

Wearable Single-Lead ECG Detects Fine-Grained Structural Heart Disease Through Echo-Report Supervision

Shenda Hong

hongshenda@pku.edu.cn

Peking University <https://orcid.org/0000-0001-7521-5127>

Chenyang He

Peking University

Qinghao Zhao

Peking University People's Hospital

Shun Huang

Peking University

Jun Li

Peking University

Gongzheng Tang

Peking University

Hao Zhang

The Second Hospital of Tianjin Medical University

Tong Liu

Second Hospital of Tianjin Medical University

Zhengkai Xue

The Second Hospital of Tianjin Medical University

Jian Liu

Peking University People's Hospital

Kangyin Chen

Second Hospital of Tianjin Medical University

Cheng Ding

Nanjing University of Aeronautics and Astronautics

Article

Keywords: Structural Heart Disease, Multimodal Fusion, Single-lead Electrocardiogram

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11313095/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Wearable Single-Lead ECG Detects Fine-Grained Structural Heart Disease Through Echo-Report Supervision

Chenyang He^{1, #}, Qinghao Zhao^{2, #}, Shun Huang¹, Jun Li^{1, 3}, Gongzheng Tang^{1, 3}, Hao Zhang⁴, Tong Liu⁴, Zhengkai Xue⁴, Jian Liu², Kangyin Chen^{4, *}, Cheng Ding^{5, *}, and Shenda Hong^{1, 3, 6, 7, *, }

¹National Institute of Health Data Science, Peking University, Beijing, China

²Department of Cardiology, Peking University People's Hospital, Beijing, China

³Institute of Medical Technology, Peking University Health Science Center, Beijing, China

⁴Tianjin Key Laboratory of Ionic-Molecular Function of Cardiovascular Disease, Department of Cardiology, Tianjin Institute of Cardiology, The Second Hospital of Tianjin Medical University, Tianjin, China

⁵Nanjing University of Aeronautics and Astronautics, Nanjing, China

⁶Institute for Artificial Intelligence, Peking University, Beijing, China

⁷State Key Laboratory of Vascular Homeostasis and Remodeling, NHC Key Laboratory of Cardiovascular Molecular Biology and Regulatory Peptides, Peking University, Beijing, China

#These authors contributed equally

*Correspondence: hongshenda@pku.edu.cn, chengding@nuaa.edu.cn, chen kangyin@vip.126.com

ABSTRACT

Structural heart disease (SHD) is a primary driver of heart failure and cardiovascular mortality, yet early detection remains constrained by the limited accessibility of echocardiography. While single-lead electrocardiogram (ECG) is ubiquitous through wearables, existing AI screening models often depend on 12-lead inputs, generalize poorly across institutions, or require massive, condition-specific labeled datasets. Recent work has demonstrated the feasibility of contrastive pre-training between single-lead ECGs and echocardiography reports within a single health system. Here, we present AnyECG-Echo, a framework that advance this paradigm toward clinical translation through three key developments: (1) evaluation in a geographically independent external cohort ($n = 16,621$); (2) diagnostic coverage of 13 fine-grained SHD subtypes spanning myocardial, chamber, valvular, and great-vessel pathologies; and (3) dual-axis mechanistic interpretability combining electrophysiology-grounded Shapley attribution with emergent correlations to quantitative measurements. Across validation cohorts totaling $n = 25,222$, the model demonstrated high AUROC for high-impact subtypes, including reduced left ventricular systolic function (AUROC 0.866–0.924), global heart enlargement (0.877–0.931), and mitral stenosis (0.836–0.906). Furthermore, we successfully validated the alignment of model outputs with established medical physiological traits, thereby enhancing interpretability. Notably, we discovered that AnyECG-Echo's outputs function as physiologically grounded digital biomarkers that accurately track objective metrics such as *LVEF* and myocardial wall thickness. These findings prove that wearable single-lead ECGs can effectively detect fine-grained structural heart disease, offering a practical solution for population-scale screening.

KEYWORDS

Structural Heart Disease, Multimodal Fusion, Single-lead Electrocardiogram

INTRODUCTION

Structural heart disease (SHD)^{1,2}, encompassing myocardial³, chamber⁴, valvular⁵, and great-vessel pathologies⁶, is a primary driver of cardiovascular mortality worldwide. A defining challenge of SHD is its prolonged, clinically silent prodrome⁷. By the time overt symptoms manifest, irreversible cardiac remodeling has often occurred, rendering disease-modifying therapies less effective. Consequently, improving population-level outcomes hinges on early detection through scalable screening infrastructure⁸. While echocardiography remains the diagnostic reference standard⁹, its reliance on specialized equipment and certified operators creates a profound clinical bottleneck¹⁰. These constraints concentrate diagnostic capacity within tertiary centers, leaving much of the early SHD burden invisible to health systems until structural compromise is advanced¹¹.

The electrocardiogram (ECG) represents the only cardiac modality available at population scale. It is inexpensive, non-invasive, and increasingly accessible through consumer wearables^{12–15} (Fig. 1a). However, three major barriers impede its clinical translation for SHD screening¹⁶. First, most high-performance models are optimized for 12-lead configurations, incompatible with the single-lead sensors used in consumer wearables^{17,18}. Second, variations in hardware and demographics across healthcare sites frequently compromise model generalizability. Third, the prevailing supervised-learning paradigm relies on massive, manually labeled datasets for each specific condition, which are expensive to assemble and lack diagnostic granularity¹⁹.

Recent work has begun exploring contrastive pre-training between single-lead ECGs and echocardiography reports as a means to bypass manual labeling entirely²⁰. While these efforts have established the feasibility of this paradigm within single health systems, three critical gaps remain before clinical deployment: the absence of evidence for cross-institutional generalization across different geographies and clinical practices; limited diagnostic granularity, with current frameworks targeting only a small set of coarse subtypes (e.g., left ventricular systolic dysfunction, composite SHD); and the absence of mechanistic verification that learned representations reflect cardiac physiology rather than center-specific spurious correlations. To address these challenges, we developed AnyECG-Echo, a multimodal contrastive learning framework that repurposes routinely generated, physician-authored echocardiography reports as an automated and high-fidelity supervisory signal. Using single-lead ECGs, the framework targets 13 echocardiography-derived SHD phenotypes across five domains (Fig. 1b). The ECG encoder is jointly pre-trained through ECG–report alignment (Fig. 1c) and ECG-only self-supervision (Fig. 1d), followed by supervised training of a linear classification head with the encoder frozen (Fig. 1e). We systematically validated this framework across three distinct validation tiers: an internal testing set ($n = 4,078$), a temporal test set ($n = 4,523$), and a geographically independent external test set ($n = 16,621$; Fig. 1).

Crucially, we interrogated the model’s clinical validity through two complementary lenses: first, by aligning ECG-segment attribution with expert-defined electrophysiological signatures, and second, by demonstrating that the model’s representations exhibit correlations with raw quantitative measurements never encountered during training. These findings suggest that the model’s output functions as a digital biomarker of SHD.

Collectively, these analyses demonstrate that AnyECG-Echo provides a scalable, mechanistically grounded foundation for SHD pre-screening. By bridging the gap between ubiquitous sensor technology and definitive diagnostics, this framework enables expert-level assessment in out-of-hospital environments, remote communities, and primary care settings where echocardiography remains inaccessible.

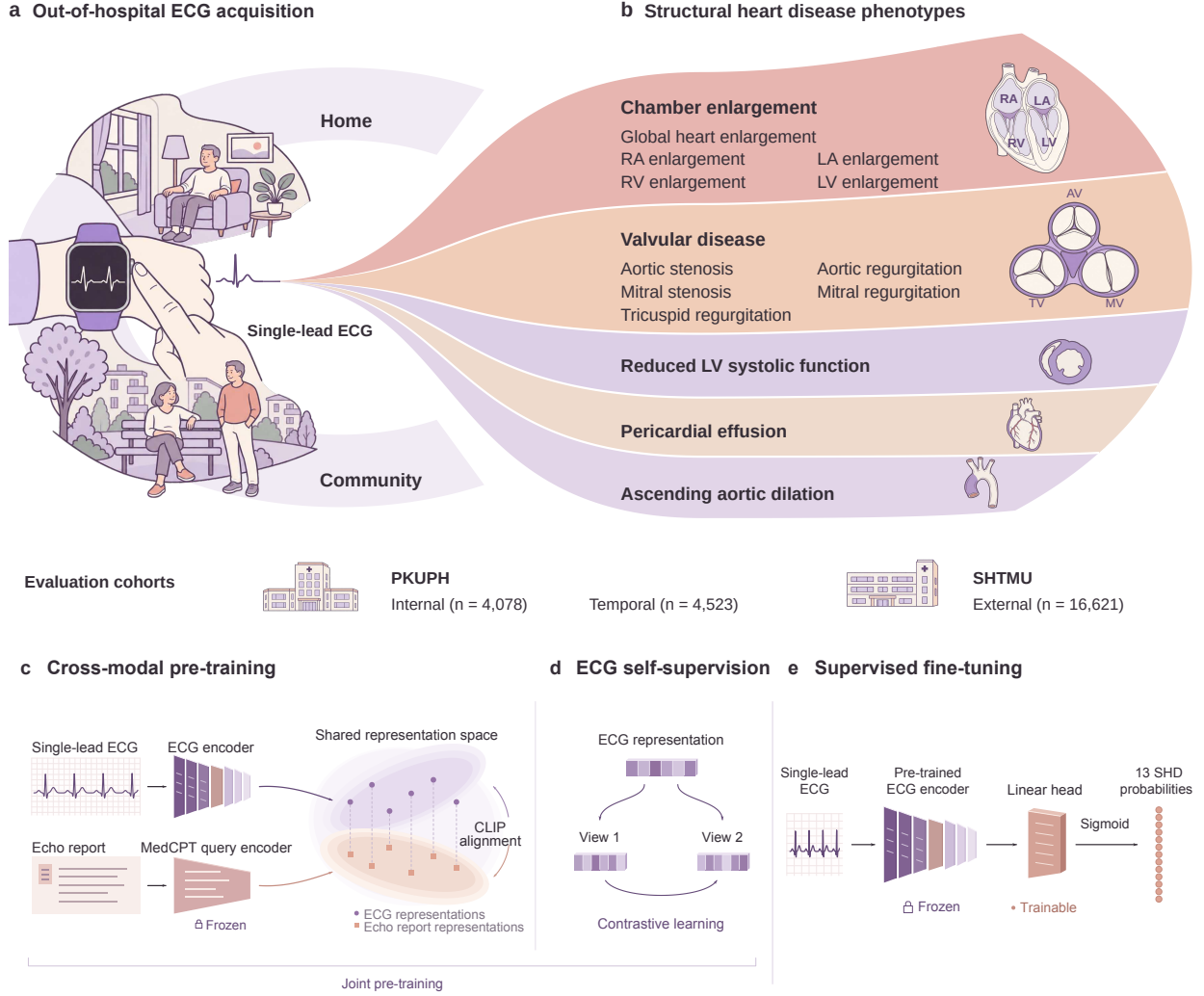


Figure 1: Single-lead ECG assessment of structural heart disease across anatomical domains. **a**, Convenient single-lead ECG acquisition outside the hospital, including home and community settings. **b**, Thirteen echocardiography-derived phenotypes grouped into five domains: chamber enlargement, valvular disease, reduced left ventricular systolic function, pericardial effusion and ascending aortic dilation. Coloured bands denote phenotype groups. The three evaluation cohorts comprise: internal testing ($n = 4,078$) and temporal validation ($n = 4,523$) at Peking University People's Hospital (PKUPH), Beijing, and external validation ($n = 16,621$) at the Second Hospital of Tianjin Medical University (SHTMU), Tianjin. All evaluations were retrospective; n denotes patients. Phenotype definitions are provided in Methods. **c**, Cross-modal contrastive pre-training. Single-lead ECG representations are aligned with paired echocardiography-report representations from a frozen MedCPT query encoder using a symmetric CLIP-style objective. **d**, ECG-only self-supervision. Two views of the same ECG representation, generated using independent dropout and linear projections, are contrasted using a unidirectional InfoNCE-style objective. The cross-modal and ECG-only objectives jointly train the ECG encoder. **e**, Supervised fine-tuning. The pre-trained ECG encoder remains frozen while a linear classification head is trained using binary cross-entropy. Elementwise sigmoid activation yields probabilities for the 13 SHD phenotypes. Architecture and training details are provided in the AnyECG-Echo architecture subsection of Methods. ECG, electrocardiogram; SHD, structural heart disease; RA, right atrium; LA, left atrium; RV, right ventricle; LV, left ventricle; AV, aortic valve; TV, tricuspid valve; MV, mitral valve.

RESULTS

AnyECG-Echo achieves high-fidelity SHD screening with single-lead inputs

AnyECG-Echo bridges the gap between accessible electrophysiology and definitive structural imaging by aligning single-lead ECG (Lead I) embeddings with the semantic narratives of physician-authored echocardiography reports. By leveraging these reports as a dense, automated supervisory signal, the framework internalizes the structural implications of electrical signals across 13 fine-grained subtypes, effectively eliminating the requirement for manual expert annotation.

Across all three validation tiers, AnyECG-Echo demonstrated reliable discriminative capacity for the 13 fine-grained SHD subtypes (Table 1). The majority of targeted subtypes achieved strong discrimination ($\text{AUROC} \geq 0.80$). While Aortic regurgitation, Ascending aortic dilation, and LA enlargement fell within a weaker range ($\text{AUROC} < 0.65$) in the external test set, the vast majority of categories across all cohorts maintained acceptable to strong predictive power ($\text{AUROC} \geq 0.65$), confirming the framework’s screening utility.

Methodological and per-label performance comparison with contemporary AI-ECG models

To contextualize both the methodological contributions and diagnostic performance of AnyECG-Echo, we conducted a systematic comparison against state-of-the-art AI-ECG models for SHD detection published between 2025 and 2026 (Table 2). The upper panel summarizes each model across five methodological dimensions; the lower panel presents per-label AUROC values on each model’s respective test cohort. These comparisons are cross-cohort rather than head-to-head.

A central finding of this comparison is the data efficiency of AnyECG-Echo. With only 37,389 ECG–echocardiography pairs—5 to 33 times fewer than its comparators—AnyECG-Echo achieves competitive or superior discrimination across the conditions where direct comparison is possible. For reduced LV systolic function, AnyECG-Echo attains an internal AUROC of 0.924 using single-lead input, compared with 0.904 for the 12-lead EchoNext trained on over 1.2 million pairs.

Equally important is the diagnostic breadth that this efficiency enables. While EchoNext and EchoNext-Mini cover 11–12 categories and Wearable-Echo-FM targets only 3, AnyECG-Echo resolves 13 fine-grained subtypes from a single-lead waveform. Seven of these—including global heart enlargement, chamber-specific enlargements (LA, LV, RA, RV), ascending aortic dilation, and mitral stenosis—have no published AI-ECG comparator to date, representing unique diagnostic coverage.

Beyond performance, the comparison reveals complementary gaps in existing approaches. EchoNext¹⁷ and AI-ECG VHD¹⁹ demonstrate strong multi-centre performance but remain dependent on clinical 12-lead ECG acquisition, limiting their deployment in home-monitoring or point-of-care settings. Wearable-Echo-FM²⁰ pioneered single-lead contrastive learning but was validated only within a single health system and without mechanistic interpretability. AnyECG-Echo is, to our knowledge, the only framework that simultaneously achieves native single-lead compatibility, geographically independent external validation, fine-grained multi-subtype coverage, and dual-axis physiological interpretability—demonstrating that a small, well-structured dataset paired with report-supervised contrastive learning can match or exceed the diagnostic utility of models trained on orders-of-magnitude more data.

Table 1: Model Performance Evaluation Across Different Test Sets

Field ID	AUROC (95% CI)	Sensitivity	Specificity	Brier Score
Internal Test Set from PKUPH				
Global heart enlargement	0.931 [0.880, 0.971]	0.810	0.933	0.005
Aortic regurgitation	0.928 [0.841, 0.979]	0.850	0.899	0.004
Reduced LV systolic fn.	0.924 [0.895, 0.950]	0.821	0.905	0.021
RA enlargement	0.914 [0.871, 0.952]	0.868	0.837	0.008
Tricuspid regurgitation	0.907 [0.838, 0.962]	0.759	0.959	0.006
RV enlargement	0.882 [0.833, 0.930]	0.738	0.837	0.009
LV enlargement	0.868 [0.843, 0.895]	0.785	0.811	0.038
Mitral stenosis	0.836 [0.685, 0.971]	0.556	0.956	0.002
Aortic stenosis	0.835 [0.734, 0.917]	0.444	0.970	0.004
Mitral regurgitation	0.808 [0.739, 0.870]	0.707	0.814	0.009
Pericardial effusion	0.728 [0.676, 0.777]	0.785	0.540	0.021
Ascending aortic dil.	0.685 [0.658, 0.713]	0.640	0.643	0.073
LA enlargement	0.665 [0.648, 0.680]	0.740	0.510	0.201
Temporal Test Set from PKUPH				
Global heart enlargement	0.929 [0.883, 0.963]	0.730	0.930	0.006
Reduced LV systolic fn.	0.920 [0.901, 0.936]	0.893	0.767	0.023
RA enlargement	0.913 [0.887, 0.936]	0.863	0.824	0.014
Tricuspid regurgitation	0.893 [0.837, 0.936]	0.733	0.876	0.008
Mitral stenosis	0.889 [0.735, 0.977]	0.636	0.952	0.002
Mitral regurgitation	0.842 [0.786, 0.889]	0.600	0.895	0.009
LV enlargement	0.838 [0.812, 0.861]	0.741	0.767	0.041
Aortic regurgitation	0.807 [0.721, 0.886]	0.517	0.911	0.005
RV enlargement	0.771 [0.689, 0.846]	0.544	0.888	0.009
Aortic stenosis	0.756 [0.616, 0.883]	0.214	0.972	0.003
Pericardial effusion	0.691 [0.646, 0.734]	0.543	0.764	0.028
Ascending aortic dil.	0.677 [0.657, 0.700]	0.633	0.629	0.089
LA enlargement	0.657 [0.641, 0.673]	0.543	0.683	0.203
External Test Set from SHTMU				
Mitral stenosis	0.906 [0.837, 0.951]	0.682	0.922	0.001
Global heart enlargement	0.877 [0.860, 0.894]	0.653	0.883	0.018
Reduced LV systolic fn.	0.866 [0.858, 0.874]	0.803	0.776	0.074
Pericardial effusion	0.832 [0.790, 0.870]	0.811	0.724	0.009
Tricuspid regurgitation	0.820 [0.800, 0.837]	0.664	0.833	0.031
Mitral regurgitation	0.806 [0.788, 0.823]	0.783	0.725	0.031
RA enlargement	0.803 [0.762, 0.843]	0.732	0.744	0.008
RV enlargement	0.748 [0.670, 0.819]	0.622	0.776	0.005
Aortic stenosis	0.714 [0.576, 0.832]	0.182	0.940	0.001
LV enlargement	0.698 [0.605, 0.777]	0.815	0.506	0.022
Aortic regurgitation	0.639 [0.609, 0.669]	0.343	0.805	0.015
Ascending aortic dil.	0.635 [0.622, 0.648]	0.648	0.545	0.094
LA enlargement	0.610 [0.602, 0.618]	0.617	0.548	0.283

LV, left ventricular; LA, left atrial; RV, right ventricular; RA, right atrial; dil., dilation; fn., function.

Table 2: Per-label AUROC comparison between AnyECG-Echo and contemporary AI-ECG models, with summary of methodological characteristics. For AnyECG-Echo, the internal cohort corresponds to the PKUPH internal test, and the external cohort corresponds to the SHTMU external validation. All AUROC values represent performance on each model’s respective test cohort. Dashes indicate that the condition was not evaluated by the corresponding model.

	AnyECG-Echo (Ours)		EchoNext ¹⁷	EchoNext-Mini ¹⁸	AI-ECG VHD ¹⁹		Wearable-Echo-FM ²⁰
	Internal	External	Multi-centre	Multi-centre	Internal	External	Internal
Model characteristics							
Input type	1-lead		12-lead	12-lead	12-lead		1-lead
Dataset scale	37,389 pairs		1,245,273 pairs	100,000 pairs	400,882 pairs		194,551 pairs
Diagnostic granularity	13 subtypes		11 categories	12 categories	3 categories		3 categories
External validation	✓		✓	—	✓		×
Interpretability	✓		×	×	✓		×
Source	-		Nature 2025	NEJM AI 2026	NEJM AI 2025		EHJ DH 2026
Per-label AUROC							
Mitral stenosis	0.836[0.685, 0.971]	0.906[0.837, 0.951]	—	—	—	—	—
Global heart enlargement	0.931[0.880, 0.971]	0.877[0.860, 0.894]	—	—	—	—	—
Reduced LV systolic function	0.924[0.895, 0.950]	0.866[0.858, 0.874]	0.904[0.896, 0.912]	0.852[0.839, 0.865]	—	—	0.894[0.884, 0.903]
Pericardial effusion	0.728[0.676, 0.777]	0.832[0.790, 0.870]	0.799[0.761, 0.834]	0.766[0.712, 0.818]	—	—	—
Tricuspid regurgitation	0.907[0.838, 0.962]	0.820[0.800, 0.837]	—	0.833[0.812, 0.852]	0.910[0.905, 0.914]	0.805[0.795, 0.816]	—
Mitral regurgitation	0.808[0.739, 0.870]	0.806[0.788, 0.823]	0.855[0.841, 0.869]	0.806[0.782, 0.829]	0.893[0.889, 0.897]	0.777[0.767, 0.785]	—
RA enlargement	0.914[0.871, 0.952]	0.803[0.762, 0.843]	—	—	—	—	—
RV enlargement	0.882[0.833, 0.930]	0.748[0.670, 0.819]	—	—	—	—	—
Aortic stenosis	0.835[0.734, 0.917]	0.714[0.576, 0.832]	0.864[0.850, 0.877]	0.859[0.839, 0.878]	—	—	—
LV enlargement	0.868[0.843, 0.895]	0.698[0.605, 0.777]	—	—	—	—	—
Aortic regurgitation	0.928[0.841, 0.979]	0.639[0.609, 0.669]	0.777[0.749, 0.805]	0.740[0.688, 0.794]	0.845[0.838, 0.851]	0.751[0.731, 0.774]	—
Ascending aortic dilation	0.685[0.658, 0.713]	0.635[0.622, 0.648]	—	—	—	—	—
LA enlargement	0.665[0.648, 0.680]	0.610[0.602, 0.618]	—	—	—	—	—

Note: Internal and External validation cohorts vary by model study definition (e.g., EchoNext’s “Multi-centre” refers to its respective multi-center test dataset).

Table 3: **Subgroup analysis of model performance in the external cohort**, stratified by age group and sex. Values are reported as AUROC [95% confidence interval]. Results for remaining subtypes are provided in *Supplementary Table: Subgroup Analysis of Low-Prevalence SHD Subtypes*.

Subgroup	SHD (<i>n</i>) (pos/neg)	Reduced LV systolic function	LV enlargement	Pericardial effusion	Ascending aortic dilation	SHD (overall)
Age group (years)						
<50	142/280	0.912 [0.807, 0.987]	0.841 [0.753, 0.919]	0.773 [0.662, 0.875]	0.823 [0.686, 0.934]	0.683 [0.628, 0.736]
50–64	347/410	0.939 [0.909, 0.964]	0.892 [0.839, 0.935]	0.845 [0.746, 0.931]	0.617 [0.542, 0.690]	0.691 [0.652, 0.729]
65–74	342/248	0.961 [0.934, 0.982]	0.790 [0.702, 0.871]	0.773 [0.593, 0.922]	0.613 [0.535, 0.691]	0.641 [0.596, 0.683]
≥75	267/136	0.855 [0.685, 0.978]	0.860 [0.783, 0.922]	0.621 [0.449, 0.795]	0.635 [0.566, 0.707]	0.677 [0.622, 0.730]
Sex						
Male	858/571	0.921 [0.886, 0.952]	0.862 [0.830, 0.895]	0.773 [0.681, 0.859]	0.655 [0.612, 0.698]	0.708 [0.681, 0.733]
Female	753/999	0.920 [0.871, 0.961]	0.817 [0.749, 0.878]	0.708 [0.636, 0.772]	0.685 [0.634, 0.733]	0.697 [0.672, 0.721]

n = positive / negative ECG recordings per subgroup (SHD overall). SHD, structural heart disease. LV, Left ventricular.

Demographic subgroup analysis and age-related performance attenuation

We evaluated model robustness across age and sex strata in the PKUPH cohort (Table 3). Sex-stratified analysis revealed consistent performance between male and female patients, with overlapping confidence intervals confirming an absence of systematic gender-related bias. In contrast, model performance exhibited a monotonic decline across most categories as age increased from < 50 to ≥ 75 years. This attenuation likely stems from age-associated electrophysiological confounders, such as conduction delays and atrial fibrillation, which accumulate in older populations and obscure subtle structural signatures.

AnyECG-Echo modeling achieves consistency with canonical medical interpretations

To verify that AnyECG-Echo prioritizes clinically valid signals over center-specific shortcuts, we performed a hierarchical interpretability analysis that bridges population-level trends and individual-level evidence. We first compared population-level median beats between high-risk and low-risk groups to capture macroscopic morphological shifts. This population perspective was then cross-validated at the individual level using occlusion-based Shapley value attribution, specifically targeting the patients identified by the model as having the highest diagnostic risk. By integrating these two perspectives, we can evaluate whether the general pathological signatures learned by the model are consistently utilized to drive its most confident clinical predictions.

Specifically, six physiologically defined ECG segments were delineated using fixed temporal offsets relative to each detected R-peak: P wave (−250 to −100 ms), PR segment (−100 to −20 ms), QRS complex (−20 to +80 ms), ST segment (+80 to +200 ms), T wave (+200 to +360 ms), and terminal segment (+360 to +400 ms). For each segment, all corresponding intervals across every detected heartbeat in a given recording were simultaneously replaced with zero values. Because the input signal undergoes z-score normalization prior to inference, zero-fill replacement is equivalent to substituting the population-mean amplitude, serving as a physiologically neutral baseline. The attribution score for each segment was then defined as the decrease in predicted probability divided by the total masked duration in seconds (termed Shapley value density), enabling fair comparison across segments of different lengths. For each SHD subtype, this analysis was performed on the top-5 patients ranked by predicted probability (with a minimum threshold of 0.5), and scores were averaged across this cohort; the mean segment-level scores are shown in the circle matrix in Fig. 2b. Population-level median beats were extracted by segmenting individual heartbeats within 800 ms windows centered on R-peaks, excluding morphological outliers (Pearson $r < 0.8$ relative to the initial median template), and computing the element-wise median of the remaining waveforms.

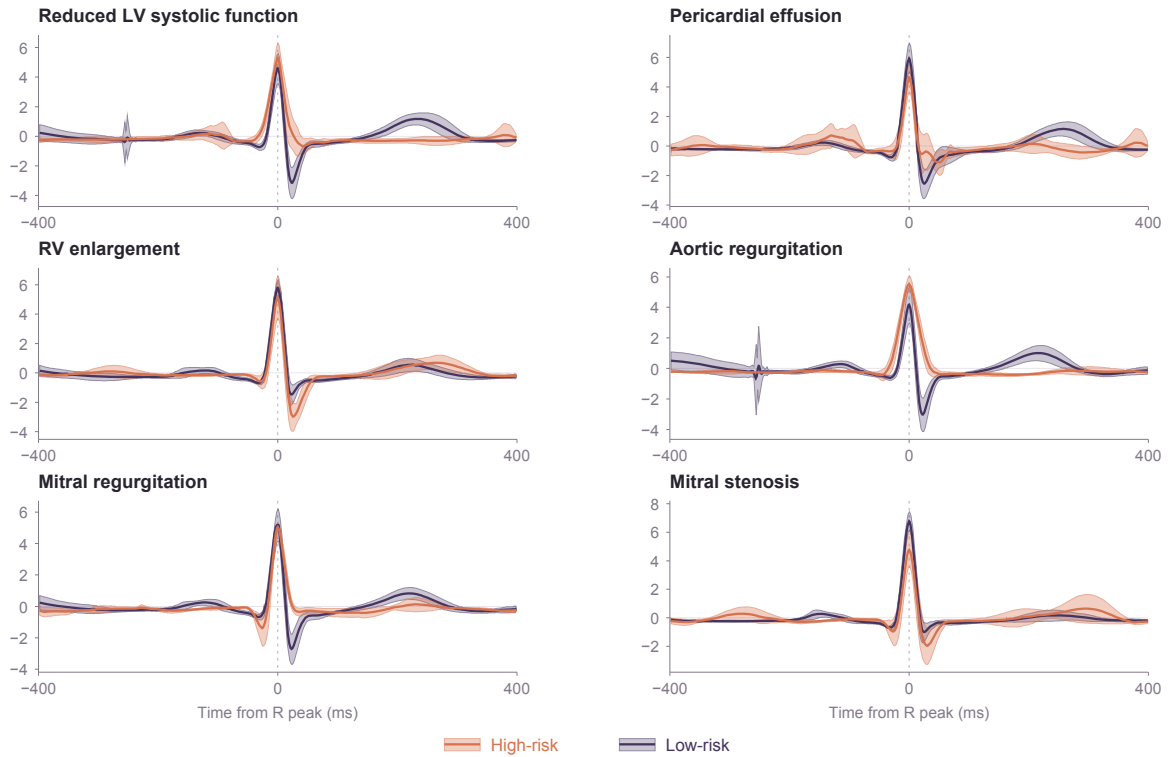
The primary mechanistic evidence is presented in Fig. 2. AnyECG-Echo effectively captures diagnostic signatures across the QRS complex, ST segment, and T wave, demonstrating high consistency between population-level morphological shifts and individual-level attributions.

For **Reduced LV systolic function**, although no direct canonical ECG correlate exists, our findings align with established explainability results for $LVEF^{21}$, with both median beats and Shapley values primarily localized to the *R*-wave. For **pericardial effusion**, we observed diminished QRS amplitude in median beats, with the highest Shapley attribution concentrated on the QRS complex. **RV enlargement** was characterized by deepened *S*-waves and reduced *R*-waves, with model attribution again localized to the QRS complex.

For valvular conditions, **aortic regurgitation** median beats exhibited increased *R*-wave voltage alongside *ST*-segment depression and *T*-wave inversion; accordingly, Shapley values were elevated across the QRS, *ST*, and *T* segments. **Mitral regurgitation** showed similar *ST*-*T* perturbations, with substantial attribution distributed across the QRS and *ST*-*T* segments. Finally, for **mitral stenosis**, median beats displayed widened *P*-waves, reduced *R*-waves, and deepened *S*-waves, with model attribution correctly prioritizing the *P*-wave and QRS complex.

Collectively, these findings demonstrate that AnyECG-Echo achieves fundamental consistency between deep-learning modeling and established medical interpretation. By grounding its predictions in canonical electrophysiology at both the population and individual scales, the framework ensures that its structural insights are mechanistically valid, transparent, and clinically trustworthy.

a Population morphology



b Top-5 mean attribution

	P	PR	QRS	ST	T	Term.
Reduced LV systolic function	• 0.027	• 0.116	• 0.237	• 0.002	• 0.031	• 0.009
Pericardial effusion	• 0.009	• 0.001	• 0.113	• -0.003	• 0.002	• 0.015
RV enlargement	• 0.112	• 0.057	• 0.155	• 0.099	• 0.112	• 0.020
Aortic regurgitation	• 0.040	• 0.005	• 0.085	• 0.076	• 0.043	• 0.095
Mitral regurgitation	• 0.113	• 0.054	• 0.207	• 0.118	• 0.097	• 0.479
Mitral stenosis	• 0.072	• 0.035	• 0.184	• 0.059	• 0.080	• 0.030

Figure 2: Population ECG morphology and segment-level occlusion attribution. **a**, Population median beats for high-risk (coral) and low-risk (purple) groups across six structural heart disease phenotypes, aligned to the R peak at 0 ms. Shaded regions indicate 95% confidence intervals. **b**, Mean segment-level attribution across the five patients with the highest predicted probability for each phenotype. For each patient, all intervals corresponding to a given segment across detected heartbeats were simultaneously replaced with zeros in the normalised input. Occlusion attribution density was calculated as the original minus the occluded prediction probability, divided by the total masked duration in seconds, and then averaged across the five patients. Columns represent the P wave, PR segment, QRS complex, ST segment, T wave and terminal segment (Term.). Circle area is proportional to the absolute mean attribution density on a common scale; numbers give the signed values in s^{-1} . LV, left ventricular; RV, right ventricular.

Predictive probabilities from AnyECG-Echo function as quantitative digital biomarkers

We calculated Spearman rank correlations between predicted disease probabilities and three gold-standard metrics: left ventricular ejection fraction ($LVEF$), interventricular septum thickness ($IVSs$), and LV posterior wall thickness ($LVPWS$).

As shown in Fig. 3, predicted probabilities for reduced LV systolic function and global heart enlargement demonstrated robust negative correlations with $LVEF$ ($\rho = -0.220$ and -0.195 , respectively; all $p < 0.001$). For pressure-overload conditions such as aortic stenosis and LV enlargement, predicted probabilities accurately tracked increases in wall thickness ($IVSs$ and $LVPWS$).

Crucially, for these examples, we observed in the quartile summaries in Fig. 3b–d that the predicted probability quartiles (Q1–Q4) maintain a monotonic correspondence with the echocardiographic parameters across different intervals. This confirms that AnyECG-Echo tracks the graded severity of systolic impairment, allowing the single-lead waveform to serve as a continuous digital biomarker of cardiac output.

We also identified a near-zero correlation with myocardial thickness ($IVSs$: $\rho = -0.01$; $LVPWS$: $\rho = +0.02$). This selective sensitivity ensures that the model discriminates between distinct disease mechanisms—fluid accumulation versus structural remodeling—rather than relying on non-specific signals of abnormality.

These findings establish AnyECG-Echo as a physiologically grounded digital biomarker that transcends binary diagnostic categories. This capability enables continuous, data-driven risk stratification in community and home-monitoring settings, identifying clinically significant structural decline long before the onset of symptomatic decompensation.

Data efficiency and asymptotic performance in downstream SHD detection

To evaluate the scaling laws of AnyECG-Echo, we fine-tuned the model on varying subsets of the development cohort (total patients $N = 12,726$), ranging from ultra-low sample sizes (100, 500, and 1,273 patients) to larger fractions (30%, 50%, 80%, and 100%) of the total dataset.

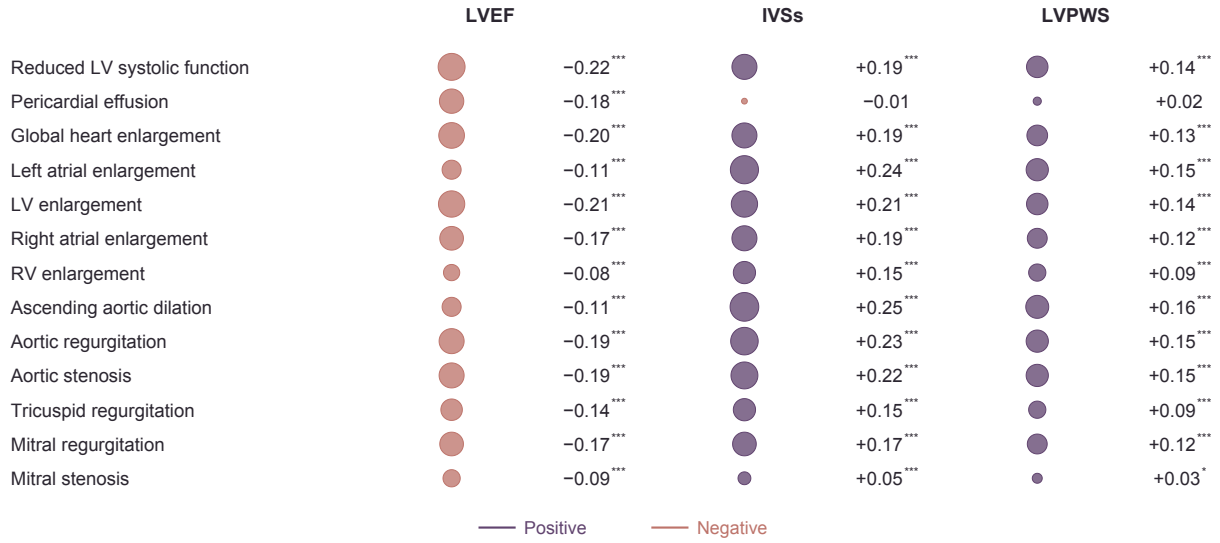
As shown in Fig. 4, AnyECG-Echo exhibited remarkable data efficiency across all cohorts. In the small-data regime, we observed a sharp and rapid increase in discriminative performance. Notably, with an ultra-minimal budget of only 100 labeled samples, the model’s performance already exceeded 75% of the level achieved by the full dataset.

As the sample size increased into the large-data regime, the performance followed a clear asymptotic trajectory. By the 50% sampling mark (representing 6,363 patients), AnyECG-Echo essentially reached its diagnostic ceiling, matching the performance attained by the full-dataset configuration. These results demonstrate that the multimodal pre-training successfully internalized a high-quality representation of cardiac physiology, enabling robust downstream adaptation with minimal supervised fine-tuning. This property is critical for deploying the framework in specialized clinical settings where labeled echocardiographic data are scarce.

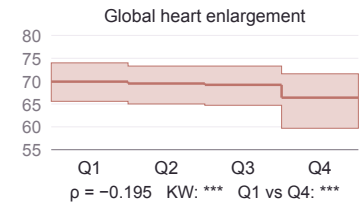
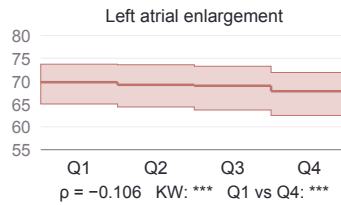
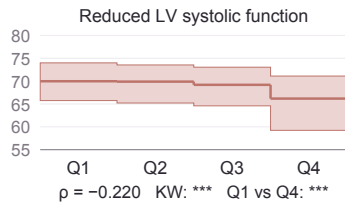
AnyECG-Echo provides net clinical benefit across the structural heart disease spectrum

To evaluate the clinical utility of AnyECG-Echo beyond discrimination metrics, we performed decision curve analysis for 12 SHD subtypes (Fig. 5a) and estimated screening efficiency

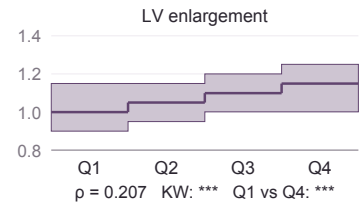
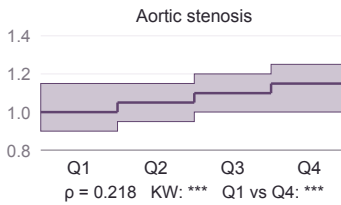
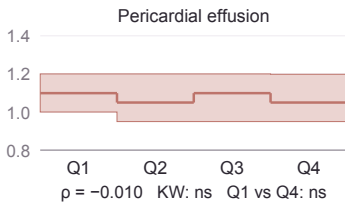
a Spearman correlations with echocardiographic measurements



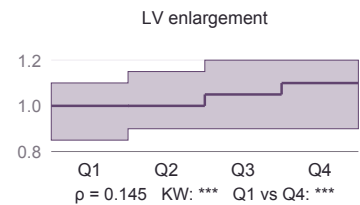
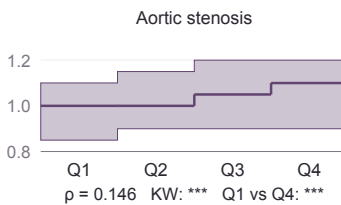
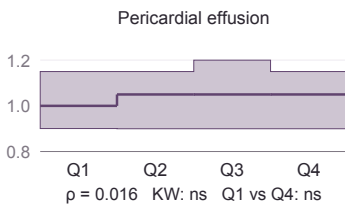
b LVEF (%)



c IVSs



d LVPWS



■ Interquartile range — Median

Figure 3: Associations between AnyECG-Echo predictions and quantitative echocardiographic measurements. **a**, Spearman rank correlations between the predicted probabilities of 13 structural heart disease phenotypes and left ventricular ejection fraction (LVEF), interventricular septal thickness (IVSs) and left ventricular posterior wall thickness (LVPWS). Circle area is proportional to $|\rho|$; purple and coral indicate positive and negative correlations, respectively. Numbers report the signed correlation coefficients. **b–d**, Echocardiographic measurements stratified by quartiles of the corresponding phenotype-specific predicted probability, from Q1 (lowest) to Q4 (highest): LVEF (**b**), IVSs (**c**) and LVPWS (**d**). Solid step lines show the median within each quartile, and shaded bands span the 25th to 75th percentiles. Colours follow the sign of the corresponding correlation. Statistical annotations accompany the correlation coefficients in **a**; annotations in **b–d** report Spearman's ρ and the significance of the across-quartile (KW) and Q1-versus-Q4 comparisons. ns, not significant.

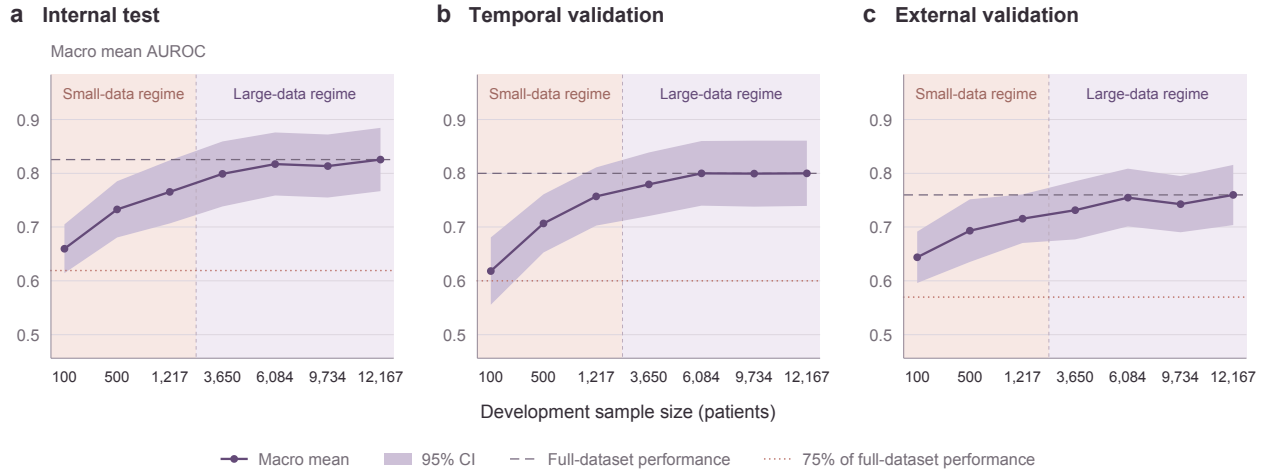


Figure 4: Data efficiency across evaluation cohorts. Macro-averaged area under the receiver operating characteristic curve (AUROC) is shown for the **a**, internal test; **b**, temporal validation; and **c**, external validation cohorts as the number of patients used for fine-tuning increases. Purple points indicate performance estimates, and the surrounding purple ribbons denote 95% confidence intervals. Training sample sizes are displayed as discrete, equally spaced categories. Pale coral and pale lavender backgrounds identify the small-data (100–1,273 patients) and larger-data (3,818–12,726 patients) regimes, respectively. Horizontal dashed lines indicate performance at the largest displayed sample size ($n = 12,726$); coral dotted lines indicate 75% of the corresponding full-dataset AUROC.

for the 13 phenotypes shown (Fig. 5b) on the internal test cohort. Across the full range of clinically plausible threshold probabilities, AnyECG-Echo consistently yielded a higher net benefit than both the default strategies of screening all patients and screening none, confirming that model-guided referral for echocardiography reduces unnecessary testing without sacrificing case detection.

Screening efficiency, quantified as the number of single-lead ECGs required to identify one true-positive case, further supports practical deployment. Under hospital prevalence, AnyECG-Echo required fewer than 10 screens per detection for LV enlargement, and Global heart enlargement. When using literature-derived community prevalence, the number of screens per detection increased as expected but remained within a practically actionable range (median across most subtypes < 100). Given that a single-lead ECG can be acquired in under 30 seconds at negligible cost using a consumer-grade wearable device, these figures compare favourably with established screening programmes for other asymptomatic conditions and support the feasibility of large-scale, opportunistic SHD screening outside the hospital setting.

DISCUSSION

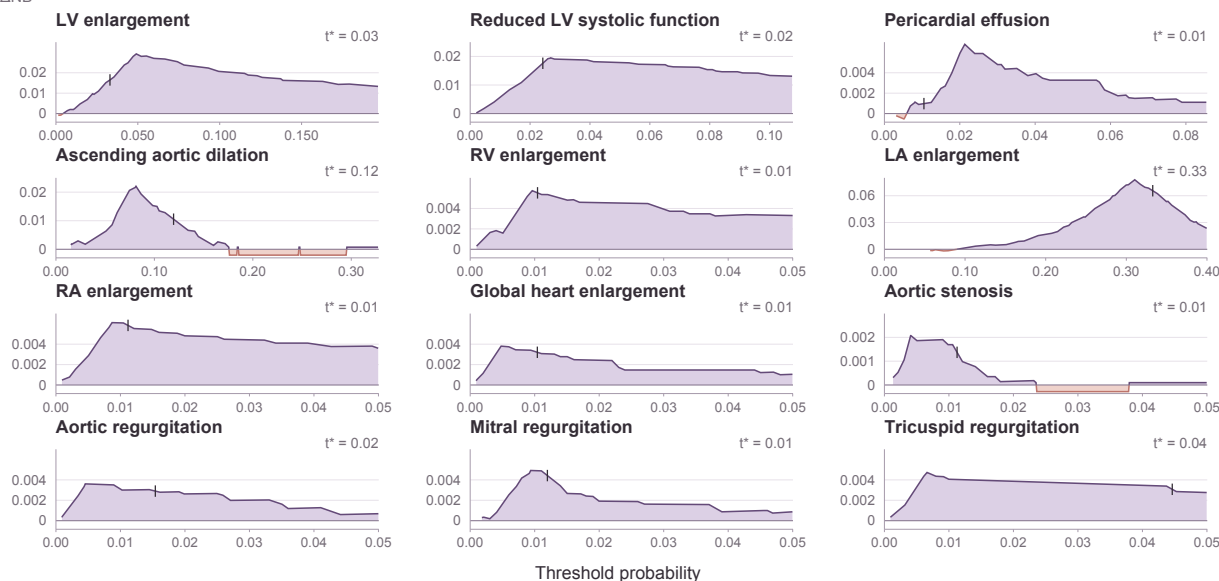
AnyECG-Echo combines ECG-report alignment with ECG-only self-supervision to detect 13 SHD phenotypes from single-lead recordings. Evaluation across internal, temporal and geographically independent external cohorts combines broad diagnostic coverage with testing across time and institutions. Routinely generated echocardiography reports provide a practical supervisory signal for learning structural cardiac information from electrical recordings, supporting the development of a single-lead screening framework across multiple anatomical domains.

Population-level median-beat comparisons and segment-level occlusion attribution provided complementary evidence of physiological consistency. Attribution scores averaged over the five patients with the highest predicted probability for each phenotype identified

a Incremental net benefit

$$\Delta\text{NB} = \text{model} - \max(\text{treat all}, 0)$$

ΔNB



b Screening efficiency

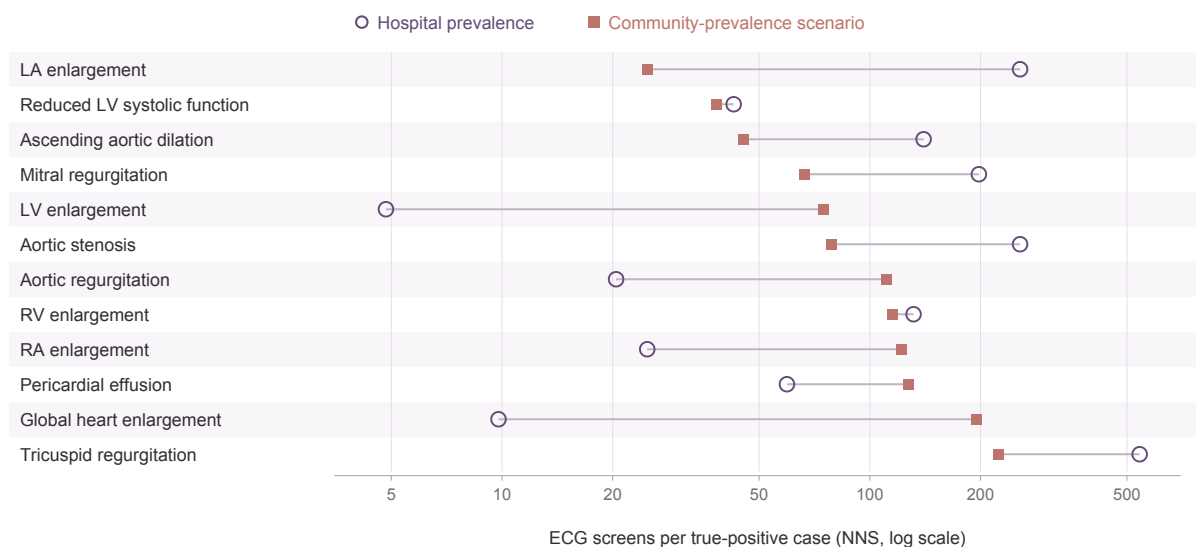


Figure 5: Incremental net benefit and screening efficiency of AnyECG-Echo. **a**, Incremental decision curves for the 12 phenotypes shown, evaluated in the internal test cohort. At each threshold probability t , incremental net benefit is defined as $\Delta\text{NB}(t) = \text{NB}_{\text{model}}(t) - \max\{\text{NB}_{\text{all}}(t), 0\}$, comparing model-guided referral for echocardiography with the better of referring all patients or referring none^{22?}. Positive values indicate greater net benefit than both default strategies. Purple and coral shading denote positive and negative increments, respectively. Short vertical ticks mark the annotated phenotype-specific thresholds t^* . Threshold ranges and vertical scales vary between phenotypes. **b**, Estimated numbers of ECG screens per true-positive case (number needed to screen, NNS) for the 13 phenotypes shown, displayed on a shared logarithmic axis. Outlined purple circles indicate estimates at hospital prevalence, and filled coral squares indicate estimates recalculated using literature-derived community prevalence. Grey lines connect paired estimates; rows are ordered by increasing community-scenario NNS. Lower NNS indicates fewer screens per detected case. Filled coral squares indicate community NNS estimates calculated using literature-derived prevalence, with sources provided in *Supplementary Table: Community prevalence estimates used for NNS calculation*. LA, left atrial; LV, left ventricular; RA, right atrial; RV, right ventricular.

ECG segments contributing to the model's most confident predictions. These findings complemented the morphological differences observed between high-risk and low-risk groups. Phenotype-specific associations with left ventricular ejection fraction and myocardial wall thickness further linked model outputs to quantitative echocardiographic measurements. Together, these analyses support the physiological interpretability of the predictions and their potential use as digital biomarkers of SHD.

An ECG-first screening pathway would use these predictions to prioritize patients for echocardiographic assessment. Decision-curve analysis and prevalence-based number-needed-to-screen estimates complement discrimination metrics with assessments of net benefit and screening efficiency. Potential applications include opportunistic screening in primary care, wearable-based home monitoring and pre-hospital triage. These settings provide opportunities to evaluate whether accessible single-lead acquisition can improve referral decisions and the allocation of echocardiographic resources.

Several limitations warrant consideration. Development and validation were conducted in cohorts of East Asian descent, and cross-ethnic validation, including Western populations, is needed to assess broader generalizability. Although temporal and geographic validation addressed concerns about overfitting and generalization, the retrospective design requires prospective, real-world studies to establish the model's impact on clinical workflows and patient outcomes.

The supervisory signal depends on the diagnostic consistency of physician-authored echocardiography reports. Although contrastive learning can tolerate some noise, sensitivity to interval cardiac changes within the ± 10 -day matching window has not been assessed through dedicated ablation studies. Missing sex data (41.5% in specific cohort subsets) may also limit the robustness and interpretability of sex-stratified bias analyses.

Anatomical and hardware-specific constraints further define the model's scope. Discrimination was limited for atrial pathologies, potentially reflecting the difficulty of capturing subtle *P*-wave morphologies from a single lead compared with the more prominent ventricular signals of the QRS complex. Hospital-recorded Lead I signals provide a proxy for wearable recordings but do not capture the full range of motion artifacts, signal-quality variation and longitudinal dynamics encountered with consumer devices. Evaluation using noisy, continuous recordings from diverse hardware platforms is therefore needed to establish performance in wearable-based home monitoring.

Referral-based recruitment introduces spectrum bias. Both cohorts comprised patients clinically referred for echocardiography, enriching the dataset for symptomatic and advanced SHD. The reported AUROC values therefore describe performance in these referral populations. Community number-needed-to-screen estimates were calculated using literature-derived prevalence and have not been validated against observed community screening outcomes. Prospective, community-based cohorts remain necessary to assess performance in asymptomatic populations with lower disease prevalence.

Temporal validation used patient-disjoint data from a later period at PKUPH, defined by the November 2019 cutoff. This cohort assessed performance over time within the development centre, sharing its institutional infrastructure, hardware environment and physician workforce. The design therefore provides limited evidence about robustness to longer-term changes in clinical practice, equipment upgrades and staff turnover. Geographically independent external validation at SHTMU partially addresses concerns about institution-specific bias, while multicentre, prospectively enrolled temporal cohorts remain necessary to evaluate long-term robustness across diverse clinical environments.

METHODS

310

Definitions of 13 SHD subtype labels

311

Labels were derived from physician-authored echocardiography reports. Each label was assigned as positive when the corresponding echocardiographic finding met or exceeded the threshold criteria defined below. Definitions follow current American Society of Echocardiography (ASE) where applicable²³. Each valvular label denotes moderate or greater severity, defined by the qualitative or semi-quantitative thresholds above, derived from the integrative grading scheme of the 2017 ASE/SCMR (regurgitation) and 2017 EACVI/ASE (stenosis) guidelines^{24,25}. As labels were extracted by rule-based nature language process (NLP) from physician-authored reports, the dominant signal is the reported severity term; the numeric thresholds are listed for transparency rather than as the literal extraction rule.

312
313
314
315
316
317
318
319
320

Table 4: Definitions and diagnostic criteria for the 13 SHD subtype labels.

Label	Echocardiographic criteria
Reduced LV systolic function	LVEF < 53%
Pericardial effusion	Echo-free space in diastole > 10 mm
LV enlargement	LVIDd > 5.8 cm (male) or > 5.2 cm (female); or LV EDV/BSA > 74 mL/m ² (male) or > 61 mL/m ² (female)
LA enlargement	LAD > 4.0 cm (male) or > 3.8 cm (female); or LA volume index > 34 mL/m ²
RV enlargement	RVD1 > 41 mm at basal level; or RVD2 > 35 mm at mid-level
RA enlargement	RA volume index > 39 mL/m ² (male) or > 33 mL/m ² (female); or RA minor axis > 2.5 cm/m ²
Global heart enlargement	Enlargement of ≥3 chambers reported simultaneously
Ascending aortic dilation	Ascending aortic diameter > 4.0 cm
Aortic stenosis (mod+)	Mean gradient ≥ 20 mmHg, or peak aortic jet velocity (V_{max}) ≥ 3.0 m/s, or AVA ≤ 1.5 cm ² (indexed AVA ≤ 0.85 cm ² /m ²)
Aortic regurgitation (mod+)	VC ≥ 3 mm, or EROA ≥ 0.10 cm ² , or RVol ≥ 30 mL, or RF ≥ 30%, or PHT ≤ 500 ms, or any holodiastolic flow reversal in the descending aorta
Mitral stenosis (mod+)	MVA ≤ 1.5 cm ² (by planimetry, PHT, or continuity equation), or mean transmitral gradient ≥ 5–10 mmHg at resting heart rate, or PHT ≥ 150 ms
Mitral regurgitation (mod+)	VC ≥ 3 mm, or EROA ≥ 0.20 cm ² (primary MR; ≥ 0.20 cm ² also used for severe secondary MR per some criteria), or RVol ≥ 30 mL, or RF ≥ 30%
Tricuspid regurgitation (mod+)	VC ≥ 3 mm, or EROA ≥ 0.20 cm ² , or RVol ≥ 30 mL, or PISA radius > 6 mm (Nyquist ~28 cm/s)

Note: SHD, structural heart disease; LV, left ventricular; LVEF, left ventricular ejection fraction; LVIDd, left ventricular internal diameter in diastole; EDV, end-diastolic volume; BSA, body surface area; LA, left atrial; LAD, left atrial diameter; RV, right ventricular; RVD1, right ventricular basal diameter; RVD2, right ventricular mid-cavity diameter; RA, right atrial; VC, vena contracta; PHT, pressure half-time; V_{max} , maximum velocity; EROA, effective regurgitant orifice area; AVA, aortic valve area; PISA, proximal isovelocity surface area; MVA, mitral valve area.

Data sources and cohort construction

321

This is a retrospective, multicenter diagnostic accuracy study with an exploratory longitudinal sub-analysis. This study utilized longitudinal ECG and transthoracic echocardiography data from two tertiary medical centers in China: Peking University People's Hospital (PKUPH; Beijing, development center) and the Second Hospital of Tianjin Medical University (SHTMU; Tianjin, external validation center). The external site provided a geographically independent cohort located approximately 140 km from the development site. The study protocol was approved by the Institutional Review Boards of PKUPH (No. 2024PHB428-001) and SHTMU (No. KY2025K386). The requirement for informed consent was waived in accordance with relevant ethical guidelines.

322
323
324
325
326
327
328
329
330

Matched ECG–echocardiography pairs were identified at the patient level using a hierarchical temporal strategy. For each report, we prioritized same-day ECGs, expanding the search window up to ±10 days to select the temporally closest recording. Pairs exceeding this 10-day window were excluded to minimize interval cardiac changes. Across both centers, this strategy yielded matched data for a total of 37,389 unique patients. To prevent data leakage, all partitions were conducted at the patient level, ensuring no individual's records overlapped between the training, validation, and testing sets.

331
332
333
334
335
336
337

At PKUPH (June 2015–May 2023), 20,768 patients were identified from 74,220 screened individuals, including 11,758 with structural heart disease (SHD) and 9,010 without SHD. To evaluate model robustness against distributional shifts, data were stratified temporally using a November 2019 cutoff. Patients from the earlier period were allocated to model development ($n = 12,167$) and internal testing ($n = 4,078$). The development cohort was further divided into training ($n = 10,067$) and validation ($n = 2,100$) subsets. Patients from the later period formed a distinct temporal validation set ($n = 4,523$) to assess performance over time. At the external center, 16,621 patients were selected from an institutional repository of 479,089 individuals for geographically independent validation. Patient selection and cohort partitioning are summarized in Fig. 6, and detailed cohort characteristics are provided in *Supplementary Table: Study Population*.

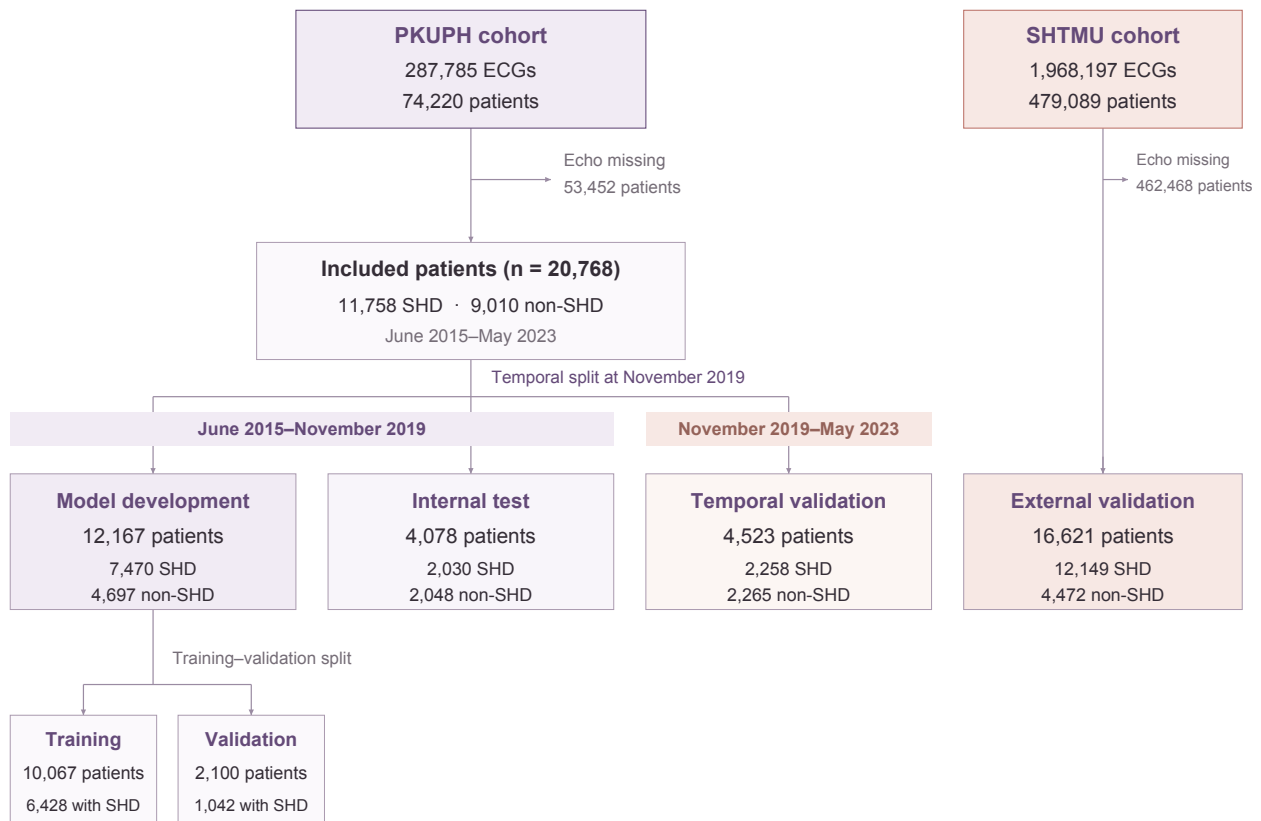


Figure 6: **Patient selection and cohort partitioning across the two study centers.** PKUPH, Peking University People's Hospital; SHTMU, Second Hospital of Tianjin Medical University; ECG, electrocardiogram; SHD, structural heart disease.

Data preprocessing

All ECG recordings were processed through a standardised pipeline before being used for model development. Records containing corrupted waveforms, incomplete lead data, or inconsistent metadata were first discarded. ECG signals acquired at sampling rates below 500 Hz were upsampled to 500 Hz through linear interpolation to ensure a consistent temporal resolution across all recordings. A notch filter centred at 50/60 Hz was applied to each signal to attenuate power-line interference, after which a fourth-order Butterworth bandpass filter with a passband of 0.67–40 Hz was employed to retain physiologically meaningful frequency components while suppressing both high-frequency noise and low-frequency artefacts. Residual baseline wander was corrected by subtracting a median-filtered

baseline estimate computed over a 0.4-second sliding window. The resulting signal was thus constrained to the 0.67–40 Hz frequency band with power-line interference and baseline drift minimised. For recordings exceeding 10 seconds in duration, a 10-second segment was extracted from the beginning of the signal; recordings shorter than 10 seconds were zero-padded to reach the target length. Prior to model input, each 10-second segment was normalised to zero mean and unit variance computed over that segment.

AnyECG-Echo architecture

AnyECG-Echo is a deep learning framework designed for structural heart disease (SHD) detection from single-lead ECG. The framework operates in two stages: multimodal contrastive pre-training and supervised fine-tuning. During pre-training, the model aligns raw single-lead waveforms with physician-authored echocardiography reports in a shared latent space. In the fine-tuning stage, the pre-trained, frozen ECG encoder serves as a feature extractor for a linear classification head to detect 13 specific SHD subtypes.

Single-lead ECG encoder

The ECG encoder maps a single-lead recording $\mathbf{X}_e \in \mathbb{R}^{1 \times L}$ (L denotes temporal length) into a low-dimensional representation. We first employ a one-dimensional residual network (ResNet18) to extract hierarchical temporal features:

$$\mathbf{H} = f_{\text{ResNet}}(\mathbf{X}_e; \theta_{\text{res}}) \in \mathbb{R}^{C \times T} \quad (1)$$

where C and T denote channel and downsampled temporal dimensions. A 1×1 convolution projects $\mathbf{H}$ into an embedding sequence $\mathbf{Z} \in \mathbb{R}^{d_{\text{proj}} \times T}$.

To capture global context, we process $\mathbf{Z}$ using a multi-head self-attention layer. A learnable class token [CLS] and positional embeddings are added, and the output corresponding to the [CLS] token, $\mathbf{e}_e \in \mathbb{R}^{d_{\text{out}}}$, serves as the global ECG embedding.

Contrastive pre-training objective

The encoder is optimized using a weighted sum of unimodal and cross-modal contrastive losses: $\mathcal{L}_{\text{pre}} = \lambda_{\text{ECG}} \mathcal{L}_{\text{ECG}} + \lambda_{\text{EM}} \mathcal{L}_{\text{EM}}$.

Unimodal Loss ($\mathcal{L}_{\text{ECG}}$): To ensure representation invariance, we generate two augmented views, $\mathbf{z}_e^{(1)}$ and $\mathbf{z}_e^{(2)}$, via independent dropout and linear projections. We minimize an InfoNCE-style loss:

$$\mathcal{L}_{\text{ECG}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\text{sim}(\mathbf{z}_{e,i}^{(1)}, \mathbf{z}_{e,i}^{(2)})/\tau)}{\sum_{j=1}^N \exp(\text{sim}(\mathbf{z}_{e,i}^{(1)}, \mathbf{z}_{e,j}^{(2)})/\tau)} \quad (2)$$

Cross-modal Loss ($\mathcal{L}_{\text{EM}}$): Single-lead ECG embeddings are aligned with semantic text embeddings $\tilde{\mathbf{e}}_t$ derived from frozen MedCPT query encoder²⁶. We employ a symmetric CLIP-style objective:

$$\mathcal{L}_{\text{EM}} = -\frac{1}{2N} \left(\sum_{i=1}^N \log \frac{\exp(s_{ii})}{\sum_{j=1}^N \exp(s_{ij})} + \sum_{j=1}^N \log \frac{\exp(s_{jj})}{\sum_{i=1}^N \exp(s_{ij})} \right) \quad (3)$$

where s_{ij} is the scaled cosine similarity between ECG i and MedCPT-encoded report j .

Supervised fine-tuning

For downstream detection, we append a linear classification head to the frozen single-lead ECG encoder. The average-pooled representation $\bar{z}$ is mapped to predicted logits $\mathbf{o} \in \mathbb{R}^M$ ($M = 13$):

$$\mathbf{o} = \mathbf{W}_{\text{cls}}\bar{z} + \mathbf{b}_{\text{cls}} \quad (4)$$

The classification head is trained using binary cross-entropy loss, while the encoder remains fixed to preserve the foundational physiological features learned during pre-training.

Data Availability

The data that support the findings of this study are not publicly available due to restrictions imposed by institutional ethics committees and data governance policies of the participating hospitals. Access to the data may be considered upon reasonable request to the corresponding author, subject to approval by the relevant ethics committees.

Code Availability

The code used for model development and evaluation in this study is publicly available at <https://github.com/gody123gody/AnyECG-Echo>, or <https://github.com/PKUDigitalHealth/AnyECG-Echo>.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (62102008, 82470527), CCF-Tencent Rhino-Bird Open Research Fund (CCF-Tencent RAGR20250108), CCF-Zhipu Large Model Innovation Fund (CCF-Zhipu202414), PKU-OPPO Fund (BO202301, BO202503), Research Project of Peking University in the State Key Laboratory of Vascular Homeostasis and Remodeling (2025-SKLVHR-YCTS-02), Beijing Municipal Science and Technology Commission (Z251100000725008), Prevention and Control of Emerging and Major Infectious Diseases-National Science and Technology Major Project (2025ZD01906000, 2025ZD01906004).

Author Contributions

Chenyang He and Qinghao Zhao contributed equally to this work. Kangyin Chen, Cheng Ding, and Shenda Hong are co-corresponding authors.

Chenyang He: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft. Qinghao Zhao: Data curation, Investigation, Validation, Writing – review & editing. Shun Huang, Jun Li, and Gongzheng Tang: Software, Validation, Formal analysis. Hao Zhang, Tong Liu, and Zhengkai Xue: Data curation, Resources. Kangyin Chen: Resources, Data curation, Supervision – review & editing. Cheng Ding: Methodology, Supervision – review & editing. Shenda Hong: Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Competing Interests

425

The authors declare no competing interests.

426

References

427

1. Dhingra, L.S., Aminorroaya, A., Sangha, V., Pedroso, A.F., Shankar, S.V., Coppi, A., Foppa, M., Brant, L.C., Barreto, S.M., Ribeiro, A.L.P. et al. (2025). Ensemble deep learning algorithm for structural heart disease screening using electrocardiographic images: Present shd. *Journal of the American College of Cardiology* 85, 1302–1313. 428
429
430
431
2. Aminorroaya, A., Dhingra, L.S., Pedroso, A.F., Shankar, S.V., Coppi, A., Khunte, A., Foppa, M., Brant, L.C., Barreto, S.M., Ribeiro, A.L.P. et al. (2025). Development and multinational validation of an ensemble deep learning algorithm for detecting and predicting structural heart disease using noisy single-lead electrocardiograms. *European Heart Journal-Digital Health* 6, 554–566. 432
433
434
435
436
3. Taggart, C., Ferry, A., Barker, S., Williams, K., Souter, G., Bularga, A., Wereski, R., McDermott, M.J., Williams, M.C., Boeddinghaus, J. et al. (2025). Targeting investigation and treatment in type 2 myocardial infarction: a pilot randomized controlled trial. *JACC: Advances* 4, 101738. 437
438
439
440
4. Ortiz-Garrido, A., Piñero-Urbe, I., Llebana-Carpio, L., Cano-Nieto, J., Gaitan-Román, D., and Picazo-Angelín, B. (2025). Double chambered left ventricle: an incidental rare congenital heart disease. report of a case within the context of a literature review. *The International Journal of Cardiovascular Imaging* pp. 1–6. 441
442
443
444
5. Unger, P., Galloo, X., and Pibarot, P. (2025). Mixed valvular heart disease: diagnosis and management. *European Heart Journal* 46, 2261–2274. 445
446
6. Yao, Z., Xie, W., Zhang, J., Yuan, H., Huang, M., Shi, Y., Xu, X., and Zhuang, J. (2023). Graph matching and deep neural networks based whole heart and great vessel segmentation in congenital heart disease. *Scientific reports* 13, 7558. 447
448
449
7. Lee, J.C., Geske, J.B., Narang, A., Khalique, O.K., Choi, A.D., Sun, Y.P., Cavalcante, J.L., Pinto, D.S., Gafoor, S.A., Jagasia, D.H. et al. (2023). Structural heart imaging survey highlights: training, challenges, and practice patterns in interventional imaging. *Cardiovascular Imaging* 16, 255–258. 450
451
452
453
8. Antoniadou, C., and Chan, K. (2025). Scalable screening for structural heart disease: promises from artificial intelligence-electrocardiogram tools. Oxford University Press UK. 454
455
9. Hickson, L.J., Negrotto, S.M., Onuigbo, M., Scott, C.G., Rule, A.D., Norby, S.M., Albright, R.C., Casey, E.T., Dillon, J.J., Pellikka, P.A. et al. (2016). Echocardiography criteria for structural heart disease in patients with end-stage renal disease initiating hemodialysis. *Journal of the American College of Cardiology* 67, 1173–1182. 456
457
458
459
10. Jiang, H., Wang, T., Sun, Z., Wang, Y., Yue, Y., Sun, Y., Jia, N., Li, M., Luo, S., Song, S. et al. (2025). Ultrasep: Sequence-aware pre-training for echocardiography probe movement guidance. *Pattern Recognition* pp. 112600. 460
461
462

11. Krammel, M., Hamp, T., Hafner, C., Magnet, I., Poppe, M., and Marhofer, P. (2023). Feasibility of resuscitative transesophageal echocardiography at out-of-hospital emergency scenes of cardiac arrest. *Scientific Reports* 13, 20085. 463 464 465
12. Palermi, S., Vecchiato, M., Ng, F.S., Attia, Z., Cho, Y., Anselmino, M., De Ferrari, G.M., Saglietto, A., Sau, A., Chiu, I.M. et al. (2025). Artificial intelligence and the electrocardiogram: A modern renaissance. *European Journal of Internal Medicine*. 466 467 468
13. Alimbayeva, Z.N., Alimbayev, C.A., Bayanbay, N.A., Ozhikenov, K.A., Bodin, O.N., and Mukazhanov, Y.B. (2022). Portable ecg monitoring system. *International Journal of Advanced Computer Science and Applications* 13. 469 470 471
14. Wang, R., Veera, S.C.M., Asan, O., and Liao, T. (2024). A systematic review on the use of consumer-based ecg wearables on cardiac health monitoring. *IEEE Journal of Biomedical and Health Informatics* 28, 6525–6537. 472 473 474
15. Lin, S., Li, Z., Wu, Q., Chen, Y., Yang, Y., and Zhao, H. (2026). A self-supervised electrocardiogram foundation model for empowering cardiovascular disease prediction and genetic factor discovery. *Nature Communications*. 475 476 477
16. Zhu, H., Luo, Q., Wang, Y., Maimaiti, M., Zhang, H., Xiong, Y., Ruan, C., Wang, J., Liu, Y., Zhou, M. et al. (2026). Artificial intelligence-enabled 8-channel ecg diagnosing of abnormalities with wide qrs complexes: a cohort study. *Health Data Science*. 478 479 480
17. Poterucha, T.J., Jing, L., Ricart, R.P., Adjei-Mosi, M., Finer, J., Hartzel, D., Kelsey, C., Long, A., Rocha, D., Ruhl, J.A. et al. (2025). Detecting structural heart disease from electrocardiograms using ai. *Nature* 644, 221–230. 481 482 483
18. Hughes, J.W., Jing, L., Finer, J., Hartzel, D., Kelsey, C., Long, A., Rocha, D., Ruhl, J., Poterucha, T., and Elias, P. (2026). Echonext-mini: A dataset and baseline ai model for detecting structural heart disease from electrocardiograms. *NEJM AI* pp. Aldbp2500516. 484 485 486
19. Liang, Y., Sau, A., Zeidaabadi, B., Barker, J., Patlatzoglou, K., Pastika, L., Sieliwonczyk, E., Whinnett, Z., Peters, N.S., Yu, Z. et al. (2025). Artificial intelligence-enhanced electrocardiography to predict regurgitant valvular heart diseases: an international study. *European Heart Journal* 46, 4823–4837. 487 488 489 490
20. Knight, E., Oikonomou, E.K., Aminorroaya, A., Pedroso, A.F., and Khera, R. (2026). Wearable-echo-fm: an ecg echo foundation model for 1-lead electrocardiography. *European Heart Journal-Digital Health* 7, ztag049. 491 492 493
21. van der Valk, V., Atsma, D., Scherptong, R., and Staring, M. (2025). Explainable ecg analysis by explicit information disentanglement with vaes. *IEEE Transactions on Biomedical Engineering*. 494 495 496
22. Vickers, A.J., Van Claster, B., Wynants, L., and Steyerberg, E.W. (2023). Decision curve analysis: confidence intervals and hypothesis testing for net benefit. *Diagnostic and prognostic research* 7, 11. 497 498 499
23. Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T. et al. (2015). Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging. *European Heart Journal-Cardiovascular Imaging* 16, 233–271. 500 501 502 503 504

24. Zoghbi, W.A., Adams, D., Bonow, R.O., Enriquez-Sarano, M., Foster, E., Grayburn, P.A., Hahn, R.T., Han, Y., Hung, J., Lang, R.M. et al. (2017). Recommendations for noninvasive evaluation of native valvular regurgitation: a report from the american society of echocardiography developed in collaboration with the society for cardiovascular magnetic resonance. *Journal of the American Society of Echocardiography* *30*, 303–371.
25. Baumgartner, H., Hung, J., Bermejo, J., Chambers, J.B., Edvardsen, T., Goldstein, S., Lancellotti, P., LeFevre, M., Miller Jr, F., Otto, C.M. et al. (2017). Recommendations on the echocardiographic assessment of aortic valve stenosis: a focused update from the european association of cardiovascular imaging and the american society of echocardiography. *European Heart Journal-Cardiovascular Imaging* *18*, 254–275.
26. Jin, Q., Kim, W., Chen, Q., Comeau, D.C., Yeganova, L., Wilbur, W.J., and Lu, Z. (2023). Medcpt: Contrastive pre-trained transformers with large-scale pubmed search logs for zero-shot biomedical information retrieval. *Bioinformatics* *39*, btad651.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplyment.pdf](#)