Article

Video-based AI for beat-to-beat assessment of cardiac function

https://doi.org/10.1038/s41586-020-2145-8

Received: 11 November 2019

Accepted: 20 February 2020

Published online: 25 March 2020

Check for updates

David Ouyang[1]✉, Bryan He[2], Amirata Ghorbani[3], Neal Yuan[4], Joseph Ebinger[4], Curtis P. Langlotz[1,5], Paul A. Heidenreich[1], Robert A. Harrington[1], David H. Liang[1,3], Euan A. Ashley[1,6,7] & James Y. Zou[2,3,6,7]✉

Accurate assessment of cardiac function is crucial for the diagnosis of cardiovascular disease[1], screening for cardiotoxicity[2] and decisions regarding the clinical management of patients with a critical illness[3]. However, human assessment of cardiac function focuses on a limited sampling of cardiac cycles and has considerable inter-observer variability despite years of training[4,5]. Here, to overcome this challenge, we present a video-based deep learning algorithm—EchoNet-Dynamic—that surpasses the performance of human experts in the critical tasks of segmenting the left ventricle, estimating ejection fraction and assessing cardiomyopathy. Trained on echocardiogram videos, our model accurately segments the left ventricle with a Dice similarity coefficient of 0.92, predicts ejection fraction with a mean absolute error of 4.1% and reliably classifies heart failure with reduced ejection fraction (area under the curve of 0.97). In an external dataset from another healthcare system, EchoNet-Dynamic predicts the ejection fraction with a mean absolute error of 6.0% and classifies heart failure with reduced ejection fraction with an area under the curve of 0.96. Prospective evaluation with repeated human measurements confirms that the model has variance that is comparable to or less than that of human experts. By leveraging information across multiple cardiac cycles, our model can rapidly identify subtle changes in ejection fraction, is more reproducible than human evaluation and lays the foundation for precise diagnosis of cardiovascular disease in real time. As a resource to promote further innovation, we also make publicly available a large dataset of 10,030 annotated echocardiogram videos.

Cardiac function is essential for the maintenance of normal systemic tissue perfusion; cardiac dysfunction manifests as dyspnea, fatigue, exercise intolerance, fluid retention and increased risk of mortality[1–3,5–8]. Impairment of cardiac function is described as cardiomyopathy or heart failure and is a leading cause of hospitalization in the United States and a growing global health issue[1,9,10]. A variety of methodologies have been used to quantify cardiac function and diagnose dysfunction. In particular, measurement of left ventricular ejection fraction, the ratio of change in the left ventricular end-systolic and end-diastolic volumes, is one of the most important metrics of cardiac function, as it identifies patients who are eligible for life-prolonging therapies[7,11]. However, the assessment of ejection fraction is associated with considerable inter-observer variability as well as inter-modality discordance based on methodology and modality[2,4,5,11–14].

Human assessment of the ejection fraction has variance in part due to the common finding of irregularity in the heart rate and the laborious nature of a calculation that requires manual tracing of the size of the ventricle to quantify every beat[4,5]. Although the American Society of Echocardiography and the European Association of Cardiovascular Imaging guidelines recommend tracing and averaging up to five consecutive cardiac cycles if variation is identified, the ejection fraction is often evaluated from tracings of only one representative beat or visually approximated if a tracing is deemed to be inaccurate[5,15]. This results in high variance and limited precision with inter-observer variation[4,12–15] ranging from 7.6% to 13.9%. More-precise evaluation of cardiac function is necessary, as even patients with a borderline reduction in ejection fraction have been shown to have considerably increased morbidity and mortality[16–18].

With rapid image acquisition and relatively low cost, and without ionizing radiation, echocardiography is the most widely used modality for cardiovascular imaging[19,20]. There is great interest in using deep learning techniques for echocardiography to determine the ejection fraction[21–23]. Previous attempts to algorithmically assess cardiac function with deep learning models relied on manually curated still images at systole and diastole instead of using the actual echocardiogram videos and these models had substantial error compared to human evaluation of cardiac function[21,22], with R^2 ranging between 0.33 and 0.50. Limitations in human interpretation, including laborious manual

[1]Department of Medicine, Stanford University, Stanford, CA, USA. [2]Department of Computer Science, Stanford University, Stanford, CA, USA. [3]Department of Electrical Engineering, Stanford University, Stanford, CA, USA. [4]Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [5]Department of Radiology, Stanford University, Stanford, CA, USA. [6]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [7]These authors jointly supervised this work: Euan A. Ashley, James Y. Zou. ✉e-mail: ouyangd@stanford.edu; jamesz@stanford.edu

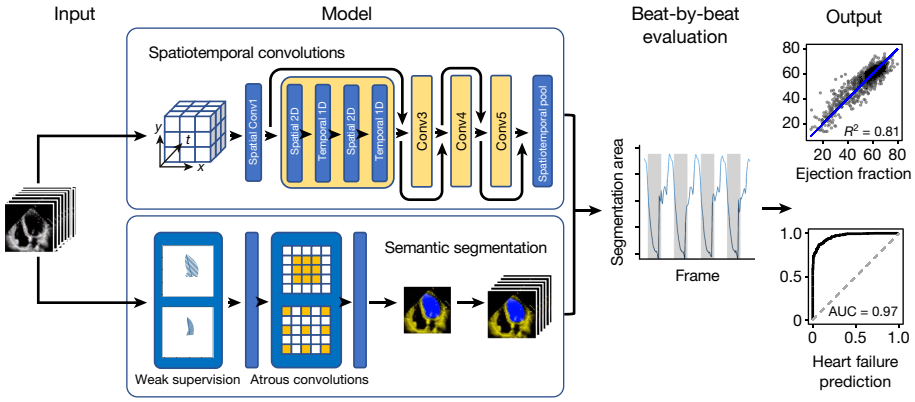


Fig. 1 | EchoNet-Dynamic workflow. For each patient, EchoNet-Dynamic uses standard apical four-chamber view echocardiogram videos as input. The model first predicts the ejection fraction for each cardiac cycle using spatiotemporal convolutions with residual connections and generates frame-level semantic segmentations of the left ventricle using weak supervision from expert human tracings. These outputs are combined to create beat-to-beat predictions of the ejection fraction and to predict the presence of heart failure with reduced ejection fraction. AUC, area under the curve.

segmentation and the inability to perform beat-to-beat quantification, may be overcome by sophisticated automated approaches[5,22,23]. Recent advances in deep learning suggest that it can accurately and reproducibly identify human-identifiable phenotypes as well as characteristics that are not recognized by human experts[24–28].

To overcome current limitations in the human assessment of cardiac function, we present EchoNet-Dynamic, an end-to-end deep learning approach for labelling of the left ventricle and estimation of the ejection fraction from input echocardiogram videos alone. We first perform frame-level semantic segmentation of the left ventricle with weakly supervised learning from clinical expert labelling. Then, a three-dimensional convolutional neural network (CNN) with residual connections predicts clip-level ejection fraction from the native echocardiogram videos. Finally, the segmentations results are combined with clip-level predictions to produce beat-to-beat evaluation of the ejection fraction. This approach provides interpretable tracings of the ventricle, which facilitate human assessment and downstream analysis, while leveraging the three-dimensional CNN to fully capture spatiotemporal patterns in the video[5,29,30].

Video-based deep learning model

EchoNet-Dynamic has three key components (Fig. 1). First, we constructed a CNN model with atrous convolutions for frame-level semantic segmentation of the left ventricle. The technique of atrous convolutions enables the model to capture larger patterns and has previously been shown to perform well on non-medical imaging datasets[29]. The standard human clinical workflow for estimating the ejection fraction requires manual segmentation of the left ventricle during end systole and end diastole. We generalize these labels in a weak supervision approach with atrous convolutions to generate frame-level semantic segmentation throughout the cardiac cycle in a 1:1 pairing with frames from the original video. The automatic segmentation is used to identify ventricular contractions and provides a clinician-interpretable intermediary that mimics the clinical workflow.

Second, we trained a CNN model with residual connections and spatiotemporal convolutions across frames to predict the ejection fraction. In contrast to previous CNN architectures for machine learning of medical images, our approach integrates spatial as well as temporal information in our network convolutions[25,29,30]. Spatiotemporal convolutions, which incorporate spatial information in two dimensions as well as temporal information in the third dimension, have previously been used in non-medical video-classification tasks[29,30]. However, this approach has not previously been used for medical data given the relative scarcity of labelled medical videos. We additionally performed a model architecture search to identify the optimal base architecture (Extended Data Fig. 1).

Finally, we make video-level predictions of the ejection fraction for beat-to-beat estimations of cardiac function. Given that variation in cardiac function can be caused by changes in loading conditions as well as heart rate in a variety of cardiac conditions, it is recommended to perform estimations of the ejection fraction for up to five cardiac cycles; however, this is not always done in clinical practice given the tedious and laborious nature of the calculation[5,15]. Our model identifies each cardiac cycle, generates a clip of 32 frames and averages clip-level estimates of the ejection fraction for each beat as test-time augmentation. EchoNet-Dynamic was developed using 10,030 apical four-chamber echocardiogram videos obtained during the course of routine clinical practice at Stanford Medicine. Extended Data Table 1 contains the summary statistics of the patient population. Details of the model and hyperparameter search are further described in the Methods and Extended Data Table 2.

Evaluation of model performance

For the test dataset from Stanford Medicine that was not previously seen during model training, the prediction of the ejection fraction by EchoNet-Dynamic had a mean absolute error of 4.1%, root mean squared error of 5.3% and R^2 of 0.81 compared with the annotations by human experts. This is well within the range of typical measurement variation between different clinicians, which is usually described as inter-observer variation[5,13–16] and can be as high as 13.9% (Fig. 2a). Using a common threshold of an ejection fraction of less than 50% to classify cardiomyopathy, the prediction by EchoNet-Dynamic had an area under the curve of 0.97 (Fig. 2b). We compared the performance of EchoNet-Dynamic to that of several additional deep learning architectures that we trained on this dataset, and EchoNet-Dynamic was consistently more accurate, suggesting the power of its specific architecture (Extended Data Table 2). In addition, we performed re-evaluation of the videos by blinded clinicians in cases in which the prediction of the ejection fraction by EchoNet-Dynamic diverged the most from the original human annotation. Many of these videos had inaccurate initial human labels (in 43% of the videos, the blinded clinicians preferred the prediction of the model), poor image quality, or arrhythmias and variations in the heart rate (Extended Data Table 3).

Generalization to a different hospital

To assess the cross-healthcare-system reliability of the model, EchoNet-Dynamic was additionally tested, without any tuning, on an external

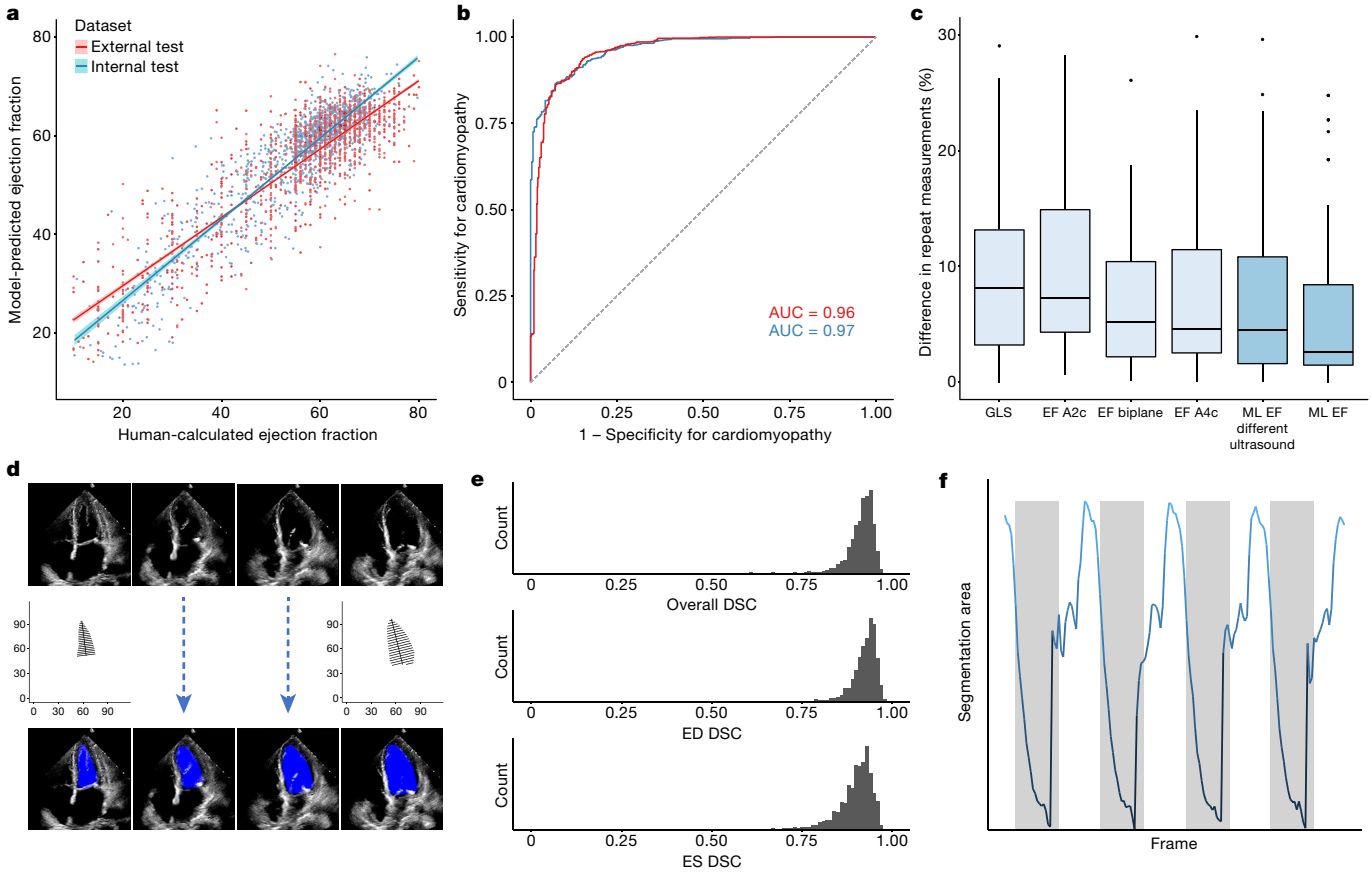


Fig. 2 | Model performance. a, The predicted ejection fraction of EchoNet-Dynamic compared with the reported ejection fraction for the internal test dataset from Stanford (blue, $n = 1,277$) and the external test dataset from Cedars-Sinai (red, $n = 2,895$). The blue and red lines indicate the least-squares regression line between model prediction and human-calculated ejection fraction. **b**, Receiver-operating characteristic curves for the diagnosis of heart failure with reduced ejection fraction for the internal test dataset (blue, $n = 1,277$) and external test dataset (red, $n = 2,895$). **c**, Variance of the metrics of cardiac function on repeat measurements. The left four box plots highlight the variation between clinicians using different techniques including global longitudinal strain (GLS) and ejection fraction (EF) by various methods ($n = 55$); the right two box plots show the variance of EchoNet-Dynamic (machine learning (ML) prediction for EF) with standard ultrasound machines ($n = 55$) and an ultrasound machine that had not previously been seen by the model ($n = 49$). Box plots show the median as a thick line, the 25th and 75th percentiles as upper and lower bounds of the box, and individual points are included for data more than 1.5× the interquartile range from the median. **d**, Weak supervision with human expert tracings of the left ventricle at end systole (ES) and end diastole (ED) is used to train a semantic segmentation model with input video frames throughout the cardiac cycle. The human expert tracings (with x, y coordinates designating the region of the image) are paired with the video frame (top row) for model training, allowing for segmentation of video frames that did not have prior human tracings (bottom row). **e**, The Dice similarity coefficient (DSC) was calculated for each end systole/end diastole frame ($n = 1,277$). **f**, The area of the left ventricle segmentation was used to identify the heart rate and bin clips for beat-to-beat evaluation of the ejection fraction.

test dataset of 2,895 echocardiogram videos from 1,267 patients from an independent hospital system (Cedars-Sinai Medical Center). On this external test dataset, EchoNet-Dynamic showed a robust prediction of the ejection fraction with a mean absolute error of 6.0%, root mean squared error of 7.7%, R^2 of 0.77 and an area under the curve of 0.96 compared with the annotations by the Cedars-Sinai cardiologists.

Comparison with human variation

To investigate the prediction variability of the model, we performed a prospective study that compared the variation in the predictions by EchoNet-Dynamic with the variation in human measurements on 55 patients that were evaluated by two different sonographers on the same day. Each patient was independently evaluated for metrics of cardiac function by multiple methods as well as our model for comparison (Fig. 2c). The assessment of cardiac function by EchoNet-Dynamic had the least variance on repeated testing (median difference of 2.6%, s.d. of 6.4) compared with the ejection fraction obtained by Simpson's biplane method (median difference of 5.2%, s.d. of 6.9, $P < 0.001$ for non-inferiority), ejection fraction from Simpson's monoplane method (median difference of 4.6%, s.d. of 7.3, $P < 0.001$ for non-inferiority) and global longitudinal strain (median difference of 8.1%, s.d. of 7.4%, $P < 0.001$ for non-inferiority). Of the initial 55 patients, 49 patients were also assessed with a different ultrasound system that was never seen during model training and the assessment by EchoNet-Dynamic had similar variance in this case (median difference of 4.5%, s.d. of 7.0, $P < 0.001$ for non-inferiority for all comparisons with human measurements).

Analysis of left ventricle segmentation

EchoNet-Dynamic automatically generates segmentations of the left ventricle, which enables clinicians to better understand how it makes predictions. The segmentation is also useful because it provides a relevant point for human interjection in the workflow and for physician oversight of the model in clinical practice. For the semantic segmentation task, the labels included 20,060 frame-level segmentations of the left ventricle by expert human sonographers and cardiologists. These manual segmentations were obtained during the course of the standard clinical workflow during end systole and end diastole. Implicit in the echocardiogram videos is that, in all intermediate frames, the

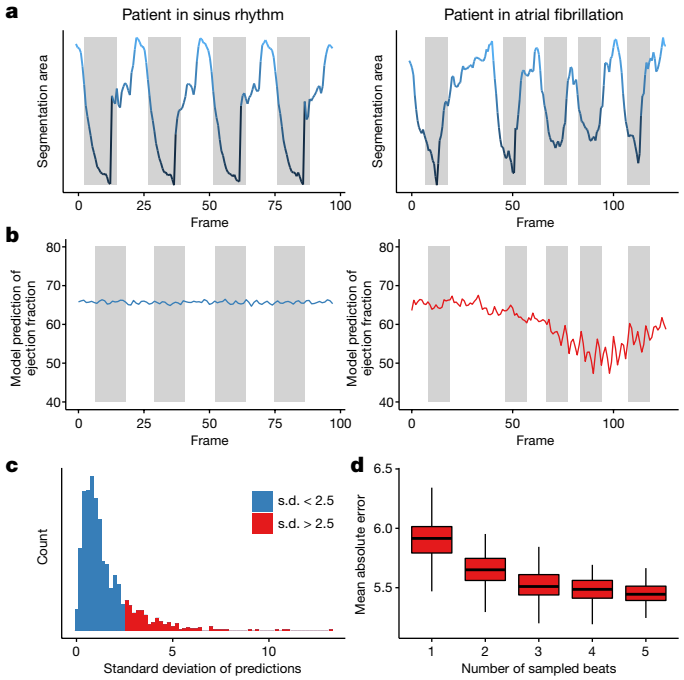


Fig. 3 | Beat-to-beat evaluation of the ejection fraction. a, Atrial fibrillation and other arrhythmias can be identified by marked variation in the intervals between ventricular contractions. **b**, Considerable variation in the heart rate was associated with higher variance in the prediction of the ejection fraction. **c**, Histogram of the s.d. of beat-to-beat evaluation of ejection fraction (n = 1,277) across the internal test videos. **d**, Assessing the effect of beat-to-beat prediction based on the number of sampled beats averaged for each prediction. Each box plot represents 100 random samples of a set number of beats and the comparison with the reported ejection fraction. Box plots show the median as a thick line, the 25th and 75th percentiles as upper and lower bounds of the box, and whiskers extend to 1.5× the interquartile range from the median.

left ventricle is constrained in shape and size between the labels at end systole and end diastole. We used these sparse human labels to train EchoNet-Dynamic to generate frame-level segmentations for the entire video (Fig. 2d). On the test dataset, the Dice similarity coefficient for the end-systolic tracing was 0.903 (95% confidence interval of 0.901–0.906) and the Dice similarity coefficient for the end-diastolic tracing was 0.927 (95% confidence interval of 0.925–0.928) (Fig. 2d). There was significant concordance in performance of end-systolic and end-diastolic semantic segmentation and the change in segmentation area was used to identify each cardiac contraction (Fig. 2e, f).

Variation in beat-to-beat model interpretation was seen in echocardiogram videos of patients with arrhythmias and ectopy (Fig. 3). When undergoing an individual beat-to-beat evaluation of the Stanford test dataset, videos with higher variance had fewer beats with an ejection fraction close to the human estimate (Extended Data Fig. 2; 51% versus 72% of beats within 5% of the human estimates of the ejection fraction, respectively, for high-variance and low-variance videos, $P < 0.0001$). In addition to the correlation with irregularity in intervals between ventricular contractions, these videos were independently reviewed by cardiologists and found to have atrial fibrillation, premature atrial contractions and premature ventricular contractions. By aggregating across multiple beats, EchoNet-Dynamic significantly reduces the estimation error of the ejection fraction (Fig. 3d). In addition, even with only one GPU, EchoNet-Dynamic rapidly performs the predictions (less than 0.05 s per prediction) and enables the real-time segmentation of the left ventricle and prediction of the ejection fraction (Extended Data Table 4).

Discussion

EchoNet-Dynamic is a video-based deep learning algorithm that achieves state-of-the-art assessment of cardiac function. It uses expert human tracings for weakly supervised learning of left ventricular segmentation and spatiotemporal convolutions on video data to obtain a beat-to-beat cumulative evaluation of the ejection fraction across the entire video. EchoNet-Dynamic is, to our knowledge, the first video-based deep learning model for echocardiogram and its performance in assessing the ejection fraction is substantially better than that of previous image-based deep learning attempts[20,22]. The variance in predictions of EchoNet-Dynamic is comparable to or less than measurements of cardiac function by human experts[5]. Moreover, its performance in predicting the ejection fraction was robustly accurate when used on a validation dataset of echocardiogram videos from an independent medical centre without additional model training. With only one GPU, EchoNet-Dynamic completes these tasks in real time; each prediction task takes only 0.05 s per frame and is much more rapid than the human assessment of ejection fraction. EchoNet-Dynamic could potentially aid clinicians with a more-precise and reproducible assessment of cardiac function and can detect subclinical changes in ejection fraction beyond the precision of human readers.

Some of the difference between the model and human evaluation is, in part, a feature of the comparison of the beat-to-beat evaluation of the ejection fraction across the video by EchoNet-Dynamic with human evaluations of only one 'representative' beat while ignoring additional beats. Choosing the representative beat can be subjective, contributing to human intra-observer variability and ignoring the guideline recommendation of averaging five consecutive beats. This five-beat workflow is rarely completed, in part because of the laborious and time-intensive nature of the human tracing task. EchoNet-Dynamic greatly decreases the labour of the cardiac function assessment by automating the segmentation task and provides the opportunity for more-frequent, rapid evaluations of cardiac function. Our end-to-end approach generates beat- and clip-level predictions of the ejection fraction as well as the segmentation of the left ventricle throughout the cardiac cycle for visual interpretation of the modelling results. In settings in which the sensitive detection of change in cardiac function is critical, early detection of change can substantially affect clinical care[2,3].

We worked with stakeholders across Stanford Medicine to release our full dataset of 10,030 de-identified echocardiogram videos as a resource for the medical machine learning community for future comparison and validation of deep learning models. To our knowledge, this is one of the largest labelled medical video datasets to be made publicly available and the first large release of echocardiogram data with matched labels of human expert tracings, volume estimates and calculations of the left ventricular ejection fraction. We expect that this dataset will greatly facilitate new echocardiogram and medical video-based machine learning approaches. We have also released the full code for our algorithm and data-processing workflow.

Our model was trained on videos obtained by trained sonographers at an academic medical centre that reflect the variation in clinical practice. With expansion in the use of point-of-care ultrasound for evaluation of cardiac function by non-cardiologists, further work needs to be done to understand model performance with input videos of more-variable quality and acquisition expertise as well as comparison with other imaging modalities. Our experiments to simulate degraded video quality and analyses across health systems suggest that EchoNet-Dynamic is robust to variation in video acquisition; however, further analyses in diverse clinical environments remain to be done.

Our results represent an important step towards the automated evaluation of cardiac function from echocardiogram videos through deep learning. EchoNet-Dynamic could augment current methods with improved precision to enable earlier detection of subclinical cardiac dysfunction, and the underlying open dataset can be used to advance

future work in deep learning for medical videos and lay the foundation for further applications of medical deep learning.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2145-8.

1. Ziaeian, B. & Fonarow, G. C. Epidemiology and aetiology of heart failure. *Nat. Rev. Cardiol.* **13**, 368–378 (2016).
2. Shakir, D. K. & Rasul, K. I. Chemotherapy induced cardiomyopathy: pathogenesis, monitoring and management. *J. Clin. Med. Res.* **1**, 8–12 (2009).
3. Dellinger, R. P. et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med.* **39**, 165–228 (2013).
4. Farsalinos, K. E. et al. Head-to-head comparison of global longitudinal strain measurements among nine different vendors: the EACVI/ASE Inter-Vendor Comparison Study. *J. Am. Soc. Echocardiogr.* **28**, 1171–1181 (2015).
5. Lang, R. M. et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur. Heart J. Cardiovasc. Imaging* **16**, 233–271 (2015).
6. McMurray, J. J. et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012. *Eur. J. Heart Fail.* **14**, 803–869 (2012).
7. Loehr, L. R., Rosamond, W. D., Chang, P. P., Folsom, A. R. & Chambless, L. E. Heart failure incidence and survival (from the Atherosclerosis Risk in Communities study). *Am. J. Cardiol.* **101**, 1016–1022 (2008).
8. Bui, A. L., Horwich, T. B. & Fonarow, G. C. Epidemiology and risk profile of heart failure. *Nat. Rev. Cardiol.* **8**, 30–41 (2011).
9. Roizen, M. F. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Yearbook Anesthesiol. Pain Manage.* **2012**, 12–13 (2012).
10. Yancy, C. W. et al. 2013 ACCF/AHA guideline for the management of heart failure. *Circulation* **128**, e240–e327 (2013).
11. Huang, H. et al. Accuracy of left ventricular ejection fraction by contemporary multiple gated acquisition scanning in patients with cancer: comparison with cardiovascular magnetic resonance. *J. Cardiovasc. Magn. Reson.* **19**, 34 (2017).
12. Pellikka, P. A. et al. Variability in ejection fraction measured by echocardiography, gated single-photon emission computed tomography, and cardiac magnetic resonance in patients with coronary artery disease and left ventricular dysfunction. *JAMA Netw. Open* **1**, e181456 (2018).
13. Malm, S., Frigstad, S., Sagberg, E., Larsson, H. & Skjaerpe, T. Accurate and reproducible measurement of left ventricular volume and ejection fraction by contrast echocardiography: a comparison with magnetic resonance imaging. *J. Am. Coll. Cardiol.* **44**, 1030–1035 (2004).
14. Cole, G. D. et al. Defining the real-world reproducibility of visual grading of left ventricular function and visual estimation of left ventricular ejection fraction: impact of image quality, experience and accreditation. *Int. J. Cardiovasc. Imaging* **31**, 1303–1314 (2015).
15. Koh, A. S. et al. A comprehensive population-based characterization of heart failure with mid-range ejection fraction. *Eur. J. Heart Fail.* **19**, 1624–1634 (2017).
16. Chioncel, O. et al. Epidemiology and one-year outcomes in patients with chronic heart failure and preserved, mid-range and reduced ejection fraction: an analysis of the ESC Heart Failure Long-Term Registry. *Eur. J. Heart Fail.* **19**, 1574–1585 (2017).
17. Shah, K. S. et al. Heart failure with preserved, borderline, and reduced ejection fraction: 5-year outcomes. *J. Am. Coll. Cardiol.* **70**, 2476–2486 (2017).
18. Papolos, A., Narula, J., Bavishi, C., Chaudhry, F. A. & Sengupta, P. P. U.S. hospital use of echocardiography: insights from the nationwide inpatient sample. *J. Am. Coll. Cardiol.* **67**, 502–511 (2016).
19. Douglas, P. S. et al. ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 Appropriate use criteria for echocardiography. *J. Am. Soc. Echocardiogr.* **24**, 229–267 (2011).
20. Zhang, J. et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation* **138**, 1623–1635 (2018).
21. Madani, A., Arnaout, R., Mofrad, M. & Arnaout, R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit. Med.* **1**, 6 (2018).
22. Ghorbani, A. et al. Deep learning interpretation of echocardiograms. *NPJ Digit. Med.* **3**, 10 (2020).
23. Behnami, D. et al. in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 65–73 (Springer, 2018).
24. Ardila, D. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **25**, 954–961 (2019).
25. Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018).
26. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
27. Coudray, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
28. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. Preprint at https://arxiv.org/abs/1706.05587 (2017).
29. Tran, D. et al. A closer look at spatiotemporal convolutions for action recognition. In *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 6450–6459 (2018).
30. Tran, D., Bourdev, L., Fergus, R., Torresani, L. & Paluri, M. Learning spatiotemporal features with 3D convolutional networks. In *Proc. IEEE International Conference on Computer Vision* 4489–4497 (2015).

Methods

Data curation

A standard full resting echocardiogram study consists of a series of 50–100 videos and still images visualizing the heart from different angles, locations and image acquisition techniques (two-dimensional images, tissue Doppler images, colour Doppler images and others). Each echocardiogram video corresponds to a unique patient and a unique visit. In this dataset, one apical four-chamber two-dimensional greyscale video is extracted from each study. Each video represents a unique individual as the dataset contains 10,030 echocardiography videos from 10,030 unique individuals who underwent echocardiography between 2016 and 2018 as part of clinical care at Stanford Health Care. Videos were randomly split into 7,465, 1,277 and 1,288 patients, respectively, for the training, validation and test sets.

The randomly selected patients in our data have a range of ejection fractions representative of the patient population who visit the echocardiography laboratory (Extended Data Table 1). Videos were acquired by skilled sonographers using iE33, Sonos, Acuson SC2000, Epiq 5G or Epiq 7C ultrasound machines and processed images were stored in a Philips Xcelera picture archiving and communication system. Video views were identified through implicit knowledge of view classification in the clinical database by identifying images and videos labelled with measurements done in the corresponding view. For example, apical four-chamber videos were identified by selecting videos from the set of videos in which a sonographer or cardiologist traced left ventricle volumes and labelled these for analysis to calculate ejection fraction. The apical four-chamber view video was thus identified by extracting the Digital Imaging and Communications in Medicine (DICOM) file linked to the measurements of the ventricular volume used to calculate the ejection fraction.

An automated preprocessing workflow was used to remove identifying information and eliminate unintended human labels. Each subsequent video was cropped and masked to remove text, electrocardiogram and respirometer information, and other information outside of the scanning sector. The resulting square images were either 600 × 600 or 768 × 768 pixels depending on the ultrasound machine and down sampled by cubic interpolation using OpenCV into standardized 112 × 112 pixel videos. Videos were spot-checked for quality control, to confirm view classification and to exclude videos with colour Doppler.

This research was approved by the Stanford University Institutional Review Board and data privacy review through a standardized workflow by the Center for Artificial Intelligence in Medicine and Imaging (AIMI) and the University Privacy Office. In addition to masking of text, electrocardiogram information and extra data outside of the scanning sector in the video files as described above, the video data of each DICOM file was saved as an AVI file to prevent any leakage of identifying information through public or private DICOM tags. Each video was subsequently manually reviewed by an employee of the Stanford Hospital familiar with imaging data to confirm the absence of any identifying information before public release.

EchoNet-Dynamic development and training

Model design and training was done in Python using the PyTorch deep learning library. Semantic segmentation was performed using the Deeplabv3 architecture[30]. The segmentation model had a base architecture of a 50-layer residual net and minimized pixel-level binary cross-entropy loss. The model was initialized with random weights and was trained using a stochastic gradient descent optimizer (Extended Data Fig. 3). Our model with spatiotemporal convolutions was initialized with pretrained weights from the Kinetics-400 dataset[31]. We tested three model architectures with variable integration of temporal convolutions (R3D, MC3 and R2+1D) and ultimately chose decomposed R2+1D spatiotemporal convolutions as the architecture with the best performance to use for EchoNet-Dynamic[29,30] (Extended Data Fig. 1 and Extended Data Table 2). In the R3D architecture, all convolutional layers considered the spatial and temporal dimensions jointly and these consisted of five convolutional blocks. The MC3 and R2+1D architectures were introduced as a middle ground between two-dimensional convolutions that considered only spatial relationships and the full three-dimensional convolutions used by R3D[29]. The MC3 architecture replaced the convolutions in the final three blocks with two-dimensional convolutions, and the R2+1 architecture explicitly factored all of the three-dimensional convolutions into a two-dimensional spatial convolution followed by a one-dimensional temporal convolution.

For predicting ejection fraction, the models were trained to minimize the squared loss between the prediction and true ejection fraction using a stochastic gradient descent optimizer with an initial learning rate of 0.0001, momentum of 0.9 and batch size of 16 for 45 epochs. The learning rate was decayed by a factor of 0.1 every 15 epochs. For model input, video clips of 32 frames were generated by sampling every other frame (sampling period of 2) with both clip length and sampling period determined by hyperparameter search (Extended Data Fig. 1). During training, to augment the size of the dataset and increase the variation of exposed training clips, each training video clip was padded with 12 pixels on each side, and a random crop of the original frame size was taken to simulate slight translations and changes in camera location. For all models, the weights from the epoch with the lowest validation loss was selected for final testing. Model computational cost was evaluated using one NVIDIA GeForce GTX 1080 Ti GPU (Extended Data Fig. 4).

Test time augmentation with beat-to-beat assessment

There can be variation in the ejection fraction, end-systolic and end-diastolic volumes during atrial fibrillation, and in the setting of premature atrial contractions, premature ventricular contractions and other sources of ectopy. The clinical convention is to identify at least one representative cardiac cycle and use this representative cardiac cycle to perform measurements, although an average of the measurements of up to five cardiac cycles is recommended when there is considerable ectopy or variation. For this reason, our final model used test time augmentation by providing individual estimates for each ventricular beat throughout the entire video and outputs the average prediction as the final model prediction. We use the segmentation model to identify the area of the left ventricle and threshold-based processing to identify ventricular contractions during each cardiac cycle. Each ventricular contraction (systole) was identified by choosing the frames of the smallest left ventricle size as identified by the segmentation arm of EchoNet-Dynamic. For each beat, a subsampled clip centred around the ventricular contraction was obtained and used to produce a beat-to-beat estimate of ejection fraction. The mean ejection fraction of all ventricular contractions in the video was used as the final video-level model prediction.

Assessing model performance and prospective clinical validation

We evaluated the relationship between model performance and the quality of the echocardiogram video. Our dataset was not curated on clinical quality and we did not exclude any videos due to insufficient image quality. On the internal Stanford test dataset, we evaluated the model performance with variation in video saturation and gain, and the performance of EchoNet-Dynamic is robust to the range of the acquisition quality of the clinical images (Extended Data Fig. 5). To further test the effect of variable video quality, we simulated noise and degraded video quality by randomly removing a proportion of pixels from videos in the test dataset and evaluated model performance on the degraded images (Extended Data Fig. 6). EchoNet-Dynamic is also robust to a wide range of synthetic noise and image degradation.

Prospective validation was performed by two senior sonographers with advanced cardiac certification and more than 15 years of experience each. For each patient, measurements of cardiac function were independently acquired and assessed by each sonographer on the same day. Every patient was scanned using Epiq 7C ultrasound machines, the standard instrument in the Stanford Echocardiography Laboratory, and a subset of patients were also rescanned by the same two sonographers using a GE Vivid 95E ultrasound machine. Tracing and measurements were done on a dedicated workstation after image acquisition. For comparison, the independently acquired apical four-chamber videos were fed into the model and the variance in measurements assessed.

External healthcare system test dataset

Transthoracic echocardiogram studies from November 2018 to December 2018 from an independent external healthcare system, Cedars-Sinai Medical Center, were used to evaluate the performance of EchoNet-Dynamic in predicting ejection fraction. The same automated preprocessing workflow was used to convert DICOM files to AVI files, mask information outside of the scanning sector and resize input to 112 × 112-pixel videos of variable length. Previously described methods were used to identify apical four-chamber view videos[22]. After manual exclusion of incorrect classifications, long cine loops of bubble studies, videos with injection of ultrasonic contrast agents and videos with colour Doppler, we identified 2,895 videos from 1,267 patients. These videos were used as the input for EchoNet-Dynamic trained on the Stanford dataset and model predictions were compared with human interpretations from physicians at Cedars-Sinai Medical Center. The input video sampling period was set to one as the frame rate of the external dataset was roughly half that of videos from the Stanford dataset. Model predictions from multiple videos of the same patient were averaged to produce a composite estimate of ejection fraction.

Re-evaluation by expert clinicians

Recognizing the inherent variation in human assessment of ejection fraction[5,13–16], five expert sonographers and cardiologists who specialize in cardiovascular imaging performed a blinded review of the echocardiogram videos with the highest absolute difference between the initial human label and the prediction by EchoNet-Dynamic (mean absolute difference of 15.0%, s.d. of 3.79%). Each expert independently received the relevant echocardiogram video and a set of two blinded measurements of ejection fractions that corresponded to the initial human label and the prediction by EchoNet-Dynamic. The experts were asked to select which ejection fraction corresponded more closely to their evaluation of ejection fraction as well as to note any limitations in echocardiogram video quality that would hinder their interpretation. In the blinded review, experts noted that 38% (15 out of 40) of videos had considerable issues with video quality or acquisition and that 13% (5 out of 40) of videos had marked arrhythmia, limiting human assessment of ejection fraction (Extended Data Table 3). In this setting, the consensus interpretation of the expert clinicians preferred the prediction by EchoNet-Dynamic over the initial human label in 43% (17 out of 40) of the echocardiogram videos.

Statistical analysis

No statistical methods were used to predetermine sample size. Confidence intervals were computed using 10,000 bootstrapped samples and obtaining 95 percentile ranges for each prediction. The performance of the semantic segmentation task was evaluated using the Dice similarity coefficient compared with the human labels from the held-out test dataset. The performance of the ejection fraction task was evaluated by calculating the mean absolute difference between the prediction of EchoNet-Dynamic and the human calculation of ejection fraction as well as calculating the R^2 between the prediction by EchoNet-Dynamic and the human calculation. Prospective comparison with human readers was performed with the uniformly most powerful invariant equivalence test for two-sample problems.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

This project introduces the EchoNet-Dynamic dataset, a publicly available dataset of de-identified echocardiogram videos, which are available at https://echonet.github.io/dynamic/.

Code availability

The code for EchoNet-Dynamic is available at https://github.com/echonet/dynamic.

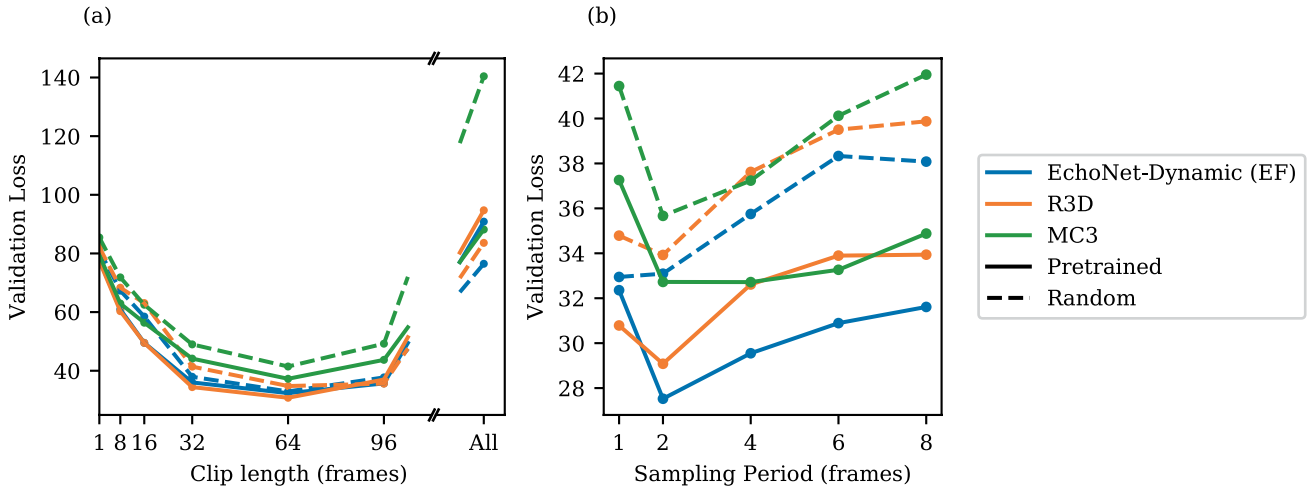
31. Kay, W. et al. The kinetics human action video dataset.Preprint at https://arxiv.org/abs/1705.06950 (2017).

Acknowledgements This work is supported by a Stanford Translational Research and Applied Medicine pilot grant, a Stanford Cardiovascular Institute pilot grant and a Stanford Artificial Intelligence in Imaging and Medicine Center seed grant. D.O. is supported by the American College of Cardiology Foundation – Merck Research Fellowship and NIH F32HL149298. B.H. is supported by a NSF Graduate Research Fellowship. A.G. is supported by the Stanford Robert Bosch Graduate Fellowship in Science and Engineering. J.Y.Z. is supported by NSF CCF 1763191, NIH R21 MD012867-01, NIH P30AG059307 and by a Chan-Zuckerberg Biohub Fellowship.

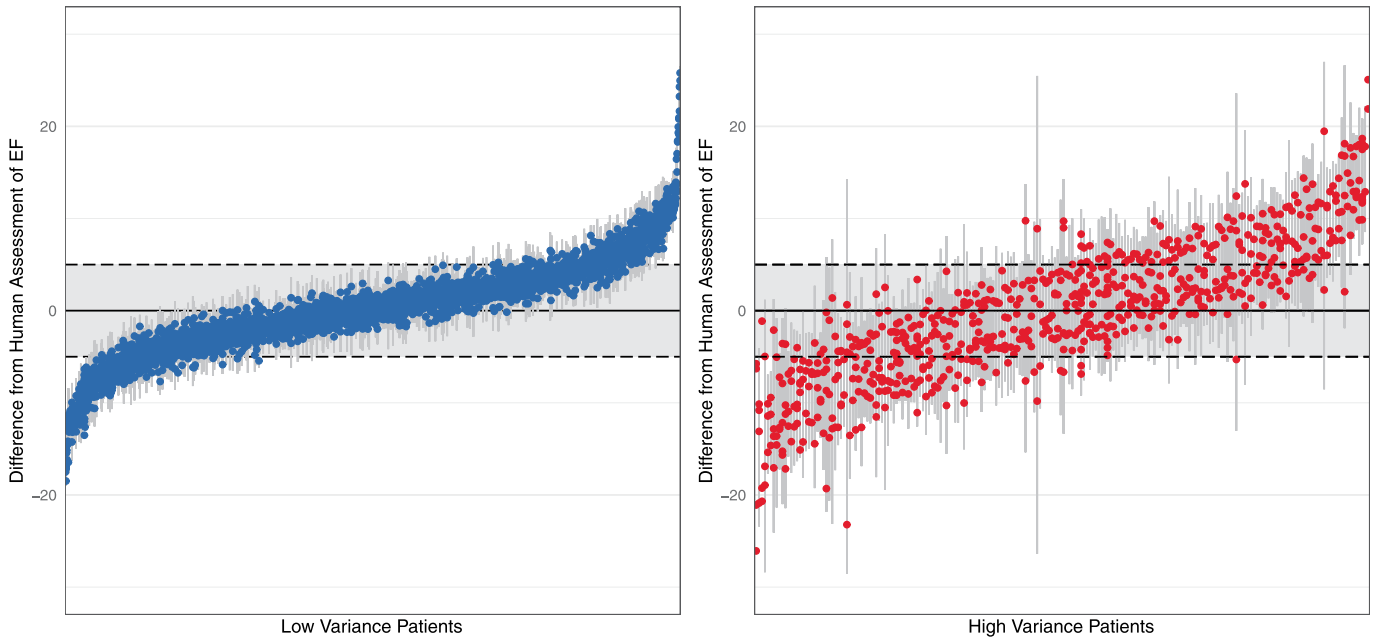
Author contributions D.O. retrieved, preprocessed and quality-controlled Stanford videos and merged electronic medical record data. D.O., B.H., A.G. and J.Y.Z. developed and trained the deep learning algorithms, performed statistical tests and created all of the figures. D.O., C.P.L., P.A.H. and R.A.H. coordinated the public release of the de-identified echocardiogram dataset. D.O., P.A.H., D.H.L. and E.A.A. performed the clinical evaluation of model performance. N.Y. and J.E. retrieved, preprocessed and quality-controlled data from Cedar-Sinai Medical Center for model testing. D.O., B.H., E.A.A. and J.Y.Z. wrote the manuscript with feedback from all authors.

Competing interests The authors declare no competing interests.

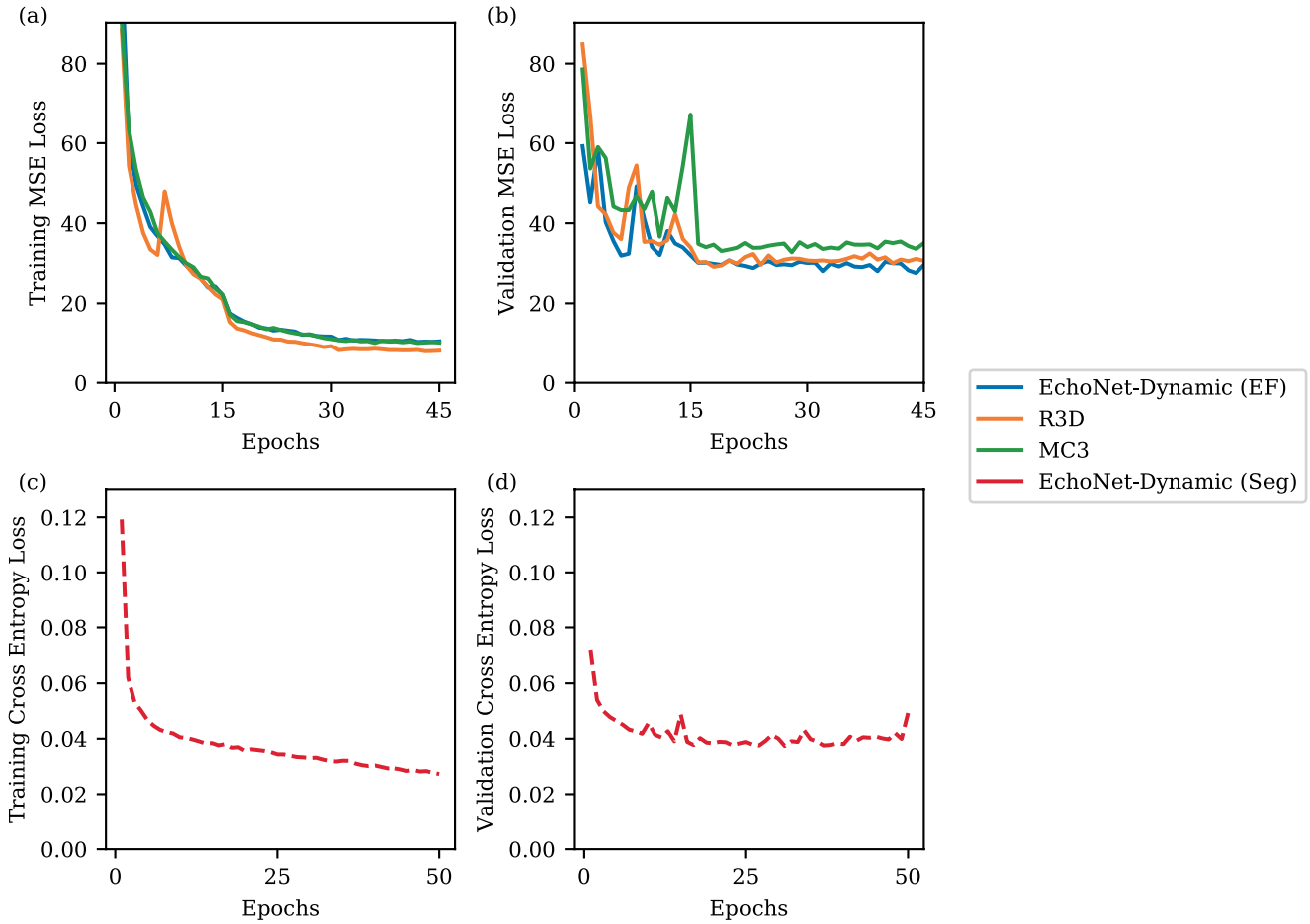
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2145-8.
Correspondence and requests for materials should be addressed to D.O. or J.Y.Z.
Peer review information *Nature* thanks Giorgio Quer, Partho Sengupta and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



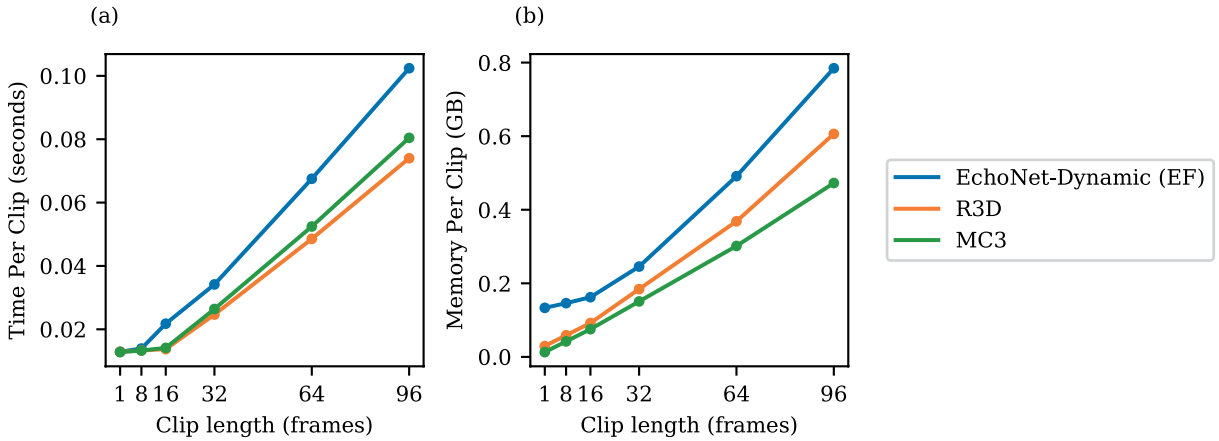
Extended Data Fig. 1 | Hyperparameter search for spatiotemporal convolutions on the video dataset to predict ejection fraction. Model architecture (R2+1D, which is the architecture selected by EchoNet-Dynamic for ejection fraction prediction, R3D and MC3), initialization (solid line, Kinetics-400 pretrained weights; dotted line, random initial weights), clip length (1, 8, 16, 32, 64, 96 and all frames) and sampling period (1, 2, 4, 6 and 8) were considered. **a**, When varying clip lengths, performance is best at 64 frames (corresponding to 1.28 s) and starting from pretrained weights improves performance slightly across all models. **b**, Varying sampling period with a length to approximately correspond to 64 frames before subsampling. Performance is best with a sampling period of 2.



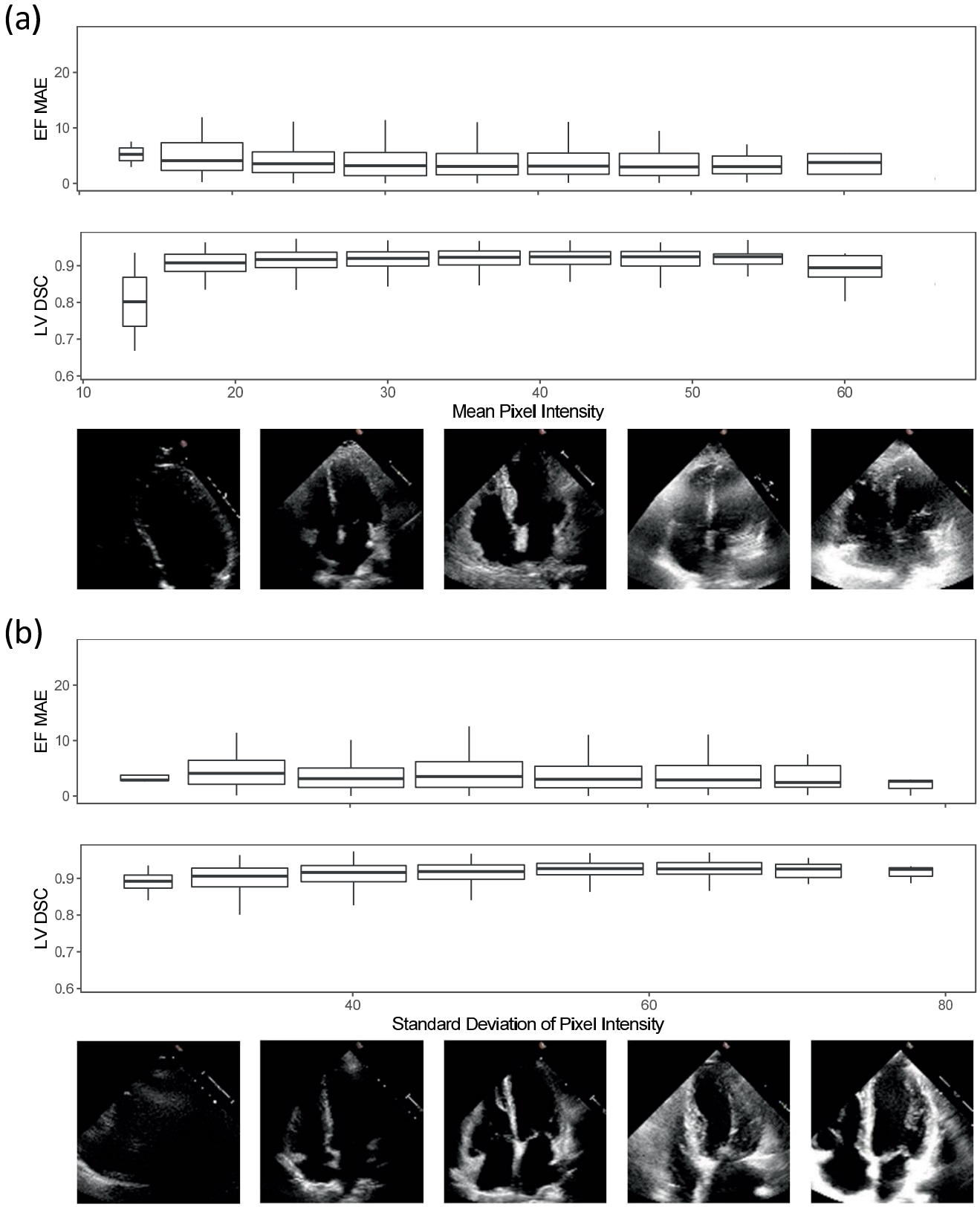
Extended Data Fig. 2 | Individual beat assessment of ejection fraction for each clip in the test dataset. Left, patients with low variance across beats (s.d. < 2.5, $n = 3,353$); right, patients with high variance across beats (s.d. > 2.5, $n = 717$). Each patient video is represented by multiple points that represent the estimate of each beat and a line that indicates 1.96 s.d. from the mean. A greater proportion of beats are within 5% of the ejection fraction estimate made by the human observer (the shaded regions) in videos with low variance compared with individual beat assessment of ejection fraction in high-variance patients.



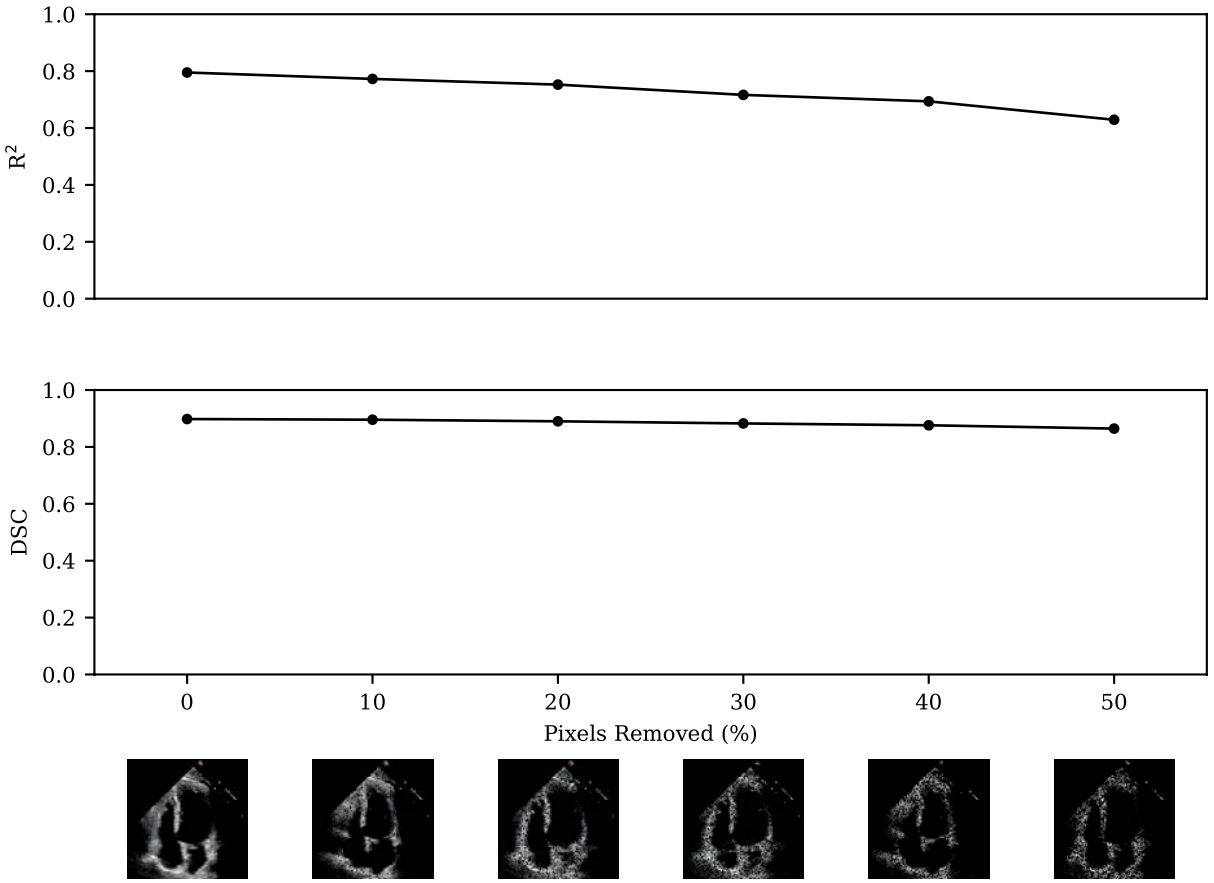
Extended Data Fig. 3 | Model performance during training. a, **b**, Mean square error (MSE) loss for the prediction of left ventricular ejection fraction during training on the training (**a**) and validation (**b**) dataset. Pixel-level cross-entropy loss for semantic segmentation of the left ventricle during training on the training (**c**) and validation (**d**) dataset.



Extended Data Fig. 4 | Relationship between clip length, and speed and memory. Hyperparameter search for model architecture (R2+1D, which is used by EchoNet-Dynamic for ejection fraction prediction, R3D and MC3) and input video clip length (1, 8, 16, 32, 64 and 96 frames) and impact on model processing time and model memory usage.



Extended Data Fig. 5 | Variation in echocardiogram video quality and relationship with EchoNet-Dynamic model performance. a, **b**, Representative quintile video frames are shown with the respective mean pixel intensity (**a**) and the s.d. of the pixel intensity (**b**) compared with the mean absolute error of the ejection fraction prediction of EchoNet-Dynamic (EF MAE) and the Dice similarity coefficient for segmentation of the left ventricle (LV DSC). Box plots show the median as a thick line, the 25th and 75th percentiles as upper and lower bounds of the box, and whiskers extend to 1.5× the interquartile range from the median. $n = 1,277$.



Extended Data Fig. 6 | Impact of degraded image quality on model performance. Random pixels were removed and replaced with pure black pixels to simulate ultrasound dropout. Representative video frames with dropout are shown across a range of dropouts. The proportion of dropout was compared with model performance with respect to the R^2 of the prediction of ejection fraction and the Dice similarity coefficient (DSC) compared with human segmentation of the left ventricle.

Extended Data Table 1 | Summary statistics of patient and device characteristics in the Stanford dataset

Statistic	Total	Training	Validation	Test
Number of Patients	10,030	7,465	1,288	1,277
Demographics				
Age, years (SD)	68 (21)	70 (22)	66 (18)	67 (17)
Female, n (%)	4,885 (49%)	3,662 (49%)	611 (47%)	612 (48%)
Heart Failure, n (%)	2,874 (29%)	2,113 (28%)	356 (28%)	405 (32%)
Diabetes Mellitus, n (%)	2,018 (20%)	1,474 (20%)	275 (21%)	269 (21%)
Hypercholesterolemia, n (%)	3,321 (33%)	2,463 (33%)	445 (35%)	413 (32%)
Hypertension, n (%)	3,936 (39%)	2,912 (39%)	525 (41%)	499 (39%)
Renal Disease, n (%)	2,004 (20%)	1,475 (20%)	249 (19%)	280 (22%)
Coronary Artery Disease, n (%)	2,290 (23%)	1,674 (22%)	302 (23%)	314 (25%)
Metrics				
Ejection Fraction, % (SD)	55.7 (12.5)	55.7 (12.5)	55.8 (12.3)	55.3 (12.4)
End Systolic Volume, mL (SD)	43.3 (34.5)	43.2 (36.1)	43.3 (34.5)	43.9 (36.0)
End Diastolic Volume, mL (SD)	91.0 (45.7)	91.0 (46.0)	91.0 (43.8)	91.4 (46.0)
Machine				
Epiq 7C, n (%)	6,505 (65%)	4,832 (65%)	843 (65%)	830 (65%)
iE33, n (%)	3,329 (33%)	2,489 (33%)	421 (33%)	419 (33%)
CX50, n (%)	83 (1%)	62 (1%)	12 (1%)	9 (1%)
Epiq 5G, n (%)	60 (1%)	44 (1%)	5 (0%)	11 (1%)
Other, n (%)	53 (1%)	38 (1%)	7 (1%)	8 (1%)
Transducer				
X5, n (%)	6,234 (62%)	4,649 (62%)	794 (62%)	791 (62%)
S2, n (%)	2,590 (26%)	1,913 (26%)	345 (27%)	332 (26%)
S5, n (%)	1,149 (12%)	863 (12%)	141 (11%)	145 (11%)
Other or Unspecified, n (%)	57 (1%)	40 (1%)	8 (1%)	9 (1%)
Day of the Week				
Monday, n (%)	1,555 (16%)	1,165 (16%)	210 (16%)	180 (14%)
Tuesday, n (%)	1,973 (20%)	1,411 (19%)	269 (21%)	293 (23%)
Wednesday, n (%)	2,078 (21%)	1,522 (20%)	270 (21%)	286 (23%)
Thursday, n (%)	2,144 (21%)	1,642 (22%)	248 (19%)	254 (20%)
Friday, n (%)	2,018 (20%)	1,461 (20%)	237 (18%)	221 (17%)
Saturday, n (%)	221 (2%)	155 (2%)	35 (3%)	31 (2%)
Sunday, n (%)	140 (1%)	109 (1%)	19 (1%)	12 (1%)

Data obtained from visits to Stanford Hospital between 2016 and 2018.

Extended Data Table 2 | Performance of EchoNet-Dynamic compared with alternative deep learning architectures in assessing cardiac function

Model	Evaluation	Sampling Period	MAE	RMSE	R^2
EchoNet-Dynamic	Beat-by-beat	1 in 2	4.05	5.32	0.81
EchoNet-Dynamic (EF)	32 frame sample	1 in 2	4.22	5.56	0.79
R3D	32 frame sample	1 in 2	4.22	5.62	0.79
MC3	32 frame sample	1 in 2	4.54	5.97	0.77
EchoNet-Dynamic (EF)	All frames	All	7.35	9.53	0.40
R3D	All frames	All	7.63	9.75	0.37
MC3	All frames	All	6.59	9.39	0.42

EchoNet-Dynamic with beat-by-beat evaluation refers to the full model, which included using segmentation of the left ventricle to identify each ventricular contraction for prediction aggregation. Frame sampling refers to the performance of the base architecture on individual video clips or simple averaging across the entire video. We trained all of these architectures on the same set of Stanford videos. n = 1,277.

Extended Data Table 3 | Videos with the most discordance between model prediction and human label of ejection fraction

Video File	Model EF	Human EF	Difference	Rev 1	Rev 2	Rev 3	Rev 4	Rev 5	Notes
0X4EFB94EA8F9FC7C2	44.6	17.7	26.9	B	B	B	B	A	Poor image quality
0XB3FFC4AE334E9F4	55.4	30.1	25.4	B	A	B	B	B	
0X15C0D7DBFF8E4FC8	57.8	34.8	23.1	A	A	A	A	A	Poor image quality, incorrect label
0X41AC5C5FC2E3352A	59.5	37.9	21.6	A	A	A	A	A	Arrhythmia, incorrect label
0X211D307253ACBEE7	31.0	10.7	20.3	A	B	B	B	B	Poor image quality, foreshortening
0X5A8D9673920F03FE	46.3	26.7	19.6	A	B	B	B	B	
0X75AF130134AADF00	46.8	28.4	18.4	B	A	B	A	B	Arrhythmia
0X6703FCFAD2E7CBCA	37.1	54.5	17.4	A	A	A	A	A	Poor image quality, incorrect label
0X345D4E0B1B2EBAA1	67.2	84.5	17.3	B	A	A	A	A	Incorrect label
0X1B2BCDAE290F6015	43.8	28.4	15.4	B	A	B	A	B	Poor image quality
0X15CC6C50F1763B61	34.9	50.2	15.2	A	A	B	B	B	
0X7D567F2A870FD8F0	44.1	29.1	15.0	B	B	B	A	A	
0X2507255D8DC30B4E	52.0	66.8	14.8	A	A	A	A	A	Poor image quality, incorrect label
0X493E34208D40DBB5	51.8	37.1	14.7	B	B	B	A	A	
0X56FD0409BFA202DF	39.1	53.7	14.6	A	A	B	A	A	Incorrect label
0X2D4304FA6A09F93E	37.2	23.2	14.0	A	B	A	B	B	
0X1EDA0F3F33F97A9D	49.0	35.0	14.0	B	B	B	B	B	Poor image quality
0X66C8EAE88FFB77EE	54.0	40.1	13.9	B	B	A	B	B	
0X1EF35FFC92F4F554	47.7	61.4	13.7	B	A	A	B	A	Incorrect label
0X1CDE7FECA3A1754B	37.2	50.9	13.7	B	A	A	A	B	Arrhythmia, incorrect label
0X777692B30E35465A	39.4	53.0	13.6	B	A	B	B	B	
0X29E66C557C99EC32	52.0	65.5	13.6	B	B	B	B	B	
0X31B6E6B67B97806A	54.0	40.4	13.5	A	A	A	A	A	Incorrect label
0X30DF42C999969D67	43.4	30.0	13.3	B	B	B	B	B	
0X36715FD73D74BF39	49.2	36.0	13.2	A	B	B	B	A	Poor image quality, Effusion
0XAA3E06425E1A23E	46.3	33.1	13.1	A	B	A	B	A	Incorrect label
0X60361B7F301DEBB7	55.2	68.3	13.0	B	A	A	A	A	Incorrect label
0X32AFF6A0BED73A67	74.2	61.6	12.6	A	A	A	B	A	Incorrect label
0X8558D35ED09F890	52.7	40.4	12.4	A	B	B	A	A	Poor image quality, incorrect label
0X868028466F66DE2	43.9	56.2	12.3	B	A	B	B	B	
0X41130893A44122AB	54.0	66.3	12.3	B	A	B	B	A	Poor image quality
0X69447E46FEDD2A3F	49.3	61.5	12.3	A	A	A	B	A	Poor image quality, incorrect label
0X797CA10A7CDE384B	62.8	75.0	12.2	B	A	B	A	A	Poor image quality, incorrect label
0XBCEAB22A81A23C1	25.4	13.3	12.2	A	B	B	B	A	Foreshortening
0X2889D8C33077C148	57.0	69.1	12.1	B	A	A	B	B	Arrhythmia
0X6DFE8F195ACC3BA4	18.0	30.1	12.1	B	A	B	B	B	
0X62431BB9CF3A33EE	44.4	56.4	12.1	A	A	A	A	A	Arrhythmia, incorrect label
0X27250C8B6DF1D971	67.7	79.7	12.0	B	A	B	B	A	Poor image quality, foreshortening
0X79DFCFF4867CB797	31.9	43.9	12.0	B	A	B	B	B	
0X2DF88C27BB20C25D	55.9	43.9	12.0	B	A	A	B	A	Foreshortening, incorrect label

'A' indicates expert preference for the prediction by EchoNet-Dynamic and 'B' indicates preference for the initial human label. The label of 'Incorrect label' was used when at least three out of five blinded experts preferred the prediction of the ejection fraction made by EchoNet-Dynamic over the initial human label in a side-by-side comparison.

Extended Data Table 4 | Model parameters and computational cost

Task	Model	Parameters (millions)	Time per prediction (sec)		Memory per prediction (GB)	
			Train	Test	Train	Test
End-to-end	EchoNet-Dynamic	71.1	0.221	0.048	1.191	0.276
EF Prediction	EchoNet-Dynamic (EF)	31.5	0.150	0.034	1.055	0.246
	R3D	33.4	0.084	0.025	0.394	0.184
	MC3	11.7	0.110	0.035	0.489	0.151
Segmentation	EchoNet-Dynamic (Seg)	39.6	0.071	0.014	0.136	0.030

nature research

Corresponding author(s): David Ouyang and James Zou

Last updated by author(s): Nov 15, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The code used for this study is made publicly available at https://echonet.github.io/dynamic/
Data analysis	Data analysis was performed using R 3.6.1 and Python 3.0. The most relevant Python packages used were pytorch (1.3.1), torchvision (0.4.2), scipy (1.3.2), opencv (4.1.1), numpy (1.17.4), scikit-image (0.16.2), and scikit-learn(0.21.3). The most relevant R packages include ggplot2 (3.2.1) and dplyr (0.8.3). Our code is available at https://github.com/echonet/. A comprehensive list of dependencies is available at https://github.com/echonet/dynamic/blob/master/requirements.txt

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data is made publicly available at https://echonet.github.io/dynamic/ under a non-commercial data use agreement.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. Deep learning models were trained while varying the number of input videos until asymptotic improvement of model performance was achieved to suggest an appropriate sample size was obtained.
Data exclusions	No data was excluded for the database.
Replication	Hyperparameter search was performed on the validation set for multiple iterations of the dataset for replication of model results. All replications were successful and confidence intervals when applicable represent variation of results or performance during bootstrapping.
Randomization	The training, validation, and test datasets were randomly split to approximately (75%/12.5%/12.5%). The split generated is available as part of our dataset release available at http://echonet.github.io/dynamic.
Blinding	During prospective validation with two sonographers, both sonographers were blinded to the interpretation of the other sonographer.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Extended data table 1 detail relevant patient demographics and imaging characteristics. Patients were randomly sampled from the imaging database. Data shown to editors and reviewers show that the patient demographics and imaging characteristics are consistent with demographics of all patients who undergo imaging at the hospital.
Recruitment	Patients were not directly involved or recruited in the study. This study involved retrospective evaluation of imaging studies from patients obtained during standard clinical care. Patient imaging was de-identified. A waiver of consent was obtained from the Stanford University IRB.
Ethics oversight	The study was approved by the Stanford University IRB

Note that full information on the approval of the study protocol must also be provided in the manuscript.