PanEcho: Complete AI-enabled echocardiography interpretation with multi-task deep learning

Gregory Holste MSE[1,2]*, Evangelos K. Oikonomou MD DPhil[2]*, Zhangyang Wang PhD[1], Rohan Khera MD MS[2-5]

[1] Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA

[2] Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

[3] Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

[4] Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA

[5] Section of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA

*Contributed equally as co-first authors

Address for correspondence:

Rohan Khera MD, MS

195 Church St, 6th Floor, New Haven, CT 06510

203-764-5885; rohan.khera@yale.edu; @rohan_khera

ABSTRACT

Echocardiography is a mainstay of cardiovascular care offering non-invasive, low-cost, increasingly portable technology to characterize cardiac structure and function[1]. Artificial intelligence (AI) has shown promise in automating aspects of medical image interpretation[2,3], but its applications in echocardiography have been limited to single views and isolated pathologies[4–7]. To bridge this gap, we present *PanEcho*, a view-agnostic, multi-task deep learning model capable of simultaneously performing 39 diagnostic inference tasks from multi-view echocardiography. *PanEcho* was trained on >1 million echocardiographic videos with broad external validation across an internal temporally distinct and two external geographically distinct sets. It achieved a median area under the receiver operating characteristic curve (AUC) of 0.91 across 18 diverse classification tasks and normalized mean absolute error (MAE) of 0.13 across 21 measurement tasks spanning chamber size and function, vascular dimensions, and valvular assessment. *PanEcho* accurately estimates left ventricular (LV) ejection fraction (MAE: 4.4% internal; 5.5% external) and detects moderate or greater LV dilation (AUC: 0.95 internal; 0.98 external) and systolic dysfunction (AUC: 0.98 internal; 0.94 external), severe aortic stenosis (AUC: 0.99), among others. *PanEcho* is a uniquely view-agnostic, multi-task, open-source model that enables state-of-the-art echocardiographic interpretation across complete and limited studies, serving as an efficient echocardiographic foundation model.

INTRODUCTION

Echocardiography is one of the pillars of modern cardiovascular diagnostics thanks to its low cost, broad accessibility, and ability to provide in-depth phenotyping of cardiac, valvular, and vascular structure and function[1]. More than 7.5 million echocardiographic studies are performed every year in the United States alone, and increasing referrals for echocardiography are contributing to rising healthcare expenditures across most nations[8,9]. Accurate reporting of echocardiography requires time, skilled acquisition, and expert readers, and is frequently noted to be subject to inter-rater variability[10,11]. Artificial intelligence (AI) algorithms have shown promise in automating various aspects of this process, from detecting valvular abnormalities[7,12–14] to quantifying key measurements such as the left ventricular (LV) ejection fraction (EF)[4,15–19], among others[20–25]. However, existing solutions typically rely on curated single-view inputs and are limited to single tasks[4–7,14,22,24–26]. This process is discordant with echocardiographic interpretation in real-world practice, in which multiple views and imaging modes, such as color Doppler imaging, are integrated to form a comprehensive evaluation, spanning functional and structural metrics of all major chambers, valves, and vessels. Versatile AI systems that handle this multi-view, multi-task workflow would enable efficient, reader-independent phenotyping of echocardiographic studies but are currently lacking.

To bridge this gap and provide a scalable solution for fully automated echocardiographic interpretation, we present *PanEcho*, an end-to-end, view-agnostic deep learning model capable of simultaneously performing 39 key echocardiographic reporting tasks. Our model was trained on over one million standard 2D B-mode and color Doppler echocardiogram videos from all views to perform a diverse mixture of 18 classification and 21 continuous regression tasks, spanning the full spectrum of structural and functional myocardial and valvular parameters. The

model demonstrated excellent predictive performance across hospital systems and under both complete and abbreviated imaging protocols, enabling flexible inference competitive with existing single-view methods dedicated to individual tasks. Further, through its unique multi-view training, *PanEcho* enables interpretable predictions by correctly identifying the echocardiographic views and imaging modes most relevant for each task. Finally, *PanEcho* exhibits robust transfer learning capabilities, outperforming other methods in both predictive performance and training efficiency when fine-tuned for downstream quantification tasks, such as EF estimation in both adult and out-of-domain pediatric populations. Given the increasing accessibility of portable ultrasound technology in point-of-care settings[27,28], *PanEcho* has the potential to enable complete AI-assisted echocardiography screening even with abbreviated imaging protocols and variable acquisition quality. Our method is the first multi-view and multi-task AI model for echocardiography, and we publicly release the model weights and source code to accelerate research on AI-enabled echocardiographic interpretation.

RESULTS

Multi-task deep learning model development

PanEcho is a view-agnostic, multi-task deep learning model for comprehensive automated interpretation of multi-view transthoracic echocardiography (**Fig. 1**). Our model can simultaneously perform 39 core echocardiographic reporting tasks and consists of (i) a two-dimensional (2D) image encoder, (ii) a temporal frame Transformer, and (iii) task-specific output heads. First, the 2D image encoder, a convolutional neural network (CNN), learns embeddings of individual echocardiographic video frames. Second, the frame-wise embeddings become inputs to a Transformer, which models temporal patterns across the frames within a video and outputs a

pooled video-level representation. Third, this video embedding is used as input to task-specific output heads to simultaneously perform a wide variety of classification and regression tasks. Finally, predictions are compared with the ground truth to compute task-specific losses, which are aggregated into a multi-task objective that the model learns to minimize. *PanEcho* is trained to perform 21 regression tasks (e.g., EF estimation) and 18 classification tasks (e.g., detecting valvular stenosis) from individual echocardiographic videos, with view-specific information aggregated to form study-level predictions that integrate multi-view information for each task.

This work leveraged 1.23 million echocardiographic videos comprising multiple views from 33,927 transthoracic echocardiography studies of 26,067 unique patients across five hospitals and a network of outpatient clinics affiliated with the Yale-New Haven Health System (YNHHS) during 2016-2022 as a part of routine clinical care. Using our previously published pipeline[7], echocardiographic videos were de-identified before being processed by a pretrained view classifier[15] to determine the echocardiographic view and whether color Doppler imaging was used. *PanEcho* was trained on a random partition of 1.03 million YNHHS echocardiograms acquired from January 2016 to June 2022 and internally evaluated on a temporally held-out test set of data from July to December 2022, with no patient overlap across the two sets. The **Methods** and **Extended Data Table 1** contain detailed descriptions of the data processing and YNHHS cohort, respectively.

Automated echocardiography interpretation performance

On a temporally distinct test set of 5,130 echocardiographic studies from YNHHS, *PanEcho* achieved a median area under the receiver operating characteristic curve (AUC) of 0.91 (mean ± standard deviation [sd]: 0.90 ± 0.06) across all 18 classification tasks (**Fig. 2a, Extended Data**

Table 2). The model accurately assessed ventricular structure and function, with temporally valid AUCs of 0.95 for moderate or greater increased LV size, 0.98 for moderate or greater LV systolic dysfunction, 0.93 for moderate or greater LV diastolic dysfunction, 0.91 for moderate or greater LV hypertrophy, 0.88 for any LV wall motion abnormalities, as well as moderate or greater increased right ventricle (RV) size and RV systolic dysfunction with AUCs of 0.87 and 0.93, respectively. *PanEcho* also achieved excellent performance on valvular disease diagnosis, reaching an AUC of 0.99 for severe aortic stenosis and 0.96 for any mitral stenosis, in addition to 0.93 AUC for moderate or greater aortic regurgitation, 0.96 AUC for moderate or greater mitral regurgitation, and 0.89 AUC for moderate or greater tricuspid regurgitation. Additional phenotypes, such as pericardial effusion, and Doppler-derived parameters, such as LV outflow tract (LVOT) obstruction, were classified with AUCs of 0.91 and 0.94, respectively.

Beyond categorical classification, *PanEcho* estimated continuous echocardiographic parameters with a median normalized mean absolute error (MAE) of 0.13 (mean ± sd: 0.14 ± 0.05) across all 21 regression tasks in the YNHHS test set (**Fig. 2b, Extended Data Table 3**). The model accurately quantified LV dimensions and function, with MAE ranging from 4.4% for estimating LVEF to 1.3 mm for LV intraventricular septum thickness (IVSd), 1.2 mm for LV posterior wall thickness (LVPWd), and 3.8 mm for LV internal diameter at diastole (LVIDd). Similarly, for the RV, *PanEcho* estimated RVIDd with 4.0 mm MAE, tricuspid annular plane excursion velocity (TAPSE) with 3.4 m/s MAE, and RV systolic excursion velocity (RV S') with 1.9 cm/s MAE. Atrial dimensions such as LA internal diameter at systole (LAIDs), LA volume, and RA transverse dimension were also estimated with 4.0 mm, 9.4 cm^3, and 4.7 mm MAE, respectively. Finally, *PanEcho* quantified Doppler imaging-derived measurements such as aortic

peak velocity with 0.31 m/s MAE, tricuspid peak gradient with 5.6 mmHg MAE, and E/e' ratio with 1.97 MAE.

To illustrate the versatility of *PanEcho* across imaging protocols, we evaluated its performance in a simulated abbreviated acquisition – increasingly performed at the point of care and on handheld devices[29] – where the model only had access to a single video from each of the following key views per study: parasternal long axis (PLAX), mid-chamber parasternal short axis (PSAX), apical 4-chamber (A4C), apical 5-chamber (A5C), and apical 2-chamber (A2C). *PanEcho* maintained strong predictive performance in this simplified setting, reaching a median 0.85 AUC (mean ± sd: 0.87 ± 0.06) across all classification tasks and 0.14 normalized MAE (mean ± sd: 0.15 ± 0.06) across regression tasks. Detailed results on the YNHHS test set under an abbreviated imaging protocol are depicted in **Extended Data Fig. 1**.

External validation of *PanEcho*

To demonstrate our model's generalizability across geographically distinct cohorts and robustness to varying input views, we evaluated *PanEcho* on a variety of tasks in two large, external echocardiography datasets (**Fig. 3**). First, *PanEcho* maintained strong external performance in assessing LV size and structure in EchoNet-LVH,[6] a dataset of 12,000 PLAX echocardiograms performed at Stanford Health Care. Our model reached an AUC of 0.98 for moderate or greater increased LV size detection and estimated LVID at systole with 3.6 mm MAE and LVID at diastole with 3.8 mm MAE. Regarding LV structure, our model classified moderate or greater increased LV wall thickness with 0.89 AUC and estimated both IVSd and LVPWd with 1.3 mm MAE, consistent with internal validation results. Next, *PanEcho* accurately evaluated LV function in EchoNet-Dynamic,[30] a dataset of over 10,000 A4C

echocardiograms from Stanford University Hospital. Here, *PanEcho* classified moderate or severe LV systolic dysfunction with 0.94 AUC and estimated LVEF with 5.5% MAE. Since both external datasets consisted of single-view echocardiography, all study-level predictions were derived from a single echocardiogram video, unlike during internal evaluation. Full EchoNet-LVH and EchoNet-Dynamic results can be found in **Extended Data Table 4** and **Extended Data Table 5**, respectively.

Analysis of task-specific view relevance

Since *PanEcho* is view-agnostic, its performance when using individual echocardiographic views or imaging modes (color Doppler vs. 2D B-mode) can serve as a proxy for that view's relevance to a given task. To enhance model interpretability, we described the echocardiographic views *PanEcho* learned to be most relevant for each task. We found that its task-specific view relevance scores corresponded to guideline-recommended best practices on characterizing cardiac and valvular structure and function[1] (**Fig. 4**). For instance, in line with standard echocardiographic interpretation, the PLAX view was most informative for LV dimension measurements (IVSd, LVPWd, LVIDs, and LVIDd) as well as aortic valve and aortic root characterization (severe AS classification and aortic root dimension estimation). Similarly, A4C was most informative for estimating LV EF and classifying LV dysfunction, also ranking as one of the top two views for detecting abnormal LV wall thickness and motion. While RV inflow ranked lower than standard apical or parasternal views – focusing on the left ventricle – for most tasks, this view was deemed highly relevant for estimating RV systolic pressure. Similarly, the subcostal view ranked among the least informative for many tasks but proved informative for detecting elevated RA pressure and moderately informative for detecting increased RA size and

estimating RA transverse dimension. Finally, color Doppler videos were the most informative for all valvular regurgitation tasks and highly relevant for abnormalities like valvular stenosis, which often involves assessment with color Doppler imaging. Full task-specific view relevance scores are depicted in **Extended Data Fig. 2**.

Transfer learning capabilities of *PanEcho*

While we have shown that *PanEcho* generalizes "out-of-the-box" across geography and time, we also assess its ability to efficiently transfer knowledge to new echocardiography datasets and tasks via transfer learning. On in-distribution and out-of-distribution regression tasks, *PanEcho* pretraining outperformed other transfer learning and initialization methods in both predictive performance and training efficiency – this included a randomly initialized model, an image-based transfer learning model (ImageNet[31] pretraining), a video-based transfer learning model (Kinetics-400[32] pretraining), and a domain-specific transfer learning model (EchoCLIP[5] pretraining on echocardiographic videos and cardiology reports). Using the official training/validation/test split of EchoNet-Dynamic, a *PanEcho*-pretrained model estimated LV EF with 4.7% MAE after just 2 epochs of fine-tuning, outperforming and converging more rapidly than an identical ImageNet-pretrained model (5.4% MAE; 9 epochs) and randomly initialized model (5.6% MAE; 15 epochs). *PanEcho* pretraining also outperformed a model with a spatiotemporal 3D CNN pretrained on the large-scale Kinetics-400[32] video dataset (5.6% MAE; 5 epochs) and a 2D image encoder pretrained on over one million A4C echocardiograms via EchoCLIP[5] (5.4% MAE; 17 epochs). Detailed EchoNet-Dynamic transfer learning results can be found in **Extended Data Table 6**.

Demonstrating its out-of-distribution transfer abilities, *PanEcho* pretraining also outperformed other initialization strategies on the novel task of pediatric EF estimation from multi-view echocardiography in EchoNet-Pediatric,[17] a dataset of over 7,000 A4C and PSAX echocardiograms. Using the official 10-fold cross-validation splits of EchoNet-Pediatric, a *PanEcho*-pretrained model reached 3.9% MAE on held-out data in 5.5 ± 1.9 epochs (mean ± sd over the 10 folds), again outperforming an identical randomly initialized model (4.9% MAE; 9.6 ± 3.1 epochs) and ImageNet-pretrained model (4.5% MAE; 10.6 ± 3.4 epochs). The *PanEcho*-pretrained backbone also outperformed a domain-specific EchoCLIP-pretrained backbone (5.2% MAE; 12.7 ± 6.0 epochs) as well as a standard 3D transfer learning approach (4.8% MAE; 13.7 ± 5.6 epochs) in terms of both performance and convergence time. See **Extended Data Table 7** for detailed EchoNet-Pediatric transfer learning results.

DISCUSSION

We present *PanEcho*, a view-agnostic deep learning model for automated echocardiography interpretation developed on over one million videos spanning a broad range of views, acquisitions, and patient phenotypes. *PanEcho* advances the current state-of-the-art in AI-enabled echocardiography, enabling flexible estimation of nearly all key parameters of cardiac function and structure from any combination of available views. The method and related algorithm leverage a computationally efficient backbone and a multi-view, multi-task training scheme, allowing their prospective and retrospective deployment across both complete and limited echocardiographic studies. Critically, our model reproduces known patterns in echocardiography reporting by learning to recognize the importance of specific views and modalities for each task. Finally, *PanEcho* exhibits several key properties of a foundation model,

learning powerful representations of echocardiographic videos that efficiently transfer to downstream and even out-of-distribution tasks and populations. The model weights and source code are publicly released in the hope that they will support research teams and investigators in leveraging the power of multi-view and multi-task AI models in echocardiography.

PanEcho was developed to address a critical gap in the field of AI-assisted echocardiography driven by a predominance of single-view and single-task models. This reflects a broader need for flexible approaches that can accommodate heterogeneous protocols and acquisitions while enabling inference for the broadest set of clinical labels. While prior work has primarily been limited to single-view echocardiography and specialized single-task models[4,6,7,16,22,24,25], *PanEcho* is unique in its multi-task modeling of all variables forming the core of standard echocardiography reporting. Unlike prior approaches that require acquisition of a particular echocardiographic view or sequence[6,15,16,18,33], *PanEcho* provides inference from any set of available echocardiograms. Here we show that across the complete set of imaging acquired as part of a standard echocardiographic study, our approach provides study-level estimates that reach performance on par with state-of-the-art specialized models for individual labels. Perhaps more importantly, *PanEcho* enables accurate diagnostic inference through abbreviated five-video protocols, which can play a critical role in simplified, automated, rapid screening echocardiograms.

To understand the value of *PanEcho*, our contribution should be evaluated in the context of recent efforts toward AI-enabled echocardiography analysis. Several prior studies exhibit robust performance across multiple echocardiographic labels, but leverage single-view echocardiography to develop *independent* models specialized for each task[6,15,16]. The unique multi-task nature of *PanEcho* immediately scales to clinical deployment by simultaneously

inferring all key clinical labels; meanwhile, single-task models would pose significant practical challenges, especially in memory-constrained environments such as on-device deployment in a point-of-care ultrasound setting. Recent approaches like EchoCLIP[5] offer a new perspective toward automated echocardiography analysis by leveraging self-supervised learning (SSL) to build a multimodal foundation model, with multi-faceted zero-shot image retrieval and interpretation capabilities by incorporating natural language. Despite the promise of SSL for efficient echocardiographic representation learning[5,26], the computational overhead of the task has so far limited its use to a single echocardiographic view, without optimized performance for any specific clinical labels. In contrast, *PanEcho*'s large-scale multi-view, multi-task learning training makes it a standalone approach for comprehensive echocardiographic interpretation from any set of echocardiograms, while maintaining foundation model properties such as efficient knowledge transfer. Its shared image encoder was trained on over 50 million echocardiogram frames from different views and modalities, learning rich features that are simultaneously informative for disparate reporting tasks. This scale and diversity of multi-view inputs and multi-task outputs is perhaps the key ingredient to learning transferable echocardiographic features, outperforming alternative approaches in both in- and out-of-distribution transfer learning applications.

Overall, *PanEcho* represents both a clinical and methodological advance. With millions of echocardiographic studies performed in the United States alone each year, and increasing availability of portable ultrasound systems enabling greater accessibility, there is a growing need for systems that enable screening and phenotyping of the full spectrum of key echocardiographic labels, from detecting ventricular and atrial chamber remodeling to valvular abnormalities and their severity. These systems can be deployed as adjuncts to abbreviated protocols (e.g.,

acquiring one video from each key view followed by AI-enabled interpretation), but also can leverage the greater breadth of acquisitions found in standard, protocoled studies where they reach clinical-level accuracy for all major labels that form a modern echocardiographic report. This versatility suggests a key value of *PanEcho* as an efficient pre-reading step to maximize efficiency in the echocardiography lab, potentially accelerating standard clinical workflows while offering an additional layer of support to expert readers. Furthermore, in areas where expert readers might not be readily available, simplified *PanEcho*-supported protocols may be used to rule out significant structural abnormalities that may necessitate urgent referral.

Certain limitations merit consideration. First, our model is trained on individual echocardiograms and averages predictions from all videos acquired during a study, applying equal weight to each video. Since we know that view relevance is task-dependent, there is an opportunity to enhance *PanEcho* by allowing the model to adaptively learn which views and specific videos in a study are most influential for a given task. Second, unlike other approaches[4,17–19,33], our method does not incorporate a segmentation step for echocardiographic measurements yet achieves comparable downstream estimation performance. This decision was made to ease multi-task learning of relatively similar classification and regression tasks and to learn representations less likely to be affected by noise or variations in acquisition quality than those from pixel-wise segmentation models. Finally, prospective validation of *PanEcho* in a real-world clinical workflow would provide further insights into its clinical applicability. Upon clinical deployment, our learned task-dependent view relevance scores could provide uncertainty quantification and prioritize the prediction of high-confidence labels given the views acquired in a given study.

In summary, *PanEcho* represents a first-of-its-kind deep learning system for flexible interpretation of a broad range of echocardiographic parameters from protocols incorporating any combination of echocardiographic views. Evidenced by its strong multi-task performance in internal and external cohorts and powerful transfer learning capabilities to new downstream tasks, *PanEcho* addresses the key need for scalable and efficient echocardiography interpretation while also serving as a foundational model to facilitate the transition from single-view to multi-view analysis. This work represents a meaningful advance toward fully automated echocardiographic assessment, and the public release of *PanEcho* model weights and source code should accelerate research on deep learning for echocardiography and computer-aided diagnosis more broadly.

METHODS

Data source

A transthoracic echocardiogram study consists of dozens of ultrasound videos acquired using multiple imaging modes (2D B-mode, color Doppler, pulsed-wave Doppler, etc.) from a variety of canonical views, achieved by placing the transducer in a specific location and orientation against the patient's ribcage. While most prior work on automated echocardiography interpretation uses still frames[13,16] videos from a single echocardiographic view[5–7,14,34] or imaging mode,[22,24,25] this study leverages both 2D B-mode and color Doppler videos from all major views. Data for internal model development and evaluation was derived from transthoracic echocardiography studies performed at Yale-New Haven Health System (YNHHS) hospitals from 2016-2022 during routine clinical care. This study was approved by the Yale University

Institutional Review Board (IRB), and the need for informed consent was waived since this research represents secondary analysis of existing data.

Echocardiography data preprocessing

Similar to our previously published echocardiography processing pipeline[7], pixel data from three-dimensional echocardiographic videos was extracted from the raw Digital Imaging and Communications in Medicine (DICOM) files, deidentified by masking out peripheral pixels containing protected health information, and saved to Audio Video Interleave (AVI) format at full resolution for rapid loading. All valid videos were processed by a pretrained view classifier[15] to determine both the echocardiographic view and imaging mode by randomly selecting ten frames and averaging predicted view probabilities over the ten frames. While the view classifier could discriminate 23 fine-grained view variations, we considered the following key views: apical 2-, 3-, 4-, and 5-chamber (A2C, A3C, etc.), parasternal long axis (PLAX), parasternal short axis (PSAX), right ventricle (RV) inflow, subcostal, and suprasternal.

To detect color Doppler, we performed a three-step process of identifying videos that were (i) classified as "Other" by the view classifier, (ii) classified as color Doppler by a custom color Doppler detection model, and (iii) contained a nontrivial amount of red pixels. For step (ii), we developed a dedicated color Doppler detection model on a manually curated dataset of echocardiogram frames derived from studies not present in the YNHHS dataset used for *PanEcho* development. Specifically, we manually labeled the presence of color Doppler in all videos from five studies and included videos from another five studies that were known to not contain color Doppler as determined by the view classifier. This dataset of 11,240 labeled frames was then randomly split intro training (80%) and validation (20%) sets at the study level. An

ImageNet-pretrained ConvNeXt-T[35] convolutional neural network (CNN) was trained to classify the presence of color Doppler using a batch size of 128, the Adam optimizer[36] with a learning rate of 0.0001, and a weighted binary cross-entropy loss for ten epochs. All frames were downsampled to 256 x 256 resolution, center cropped to 224 x 224, and normalized with ImageNet channel-wise means and standard deviations. The model achieved 100% accuracy on the validation set and was then applied to all videos classified as "Other" by the view classifier. Similar to view classification, ten randomly selected frames from each video were passed to the color Doppler detection model, and predictions were averaged over the ten frames; videos not classified as color Doppler were excluded from the cohort.

As a final quality check, for step (iii), the candidate color Doppler videos underwent color detection to assert the presence of the hue of red typically present in color Doppler echocardiography to indicate blood flow toward the ultrasound probe. Frames in each video were converted to the HSV color space, and individual pixels were determined to be red if their HSV values fell between (-10, 150, 150) and (10, 255, 255). Videos were deemed to contain a nontrivial number of red pixels if the total fraction of unique pixels that were red at any point in the video exceeded 1%; all other videos were discarded. Beyond filtering out videos that were neither color Doppler nor 2D B-mode, we did not perform any further quality control to encourage robustness to variations in acquisition quality (e.g., low-contrast or off-axis images), ultrasound machine settings, etc. encountered in real-world clinical practice.

After color Doppler detection, we limited our dataset to contain at most four unique studies per patient – randomly selecting four studies to keep for patients examined at least five times – to prevent overrepresentation of specific patients and outcomes. Next, the resulting cohort was split into development and internal test sets, with studies performed from July to

December 2022 set aside as a temporally distinct test set. The remaining studies from January 2016 to June 2022 were to be used for model development after removing studies from all patients present in the test set to prevent data leakage. The development set was randomly partitioned into training (92.5%) and validation (7.5%) sets at the patient level for model training. Finally, all videos underwent more thorough deidentification by masking out pixels beyond the central image content – namely, we retained pixels from within the convex hull of the largest contour in each frame using *opencv* (https://opencv.org/). Videos were then cropped to the central image content in a temporally consistent manner and downsampled to 256 x 256 resolution with bicubic interpolation. The final YNHHS cohort consisted of 1,230,490 TTE videos from 33,927 videos of 26,067 unique patients (**Extended Data Table 1**).

Echocardiographic reporting labels

For each study in the YNHHS cohort, we extracted labels for a total of 39 reporting tasks, representing a wide variety of categorical classification (e.g., disease diagnosis) and continuous regression tasks (e.g., echocardiographic parameter estimation). This included 18 classification tasks encapsulating size, structure, and function of all four heart chambers, valvular disease, etc. and 21 regression tasks quantifying key dimensions of each chamber, blood flow velocities, etc. All labels were directly extracted from the local electronic echocardiography reporting system (Lumedx®, Oakland, CA) and reflected the final measurements and reporting confirmed by a certified echocardiographer in line with the guidelines of the American Society of Echocardiography[1]. To minimize the effect of extreme outliers on regression tasks, we applied winsorization to all continuous variables, limiting the lowest and highest values to the 0.5 and 99.5 percentile values, respectively. Additionally, given the relatively low prevalence of severe

phenotypes across certain categorical labels in classification tasks, we pooled moderate and severe phenotypes into shared severity groups for selected tasks. See **Extended Data Table 8** for a comprehensive list and description of all tasks used in this study.

PanEcho **model development**

As depicted in **Fig. 1**, our model consists of a 2D image encoder, a temporal Transformer, and task-specific output heads. We adopted a decoupled "2+1D" approach to modeling echocardiogram videos – with separate modules to learn spatial and temporal features – primarily for downstream flexibility; for instance, our 2D image backbone can be readily adapted for any echocardiographic task, while a 3D backbone would be more difficult to retrofit to a 2D image-only task such as segmentation. *PanEcho* takes an echocardiogram video clip as input and outputs predictions for all 39 echocardiographic reporting tasks described above. Each video frame is first processed by the 2D image encoder, an ImageNet[31]-pretrained ConvNeXt-T[35] CNN, which produces a learned feature vector, or representation, of each frame. These frame-wise representations are then interpreted as an ordered sequence – like words in a sentence in natural language processing – and modeled using self-attention[37] to learn time-varying associations over the frames. Frame order is embedded via sinusoidal positional encoding, which is then elementwise added to the frame-wise feature vectors and fed to a Transformer encoder consisting of four layers, each with eight self-attention heads. Mean pooling is then used to aggregate frame-wise feature vectors into a single video-level representation, which is used as input to the task-specific output heads. Each output head consists of a Dropout[38] layer with probability 0.25 and a fully-connected layer. Both regression and binary classification tasks used one output neuron, the latter followed by a sigmoid activation. Multi-class classification tasks

with k classes used k output neurons with softmax activation, and multi-label classification tasks (in our case, only Increased LV Wall Thickness) were modeled with separate binary classification heads for each class. See **Extended Data Table 8** for a description of how each task was modeled. *PanEcho* was trained to minimize the mean of all valid task-specific losses – cross-entropy for classification tasks and mean squared error for regression tasks. To control for varying units and scales of regression tasks, we first divided each regression loss by the mean observed value of that measurement in the training set before loss aggregation.

PanEcho was implemented and trained in PyTorch[39] with distributed training across eight NVIDIA A100 graphics processing units (GPUs) with automatic mixed precision to maximize throughput. During training, the model received as input a randomly sampled video clip of 16 consecutive frames from an echocardiogram, following prior work[7]. To increase robustness to variations in acquisition and increase effective sample size, the following augmentations were performed to all video frames in a temporally consistent manner: random crop to 224 x 224 resolution, random horizontal flip with probability 0.5, random rotation within (-15°, 15°), then followed by ImageNet normalization. The model was trained with a batch size of 16 per GPU, the Adam optimizer[36], and minimized the multi-task loss described above with learning rate 0.0001. The learning rate was reduced by a factor of 0.5 if the validation metric (mean classification AUC and regression R^2 across all tasks) did not improve for three consecutive epochs; though MAE was the primary evaluation metric for regression tasks, this validation metric was chosen because AUC and R^2 are both increasing and bounded to $[0, 1]$. The model was trained for a maximum of 30 epochs with early stopping if validation metric did not improve for 10 consecutive epochs. At test time, four 16-frame clips are randomly sampled from each video and task-wise predictions are averaged over all clips to produce video-level predictions.

Since *PanEcho* is view-agnostic and labels are determined at the study level, predictions from all videos acquired during the same study (regardless of imaging mode or view) were averaged to form a single study-level prediction for each task.

Multi-task performance evaluation

Since task labels are unique to each echocardiographic study, evaluation was performed at the study level using all available videos and tasks. For internal YNHHS evaluation, this meant that multi-view aggregation could be leveraged for inference on all 39 tasks. Evaluation on external cohorts, however, was limited to the use of one or two echocardiographic views for a certain subset of labels present in the given dataset. Classification tasks were evaluated primarily by area under the receiver operating characteristic curve (AUC) and average precision (AP), and regression tasks were evaluated by mean absolute error (MAE) and R^2. For multi-class classification tasks, we present AUC results on the most severe class in the main text primarily to simplify presentation; further, there is likely significant uncertainty in intermediate designations such as "mild-moderate", and our prior work on severe aortic stenosis detection[7,14] has demonstrated that models trained for severe disease detection naturally produce probabilities that stratify the spectrum of severity. For regression tasks, we report task-wise MAE in the main text as well as the normalized MAE – MAE divided by the mean of ground truth measurements – averaged over all regression tasks to summarize overall performance while accounting for the vastly different units and scales across tasks. We computed 95% confidence intervals for all metrics with 1,000 bootstrap samples of the given test set at the study level using the percentile method.

External validation cohorts

To ensure generalizability to new patient cohorts, *PanEcho* was validated externally on two large echocardiography datasets from other hospital systems, EchoNet-LVH[6] and EchoNet-Dynamic[30], on a total of 10 tasks assessing LV size, structure, and function. EchoNet-LVH consists of 12,000 PLAX echocardiograms performed at Stanford Health Care from 2008-2020, including echocardiographic measurements of LV intraventruclar septum thickness at diastole (IVSd), LV posterior wall thickness at diastole (LVPWd), LV internal diameter at systole (LVIDs), and LVID at diastole (LVIDd). Since categorical labels for increased LV size and wall thickness were not explicitly provided, we determined increased LV size labels via "Moderate or greater" = LVIDd ≥ 6.4 cm, "Normal" = LVIDd ≤ 5.2 cm, and "Mild" otherwise, as well as increased LV wall thickness labels via "Moderate or greater" = IVSd ≥ 1.3 cm & LVPWd ≥ 1.3 cm, "Any" = IVSd ≥ 1.1 cm & LVPWd ≥ 1.1 cm, and "None" otherwise. EchoNet-Dynamic consists of 10,030 A4C echocardiograms acquired at Stanford University Hospital from 2016-2018 with labels for LV EF, end-diastolic volume, and end-systolic volume. Much like EchoNet-LVH, since only continuous measurements were provided, we determined LV systolic dysfunction labels as follows: "None-Hyperdynamic" = LV EF ≥ 54%, "Moderate or greater" = LV EF ≤ 40%, and "Mild" otherwise.

While categorical cutoffs for these conditions are sex-dependent, these conservative thresholds were chosen since patient sex was not provided in EchoNet-LVH nor EchoNet-Dynamic. For both datasets, external validation was performed using all available labels for each task. Unlike the YNHHS dataset, both external datasets contain a single echocardiogram video from a single view per study, so multi-view integration and analysis could not be performed.

Task-specific view relevance

Different echocardiographic views are used to visualize distinct aspects of the cardiovascular anatomy and function; this means that while key views like PLAX and A4C might be useful for many tasks, they may be completely irrelevant to others. Additionally, while the standard imaging mode of 2D B-mode ultrasound is most used, color Doppler imaging – which quantifies blood flow, often with a red-blue color overlay – is the gold standard for echocardiographic interpretation tasks like valvular regurgitation diagnosis. Since *PanEcho* is view-agnostic, having been trained on both 2D B-mode and color Doppler videos from all major views, we were able to use its predictive ability on individual view types as a proxy for task-dependent relevance. Specifically, we defined a normalized view relevance score

$$R_{v,t} = \begin{cases} \frac{m_{v,t}}{\max_v (m_{v,t})}, & t \text{ is a classification task} \\ \frac{m_{v,t}}{\min_v (m_{v,t})}, & t \text{ is a regression task} \end{cases},$$

where $R_{v,t}$ is the relevance of view v for task t, and $m_{v,t}$ is the performance metric on task t when only using view v (AUC for classification tasks and MAE for regression tasks). This produces a task-normalized score where, for a given task, 1 represents the most informative view, and each score can be interpreted as the "fractional importance relative to the best view." This analysis was performed on the YNHHS test set and metrics were computed after selecting a maximum of three videos per view in a given study with the most confident predicted view probability by the view classifier; this was done to control for the variable prevalence of views – without this, the most common views would be overrepresented within each study and unfairly benefit from a greater ensembling effect after video-level aggregation. For tasks typically performed with or aided by some form of Doppler imaging, we performed this analysis again

after including color Doppler videos (from any echocardiographic view) as an additional "view" to assess the task-dependent value of color Doppler imaging.

Transfer learning experiments

Beyond evaluating "out-of-the-box" generalizability of *PanEcho* to new patient populations, we also investigated its transfer learning capabilities when fine-tuned on new echocardiography data and tasks. We hypothesized that *PanEcho*'s large-scale multi-task and multi-view training would make it an ideal candidate for efficient transfer learning to downstream echocardiographic interpretation tasks. To evaluate transfer learning ability, we fine-tune the 2+1D *PanEcho* model architecture for downstream LV EF estimation in new patient cohorts while varying the initialization of the 2D image encoder, assessing both predictive performance on test data and training efficiency (defined as the number of epochs before convergence, as determined by early stopping). Specifically, we consider a 2+1D *PanEcho* architecture with a YNHHS-pretrained ConvNeXt-T, a randomly initialized ConvNeXt-T, and an ImageNet-pretrained ConvNeXt-T image encoder. While this represents a controlled experiment in which the only variable is the initialization of the 2D image encoder, we also consider (i) an "in-domain" 2+1D transfer learning approach leveraging a ConvNeXt-B backbone pretrained on one million A4C echocardiograms from EchoCLIP[5] and (ii) a 3D transfer learning approach leveraging a 3DResNet-18[40] pretrained on the large-scale Kinetics-400[32] video dataset; for the latter model, the spatiotemporal 3D CNN removes the need for the temporal Transformer of the *PanEcho* architecture.

Transfer learning experiments were performed on EchoNet-Dynamic and EchoNet-Pediatric[17] for single-view EF and multi-view pediatric EF estimation, respectively. EchoNet-

Dynamic fine-tuning was conducted using the official training/validation/test splits leveraging all available cases with LV EF labels. Results are reported on the official test set leveraging a single A4C echocardiogram per study. EchoNet-Pediatric consists of 3,176 A4C and 4,424 parasternal short axis (PSAX) echocardiograms collected from patients at Lucile Packard Children's Hospital from 2014-2021. Using the official 10-fold cross-validation splits of EchoNet-Pediatric, 10 models were fine-tuned by treating the first consecutive 8 folds as a training set, the next fold as a validation set, and the next fold as a held-out test set. Since EchoNet-Pedatric is a multi-view dataset, EF estimates were averaged over A4C and PSAX views acquired in the same study, whenever available, at test time. Results are reported by aggregating all held-out test fold predictions, and training time is summarized by mean and standard deviation number of epochs to convergence across the 10 cross-validation experiments. All transfer learning models were trained with the same procedure as PanEcho except that only the EF output head and loss were used, loss was used as the validation metric, no augmentation was used, and no learning rate reduction was used to simplify training.

DATA AVAILABILITY

The YNHHS data used in this study is not available for public sharing due to the restrictions in our IRB agreement. However, deidentified test data may be made available to researchers under a data use agreement upon publication in a peer-reviewed journal. The external datasets EchoNet-LVH, EchoNet-Dynamic, and EchoNet-Pediatric can be accessed through the Stanford AIMI Shared Datasets repository at the following links, respectively: https://stanfordaimi.azurewebsites.net/datasets/5b7fcc28-579c-4285-8b72-e4238eac7bd1,

https://stanfordaimi.azurewebsites.net/datasets/834e1cd1-92f7-4268-9daa-d359198b310a, and https://stanfordaimi.azurewebsites.net/datasets/a84b6be6-0d33-41f9-8996-86e5df53b005.

CODE AVAILABILITY

The code repository for this study will be made available at https://github.com/CarDS-Yale/PanEcho.

ACKNOWLEDGMENTS

National Heart, Lung, And Blood Institute of the National Institutes of Health (under award numbers R01HL167858 and K23HL153775 to R.K., and F32HL170592 to E.K.O.), National Institute on Aging of the National Institutes of Health (under award number R01AG089981 to R.K.), and the Doris Duke Charitable Foundation (under award number 2022060 to R.K.).

COMPETING INTERESTS

R.K. is an Associate Editor of JAMA and receives research support, through Yale, from the Blavatnik Foundation, Bristol-Myers Squibb, Novo Nordisk, and BridgeBio. He is a coinventor of U.S. Provisional Patent Applications 63/177,117, 63/428,569, 63/346,610, 63/484,426, 63/508,315, 63/580,137, 63/606,203, 63/562,335, and a co-founder of Ensight-AI, Inc and Evidence2Health, LLC. E.K.O. is a co-founder of Evidence2Health LLC, a co-inventor in patent applications (18/813,882, 17/720,068, 63/619,241, 63/177,117, 63/580,137, 63/606,203, 63/562,335, US11948230B2), and has served as consultant for Caristo Diagnostics Ltd and Ensight-AI Inc, outside the submitted work. All other authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization: G.H., E.K.O, and R.K.; Data Curation: G.H., E.K.O, and R.K.; Methodology: G.H and E.K.O.; Data Analysis: G.H. and E.K.O.; Writing, Review, and Editing: G.H., E.K.O., Z.W., R.K.; Supervision: Z.W. and R.K.

REFERENCES

1. Mitchell, C. *et al.* Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: Recommendations from the American society of echocardiography. *J. Am. Soc. Echocardiogr.* **32**, 1–64 (2019).
2. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat. Med.* **28**, 31–38 (2022).
3. Esteva, A. *et al.* Deep learning-enabled medical computer vision. *NPJ Digit Med* **4**, 5 (2021).
4. Ouyang, D. *et al.* Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252–256 (2020).
5. Christensen, M., Vukadinovic, M., Yuan, N. & Ouyang, D. Vision–language foundation model for echocardiogram interpretation. *Nat. Med.* **30**, 1481–1488 (2024).
6. Duffy, G. *et al.* High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy With Cardiovascular Deep Learning. *JAMA Cardiol* **7**, 386–395 (2022).
7. Holste, G. *et al.* Severe aortic stenosis detection by deep learning applied to echocardiography. *Eur. Heart J.* (2023) doi:10.1093/eurheartj/ehad456.

8. Wei, C., Milligan, M., Lam, M., Heidenreich, P. A. & Sandhu, A. Variation in cost of echocardiography within and across United States hospitals. *J. Am. Soc. Echocardiogr.* **36**, 569-577.e4 (2023).

9. Virnig, B. A., Shippee, N. D., O'Donnell, B., Zeglin, J. & Parashuram, S. *Trends in the Use of Echocardiography, 2007 to 2011*. (Agency for Healthcare Research and Quality (US), 2014).

10. Pillai, B., Salerno, M., Schnittger, I., Cheng, S. & Ouyang, D. Precision of echocardiographic measurements. *J. Am. Soc. Echocardiogr.* **37**, 562–563 (2024).

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

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

13. Huang, Z., Long, G., Wessler, B. & Hughes, M. C. A New Semi-supervised Learning Benchmark for Classifying View and Diagnosing Aortic Stenosis from Echocardiograms. in *Proceedings of the 6th Machine Learning for Healthcare Conference* (eds. Jung, K., Yeung, S., Sendak, M., Sjoding, M. & Ranganath, R.) vol. 149 614–647 (PMLR, 06--07 Aug 2021).

14. Oikonomou, E. K. *et al.* A Multimodality Video-Based AI Biomarker For Aortic Stenosis Development And Progression. *JAMA Card* **9**, 534–544 (2024).

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

16. Ghorbani, A. *et al.* Deep learning interpretation of echocardiograms. *NPJ Digit Med* **3**, 10 (2020).

17. Reddy, C. D., Lopez, L., Ouyang, D., Zou, J. Y. & He, B. Video-Based Deep Learning for Automated Assessment of Left Ventricular Ejection Fraction in Pediatric Patients. *J. Am. Soc. Echocardiogr.* **36**, 482–489 (2023).

18. Tromp, J. *et al.* A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *Nat. Commun.* **13**, 6776 (2022).

19. Zeng, Y. *et al.* MAEF-Net: Multi-attention efficient feature fusion network for left ventricular segmentation and quantitative analysis in two-dimensional echocardiography. *Ultrasonics* **127**, 106855 (2023).

20. Khera, R. *et al.* Transforming Cardiovascular Care With Artificial Intelligence: From Discovery to Practice: JACC State-of-the-Art Review. *J. Am. Coll. Cardiol.* **84**, 97–114 (2024).

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

22. Long, A. *et al.* Deep Learning for Echo Analysis, Tracking, and Evaluation of Mitral Regurgitation (DELINEATE-MR). *Circulation* (2024) doi:10.1161/CIRCULATIONAHA.124.068996.

23. Ferreira, D. L., Salaymang, Z. & Arnaout, R. Label-free segmentation from cardiac ultrasound using self-supervised learning. *arXiv [eess.IV]* (2022).

24. Vrudhula, A., Duffy, G., Vukadinovic, M., Liang, D. & Cheng, S. High Throughput Deep Learning Detection of Mitral Regurgitation. *medRxiv* (2024).

25. Vrudhula, A. *et al.* Deep Learning Phenotyping of Tricuspid Regurgitation for Automated High Throughput Assessment of Transthoracic Echocardiography. *medRxiv* (2024) doi:10.1101/2024.06.22.24309332.

26. Holste, G., Oikonomou, E. K., Mortazavi, B. J., Wang, Z. & Khera, R. Efficient deep learning-based automated diagnosis from echocardiography with contrastive self-supervised learning. *Commun. Med.* **4**, 133 (2024).

27. Díaz-Gómez José L., Mayo Paul H. & Koenig Seth J. Point-of-Care Ultrasonography. *N. Engl. J. Med.* **385**, 1593–1602 (2021).

28. Ginsburg, A. S., Liddy, Z., Khazaneh, P. T., May, S. & Pervaiz, F. A survey of barriers and facilitators to ultrasound use in low- and middle-income countries. *Sci. Rep.* **13**, 1–11 (2023).

29. Narang, A. *et al.* Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use. *JAMA Cardiol* **6**, 624–632 (2021).

30. Ouyang, D. *et al.* EchoNet-dynamic: A large new cardiac motion video data resource for medical machine learning. *NeurIPS ML4H* (2019).

31. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (IEEE, 2009).

32. Kay, W. *et al.* The Kinetics Human Action Video Dataset. *arXiv [cs.CV]* (2017).

33. Tromp, J. *et al.* Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *Lancet Digit Health* **4**, e46–e54 (2022).

34. Hughes, J. W. *et al.* Deep learning evaluation of biomarkers from echocardiogram videos. *EBioMedicine* **73**, 103613 (2021).

35. Liu, Z. *et al.* A ConvNet for the 2020s. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 11966–11976 (2022).

36. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv [cs.LG]* (2014).

37. Vaswani, A. *et al.* Attention is All you Need. *Adv. Neural Inf. Process. Syst.* 5998–6008 (2017).

38. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).

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

40. Tran, D. *et al.* A Closer Look at Spatiotemporal Convolutions for Action Recognition. in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018). doi:10.1109/cvpr.2018.00675.

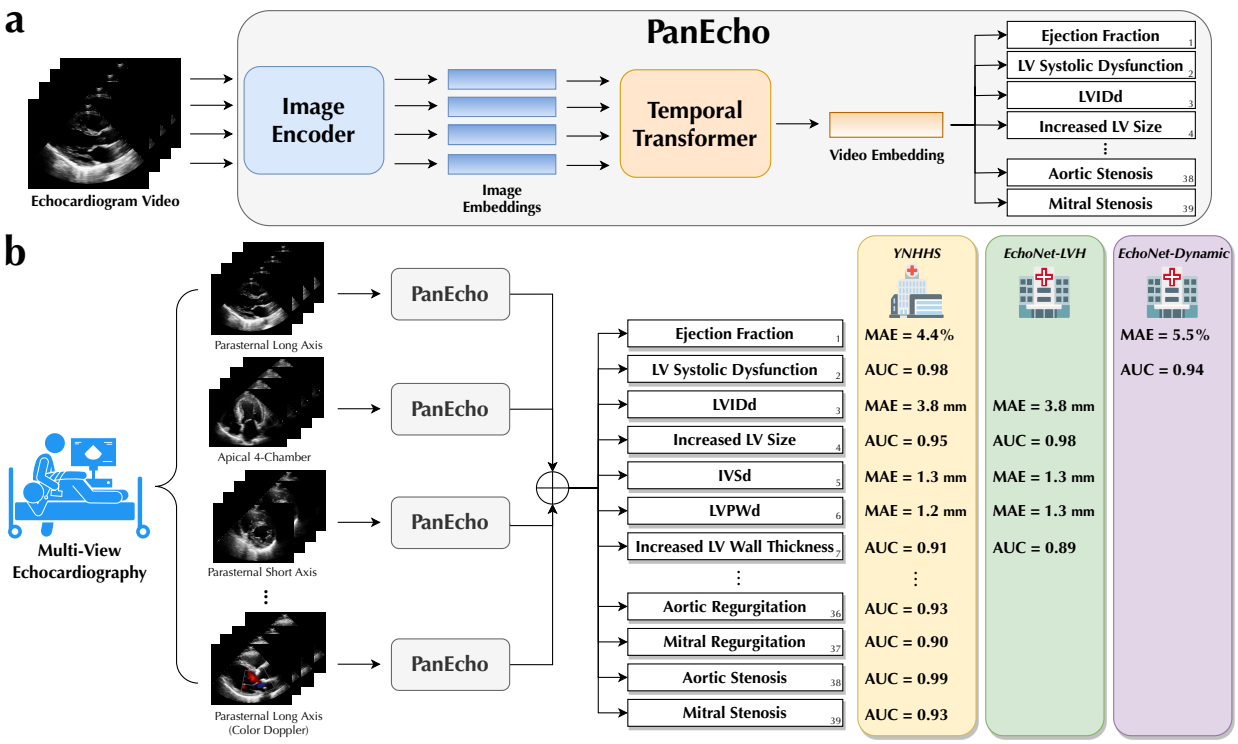

Fig. 1 | Overview of PanEcho. (a) Schematic of *PanEcho*, a view-agnostic, multi-task deep learning model that automatically performs 39 transthoracic echocardiography interpretation tasks from echocardiogram videos. *PanEcho* consists of an image encoder to embed individual video frames, a temporal Transformer to learn temporal associations over frames in a video, and task-specific output heads to perform a wide variety of classification and regression tasks. The model was trained end-to-end on over one million echocardiograms from YNHHS hospitals with a multi-task learning objective. **(b)** Since *PanEcho* is view-agnostic, multi-view echocardiography can be leveraged to integrate information across views. At test time, predictions are aggregated across echocardiograms acquired during the same study to generate study-level predictions for each task. For all 39 tasks, *PanEcho* was validated internally on temporally held-out data from YNHHS hospitals. For tasks with publicly available labels, *PanEcho* was also externally validated on EchoNet-LVH and EchoNet-Dynamic, two large-scale single-view echocardiography datasets for assessing LV structure and function.
AUC = area under the receiver operating characteric curve; IVSd = intraventricular septum thickness at diastole; LV = left ventricle; LVIDd = left ventricular internal diameter at diastole; LVPWd = left ventricular posterior wall thickness at diastole; MAE = mean absolute error; YNHHS = Yale-New Haven Health System.

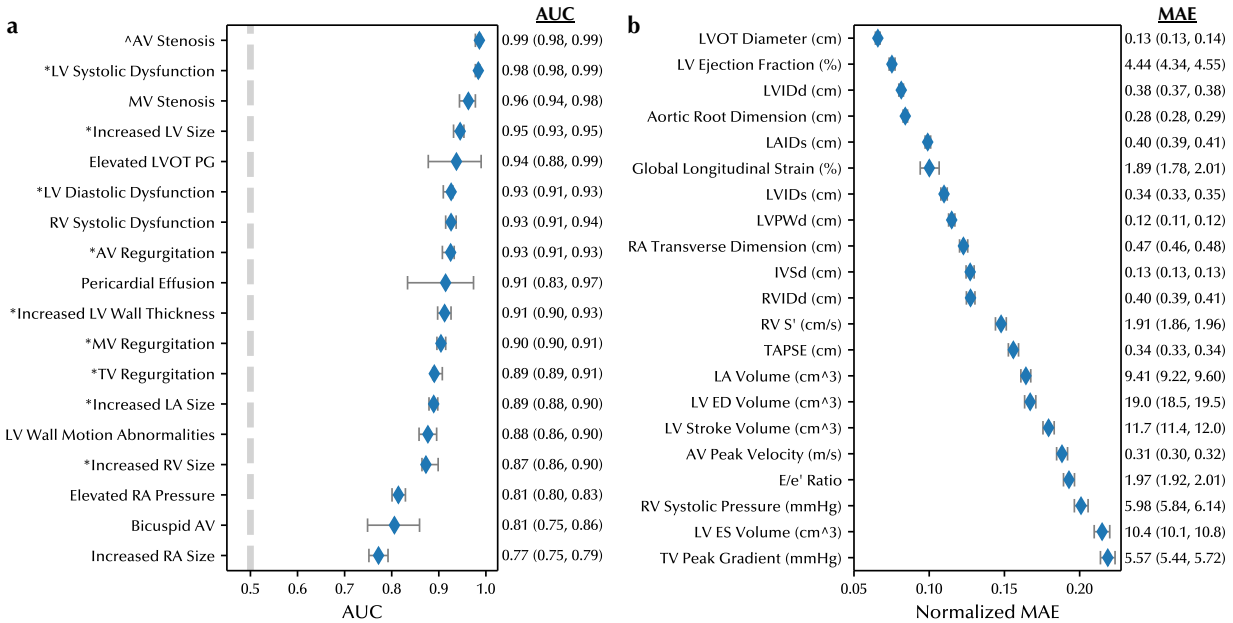

Fig. 2 | Multi-task performance evaluation. Multi-task evaluation of *PanEcho* on the internal YNHHS test set, consisting of temporally held-out echocardiograms acquired from June-December 2022. Error bars and values in parentheses represent bootstrapped 95% confidence intervals. **(a)** For classification tasks, AUC values are presented. The grey dashed line represents the performance of random guessing. * = moderate or greater; ^ = severe. **(b)** For regression tasks, normalized MAE (using the mean of ground truth measurements) is visually presented to account for varying scales and units. Raw MAE is also presented in text beside each task.
AUC = area under the receiver operating characteristic curve; AV = aortic valve; ED = end-diastolic; ES = end-systolic; IVSd = intraventricular septum thickness at diastole; LA = left atrium; LV = left ventricle; LAIDs = left atrial internal diameter at systole; LVIDd = left ventricular internal diameter at diastole; LVIDs = left ventricular internal diameter at systole; LVOT = left ventricular outflow tract; LVPWd = left ventricular posterior wall thickness at diastole; MAE = mean absolute error; PG = pressure gradient; RA = right atrium; RV = right ventricle; RVIDd = right ventricular internal diameter at diastole; RV S' = right ventricular systolic excursion velocity; TV = tricuspid valve; YNHHS = Yale-New Haven Health System.

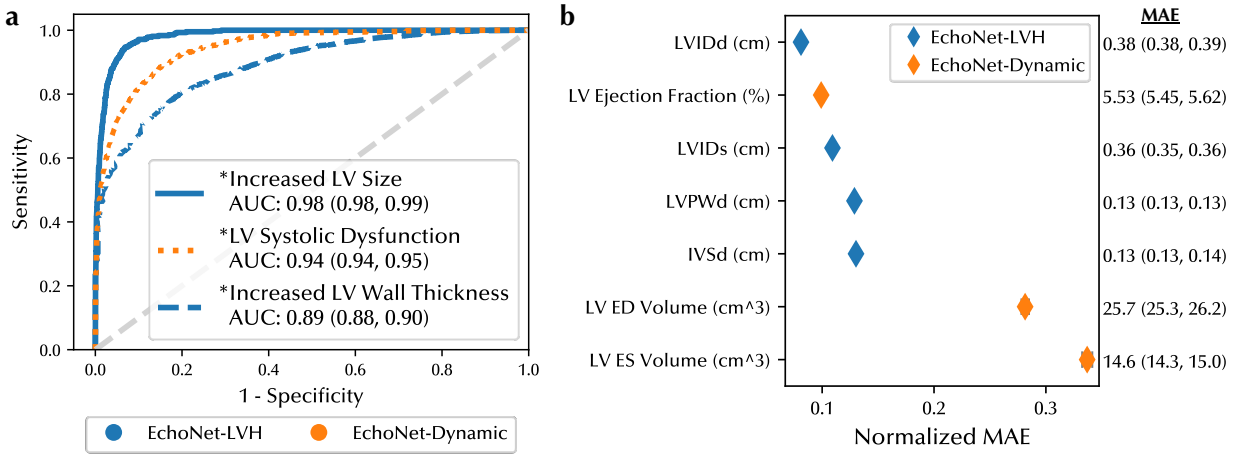

Fig. 3 | External performance evaluation. Multi-task evaluation of *PanEcho* on two external echocardiography datasets, EchoNet-LVH (blue) and EchoNet-Dynamic (orange). Error bars and values in parentheses represent bootstrapped 95% confidence intervals. **(a)** For classification tasks, receiver operating characteristic curves are presented. The grey dashed line represents the performance of random guessing. * = moderate or greater. **(b)** For regression tasks, normalized MAE (using the mean of ground truth measurements) is visually presented to account for varying scales and units. Raw MAE is also presented beside each task.

AUC = area under the receiver operating characteristic curve; ED = end-diastolic; ES = end-systolic; IVSd = intraventricular septum thickness at diastole; LV = left ventricle; LVIDd = left ventricular internal diameter at diastole; LVIDs = left ventricular internal diameter at systole; LVPWd = left ventricular posterior wall thickness at diastole; MAE = mean absolute error.

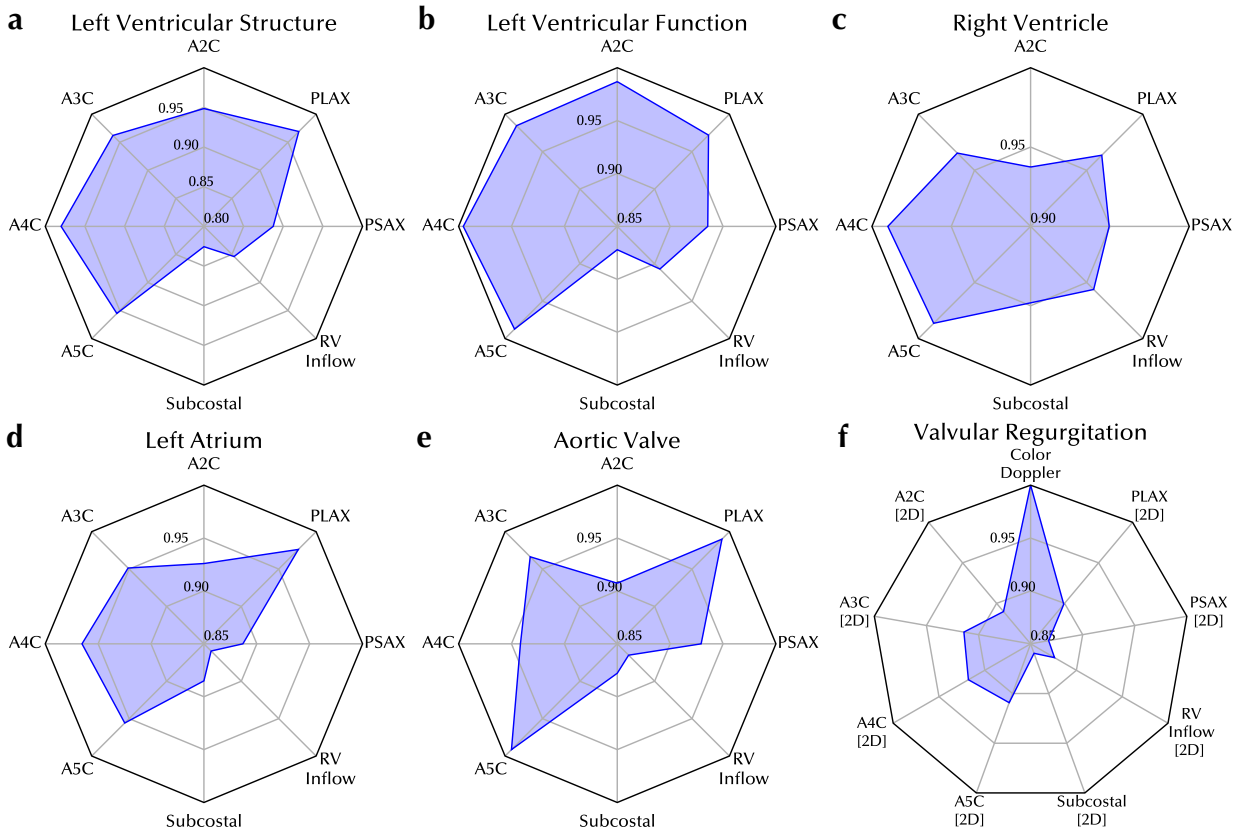

Fig. 4 | Task-specific view relevance. Radar plots depicting the relative importance of each echocardiographic view for a given aspect of cardiovascular diagnosis. For each task, a normalized view relevance score is computed, where 1 indicates the most relevant view; presented view relevance scores are then averaged over all tasks falling under a given category in each subplot. Analysis is performed on the internal YNHHS test set using up to three videos from a given echocardiographic view per study.
A2C = apical 2-chamber; A3C = apical 3-chamber; A4C = apical 4-chamber; A5C = apical 5-chamber PLAX = parasternal long axis; PSAX = parasternal short axis; RV = right ventricle; YNHHS = Yale-New Haven Health System.