Supplementary information for "GS-ECG: a generalist–specialist approach to structural heart disease prediction from electrocardiograms"

Supplementary Tables

Table S1 Overall model performance on the MIMIC-IV-Ext-SHD test set.

Model	Mean AUROC	Mean AUPRC	SHD AUROC	SHD AUPRC
Tabular only				
Logistic regression	0.7110 [0.6973–0.7248]	0.2278 [0.2200–0.2389]	0.7428 [0.7312–0.7540]	0.7683 [0.7553–0.7819]
XGBoost	0.7207 [0.7075–0.7332]	0.2362 [0.2304–0.2473]	0.7513 [0.7400–0.7623]	0.7769 [0.7637–0.7913]
Multilayer perceptron	0.7206 [0.7071–0.7335]	0.2298 [0.2238–0.2401]	0.7469 [0.7355–0.7582]	0.7720 [0.7589–0.7864]
Waveform only				
CNN1D	0.7903 [0.7787–0.8003]	0.2800 [0.2729–0.2935]	0.7892 [0.7784–0.7989]	0.8198 [0.8075–0.8327]
CRNN	0.7896 [0.7785–0.8002]	0.2924 [0.2848–0.3064]	0.8015 [0.7910–0.8119]	0.8267 [0.8139–0.8399]
Transformer	0.7760 [0.7641–0.7868]	0.2850 [0.2778–0.2962]	0.7976 [0.7870–0.8072]	0.8242 [0.8115–0.8366]
Tabular + Waveform				
ECGFounder-Frozen	0.7951 [0.7835–0.8061]	0.2995 [0.2922–0.3118]	0.8116 [0.8022–0.8204]	0.8380 [0.8262–0.8489]
ECGFounder-FT	0.7942 [0.7821–0.8058]	0.3102 [0.3024–0.3250]	0.8135 [0.8041–0.8230]	0.8420 [0.8309–0.8528]
EchoNext-Mini Model	0.7926 [0.7816–0.8031]	0.2827 [0.2755–0.2956]	0.8032 [0.7932–0.8128]	0.8306 [0.8183–0.8418]
DGRNet	0.8149 [0.8031–0.8253]	0.3293 [0.3189–0.3460]	0.8220 [0.8121–0.8312]	0.8505 [0.8392–0.8608]
GS-ECG	**0.8231** [0.8123–0.8332]	**0.3353** [0.3249–0.3530]	**0.8248** [0.8154–0.8339]	**0.8524** [0.8410–0.8624]

Note. Values are point estimates with 95% CIs from 1,000 patient-level bootstrap resamples. Mean metrics are macro-averages over all 11 outputs, including composite SHD. ECGFounder-Frozen, ECGFounder-FT, and DGRNet were trained on the same predictor-training patients. GS-ECG calibrated and fused ECGFounder-Frozen and DGRNet with parameters fitted on the held-out fusion-training patients. The EchoNext-Mini Model was retrained on MIMIC-IV-Ext-SHD. Bold indicates the highest reported value in each column.

Table S2 Overall model performance on the EchoNext-Mini Dataset test set.

Model	Mean AUROC	Mean AUPRC	SHD AUROC	SHD AUPRC
Tabular only				
Logistic regression	0.7237 [0.7074–0.7385]	0.2171 [0.2086–0.2308]	0.7480 [0.7346–0.7605]	0.6770 [0.6560–0.6988]
XGBoost	0.7263 [0.7085–0.7435]	0.2337 [0.2251–0.2471]	0.7630 [0.7500–0.7745]	0.7039 [0.6831–0.7236]
Multilayer perceptron	0.7310 [0.7144–0.7470]	0.2314 [0.2220–0.2454]	0.7607 [0.7478–0.7726]	0.7011 [0.6808–0.7210]
Waveform only				
CNN1D	0.8050 [0.7916–0.8175]	0.3216 [0.3075–0.3432]	0.8225 [0.8111–0.8332]	0.7905 [0.7749–0.8049]
CRNN	0.7978 [0.7842–0.8103]	0.3017 [0.2897–0.3205]	0.8129 [0.8015–0.8245]	0.7795 [0.7620–0.7946]
Transformer	0.7932 [0.7796–0.8060]	0.3021 [0.2911–0.3180]	0.8108 [0.7992–0.8218]	0.7769 [0.7600–0.7939]
Tabular + Waveform				
ECGFounder-Frozen	0.8034 [0.7917–0.8148]	0.3187 [0.3076–0.3355]	0.8279 [0.8162–0.8390]	0.7988 [0.7817–0.8149]
ECGFounder-FT	0.8063 [0.7947–0.8176]	0.3227 [0.3115–0.3387]	0.8304 [0.8188–0.8418]	0.8033 [0.7863–0.8191]
EchoNext-Mini Model	0.8026 [0.7903–0.8148]	0.3115 [0.2974–0.3346]	0.8197 [0.8081–0.8314]	0.7888 [0.7719–0.8054]
DGRNet	0.8213 [0.8098–0.8329]	0.3434 [0.3307–0.3650]	0.8360 [0.8248–0.8466]	0.8033 [0.7869–0.8189]
GS-ECG	**0.8250** [0.8139–0.8360]	**0.3487** [0.3361–0.3696]	**0.8394** [0.8285–0.8506]	**0.8112** [0.7952–0.8267]

Note. Values are point estimates with 95% CIs from 1,000 patient-level bootstrap resamples on the same 5,442 test patients. The EchoNext-Mini Model is the authors' released checkpoint. Mean metrics are macro-averages over all 12 outputs, including composite SHD. Bold indicates the highest reported value in each column.

Table S3 Label-wise AUPRC and AUROC of the EchoNext-Mini Model, ECGFounder-Frozen, ECGFounder-FT, DGRNet, and GS-ECG on the MIMIC-IV-Ext-SHD test set.

Label	Positive patients	EN-Mini	Frozen	FT	DGRNet	GS-ECG
a. AUPRC						
LVEF $\leq$ 45%	1,538	0.5709	0.6345	0.6583	0.6913	**0.6946**
LV wall thickness $\geq$ 1.3 cm	2,121	0.4120	0.4694	0.4853	0.5090	**0.5121**
Aortic stenosis	215	0.1242	0.1194	0.1174	0.1440	**0.1503**
Aortic regurgitation	29	0.0315	0.0298	0.0522	0.0853	**0.0913**
Mitral regurgitation	277	0.1353	0.1276	0.1494	0.1661	**0.1684**
Tricuspid regurgitation	382	0.2108	0.2369	0.2412	0.2561	**0.2682**
Pulmonary regurgitation	157	0.0741	0.0665	0.0686	0.0803	**0.0830**
RV systolic dysfunction	580	0.3124	0.3195	0.3341	0.3670	**0.3746**
Pericardial effusion	103	0.0516	0.0666	0.0675	0.0721	**0.0810**
TR maximum velocity $\geq$ 3.2 m/s	1,142	0.3567	0.3865	0.3965	0.4002	**0.4124**
Composite SHD	4,136	0.8306	0.8380	0.8420	0.8505	**0.8524**
b. AUROC						
LVEF $\leq$ 45%	1,538	0.8310	0.8596	0.8684	0.8783	**0.8811**
LV wall thickness $\geq$ 1.3 cm	2,121	0.6758	0.7060	0.7163	0.7273	**0.7303**
Aortic stenosis	215	0.8141	0.8116	0.8196	0.8387	**0.8444**
Aortic regurgitation	29	0.7904	0.7687	0.7687	0.8113	**0.8348**
Mitral regurgitation	277	0.8095	0.8141	0.8138	0.8368	**0.8427**
Tricuspid regurgitation	382	0.8313	0.8346	0.8381	0.8538	**0.8584**
Pulmonary regurgitation	157	0.7864	0.7624	0.7556	0.7887	**0.7931**
RV systolic dysfunction	580	0.8319	0.8300	0.8385	0.8540	**0.8568**
Pericardial effusion	103	0.7753	0.7745	0.7231	0.7616	**0.7934**
TR maximum velocity $\geq$ 3.2 m/s	1,142	0.7700	0.7728	0.7809	0.7909	**0.7939**
Composite SHD	4,136	0.8032	0.8116	0.8135	0.8220	**0.8248**

Note. Values are point estimates on the same 7,385 test patients. The EchoNext-Mini Model was retrained on MIMIC-IV-Ext-SHD. Bold indicates the highest value for each output and metric. Positive patients, number of test patients with the outcome. Abbreviations: EN-Mini, EchoNext-Mini Model; Frozen, ECGFounder-Frozen; FT, ECGFounder-FT; LV, left ventricular; LVEF, left ventricular ejection fraction; RV, right ventricular; SHD, structural heart disease; TR, tricuspid regurgitation.

Table S4 Label-wise AUPRC and AUROC of the EchoNext-Mini Model, ECGFounder-Frozen, ECGFounder-FT, DGRNet, and GS-ECG on the EchoNext-Mini Dataset test set.

Label	Positive patients	EN-Mini	Frozen	FT	DGRNet	GS-ECG
a. AUPRC						
LVEF $\leq$ 45%	962	0.5989	0.6792	0.6926	0.7092	**0.7218**
LV wall thickness $\geq$ 1.3 cm	1,061	0.3715	0.4114	0.4169	0.4334	**0.4401**
Aortic stenosis	286	0.2478	0.2671	0.2682	0.3115	**0.3202**
Aortic regurgitation	66	0.0325	0.0245	0.0253	**0.0384**	0.0331
Mitral regurgitation	337	0.2220	0.2657	0.2720	0.2637	**0.2807**
Tricuspid regurgitation	353	0.2924	0.3024	0.3055	0.3452	**0.3466**
Pulmonary regurgitation	20	**0.1165**	0.0362	0.0310	0.0764	0.0774
RV systolic dysfunction	419	0.4280	0.4523	0.4562	0.4742	**0.4832**
Pericardial effusion	69	**0.0725**	0.0335	0.0357	0.0458	0.0405
PASP $\geq$ 45 mmHg	699	0.3573	0.3462	0.3486	0.3801	**0.3865**
TR maximum velocity $\geq$ 3.2 m/s	375	0.2105	0.2067	0.2172	0.2395	**0.2425**
Composite SHD	2,318	0.7888	0.7988	0.8033	0.8033	**0.8112**
b. AUROC						
LVEF $\leq$ 45%	962	0.8517	0.8840	0.8886	0.8939	**0.8976**
LV wall thickness $\geq$ 1.3 cm	1,061	0.7343	0.7573	0.7605	0.7680	**0.7731**
Aortic stenosis	286	0.8587	0.8359	0.8397	**0.8772**	0.8768
Aortic regurgitation	66	0.7396	0.7049	0.7021	**0.7564**	0.7516
Mitral regurgitation	337	0.8061	0.8278	0.8323	0.8257	**0.8362**
Tricuspid regurgitation	353	0.8334	0.8341	0.8371	0.8517	**0.8539**
Pulmonary regurgitation	20	0.8317	0.8304	0.8334	0.8357	**0.8491**
RV systolic dysfunction	419	0.8663	0.8781	0.8812	0.8881	**0.8901**
Pericardial effusion	69	**0.7662**	0.7289	0.7253	0.7405	0.7472
PASP $\geq$ 45 mmHg	699	0.7703	0.7714	0.7761	0.7927	**0.7949**
TR maximum velocity $\geq$ 3.2 m/s	375	0.7536	0.7606	0.7691	0.7894	**0.7905**
Composite SHD	2,318	0.8197	0.8279	0.8304	0.8360	**0.8394**

Note. Values are point estimates on the same 5,442 test patients. The EchoNext-Mini Model used the authors' released checkpoint. Bold indicates the highest value for each output and metric. Positive patients, number of test patients with the outcome. Abbreviations: EN-Mini, EchoNext-Mini Model; Frozen, ECGFounder-Frozen; FT, ECGFounder-FT; LV, left ventricular; LVEF, left ventricular ejection fraction; PASP, pulmonary artery systolic pressure; RV, right ventricular; SHD, structural heart disease; TR, tricuspid regurgitation.

Table S5 Label-wise AUROC and AUPRC for GS-ECG and the EchoNext-Mini Model on the MIMIC-IV-Ext-SHD test set.

Label	EchoNext-Mini Model AUROC	GS-ECG AUROC	EchoNext-Mini Model AUPRC	GS-ECG AUPRC
LVEF $\leq 45\%$	0.8310 [0.8208–0.8416]	**0.8811** [0.8728–0.8904]	0.5709 [0.5463–0.5980]	**0.6946** [0.6707–0.7178]
LV wall thickness ≥ 1.3 cm	0.6758 [0.6633–0.6895]	**0.7303** [0.7190–0.7418]	0.4120 [0.3928–0.4322]	**0.5121** [0.4915–0.5357]
Aortic stenosis	0.8141 [0.7888–0.8387]	**0.8444** [0.8203–0.8669]	0.1242 [0.0956–0.1663]	**0.1503** [0.1171–0.1990]
Aortic regurgitation	0.7904 [0.6951–0.8661]	**0.8348** [0.7456–0.9162]	0.0315 [0.0129–0.0765]	**0.0913** [0.0212–0.2118]
Mitral regurgitation	0.8095 [0.7859–0.8332]	**0.8427** [0.8223–0.8615]	0.1353 [0.1135–0.1660]	**0.1684** [0.1389–0.2082]
Tricuspid regurgitation	0.8313 [0.8127–0.8479]	**0.8584** [0.8411–0.8742]	0.2108 [0.1795–0.2499]	**0.2682** [0.2272–0.3137]
Pulmonary regurgitation	0.7864 [0.7514–0.8201]	**0.7931** [0.7560–0.8245]	0.0741 [0.0575–0.1015]	**0.0830** [0.0623–0.1189]
RV systolic dysfunction	0.8319 [0.8150–0.8471]	**0.8568** [0.8408–0.8715]	0.3124 [0.2775–0.3517]	**0.3746** [0.3349–0.4182]
Pericardial effusion	0.7753 [0.7391–0.8112]	**0.7934** [0.7519–0.8337]	0.0516 [0.0343–0.0951]	**0.0810** [0.0500–0.1300]
TR maximum velocity ≥ 3.2 m/s	0.7700 [0.7560–0.7842]	**0.7939** [0.7804–0.8066]	0.3567 [0.3312–0.3864]	**0.4124** [0.3856–0.4427]
Composite SHD	0.8032 [0.7932–0.8128]	**0.8248** [0.8154–0.8339]	0.8306 [0.8183–0.8418]	**0.8524** [0.8410–0.8624]

Note. Values are point estimates with 95% CIs from 1,000 patient-level bootstrap resamples on the same 7,385 test patients. The EchoNext-Mini Model was retrained on MIMIC-IV-Ext-SHD. Bold indicates the higher value for each label and metric. Abbreviations: LV, left ventricular; LVEF, left ventricular ejection fraction; RV, right ventricular; TR, tricuspid regurgitation.

a. Thresholds and confusion-matrix counts

Table S6 Validation-selected operating points for GS-ECG composite SHD prediction on the MIMIC-IV-Ext-SHD test set.

Operating point	Score cutoff	Referred, n (%)	TP	FP	TN	FN
NPV ≥90% target	0.127	6,739 (91.3)	4,090	2,649	600	46
Sensitivity ≥90% target	0.273	5,455 (73.9)	3,778	1,677	1,572	358
Youden index	0.489	3,522 (47.7)	2,843	679	2,570	1,293
30% referral target	0.653	2,227 (30.2)	1,955	272	2,977	2,181
PPV ≥90% target	0.724	1,724 (23.3)	1,567	157	3,092	2,569

b. MIMIC-IV-Ext-SHD test performance

Operating point	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	NNE
NPV ≥90% target	98.9 [98.6–99.2]	18.5 [17.2–19.8]	60.7 [59.5–61.8]	92.9 [90.9–94.9]	1.65 [1.62–1.68]
Sensitivity ≥90% target	91.3 [90.5–92.2]	48.4 [46.7–49.9]	69.3 [68.1–70.4]	81.5 [79.8–83.1]	1.44 [1.42–1.47]
Youden index	68.7 [67.4–70.1]	79.1 [77.7–80.4]	80.7 [79.4–82.0]	66.5 [65.1–68.1]	1.24 [1.22–1.26]
30% referral target	47.3 [45.7–48.7]	91.6 [90.7–92.6]	87.8 [86.4–89.1]	57.7 [56.5–59.0]	1.14 [1.12–1.16]
PPV ≥90% target	37.9 [36.3–39.4]	95.2 [94.4–95.9]	90.9 [89.5–92.2]	54.6 [53.4–55.9]	1.10 [1.08–1.12]

Note. Cutoffs were selected on validation data and applied unchanged to the test predictions for 7,385 patients, 4,136 with composite SHD. Cutoffs are displayed to three decimal places; all analyses used their full precision. Panel b reports point estimates with 95% CIs in brackets from 1,000 patient-level bootstrap resamples with thresholds held fixed. NNE = referred patients/TP. Abbreviations: FN, false negative; FP, false positive; NNE, number needed to echo; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.

Table S7 MIMIC-IV-Ext-SHD ECG eligibility and quality-control criteria, with the number of ECGs excluded by each.

Item	Criterion or action	Failing ECGs	Newly excluded
Acquisition time	Valid acquisition time.	0	0
Sex and age	Sex and age at ECG available.	4,489	4,489
Adult age	Patients aged at least 18 years. MIMIC-IV records an anchor age of 91 for patients older than 89 years; age at ECG, derived from the anchor age and year, was capped at 90 years.	45	45
Automated measurements	Required machine measurements available.	1	1
RR interval	Retained within 200–3000 ms.	1,541	1,531
Ventricular rate	Retained within 20–300 bpm.	1,541	0
QRS duration	Retained within 20–300 ms.	1,579	62
QTc interval	Retained within 200–800 ms.	1,537	18
Pacing	Excluded when machine report text indicated ventricular pacing or a pacemaker rhythm.	19,383	19,288
Waveform format	Diagnostic 12-lead ECG with 5,000 samples per lead at 500 Hz and matching WFDB header and signal files.	0	0
PR interval	Derived from P-wave onset to QRS onset; treated as missing when timing was invalid or the value was outside 40–400 ms.	–	–
Total			25,434

Note. Criteria are listed in the order applied to the 800,035 MIMIC-IV-ECG records from 161,352 patients, of which 774,601 records from 158,872 patients remained. An ECG can fail several criteria, so only the newly excluded counts sum to the total. Measurements outside the specified ranges were excluded unless the criterion explicitly states that the value was retained as missing; PR intervals set to missing did not exclude records. Abbreviations: ECG, electrocardiogram; QTc, corrected QT interval; RR, R–R interval; WFDB, Waveform Database.

Table S8: MIMIC-IV-Ext-SHD echocardiography-derived outcome definitions.

Outcome	Positive criterion	MIMIC-IV-ECHO field(s)
LVEF $\leq 45\%$	Any available conventional, biplane, three-dimensional, or resting LVEF measurement at or below 45%.	`lvef`, `biplane_lvef`, `lvef_3d`, `rest_lvef`
LV wall thickness ≥ 1.3 cm	Septal or inferolateral/posterior wall thickness at or above 1.3 cm. Values on a millimetre scale were converted to centimetres before thresholding.	`septal_thickness`, `inf_lat_thickness`
Aortic stenosis	Structured severity indicating moderate, moderate-to-severe, or severe aortic stenosis.	`aortic_stenosis`
Aortic regurgitation	Structured severity indicating moderate, moderate-to-severe, or severe aortic regurgitation.	`aortic_regurg`
Mitral regurgitation	Structured severity indicating moderate, moderate-to-severe, or severe mitral regurgitation.	`mitral_regurg`
Tricuspid regurgitation	Structured severity indicating moderate, moderate-to-severe, or severe tricuspid regurgitation.	`tricuspid_regurg`
Pulmonary regurgitation	Structured severity indicating significant, moderate, moderate-to-severe, or severe pulmonary regurgitation; the dataset-specific *significant* category was retained as positive.	`pulm_regurg`
RV systolic dysfunction	Structured RV function indicating moderate dysfunction, severe dysfunction, or depressed function.	`rv_function`
Pericardial effusion	Structured severity indicating moderate or large pericardial effusion.	`pericardial_effusion`
TR maximum velocity $\geq$ 3.2 m/s	TR maximum velocity $\geq$ 3.2 m/s. When velocity was unavailable, TR pressure gradient $\geq$ 40.96 mmHg was used as the equivalent criterion.	`tr_velocity`; `tr_mmhg` when `tr_velocity` was unavailable
Composite SHD	At least one of the 10 component labels was positive.	Component labels

Note. Fields are values of the `measurement` column in the MIMIC-IV-ECHO structured-measurement table; only TTE studies were used. TTE studies with a prosthetic or repaired valve were excluded, identified from the numbered `avr_structure`, `mvr_structure`, `pvr_structure`, and `tvr_structure` fields and from text matches such as prosthesis, replacement, repair, annuloplasty, TAVR, or valve clip. The PASP $\geq$45 mmHg label used in EchoNext was not included because MIMIC-IV-ECHO records TR velocity and gradient but not PASP; TR maximum velocity $\geq$3.2 m/s, a gradient of about 41 mmHg, corresponds to a PASP of about 46 mmHg with a right atrial pressure of 5 mmHg. Abbreviations: LV, left ventricular; LVEF, left ventricular ejection fraction; PASP, pulmonary artery systolic pressure; RV, right ventricular; TAVR, transcatheter aortic valve replacement; TR, tricuspid regurgitation; TTE, transthoracic echocardiography.

Supplementary Figures

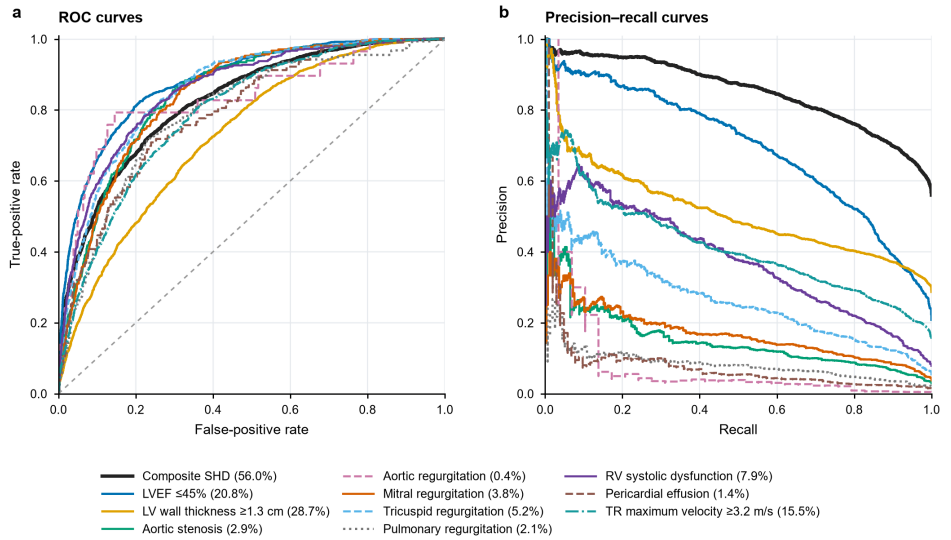


Fig. S1 Label-wise receiver operating characteristic (ROC) and precision–recall curves for the 11 GS-ECG outputs on the MIMIC-IV-Ext-SHD test set. Panel a plots true-positive rate against false-positive rate; the diagonal denotes chance ranking. Panel b plots empirical precision against recall. Legend values give test-set prevalence, which equals the precision of random ranking and the end point of each precision–recall curve at recall 1. Curves use GS-ECG predictions for all 7,385 test patients, with one ECG per patient.

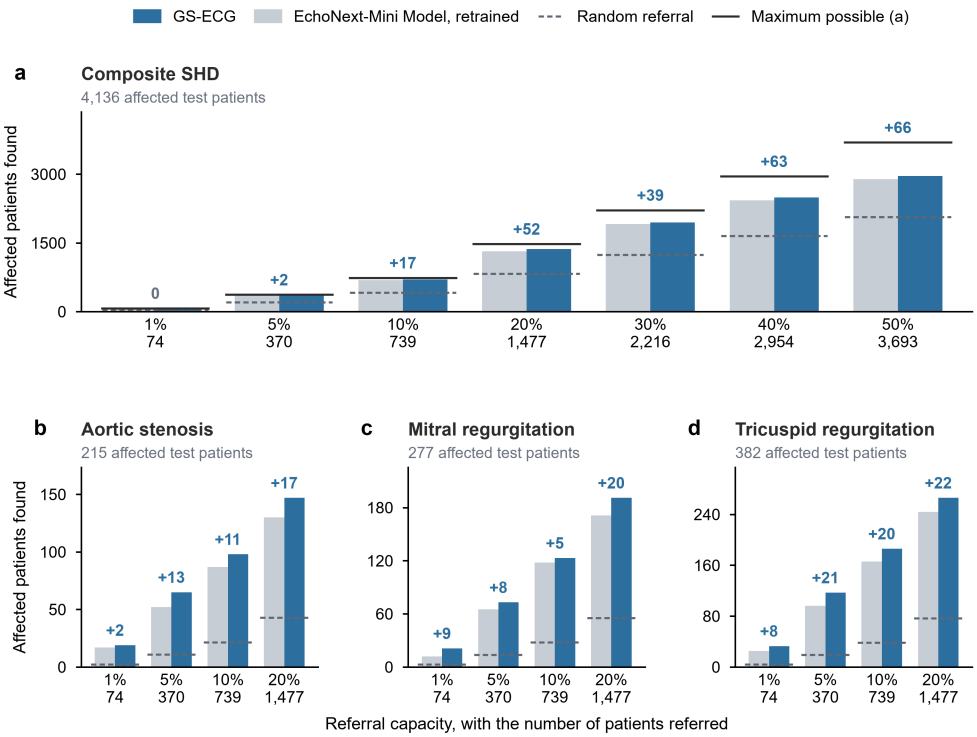


Fig. S2 Patients found by GS-ECG and the retrained EchoNext-Mini Model at fixed referral capacities on MIMIC-IV-Ext-SHD. Panel a shows composite SHD at seven capacities from 1% to 50%, and panels b–d show aortic stenosis, mitral regurgitation, and tricuspid regurgitation at four capacities from 1% to 20%. At each capacity, each model referred that percentage of the 7,385 test patients, rounded up and ranked separately for each output; the number referred is given below each capacity, and annotations give GS-ECG minus the EchoNext-Mini Model. Dashed lines give the number expected under random referral, and solid lines in panel a the most patients that could be found at each capacity. For the three valvular abnormalities, GS-ECG also found more affected patients at every intermediate capacity in one-percentage-point steps. Y-axes are scaled separately.