SUPPLEMENTAL MATERIAL

Supplemental Methods

Note 1: Preprocessing

Typical echocardiograms consist of a combination of 70-120 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 *pydicom* Python library to count the number of frames within each file thus enabling separation of still images from videos. We used the RSNA Clinical Trial Processor (https://www.rsna.org/ctp.aspx) for deidentification of videos, as this tool allows specification of location of identifying patient information for specific ultrasound manufacturer and model combinations. We next used the *gdmconv* utility from the Grassroots DICOM Library (GDCM) to convert compressed DICOM format videos into a raw DICOM format. This allowed use of the *pydicom* library for conversion of DICOM videos into numerical arrays. Numerical arrays were compressed for subsequent use. A subset of these were converted into Audio Video Interleaved (avi) format for manual segmentation.

Note 2: Architecture of the View Classification Model

The VGG network[1] takes a fixed-sized input of grayscale images with dimensions 224x224 pixels (we use scikit-image to resize by linear interpolation). Each image is passed through ten convolution layers, five max-pool layers, and three fully connected layers. (We experimented with a larger number of convolution layers but saw no improvement for our task). All convolutional layers consist of 3x3 filters with stride 1 and all max-pooling is applied over a 2x2 window with stride 2. The convolution layers consist of 5 groups of 2 convolution layers, which are each followed by 1 max pool layer. The stack of convolutions is followed by two fully connected layers, each with 4096 hidden units, and a final fully connected layer with 23 output units. The output is fed into a 23-way softmax layer to represent 23 different echocardiographic views. This final step represents a standard multinomial logistic regression with 23 mutually exclusive classes. The predictors in this model are the output nodes of the neural network. The view with the highest probability was selected as the predicted view.

Additionally, each echocardiogram 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. To address this issue, every image is randomly cropped between 0-20 pixels from each edge and resized to 224x224 during training. This provides variation in the periphery information, which guides the network to target more relevant features and improves the overall robustness of our view classification models.

Training data comprised of 10 random frames from each manually labeled echocardiographic video. We trained our network on approximately 70,000 pre-processed images. For stochastic optimization, we used the ADAM optimizer[2] with an initial learning rate of 1e-5 and mini-batch size of 64. For regularization, we applied a weight decay of 1e-8 on all network weights and

dropout with probability 0.5 on the fully connected layers. We ran our tests for 20 epochs or ~20,000 iterations, which takes ~3.5 hours on a Nvidia GTX 1080. Runtime per video was 600 ms on average.

Accuracy was assessed by 5-fold cross-validation at the individual image level. When deploying the model, we would average the prediction probabilities for 10 randomly selected images from each video.

An important and unaddressed quantity is the fraction of echocardiograms for which we could generate a usable view. One way to address this is to look at the probabilities for a view that we would expect in every study, such as A4c. We took all studies we had downloaded for which there was at least one measurement in the UCSF database (mass, volume or function) and had at least 40 videos (suggesting it was likely not a limited study). For each of these, we looked at the probability of the A4c class for all videos and took the maximum (Figure S2). Only 3% of these had a maximum probability of the A4c class that was less than 0.25 (a value that we found empirically was still interpretable for segmentation) and 5% had a maximum probability less than 0.40. If we include probabilities for both A4c and A4c with occlusion of the left atrium, these values drop to 1% and 2% respectively.

Note 3: Convolutional Neural Networks for Image Segmentation

To train image segmentation models, we derived a CNN based on the U-net architecture described by Ronneberger et al[3]. The U-net-based network we used accepts a 384x384 pixel fixed-sized image as input, and is composed of a contracting path and an expanding path with a total of 23 convolutional layers. The contracting path is composed of twelve convolutional layers with 3x3 filters followed by a rectified linear unit and four max pool layers each using a 2x2 window with stride 2 for down-sampling. The expanding path is composed of ten convolutional layers with 3x3 filters followed by a rectified linear unit, and four 2x2 up-convolution layers. Every up-convolution in the expansion path is concatenated with a feature map from the contracting path with same dimension. This is performed to recover the loss of pixel and feature locality due to downsampling images, which in turn enables pixel-level classification. The final layer uses a 1x1 convolution to map each feature vector to the output classes.

Separate U-net CNN networks were trained to perform segmentation on images from PLAX, PSAX (at the level of the papillary muscle), A4c, A3c, and A2c views. Training data was derived for each class of echocardiographic view via manual segmentation. We performed data augmentation techniques including cropping and blacking out random areas of the echocardiographic image in order to improve model performance in the setting of a limited amount of training data. The rationale is that models that are robust to such variation are likely to generalize better to unseen data. Training data underwent varying degrees of cropping (or no cropping) at random amounts for each edge of the image. Similarly, circular areas of random size set at random locations in the echocardiographic image were set to 0-pixel intensity to achieve ``blackout''. This U-net architecture and the data augmentation techniques enabled highly efficient training, achieving accurate segmentation from a relatively low number of training examples. Finally, in addition to pixelwise cross-entropy loss, we included a distance-based loss penalty for

misclassified pixels. The loss function was based on the distance from the closest pixel with the same misclassified class in the ground truth image. This helped mitigate erroneous pixel predictions across the images.

We used an Intersection Over Union (IoU) metric for assessment of results. The IoU takes the number of pixels which overlap between the ground truth and automated segmentation (for a given class, such as left atrial blood pool) and divides them by the total number of pixels assigned to that class by either method. It ranges between 0 and 100.

Note 4: Derivation of measures of cardiac structure and function

We used the output of the CNN-based segmentation to compute chamber dimensions and ejection fraction. A typical echocardiographer typically filters through many videos to choose specific frames for measurement. They also rely on the electrocardiogram (ECG) tracing to phase the study and thus choose end-systole and end-diastole. Since our goal is to enable use of handheld echocardiographic devices without ECG capabilities, we needed to rely on segmentation (i.e. variation in ventricular area) to identify end-systole and end-diastole. Since there are likely to be chance errors in any CNN model, we emphasized averaging as many cardiac cycles as possible, both within one video and across videos. For each study, we used all videos that included the chamber of interest, provided that it was not occluded, relying upon our view classification model to identify those videos.

LVEDV,LVESV,LVEF: We first used the time interval between frames and the patient heart rate to estimate the duration of the cardiac cycle (we thus restricted our analyses to echocardiograms for which heart rate and frame rate were available within the metadata). We then moved a sliding window across the video with a window length of 90% of a cardiac cycle (thus avoiding seeing end-systole or end-diastole more than once). Within a window, we selected the 90% and 10% percentile of the left ventricular volumes to serve as LV end-diastolic area and end-systolic areas, respectively. We derived LVEDV and LVESV using the area-length formula. We also used these to compute an EF for that cycle. To enable making multiple measurements per study, we moved a sliding window across the video with a step size of half of a cardiac cycle. We selected two additional percentile values for each metric: one percentile applied to measurements from multiple cycles within one video, and a second across all videos in a study. We selected the first percentile based on intuition regarding how the typical echocardiographer scans through images to select one for manual segmentation. We also avoided minimum and maximum values to exclude outliers from poor quality segmentation. We selected the second percentile to minimize bias between measured and automated values, although in most cases there was relatively little difference with choice of threshold and we used the median as default. For the first cutoff (i.e. multiple measurements from one video), we used 90% percentile for LVEDV and 50% percentile values (i.e. the median) for LVESV and LVEF. For the second cutoff (across multiple videos in a study), we selected median values for LVEDV, 25th percentile for LVESV, and 75th percentile for LVEF.

LAVOL: For LAVOL, we took a similar approach, again taking the 90% percentile of the LA area for each window. If there were multiple LAVOL measurements from one video we took the median value, and if there were multiple videos per study, we took the 25th percentile of these values.

We found that erroneous LAVOL values would arise from videos with an occluded LA. Although our view classification CNN was trained to discriminate these, some videos slipped through. We thus imposed an additional heuristic of excluding measurements from videos where LAVOL/LVEDV was less than 30%, as we found empirically that fewer than 5% of non-occluded studies had a ratio this extreme.

LV mass: For LV mass we again took a sliding window approach, using the 90% percentile value for the LV outer (myocardial) area and computed LV mass using the Area-Length formula[4]. If there were multiple LV mass measurements from one video we took the median value, and if there were multiple videos per study, we took the 25th percentile of these values.

Note 5: Automated Longitudinal Strain Measurements Using Speckle Tracking
We opted to write our own algorithm for strain computation, adapting an approach previously described by Rappaport and colleagues[5]. 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, a particle tracking software package, to locate speckles. The *trackpy* locate function 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 TMCCOEFFNORMED 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[6] and excluded matches that fell below a threshold level of agreement (0.85). 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 smoothing and integrated the strain rate to obtain longitudinal strain. Analysis was performed within windows corresponding to one cardiac cycle (as estimated from the heart rate and frame rate). For each window, we selected the frame with the lowest (most negative) strain value across all segments to compute the global longitudinal strain, integrating both the medial and lateral portions of the ventricle. The use of a sliding window allowed multiple strain estimates per video.

We also computed average longitudinal strain, deriving the minimum strain value across 25-30 positions along the length of the left or right ventricle, taken separately, and then computing a median across all positions.

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 was stored as a measure of quality for each video.

Estimation of strain typically required 1-4 minutes per video, depending on the image size and

the number of frames.

Note 6: CNNs to detect disease.

Just as with view classification, we used a 13 layer VGG model[1]. Our method and architecture were identical to our view classification network described in Note 2, but with a final fully connected layer of 2 output units. This final layer is fed into a 2-class softmax layer to represent probabilities for disease vs. control. For stochastic optimization, we used the ADAM optimizer[2] with an initial learning rate of 1e-5 and mini-batch size of 64. For regularization, we applied a weight decay of 1e-8 and dropout with probability 0.5 on the fully connected layers. We ran our tests on 20000 training images for 20 epochs, which took two hours to run on a Nvidia GTX 1080. Run-time performance was approximately 600ms per video.

Supplemental Tables

Table S1. Characteristics of Echocardiograms Used in this Manuscript

Year		
	2017	5471
	2016	2865
	2015	919
	2014	2088
	2013	607
	2012	634
	2011	532
	2010 and before	918
	Total	**14035**
Manufacturer (%)		
	Philips Medical Systems ie33	44
	Acuson Sequoia	14
	Philips Medical Systems EPIQ 7C	14
	GE Vingmed Ultrasound Vivid E9	11
	Philips Medical Systems HD15	11
	GE Vingmed Ultrasound Vivid E95	4
	GE Vingmed Ultrasound Vivid i	2
	Other	<1
Patients		
	Age (years)	57±17
	Sex (% Female)	55

Table S2: Characteristics of Studies Used to Train View Classification Model

Year		
	2017	76
	2016	66
	2015	15
	2014	35
	2013	38
	2012	19
	2011	10
	2010 and before	18
	Total	**277**
Manufacturer (%)		
	Philips Medical Systems ie33	36
	Acuson Sequoia	36
	GE Vingmed Ultrasound Vivid E9	11
	Philips Medical Systems HD15	9
	Philips Medical Systems EPIQ 7C	4
	Other	4
Patients		
	Age (years)	57±16
	Sex (% Female)	68

Table S3: Characteristics of Studies Used to Train Segmentation Models for Individual Views

View			
A2c			
	Number of images		214
	Year of study		
		2017	2
		2016	20
		2015	18
		2014	43
		2013	61
		2012	29
		2011	15
		2010 and before	26
	Patients		
		Age (years)	62±12
		Sex (% Female)	77
A3c			
	Number of images		141
	Year of study		
		2017	141
	Patients		
		Age	61±15
		Sex (% Female)	52
A4c			
	Number of images		182
	Year of study		
		2016	26
		2015	9
		2014	26
		2013	53
		2012	19
		2011	19
		2010 and before	30
	Patients		
		Age (years)	59±14
		Sex (% Female)	65
PSAX			
	Number of images		124
	Year of study		
		2017	6
		2016	20
		2015	6
		2014	9
		2013	36
		2012	12
		2011	13
		2010 and before	22
	Patients		
		Age (years)	59±16
		Sex (% Female)	52
PLAX			
	Number of images		130
	Year of study		
		2017	3
		2016	50
		2015	18
		2014	24
		2013	8
		2012	8
		2011	8
		2010 and before	9
	Patients		
		Age (years)	61±15
		Sex (% Female)	85

Table S4: Characteristics of Echocardiograms Used to Validate Measurements of Structure and Function

Year		
	2017	4797
	2016	2085
	2015	335
	2014	1352
	2013	54
	2012	37
	2011	5
	2010 and before	1
	Total	**8666**
Manufacturer (%)		
	Philips Medical Systems ie33	49
	GE Vingmed Ultrasound Vivid E9	11
	GE Vingmed Ultrasound Vivid E95	5
	Philips Medical Systems HD15	13
	Philips Medical Systems EPIQ 7C	20
	GE Vingmed Ultrasound Vivid i	2
	Other	<1
Patients		
	Age (years)	58±18
	Sex (% Female)	50

Table S5: Characteristics of Echocardiograms Used to Train HCM Classification Model

		Cases	Controls	p-value
Year - Number of studies (%)				
	2017	62 (13)	364 (16)	
	2016	81 (16)	414 (18)	
	2015	59 (12)	240 (11)	
	2014	42 (8)	210 (9)	0.21
	2013	55 (11)	189 (8)	
	2012	47 (9)	209 (9)	
	2011	52 (11)	223 (10)	
	2010 and before	97 (20)	399 (18)	
Manufacturer (%)				
	Philips Medical Systems iE33	28	27	
	Acuson Sequoia	33	32	
	GE Vingmed Ultrasound Vivid E9	16	16	0.84
	Philips Medical Systems HD15	12	14	
	Philips Medical Systems EPIQ 7C	5	6	
	GE Vingmed Ultrasound Vivid i	3	2	
	GE Vingmed Ultrasound Vivid E95	2	2	
	Other	0	1	
Patients				
	N (unique patients/studies)	260/495	2064/2244	
	Age (years)	58±14	57±15	0.63
	Sex (% Female)	44	47	
	Genotype Positive (%)	18	0	

Table S6: Characteristics of Echocardiograms Used to Train Amyloid Classification Model

		Cases	Controls	p-value
Year - Number of studies (%)				
	2017	21 (12)	121 (15)	
	2016	29 (16)	147 (18)	
	2015	37 (21)	144 (18)	
	2014	17 (9)	86 (7)	0.89
	2013	14 (8)	57 (8)	
	2012	14 (8)	63 (7)	
	2011	14 (8)	55 (7)	
	2010 and before	33 (18)	132 (16)	
Manufacturer (%)				
	Philips Medical Systems iE33	52	52	
	Acuson Sequoia	32	30	
	GE Vingmed Ultrasound Vivid E9	5	6	0.99
	Philips Medical Systems HD15	3	3	
	Philips Medical Systems EPIQ 7C	7	8	
	GE Vingmed Ultrasound Vivid i	1	1	
Patients				
	N (unique patients/studies)	81/179	771/804	
	Age (years)	65±11	66±11	0.36
	Sex (% Female)	22	23	

Table S7: Characteristics of Echocardiograms Used to Train PAH Classification Model

	Cases	Controls	
Year - Number of studies (%)			
2017	117 (20)	564 (23)	
2016	108 (18)	503 (20)	
2015	85 (15)	285 (11)	
2014	64 (11)	297 (12)	0.20
2013	54 (9)	190 (8)	
2012	54 (9)	194 (8)	
2011	33 (6)	135 (5)	
2010 and before	69 (12)	323 (13)	
Manufacturer (%)			
Philips Medical Systems iE33	50	51	
Acuson Sequoia	24	25	
GE Vingmed Ultrasound Vivid E9	10	9	0.65
Philips Medical Systems HD15	6	5	
Philips Medical Systems EPIQ 7C	8	8	
GE Vingmed Ultrasound Vivid i	2	1	
Other	0	1	
Patients			
N (unique patients/studies)	104/584	2180/2487	
Age (years)	51±12	51±13	0.62
Sex (% Female)	82	80	

Table S8: Characteristics of Echocardiograms Used for Cardiotoxicity Study of Chemotherapy

Year		
	2017	18
	2016	53
	2015	119
	2014	220
	2013	173
	2012	137
	2011	83
	2010 and before	102
	Total	**890**
Manufacturer (%)		
	Philips Medical Systems ie33	49
	GE Vingmed Ultrasound Vivid E9	9
	GE Vingmed Ultrasound Vivid E95	<1
	Philips Medical Systems HD15	4
	Philips Medical Systems EPIQ 7C	1
	Acuson Sequoia	37
Patients		
	Age (years)	55±11
	Sex (% Female)	100

Table S9: Numbers of Echocardiogram Videos Labeled to Train View Classification Model

View		Number of videos labeled
Parasternal		
	Long axis – remote	91
	Long axis	456
	Long axis – zoom of left atrium	88
	Long axis – centered over left atrium	78
	RV inflow	95
	Short axis at apex	21
	Short axis at papillary muscle	458
	Short axis at mitral valve	114
	Short axis at aortic valve	263
	Short axis at aortic valve - zoom	106
Apical		
	Apical 2-chamber – no occlusions	465
	Apical 2-chamber – occluded left atrium	266
	Apical 2-chamber – occluded left ventricle	30
	Apical 3-chamber – no occlusions	235
	Apical 3-chamber – occluded left atrium	103
	Apical 3-chamber – occluded left ventricle	25
	Apical 4-chamber – no occlusions	770
	Apical 4-chamber – occluded left atrium	314
	Apical 4-chamber – occluded left ventricle	57
	Apical 5-chamber	191
Subcostal	All subcostal views	422
Suprasternal	All suprasternal views	85
Other	Views with Doppler, IV contrast, could not classify	2530
Total		**7168**

Table S10: Internal measures of consistency for echocardiographic structure and function measurements. Spearman correlation coefficients are listed along with a p-value for a null hypothesis significance test.

Comparison	**N**	**Correlation – Manual vs. Manual (p-value)**	**Correlation – Automated vs. Automated (p-value)**
Left atrial volume vs. left ventricular mass	4012	0.54 (<2e-16)	0.56 (<2e-16)
Left ventricular mass vs. left ventricular diastolic volume	5874	0.62 (<2e-16)	0.61 (<2e-16)
Left ventricular mass vs. left ventricular systolic volume	5856	0.58 (<2e-16)	0.55 (<2e-16)
Left atrial volume vs. left ventricular diastolic volume	4748	0.48 (<2e-16)	0.56 (<2e-16)
Left atrial volume vs. left ventricular systolic volume	4738	0.46 (<2e-16)	0.49 (<2e-16)
Left atrial volume vs. left ejection fraction	4720	-0.22 (<2e-16)	-0.23 (<2e-16)
Left ventricular mass vs. global longitudinal strain	243	-0.16 (0.01)	-0.27 (<2e-16)
Left ventricular mass vs. left ejection fraction	5123	-0.28 (<2e-16)	-0.28 (<2e-16)
Left ventricular diastolic volume vs. global longitudinal strain	326	-0.15 (0.006)	-0.17 (0.002)
Left ventricular systolic volume vs. global longitudinal strain	326	-0.29 (<2e-16)	-0.27 (<2e-16)
Left ventricular ejection fraction vs. global longitudinal strain	251	0.37 (<2e-16)	0.32 (<2e-16)

Supplemental Figures

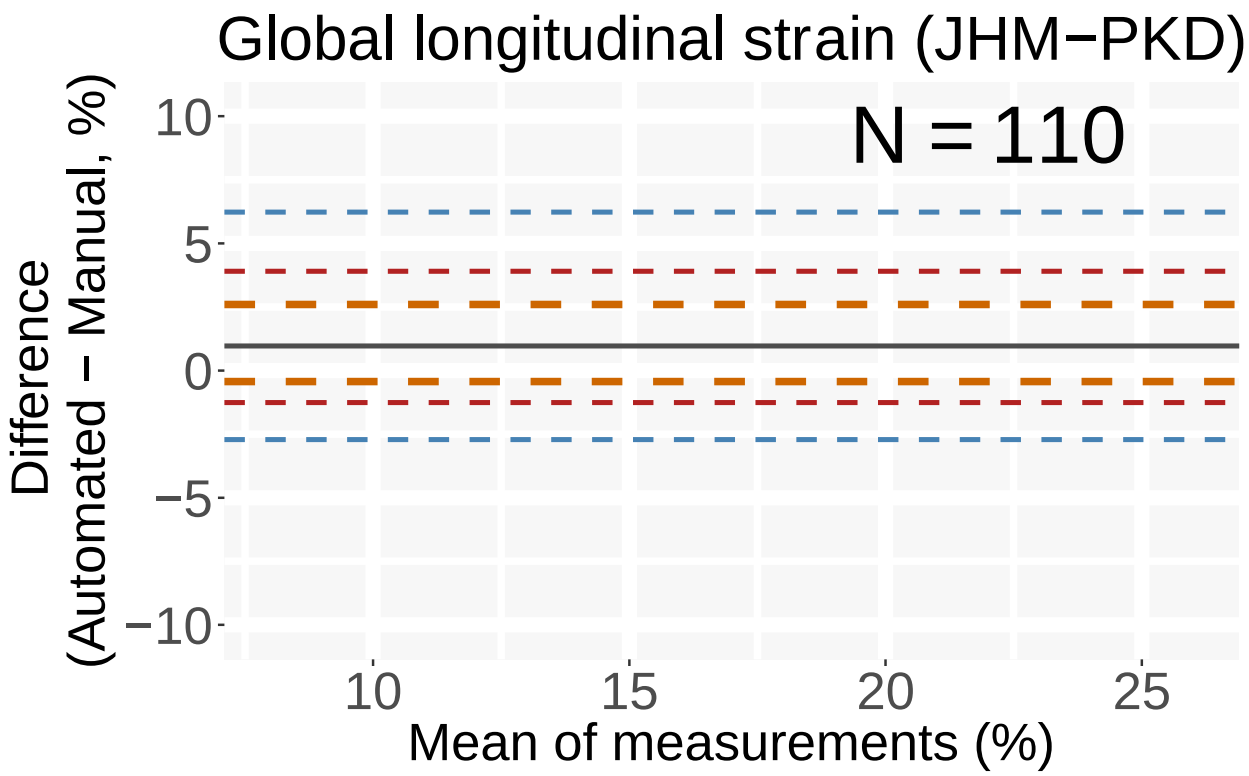

Figure S1: Bland-Altman plot for 110 studies from a polycystic kidney disease cohort (PKD).

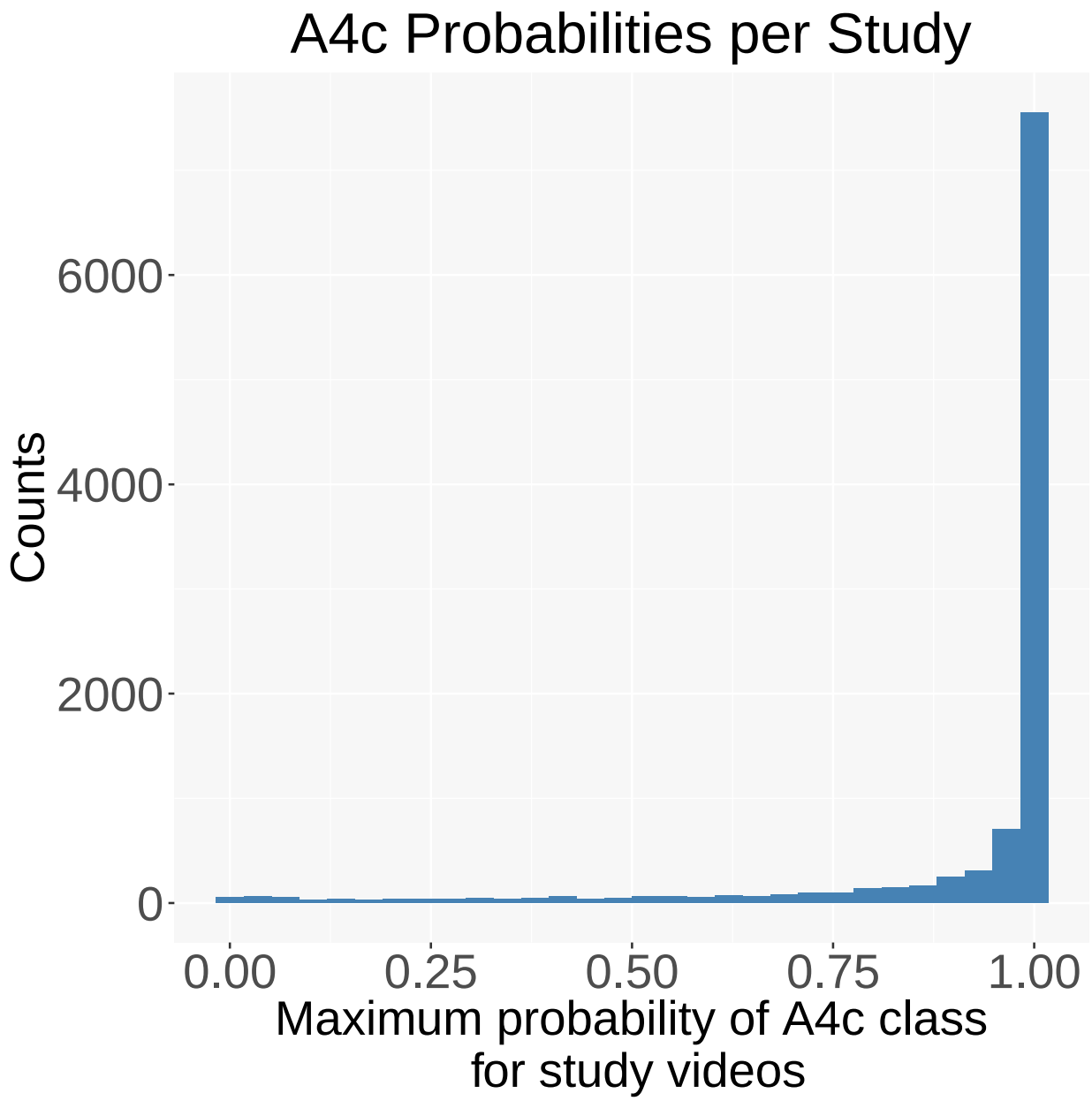

Figure S2: Frequency distribution of maximum probabilities of A4c view class for 10524 studies.

1. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv. 2014;1409.1556v6 [cs.CV], 10 Apr 2015. URL: https://arxiv.org/abs/1409.1556.

2. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv. 2014;1412.6980v9 [cs.LG], 30 Jan 2017. URL: https://arxiv.org/abs/1412.6980.

3. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv. 2015;1505.04597v1 [cs.CV], 18 May 2015. URL: https://arxiv.org/abs/1505.04597.

4. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, Lancellotti P, Muraru D, Picard MH, Rietzschel ER, Rudski L, Spencer KT, Tsang W, Voigt J-U. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr*. 2015;28:1–39.e14.

5. Rappaport D, Adam D, Lysyansky P, Riesner S. Assessment of myocardial regional strain and strain rate by tissue tracking in B-mode echocardiograms. *Ultrasound Med Biol*. 2006;32:1181–1192.

6. Wilkenshoff UM, Sovany A, Wigström L, Olstad B, Lindström L, Engvall J, Janerot-Sjöberg B, Wranne B, Hatle L, Sutherland GR. Regional mean systolic myocardial velocity estimation by real-time color Doppler myocardial imaging: a new technique for quantifying regional systolic function. *J Am Soc Echocardiogr*. 1998;11:683–692.