
Supplementary information

**Detecting structural heart disease from
electrocardiograms using AI**

In the format provided by the
authors and unedited

Supplemental Information Guide

Manuscript Title:

Detecting structural heart disease from electrocardiograms using AI

Authors:

Pierre Elias, MD et al.

Corresponding Author:

Pierre Elias, MD

622 W 168th St

PH20

New York, New York 10032

United States of America

Pae2115@cumc.columbia.edu

Table of Contents

Supplemental Methods.....	3
Model Design.....	3
Severe SHD.....	4
Internal Sensitivity Analyses	4
External Validation Cohorts	5
Supplemental Figures.....	6
Supplemental Figure 1: EchoNext Calibration for SHD in the NYP Multicenter Cohort.....	6
Supplemental Figure 2 Model results for Severe SHD.....	7
Supplemental Figure 3: Comparison of models trained with varied data splits by NYP Hospitals	8
Supplemental Figure 4: Correlation matrix with correlation coefficients between prediction scores of all component labels as well as the composite SHD label.....	9
Supplemental Tables.....	10
Supplemental Table 1 EchoNext Model Architecture and Parameters	10
Supplemental Table 2: Model Performance for all component labels	11
Supplemental Table 3: Severe SHD Prevalence in the NYP Multicenter Cohort	12
Supplemental Table 4. Cohort characteristics by NYP Campus (Combined, East, West)....	13
Supplemental Table 5: Patient characteristics for internal cross-validation cohorts	15
Supplemental Table 6: Comparison of results for models with variable training sets. All models evaluated on common test set comprising data from East and West campuses	17

Supplemental Table 7: Model performance in ECGs acquired within 0-3 months, 3-6 months, 6-9 months and 9-12 months prior to the echocardiogram	18
Supplemental Table 8: Model performance in various combinations of individual diseases	19
Supplemental Table 9: Characteristics for external validation cohorts	20
Supplemental Table 10: Performance metrics at fixed sensitivity of 70% in NYP multi-center cohort and external validation cohorts	21
Supplemental Table 11: Cardiologist Survey Results	22
Supplemental Table 12: Cardiologists survey results in ECGs with and without normal finding	23
Supplemental Table 13: Patient characteristics for the Columbia ECG Cohort	24
Supplemental Table 14: Component label performance of the Columbia Mini-Model in NYP Multicenter test set.....	25
Supplemental Table 15: Component label performance of the Columbia Mini-Model in the Columbia ECG test set.....	26
Supplemental Table 16: DISCOVERY clinical trial patients characteristics	27
Supplemental Table 17: Pilot clinical trial outcomes stratified by EchoNext and ValveNet prediction score categories.....	28
Supplemental Table 18: Training site characteristics	29

Supplemental Methods

Model Design

EchoNext is a convolutional neural network, based on design and approach previously described for ValveNet (Supplemental Table 1).³ A 1-Dimensional convolutional neural network with twelve channels (one for each ECG lead) applies the same filter with the same time width across all leads at once, increasing model interpretability and feature learning from well understood components such as QRS complexes. Residual network (ResNet) blocks using shortcut connections are employed to learn disparate features across many layers of the network. EchoNext fuses demographic and ECG-specific tabular data directly to the linear layer late in the model to maximize the value of these variables and allow the neural network to focus on learning unique discriminative features from the ECG not well captured by clinical variables such as age. ECG data were pre-processed to exclude all studies with missing age or sex information, all studies with poor quality traces, or measurements acquired during ventricular pacing. Missing atrial rate or PR intervals were imputed as 0; other ECG-derived measures were imputed with the median value. Additionally, ECGs were preprocessed for input into the model by undergoing a per-lead median signal filter, truncation of severe outliers beyond 0.1st and 99.9th percentile, and per-lead normalization based on means and standard deviation. The output for this model is a number from 0 to 1, with numbers closer to 1 indicating a higher model confidence that the disease is present.

The model was implemented using the Pytorch framework in Python and trained on V100 GPUs within the Azure Machine Learning environment. Data were split into train, validation, and test sets, as described in the main text. The model was optimized based on validation AUC. Following hyperparameter grid search experiments, the following parameters were retained in the final model: batch size: 16, filter size: 16, weight decay: 0.01, and dropout: 0.5. Additionally, we used the Adam Optimizer with learning rate: 5e-5.

Severe SHD

As a secondary analysis, we defined alternative thresholds for all structural heart disease component labels to represent a more severe phenotype. Briefly, these were LVEF $\leq$ 35%, maximum LV wall thickness $\geq$ 1.6cm, 'severe' regurgitation (any valve), 'severe' aortic stenosis, 'severe' RV systolic dysfunction, 'large' pericardial effusion, PASP $\geq$ 60mmHg, or TR maximum velocity $\geq$ 3.6m/s. The prevalence of each of these labels within the NYP Multicenter Cohort is provided in Supplemental Table 2. In particular, the overall prevalence of Severe SHD in the test set was 16.1%. The model trained against this severe phenotype yielded marginally higher AUROC and AUPRC (87.7% and 60.8%, respectively; Supplementary Figure 1) compared with the primary EchoNext results for moderate or severe SHD, as presented in main text.

Internal Sensitivity Analyses

In addition to the primary data partition, we performed a sensitivity analysis generating and comparing several distinct models based on variations in the training and validation sets. Specifically, we leveraged existing institutional internal data partitions to train three separate models: 'East' using data exclusively from Weill Cornell Medical Center, NYP-Brooklyn Methodist Hospital, NYP-Queens Hospital, and NYP-Lower Manhattan Hospital; 'West' using

data exclusively from Columbia University Irving Medical Center, NYP-Westchester Hospital, NYP-CHONY, and NYP-Allen Hospital; and 'Blend' using a subset of the combined data from both East and West (Supplemental Table 3). A single test set was created with data from East and West to compare the various models. It is worth emphasizing that approximately half of this blended test set represents "external" data for the models trained on East or West data alone, offering an opportunity to assess generalizability compared with the model trained with Blended data. Details of the different data partitions used for training are provided in Supplemental Table 4; the 'Blend' set was sub-set from the available data to approximate training sample size.

Model performance results are shown in Supplemental Table 5 and Supplemental Figure 3. Overall, differences in performance were minimal across models, with differences in AUROC and AUPRC of 0.2% and 0.1%, respectively between West and Blended models. These findings support excellent external generalizability of models trained on these data for this task, as the performance drop with or without external test data was minimal. To ensure that 12 months was an appropriate maximum period between ECG (feature) and echo (label) pairing we looked at performance by 3 month blocks as shown in Supplemental Table 7. Time between ECG and echo of 6-9 months had the highest AUROC (85.0%) and 3-6 months had the lowest AUROC (82.5%).

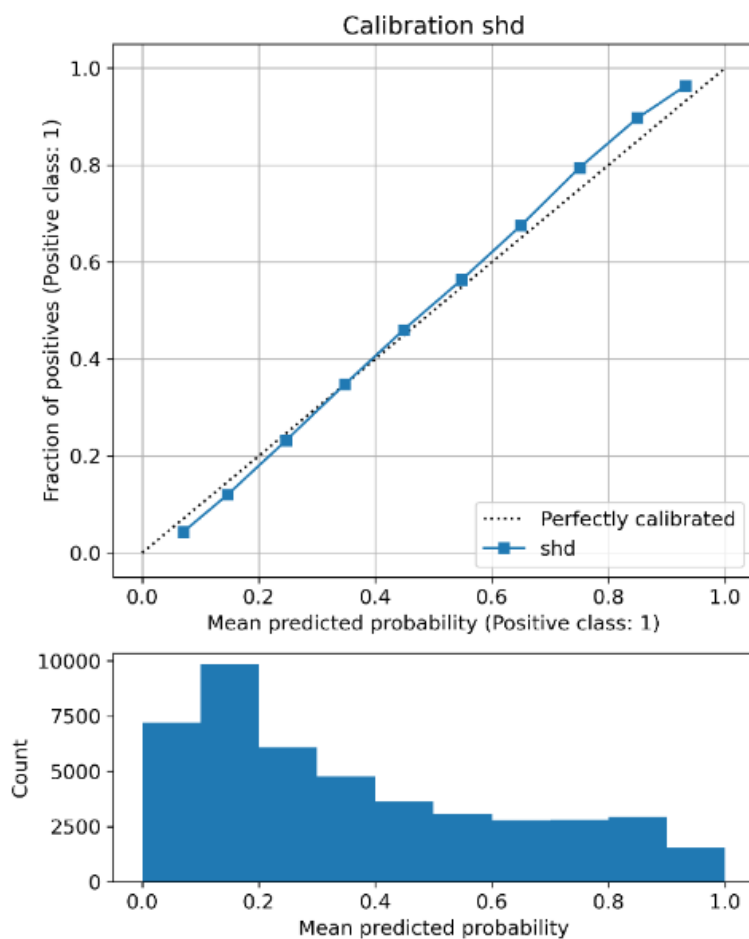
External Validation Cohorts

To evaluate the performance of EchoNext on data from institutions external to NYP, a docker container with the scripts and artifact for the model were sent to Cedars-Sinai Medical Center, the Montreal Heart Institute, and the University of California San Francisco Medical Center for additional validation. Local cohorts were independently curated following the same basic

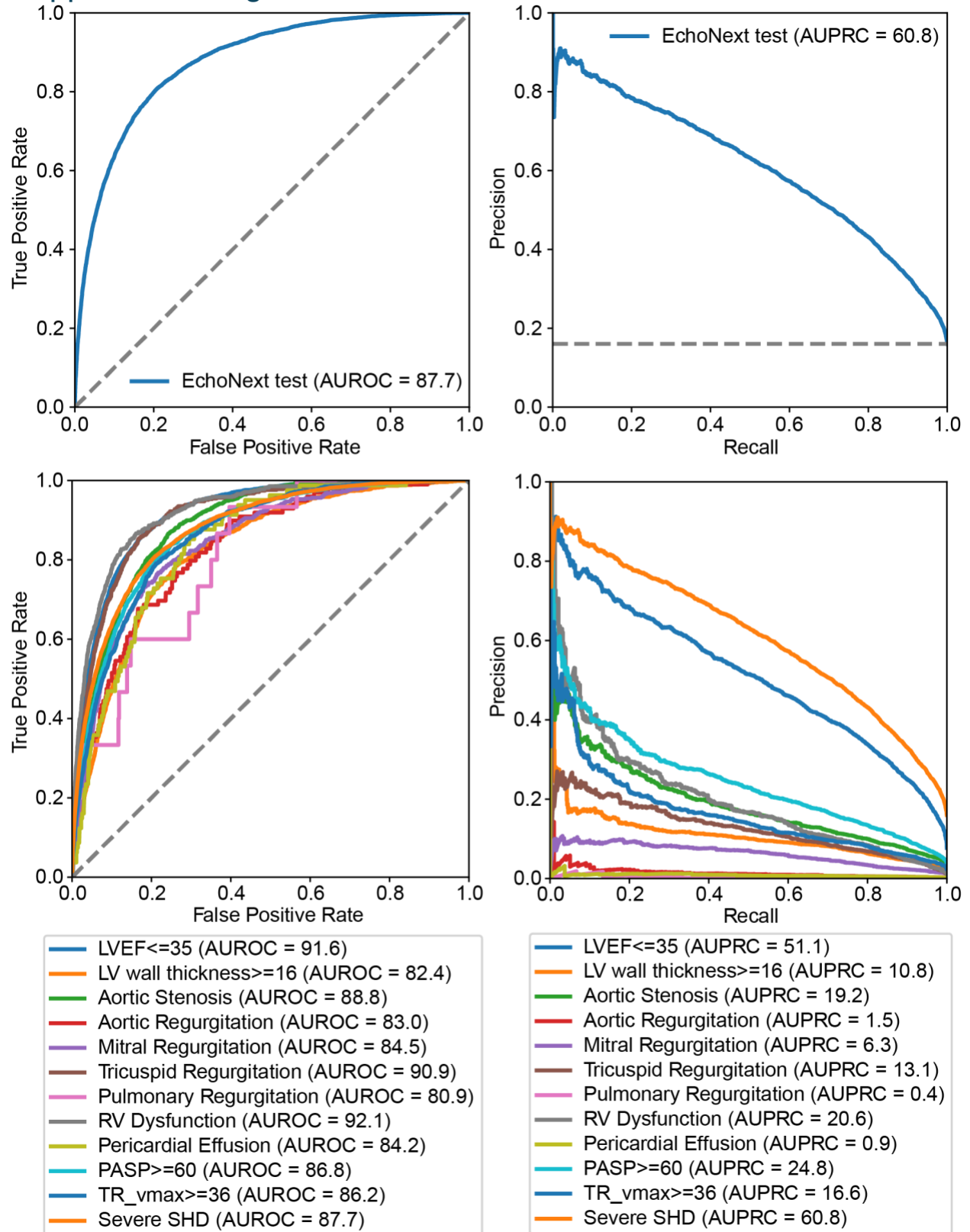
approach described above, with some variability permitted related to data exclusions for label missingness. Full cohort-specific details are as follows available in Supplemental Table 9.

Supplemental Figures

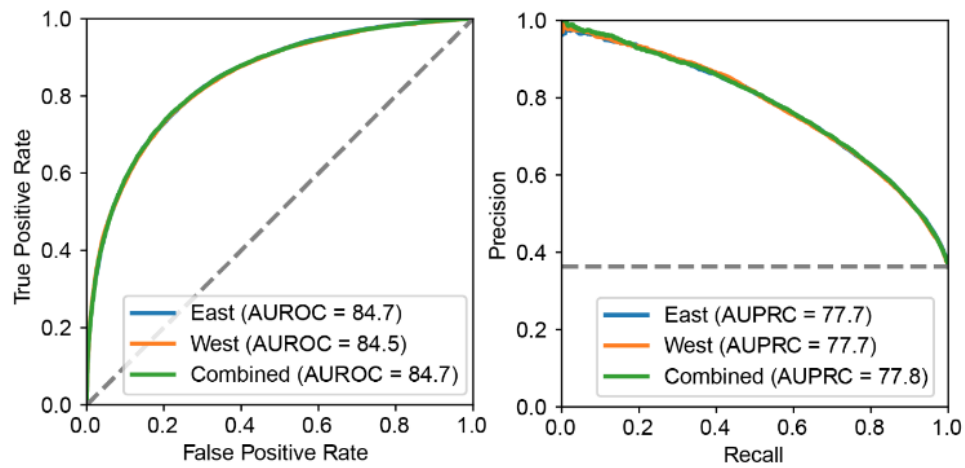
Supplemental Figure 1: EchoNext Calibration for SHD in the NYP Multicenter Cohort



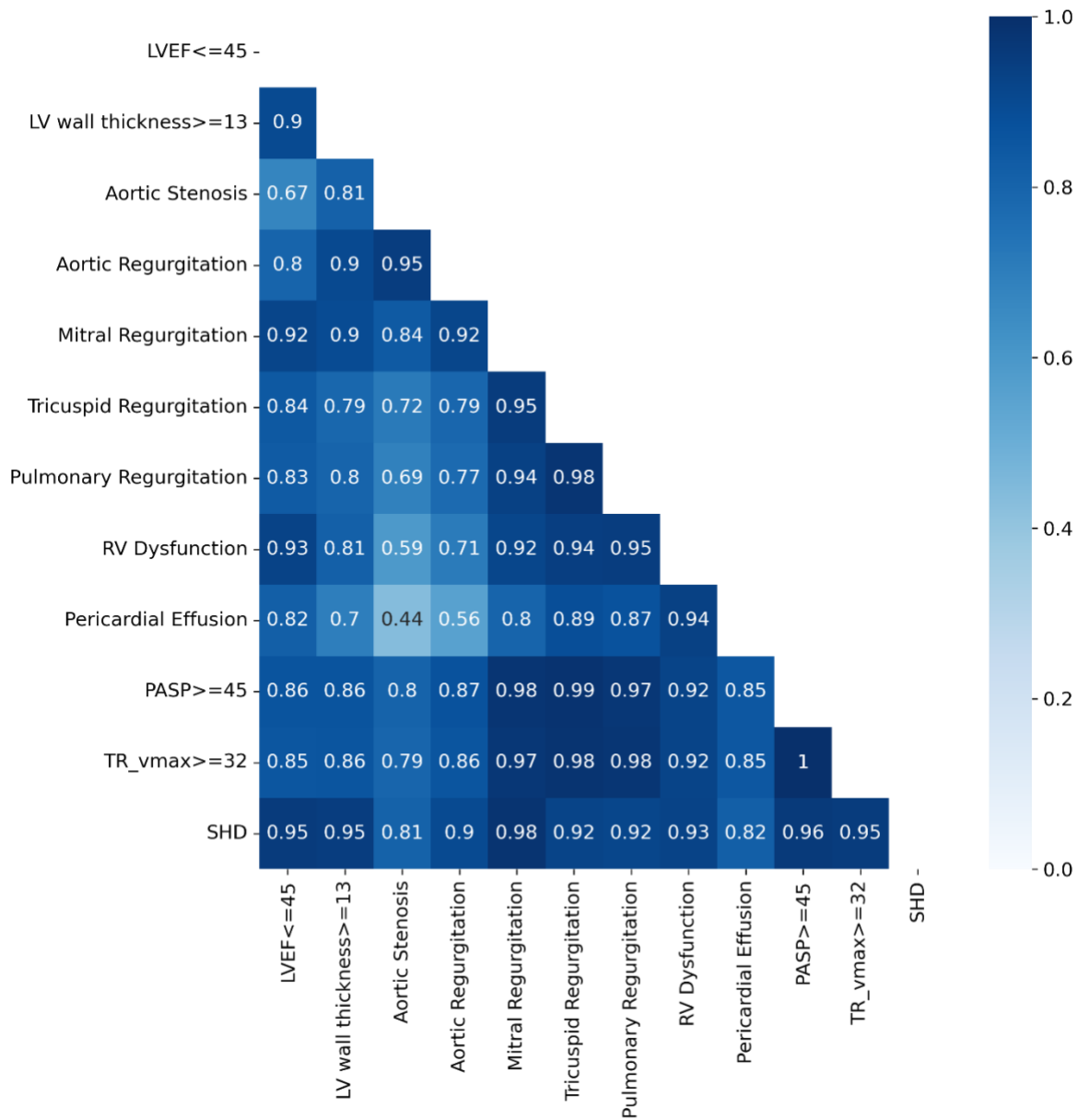
Supplemental Figure 2 Model results for Severe SHD



Supplemental Figure 3: Comparison of models trained with varied data splits by NYP Hospitals



Supplemental Figure 4: Correlation matrix with correlation coefficients between prediction scores of all component labels as well as the composite SHD label



Supplemental Tables

Supplemental Table 1 EchoNext Model Architecture and Parameters

Layer (batch x data size)	Input Shape (batch x data size)	Output Shape	Param #
=			
ResNet1dWithTabular	[16, 1, 2500, 12]	[16, 12]	--
└─Conv1d: 1-1	[16, 12, 2500]	[16, 16, 1250]	2,880
└─BatchNorm1d: 1-2	[16, 16, 1250]	[16, 16, 1250]	32
└─ReLU: 1-3	[16, 16, 1250]	[16, 16, 1250]	--
└─MaxPool1d: 1-4	[16, 16, 1250]	[16, 16, 625]	--
└─Sequential: 1-5	[16, 16, 625]	[16, 16, 625]	--
└─BasicBlock1d: 2-1	[16, 16, 625]	[16, 16, 625]	3,648
└─BasicBlock1d: 2-2	[16, 16, 625]	[16, 16, 625]	3,648
└─BasicBlock1d: 2-3	[16, 16, 625]	[16, 16, 625]	3,648
└─Sequential: 1-6	[16, 16, 625]	[16, 32, 313]	--
└─BasicBlock1d: 2-4	[16, 16, 625]	[16, 32, 313]	11,456
└─BasicBlock1d: 2-5	[16, 32, 313]	[16, 32, 313]	14,464
└─BasicBlock1d: 2-6	[16, 32, 313]	[16, 32, 313]	14,464
└─BasicBlock1d: 2-7	[16, 32, 313]	[16, 32, 313]	14,464
└─Sequential: 1-7	[16, 32, 313]	[16, 64, 157]	--
└─BasicBlock1d: 2-8	[16, 32, 313]	[16, 64, 157]	45,440
└─BasicBlock1d: 2-9	[16, 64, 157]	[16, 64, 157]	57,600
└─BasicBlock1d: 2-10	[16, 64, 157]	[16, 64, 157]	57,600
└─BasicBlock1d: 2-11	[16, 64, 157]	[16, 64, 157]	57,600
└─BasicBlock1d: 2-12	[16, 64, 157]	[16, 64, 157]	57,600
└─BasicBlock1d: 2-13	[16, 64, 157]	[16, 64, 157]	57,600
└─Sequential: 1-8	[16, 64, 157]	[16, 128, 79]	--
└─BasicBlock1d: 2-14	[16, 64, 157]	[16, 128, 79]	180,992
└─BasicBlock1d: 2-15	[16, 128, 79]	[16, 128, 79]	229,888
└─BasicBlock1d: 2-16	[16, 128, 79]	[16, 128, 79]	229,888
└─AdaptiveAvgPool1d: 1-9	[16, 128, 79]	[16, 128, 1]	--
└─AdaptiveMaxPool1d: 1-10	[16, 128, 79]	[16, 128, 1]	--
└─Dropout: 1-11	[16, 256, 1]	[16, 256, 1]	--
Tabular Feature Input	[16, 7]	--	--
└─Linear: 1-12	[16, 263]	[16, 12]	3,168
=====			
Total params: 1,046,080			
Trainable params: 1,046,080			
Non-trainable params: 0			
Total mult-adds (G): 2.08			
=====			
Input size (MB): 1.92			
Forward/backward pass size (MB): 95.12			
Params size (MB): 4.18			
Estimated Total Size (MB): 101.22			
=====			
=====			

Supplemental Table 2: Model Performance for all component labels

	AUROC	AUPRC	F1-SCORE	OR
SHD	85.2 [84.5, 85.9]	78.5 [77.3, 79.7]	67.9 [66.6, 69.1]	12.7 [11.5, 14.0]
LVEF≤45%	90.4 [89.6, 91.2]	64.8 [62.0, 67.2]	57.9 [56.0, 59.8]	22.7 [20.0, 26.0]
LVWT≥1.3CM	77.2 [76.0, 78.4]	32.9 [30.7, 35.2]	39.3 [37.5, 41.1]	5.4 [4.8, 6.1]
AORTIC STENOSIS	86.4 [85.0, 87.7]	26.5 [23.1, 30.2]	23.5 [21.7, 25.5]	12.9 [10.3, 16.2]
AORTIC REGURGITATION	77.7 [74.9, 80.5]	5.8 [4.5, 7.4]	8.0 [6.7, 9.3]	5.8 [4.3, 7.8]
MITRAL REGURGITATION	85.5 [84.1, 86.9]	28.5 [25.2, 32.1]	30.2 [28.1, 32.2]	12.0 [10.0, 14.3]
TRICUSPID REGURGITATION	86.6 [85.4, 87.8]	34.4 [30.9, 37.7]	36.6 [34.4, 38.9]	14.9 [12.4, 17.8]
PULMONARY REGURGITATION	78.6 [71.1, 85.3]	3.2 [1.1, 7.3]	2.7 [1.7, 3.8]	7.0 [3.5, 12.9]
RV SYSTOLIC DYSFUNCTION	90.8 [89.6, 91.9]	40.8 [36.4, 45.5]	33.9 [31.4, 36.3]	23.2 [18.5, 28.4]
PERICARDIAL EFFUSION	79.9 [76.1, 83.4]	3.3 [2.1, 5.0]	5.3 [3.9, 6.7]	5.7 [3.7, 8.8]
PASP≥45MMHG	82.7 [81.7, 83.8]	43.6 [40.9, 46.2]	46.0 [44.1, 47.9]	8.8 [7.8, 10.0]
TR VMAX≥3.2CM/S	84.8 [83.5, 86.1]	29.2 [26.1, 32.6]	32.5 [30.3, 34.5]	10.7 [9.0, 12.7]

Supplemental Table 3: Severe SHD Prevalence in the NYP Multicenter Cohort

	TRAIN (N=796,816)	VALIDATION (N=35,780)	TEST (N=44,719)
SEVERE SHD PREVALENCE	190467 (23.9%)	5636 (15.8%)	7180 (16.1%)
LVEF≤35%	109272 (13.7%)	2762 (7.7%)	3459 (7.7%)
LVWT≥1.6CM	26919 (3.4%)	911 (2.5%)	1066 (2.4%)
SEVERE AORTIC STENOSIS	14318 (1.8%)	866 (2.4%)	1135 (2.5%)
SEVERE AORTIC REGURGITATI ON	1781 (0.2%)	72 (0.2%)	99 (0.2%)
SEVERE MITRAL REGURGITATI ON	12223 (1.5%)	327 (0.9%)	508 (1.1%)
SEVERE TRICUSPID REGURGITATI ON	35854 (4.5%)	808 (2.3%)	1086 (2.4%)
SEVERE PULMONARY REGURGITATI ON	621 (0.1%)	16 (0.0%)	15 (0.0%)
SEVERE RV SYSTOLIC DYSFUNCTION	25623 (3.2%)	432 (1.2%)	555 (1.2%)
LARGE PERICARDIAL EFFUSION	6164 (0.8%)	72 (0.2%)	81 (0.2%)
PASP≥60MMH G	58350 (7.3%)	1376 (3.8%)	1785 (4.0%)
TR VMAX≥3.6CM/S	35854 (4.5%)	808 (2.3%)	1086 (2.4%)

Supplemental Table 4. Cohort characteristics by NYP Campus (Combined, East, West)

	COMBINED	EAST	WEST
N ECGS	877315	407876	469439
N PATIENTS	230318	110439	124208
AGE (YR, MEDIAN [IQR])	64 [52, 75]	65 [52, 75]	63 [51, 74]
AGE GROUPS			
18-59	349473 (39.8%)	156516 (38.4%)	192957 (41.1%)
60-69	209499 (23.9%)	96092 (23.6%)	113407 (24.2%)
70-79	181318 (20.7%)	86792 (21.3%)	94526 (20.1%)
80+	137025 (15.6%)	68476 (16.8%)	68549 (14.6%)
MALE SEX	447342 (51.0%)	206398 (50.6%)	240944 (51.3%)
RACE/ETHNICITY			
HISPANIC	200171 (22.8%)	52647 (12.9%)	147524 (31.4%)
WHITE	298568 (34.0%)	159030 (39.0%)	139538 (29.7%)
BLACK	153420 (17.5%)	74108 (18.2%)	79312 (16.9%)
ASIAN	48696 (5.6%)	36106 (8.9%)	12590 (2.7%)
OTHER	74801 (8.5%)	39489 (9.7%)	35312 (7.5%)
UNKNOWN	101659 (11.6%)	46496 (11.4%)	55163 (11.8%)
CLINICAL CONTEXT			
EMERGENCY	310680 (35.4%)	136615 (33.5%)	174065 (37.1%)
INPATIENT	378789 (43.2%)	158988 (39.0%)	219801 (46.8%)
OUTPATIENT	166806 (19.0%)	106536 (26.1%)	60270 (12.8%)
PROCEDURAL	15752 (1.8%)	3285 (0.8%)	12467 (2.7%)
UNKNOWN	5288 (0.6%)	2452 (0.6%)	2836 (0.6%)
SHD PREVALENCE	386936 (44.1%)	165409 (40.6%)	221527 (47.2%)
LVEF≤45%	173003 (19.7%)	71376 (17.5%)	101627 (21.6%)
LVWT≥1.3CM	143753 (16.4%)	44214 (10.8%)	99539 (21.2%)
AORTIC STENOSIS*	35539 (4.1%)	16513 (4.0%)	19026 (4.1%)
AORTIC REGURGITATION*	16446 (1.9%)	9661 (2.4%)	6785 (1.4%)
MITRAL REGURGITATION*	70549 (8.0%)	33317 (8.2%)	37232 (7.9%)
TRICUSPID REGURGITATION*	84949 (9.7%)	39610 (9.7%)	45339 (9.7%)
PULMONARY REGURGITATION*	6269 (0.7%)	2813 (0.7%)	3456 (0.7%)
RV SYSTOLIC DYSFUNCTION**	80425 (9.2%)	29898 (7.3%)	50527 (10.8%)
PERICARDIAL EFFUSION***	20814 (2.4%)	10092 (2.5%)	10722 (2.3%)
PASP≥45MMHG	166204 (18.9%)	87071 (21.3%)	79133 (16.9%)
TR VMAX≥3.2CM/S	88045 (10.0%)	45557 (11.2%)	42488 (9.1%)

Note: 1. For validation and test set, only 1 ECG per patient was used, therefore total number of ECGs is smaller than the number reported in results and figure 1 (N=1,245,273), however the total number of patients stays the same. 2. There are a small percentage of patients who had ECGs from both East and West campus; therefore, the sum of east and west campus patients is greater than the total number of patients. However, there is no patient overlap between east and west campus test set.

*Valvular heart disease labels were defined as moderate or severe in severity

**Moderate or severe RV systolic dysfunction

***Moderate or large pericardial effusion

Supplemental Table 5: Patient characteristics for internal cross-validation cohorts

Model Training Set	Combined		East		West	
	Train	Validation	Train	Validation	Train	Validation
<i>N</i> ECGs	414079	17891	370157	16759	397929	19021
<i>N</i> patients	74909	17891	72720	16759	77099	19021
Age (yr, median[IQR])	64 [52, 74]	64 [51, 76]	65 [53, 75]	65 [51, 76]	63 [52, 74]	64 [52, 76]
Age groups						
18-59	168055 (40.6%)	7065 (39.5%)	141446 (38.2%)	6627 (39.5%)	160755 (40.4%)	7467 (39.3%)
60-69	98947 (23.9%)	3901 (21.8%)	88044 (23.8%)	3622 (21.6%)	97274 (24.4%)	4208 (22.1%)
70-79	85272 (20.6%)	3685 (20.6%)	79015 (21.3%)	3466 (20.7%)	81627 (20.5%)	3892 (20.5%)
80+	61805 (14.9%)	3240 (18.1%)	61652 (16.7%)	3044 (18.2%)	58273 (14.6%)	3454 (18.2%)
Male sex	213078 (51.5%)	8591 (48.0%)	188535 (50.9%)	7917 (47.2%)	204844 (51.5%)	9233 (48.5%)
Race/Ethnicity						
Hispanic	93669 (22.6%)	3547 (19.8%)	48513 (13.1%)	1821 (10.9%)	128386 (32.3%)	5252 (27.6%)
White	139538 (33.7%)	6219 (34.8%)	144172 (38.9%)	6605 (39.4%)	117856 (29.6%)	5977 (31.4%)
Black	76551 (18.5%)	2701 (15.1%)	68794 (18.6%)	2402 (14.3%)	64599 (16.2%)	2859 (15.0%)
Asian	22156 (5.4%)	1125 (6.3%)	32165 (8.7%)	1752 (10.5%)	10574 (2.7%)	518 (2.7%)
Other	34746 (8.4%)	1633 (9.1%)	35770 (9.7%)	1630 (9.7%)	29811 (7.5%)	1601 (8.4%)
Unknown	47419 (11.5%)	2666 (14.9%)	40743 (11.0%)	2549 (15.2%)	46703 (11.7%)	2814 (14.8%)
Clinical Context						
Emergency	148035 (35.8%)	5917 (33.1%)	126326 (34.1%)	4617 (27.5%)	146272 (36.8%)	7317 (38.5%)
Inpatient	181291 (43.8%)	6999 (39.1%)	145640 (39.3%)	5884 (35.1%)	188913 (47.5%)	7960 (41.8%)
Outpatient	74971 (18.1%)	4462 (24.9%)	92996 (25.1%)	5995 (35.8%)	50072 (12.6%)	2976 (15.6%)
Procedural	7336 (1.8%)	374 (2.1%)	2981 (0.8%)	139 (0.8%)	10406 (2.6%)	583 (3.1%)
Unknown	2446 (0.6%)	139 (0.8%)	2214 (0.6%)	124 (0.7%)	2266 (0.6%)	185 (1.0%)
SHD prevalence	185156 (44.7%)	6594 (36.9%)	152757 (41.3%)	5637 (33.6%)	192567 (48.4%)	7357 (38.7%)
LVEF≤45%	84062 (20.3%)	2284 (12.8%)	67420 (18.2%)	1767 (10.5%)	89975 (22.6%)	2798 (14.7%)

<i>LVWT≥1.3cm</i>	69685 (16.8%)	2439 (13.6%)	40603 (11.0%)	1596 (9.5%)	86277 (21.7%)	3171 (16.7%)
<i>Aortic Stenosis*</i>	16120 (3.9%)	827 (4.6%)	14981 (4.0%)	621 (3.7%)	16328 (4.1%)	977 (5.1%)
<i>Aortic Regurgitation*</i>	7344 (1.8%)	324 (1.8%)	8842 (2.4%)	381 (2.3%)	5791 (1.5%)	258 (1.4%)
<i>Mitral Regurgitation*</i>	33807 (8.2%)	997 (5.6%)	31318 (8.5%)	884 (5.3%)	32917 (8.3%)	1048 (5.5%)
<i>Tricuspid Regurgitation*</i>	41492 (10.0%)	1082 (6.0%)	37319 (10.1%)	984 (5.9%)	39366 (9.9%)	1195 (6.3%)
<i>Pulmonary Regurgitation*</i>	3367 (0.8%)	76 (0.4%)	2652 (0.7%)	74 (0.4%)	3056 (0.8%)	68 (0.4%)
<i>RV Systolic Dysfunction**</i>	40529 (9.8%)	806 (4.5%)	28746 (7.8%)	493 (2.9%)	44804 (11.3%)	1069 (5.6%)
<i>Pericardial Effusion***</i>	10525 (2.5%)	177 (1.0%)	9687 (2.6%)	198 (1.2%)	9555 (2.4%)	175 (0.9%)
<i>PASP≥45mmHg</i>	79885 (19.3%)	2369 (13.2%)	81374 (22.0%)	2572 (15.3%)	69358 (17.4%)	2070 (10.9%)
<i>TR Vmax≥3.2cm/s</i>	43202 (10.4%)	1143 (6.4%)	42755 (11.6%)	1253 (7.5%)	37135 (9.3%)	1022 (5.4%)

*Valvular heart disease labels were defined as moderate or severe in severity

**Moderate or severe RV systolic dysfunction

***Moderate or large pericardial effusion

Supplemental Table 6: Comparison of results for models with variable training sets. All models evaluated on common test set comprising data from East and West campuses

<i>Model Training Set</i>	<i>AUROC</i>	<i>AUPRC</i>	<i>F1-score</i>	<i>OR</i>
<i>East</i>	84.7 [84.0, 85.4]	77.7 [76.4, 78.9]	68.9 [67.7, 70.1]	11.6 [10.6, 12.8]
<i>West</i>	84.5 [83.8, 85.3]	77.7 [76.5, 79.0]	68.0 [66.6, 69.2]	11.8 [10.6, 13.0]
<i>Combined</i>	84.7 [83.9, 85.4]	77.8 [76.5, 79.1]	68.2 [67.0, 69.4]	12.0 [10.9, 13.2]

Supplemental Table 7: Model performance in ECGs acquired within 0-3 months, 3-6 months, 6-9 months and 9-12 months prior to the echocardiogram

TIME FROM ECG TO ECHO	N ECG	AUROC	AUPRC
0-3 MONTHS	33244	84.5	79.1
3-6 MONTHS	2166	82.5	80.8
6-9 MONTHS	3047	85.0	81.8
9-12 MONTHS	2090	83.1	84.04

Supplemental Table 8: Model performance in various combinations of individual diseases

	N(%) ECG	AUROC	AUPRC	F1-SCORE	OR
LV/RV DYSFUNCTION	6402 (14.3%)	90.1 [89.2, 90.8]	66.3 [63.9, 68.6]	59.5 [57.6, 61.2]	21.4 [18.7, 24.2]
VALVULAR DISEASE	6634 (14.8%)	84.8 [83.8, 85.7]	50.1 [47.5, 52.7]	47.9 [46.3, 49.4]	11.3 [10.2, 13.1]
LEFT-SIDED HEART DISEASE	12770 (28.5%)	84.8 [84.0, 85.6]	70.6 [69.0, 72.1]	65.5 [64.3, 66.8]	11.2 [10.2, 12.4]
RIGHT-SIDED HEART DISEASE	7776 (17.4%)	84.5 [83.5, 85.4]	57.7 [55.3, 60.1]	55.3 [53.6, 56.9]	10.7 [9.6, 12.0]

Valvular disease: any of moderate or severe aortic, mitral, tricuspid, or pulmonary regurgitation, or aortic stenosis;

Left-sided heart disease: any of LVEF $\leq$ 45%, LVWT $\geq$ 1.3cm, moderate or severe aortic stenosis or regurgitation, or mitral regurgitation;

Right-sided heart disease: any of moderate or severe RV systolic dysfunction, moderate or severe tricuspid or pulmonary regurgitation, PASP $\geq$ 45mmHg, or TR Vmax $\geq$ 3.2cm/s.

Supplemental Table 9: Characteristics for external validation cohorts

	CEDARS-SINAI	MONTREAL HEART INSTITUTE	UCSF
N PATIENTS / ECGS	10177	10862	6106
AGE (YR, MEDIAN[IQR])	70 [58, 80]	66 [54, 76]	66 [53, 76]
AGE GROUPS			
18-59	2706 (26.6%)	3798 (35.0%)	2238 (36.7%)
60-69	2217 (21.8%)	2455 (22.6%)	1414 (23.2%)
70-79	2593 (25.5%)	2804 (25.8%)	1363 (22.3%)
80+	2661 (26.1%)	1805 (16.6%)	1091 (17.9%)
MALE SEX	5340 (52.5%)	6696 (61.6%)	3133 (51.3%)
SHD PREVALENCE	5312 (52.2%)	5904 (54.4%)	2782 (45.6%)
LVEF≤45%	1516 (14.9%)	2079 (19.1%)	761 (12.5%)
LVWT≥1.3CM	2432 (23.9%)	1030 (9.5%)	1142 (18.7%)
AORTIC STENOSIS*	128 (1.26%)	915 (8.4%)	245 (4.0%)
AORTIC REGURGITATION*	387 (3.8%)	396 (3.6%)	78 (1.3%)
MITRAL REGURGITATION*	1425 (14.0%)	1587 (14.6%)	324 (5.3%)
TRICUSPID REGURGITATION*	1816 (17.8%)	2070 (19.1%)	518 (8.4%)
PULMONARY REGURGITATION*	224 (2.2%)	208 (1.9%)	39 (0.6%)
RV SYSTOLIC DYSFUNCTION**	147 (1.44%)	228 (2.1%)	270 (4.4%)
PERICARDIAL EFFUSION***	71 (0.7%)	650 (6.0%)	38 (0.6%)
PASP≥45MMHG	1313 (12.9%)	1017 (9.4%)	1079 (17.7%)
TR VMAX≥3.2CM/S	845 (8.3%)	7 (0.1%)	911 (14.9%)

*Valvular heart disease labels were defined as moderate or severe in severity

**Moderate or severe RV systolic dysfunction

***Moderate or large pericardial effusion

Supplemental Table 10: Performance metrics at fixed sensitivity of 70% in NYP multi-center cohort and external validation cohorts

	PPV (%)	NPV (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
NYP multi-center test	70.3	83	70	83.2	78.4
Montreal	75.6	67.3	70	73.1	71.5
Cedars-Sinai	73.8	69	70	72.8	71.4
UCSF	69.8	74.9	70	74.6	72.6

Supplemental Table 11: Cardiologist Survey Results

CARDIOLOGIST INTERPRETATION - WITHOUT ECHONEXT

	Structural Disease	No Structural Disease	Total
STRUCTURAL DISEASE PREDICTED	408	316	724
NO STRUCTURAL DISEASE PREDICTED	260	616	876
TOTAL	668	932	

CARDIOLOGIST INTERPRETATION - WITH ECHONEXT

	Structural Disease	No Structural Disease	Total
STRUCTURAL DISEASE PREDICTED	432	257	689
NO STRUCTURAL DISEASE PREDICTED	236	675	911
TOTAL	668	932	

ECHONEXT PERFORMANCE - ALONE

	Structural Disease	No Structural Disease	Total
STRUCTURAL DISEASE PREDICTED	45	17	62
NO STRUCTURAL DISEASE PREDICTED	17	71	88
TOTAL	62	88	

Supplemental Table 12: Cardiologists survey results in ECGs with and without normal finding

	<i>Accuracy (%)</i>			
	N (%) ECG	% SHD	EchoNext - alone	Cardiologist – without EchoNext
<i>Normal ECG</i>	36 (24%)	25	77.8	69±6
<i>Borderline or Abnormal ECG</i>	114 (76%)	46.5	77.2	62.5±4.7

Supplemental Table 13: Patient characteristics for the Columbia ECG Cohort

	ALL	TRAIN	VALIDATION	TEST
N ECGS	100,000	72,475	4,626	5,442
N PATIENTS	36,286	26,218	4,626	5,442
AGE (YR, MEDIAN[IQR])	63 [51, 73]	63 [52, 73]	64 [52, 75]	64 [52, 74]
AGE GROUPS				
18-59	41,253 (41.3%)	29,783 (41.1%)	1,787 (38.6%)	2,124 (39.0%)
60-69	25,982 (26.0%)	18,745 (25.9%)	1,093 (23.6%)	1,318 (24.2%)
70-79	20,190 (20.2%)	14,898 (20.6%)	975 (21.1%)	1,154 (21.2%)
80+	12,575 (12.6%)	9,049 (12.5%)	771 (16.7%)	846 (15.5%)
MALE SEX	53,581 (53.6%)	38,951 (53.7%)	2,270 (49.1%)	2,711 (49.8%)
RACE/ETHNICITY				
HISPANIC	31,013 (31.0%)	22,806 (31.5%)	1,351 (29.2%)	1,649 (30.3%)
WHITE	29,211 (29.2%)	21,289 (29.4%)	1,385 (29.9%)	1,569 (28.8%)
BLACK	16,461 (16.5%)	11,559 (15.9%)	728 (15.7%)	846 (15.5%)
ASIAN	3,428 (3.4%)	2,602 (3.6%)	134 (2.9%)	153 (2.8%)
OTHER	7,429 (7.4%)	5,272 (7.3%)	380 (8.2%)	457 (8.4%)
UNKNOWN	12,458 (12.5%)	8,947 (12.3%)	648 (14.0%)	768 (14.1%)
CLINICAL CONTEXT				
EMERGENCY	31453 (31.5%)	22811 (31.5%)	1688 (36.5%)	1971 (36.2%)
INPATIENT	48162 (48.2%)	34906 (48.2%)	1903 (41.1%)	2203 (40.5%)
OUTPATIENT	17194 (17.2%)	12423 (17.1%)	858 (18.5%)	1059 (19.5%)
PROCEDURAL	3191 (3.2%)	2335 (3.2%)	177 (3.8%)	209 (3.8%)
SHD PREVALENCE	52188 (52.2%)	37958 (52.4%)	1990 (43.0%)	2318 (42.6%)
LVEF≤45%	23892 (23.9%)	16962 (23.4%)	866 (18.7%)	962 (17.7%)
LVWT≥1.3CM	24220 (24.2%)	17667 (24.4%)	877 (19.0%)	1061 (19.5%)
AORTIC STENOSIS	4054 (4.1%)	2919 (4.0%)	252 (5.4%)	286 (5.3%)
AORTIC REGURGITATION	1264 (1.3%)	878 (1.2%)	62 (1.3%)	66 (1.2%)
MITRAL REGURGITATION	8451 (8.5%)	6137 (8.5%)	282 (6.1%)	337 (6.2%)
TRICUSPID REGURGITATION	10651 (10.7%)	7707 (10.6%)	305 (6.6%)	353 (6.5%)
PULMONARY REGURGITATION	821 (0.8%)	603 (0.8%)	21 (0.5%)	20 (0.4%)
RV SYSTOLIC DYSFUNCTION	13243 (13.2%)	9597 (13.2%)	368 (8.0%)	419 (7.7%)
PERICARDIAL EFFUSION	3023 (3.0%)	2079 (2.9%)	52 (1.1%)	69 (1.3%)
PASP≥45MMHG	18993 (19.0%)	13727 (18.9%)	581 (12.6%)	699 (12.8%)
TR VMAX≥3.2CM/S	10212 (10.2%)	7492 (10.3%)	267 (5.8%)	375 (6.9%)

Supplemental Table 14: Component label performance of the Columbia Mini-Model in NYP Multicenter test set

	AUROC	AUPRC	F1-SCORE	OR
SHD	83.1 [82.3, 83.8]	75.0 [73.6, 76.3]	67.7 [66.5, 68.8]	9.8 [9.0, 10.7]
LVEF≤45%	87.4 [86.4, 88.2]	52.6 [49.8, 55.3]	49.0 [47.2, 50.7]	15.5 [13.6, 17.8]
LVWT≥1.3CM	74.8 [73.5, 76.0]	28.4 [26.4, 30.3]	37.3 [35.6, 39.2]	4.8 [4.3, 5.5]
AORTIC STENOSIS	84.3 [82.8, 85.7]	21.5 [18.7, 24.7]	22.2 [20.4, 23.8]	10.4 [8.4, 12.7]
AORTIC REGURGITATION	74.6 [71.6, 77.7]	5.1 [3.8, 6.6]	7.0 [5.9, 8.2]	4.4 [3.2, 5.9]
MITRAL REGURGITATION	82.8 [81.3, 84.4]	23.5 [20.5, 26.4]	25.6 [23.8, 27.4]	9.7 [8.1, 11.7]
TRICUSPID REGURGITATION	83.2 [81.7, 84.8]	28.4 [25.3, 31.5]	31.0 [28.9, 33.0]	10.6 [9.0, 12.7]
PULMONARY REGURGITATION	74.6 [66.3, 82.5]	2.7 [0.9, 7.0]	2.3 [1.4, 3.2]	6.1 [3.0, 11.3]
RV SYSTOLIC DYSFUNCTION	89.0 [87.7, 90.3]	33.7 [29.6, 37.9]	27.0 [24.9, 29.0]	18.4 [14.5, 23.2]
PERICARDIAL EFFUSION	76.3 [71.3, 80.5]	2.9 [1.9, 4.1]	4.2 [3.3, 5.3]	5.6 [3.7, 8.6]
PASP≥45MMHG	79.2 [77.9, 80.4]	36.4 [33.9, 38.9]	41.2 [39.5, 42.8]	6.7 [5.9, 7.5]
TR VMAX≥3.2CM/S	80.8 [79.3, 82.2]	21.6 [19.1, 24.1]	26.8 [24.9, 28.6]	7.7 [6.5, 9.0]

Supplemental Table 15: Component label performance of the Columbia Mini-Model in the Columbia ECG test set

	AUROC	AUPRC	F1-SCORE	OR
SHD	82.0 [80.9, 83.0]	78.9 [77.3, 80.4]	69.0 [67.2, 70.6]	9.3 [8.1, 10.6]
LVEF≤45%	85.2 [83.9, 86.5]	60.0 [56.7, 63.3]	54.8 [52.4, 57.0]	11.3 [9.6, 13.3]
LVWT≥1.3CM	73.4 [71.9, 75.0]	37.2 [34.3, 40.1]	45.4 [43.1, 47.6]	4.5 [3.9, 5.1]
AORTIC STENOSIS	85.9 [84.0, 87.7]	25.2 [21.0, 29.5]	25.2 [22.5, 27.5]	13.2 [9.8, 18.2]
AORTIC REGURGITATION	73.9 [68.4, 78.9]	3.5 [2.2, 5.1]	4.8 [3.5, 6.1]	4.4 [2.6, 7.8]
MITRAL REGURGITATION	80.6 [78.1, 83.1]	22.5 [18.9, 26.1]	26.2 [23.6, 28.6]	8.6 [6.7, 11.3]
TRICUSPID REGURGITATION	83.3 [81.3, 85.4]	29.6 [24.8, 34.5]	29.2 [26.6, 32.0]	8.9 [7.0, 11.4]
PULMONARY REGURGITATION	82.9 [72.0, 93.4]	14.5 [2.2, 30.8]	2.5 [1.4, 3.8]	11.3 [3.3, 32.8]
RV SYSTOLIC DYSFUNCTION	86.6 [84.8, 88.5]	42.9 [38.0, 47.7]	36.8 [34.1, 39.4]	15.0 [11.7, 19.3]
PERICARDIAL EFFUSION	76.6 [71.2, 82.0]	7.6 [3.7, 12.7]	5.2 [3.7, 6.7]	4.4 [2.6, 7.1]
PASP≥45MMHG	77.0 [75.2, 78.8]	35.7 [32.3, 39.1]	38.8 [36.1, 41.3]	5.6 [4.7, 6.6]
TR VMAX≥3.2CM/S	75.4 [73.0, 77.8]	21.3 [17.4, 25.3]	23.7 [21.2, 26.1]	4.5 [3.7, 5.7]

Supplemental Table 16: DISCOVERY clinical trial patients characteristics

Variable	Value
Age, yr (median[IQR])	80 [72, 86]
Male sex	43
Race/Ethnicity	
Hispanic	57
White	26
Black	17

Supplemental Table 17: Pilot clinical trial outcomes stratified by EchoNext and ValveNet prediction score categories

	ECHONEXT			VALVENET	
	(0.0, 0.3]	(0.3, 0.6]	(0.6, 1.0]	(0.3, 0.6]	(0.6, 1.0]
N	17	50	33	47	53
LVEF≤45%	0	4	10	2	12
LVWT≥1.3CM	0	7	9	6	10
AORTIC STENOSIS*	0	0	2	0	2
AORTIC REGURGITATION*	0	0	0	0	0
MITRAL REGURGITATION*	0	1	6	0	7
TRICUSPID REGURGITATION*	0	3	1	0	4
PULMONARY REGURGITATION*	0	0	0	0	0
RV SYSTOLIC DYSFUNCTION**	0	1	3	1	3
PASP≥45MMHG	1	1	7	1	8
PERICARDIAL EFFUSION***	0	0	0	0	0
LEFT-VHD¹	0	1	8	0	9
SHD²	1	14	24	9	30

*Valvular heart disease labels were defined as moderate or severe in severity

**Moderate or severe RV systolic dysfunction

***Moderate or large pericardial effusion

¹Left-VHD: Left-sided valvular heart disease, including any of moderate or severe aortic stenosis, aortic regurgitation, or mitral regurgitation

²SHD: Structural heart disease, including any of LVEF≤45%, LV wall thickness ≥1.3 cm, moderate or severe RV dysfunction, any moderate or severe valvular heart disease, PASP≥45mmHg, or a moderate or large pericardial effusion

Supplemental Table 18: Training site characteristics

	Columbia	Weill Cornell	Allen	Westchester	Lower Manhattan	Brooklyn Methodist	Queens	Hudson Valley
Number of Beds	738	862	196	291	180	591	535	128
Hospital Type	Academic	Academic	Community	Community	Community	Tertiary	Tertiary	Community
ED Visits/yr	81,848	73,692	40,290	34,946	35,583	71,884	75,828	32,882
Inpatient Visits/yr	32,012	48,402	11,663	11,666	9,618	30,078	28,838	8,466
Payor								
Medicare	34.5%	30.2%	38.6%	49.8%	29.0%	38.0%	45.7%	51.1%
Medicaid	38.3%	17.9%	48.2%	19.9%	21.8%	33.5%	34.5%	17.1%
Commercial	26.0%	50.5%	13.4%	28.8%	41.5%	26.9%	17.5%	29.6%
Race/Ethnicity								
Asian	2.3%	11.4%	0.9%	2.0%	34.6%	4.0%	34.9%	1.3%
Black	20.5%	12.7%	19.6%	23.6%	15.0%	39.9%	12.3%	10.6%
Hispanic	35.3%	12.2%	44.9%	20.3%	19.7%	11.0%	24.6%	18.0%
White	32.6%	53.6%	25.9%	44.1%	26.9%	35.9%	25.9%	65.3%
Unknown	9.1%	9.8%	8.4%	9.8%	3.6%	9.1%	2.1%	4.6%