An End-to-End Computer Vision Pipeline for Automated Cardiac Function Assessment by Echocardiography

Rahul C. Deo MD PhD1,2,3,4,5*, Jeffrey Zhang6, Laura A. Hallock BSc6, Sravani Gajjala1, Lauren Beussink-Nelson MHS, RDCT7, Geoffrey H. Tison MD, MPH2, Eugene Fan MD2, Mandar A. Aras MD PhD1,2, ChaRandle Jordan MD, PhD2, Kirsten E. Fleischmann MD, MPH2, Michelle Melisko MD2, Atif Qasim MD, MSCE2, Sanjiv J. Shah MD5, Ruzena Bajcsy PhD6

1Cardiovascular Research Institute, University of California, San Francisco, USA
2Department of Medicine, University of California, San Francisco, USA
3Institute for Human Genetics, University of California, San Francisco, USA
4California Institute for Quantitative Biosciences, San Francisco, California, USA
5Institute for Computational Health Sciences, University of California, San Francisco, USA
6Department of Electrical Engineering and Computer Science, University of California at Berkeley, USA
7Division of Cardiology, Department of Medicine; and Feinberg Cardiovascular Research Institute, Northwestern University
Feinberg School of Medicine, Chicago, IL

Abstract

Background: Automated cardiac image interpretation has the potential to transform clinical practice in multiple ways including enabling low-cost assessment of cardiac function in the primary care setting. We hypothesized that advances in computer vision could enable building a fully automated, scalable analysis pipeline for echocardiogram (echo) interpretation, with a focus on evaluation of left ventricular function.

Methods: Our approach entailed: 1) preprocessing, which includes auto-downloading of echo studies, metadata extraction, de-identification, and conversion of images into numerical arrays; 2) convolutional neural networks (CNN) for view identification; 3) localization of the left ventricle and delineation of cardiac boundaries using active appearance models (AAM); 4) identification of properly segmented images using gradient boosting; and 5) particle tracking to compute longitudinal strain.

Results: CNNs identified views with high accuracy (e.g. 98.6% for apical 4-chamber) and the combination of CNN/bounding box/AAM accurately segmented 67-88% of videos. We analyzed 2775 apical videos from patients with heart failure and found good concordance with vendor-derived longitudinal strain measurements at the individual video level ($\rho = 0.77$) and at the patient level ($\rho = 0.51$). We also analyzed 9402 videos from breast cancer patients undergoing serial monitoring for trastuzumab cardiotoxicity to illustrate the potential for automated, quality-weighted modeling of patient trajectories.

Conclusions: We demonstrate the feasibility of a fully automated echocardiography analysis pipeline for assessment of left ventricular function. Our work lays the groundwork for using automated interpretation to support point-of-care handheld cardiac ultrasound and may enable large-scale analysis of the millions of echos currently archived within healthcare systems.

I. INTRODUCTION

CARDIAC muscle disease often progresses for years prior to the onset of symptoms. This process, known as cardiac remodeling, can accompany conditions such as valvular disease, hypertension and diabetes mellitus, and result in pathologic changes to the heart that are difficult to reverse once established (1). Although early evidence of remodeling is often detectable by imaging (2) and could in principle be tracked longitudinally in a personalized manner, the cost of imaging all individuals with cardiac risk factors would be prohibitive.

Automated image interpretation could enable such monitoring at far lower costs, especially when coupled with inexpensive data acquisition. For echocardiography one such strategy could involve handheld ultrasound devices used by non-experts (3) at point of care locations (e.g. primary care clinic) and a cloud-based automated interpretation system that assesses cardiac structure and function and compares results to one or more prior studies. Automated image interpretation could also enable surveillance of echo data collected at a given center and could be coupled with statistical models to highlight early evidence of dysfunction or detect rare myocardial diseases. Such an approach could, for example, enable systematic comparison across the tens of millions of echocardiograms completed each year in the Medicare population alone (4).

Automated image interpretation falls under the discipline of computer vision, which, in turn, is a branch

*To whom correspondence should be addressed:
rahul.deo@ucsf.edu
of machine learning where computers learn to mimic human vision (5). Although the application of computer vision to medical imaging has been longstanding (6), recent advances in computer vision algorithms, processing power, and a massive increase in digital labeled data has resulted in a striking improvement in classification performance for several test cases, including retinal (7) and skin disease (8). Echocardiography, nonetheless, presents challenges beyond these examples. Rather than comprising a single still image, a typical echo study consists of up to 50 videos collected from different viewpoints, and viewpoints are not labeled in each study. Furthermore, measurements can vary from video to video because of intrinsic beat-to-beat variability in cardiac performance as well as variability from the process of approximating a three-dimensional object using two-dimensional cross-sectional images. Given the extent of this variability as well the sheer amount of multidimensional information in each study that often goes unused, it appears that echo would benefit automated learning approach to assist human interpretation.

In this manuscript, we present the first fully automated end-to-end computer vision pipeline for automated interpretation of cardiac function, using a combination of machine learning strategies including convolutional neural networks, active appearance models, particle tracking, and gradient boosting. We demonstrate the scalability of our approach by analyzing over 10,000 videos and validate our accuracy against commercial vendor packages. We describe some of the challenges we encountered in the process as well as potential promising applications.

II. METHODS

Human Subjects Research

The institutional review board at the University of California San Francisco granted permission to download echocardiographic data as well as review relevant clinical data for this study.

Auto-downloading of DICOM format echo studies from the Syngo client

We realized at the onset of this project that studies stored within our echo database (Syngo, Siemens Healthcare) were in a proprietary format that could not be used for image analysis. To avoid manual download of the thousands of studies used for this project, we wrote scripts using AutoIt software (https://www.autoitscript.com/site/autoit/) to mimic human interaction with the web-based client. This enabled downloading individual studies in Digital Imaging and Communications in Medicine (DICOM) format specified by date or medical record number at a rate of approximately 1 study per 2 minutes.

Preprocessing

Typical echo studies consist of a combination of 80-120 still images and videos. The still images are usually used for manual measurements and thus our primary interest was in the videos. We first used the exiftool utility to count the number of frames within each file thus enabling separation of still images from videos. We next used the gdcmconv utility from the Grassroots Dicom Library (GDCM) to convert compressed DICOM format videos into a raw DICOM format. This allowed use of the pydicom Python library, which enabled conversion of DICOM videos into single channel (i.e. grayscale) numerical arrays. In doing so, we also “blacked out” the identifying patient information on the videos by setting the corresponding pixel intensities to minimal intensity. Numerical arrays were compressed for subsequent use. A subset of these were converted into Audio Video Interleaved (avi) format for manual segmentation.

To extract metadata corresponding to each file, we used the gdcmdump utility from the GDCM library. We were particularly interested in frame time (i.e. time interval between adjacent frames), heart rate, number of columns and rows, and the dimensions in physical units (i.e. centimeters) corresponding to each pixel, as these would be needed for strain measurements. We removed identifying information (name, birth date) and created a compressed metadata file corresponding to each study.

Convolutional Neural Network Based View Identification

Videos in echo studies do not specify which view is depicted. We thus set out to develop a classifier to automate view identification. We elected to use convolutional neural networks (CNNs) based on their superior performance in image classification. CNNs are one algorithm used for "deep learning", a machine learning approach which employs multiple layers of learning, where each layer develops a successively higher order abstraction of the layer below (9).

We based our approach on the VGG architecture by Simonyan & Zisserman (10). The network takes in a fixed-sized input of grayscale images with dimensions 224x224 pixels. Each image is passed through thirteen convolution layers, five max-pool layers, and three fully connected layers. All convolutional layers consist of 3x3 filters with stride 1 and all max-pooling is applied over a 2x2 window with stride 2. The stack of convolutions is followed by two fully connected layers, each with 500 hidden units, and a final fully connected layer with 8 output units. The output is fed into a 6-way softmax to represent six different echo views: parasternal long-axis (PLAX), parasternal short-axis at the papillary muscle (PSAX), apical 2-, 3-, and 4-chamber (A2c, A3c, and A4c), and inferior vena cava (IVC). The view with the highest probability was selected as the predicted view.
(The final model had 9 classes, including “other” and “zoomed-in” versions of A2c, A3c, and A4c).

Additionally, each echo contains periphery information unique to different output settings on ultrasound machines used to collect the data. This periphery information details additional details collected (i.e. electrocardiogram, blood pressure, etc.). To improve generalizability across institutions, we wanted the classification of views to use ultrasound data and not metadata presented in the periphery. Because periphery information is predominantly static between frames, we tracked pixels that do not change intensity over frames and create a mask to remove such pixels. Because not all periphery information is completely static (i.e. ECG activity), we sampled multiple frames for a robust periphery removal algorithm.

Training data comprised 10 random frames from each manually labeled echo video. We convert all images to grayscale and resized the images to 224x224. We calculated the difference between each pair of adjacent images and store the 9 resulting “difference” images. We blur all 9 “difference” images by applying a Gaussian filter with sigma set to 5. To create our mask, we set each pixel to 1 most corresponding pixels in the 9 “difference” images are nonzero, and 0 otherwise. Applying this mask to each image in the video sets pixels corresponding to periphery information to 0. Finally, we zero-centered the data by subtracting each pixel by the mean pixel value of the training set.

We trained our network on approximately 40,000 pre-processed images. For stochastic optimization, we used the ADAM optimizer (11) with an initial learning rate of 1x10^-3 and mini-batch size of 64. For regularization, we applied a weight decay of 1x10^-5 and dropout with probability 0.5 on the fully connected layers. We ran our tests for 10-20 epochs or 10-20,000 iterations.

For final classification, we assigned the view of a video by taking the majority decision of predicted view labels on the 10 frames extracted from the video. Our results consistently showed a 1-2% increase in accuracy with this consensus approach over using 1 frame per video. Performance was assessed using cross-validation.

**Active Appearance Models for Image Segmentation**

Image segmentation is the process by which one delineates the boundaries of objects of interest within an image. It is often performed manually in the context of echo interpretation – but a computer can also be taught to identify these same structures. Automated image segmentation approaches typically rely on some initial estimate as to where the structure of interest is located. For example, facial recognition strategies take advantage of a finely tuned “face detector” that was trained on thousands of images and places a bounding box around one or more generic faces located in an image. For our specific task, we developed an effective fully automated bounding box determination for the cardiac chambers in the A2c, A3c, and A4c views. To accomplish this, we made use of the natural contrast in image intensity between the darker blood pool and echogenic structures - specifically the ventricular and atrial free walls, the interventricular and interatrial septa, and the mitral/tricuspid valvular apparatus (when closed). To avoid spurious peaks, we performed a Gaussian blur of the image and used a peak finding function in the Scientific Python (scipy) package to identify peaks in intensity in both the horizontal and vertical direction with constraints regarding the location and spacing of peaks.

Once a bounding box is established, several strategies can be used to automatically segment an image. Many of these algorithms were originally developed with the goal of recognizing specific faces and have since been applied to a broader group of problems, including medical image segmentation. The active appearance model (AAM) algorithm builds a statistical model of shape and gray-scale appearance that can be readily fit to new examples (12). We made this selection because of the availability of an excellent open-source library (Menpo) (13), prior use of AAM on echo images (14), and a need for only modest amounts of training data.

To train AAMs for the apical echo views, we selected videos from different echo studies and manually segmented 2-3 still frames representing different phases within the cardiac cycle. For A4c, this involves placing points along the inner LV border, the outer LV border, the left atrium (LA) endocardial border, and the right atrial endocardial border (done with ImageJ, requiring approximately 2 minutes per image). We manually segmented a total of 58 echo images for this task. Additional scripts then convert these points into standardized “landmarks”, such as regularly spaced points from the apex to the mitral annulus, or along the length of the LA.

The AAM algorithm then takes landmarks from multiple individual images and aligns them (the standardized positions are necessary for this step) and uses principal components analysis (PCA) to find a series of basis vectors that represent positional variation across the landmark points from image to image. Next, the actual images corresponding to these landmarks are used to extract the grayscale (i.e. texture) information surrounding each landmark. A second round of PCA is then performed to describe the texture variation from image to image in these areas. This is the “appearance model”.

When a new image is considered, the bounding box detector is applied to localize one or more chambers of
interest. The AAM algorithm then derives coefficients that can be applied to the basis vectors to minimize the discordance between the mean appearance model from the training data and the new image. The output of this approach is the location of the same stereotyped landmarks onto the new image.

**Classifying Successful and Unsuccessful Segmentation Using Gradient Boosting**

Automated segmentation can be challenging, especially with poor quality ultrasound images. In fact, manual segmentation can also be difficult in many cases. Although ultrasound image quality continues to improve with technological advances, our long-term goal was to be able to work with archived images dating as far back as 15-20 years and thus we needed an efficient way to identify and filter poorly segmented images. We reasoned that since each echo study typically had 8-12 videos depicting apical views, we would still be able to calculate longitudinal strain accurately on a subset of videos for each study – but it would be important to exclude those that would result in erroneous estimates.

We applied preprocessing, automated view identification, bounding box determination and automated image segmentation (locating the ventricle) to echo studies from a cohort of 420 patients with heart failure and preserved ejection fraction (15), collected at Northwestern University. The former set were all collected between 2006-2009 and thus were typically inferior in resolution to today’s images. We selected 1842 videos to judge segmentation of the left ventricle and found 67% of these were well segmented and would be useful for ventricle strain calculation (our pipeline produces images of the AAM output and thus manual assessment of the fit is rapid). Images with poor segmentation were predictable, and occurred in the setting of inaccurate boundary box localization, markedly off-axis ventricles, a prominent “false tendon”, portions of the ventricle that were cut-off, inaccurate view classification, or very poor acoustic windows. We also analyzed all the apical viewpoint images (108 in total) from a second set of 10 patients collected at Northwestern University which we had selected based on their high-quality longitudinal strain measurements (described below). The proportion of images that were well-segmented in this group was much higher, at 88%.

We used the well- and poorly-segmented videos as training data to derive a classifier for effective segmentation of the left ventricle. As described above, the output of AAM is the position of landmarks corresponding to the segmented structure of interest. In the case of the left ventricle, this consisted of 22 equally spaced points describing the location of the inner and outer boundaries of the heart. We used this output to extract a series of 30 quantitative features to characterize the segmentation – including the spacing between inner and outer landmarks at different positions (i.e. the wall thickness) and properties of the image inertial moments, which are simple functions of the weighted average of pixel intensities. These included the total ventricular length and width, the ratio of length to width, the position of the center of the ventricle, and the angle of the major and minor axes.

We used stochastic gradient boosting to develop a classifier (16). Gradient boosting is a machine learning algorithm which fits an additive expansion of decision trees to perform a classification or prediction task (i.e. supervised learning). We used the gbm (17) and caret (18) packages in R for this task, and developed a classifier with 80% and 77% accuracy for a4c and a2c, respectively, as judged by cross-validation.

**Automated Longitudinal Strain Measurements Using Speckle Tracking**

Longitudinal strain is a dimensionless measure of the longitudinal function of the heart (19). It is increasingly seen as a sensitive and accurate measure of cardiac function, and is now considered the standard for detecting abnormalities such as cardiotoxicity from chemotherapy (20). Strain is typically estimated by tracking the motion of echogenic particles or speckles from frame to frame. Although there are multiple commercial vendor packages available to compute strain, all require some amount of manual user intervention, thus drastically limiting scalability. Furthermore, because the vendor algorithm remains hidden from view, it is difficult to gauge which measurements may be problematic.

We opted to write our own algorithm for strain computation, adapting an approach previously described by Rappaport and colleagues (21). Using the results of our image segmentation, we split the left ventricle along its long axis, and output images focused on the endocardial border of the hemi-ventricle. For a given frame, we used the trackpy python package (22), a particle tracking software package, to locate speckles. The trackpy locate command allows the user to modify parameters involved in particle localization including particle diameter and minimum inter-particle separation. To track a given speckle from frame to frame, we selected a multipixel patch surrounding it and then located the best match for that patch in the next frame using the matchTemplate function in the OpenCV package (with the TM_CCOEFF_NORMED statistic). Importantly, we limited the search space to that region that could be attained based on the maximum predicted velocity of the corresponding myocardial segment (23).
and excluded matches that fell below a threshold level of agreement. We then computed the displacement (in pixels) of the patch and projected the displacement onto the long axis of the ventricular segment. We fit a cubic polynomial function to estimate the variation in frame-to-frame longitudinal displacement with position along the long axis and computed its first derivative to obtain the strain rate. We next performed median weighted smoothing and integrated the strain rate to obtain longitudinal strain. Finally, we implemented drift correction and performed further Fourier transformation of the resulting strain curve both to localize peaks and to take advantage of periodicity for additional signal smoothing.

We noted that images with very few successfully tracked speckles gave unstable estimates of longitudinal strain and thus we adaptively lowered the threshold level of agreement to include sufficient particles for function estimation for each frame. The median number of particles that passed the original filter, \( n \), was stored as a measure of quality for each video’s strain estimate. We noted that \( n \) tended to be lower for myocardial segments at the lateral image boundary and we thus primarily focused on strain estimates derived from the semi-ventricle region closer to the image midline.

**Comparison with Commercial Vendor Derived Strain Measurements**

Commercial vendor derived strain measurements were performed as previously described (24). Two sets of measurements were used for comparison. Set A comprises 240 echo studies of patients with HFP EF collected from 2006-2009; importantly these images were obtained without being optimized for evaluation of strain. We used the Tomtec package for post-hoc strain estimation, and computed an average value across one A4c and one A2c video. Set B was a more recent set comprising 10 echo studies from 2016, all of which were optimized for evaluation of strain. For each study, longitudinal strain for one A4c video and one A2c video was estimated using EchoPAC (GE Healthcare) software.

We examined concordance between automated strain estimates and those derived with the aid of vendor software using a variety of metrics. For Set A, we computed a weighted median across all A4c, A2c and A3c images, using \( \pi \) values (described above) as weights to obtain an estimate of global longitudinal strain. For Set B, we used the same video and same beat for comparison. Both sets of measurements were obtained blinded to each other. We computed Pearson and Spearman correlation coefficients and computed a mean average difference across comparisons. Variability in the correlation estimate was estimated using bootstrap confidence intervals (25).

**Analysis of serial echocardiograms from Trastuzumab-and Pertuzumab-treated patients**

Patients who received trastuzumab or pertuzumab for adjuvant or metastatic disease or received a screening echocardiogram between 2011 and 2015 were identified using the UCSF pharmacy and echocardiogram databases. Patients with a transthoracic echocardiogram at baseline, early in therapy (< 5 months, mean 3.0 months), and at 12 months were
included in the cohort (n = 152, mean age 54.3 years, all female).

Ejection fraction values were extracted from the echocardiogram reports. Patient demographics, co-morbidities, current medications, and oncological history were obtained from chart review.

We downloaded all available echos from each of these individuals and processed these through our entire pipeline. Plots of variation of longitudinal strain with time were generated using the ggplot2 package in RStudio. In addition to plotting strain values, we generated a smoothing spline curve using the smooth.spline function in R, with \( \pi \) values as weights.

III. RESULTS

Overview: An End-to-End Pipeline for Automated Function Interpretation

Our primary goal was to develop an analytic pipeline for assessment of cardiac function that required no user intervention and thus could be deployed on a high-performance computing cluster. We divided our approach into 5 steps (Figure 1). Such a modular approach allowed simultaneously advancing each part of the project – and readily lends itself to future improvements at one or more steps either through implementing more efficient algorithms (e.g. faster computation of strain) or trying alternate approaches (e.g. convolutional neural networks for image segmentation). It is also straightforward to envision how other clinical applications could be integrated – such as automation of routine measurements or detection of rare structural diseases.

Preprocessing entailed automated downloading of echo studies in DICOM format, separating videos from still images, extracting metadata (such as frame rate, heart rate), converting them into numerical arrays for matrix computations, and de-identifying images by overwriting patient health information. We next used convolutional neural networks (described below) for automatically determining echo views. After performing automated localization of cardiac chambers, we used active appearance models for image segmentation. We filtered away poorly segmented images using a classifier based on the gradient boosting algorithm. We next used the output of the image segmentation to delineate boundaries for computation of longitudinal strain. Finally, we deployed a particle tracking algorithm to compute strain. Below we describe several steps in greater detail.

Convolutional Neural Networks (“Deep Learning”) for View Identification

---

**Fig 2: Convolutional neural networks successfully discriminate echo views.** (Left) tSNE visualization of view classification. tSNE is an algorithm used to visualize high-dimensional data in lower dimensions. It depicts the successful grouping of test images corresponding to 6 different echocardiographic views. (Right) Confusion matrix demonstrating successful and unsuccessful view classifications within test data set. Numbers along the diagonal represent successful classifications while off-diagonal entries are misclassifications. Views are numbered as follows: 0) PLAX; 1) PSAX; 2) A2c; 3) A3c; 4) A4c; 5) IVC.
Typical echo studies consist of up to 50 separate videos representing up to a dozen different viewpoints. For example, several different views are taken with the transducer placed beside the sternum (e.g. parasternal long axis and short axis views), at the cardiac apex (apical views), or below the xiphoid process (subcostal views). Unfortunately, none of these views are labeled explicitly. Thus, the first learning step involves teaching the machine to recognize individual echo views. Since our primary interest was to compute longitudinal strain from the apical views, it was not necessary to distinguish all views. For training data, we manually labeled six different views: apical 2-, 3-, and 4-chamber (A2c, A3c, and A4c), parasternal long axis (PLAX), parasternal short axis at the level of the papillary muscles (PSAX), the inferior vena cava (IVC) and labeled all as “others”. We next used a multi-layer convolutional neural network, an algorithm commonly used for “deep learning”, to distinguish between the different views.

Deep learning is a form of machine learning devised to mimic the way the visual system works (9). The “deep” adjective refers to multiple layers of “neurons”, processing nodes tuned to recognize features within an image (or other complex input). The lower layers typically recognize simple features such as edges. The neurons in subsequent layers recognize combinations of simple features and thus each layer provides increasing levels of abstraction. The features in the top layer are typically used in a multiclass logistic regression model, which provides a final probabilistic output for classification.

We trained a 13–layer network and found an extremely high level of accuracy for view classification as judged by cross-validation (Figure 2, e.g. 99% for parasternal long axis). Clustering of the top layer features by t-Distributed Stochastic Neighbor Embedding (tSNE) (27), a useful algorithm for visualizing high-dimensional data, revealed clear separation of the different classes, with intuitive closer groupings of some pairs (e.g. A2c and A3c). Videos with apical classified views served as input for the next step.

Active Appearance Models for Image Segmentation

Image segmentation involves identifying the location of objects of interest within an image. For example, one could identify the faces of people in a surveillance camera video or the location of other automobiles on the road in front of a self-driving car. Given that image segmentation represents a critical component of computer vision and robotics, computer scientists have developed multiple different algorithms to carry out this task.

Prior to precisely locating a structure of interest, many of these algorithms require some broad guidance on where the structure may be found. This is because segmentation is ultimately an optimization problem, and optimization often fails or is unacceptably slow in settings where initial parameter values are too far from the optimal values (28). We derived a simple method to construct a bounding box around the left ventricle (and other structures such as the left atrium) by recognizing that there are characteristic variations in image intensities as one traverses from the right ventricular...
free wall to the right ventricular blood pool across the interventricular septum, left ventricle and left ventricular free wall. Variation is also seen in the vertical direction (Figure 3). We used a peak finding algorithm to localize these structures and were thus able to delineate accurate ventricular locations for >90% of images.

Next, we focused on image segmentation. We chose to use an approach called active appearance models (AAMs). AAMs build a template of the structure of interest, which reflects both structural variation within the training data and variation in image intensities. For example, an AAM model of the left ventricle could track the endocardial and epicardial boundaries, whose location and spacing vary throughout the cardiac cycle and whose echogenicity may vary from individual to individual and potentially from study to study. Our preference for AAMs was driven partly because we anticipated considerable shape similarities across different images and thus would be able to use relatively modest amounts of training data.

Training consisted of manually tracing individual structures of interest within echo images. We then converted these contours into regularly spaced landmarks. These landmarks and the accompanying image were then used to train the AAM model for both A4c and A2c images. We applied our model (preceded by bounding box determination) to a series of 2775 videos collected between 2006 and 2009 and to a second set of 108 videos collected in 2016 (Figure 3).

We inspected the AAM output for 1842 videos in the older data set and found 67% of these were well segmented (as judged by correct delineation of the endocardial boundary) whereas 88% of the more recent data set was well-segmented. Poor segmentation was predictable, and seen in circumstances of inaccurate view classification, inaccurate bounding box localization, markedly off-axis ventricles, a prominent band in the left ventricle, occlusion of portions of the ventricle or images with poor signal to noise.

Building a Classifier to Separate Well- and Poorly-Segmented Structures

The primary goal of the steps to this point was to generate input data to assist longitudinal strain calculation, specifically an approximate localization of the endocardial boundary. Since typical echo studies have 8-12 videos characterizing motion of the left ventricle from the apical viewpoint, we could compute longitudinal strain across all the well-segmented images from each study and derive a weighted median estimate across all of them. This multi-video summary would help deal with some of the difficulties that arises from beat-to-beat and video-to-video variation as well as buffer against some frames being more challenging to assess than others (discussed below). It was nonetheless important to eliminate poorly segmented images. Since manually assessing this step would be inefficient, we used supervised learning to derive a classifier to detect the subset of images from Set A that were poorly segmented.

![Fig. 4: Comparison between vendor and automated longitudinal strain measurements.](image)

(Left) Scatterplot comparing automated and vendor (EchoPAC - GE) estimates of absolute longitudinal strain on the exact same videos. (Right) Scatterplot comparing automated and vendor (Tomtec) estimates of global longitudinal strain averaging across multiple videos from each study. For this "patient-level" comparison, the exact same videos may not have been used for the strain estimate of the two methods. Size of points in both images represent the median number of speckles successfully tracked from frame to frame. The blue line represents a linear regression fit with 95% confidence interval for predicted values delineated by the shaded area.
Fig. 5: Illustrative longitudinal strain trajectories of patients undergoing trastuzumab chemotherapy. Automated strain values were computed for 9421 (apical) videos of 152 breast cancer patients undergoing serial echo monitoring during chemotherapy. Individual plots were generated for each patient: plots for two patients are shown, a 44 year old woman (left) receiving trastuzumab therapy with 4 cycles of doxorubicin and a 58 year old woman (right) receiving trastuzumab therapy only. Each cluster of colored dots represents an individual echo study, with individual points representing distinct videos used to estimate longitudinal strain. The size of each point reflects the quality of the estimate, as judged by the median number of successfully tracked speckles. A weighted smoothing spline was fit to the data. Ejection fractions in the published echo report are shown. Vertical blue dashed lines represent initiation and cessation of trastuzumab therapy; red dashed lines, for the patient on the left, represent doxorubicin infusions. A horizontal dashed line at longitudinal strain of 16% indicates a commonly used threshold for abnormal strain.

Longitudinal Strain Determination Using Particle Tracking

Our final step was to compute longitudinal strain for the left ventricle using the output of the AAMs. Longitudinal strain is an increasingly popular method to assess the longitudinal function of the heart (19). It is a sensitive measure of early evidence of cardiac dysfunction and is tolerant of errors in mapping of the endocardial border, in contrast with ejection fraction, which requires perfect delineation of this boundary.

Although commercial packages to measure strain have been available for many years, they invariably require some user intervention and thus cannot be implemented in a scalable, fully automated pipeline. Furthermore, the black-box nature of these packages has made it difficult to interpret how the measurement is made and what limitations there may be. Although strain is a robust measure of function, inter-vendor agreement in strain measurements is variable. For example, one recent study found correlation coefficients of 0.23 for strain measurements by software packages from GE vs. Toshiba, 0.42 for Philips vs Toshiba, and 0.72 for GE vs. Philips (19, 29, 30). The source of such variability is not clear, but probably reflects differences in image acquisition, processing, and algorithmic differences in how strain is computed, which makes it difficult to describe any of these as a “gold standard”.

We wrote our own algorithm for strain estimation, adapted from a previously published approach (16). We tracked echogenic particles from frame to frame to estimate velocities of particles across the length of the ventricle. Fitting this variation in particle velocity with position permitted estimates of myocardial velocity, strain rate, and strain. We compared our results to measurements based on commercial vendor packages (Figure 4), and found good agreement at the individual beat level ($\rho=0.77$, 95% bootstrap CI 0.58 – 0.92). Agreement at the overall patient level (i.e. where the actual videos analyzed may differ) was also good ($\rho=0.51$, 95% bootstrap CI 0.38 – 0.63). Spearman correlation coefficients were 0.77 and 0.50 respectively and the mean differences between methods on the two datasets were -2.6% (95% CI -1.46 – 0.93) and -0.25% (95% CI -0.32 – -0.18), respectively. These numbers are consistent with or superior to most published estimates of strain concordance across different algorithms.

Although we could readily compute strain estimate on all images in the former data set, for the latter data set, we could not register high quality strain measurements on 14% of images because of an insufficient number of confidently tracked particles. The median number of well tracked particles across a study (we refer to this as $\pi$) turned out to be a very useful quality metric as low $\pi$ values (<10) coincided with poor agreement in strain measurement between 1) different regions of the heart;
Mapping Patient Trajectories During Trastuzumab or Pertuzumab Treatment

As described in the introduction, the primary motivation of this work is to facilitate early, low-cost detection of cardiac dysfunction in asymptomatic individuals to motivate initiation or intensification of therapy. Such dysfunction can be a consequence of cardiac risk factors including diabetes mellitus and hypertension but can also arise from cardiotoxic agents used to treat cancer (31). In fact, many of these agents require serial cardiac monitoring and oncologists may opt to change treatment course depending on evidence of cardiac deterioration, as judged by changes in ejection fraction or longitudinal strain (20).

Given our ability to estimate longitudinal strain accurately, we hypothesized that we should be able to use our analytic pipeline to generate quantitative patient trajectories for breast cancer patients treated with cardiotoxic agents. We identified 152 patients treated with trastuzumab or pertuzumab, antibody inhibitors of the Her2 protein, which are known to cause cardiotoxicity in a subset of patients. We downloaded 1047 echo studies from these patients and processed these (9402 apical videos) through our pipeline. We generated automated plots of strain trajectories, overlaying chemotherapy usage and reported ejection fractions onto our visualization. We found interpretation was greatly assisted by characterizing each estimate by its value of $\pi$, and used a $\pi$-weighted spline fit of the data.

We observed a breadth of patient trajectories. Figure 5 reveals illustrative examples. The patient on the left, a 44 year old woman who had already been receiving trastuzumab therapy for Stage III ER+/HER2+ breast cancer, had a rapid drop in longitudinal strain after initiation of doxorubicin therapy. Although there is tight clustering of the nadir strain values (Figure 5, mustard-colored dots), there is only moderate concordance with the reported ejection fraction in the study. The patient on the right-hand side is a 58 year old breast cancer patient with Type 2 diabetes mellitus and hyperlipidemia who experienced cardiac dysfunction that improved after cessation of trastuzumab, although both the baseline and final strain values are at the lower limit of normal. Such plots (with accompanying statistics) could be generated by a cloud-based interpretation system that stores prior estimates, thus allowing depiction of longitudinal trends.

To further validate our approach, we also compared strain values in patients who did or did not receive doxorubicin-cyclophosphamide neo-adjuvant therapy prior to receiving trastuzumab/pertuzumab. Consistent with prior results (32), pretreatment with anthracyclines worsened cardiac function, as represented by lower median (19.7 vs 21.1%, $p = 0.01$) and nadir (16.2 vs. 17.8%, $p = 0.02$) absolute strain values (Figure 6).

IV. DISCUSSION

We achieved our primary objective, namely to construct an end-to-end automated pipeline for assessment of left ventricular function. This pipeline is fully scalable, as evidenced by our analysis of over 10,000 videos for this

---

**Fig. 6: Automated strain measurements confirm the more severe toxicity that occurs when combining trastuzumab/pertuzumab with anthracyclines.** Violin plots showing median (left) and nadir (right) longitudinal strain values for patients pretreated (red) or not pretreated (blue) with neo-adjuvant doxorubicin/cyclophosphamide prior to therapy with trastuzumab (and/or pertuzumab). For both statistics, the addition of the anthracycline doxorubicin resulted in more severe cardiac toxicity, as judged by significantly lower absolute strain values.
manuscript, all in a period of less than one month. Its modular nature provides multiple points for quality assessment and enables parallel improvement on multiple fronts. Although this represents the first example of a fully automated pipeline, it is important to acknowledge that many groups have made advances at each of these steps separately and insights from these efforts were invaluable for our work (14, 21, 33-35).

Even though we have developed a working product, there is room for improvement at nearly all steps. In some cases, we anticipate that more training data will be sufficient for improved performance. For example, most of the instances of poor segmentation arose in images that were not represented in our training examples, such as extreme off-axis views and prominent bands (“false tendons”) within the left ventricle. Although this currently limits the fraction of images with successful interpretation, our workflow in fact welcomes these “failures” as they can be manually segmented and added to our training data. In other cases, altogether different algorithms can be used for some of the tasks. For example, with enough training data, convolutional neural networks have been used for successful image segmentation (36), and may obviate our need for a two-step process involving a bounding box followed by active appearance models.

Developing our own strain algorithm was also instructive as it gave information on which factors resulted in variation in strain estimates. Most importantly, we found that in the setting of enough well-tracked particles, one can obtain results consistent with those derived using commercial packages, presumably because most of these packages are implementing a very similar algorithm. Nonetheless, we did notice that a significant proportion of images generated too few well-tracked particles for accurate strain estimation. Although we learned that one can always relax thresholds and obtain some estimate, it is likely to be inaccurate and inconsistent. Thresholding and/or weighting by this index (r) provided a valuable method to integrate data from many views and many studies. Along those lines, it should be possible to provide point-of-care user directives when data are collected to provide feedback on the likelihood of a good strain measurement from a given video.

We see this work as taking a step towards augmenting clinical practice rather than replacing current approaches. Specifically, we would like to see more measurements taken when patients are asymptomatic but at risk of cardiac dysfunction, with quantitative comparisons made to prior studies to obtain personalized longitudinal trajectories. Such an approach would shift evaluation to the primary care setting, with data collected by non-experts – and the resulting initiation and tailoring of care would hopefully reduce the alarming increase in heart failure incidence that has taken place in recent decades (37). A similar approach could be taken with point-of-care ultrasound at oncology infusion centers – both reducing the cost and increasing the timeliness of diagnoses of cardiotoxicity. In anticipation of such an eventuality, we deliberately avoided using any ECG information in our pipeline to accommodate analysis of data from low-cost portable handheld ultrasound devices.

Although the primary emphasis in this work was on measurement of ventricular dysfunction using longitudinal strain, we have established the groundwork for applying computer vision approaches to other aspects of echo interpretation (Figure 1). For example, accurate segmentation of the atria, ventricles, and aorta would assist in carrying out routine measurements commonly made in echo studies. Computer vision approaches could also be used for detection of myopathic disease such as hypertrophic cardiomyopathy or cardiac amyloidosis.

We have found that the combination of automated preprocessing and the ability to identify individual echo views using deep learning allows rapid accrual of training data for specific tasks, such as accumulating parasternal long axis videos for the detection of mitral valve prolapse. We are optimistic that this approach will have an impact by 1) extracting knowledge from the millions of archived echos available in echo laboratories and 2) enabling causal insights that require systematic longitudinal tracking of myocardial disease trajectories.

ACKNOWLEDGMENT

Work in the Deo laboratory is funded by NIH/NHLBI grant DP2 HL123228. Additional funding for a portion of echocardiographic data used for validation was provided by NIH/NHLBI grant R01 HL127028.

REFERENCES

1. J. A. Hill, E. N. Olson, Cardiac plasticity, The New England journal of medicine 358, 1370–1380 (2008).

2. T. Ishizu, Y. Seo, Y. Kameda, R. Kawamura, T. Kimura, N. Shimojo, D. Xu, N. Murakoshi, K. Aonuma, Left ventricular strain and transmural distribution of structural remodeling in hypertensive heart disease, Hypertension 63, 500–506 (2014).

3. A. N. Nesvokic, T. Edvardsen, M. Galderisi, M. Garbi, G. Gullace, R. Jurcut, H. Dalen, A. Hagendorff, P. Lancellotti, for the European Association of Cardiovascular Imaging Document Reviewers:, B. A. Popescu, R. Sicari, A. Stefanidis, Focus cardiac ultrasound: the European Association of Cardiovascular
Imaging viewpoint, *Eur Heart J Cardiovasc Imaging* **15**, 956–960 (2014).

4. B. W. Andrus, H. G. Welch, Medicare services provided by cardiologists in the United States: 1999-2008, *Circ Cardiovasc Qual Outcomes* **5**, 31–36 (2012).

5. R. Szeliński, *Computer Vision* (Springer Science & Business Media, 2010).

6. R. Bajcsy, S. Kovačič, Multiresolution elastic matching, *Computer Vision, Graphics, and Image Processing* **46**, 1–21 (1999).

7. V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, D. R. Webster, Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, *Jama* **316**, 2402–2410 (2016).

8. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* **542**, 115–118 (2017).

9. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* **521**, 436–444 (2015).

10. K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv 1409.1556 [cs.CV]* (2014).

11. D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization. *arXiv 1412.6980 [cs.LG]* (2014).

12. T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 681–685 (2001).

13. J. Alabort-i-Medina, S. Zafeiriou, A Unified Framework for Compositional Fitting of Active Appearance Models, *International Journal of Computer Vision*, 1–39 (2016).

14. J. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, J. H. C. Reiber, Automatic segmentation of echocardiographic sequences by active appearance motion models, *IEEE Trans. Med. Imaging* **21**, 1374–1383 (2002).

15. S. J. Shah, D. H. Katz, S. Selvaraj, M. A. Burke, C. W. Yancy, M. Gheorghide, R. O. Bonow, C.-C. Huang, R. C. Deo, Phenomapping for novel classification of heart failure with preserved ejection fraction, *Circulation* **131**, 269–279 (2015).

16. J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Statist.* **29**, 1189–1232 (2001).

17. G. Ridgeway, *CRAN - Package gbm* (R package version, 2006).

18. M. Kuhn, Caret package, *Journal of Statistical Software* (2008).

19. J. U. Voigt, G. Pedrizzetti, P. Lysyansky, T. H. Marwick, H. Houle, R. Baumann, S. Pedri, Y. Ito, Y. Abe, S. Metz, J. H. Song, J. Hamilton, P. P. Sengupta, T. J. Kolas, J. D’hooge, G. P. Aurigemma, J. D. Thomas, L. P. Badano, Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging, *Eur Heart J Cardiovasc Imaging* **16**, 1–11 (2015).

20. J. C. Plana, M. Galderisi, A. Barac, M. S. Ewer, B. Ky, M. Scherrer-Crosbie, J. Ganame, I. A. Sebag, D. A. Agler, L. P. Badano, J. Banchs, D. Cardinale, J. Carver, M. Cerqueira, J. M. DeCara, T. Edvardsen, S. D. Flamm, T. Force, B. P. Griffin, G. Jerusalem, J. E. Liu, A. Magalhaes, T. Marwick, L. Y. Sanchez, R. Sicari, H. R. Villarraga, P. Lancellotti, Expert consensus for multimodality imaging evaluation of adult patients during and after cancer therapy: a report from the American Society of Echocardiography and the European Association of Cardiovascular Imaging, *Eur Heart J Cardiovasc Imaging* **15**, 1063–1093 (2014).

21. D. Rappaport, D. Adam, P. Lysyansky, S. Riesner, Assessment of myocardial regional strain and strain rate by tissue tracking in B-mode echocardiograms, *Ultrasound in Medicine & Biology* **32**, 1181–1192 (2006).

22. D. B. Allan, T. A. Caswell, N. C. Keim, *Allan: Trackpy v0. 2 (doi: 10.5281/zenodo. 9971) - Google Scholar* (2014).

23. U. M. Wilkenshoff, A. Sovany, L. Wigström, B. Olstad, L. Lindström, J. Engvall, B. Janerot-Sjöberg, B. Wranne, L. Hatle, G. R. Sutherland, Regional mean systolic myocardial velocity estimation by real-time color Doppler myocardial imaging: a new technique for quantifying regional systolic function, *Journal of the American Society of Echocardiography* **11**, 683–692 (1998).

24. B. H. Freed, V. Daruwalla, J. Y. Cheng, F. G.
Aguilar, L. Beussink, A. Choi, D. A. Klein, D. Dixon, A. Baldridge, L. J. Rasmussen-Tovrik, K. Maganti, S. J. Shah, Prognostic Utility and Clinical Significance of Cardiac Mechanics in Heart Failure With Preserved Ejection Fraction, *Circ Cardiovasc Imaging* **9**, 1–10 (2016).

25. B. Efron, Bootstrap Confidence-Intervals for a Class of Parametric Problems, *Biometrika* **72**, 45–58 (1985).

26. H. Wickham, *ggplot2* (Springer Science & Business Media, New York, NY, 2009).

27. L. Maaten, G. Hinton, Visualizing data using t-SNE, *Journal of Machine Learning Research*, 2579–2605 (2008).

28. R. C. Deo, Machine Learning in Medicine, *Circulation* **132**, 1920–1930 (2015).

29. H. Yang, T. H. Marwick, N. Fukuda, H. Oe, M. Saito, J. D. Thomas, K. Negishi, Improvement in Strain Concordance between Two Major Vendors after the Strain Standardization Initiative, *Journal of the American Society of Echocardiography* **28**, 642–6e7 (2015).

30. Y. Nagata, M. Takeuchi, K. Mizukoshi, V. C.-C. Wu, F.-C. Lin, K. Negishi, S. Nakatani, Y. Otsuji, Intervendor variability of two-dimensional strain using vendor-specific and vendor-independent software, *Journal of the American Society of Echocardiography* **28**, 630–641 (2015).

31. J. J. Moslehi, Cardiovascular Toxic Effects of Targeted Cancer Therapies, *The New England journal of medicine* **375**, 1457–1467 (2016).

32. H. K. Narayan, B. Finkelman, B. French, T. Plappert, D. Hyman, A. M. Smith, K. B. Margulies, B. Ky, Detailed Echocardiographic Phenotyping in Breast Cancer, *Circulation* **135**, 1397–1412 (2017).

33. J. R. Hussan, P. J. Hunter, P. A. Gladding, N. Greenberg, R. Christie, A. Wu, H. Sorby, J. D. Thomas, ICMA: an integrated cardiac modeling and analysis platform, *Bioinformatics* **31**, 1331–1333 (2015).

34. S. Narula, K. Shameer, A. M. Salem Omar, J. T. Dudley, P. P. Sengupta, Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography, *Journal of the American College of Cardiology* **68**, 2287–2295 (2016).

35. X. Gao, W. Li, M. Loomes, W. Lianyi, A fused deep learning architecture for viewpoint classification of echocardiography, *Information Fusion* **36**, 103–113 (2017).

36. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **1706.07342**[cs.CV] (2015).

37. E. Braunwald, The war against heart failure: the Lancet lecture, *The Lancet* **385**, 812–824 (2015).