Supplemental information:

Supplemental Table 1 and Supplemental Table 2: View label agreement

Kappa Statistics				
Statistic	**Estimate**	**Standard Error**	**95% Confidence Limits**	
Simple Kappa	0.8262	0.0094	0.8077	0.8447
Weighted Kappa	0.8987	0.0065	0.8859	0.9115

Table										
E(E)	**L(L)**									
Frequency	**A2C**	**A4C**	**CW AoV**	**No Label**	**PLAX**	**PSAX AoV**	**PW AoV**	**TEE**	**Vascular**	**Total**
A2C	117	1	0	8	0	0	0	0	0	126
A4C	2	208	0	23	0	0	0	0	0	233
CW AoV	0	0	109	124	0	0	0	0	0	233
No Label	9	29	26	1399	44	1	5	0	0	1513
PLAX	0	0	0	12	220	0	0	0	0	232
PSAX AoV	0	0	0	10	0	88	0	0	0	98
PW AoV	0	0	0	18	0	0	78	0	0	96
TEE	0	0	0	0	0	0	0	158	0	158
Vascular	0	0	0	0	0	0	0	0	60	60
Total	128	238	135	1594	264	89	83	158	60	2749

Supplemental Table 1 and Supplemental Table 2. View label agreement 50 studies were selected as a test set. These studies were heldout and each labeled by two sonographers. View labels representing A2C (apical 2 chamber), A4C (apical 4 chamber), CW AoV (continuous wave Doppler across the aortic valve), PW AoV (pulse wave Doppler across the aortic valve), PLAX (parasternal long axis), PSAX (parasternal short axis of the aortic valve), TEE (transesophageal images), and vascular (non-cardiac images) were labeled. Kappa for agreement was calculated.

Supplemental Table 3: View Label Cohorts across Train/Test Splits

	Number of Labeled Images						Unlabeled Images
	Total	PLAX	PSAX	A4C	A2C	Other-or-A4C-or-A2C	**Total**
Train	10253	2879	1008	1380	1050	3936	16112
Valid	3505	965	357	399	309	1473	5150
Test	3511	964	359	426	311	1451	5332
View-only Train	7694	2564	1005	2359	1766	0	37576
Unlabeled Train	353500	-	-	-	-	-	-

Supplemental Table 3: View Label Cohorts. Dataset Counts for 5-way view classification task. Arranged by Image number according to view label, reporting the mean over 3 splits. No split deviates more than 9% from the mean.

Supplemental Table 4: Internal validation of view classifier on TMED-2 test sets

Development Set Size number studies (train/valid)	Split1 balanced accuracy (2.5, 97.5th)	Split2 balanced accuracy (2.5, 97.5th)	Split3 balanced accuracy (2.5, 97.5th)	**Average balanced accuracy (2.5, 97.5th)**
479 (360 / 119)	96.72 (95.84, 97.53)	97.23 (96.41, 98.00)	97.14 (96.61, 98.09)	**97.03** **(95.85, 97.54)**
165	95.47 (94.47, 96.41)	95.46 (94.44, 96.43)	96.38 (95.49, 97.24)	**95.77** **(94.46, 96.44)**
56	88.99 (87.55, 90.34)	89.99 (88.47, 91.42)	92.03 (90.72, 93.28)	**90.34** **(87.52, 90.35)**

Supplemental Table 4: Internal validation of view classifier on TMED-2 test sets. The view classifier is trained to produce a 5-way probabilistic view classification (PLAX, PSAX AoV, A4C, A2C, or Other) given a single image. We report balanced accuracy across 3 training/test splits of the TMED-2 dataset with a 95% bootstrap CI in parentheses.

Supplemental Table 5: External validation of view classifier on Stanford EchoNet

Development Set Size number studies (train/valid)	Split1 A4C accuracy (2.5, 97.5th)	Split2 A4C accuracy (2.5, 97.5th)	Split3 A4C accuracy (2.5, 97.5th)	**Average A4C accuracy (2.5, 97.5th)**
479 (360 / 119)	95.02 (94.61, 95.45)	92.99 (92.31, 93.33)	92.21 (92.11, 93.17)	**93.41 (93.21, 93.77)**
165	70.57 (69.69, 71.48)	90.02 (89.43, 90.60)	82.84 (82.11, 83.57)	**81.14 (80.72, 81.58)**
56	61.30 (60.33, 62.24)	55.29 (54.34, 56.28)	81.66 (80.90, 82.41)	**66.08 (65.55, 66.61)**

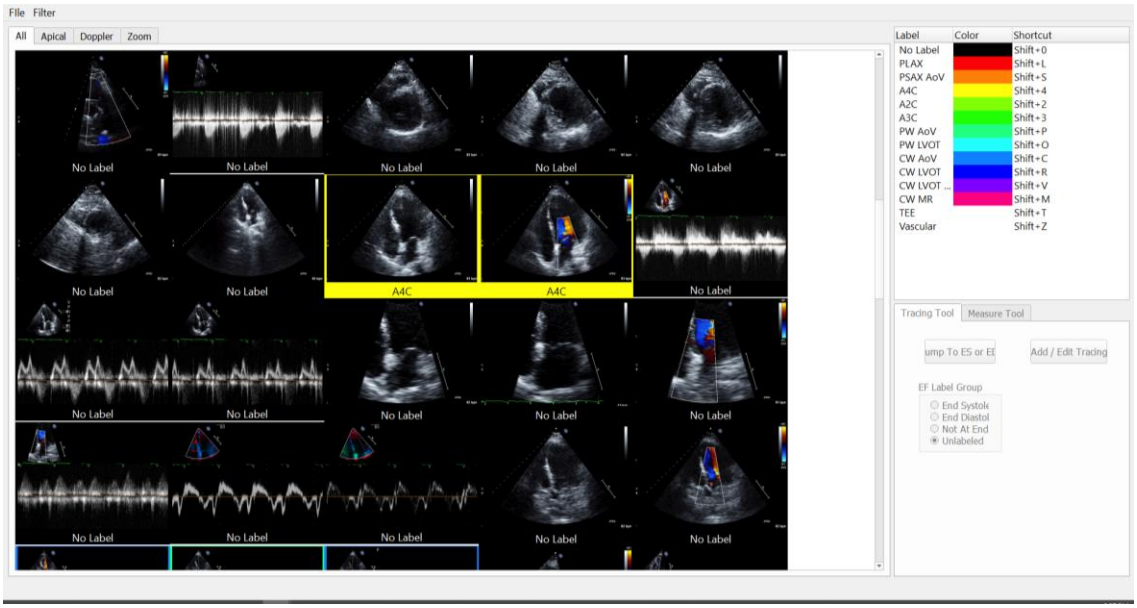
Supplemental Table 5: External validation of view classifier on Stanford EchoNet. We evaluate the TMED-2 trained classifier on all 10030 A4C view images in the Stanford EchoNet Dynamic dataset. We report accuracy (fraction of A4C views correctly identified) across 3 training splits of the TMED-2 training set, with a 95% bootstrap CI in parentheses.

Supplemental Table 6: AS diagnosis classifier balanced accuracy on TMED-2 test set

Dev Set Size number studies (train/valid)	Method	Split1 balanced accuracy (2.5, 97.5th)	Split2 balanced accuracy (2.5, 97.5th)	Split3 balanced accuracy (2.5, 97.5th)	**Average** balanced accuracy
479 (360/119)	Simple Average	35.16 (31.82, 38.97)	35.29 (33.33, 38.27)	34.18 (31.37,37.50)	34.88
	Prioritized View	74.63 (66.73, 82.24)	72.61 (66.05,79.24)	76.24 (69.62,82.87)	74.49

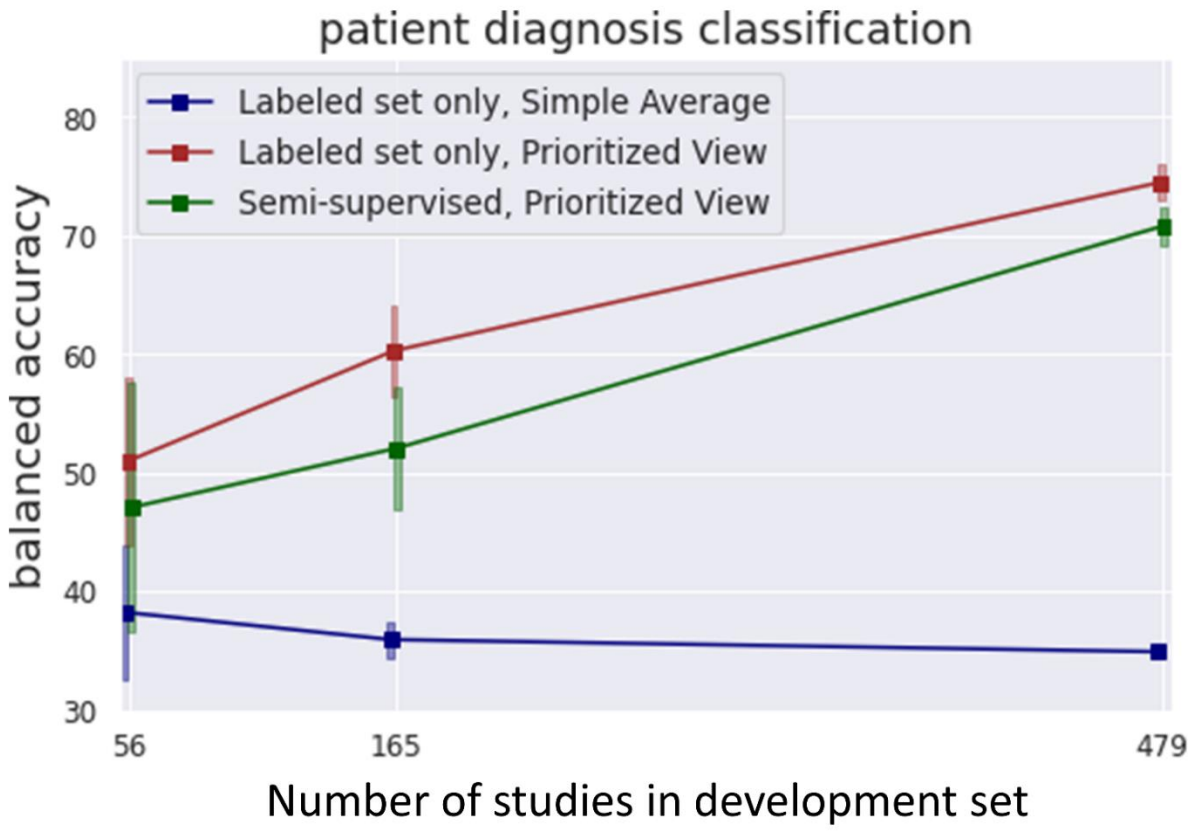
Supplemental Table 6: AS diagnosis classifier balanced accuracy on TMED-2 test set. We report balanced accuracy for this 3-class problem (no AS, early AS, significant AS), with 95% bootstrap CI in parentheses. Random chance would have 33.33% balanced accuracy.

Supplemental Figure 1: Image Labeling Tool Diagram



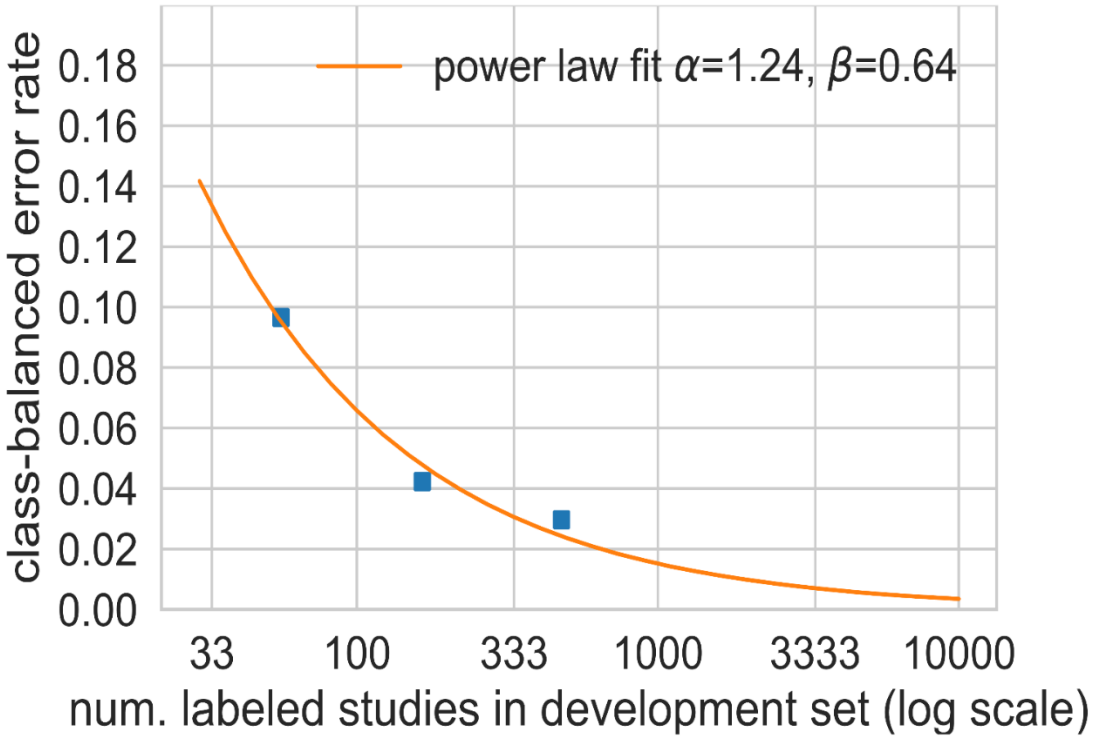
Supplemental Figure 1: Image labeling tool diagram. Labeling tool displays de-identified images in an array that allows rapid assignment of view labels using quick keys for the labels of interest.

Supplemental Figure 2. Diagnosis Classification Accuracy as a Function of Sample Size



Supplemental Figure 2. Diagnosis Classification Accuracy as a Function of Sample Size. We plot balanced accuracy across all studies in the TMED-2 test set (*y-axis*) versus the number of studies available for model development (training and validation, *x-axis*). Each line gives the performance of one prediction strategy for aggregating across all images in a study: Prioritized View and Simple Average. Square markers give the average over 3 splits, and the color bar represents standard deviation.

Supplemental Figure 3. Projections of View Classifier Error Rate vs Sample Size



Supplemental Figure 3. Projections of View Classifier Error Rate as Training Set Size Increases. As amount of available labeled data increases, using the power law scaling rule $\alpha n^{-\beta}$ fit to minimize squared error with the three measurements shown. With this prediction, 1000 labeled studies available for training and validation (double our current release's amount) would yield a balanced error rate of 1.5%, equivalent to balanced accuracy of 98.5%.