additional physiologic work, ~~while~~ and ~~hypertrophic cardiomyopathy~~HCM and cardiac amyloidosis can similarly manifest with an increase in ~~left ventricular~~LV mass ~~but~~ in the absence of physiologic stress**»**.

Commented [SB(17]: Please add citations.

In addition to the presence of LVH, the degree of ventricular thickness also has **«SB18:** ~~significant~~**»** ~~-~~substantial prognostic value in many diseases.[10-12] Ventricular wall thickness is used to ~~risk~~ risk-stratify patients for risk of sudden cardiac death and help determine which patients should undergo defibrillator implantation.[12] Nevertheless, quantification of ventricular wall thickness remains subject to ~~significant~~ substantial intra-provider and inter-provider variability across imaging modalities.[13,14] Even with the high image resolution and signal-to-noise ratio of cardiac magnetic

Commented [SB(18]: The journal only uses “significant” when referring to statistical significance.

resonance imaging, there is ~~significant~~ marked test-retest variability ~~due~~ 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 techinques.[19,20]

Recent work[21-24] has shown that deep learning applied to medical imaging can identify clinical phenotypes beyond conventional image interpretation and with higher accuracy than «**RK19:** interpretation by» human experts.[~~21-24~~] We hypothesized that echocardiography, ~~As~~ the most common form of cardiovascular imaging and the basis of European ~~society~~ Society of Cardiology guidelines for diagnosing hypertrophy,[12] ~~we hypothesize that echocardiography,~~ when enhanced with artificial intelligence (AI) models, ~~can~~ could provide additional value in understanding disease states by predicting both the presence and the potential cause of ~~left ventricular hypertrophy~~LVH in a screening population~~as well asthe potential etiology of diagnosis~~. To ~~overcome~~ address current limitations in the assessment of ventricular hypertrophy and disease diagnosis, we ~~present~~ «**RK20:** developed» an end-to-end deep learning approach for labeling the ~~left ventric~~LV~~le~~ dimensions, quantifying

Commented [RK19]: AU: Addition OK?

Commented [RK20]: AU: Change OK?

ventricular wall thickness, and predicting ~~etiology~~ 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 ~~left ventricular hypertrophy~~LVH, we used a 3-dimensional convolutional neural network with residual connections to predict the ~~etiology~~ cause of the LVH, including predictions for cardiac amyloidosis and aortic stenosis among a background of other hypertrophic diseases.

«RK21: Methods»

Commented [RK21]: AU: Were the supplemental methods (now "eMethods" use in this study? If so, please add a mention of "eMethods in the Supplement" to an appropriate place in the text. If these methods were not used, they should be removed from the supplement.

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, ~~2D~~ 2-dimensional images, tissue Doppler ~~-~~images, and color Doppler images~~, and others~~). In this cohort study, ~~Patients~~ patients were identified ~~by~~ 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 **«SB22:** from X

X, XXXX, to X X, XXXX». «RK23: This research was approved by the Stanford University and CMSC institutional review boards. »«SB24: This study followed the [guideline].»

~~In this study, relevant~~ Relevant parasternal long-axis (PLAX) and apical 4-chamber ~~(A4C)~~ 2~~-D~~ dimensional videos were extracted from each echocardiogram study (Table~~s~~ 1 and ~~2~~Table 2). Human clinician annotations of intraventricular septum (IVS), ~~left ventricular~~LV internal dimension (LVID), and ~~left ventricular~~LV posterior wall (LVPW) «SB25: measurements »were used as training labels to assess ventricular hypertrophy. «RK26: Parasternal long-axis videos obtained ~~From~~ from Stanford Health Care (SHC)~~, PLAX videos~~ were split and used as follows: 9600 for the training set, 1200 for the validation set, and 1200 ~~patients, respectively,~~ for ~~the training, validation, and~~ the test set~~s~~. »An additional 7767 SHC echocardiogram studies were ~~of~~ from patients with defined disease characteristics, including cardiac amyloidosis, ~~hypertrophic cardiomyopathy~~HCM, and severe aortic stenosis. From these studies, the ~~A4C~~ apical 4-chamber videos were extracted and used as input data for the hypertrophic disease classification task. Videos were processed in a

Commented [SB(22]: Please give the study dates.

Commented [RK23]: AU: Please note that this information has been moved here to meet journal style requirements. Please provide a statement of informed consent. If obtained, was it written or oral? If waived, please indicate by what institution and why. Were data deidentified?

Commented [SB(24]: Was a study guideline followed?

Commented [SB(25]: Correct?

Commented [RK26]: AU: Changes OK?

previously described automated preprocessing workflow ~~removing~~ in which identifying information and human labels had been removed.[21,25] ~~This research was approved by the Stanford University and Cedars-Sinai Medical Center Institutional Review Boards.~~

Domestic and International External Health Care System Test Data Sets

Transthoracic echocardiogram studies from CSMC and the Unity Imaging Collaborative (UIC)[16] were used to evaluate ~~our~~ 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 ~~(DICOM)~~ files to AVI files.[22] ~~In total, we~~ We extracted a total of 3660 ~~total~~ videos from CSMC as a domestic held-out test data set. Labeled **«SB27:** echocardiograms ~~images~~ **»**from the ~~Unity Imaging Collaborative~~UIC were used as an additional held-out international test data set not seen during model training. These echocardiogram videos were obtained from the British echocardiography ~~labs~~laboratories~~,~~ and were retrospectively annotated by echocardiography-certified cardiologists.

Commented [SB(27]: Were these videos or images?

Deep Learning Algorithm Development and Training

Model design and training ~~was~~ were done in Python, **«SB28:** version X **»**(Python Software Foundation) using the PyTorch deep learning library. A modified DeepLabv~~V~~3[26] 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. ~~3D~~ Three-dimensional implementations of a segmentation model took substantially more computational resources without ~~significant~~ marked improvement in performance. An Adam optimizer ~~was used~~ 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 ResNet3D[27] architecture was used to classify videos. Given the potential for **«SB29:** overlap patients **»**with multiple ~~etiological diagnoses for~~causes of LVH,[28] parallel binary classification deep learning models were trained to predict probability of **«SB30:** amyloid**»**, ~~hypertrophic cardiomyopathy~~HCM,

Commented [SB(28]: Please give the version of Python.

Commented [SB(29]: Please clarify what is meant by “overlap patients”?

Commented [SB(30]: Should this be amyloidosis?

«RK31: and aortic stenosis~~,~~ secondary to uncontrolled hypertension~~,~~ and in the ~~setting~~ »context of end-stage kidney disease ~~(ESRD)~~ independently. Distinct from ~~prior~~ «SB32: previous literature», 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 «SB33: images from »patients with diagnosed ~~HCM~~HCM, aortic stenosis, hypertension, «RK34: and ~~in setting of ESRD~~ end-stage kidney disease» as other ~~etiologies~~ 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.

Commented [RK31]: AU: Changes OK?

Commented [SB(32]: Please add citations.

Commented [SB(33]: Correct?

Commented [RK34]: AU: OK as edited?

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 ~~significant~~ substantial change from the current study to the ~~prior~~ 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 ~~was~~ were measured for both the current and the subsequent ~~study~~studies. The variance in measurement between the previous and subsequent ~~study~~ studies was used as a surrogate of clinician variation and was compared with the variation of ~~our~~ the deep learning algorithm. ~~On~~ **«SB35:** In the CSMC data set, we identified 99 random studies; ~~and performed~~ blinded relabeling was performed by 2 ~~level~~ level ~~III~~ 3 echocardiography-certified cardiologists, and ~~compared~~ their performance was compared ~~against~~ with the performance of ~~our~~ the deep learning algorithm on the consensus label**»**.

Commented [SB(35]: Is this sentence accurate as edited? Were the cardiologists authors of this study?

Statistical Analysis

~~Confidence intervals~~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 ~~to~~ with human labels in the hold-out test data set. The centroid of each predicted key point was used to calculate **«SB36:** measurements.**»**

Commented [SB(36]: Please give the name, version, and manufacturer of the software used for statistical analysis.

Results

~~Our~~ **«SB37:** The study included 13 310 patients (X [X%] male; mean [SD] age, X [X] years).**»** The deep learning workflow for screening of ~~hypertrophic cardiomyopathy~~HCM and cardiac amyloidosis ~~has~~ had 2 components (Figure 1). First, we designed a deep learning model with atrous convolutions for semantic segmentation of ~~parasternal long axis (~~PLAX~~)~~ echocardiogram videos and identification of the ~~intraventricular septum~~ ~~(~~IVS~~)~~, ~~left ventricular~~LV internal dimension ~~(LVID)~~, and ~~left ventricular~~LV posterior wall ~~(LVPW)~~. 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, ~~often~~ only ~~labels~~ 1 or 2 frames of a video are often labeled, ~~while~~ 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.

After detection of LVH, identifying the specific ~~etiology~~ cause (eg, infiltrative disease, inherited cardiomyopathies, or chronic elevated

Commented [SB(37]: Is this the correct total sample size? Please provide demographic information.

afterload) can help guide therapy. We trained a video-based «RK38: convolutional neural network »~~CNN~~ model with spatiotemporal convolutions to predict ~~etiology~~ the cause of LVH (Figure 3). Integrating spatial ~~as well as~~and temporal information, ~~our~~ the model ~~expands~~ expanded ~~up~~on ~~our~~ «RK39: ~~prior~~ previous work »with video-based model interpretation of echocardiograms and ~~classifies~~ classified videos based on probability of hypertension, aortic stenosis, ~~hypertrophic cardiomyopathy~~HCM, or cardiac amyloidosis as ~~etiology~~ causes of ventricular hypertrophy. ~~We additionally~~In addition, we performed a video-based model architecture and hyperparameter search to identify the optimal base architecture for ~~our~~ the deep learning algorithm (eFigure 1 in the Supplement). ~~Our~~ The deep learning algorithm was trained on a data set of 17 802 echocardiogram videos from ~~Stanford Health Care (~~SHC~~)~~ and then evaluated on held-out test cohorts from SHC, ~~Cedars-Sinai Medical Center (~~CSMC~~)~~, and ~~Unity Imaging Collaborative~~UIC.

Commented [RK38]: AU: Is expansion of CNN correct?

Commented [RK39]: AU: Please provide reference citation. The journal only uses “our” when the byline of the reference matches the byline of the present study.

Evaluation of Hypertrophy Detection

From the held-out test data set from SHC ~~(n = 1200)~~ not seen during model training (n = 1200), ~~our~~ the deep learning algorithm predicted ventricular dimensions with an ~~(~~R^2~~)~~ of 0.97 compared ~~to~~ with annotations

by human experts (eFigure 2 in the Supplement). ~~Our~~ The deep learning algorithm had a mean absolute error (MAE) of **«SB40:** 1.2 mm for IVS thickness, 2.4 mm for LVID, and 1.4 mm for LVPW**»** thickness. This ~~compares~~ compared favorably with clinical inter-provider variation, which had an MAE of **«SB41:** 1.3 mm for IVS thickness, 3.7 mm for LVID, and 1.3 mm for LVPW**»** thickness. **«SB42:** EchoNet-LVH **»**also performed well ~~when~~ compared ~~to~~ with the prospective consensus annotation of 2 level -3 echocardiography-certified cardiologists in 99 random studies from CSMC (eFigure 3 in the Supplement). **«RK43:** To assess the reliability of the model across- health care- systems ~~and~~ internationally ~~reliability of the model~~, ~~our~~ the deep learning algorithm**»** was ~~additionally~~ also tested~~,~~ without any tuning~~,~~ on an external test data set of 1791 videos from ~~Unity~~ UIC~~Imaging Collaborative~~ and 3660 videos from CSMC. On the ~~Unity~~ UIC external test data set, ~~our~~ the deep learning algorithm showed a robust prediction accuracy, with an overall R^2 of 0.90~~,~~ and **«SB44:** MAEs of 1.6 mm for IVS thickness, 3.6 mm for LVID, and 2.1 mm for LVPW thickness.**»** ~~Highlighting data shift and potential variations in practice across institutions and continents, w~~When fine-tuned using the training split of the

Commented [SB(40]: Please add 95% CIs. Correct to add thickness?

Commented [SB(41]: Please add 95% CIs.

Commented [SB(42]: Please add a sentence to the Methods describing EchoNET-LVH to give background to this result.

Commented [RK43]: AU: Changes OK?

Commented [SB(44]: Please provide 95% CIs.

~~Unity~~ UIC data set, ~~our~~ the deep learning algorithm showed an improved performance with an overall R^2 of 0.92~~,~~ and ~~median absolute error~~MAEs of **«SB45:** 1.1 mm for IVS thickness, 1.7 mm for LVID, and 1.5 mm for LVPW**»** thickness on the ~~Unity~~ UIC validation data split, **«SB46:** indicating data shift and potential variations in practice across institutions and continents **»**(eTable 1 in the Supplement).

A rapid, high-throughput automated approach ~~allows~~ allowed for measurement of every individual frame, ~~that~~ 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~~ higher-fidelity overall assessments. ~~While~~ Although the SHC and ~~Unity~~ UIC data sets were directly compared on annotated individual frames, we evaluated ~~our~~ the deep learning algorithm's beat-to-beat evaluation on the CSMC data set ~~in comparison~~ 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 **«SB47:** ~~average~~ mean **»**measurements from each heartbeat ~~across~~ throughout the entire video. On the CSMC external test data set, ~~our~~ the

Commented [SB(45]: Please add 95% CIs.

Commented [SB(46]: Correct as edited?

Commented [SB(47]: OK to change "average" to "mean"?

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 ~~Etiology~~ Cause of ~~Hypertrophy~~LVH

The ~~etiology~~ cause derivation, validation, and test cohorts from SHC had 6215, 787, and 765 videos, respectively. ~~On~~ In the held-out test cohort, ~~our~~ the deep learning algorithm ~~classifies~~ «RK48: distinguished» cardiac amyloidosis with an AUC of 0.83, ~~hypertrophic cardiomyopathy~~HCM with an AUC of 0.98, and aortic stenosis with an AUC of 0.89 from other ~~etiologies~~ causes of LVH. On the held-out test cohort, the area under the precision-recall curve ~~(AUPRC)~~ of ~~our~~ the deep learning algorithm was 0.77 for cardiac amyloidosis ~~was 0.77~~, ~~hypertrophic cardiomyopathy was~~ 0.95 for HCM, and 0.79 for aortic stenosis ~~was 0.79~~. The proposed ensemble of binary classification video-based deep learning classifiers in ~~our~~ the deep learning algorithm was similar in performance to a multilabel, multiclass deep learning model for disease detection~~, however~~ but had the flexibility of being able to identify «RK49: overlap patients »with multiple diagnoses. ~~On~~ In an external test data set of 2351 ~~A4c~~ apical 4-chamber–view videos

Commented [RK48]: AU: Change OK?

Commented [RK49]: AU: As above, please clarify what is meant by "overlap patients."

from CSMC with 358 videos of cardiac amyloidosis, 146 videos of aortic stenosis, 468 videos of ~~hypertrophic cardiomyopathy~~HCM, and 1379 videos of other ~~etiologies~~ causes of LVH, ~~our~~ the deep learning algorithm had an AUC of 0.79 for predicting cardiac amyloidosis and an AUC of 0.89 for predicting ~~hypertrophic cardiomyopathy~~HCM. ~~On~~ For the CSMC cohort, the area under the precision-recall curve ~~AUPRC~~ of ~~our~~ the deep learning algorithm ~~for~~ was 0.54 for cardiac amyloidosis ~~was 0.54~~, 0.69 for ~~hypertrophic cardiomyopathy was 0.69~~HCM, and 0.08 for aortic stenosis ~~was 0.08~~. 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 ~~etiologies~~ 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)~~,~~ ~~where~~ in which a higher AUC outside the diagonal ~~shows~~ showed the model confusion and a lower AUC ~~suggests~~ suggested improved discrimination between phenotypic mimics. In this experiment, ~~while~~ although ~~such a~~the model ~~can~~ produced a **«SB50:**

higher» AUC (~~AUC of~~ 0.96 for cardiac «SB51: amyloid», 0.98 for ~~AS~~aortic stenosis, and 0.97 for HCM), there was ~~significant~~ substantial confusion when other causes were ~~introducing~~ introduced~~other etiologies~~, suggesting that a model trained only on age- and sex-matched controls would primarily identify~~ies~~ «SB52: hypertrophy».

Commented [SB(50]: OK? If not "higher" than which comparator?

Commented [SB(51]: Should this be amyloidosis?

Commented [SB(52]: LVH?

Discussion

~~Our~~ The ~~artificial intelligence~~AI-guided workflow used in this cohort study ~~is~~ was a deep learning system that automatically ~~quantifies~~ quantified ~~left ventricular wall thickness~~LV wall thickness on echocardiography while also predicting the cause of ~~etiology of~~ LVH as ~~attributable to~~ either ~~hypertrophic cardiomyopathy~~HCM or cardiac amyloidosis. The deep learning algorithm ~~performs~~ performed measurements of ventricular thickness and diameter well within the variance of human clinical test-retest assessment~~–~~ while ~~concurrently~~ aiding the detection of subtle ventricular phenotypes that tend to be challenging for human readers. This integration of ~~left ventricular~~LV measurement and prediction of ~~etiology~~ 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.

~~Rather than rare, there is reason to believe that~~ **«SB53:** ~~d~~Studies[28-30] have suggested that diseases **»**such as cardiac amyloidosis are underdiagnosed rather than rare.[28-30] ~~Particularly g~~Given the large heterogeneous population of patients with heart failure with preserved ejection fraction,[31] ~~methodologies~~ methods that might appropriately ~~as well as~~and efficiently increase suspicion for under-recognized ~~etiologies~~causes, such as subtypes of amyloidosis with newly available targeted therapies, ~~can~~ may help address a persistent unmet need. Accordingly, an ~~intriguing~~ opportunity exists in~~is~~ the application of efficient AI algorithms to increase recognition of historically underdiagnosed disease~~s conditions across~~ in stored images in databases of large echocardiography laboratories. Notwithstanding ~~the fact~~ 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 ~~across~~ in large populations. As such,

Commented [SB(53]: OK as edited?

more prospective work is needed to evaluate the potential of such algorithms to expedite appropriate clinical evaluation, targeted testing, and confirmation ~~prior to~~before eventual diagnosis and initiation of disease-modifying therapy.

«SB54: 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 work[8] assessing cardiac function, **«SB55:** the present study showed that **»**deep learning models on echocardiogram images can automate an increasingly larger proportion of tasks for assessing cardiac form and structure to provide more holistic evaluation of cardiovascular disease. With improved precision to detect ventricular remodeling and cardiac dysfunction,

Commented [SB(54]: The Strengths and Limitations section should appear directly before the conclusions; thus, this was moved up.

Commented [SB(55]: Correct?

AI systems offer the potential for earlier detection and treatment of subclinical cardiovascular disease, including less common or underdiagnosed conditions.

«RK56: Strengths and Limitations»

~~Notwithstanding the imitations, this~~This study ~~offers~~ 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 ~~all~~ sources of data set shift that limit the direct comparison of model performance.[34,35] With fine-tuning on site-specific data, our model compares favorably with **«SB57:** prior state-of-the-art approaches to assessing ventricular wall thickness and hypertrophy on open benchmarks**»**.

Expanding **«RK58:** ~~our~~ on ~~prior~~ previous work, **»**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 ~~our~~ the prior data set of 10 030 apical 4-chamber videos[21] to a total of 22 030 echocardiogram videos made publicly

Commented [RK56]: AU: Head has been added and order of paragraphs has been changed to meet journal style requirements of listing strengths first.

Commented [SB(57]: Please add citations.

Commented [RK58]: AU: Please provide reference citation.

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 ~~have~~ also released the full code for our algorithm and data-processing workflow.

This study also has ~~Several~~ limitations ~~of our analysis merit consideration~~. First, ~~since~~ 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 **«SB59:** Black ~~Americans~~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~~s~~ ~~and~~ characteristics.[32,33] 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 ~~ultrasound~~ 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

Commented [SB(59]: Correct? Or do you mean “African American individuals”?

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.[32] As such, further work ~~in addition to~~and prospective validation ~~is~~ are needed to more fully understand the ~~impact~~ effect of AI-guided screening workflows on clinical care.

~~Notwithstanding the limitations, this study offers several strengths. A key challenge in 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 all sources of data set shift that limit the direct comparison of model performance.[34,35] 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. Expanding our prior work, 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 our prior data set of 10 030 apical 4-chamber videos[21] to a total of 22 030 echocardiogram videos made publicly available, 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 have also released the full code for our algorithm and data processing workflow.~~

Conclusions

«SB60: In this cohort study, using **»**measurements across multiple heart beats, the deep learning model more accurately identified subtle changes in LV wall geometric measurements and their causes. Unlike with human experts, the deep learning workflow is fully automated, allowing for reproducible, precise measurements, and may provide the a foundation for precision diagnosis of cardiac hypertrophy.

Commented [SB(60]: A conclusions section was added. Please confirm or suggest revisions.

Article Information

Accepted for Publication: December 21, 2021.

Corresponding Authors: Susan Cheng, MD, MPH (susan.cheng@cshs.org), and David Ouyang, MD (david.ouyang@cshs.org**«RK61:**), Department of Cardiology, Smidt

Heart Institute, Cedars-Sinai Medical Center, »127 S San Vicente Blvd, Ste A3600, Los Angeles, CA 90048.

«MM62: 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. **«RK63:** Author x **»**had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

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.

Commented [RK61]: AU: The address must match an affiliation. Has the correct affiliation been added for Dr Ouyang?

Commented [MM62]: Each author certifies their specific contributions on their individual Authorship Form. Changes should not be made at this stage. If you require any changes, we will need a complete explanation from you and will need to unschedule your article from publication and request new Authorship Forms.

Commented [RK63]: AU: Please give 1-2 authors that should be added here,

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

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

«JN64: Conflict of Interest Disclosures:» Dr Duffy reported having a **«SB65:** patent ~~00000~~ pending Cedars**»**. Dr Kwan reported receiving grants from the Doris Duke Charitable Foundation during the conduct of the study. Dr Alexander reported receiving ~~other~~ consulting fees from Alnylam ~~Consulting fees~~, ~~other from~~ Eidos ~~Consulting fees~~, and ~~other from~~ Pfizer ~~Consulting fees~~ outside the submitted work. Dr Lungren reported receiving personal fees from Nines Radiology, ~~personal fees from~~ Philips Healthcare, ~~personal fees~~and ~~from~~ SegMed~~,~~ ; **«RK66:** other from **»**Centaur Equity~~,~~ ; and grants from the National Institutes of Health outside the submitted work. Dr Rader reported receiving ~~other~~ consulting fees from Medtronic ~~consulting fees~~, ~~other from~~ Recor ~~consulting fees~~, and ~~other from~~ Bristol -Myers ~~-~~ Squibb ~~consulting fees~~ outside the submitted work. Dr Ashley reported being a founder of ~~other from~~ Personalis Inc~~. (founder)~~, ~~other from~~ DeepCell Inc~~. (founder)~~, ~~other from~~ and Silicon Valley Sports Analytics ~~(founder);~~ ~~,~~ ~~other from Nuevocor Founding (advisor),~~ **«SB67:** other from **»**Astra Zeneca; ~~(non-executive director), other from~~ serving as an advisor for

Commented [JN64]: Please carefully review the conflicts of interest disclosures and approve or update.

Commented [SB(65]: Please clarify "patent Cedars" and indicate what the patent is for.

Commented [RK66]: AU: Please specify what other means.

Commented [SB(67]: Please specify other.

Nuevocor Founding, Cathay ~~(capital advisor)~~, ~~other from~~ Third Rock Ventures ~~(advisor)~~, ~~other from~~ Medical Excellence ~~(capital advisor)~~, ~~other from~~ Foresite ~~(capital advisor)~~, ~~other from~~and Novartis ~~(advisor,)~~, Genome Medical, Disney, and Sequence Bio; ~~other~~ receiving grants from Bristol Myers Squibb, ~~Academic grant, other from~~ Takeda, Google, and Verily ~~Academic grant,~~; receiving ~~other~~ «SB68: ~~from~~support in kind from »Samsung Collaborative, ~~support in kind, other from~~ Analog Devices Collaborative, Illumina Collaborative, PacBio Collaborative, and Nanopore Collaborative; and ~~support in kind,~~ «SB69: other »from Apple Advisor ~~, ;~~ ~~other from Google Grant, other from Verily Grant,; other from Illumina Collaborative, support in kind, other from PacBio Collaborative; support in kind, other from Nanopore Collaborative support in kind,; other from Genome Medical (advisor), other from Disney (advisor), and other from Sequence Bio (advisor)~~ outside the submitted work. Dr Witteles reported receiving personal fees from Pfizer, ~~personal fees from~~ Alnylam, ~~personal fees from~~ Eidos, ~~personal fees from~~ Ionis, and ~~personal fees from~~ Intelia outside the submitted work. Dr Ouyang and Dr Cheng reported having a patent pending for EchoNet-LVH ~~pending~~. No other disclosures were reported.

Commented [SB(68]: Please specify the type of support.

Commented [SB(69]: Please specify other.

Funding/Support: «**SB70:** This study was supported by grants K99-HL157421 (Dr Ouyang) and 5T32HL116273-07 (Dr Yuan) ~~is supported by~~from the National Institutes of Health~~»grants K99-HL157421, Dr Yuan is supported by 5T32HL116273-07~~.

Role of the Funder/Sponsor: «**RK71:** ~~[Name]~~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.»

Commented [SB(70]: Were both grants from the NIH? Were these grants awarded specifically for the conduct of the study?

Commented [RK71]: Please confirm or revise.

Additional Information: ~~This project introduces the~~The EchoNet-LVH ~~Dataset~~dataset, a publicly available dataset of de-identified echocardiogram videos, is available at~~:~~ «**e72:** https://echonet.github.io/plax/». ~~Code Availability:~~ The code for EchoNet-LVH is available at~~:~~ ~~https://github.com/echonet/lvh~~https://github.com/echonet/lvh.

Commented [e72]: The URL does not work. Please provide a working URL.

References

1. Mardis ER. A decade's perspective on DNA sequencing technology. *Nature*. 2011;470(7333):198-203. Medline:21307932 doi:10.1038/nature09796

2. National Human Genome Research Institute, National Institutes of Health. DNA sequencing costs: data. «RK73: Accessed month day, year.» 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. Medline:8598840 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. Medline:19584346 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*.

Commented [RK73]: AU: Please supply date accessed.

2021;12(1):2725. Medline:33976166 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. Medline:25398313 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. Medline:33976142 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. Medline:30354459 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. doi:10.1016/S0735-1097(21)01880-5

10. Lee GY, Kim K, Choi JO, et al. Cardiac amyloidosis without increased left ventricular wall thickness. *Mayo Clin Proc*. 2014;89(6):781-789. Medline:24702732 doi:10.1016/j.mayocp.2014.01.013

11. Pagourelias 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. Medline:28298286 doi:10.1161/CIRCIMAGING.116.005588

12. Elliott PM, Anastasakis 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. Medline:25173338 doi:10.1093/eurheartj/ehu284

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. Medline:28363351 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. Medline:17060807 doi:10.2459/01.JCM.0000250869.78777.09

15. 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. Medline:33735065 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. Medline:33998247 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/ASE/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. Medline:21338862

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. Medline:25344557 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. Medline:20152692 doi:10.1016/j.echo.2009.11.016

20. Farsalinos KE, Daraban AM, Ünlü 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. Medline:32269341 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. Medline:31015713 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. Medline:31993508 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. Medline:32393799 doi:10.1038/s41591-020-0870-z

25. Github. ConvertDICOMToAVI.ipynb at master echonet/dynamic. «SB74: Accesed X X, XXXX. [URL]»

Commented [SB(74]: Please provide a URL and the date you last accessed the site.

26. Yurtkulu SC, Şahin YH, Unal G. Semantic segmentation with extended DeepLabv3 architecture. Paper presented at: 27th Signal Processing and Communications Applications Conference. **«SB75:** 2019:1-4**»**.

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. Medline:29019612 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

Commented [SB(75]: Has this been published? If so, please provide the name of the journal in which is was published. If not, please give the location of the meeting and the date of the presentation.

study. *Ann Med*. 2008;40(3):232-239. Medline:18382889 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. Medline:27143678 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. Medline:32202936 doi:10.1161/CIRCULATIONAHA.119.041886

32. Wu E, Wu K, Daneshjou 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. Medline:33820998 doi:10.1038/s41591-021-01312-x

33. 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.
34. Quiñonero-Candela J, Sugiyama M, Schwaighofer A, Lawrence ND, eds. *Dataset Shift in Machine Learning*. The MIT Press; 2008.
35. 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.

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

A. For each patient, ~~our~~ the deep learning algorithm ~~uses~~ used parasternal long-axis echocardiogram video as input to derive key points and precisely estimate ventricular dimensions. After identifying patients with left ventricular hypertrophy (LVH), ~~our~~ the deep learning workflow used a video-based architecture to distinguish ~~between~~ common causes of LVH. B. Correlation of human annotations vs

model predictions for ventricular dimensions in data sets from 3 health care systems (n = 1200 for Stanford Health Care [SHC], n = 1309 for Cedars-Sinai Medical Center [CSMC], and n = 1791 for Unity Imaging Collaborative). C. Model variation on data sets from 3 health care systems compared ~~to~~ with human clinical annotation variation. The ~~Boxplot represents the mean as a~~ middle lines represent mean values; ~~thick line, 25th and 75th percentiles as~~ upper and lower bounds of the boxes, 25th and 75th percentiles; and individual points, ~~for instances~~values greater than 1.5 times the ~~interquartile range~~IQR ~~from~~ of the mean. D. Receiver operating characteristic curves for diagnosis of amyloidosis ~~on~~ in the ~~Stanford~~ SHC validation (n = 813) and test (n = 812) data sets. AUC indicates area under the curve; HCM, hypertrophic cardiomyopathy IVS, intraventricular septum; LVID, left ventricular internal dimension; LVPW, left ventricular posterior wall.

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

A. Model prediction of key points ~~(in green, purple, blue, and yellow)~~ on an individual frame of parasternal long-axis video. B. Frame-by-frame prediction of wall thickness and ventricular dimension and automated detection of systole and diastole allowing for beat-to-beat

prediction of ventricular hypertrophy. C. Waterfall plot of individual video variation in beat-to-beat evaluation of ventricular hypertrophy (n = 2320) ~~across~~ in the internal test data set. Each video is represented by multiple points along a line representing the measurement of each beat and a line signifying the range of predictions. IVS, indicates intraventricular septum; IVSd, intraventricular septum (diastole); LVID, left ventricular internal dimension; LVIDd, left ventricular internal dimension (diastole); LVPW, left ventricular posterior wall.

Figure 3. Performance of Disease ~~Etiology~~ Cause Classification Across 2 Independent Domestic Institutions

A. Receiver operating characteristic curves for detection of cardiac amyloidosis and aortic stenosis in the Stanford Health Care (SHC) internal test data set (n = 765) and Cedars-Sinai Medical Center (CSMC) external test set (n = 2351). B. Representative images for selected cases and controls for each ~~etiology~~cause. C. Precision-recall curves for detection of amyloidosis and hypertrophic cardiomyopathy (HCM) in the SHC test data set (n = 765). **«RK76:** AP indicates anteroposterior**»**; AUC, area under the curve.

«SB77: Video 1. [title]**»**

Commented [RK76]: AU: Is expansion of AP correct?

Commented [SB(77]: (1) For each video, please provide a title.
(2) Please indicate where each video should be cited in the main text.

Video 2. [title]

Video 3. [title]

Video 4. [title]

Video 5. [titles]

Table 1. Baseline Characteristics of Patients With Parasternal Long- Axis «SB78: Videos of Echocardiograms »From Stanford Health ~~care~~ Care and CSMC[a] ~~Cedars-Sinai Medical Center Echocardiograms~~

Characteristic	Stanford Health Care				CSMC ~~T~~test set ~~(CSMC)~~ (n = 1309)
	Training set ~~(Stanford)~~ (n = 9601)	Validation set (n = 1200) ~~(Stanford)~~	Test set (n = 1200) ~~(Stanford)~~	Total (N = 12 001) ~~(Stanford)~~	
Age, mean (SD), y	61.60 (17.35)	61.29 (17.79)	61.73 (17.47)	61.58 (17.41)	62.80 (17.16)
Sex					
«RK79: Female					»
Male	4333 (53.3)	582 (56.0)	577 (56.6)	5492 (53.9)	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)

Commented [SB(78]: OK as edited?

Commented [RK79]: AU: Please supply number and percentage of female patients included in each set.

«SB80: 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 ~~mellitus~~	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; LVH, left ventricular hypertrophy; LVIDd, left ventricular

Commented [SB(80]: Please indicate the racial and ethnic groups included as other. If not available, please indicate why.

internal dimension (diastole); LVIDs, left ventricular internal dimension (systole); LVPWd, left ventricular posterior wall ~~thickness~~ (diastole).

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

[b]Other racial and ethnic groups included.

Table 2. Baseline Characteristics of Patients With Apical 4-Chamber Videos of Echocardiograms From Stanford Health ~~care~~ Care and ~~Cedars-Sinai Medical Center~~CSMC[a] ~~Echocardiograms~~

Characteristic	Stanford Health Care				CSMC test set (n = 2351)
	Training set (n = 6461)	Validation set (n = 814)	Test set (n = 809)	Total (N = 8084)	
«SB81: 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	68.97 (16.90)	70.01 (16.63)	69.42 (15.95)	69.12 (16.78)	69.59 (14.66)
Female					
Male	3109 (59.5)	400 (59.9)	374 (60.5)	3883 (59.6)	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)
«SB82: 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)

Commented [SB(81]: Correct?

Commented [SB(82]: Please add the groups included as Other to the footnote. If unavailable, give the reason why.

Hypertension	3657 (82.0)	473 (82.8)	423 (80.7)	4553 (82.0)	1290 (54.9)
Diabetes ~~mellitus~~	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; LVH, left ventricular hypertrophy; LVIDd, left ventricular internal dimension (diastole); LVIDs, left ventricular internal dimension (systole); LVPWd, left ventricular posterior wall ~~thickness~~ (diastole).

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

[b]Other racial and ethnic groups included.

Supplement. **eMethods.**

eFigure 1. Hyperparameter Search of Video-based Classification Model

eFigure 2. Comparison of Model Performance With Human Variation

eFigure 3. Comparison of Model Performance With Prospective Consensus Annotation of Two Level III Echocardiography Certified Cardiologists

eTable 1. Performance on Unity Imaging Collaborative External Test Dataset With and Without fine-tuning