

Multimodal AI estimates hemodynamic echocardiographic indicators of structural heart disease from heart sounds and pulse waves

Qun Wei

Keimyung University

In-Cheol Kim

Keimyung University Dongsan Hospital

Ji-Seung Han

Keimyung University

Ji-Hyun Kim

Keimyung University

Ji-Woo Jeong

Keimyung University

Gang-oh Na

Keimyung University

Sang-Won Lee

Keimyung University

Geunhyung Kim

Keimyung University

Hyun-Jin Kim

Keimyung University Dongsan Hospital

Se-Young Yun

Keimyung University

MinJeong Kim

Keimyung University

Jaehyeong Cho

Keimyung University Dongsan Medical Center

Chenxi Yang

Southeast University

Yong-Sik Park

Clairaudience Co., Ltd.

Min-Seok Kim

Keimyung University

Hyungseop Kim

khyungseop@dsmc.or.kr

Keimyung University Dongsan Hospital

Article

Keywords:

Posted Date: September 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11038459/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Echocardiography is the reference method for cardiac hemodynamics but requires dedicated hardware and an expert operator, whereas the stethoscope reports only whether disease is likely. This study asked whether a stethoscope examination can instead yield triage-level estimates of the quantities that echocardiography measures. A multimodal (heart sound, pulse wave and demographics) model estimated continuous echocardiographic indicators from one 15-second recording of chest phonocardiography and finger photoplethysmography with a single device. In 1,502 recordings from 1,339 consecutive patients with same-day echocardiography, out-of-fold Pearson correlation averaged 0.402 across eight pre-registered indicators and 0.409 across a post hoc protocol-acquired panel; in 118 held-out recordings, 0.516 and 0.453. Predictions were compressed toward the cohort mean, so the output is a triage-level estimate: a pre-specified rule-out rule reached a sensitivity of 0.95 at a specificity of 0.27. Flow-generated quantities were recovered whereas wall thickness was not. All data come from one center; transportability is untested.

Introduction

Clinical decision-making in structural heart disease, which includes valvular lesions and the ventricular and atrial changes that accompany them, relies on hemodynamic parameters, including jet velocity across a stenotic valve, the corresponding pressure gradient, effective orifice area, regurgitant orifice size, pulmonary artery systolic pressure, ventricular ejection fraction, and ventricular filling pressure. Clinical guidelines assess disease severity and determine the appropriate timing for intervention based on these parameters¹⁻³. These parameters represent the mechanical quantities of flow and pressure generated by the heart and its valves.

Transthoracic echocardiography (TTE) with Doppler measures these parameters directly and is the reference standard for hemodynamic evaluation. It is, however, a specialized modality in two senses at once. It requires dedicated hardware, and it requires an operator: both acquisition, which depends on an adequate acoustic window and precise Doppler alignment, and interpretation demand skills cultivated over years⁴. Handheld and portable scanners have reduced the hardware cost⁵, but not the operator dependence, so the constraint has moved rather than disappeared. Hemodynamic assessment is therefore usually unavailable at the point where patients are first assessed. The consequence is measurable: valvular heart disease affects more than 2% of the general population and more than 10% of adults over 75, yet most cases remain undiagnosed⁶⁻⁸, a situation described as the next cardiac epidemic⁹. The 2025 ESC/EACTS guidelines reaffirm auscultation as the essential first examination while documenting continuing underdiagnosis and undertreatment in the community^{1,10}. The gap is therefore one of access, not of physics: the instrument present at first contact does not measure, and the instrument that measures is not present. This motivates asking what the accessible instrument can be made to measure.

This continued reliance on auscultation has a physical basis. As the natural bedside counterpart to echocardiography, the phonocardiogram (PCG) observes the same physiological phenomena that Doppler echocardiography quantifies — the closure of the cardiac valves, the flow of blood through them, and the pressures that this flow generates. The first and second heart sounds (S1, S2) mark the closure of the atrioventricular and semilunar valves, which define the boundaries of systole used in echocardiographic assessment. Murmurs constitute the acoustic signatures of turbulent flow: blood propelled through a stenotic orifice or regurgitating via an incompetent valve generates broadband vibrations, the intensity, frequency content, and phase within the cardiac cycle of which are directly related to parameters measured by Doppler echocardiography, including jet velocity, pressure gradient, and effective orifice area. Traditional physical diagnosis qualitatively exploits these correlations: the volume and late peaking of an ejection murmur correspond to the aortic pressure gradient, the prominence of the pulmonic component of the second heart sound indicates pulmonary artery pressure, and the amplitude of regurgitant murmurs is related, albeit imperfectly, to the size of the regurgitant orifice. The arterial pulse can be interpreted similarly: the slow-rising, low-amplitude surge in severe aortic stenosis and the wide pulse pressure associated with aortic regurgitation are hemodynamic deductions derived from a pressure waveform, whose peripheral manifestation is the fingertip pulse. Decades before the advent of deep learning, quantitative spectral analysis of murmurs was used to grade the severity of aortic stenosis and even estimate valve area^{11–14}. In practice, however, auscultation yields qualitative impressions rather than quantitative data. The bedside interpretation of systolic murmurs proves unreliable even among seasoned clinicians^{15,16}, and its sensitivity for echocardiographically confirmed disease is relatively low in asymptomatic patients in primary care settings¹⁷. The hemodynamic information is present in the sound; what is missing is a means of extracting it in quantitative form.

Machine learning offers a route to this kind of quantitative extraction. Deep learning applied to digital stethoscope recordings has demonstrated that heart sounds contain recoverable hemodynamic information: case–control studies have identified clinically significant murmurs¹⁸, PCG-based pulmonary hypertension screening has achieved an area under the receiver operating characteristic curve (AUROC) of 0.82 on a per-patient basis¹⁹, and public challenge datasets have facilitated murmur-detection research^{20–23}. To date, however, these systems have remained predominantly end-to-end binary classifiers targeting one or two conditions. They answer whether disease is present, which is the question a screening test asks, but not how severe it is, which is the question that determines management: guideline thresholds are stated in velocities, gradients, areas and pressures, and a probability cannot be compared with any of them. Their outputs are single, non-transparent scores rather than quantitative hemodynamic measures, and cannot be inspected, reused, or audited against the clinical measurements that inform decision-making. Additionally, enriched case–control validation is known to overstate performance relative to consecutive clinical populations¹⁸.

These limitations concern what existing systems predict; a further and less visible one concerns how the signal itself is prepared. The PCG analysis requires delineating cardiac cycles, and the standard

procedure locates the first and second heart sounds through digital filtering and envelope-based detection tuned to their predominant low-frequency components^{24,25}. However, murmurs are faint, broadband, and spectrally proximate to noise: filtering methods sufficiently aggressive to clean bedside recordings also diminish the very signals that convey pathological information. Additionally, acoustic segmentation is least reliable in the pathological recordings where murmurs obscure the heart sounds. Consequently, noise removal risks attenuating the murmurs as well. The fusion of fiducial points from the PCG and a simultaneously recorded photoplethysmographic (PPG) signal offers a solution: it reliably identifies S1 and S2 in the presence of murmurs and ambient noise by aligning the sounds to the pulse wave rather than solely to acoustics²⁶. On this basis, the fingertip PPG pulse wave was adopted as the auxiliary channel, acquired together with the PCG by a single multimodal stethoscope: beyond supplying the timing reference, it reflects stroke volume, ejection duration, and vascular load^{27–29}, and multimodal deep learning that integrates complementary data streams has advanced risk prediction in other cardiac domains^{30,31}. Existing studies have accordingly advanced along two directions: obtaining a cleaner auscultation signal, and classifying individual conditions from it. A 2025 systematic review of 199 machine learning studies on PCGs³² found the field concentrated on murmur detection and disease-specific screening; none of the reviewed studies predicted continuous echocardiographic indicators, and pulse-wave-based approaches fell outside its scope. The present study takes the same signals one step further, applying the combined PCG–PPG recording to the estimation of the echocardiographic indicators themselves; this shift in prediction target — from classifying conditions to estimating the underlying quantities — is the focus of the present report.

The approach taken here is to change what the existing examination reports, not the instrument that performs it — to repurpose the stethoscope, by analogy with drug repurposing, for a measurement it was never built to make. The stethoscope is not modified, no new sensing modality is introduced, and no additional hardware is required beyond a pulse sensor already integrated into the same instrument; what changes is the task the recording is asked to perform, from deciding whether a lesion is present to estimating the quantities by which its severity is graded. Instead of mapping sound to a diagnosis, a multimodal deep-learning model — multimodal in that it combines inputs of different physical nature, the heart sound, the pulse wave and the demographic covariates age and sex, to estimate quantities defined in a third modality, echocardiography — was developed that jointly processes the two synchronized signals and reports a panel of continuous hemodynamic echocardiographic indicators, including aortic valve peak velocity (AV Vmax), the left ventricular outflow tract velocity–time integral (LVOT VTI), pulmonary artery systolic pressure (PASP), left ventricular ejection fraction (LVEF), filling pressure (E/e'), and the left atrial volume index (LAVI) (Fig. 1). Per-cycle mel-spectrogram representations of the PCG and morphological and timing features of the PPG are encoded by a deep network, aggregated across cardiac cycles, and regressed jointly onto the echocardiographic targets extracted from the report, with each prediction structured as the sum of a linear demographic offset determined by age and sex and a deep multimodal residual derived from the PCG and the PPG. The panel is hemodynamic by construction: because heart sounds and the pulse wave are generated by blood flow, the quantities determined by flow — jet velocity, pressure gradient, orifice area, regurgitant flow, pressures, and filling —

are those physically coupled to the recorded signal, along with the chamber dimensions remodeled by chronic hemodynamic load. This target choice confers three practical benefits. First, interpretability: the output uses the same quantitative vocabulary as clinicians, so predictions can be inspected and reasoned about rather than accepted as a score. Second, extensibility: the indicator panel is designed as the substrate for downstream diagnostic modules, which can be built without retraining the signal front-end; these modules are under development and are beyond the scope of this report. Third, explicit attribution: because age and sex alone predict many echocardiographic indicators, the linear demographic baseline is architecturally separated from the auscultatory contribution, so the added value of sound is quantified explicitly rather than assumed. Stating that contribution per indicator also shows where quantification succeeds and where it does not. The model was developed on a consecutive, unselected outpatient cohort of 1,502 recordings with same-day echocardiographic ground truth, validated on a held-out set of 118 recordings from 96 patients who were withheld from development and paired with an echocardiogram within 30 days, and characterized for test–retest repeatability and prediction-interval coverage; the sources of the current performance bottleneck were then analyzed.

Results

Study cohort

The development cohort comprised 1,502 recordings from 1,339 consecutive patients (mean age 66.1 ± 11.4 years, 39.1% female) undergoing clinically indicated TTE, each paired with a same-day echocardiogram. The held-out validation cohort comprised 118 recordings from 96 patients enrolled at the same center over the same period whose recordings could not be paired with an echocardiogram of the same day; these patients were withheld from model development in their entirety, with no patient overlap with the development cohort. Their recordings were paired with the nearest echocardiogram obtained within 30 days (median 10 days, interquartile range 7–17; mean age 64.4 ± 14.9 years, 41.7% female).

Neither cohort was enriched or selected on the basis of disease status. Patients were enrolled consecutively as encountered, so that the full clinical spectrum was preserved, including predominantly mild findings and low-quality recordings, which were flagged rather than discarded. Every recording yielded usable cardiac cycles (median 13, interquartile range 11–15); three recordings, two in development and one held-out, yielded fewer than five and were flagged as low-confidence but retained. Table 1 summarizes the demographic characteristics of both cohorts, the distribution of each indicator within the reported panel, and the severity spectrum of the four graded lesions.

Table 1
Cohort characteristics, measured echocardiographic indicators, and disease spectrum.

Characteristic	Development cohort	Held-out cohort	SMD	P value
Demographics and recording characteristics				
Recordings, n	1,502	118	—	—
Patients, n	1,339	96	—	—
Age, years	66.1 ± 11.4	64.4 ± 14.9	-0.13	0.27
Female, n (%)	524 (39.1)	40 (41.7)	+ 0.05	0.62
Body-mass index, kg/m ²	24.6 ± 3.6	24.5 ± 4.0	-0.02	0.83
Body surface area, m ²	1.73 ± 0.21	1.71 ± 0.21	-0.09	0.39
PCG–echocardiography interval, days	0 (IQR 0–0)	10 (IQR 7–17)	—	By design
Heart rate during recording (pulse wave), bpm	76.2 ± 12.5	79.1 ± 13.2	+ 0.22	0.02
Measured echocardiographic indicators, mean ± SD (n measured)				
Aortic valve peak velocity, m/s	1.43 ± 0.49 (1,496)	1.92 ± 1.28 (117)	+ 0.50	< 0.001
Mean transaortic pressure gradient, mmHg	16.0 ± 11.0 (131)	37.4 ± 32.1 (28)	+ 0.89	0.002
Aortic valve area, cm ²	2.35 ± 0.72 (529)	1.58 ± 0.66 (28)	-1.11	< 0.001
Dimensionless velocity index	0.65 ± 0.17 (548)	0.37 ± 0.18 (29)	-1.58	< 0.001
LVOT velocity–time integral, cm	20.2 ± 5.1 (562)	24.8 ± 5.8 (31)	+ 0.85	< 0.001
Regurgitant vena contracta width, cm	0.48 ± 0.16 (446)	0.56 ± 0.26 (39)	+ 0.38	0.06
Regurgitant PISA radius, cm	0.56 ± 0.21 (489)	0.63 ± 0.26 (45)	+ 0.30	0.08
Regurgitant volume, mL	38.5 ± 29.3 (270)	43.8 ± 26.7 (30)	+ 0.19	0.31
PASP, mmHg	26.1 ± 9.4 (1,282)	28.9 ± 11.7 (108)	+ 0.26	0.02
LVEF, %	56.4 ± 10.6 (1,502)	52.1 ± 15.4 (117)	-0.33	0.003

Characteristic	Development cohort	Held-out cohort	SMD	P value
Demographics and recording characteristics				
Global longitudinal strain, GLS , %	19.1 ± 3.1 (593)	17.5 ± 4.5 (41)	-0.41	0.03
E/e'	9.5 ± 4.4 (1,494)	12.2 ± 7.2 (118)	+ 0.46	< 0.001
Left atrial diameter, cm	4.16 ± 0.75 (1,502)	4.46 ± 0.83 (118)	+ 0.38	< 0.001
Left atrial volume index, mL/m ²	43.6 ± 17.9 (1,493)	52.8 ± 25.5 (118)	+ 0.42	< 0.001
LV end-diastolic diameter, cm	4.81 ± 0.58 (1,501)	5.03 ± 0.76 (118)	+ 0.32	0.003
LV end-systolic diameter, cm	3.34 ± 0.69 (1,502)	3.63 ± 1.00 (118)	+ 0.35	0.002
Valvular disease spectrum, n (%)				
Aortic stenosis			+ 0.44	< 0.001
None	1,406 (93.6)	99 (83.9)		
Mild	55 (3.7)	3 (2.5)		
Moderate	29 (1.9)	4 (3.4)		
Severe	12 (0.8)	12 (10.2)		
Aortic regurgitation			+ 0.28	0.02
None	1,018 (67.8)	69 (58.5)		
Mild	409 (27.2)	35 (29.7)		
Moderate	55 (3.7)	12 (10.2)		
Severe	20 (1.3)	2 (1.7)		
Mitral regurgitation			+ 0.31	0.02
None	506 (33.7)	34 (28.8)		
Mild	898 (59.8)	65 (55.1)		
Moderate	76 (5.1)	13 (11.0)		
Severe	22 (1.5)	6 (5.1)		

Characteristic	Development cohort	Held-out cohort	SMD	P value
Demographics and recording characteristics				
Tricuspid regurgitation			+ 0.13	0.18
None	481 (32.0)	32 (27.1)		
Mild	899 (59.9)	73 (61.9)		
Moderate	86 (5.7)	9 (7.6)		
Severe	36 (2.4)	4 (3.4)		

Values are mean $\pm$ SD, median (IQR) or n (%). SMD, standardized mean difference, held-out minus development, in pooled-SD units (for the graded lesions, the multinomial SMD across the four grades). P values are from Welch's t-test (continuous variables), the χ^2 test (sex) and the Mann–Whitney U test on the ordinal grade (lesions); the recording–echocardiography interval differs by design and was not tested. Demographic rows count each patient once; indicator and lesion rows count recordings. Heart rate is 60 divided by the mean length of the accepted pulse-wave cycles of each recording. For each lesion the SMD and P value refer to the four-grade distribution.

Two features of the reference data govern what can be reported. The first is that coverage is markedly uneven, and that the unevenness is not random. Measurements acquired by protocol were available in almost every study: LVEF and left ventricular dimensions in 99.9–100% of development recordings, the left atrial volume index in 99.4%, E/e' in 99.5%, peak aortic valve velocity in 99.6%, and PASP in 85.4%. Quantitative Doppler measurements, by contrast, are performed at the discretion of the echocardiographer and only when a lesion is suspected: the LVOT velocity–time integral was available in 37.4% of development recordings, the aortic valve area in 35.2%, the dimensionless velocity index in 36.5%, and the mean transaortic gradient in 8.7%. The same asymmetry is sharper in the held-out cohort, where the aortic valve area was available in 28 of 118 recordings and the LVOT velocity–time integral in 31, against 108 to 118 for the measurements acquired by protocol. Accuracy quoted for a discretionary measurement therefore describes the subgroup in whom that measurement was judged necessary, not the screening population; the reported panel was anchored on protocol-acquired measurements for this reason.

The second feature is that the held-out cohort carries a higher burden of disease: LVEF $52.1 \pm 15.4\%$ versus $56.4 \pm 10.6\%$, left atrial volume index 52.8 ± 25.5 versus 43.6 ± 17.9 mL/m², and a higher proportion of moderate or greater lesions, up to fivefold for aortic stenosis (13.6% versus 2.7%; mitral regurgitation 16.1% versus 6.5%, aortic regurgitation 11.9% versus 5.0%, tricuspid regurgitation 11.0% versus 8.1%). Demographics did not differ (standardized mean differences within ± 0.13), whereas every protocol-acquired indicator did (0.26–0.50 in absolute value) and the discretionary aortic-flow indices

differed most (0.85–1.58; Table 1). This shift in prevalence is revisited in the conformal coverage analysis.

Echocardiographic indicator prediction and the net contribution of auscultation

Two indicator panels are reported: the pre-registered panel is the pre-specified primary result and is given in the text and in Fig. 2a, whereas the figures and Tables 2 and 3 use the protocol-acquired panel because every one of its members is measured in essentially every study. Across the eight pre-registered indicators, the out-of-fold pooled Pearson correlation averaged 0.402 (models at the final training epoch, without validation-based selection; demographic offset included; 1,502 recordings), and across the eight protocol-acquired indicators that form the reported panel of the figures it averaged 0.409. Within the reported panel the correlation was 0.52 for LAVI, 0.50 for AV Vmax, 0.49 for the LVOT VTI, 0.44 for PASP, 0.34 for LVEF, 0.34 for E/e', 0.33 for the left ventricular end-diastolic diameter, and 0.31 for the end-systolic diameter (Fig. 2a). Age and sex alone reached 0.265 on this panel, so the increment attributable to the auscultatory signals was + 0.145. Throughout, this increment is the full-model correlation minus the age–sex baseline, with the baseline counted as zero where its out-of-fold correlation is negative (four indicators, none in the reported panel; see below). This reporting panel was selected by the collaborating cardiologist after the per-indicator results had been reviewed and is therefore post hoc; the selection and its timing are declared in the Methods. The pre-registered panel — the eight indicators measured in at least 30% of studies, fixed in a dated target-definition module before any result was examined — is the pre-specified primary result and is reported wherever the reported panel appears. On the identical models and rows its increment was larger, at + 0.185, because it is composed of the regurgitant and aortic-flow indices for which demographics carry almost no information. The revision thus leaves the mean correlation essentially unchanged (+ 0.008) while reducing the measured contribution of sound, and, as reported below, it also reduces the held-out value. It was adopted for the denominator it restores rather than for any number it improves: every indicator in the revised panel is acquired by protocol in essentially every study, whereas half of the pre-registered panel is quantified only when a lesion is already suspected.

This contribution was not uniform, and its pattern is the central result of the study. Grouping the full target set by mechanism rather than by organ, the quantities generated by flow, pressure, and function gained + 0.195 from the auscultatory signals; the chamber dimensions remodeled by chronic hemodynamic load gained + 0.140; and the static structural attributes transcribed from the report — wall thickness, leaflet calcification and thickening, annular dilatation — gained + 0.018, with the two wall-thickness targets alone at – 0.007, indistinguishable from nothing. The extremes make the mechanism explicit. Regurgitant vena contracta width is predicted at $r = 0.23$ by the full model, whereas the age–sex baseline carries no usable information: its out-of-fold correlation is negative in every fold (– 0.06 to – 0.40), meaning that the relation fitted on the training folds does not transfer, and such a baseline is counted as zero rather than as inverse information (the same applies to the regurgitant PISA radius, the

regurgitant volume and the mean transaortic gradient; without this floor the flow-tier increment would read + 0.251 and the pre-registered panel + 0.230). The auscultatory signal therefore supplies all of the predictive power for regurgitant orifice size. Conversely, aortic valve area is already predicted at $r = 0.43$ by age and sex, and sound adds only + 0.07: in this cohort, aortic stenosis is largely a function of age. Within the reported panel the same ordering holds — LAVI gains + 0.25 and AV Vmax + 0.23 from sound while the left ventricular end-diastolic diameter gains + 0.04 (Fig. 2a), although both LAVI and the diastolic diameter are chamber measurements acquired in the same studies with the same protocol. The organizing variable is therefore not whether a quantity is structural or hemodynamic, but whether it is generated by flow. These attributions show where an acoustic measurement adds information beyond a demographic risk table, and they are the reason the demographic offset is modeled explicitly instead of being folded into the network.

Whether a deep network is required for this task was examined, on the earlier data build (1,171 recordings) and the pre-registered panel, by interposing a classical rung between the demographic baseline and the full model. The earlier build was used because the analysis requires a demographics-only training run that exists only for that build; it is reported as a sensitivity analysis and its absolute values are not interchangeable with those above. On 1,171 evaluation rows, under the same patient-grouped five-fold splitting and with all normalization fitted on training folds only, ridge regression on the hand-crafted per-recording descriptors the pipeline already computes — the 55-dimensional cycle feature vector and the 24 diastolic-notch descriptors, taken as medians over the cycles of a recording, together with age and sex (81 dimensions, no spectrogram) — raised the mean correlation from 0.188 to 0.258, and the full deep multimodal model reached 0.373. The increment attributable to the learned representation (+ 0.12 on average) is concentrated where the hand-crafted descriptors fail: regurgitant vena contracta width (0.05 to 0.26), regurgitant PISA radius (0.12 to 0.28), and PASP (0.15 to 0.30). Where the descriptors already capture much of the signal — AV Vmax (0.41 to 0.53), LVOT VTI, aortic valve area — the deep margin is correspondingly smaller. The pattern follows the mechanism: envelope amplitude and pulse-contour descriptors carry the ejection-flow information, whereas murmur audibility, on which the regurgitant and pulmonary-pressure indicators depend, is recovered only by the learned spectrogram representation. The ladder shows where representation learning matters; the deep network contributes a real margin, but the claim of this study is not an architectural one.

Which of the two signals carries each indicator was examined by knockout on the trained networks: in each of the 25 networks, the embedding of one channel for every validation patient was replaced by the mean embedding of that network's training patients, so that the channel contributes no patient-specific information while the input stays within the distribution the fusion trunk was trained on (Methods; Supplementary Table S2). The two channels divide the panel between them. Removing the acoustic embedding reduced the AV Vmax correlation from 0.50 to 0.29 and PASP from 0.44 to 0.31, and drove the regurgitant indices and the mean gradient to zero (vena contracta 0.23 to 0.01, PISA radius 0.27 to 0.09, regurgitant volume 0.21 to - 0.08, mean gradient 0.35 to - 0.02), whereas removing the pulse-wave embedding left these essentially unchanged (AV Vmax 0.49). Removing the pulse-wave embedding instead reduced LVEF from 0.34 to 0.27, the LVOT VTI from 0.49 to 0.40, LAVI from 0.52 to 0.41, the left

atrial diameter from 0.41 to 0.29 and global longitudinal strain from 0.26 to 0.14, with the acoustic knockout costing these indicators little; for the ventricular diameters the acoustic knockout raised the correlation (left ventricular end-diastolic diameter, LVDd, 0.33 to 0.36, difference + 0.03, 95% CI + 0.00 to + 0.05), so the sound channel added noise rather than information to them. At the panel level the two knockouts looked symmetric (0.35 and 0.36 against 0.41) because each channel carries different indicators, not because they are redundant; on the pre-registered panel, weighted toward aortic-flow and regurgitant indices, the acoustic knockout cost far more (0.28 against 0.38). Removing both embeddings returned the panel mean to 0.26, the age–sex baseline, which confirms that the replacement removed information rather than corrupting the input. The pre-specified expectations understated this division: the pulse-wave share of AV Vmax had been expected at 30–40% and was 6%, and the ventricular diameters had been expected to be shared and were carried by the pulse wave alone. The pattern matches the physiological correspondences of Table 4: quantities generated by flow and pressure are read from the sound, and stroke volume and chamber size are read from the pulse. In both conditions the cycle boundaries remained pulse-wave-anchored; the knockout removes the information in the pulse wave, not its timing reference, and whether the panel could be recovered from a stethoscope without the pulse sensor was not a question of this study.

Beyond the reported panel, the model is trained on a wider target set, and the indicators outside the panel locate its boundaries (Fig. 3a, b). Among the quantities acquired only at the echocardiographer's discretion, out-of-fold correlation reached 0.50 for aortic valve area ($n = 529$), 0.44 for the dimensionless velocity index ($n = 548$), 0.35 for the mean transaortic pressure gradient ($n = 131$), 0.27 for the regurgitant PISA radius ($n = 489$), 0.26 for global longitudinal strain ($n = 593$), 0.23 for regurgitant vena contracta width ($n = 446$), and 0.21 for regurgitant volume ($n = 270$); the left atrial diameter, acquired by protocol, reached 0.41 ($n = 1,502$). Several of these equal or exceed the weaker members of the reported panel. They are nonetheless reported as secondary, because the denominator differs in kind: a discretionary Doppler quantity is measured in roughly a third of studies, and accuracy quoted for it describes the subgroup in whom the measurement was judged necessary, not the population a screening examination would encounter.

The composition of the panel follows the underlying physical mechanism rather than the availability of measurements. Heart sounds and the pulse wave are produced by blood flow; the quantities that flow determines — jet velocity, pressure gradient, orifice area, regurgitant flow, pulmonary pressure, and ventricular filling — are therefore causally linked to the recorded signals, and chamber dimensions reflect the chronic load that flow imposes. Static structural attributes transcribed from the report — left ventricular wall thickness, leaflet calcification and thickening, annular dilatation — do not generate sound. These were retained as auxiliary training targets (Methods) but are not reported as outputs, an exclusion decided on mechanistic grounds and applied irrespective of their measured correlations. Their measured accuracy is presented as a negative control, below the dashed line in Fig. 2a and within the target set in Fig. 3b, and is consistent with that reasoning: the interventricular septum reached $r = 0.200$ and the posterior wall 0.187, but age and sex alone reached 0.211 and 0.190 respectively, so the auscultatory signal contributed -0.011 and -0.003 . The apparent correlation of wall thickness is a

demographic correlation; the sound adds nothing to it. This marks where acoustic quantification stops, and the limit is set by the physics of the signal, not by the size of the model or the cohort.

Because a predicted indicator is used in practice to judge whether the corresponding measurement is likely to be abnormal, the predicted continuous value was additionally evaluated as a detector of abnormality of the measured value of the same indicator at conventional thresholds (Fig. 3c). The AUROC was 0.80 (95% CI 0.70–0.90) for AV Vmax $\geq$ 3.0 m/s (36 positive recordings), 0.80 (0.76–0.84) for PASP $\geq$ 40 mmHg (98 positives), 0.74 (0.67–0.81) for aortic valve area $\leq$ 1.5 cm² (56 positives), 0.71 (0.68–0.74) for LAVI $>$ 34 mL/m² (1,027 positives), 0.70 (0.65–0.75) for a left ventricular end-diastolic diameter $>$ 5.6 cm (123 positives), 0.69 (0.64–0.74) for E/e' $>$ 14 (141 positives), and 0.68 (0.64–0.71) for LVEF $<$ 50% (319 positives). All twelve thresholds examined, one for each indicator with an established cut-off, discriminated above chance, the weakest being regurgitant vena contracta width $\geq$ 0.4 cm (0.63, 0.57–0.68). Two caveats apply: the LAVI threshold is exceeded in 69% of this referred population, so detection there separates the minority of normal values rather than a rare abnormality, and several thresholds rest on few positives (36 for AV Vmax $\geq$ 3.0 m/s). Global longitudinal strain is stored as a signed value, so the abnormality listed as |GLS| $<$ 16% corresponds to GLS $>$ - 16%. These are indicator-level detections scored against the measured value of the same indicator; they are not diagnostic classifications of valvular disease. Confidence intervals are nominal, with no adjustment for multiplicity, and this analysis is descriptive.

Expressed in the physical units of each measurement, the results are as follows (Fig. 4). Of the AV Vmax predictions, 90% fell within $\pm$ 0.5 m/s of the measured value (mean absolute error 0.27 m/s), and 83% of PASP predictions fell within $\pm$ 10 mmHg (6.0 mmHg). For E/e', 71% fell within $\pm$ 3 units (2.73); for the left ventricular end-diastolic diameter, 69% within $\pm$ 0.5 cm (0.41 cm); and for the LVOT VTI, 68% within $\pm$ 4 cm (3.39 cm). Conversely, 50% of LAVI predictions fell within $\pm$ 8 mL/m² (10.9 mL/m²) and 46% of LVEF predictions within $\pm$ 5 percentage points (7.5 points). These tolerances are provisional working values, chosen to approximate a clinical decision step, and are not proposed as limits of agreement; the mean absolute error is shown in Fig. 4c as a fraction of the same tolerance, so that a value above 1.0 marks an error larger than the decision step. Restated as explained variance, the figures are 27% for LAVI, 25% for AV Vmax, and 12% for LVEF. These three views of the same output – correlation, which is the metric usually reported for this task, error in physical units, and explained variance – are presented together (Fig. 4a–c): correlation is the most permissive of the three, whereas error in physical units determines whether a prediction can stand in for a measurement in a clinical decision. Read in this way, the current output is a triage-level estimate: closest to a measurement for AV Vmax and PASP, and furthest for LAVI and LVEF, where the mean absolute error exceeds the tolerance that would be considered clinically acceptable. This is the resolution at which a quantified stethoscope examination is currently useful, at the point of care where echocardiography is absent.

Agreement in physical units was examined further with Bland–Altman analysis (Fig. 5; Table 2). Mean bias was negligible for every indicator (largest, - 0.9 mL/m² for LAVI), but the limits of agreement were wide and the predictions were compressed toward the cohort mean: the slope of predicted on measured

values ranged from 0.13 (LVDs) to 0.31 (LAVI), so that a predicted value moves only one-seventh to one-third as far as the measured value it estimates, and the downward trend of the differences with the mean, visible in every panel of Fig. 5, is the mirror image of that compression. The limits of agreement are descriptive rather than a tolerance: they bound the range within which 95% of individual differences fall, so that for LVEF (– 18.7 to + 19.9 percentage points) a single estimate can place a patient with reduced ejection fraction in the normal range, or the reverse, which is why no guideline category is derived from a point estimate. The 90% prediction-interval half-widths calibrated on the out-of-fold residuals were ± 0.50 m/s for AV Vmax, equal to its working tolerance; ± 12.4 mmHg for PASP (1.2 times the tolerance); ± 0.85 and ± 0.97 cm for the ventricular diameters (1.7 and 1.9 times); ± 5.2 for E/e' and ± 6.9 cm for the LVOT VTI (1.7 times each); ± 16.6 percentage points for LVEF (3.3 times); and ± 22 mL/m² for LAVI (2.8 times). These widths, rather than the correlations, define what the output can currently be used for: AV Vmax and PASP are estimated to within about one decision step, whereas an LVEF or LAVI estimate cannot on its own separate adjacent clinical categories.

Table 2

Agreement in physical units and prediction-interval width for the eight reported indicators (protocol-acquired panel, post hoc, chosen for complete denominators).

Indicator (unit)	n	Bias (95% CI)	Limits of agreement	Slope, predicted on measured (95% CI)	90% prediction-interval half-width	Half-width ÷ tolerance
LVEF (percentage points)	1,502	+0.6 (0.0 to +1.2)	-18.7 to +19.9	0.15 (0.13–0.17)	± 16.6	3.3
AV Vmax (m/s)	1,496	-0.04 (-0.07 to -0.02)	-0.87 to +0.78	0.25 (0.20–0.30)	± 0.50	1.0
LVOT VTI (cm)	562	-0.21 (-0.64 to +0.19)	-8.85 to +8.42	0.29 (0.23–0.34)	± 6.89	1.7
E/e'	1,494	-0.4 (-0.6 to -0.1)	-8.5 to +7.7	0.15 (0.12–0.18)	± 5.2	1.7
PASP (mmHg)	1,282	-0.6 (-1.1 to 0.0)	-16.9 to +15.7	0.22 (0.19–0.25)	± 12.4	1.2
LVDD (cm)	1,501	-0.03 (-0.06 to 0.00)	-1.10 to +1.04	0.15 (0.13–0.18)	± 0.85	1.7
LVDs (cm)	1,502	-0.04 (-0.08 to -0.01)	-1.32 to +1.24	0.13 (0.11–0.16)	± 0.97	1.9
LAVI (mL/m ²)	1,493	-0.9 (-1.8 to -0.1)	-30.5 to +28.6	0.31 (0.27–0.34)	± 22.4	2.8

Validation in held-out patients

The trained ensemble was frozen after development and applied once, without refitting or renormalization, to the held-out cohort: 118 recordings from 96 patients withheld from development, with no patient overlap, each paired with the nearest echocardiogram within 30 days (median 10 days). The indicator predictions replicated (Fig. 3d). The mean correlation across the reported panel was 0.453, with 0.81 (95% CI 0.70–0.89) for AV Vmax (n = 117), 0.57 (0.35–0.80) for the LVOT VTI (n = 31), 0.49 (0.36–0.62) for LVEF (n = 117), 0.47 (0.32–0.61) for LAVI (n = 118), 0.42 (0.29–0.55) for PASP (n = 108), 0.31 (0.17–0.45) for E/e', 0.29 (0.13–0.47) for the left ventricular end-systolic diameter, and 0.26 (0.12–

0.40) for the end-diastolic diameter ($n = 118$ each); the interval for the LVOT VTI is wide because it was measured in fewer than a third of these recordings. Under the pre-registered panel the same frozen model reached 0.516 on the same recordings. Both values are reported because they differ in direction: the revised panel, adopted for the completeness of its denominator, gives the lower figure of the two.

The held-out mean (0.453) is close to the internal out-of-fold mean (0.409), but the two are not directly comparable and no claim of superiority in the held-out cohort is made. Held-out predictions are averages over the five networks of each seed (correlations computed per seed and then averaged, as internally), whereas each internal out-of-fold prediction comes from a single fold model; the held-out cohort is also smaller and carries a higher burden of disease, which widens the range of the measured values and tends to raise a correlation. What the held-out cohort supports is replication: the same indicators are predicted with similar magnitude in patients no model had seen, by a model that was not touched between the two evaluations. Figure 3d places the two side by side for that purpose only.

How far the reference echocardiogram may lag the recording was examined in a larger and more loosely paired version of the same held-out population: 240 recordings from 221 patients in which the reference preceded the recording by a median of 56 days. With the same 25 networks and the same reporting panel, the mean correlation was 0.456 (95% patient-clustered bootstrap interval 0.39–0.52) in the 97 recordings paired within 30 days, 0.320 (0.20–0.43) in the 44 pairs at 31–90 days, and 0.363 (0.27–0.44) in the 99 pairs beyond 90 days (Fig. 2b). The pre-registered expectation of a monotone decline across the three strata was not met: the middle stratum, 44 recordings from 41 patients, could not be ordered against either neighbor (difference from the > 90-day stratum -0.04 , -0.19 to $+0.10$). What the data support is a difference between recent and older references: recordings paired within 30 days exceeded those paired beyond 90 days by 0.09 (-0.01 to $+0.21$) and, in a post hoc pooling of the two later strata, all recordings paired beyond 30 days by 0.11 ($+0.01$ to $+0.20$; 0.46 versus 0.36, $n = 97$ and 143), with the same contrast under the pre-registered panel (0.494 versus 0.346). Two readings are compatible with this pattern and the data do not separate them: genuine hemodynamic change over the months between reference and recording, or a difference between the patients who receive a same-visit echocardiogram and those who do not. Either way an older reference is a weaker standard, which is why training was restricted to same-day pairs and the primary held-out analysis to pairs within 30 days. Re-pairing the same patients against their nearest study within 30 days produced the 118-recording set used above at 0.453, essentially the ≤ 30 -day stratum (0.456), so the re-pairing changed which recordings qualify rather than the accuracy of the qualifying ones. It also cost completeness, because the nearer study is often an abbreviated follow-up: among the 100 recordings common to both pairings, aortic valve area, the dimensionless velocity index and the LVOT VTI were available in 78, 90 and 80 under the original pairing but in only 20, 21 and 24 under the re-pairing. This trade-off between recency and completeness is one reason the reported panel was moved onto protocol-acquired measurements, which survive the re-pairing intact.

Measurement repeatability and prediction intervals

In 152 patients the recording was repeated during the same clinical visit, yielding 160 additional recordings and permitting a test–retest analysis. Adding these repeats to training in a pre-registered sensitivity refit on the earlier data build did not change performance ($\Delta - 0.002$, 95% CI $- 0.016$ to $+ 0.011$ on the fixed evaluation rows), consistent with repeats contributing signal without contributing labels. Their value lies elsewhere. Between two recordings of the same patient minutes apart, the predicted indicators agreed at a mean correlation of 0.815 (0.72 for LVEF to 0.89 for E/e' across the reported panel; 149 patients with a valid pair in the development cohort; the remaining three lie outside it). The reproducibility of the reference standard was estimated on the same principle, from patients in whom echocardiography had been repeated, giving 0.90 for LVEF, 0.92 and 0.89 for the left ventricular systolic and diastolic dimensions, 0.77 for E/e' and 0.73 for the pulmonary artery systolic pressure. Because an observed correlation is attenuated by the unreliability of both variables, the attainable ceiling is the geometric mean of the two reliabilities rather than the reproducibility of the prediction alone; for the five indicators with cohort-measured reproducibility this places the ceiling at 0.79 to 0.84, and for the panel mean at 0.83 to 0.85, against which the observed 0.409 represents approximately half (48–49%; Fig. 4d; Table 3). The ceiling uses the cohort's own echocardiographic reproducibility for the indicators with at least thirty repeated pairs; for LAVI, AV Vmax and the LVOT VTI, for which fewer pairs were available or the source records contained no corresponding field, published inter-observer reliabilities were substituted and the ceiling is given as a range. Correcting for both noise sources leaves a disattenuated correlation of 0.37 to 0.61 depending on the indicator (Table 3), so the residual is not accounted for by measurement noise on either side: it is what the signal does not contain. These corrected quantities are reported for interpretation only and are not performance estimates. This also answers the practical objection that a repeated measurement might give a different answer.

To convey per-patient uncertainty, split-conformal 90% prediction intervals were calibrated on the out-of-fold residuals of each indicator. In the held-out cohort, coverage was 89.3% among patients without moderate-or-greater valvular disease but 75.1% among those with it, a consequence of the higher disease prevalence of that cohort interacting with shrinkage of extreme predicted values³³, giving 82.9% overall. Recalibrating the interval widths on approximately half of the held-out cohort restored coverage to 90.8% on the held-out remainder, motivating site-level recalibration, using on the order of one hundred labeled recordings, as a standard deployment procedure analogous to periodic calibration of a measuring instrument. The half-widths of the intervals are listed in Table 2.

Table 3

Reliability of the prediction and of the reference standard, attainable ceilings, and disattenuated correlations for the reported indicator panel (protocol-acquired, post hoc, chosen for complete denominators).

Indicator	r	Reproducibility of prediction	Reproducibility of reference (pairs)	Ceiling	Fraction attained
LAVI	0.521	0.822	0.91–0.97 (literature)	0.86–0.89	58–60%
AV Vmax	0.505	0.865	0.93–0.99 (literature)	0.90–0.92	55–56%
LVOT VTI	0.492	0.842	0.78–0.94 (literature)	0.81–0.89	55–61%
PASP	0.440	0.864	0.730 (71)	0.794	55%
LVEF	0.339	0.719	0.899 (83)	0.804	42%
E/e'	0.338	0.888	0.771 (79)	0.827	41%
LVDd	0.334	0.781	0.893 (80)	0.835	40%
LVDs	0.305	0.741	0.915 (80)	0.823	37%
Mean	0.409	0.815	—	0.83–0.85	48–49%

Literature reliabilities (Methods): LAVI, inter-observer intraclass correlation 0.947 (95% CI 0.909–0.969)³⁴; AV Vmax, inter-observer coefficients of variation of 6% (95% CI 4–9%)³⁵ and 4.0%³⁶; LVOT VTI, coefficient of variation 12%³⁵ and intraclass correlation 0.939³⁴. Coefficients of variation were converted to reliabilities with the development-cohort mean and SD; where two sources were available the range spans their estimates, and for LAVI it is the confidence interval of the single source. Ceiling, geometric mean of the two reliabilities; fraction attained, r divided by the ceiling; n as in Table 2.

Output form, an illustrative referral rule, and subgroups

Because the predictions are compressed toward the cohort mean, the intended output is a point estimate accompanied by its 90% prediction interval rather than a category, and guideline cut-offs are not applied to the point estimate. An affine recalibration of predicted on measured values would restore the slope to unity but would widen the interval by the same factor; it adds no information and was not applied. To show what the output can do in a decision, one referral rule was fixed in a dated analysis plan before it was evaluated, after the primary estimates had been locked (Methods), and applied on the out-of-fold predictions with nested operating points (Methods): refer for echocardiography when the predicted AV Vmax or the predicted PASP exceeds the value that captures 90% of the recordings with a measured AV Vmax of 3.0 m/s or more, or a measured PASP of 40 mmHg or more, in the other four folds. Among 1,279 recordings with both measurements in this already-referred population (120 positive,

9.4%), this sensitivity-first OR rule reached a sensitivity of 0.95 (95% CI 0.91–0.98) at a specificity of 0.27 (0.24–0.30), a negative predictive value of 0.98 and a referral rate of 75% (Supplementary Table S3); the pre-specified expectation of a specificity between 0.35 and 0.55 was not met. The operating points fell at predicted values of about 1.3 m/s and 25 mmHg, roughly half the guideline cut-offs, which is the compression made concrete. In decision-curve terms the rule had a higher net benefit than referring every patient at every threshold probability, but a positive net benefit only below a threshold probability of about 0.12. The output therefore functions as a rule-out aid, sparing one echocardiogram in four at 95% sensitivity, and not as a rule-in test, and its reach is limited to the lesions that these two indicators reflect (Discussion); the PASP term alone was more efficient (sensitivity 0.91, specificity 0.49, referral rate 54%) and is the form to prefer. Transported unchanged to the held-out cohort (operating points fixed at the mean of the fold values, 1.32 m/s and 24.6 mmHg; 107 recordings with both measurements, 31 positive), the OR rule gave a sensitivity of 0.97 at a specificity of 0.46 and a referral rate of 66%, and the PASP term alone 1.00, 0.54 and 54%; these held-out operating characteristics are descriptive, rest on 31 and 17 positive recordings without bootstrap intervals, and the higher specificity reflects the wider spread of measured values in that cohort (Supplementary Table S3).

Accuracy was similar across patient subgroups (Supplementary Table S4). The reported-panel mean was 0.35, 0.37 and 0.37 across age tertiles (63 years or younger, 64–72, older than 72), 0.34 in women and 0.36 in men, 0.41, 0.40 and 0.42 across body-mass-index tertiles, and 0.39, 0.42 and 0.40 across tertiles of the number of usable cycles. The values within age and sex strata lie below the pooled 0.409 because the variance that the demographic offset explains is removed within a stratum, not because any stratum is predicted worse; AV Vmax was best predicted in the oldest tertile (0.53 against 0.38 and 0.41), where aortic stenosis concentrates. Cardiac rhythm was not recorded, so the coefficient of variation of the pulse-wave cycle length served as a proxy for irregularity (Supplementary Table S5). A coefficient above 0.15 marked 25% of development recordings, far more than the expected prevalence of atrial fibrillation, so the proxy also captures respiratory variation, premature beats and imperfect delineation, and the accuracy of these recordings was not lower (0.42 against 0.39). A coefficient above 0.25, closer to a fibrillating rhythm, marked 2.5% of recordings, in which the panel mean was 0.33 (95% CI 0.21–0.42) against 0.41, an estimate too imprecise to settle whether an irregular rhythm degrades the prediction; in the held-out cohort the 28 recordings above 0.15 gave 0.30 against 0.43. Finally, because the discretionary Doppler quantities are measured only when a lesion is suspected, their availability was tabulated against the graded severity of the relevant valve (Supplementary Table S6). The regurgitant indices were measured in none of the 241 recordings without regurgitation, in 7–32% of those with mild regurgitation and in 68–92% of those with moderate or severe regurgitation, and the aortic valve area in 31% of recordings without and in every recording with aortic stenosis. The accuracy reported for these indicators therefore describes the ordering of severity among patients in whom a lesion had already been seen; no imputation of normal values for unmeasured indicators was attempted, because the unmeasured majority with mild disease are not normal.

Discussion

A single-site, single-device, 15-second recording of heart sounds and the finger pulse supported quantitative prediction of echocardiographic indicators in a consecutive clinical population. No hardware was added and no new modality was introduced; an instrument that is already at every bedside was repurposed for a task it was not designed for, and the boundary of that repurposing — the quantities the signal does not carry, the compression of its estimates and the interval that must accompany them — was measured rather than assumed. The auscultatory signals added + 0.185 in mean correlation beyond age and sex across the pre-registered panel and + 0.145 across the protocol-acquired panel, and the predictions were reproducible on repeat measurement well beyond their accuracy against the reference standard. The central design decision was to predict an interpretable panel of continuous indicators rather than a diagnostic probability. It was made for reusability and auditability, and it is what allows the contribution of sound to be isolated for each indicator rather than aggregated into a single score. Several pre-registered architectural extensions failed to improve on this structure, which places the current bottleneck in the information content of a single-site signal and in the size of the labeled cohort rather than in model capacity.

These figures are not comparable to the high accuracies reported for case–control designs, and the differences from the most cited stethoscope study¹⁸ are instructive. That algorithm produces a binary output from recordings at four auscultation positions using an acoustic pipeline that applies a 30 Hz high-pass filter and downsamples to 2 kHz; it was developed in an enriched case–control sample, its authors noting that specificity is likely overstated relative to a general population, and 12.7% of recordings were excluded for inadequate signal. The system reported here instead yields a quantitative hemodynamic profile from a single 15-second recording, keeps the acoustic channel unfiltered through pulse-wave-anchored segmentation, and is evaluated on a consecutive cohort that retains mild disease and low-quality recordings, which are flagged rather than excluded. The spectrum effect therefore operates unfavorably here, and diagnostic performance is known to depend on the composition of the tested population^{37,38} — a dependence that enriched designs conceal. Whether the electrocardiogram, which is at least as accessible, would serve the same purpose was not tested here. The attribution pattern suggests that the two signals are complementary, because the auscultatory increment is concentrated in quantities generated by flow, which the electrical signal does not observe, but this remains a hypothesis. For a single target, specialization remains preferable where it is available: the PASP correlation reported here (0.44) is below that of a dedicated pulmonary-hypertension model trained on one target with multi-site recordings (0.64)¹⁹, whereas the present model reports a whole panel from one site and one recording.

The attribution analysis carries a clinical message: the value of acoustic AI is concentrated in regurgitation. Age and sex alone already stratify aortic valve area substantially ($r = 0.43$), but they carry no information at all about regurgitant orifice size — the auscultatory signal supplies it. A risk model built on demographics would therefore be blind to regurgitant lesions, which is exactly where the net contribution of the device peaks. This aligns with the guideline emphasis on auscultation as the essential first examination¹, and it identifies the indicators for which an acoustic measurement, rather

than a demographic table, is the informative input. Read across the whole target set, the ordering of accuracy follows the ordering of mechanism: the aortic-flow indicators, whose murmur genesis is most direct, are predicted best; the regurgitant indices derive essentially all of their accuracy from sound, their demographic baseline carrying no usable information; and wall thickness, which generates no sound, receives nothing from it. Both the reach and the boundary of the measurement are set by the signal it is read from. The knockout analysis adds the reason for recording two signals at once: the sound carries the flow and pressure indicators and the pulse wave carries stroke volume and chamber size, so neither channel alone would report the panel.

The reliability analysis bounds where improvement must come from. Two recordings of the same patient, minutes apart, yield predictions agreeing at 0.815, and two echocardiograms of the same patient agree at 0.73 to 0.92 depending on the indicator; the resulting ceiling of 0.79 to 0.84 is roughly twice the observed 0.409 (Fig. 4d). Measurement noise on either side is therefore not the binding constraint, and neither is it negligible: approximately half of the gap to unity is attributable to it, and the remainder to how much of the echocardiographic state is recoverable from fifteen seconds of sound at one chest position together with the simultaneous finger pulse, and to how many labeled examples exist for the indicators that are rarely measured. This moves the development priority away from network architecture and toward data. The baseline ladder supports the same reading from the other side: ridge regression on hand-crafted descriptors reaches 0.258 against the model's 0.373 on the earlier build, so the learned representation does contribute a real margin — concentrated in the regurgitant and pulmonary-pressure indicators — but no rung of the ladder approaches the repeatability bound. Retraining supports the same reading: increasing the supervision available per patient, from one-third indicator coverage to complete coverage, left internal accuracy statistically unchanged and held-out accuracy unchanged as well, a pattern more consistent with an information ceiling than with a shortage of training signal. What remains actionable is the composition of the cohort itself, above all quantitative Doppler measurement of regurgitant lesions in patients without visible disease, whose near-absence leaves the regurgitation indicators supervised almost entirely by positive cases.

Several constraints bound these claims. All data originate from one center, one device, and one recording position (chest auscultation site 2, with the pulse wave taken from the finger); the pulmonic area favors the pulmonary-pressure and aortic-flow indicators and may disadvantage mitral lesions, and multi-site recording is the next step. Inter-institutional generalization is untested, and the held-out cohort, although it contains no patient seen in development, shares the recording period, the recording environment and the echocardiography laboratory, so it tests generalization to unseen patients, not transportability to another site or period. The cohort is also a single population of Korean adults, so generalization across populations, body habitus, and recording environments is untested. No formal power calculation was performed; enrollment was consecutive over a fixed period and confidence intervals are reported throughout. The reporting panel was revised after the per-indicator results had been reviewed and is post hoc, with the pre-registered panel reported alongside it throughout; the exclusion of the wall-thickness targets from the panel was likewise informed by their measured accuracy, and both targets are shown in Figs. 2a and 3 so that the negative result remains visible.

Indicators outside the reported panel rest on fewer observations and are reported as secondary rather than as validated outputs. Predicted values are compressed toward the cohort mean (slope of predicted on measured 0.13–0.31; Table 2), as regression to the mean dictates, so published guideline cut-offs must not be applied directly to predicted indicators; correlation-based metrics are invariant to this affine compression and therefore do not reveal it, which is why agreement in physical units, prediction intervals and threshold behavior are reported separately. The attainable-ceiling analysis rests on the assumptions stated in Methods and on published reliabilities for three indicators, and its corrected values are descriptive. Accuracy was lower against reference echocardiograms older than 30 days, so applications requiring current hemodynamic values need recent reference measurements. Accuracy was similar across sex, age and body-mass-index strata (Supplementary Table S4), but cardiac rhythm was not recorded, and the pulse-interval proxy used for it cannot separate atrial fibrillation from benign variability. The referral rule reported here is illustrative: its operating points were chosen on this cohort, and its usefulness depends on a referral threshold that a deploying site must set. Because the rule rests on the two indicators that the model estimates most reliably, AV Vmax and PASP, its reach is confined to aortic stenosis and to lesions that raise pulmonary pressure. Early or low-flow mitral and aortic regurgitation, and other valve lesions that do not elevate pulmonary pressure, lie outside what these two indicators can signal, so the rule is structurally unable to screen for high-risk valvular heart disease in general and should not be read as such. Its positive state, an AV Vmax of 3.0 m/s or more or a PASP of 40 mmHg or more, is a pragmatic composite rather than a clinical definition of high risk; in patients with more than one lesion, the indicator that triggers referral need not be the lesion that determines management, and the indicator specificity of a positive result is correspondingly limited. Finally, conformal coverage was not maintained under the prevalence shift of the held-out cohort without local recalibration, which should be treated as a required deployment step rather than an optional refinement.

The predicted indicator panel is the substrate, not the end point. Downstream modules that consume it — screening for valvular disease and assessment of heart failure — are under development on the same cohort and will be reported separately, with expanded positive case counts, since the number of moderate-or-greater cases presently available bounds what any such module can be shown to achieve. Keeping the two layers separate is also a methodological commitment: the prediction model is evaluated as a measurement instrument, against continuous references and with repeatability and interval coverage reported, while diagnostic claims are deferred to the reports that can support them. Framed as quantification of an existing examination, the contribution of this study is narrower and more checkable than a diagnostic claim would be. An instrument that is already at the bedside was shown to carry recoverable quantitative information about the hemodynamic state, the amount of that information was measured indicator by indicator against the modality that defines the ground truth, and the boundary at which it vanishes was located and reported. Whether such a device is useful depends on where it is used: at the point of first contact, where echocardiography is absent, a triage-level estimate of the quantities that determine referral is a different proposition from the same estimate beside a scanner.

Methods

Study design and population

Consecutive adult patients undergoing clinically indicated TTE at the Division of Cardiology, Keimyung University Dongsan Hospital (Daegu, Republic of Korea), were enrolled between July 2025 and August 2026. The dataset was frozen on 18 August 2026, the date of the last enrolled recording, and all analyses reported here were completed by 13 September 2026. The Institutional Review Board of Keimyung University Dongsan Medical Center approved the study (DSMC IRB 2025-07-068-003), and all participants provided written informed consent. Training used only recordings paired with a same-day echocardiogram (1,502 recordings from 1,339 patients), so that the reference standard reflects the same hemodynamic state as the recording; the dependence of accuracy on the pairing interval is examined in Results. The held-out validation set comprised patients whose recordings had no same-day echocardiogram; these patients were excluded from development in their entirety, with no overlap of patient identifiers. Two pairings of this cohort are reported. The primary set contains 118 recordings from 96 patients, each paired with the nearest echocardiogram within 30 days (median 10 days, interquartile range 7–17). A larger and more loosely paired set of 240 recordings from 221 patients (median recording–echocardiography interval 56 days, interquartile range 14–126) is used only for the interval-stratified analysis, which was pre-specified; 100 recordings are common to both. Table 1 summarizes the cohort demographics, the distribution of each indicator, and the severity spectrum of the four graded lesions. In the table, age, sex, body-mass index, and body surface area are summarized at the patient level, on the basis of the first recording of each patient, whereas all other quantities are summarized at the recording level; global longitudinal strain is stored as a signed value and is reported by its magnitude, |GLS|. Because quantitative Doppler measurements are performed at the discretion of the echocardiographer rather than by a fixed protocol, indicator coverage varies, and denominators are reported per indicator. Severity grades (0, none; 1, mild; 2, moderate; 3, severe) describe the disease spectrum of the consecutive cohort and are not model targets. Cardiac rhythm was not available as a structured field; no recording was excluded for arrhythmia, and cycles that the pulse-wave delineation could not place were simply not counted. Recordings contributed a median of 13 usable cycles (interquartile range 11–15); two development recordings and one held-out recording yielded fewer than five cycles and were flagged as low-confidence but retained. Heart rate during the recording, derived from the pulse wave, is given in Table 1; the heart rate recorded at echocardiography was available in fewer than 90% of studies and is not tabulated.

Prediction targets and physiological rationale

This work rests on the premise that the reported indicators can be derived from heart sounds and the finger pulse, because both signals reflect the same hemodynamic phenomena that these indicators quantify. Turbulent flow passing through a stenotic or regurgitant orifice produces murmur energy, whose intensity, spectral content, and timing correlate with jet velocity, gradient, and orifice size^{11–14}; valve closure generates the sounds S1 and S2, with their amplitudes and intervals indicating contractile force³⁹, ejection duration²⁹, and pulmonary artery pressure⁴⁰; additionally, the peripheral arterial pulse is a pressure-wave profile shaped by stroke volume, ejection dynamics, and vascular load^{27,28,41}. Table 4

lists, for each category of panel indicators, the hypothesized acoustic and pulse-wave correlates and the physiological mechanisms linking them to the recorded signals. Its rows represent mechanism axes rather than individual indicators; panel members sharing a generating mechanism are grouped accordingly, and the two tiers of the panel, namely the flow, pressure, and function quantities and the chamber measures of chronic hemodynamic load, are distinguished. This mechanistic criterion, according to which a parameter is either produced by blood flow or modified by chronic hemodynamic load, defines the scope of the reported panel. The table records the physiological rationale by which the panel was delimited; it is not an empirical, per-feature attribution. None of the descriptors listed there are used as input features; rather, the network processes the complete log-mel representation of the unfiltered PCG together with pulse-wave features. Empirical attribution was assessed separately, through the architectural separation of the demographic offset from the auscultatory residual (Results). Because the PPG records a peripheral arterial pulse at the fingertip, the classical carotid-pulse signs cited in the table, namely *pulsus parvus et tardus* and the wide pulse pressure with rapid diastolic runoff, are physiological analogs rather than directly measured equivalents; the finger pulse is shaped by the same ejection dynamics but is delayed and damped by peripheral transmission. Measures of chamber remodeling constitute a separate category, as they produce no intrinsic sound and are inferred solely from the persistent signature of loading lesions.

Table 4

Hypothesized physiological correspondence between the predicted echocardiographic indicators and the recorded heart-sound (PCG) and pulse-wave (PPG) signals.

Indicator group (panel members)	Acoustic correlate (PCG)	Pulse-wave correlate (PPG)	Mechanism
Flow, pressure and function indicators			
Aortic-valve flow (AV Vmax, mean PG, AVA, DVI)	Systolic ejection murmur: intensity, frequency content and late systolic peaking track jet velocity and gradient. Spectral analysis of the murmur alone was shown to grade stenosis severity ^{11–14} .	Slowed, low-amplitude systolic upstroke, the peripheral counterpart of pulsus parvus et tardus.	Turbulent transvalvular jet; a fixed orifice obstructs and delays ejection ² .
Outflow and stroke volume (LVOT VTI)	Energy of the ejection sound and the duration of the systolic interval ²⁹ .	Pulse amplitude and pulse area track the volume ejected per beat ²⁷ .	The velocity–time integral is the integral of ejected flow over one beat.
Regurgitant lesions (Reg VC, Reg PISA, RegVol)	Holosystolic (mitral, tricuspid) and early-diastolic (aortic) murmur energy scales with regurgitant orifice size; murmur audibility is the dominant acoustic cue.	Wide pulse pressure and rapid diastolic runoff alter the waveform decay in aortic regurgitation.	Turbulence of the regurgitant jet scales with effective regurgitant orifice area ³ .
Pulmonary pressure (PASP)	Intensity of the pulmonic component of the second sound (P2) rises with pulmonary artery pressure, and the recording position (site 2, the pulmonic area) is where P2 is loudest; splitting of S2 narrows ⁴⁰ . Heart-sound analysis was shown to detect pulmonary hypertension ¹⁹ .	Timing shifts attributable to right ventricular load.	Abrupt pulmonic valve closure against an elevated pulmonary artery pressure.
Systolic function (LVEF, GLS)	First-sound intensity tracks the rate of pressure rise (dP/dt) ³⁹ ; the systolic time intervals lengthen as contractility falls ²⁹ .	The pulse upstroke supplies the arrival end of the pre-ejection and ejection intervals, which the sound–pulse pair fixes jointly.	Valve-closure force and ejection timing are direct functions of the contractile state.
Filling and diastolic function (E/e')	Low-frequency diastolic sounds (third and fourth heart sounds) accompany elevated filling pressure and a stiff ventricle ⁴² ; the internal structure of the diastolic interval changes.	Diastolic decay and the cycle-length structure of the pulse train.	Rapid deceleration of early filling into a non-compliant ventricle.
Chamber measures of chronic hemodynamic load			

Indicator group (panel members)	Acoustic correlate (PCG)	Pulse-wave correlate (PPG)	Mechanism
Flow, pressure and function indicators			
Chronic-load remodeling (LAD, LAVI, LVDD, LVDs)	No sound of its own; reached only through the persistent acoustic signature of the loading lesions of the rows above.	Irregularity of the pulse rhythm accompanying atrial remodeling.	Chamber dilatation represents the time-integral of the hemodynamic load measured by the rows above; these four indicators are therefore reported as a separate tier ⁴³ .

Heart-sound and pulse-wave acquisition

Physiological signal acquisition was performed using the MUST electronic stethoscope (Clairaudience Co., Ltd., Daegu, Republic of Korea), with the auscultation head capturing the PCG from a single chest position, auscultation site 2 at the left second intercostal space (pulmonic area). Concurrently, a PPG sensor embedded on the opposite face of the same device recorded the pulse wave from the fingertip resting upon it (Fig. 6). The acoustic channel was sampled at a frequency of 16 kHz, and the pulse-wave channel at 105 Hz. Each measurement lasted 15 seconds. All measurements were performed immediately prior to the echocardiographic examination of the same visit, in a quiet examination room maintained at approximately 25°C and 50% relative humidity. Before recording, each participant was instructed on the purpose of the study and the operation of the device and then rested for two minutes; the recordings were subsequently obtained with the participant seated in a comfortable position, with outer garments removed and the auscultation head applied over the innermost layer of clothing. No murmur-attenuating noise reduction was applied to the acoustic channel; instead, the pulse wave was utilized as the timing reference for all subsequent segmentation. The device was purpose-built around this pulse-wave-anchored acquisition principle.

Reference standard and labels

The reference standard was the comprehensive TTE examination^{43–46}, performed and interpreted by board-certified cardiologists at Keimyung University Dongsan Hospital, who were blinded to the device output. Transthoracic echocardiography was performed using commercial ultrasound systems (GE Vivid E95, GE Vingmed Ultrasound, Horten, Norway; Philips Epiq 7C, Philips Healthcare, Andover, MA). Indicator values were entered into the electronic medical record by the examining cardiologists and cross-verified by four cardiologists of the Division of Cardiology as part of the labeling workflow. Twenty-seven indicators were extracted programmatically from the structured electronic medical record and the corresponding report text, with no manual transcription, and all twenty-seven were used as training targets. They comprise the hemodynamic quantities determined by flow, pressure, and function; the chamber dimensions remodeled by chronic hemodynamic load; and the static structural or transcription

attributes — left ventricular septal and posterior wall thickness, calcification of the aortic and mitral leaflets, thickening of the aortic, mitral, and tricuspid leaflets, dilatation of the mitral and tricuspid annuli, and the number of regurgitant valves — together with the aortic regurgitant pressure half-time, which was measured in too few studies to be reported; these eleven targets serve solely as auxiliary supervision and are not reported as outputs. Training was not narrowed to the reported panel: earlier attempts to restrict the target set did not improve accuracy, and the full set is retained so that the auxiliary targets remain available as mechanistic controls. Because several regurgitant indicators are recorded per affected valve, the attribution of each vena contracta and PISA measurement to its valve was re-adjudicated before any model was developed.

Preprocessing and signal feature extraction

Cardiac cycles were delineated from the simultaneously acquired pulse wave, extending the PCG–PPG fiducial-fusion principle²⁶ from S1/S2 identification to full cycle delineation, and keeping the acoustic channel free of the filtering that envelope-based acoustic segmentation would require (Fig. 7). Segmentation based solely on the acoustic channel requires band-limiting of the same signal that is to be analyzed, whereas anchoring on the pulse wave confines that band-limiting to the timing path alone, namely a 25–150 Hz envelope used only to place S1 and S2, so that the signal presented to the model retains its full murmur content. In a representative recording (Fig. 7a), the systolic peaks are detected on the pulse wave and every cycle boundary is derived from it. Within each cycle (Fig. 7b), the systolic peak is located on the pulse wave alone, S2 and S1 are then placed on the band-limited acoustic envelope within physiologic windows preceding the peak, and the cycle boundary is the pulse-wave zero-crossing preceding S1. Each cycle contributes a fixed 1.0-second window from which the 32×98 log-mel spectrogram is computed (Fig. 7c); predictions are made per cycle and averaged over the recording. Morphological and timing features were computed per cycle from the PPG. Recordings with fewer than five usable cycles were flagged as low-confidence at inference. All normalization statistics were fitted on training folds only. The log-mel representation of cycle k is

$$M_k(m, \tau) = \log \sum_f H_m(f) |X_k(f, \tau)|^2$$

1

where $X_k(f, \tau)$ is the short-time Fourier transform of the unfiltered PCG of cycle k , and H_m is the m -th mel filter ($m = 1 \dots 32$, 98 time frames τ).

Model architecture

At the core of the echocardiographic indicator prediction model is a deep multimodal network, denoted g , which is trained on the paired physiological-signal and echocardiographic data and maps the PCG and the PPG of each cardiac cycle to the echocardiographic targets. The network comprises two modality-specific encoders, namely an acoustic encoder operating on the per-cycle log-mel spectrogram and a pulse-wave encoder operating on the pulse-wave cycle and its hand-crafted descriptors, followed by a

shared fusion trunk and a linear indicator head that jointly regresses all echocardiographic targets; the complete architecture is illustrated in Fig. 8. Every usable cycle of a recording is encoded independently: the per-cycle inputs are the log-mel spectrogram of the unfiltered PCG, the 256-sample pulse-wave cycle, and two hand-crafted descriptor sets, the 55-dimensional cycle feature vector and the 24 dicotic-notch descriptors; the encoders map these to a 256-dimensional acoustic embedding and a 128-dimensional pulse-wave embedding, and the concatenated 384-dimensional representation passes through the shared trunk to the indicator head. The head jointly regresses 27 targets under the masked loss: 16 echocardiographic indicators, of which eight form the reported panel and eight are reported as secondary, and 11 auxiliary targets that supervise the shared representation but are not reported. The measured same-day echocardiogram enters only as the supervision target of that loss and is never an input at inference. Per-cycle predictions are averaged over the recording, and the deployed model averages 25 networks (five seeds by five folds). To this deep component a linear demographic offset, denoted f , is added, defined as a ridge-regularized model of age and sex fitted on the training folds and subsequently frozen, so that each final prediction is the sum of g and f , and g is optimized against the portion of each echocardiographic value that the demographics leave unexplained. This separation serves the attribution of the auscultatory contribution, measured as the performance difference between the comprehensive model and the offset alone, rather than constituting the primary predictive pathway. The constituent modules of g are described below in the order of the data flow. The acoustic encoder converts the 32-band by 98-frame log-mel spectrogram of each cycle into a 256-dimensional embedding. It consists of four two-dimensional convolutional blocks, each composed of a convolution, batch normalization, and a GELU activation, with 32, 64, 128, and 128 channels in sequence, and 2×2 max-pooling follows each of the first three blocks. The resulting feature map is reduced by adaptive average pooling to a 4×4 grid and flattened to 2,048 dimensions, which two fully connected layers subsequently reduce to 512 and then to 256 dimensions. The pulse-wave encoder produces a 128-dimensional embedding from three inputs. A one-dimensional convolutional branch processes the 256-sample pulse-wave cycle through three blocks with kernel sizes of 7, 5, and 3 and with 32, 64, and 128 channels, respectively; adaptive average pooling to eight positions yields a 1,024-dimensional vector, which two fully connected layers reduce to 256 and then to 64 dimensions. In parallel, two multilayer perceptrons encode the hand-crafted per-cycle descriptors. The first reduces the 55-dimensional cycle feature vector, which comprises 25 PCG time-domain shape descriptors, such as S1/S2 amplitude and murmur duration fraction, and 30 pulse-wave descriptors (18 per-cycle morphology and timing descriptors and 12 recording-level descriptors replicated across the cycles of a recording), to 64 and then to 32 dimensions; the second maps the 24 dicotic-notch descriptors through two 32-dimensional layers. Acoustic information in the time domain is thereby incorporated into the model through these descriptors in addition to the spectrogram. The outputs of the two encoders are concatenated into a 384-dimensional representation, which the shared trunk, consisting of two fully connected layers with 256 and 128 units, processes before the linear indicator head. The head outputs are bounded by $4 \cdot \tanh(u)$ within the z-normalized target space, enabling the joint regression of all 27 echocardiographic targets. Per-cycle predictions are averaged within each recording. Targets undergo z-normalization based on training-fold statistics (with right-skewed targets log-transformed). The prediction for indicator i is calculated as

$$\hat{y}^{(i)} = f^{(i)}(a, s) + \frac{1}{K} \sum_{k=1}^K g^{(i)}(M_k, p_k), \quad g^{(i)} = 4 \tanh(u^{(i)})$$

2

where a and s denote age and sex, M_k and p_k the mel-spectrogram and PPG features of cycle k , K the number of usable cycles, and u the pre-activation network output; the tanh bounding confines the residual to ± 4 standard deviations of the target. The model is trained with a masked mean-squared error computed only over the indicators actually observed for each patient, with uniform target weighting.

$$\mathcal{L}_A = \frac{\sum_i m_i (\hat{y}^{(i)} - y^{(i)})^2}{\sum_i m_i}$$

3

where m_i represents the observation mask, and both $\hat{y}$ and y reside within the z-normalized target space. The targets outside the reported panel are retained in the loss via auxiliary supervision of the shared representation but are not presented as outputs. Owing to the masked nature of the loss, targets not present in a patient's data do not contribute gradients, ensuring that the broader target set does not compete with the primary panel for model capacity. Training on the 16 indicators alone was compared with training on all 27 targets in a pre-registered experiment; as the comparison found no material difference between the two, the broader set was retained.

Training and evaluation protocol

Networks were trained with the AdamW optimizer (learning rate 0.0002 with cosine annealing) for up to 40 epochs, with training stopped after 12 consecutive epochs without improvement of the validation-fold score (training ended after 14 to 35 epochs, median 22), using patient-grouped five-fold cross-validation with five random seeds, for 25 models in total. Two evaluation conventions were fixed in advance: absolute performance is taken from the checkpoint at the final training epoch, and out-of-fold predictions are pooled before correlations are computed. The validation fold therefore determines when training stops but not which checkpoint is kept; validation-based checkpoint selection and per-fold averaging were avoided, because preliminary analyses showed that both inflate the resulting scores. Predictions include the demographic offset unless otherwise stated. The architecture, optimizer settings, epoch budget and stopping rule were fixed in earlier development builds on smaller cohorts, before the present cohort was assembled, and were not tuned on the evaluation rows reported here; no per-indicator tuning was performed. Body size (body-mass index, height and weight) was not included in the demographic offset: in a development experiment it raised the training-fold fit only marginally (mean $R^2 + 0.02$) and was not adopted. Two reporting panels are used throughout, and the distinction between them is material. The pre-registered panel (AV Vmax, aortic valve area, the dimensionless velocity index, the LVOT VTI, regurgitant vena contracta width, regurgitant PISA radius, PASP, and LVEF) was fixed in the

analysis code before any per-indicator result was examined, by a coverage criterion of at least 30% availability, and is recorded in a single dated target-definition module from which both training and evaluation scripts import. The panel reported as primary here (LVEF, AV Vmax, the LVOT VTI, E/e', PASP, left ventricular end-diastolic and end-systolic diameters, and the left atrial volume index) was subsequently proposed by the collaborating cardiologist on clinical grounds, after the per-indicator results had been reviewed. That revision is therefore post hoc and is declared as such: it was adopted because every indicator in it is acquired by protocol in essentially every study, whereas half of the pre-registered panel is quantified only when a lesion is already suspected, and not because it improves any reported figure — it lowers the held-out value (0.453 against 0.516) and reduces the measured contribution of auscultation (+ 0.145 against + 0.185). Both panels are reported side by side wherever either appears. The exclusion of the two wall-thickness targets from the reported panel was likewise informed by their measured accuracy, and both are shown in Figs. 2a and 3 so that this negative result remains inspectable. The hand-crafted-descriptor baseline reported alongside the main model is a ridge regression ($\alpha = 10$) whose 81 inputs are the per-recording medians of the descriptors described in Model architecture together with age and sex; it requires a demographics-only training run that exists only for the earlier data build, so it was fitted on the 1,171 recordings of that build under the identical five-fold patient-grouped split, and is reported as a sensitivity analysis whose absolute values are not interchangeable with those of the primary analysis.

The reported model is itself the product of a corrected supervision build, and its stability against that correction was examined. Relative to the preceding build, in which the left-heart and diastolic targets were available for roughly one-third of recordings, the echocardiographic metadata were re-extracted to complete coverage for those targets, six mis-assigned recording-to-echocardiogram mappings and thirty-five physiologically implausible reference values were corrected, and all twenty-five networks were retrained from scratch. Coverage was verified against the training cache itself, not against the metadata, because an earlier attempt had left the expanded values outside the cache. On the same fixed evaluation rows, accuracy on the pre-registered panel changed by + 0.006 (paired patient-bootstrap 95% CI – 0.003 to + 0.016) and held-out accuracy by – 0.001, with five of the eight indicators marginally higher. Broadening the supervision available per patient therefore neither raised nor destabilized performance.

Missingness and leakage control

Within echocardiographic reports, the availability of an indicator is itself informative; for example, the vena contracta width (VC) is measured predominantly when regurgitation is observed, with 98% availability in positive cases versus 7% in negative cases. Consequently, missingness indicators are considered as label leakage and have been excluded from all models. Cross-validation folds were partitioned based on patient identifiers; an automated audit was conducted to confirm the absence of patient overlap, measurement overlap, and to ensure training-only normalization prior to the acceptance of any run. These safeguards are not incidental: on public heart-sound corpora, retaining duplicate recordings and partitioning at the record rather than the patient level has been shown to drive several

convolutional networks to 100% accuracy, and a model trained on one such corpus transferred to another at chance level once the label definition changed⁴⁷.

Validation in held-out patients

After development, the 25 networks and all normalization statistics were consolidated into a versioned artifact. The complete pipeline, from raw signals to predicted indicators, was then applied once to the held-out dataset. No component was recalibrated and no statistic was recomputed after the held-out labels were introduced. Held-out predictions are the average of the five fold networks of each seed, with correlations computed per seed and then averaged across the five seeds, as internally; each internal out-of-fold prediction, by contrast, comes from the single network of the fold in which the patient was held out, so that a held-out prediction is an ensemble average whereas an internal one is not. The two are therefore presented side by side but not compared as if measured under a single protocol. An automated audit confirmed the absence of patient overlap between the training and held-out cohorts at the identifier level. The out-of-fold and held-out prediction files were generated once, on 22 August 2026, and locked; the reporting-panel revision of 27 August and every analysis dated after it read these files without regeneration.

Repeatability and conformal prediction intervals

Test–retest repeatability was assessed on the first two same-visit recordings of each patient in whom a repeat was acquired in addition to the recording included in the primary cohort (149 patients of the development cohort); the repeats were processed by the identical pipeline. Predictions for both members of each pair were obtained out of fold and averaged across seeds; because all recordings of a patient are confined to the same cross-validation fold, no network contributing to either prediction had access to that patient’s recordings. Agreement was quantified by Pearson’s correlation coefficient (r) for each indicator. Prediction intervals were established by split-conformal calibration^{48,49}: for each indicator, the 90th percentile of absolute out-of-fold residuals defined a symmetric interval. Coverage in the held-out cohort was assessed across the entire cohort, stratified by disease status, and after recalibration of interval widths on a randomly selected half of the held-out patients. Conformal validity presumes exchangeability, a condition violated by the prevalence shift in the held-out cohort; the recalibration measures the impact of that violation and the sample size needed to recover nominal coverage.

The reproducibility of the reference standard was estimated from patients in whom echocardiography had been performed more than once, using the earliest two studies of each patient and the median of repeated measurements of the same quantity within a study. Pairs were available for LVEF ($n = 83$), the left ventricular dimensions ($n = 80$ each), E/e' ($n = 79$) and the pulmonary artery systolic pressure ($n = 71$); for the remaining reported indicators the source records contained no corresponding field or fewer than thirty pairs, and published inter-observer reliabilities were substituted^{34–36}: an intraclass correlation for LAVI and the LVOT VTI, and inter-observer coefficients of variation for AV Vmax and the LVOT VTI, converted to reliabilities with the development-cohort mean and standard deviation as $1 - (CV \times$

mean)²/SD². Provisional bands fixed before the analysis were replaced by these source-derived values on 13 September 2026, when the citation was corrected; the change moved the fraction attained by at most four percentage points (Table 3). Attainable ceilings were computed as the geometric mean of the reproducibility of the prediction and that of the reference standard, and disattenuated correlations as the observed correlation divided by that ceiling. Two opposing biases apply and are not resolved by the present data: the interval between repeated echocardiograms was substantial (median 118 days), so that genuine change is included in the estimate, which lowers it; and systematic error in the model reappears in both recordings of a patient, which raises the estimated reproducibility of the prediction. The two act in opposite directions, and the fraction is therefore reported as a range rather than a point estimate.

Referral rule, subgroup, rhythm, availability and modality-knockout analyses

These analyses were pre-specified in a dated analysis plan (13 September 2026) held in the analysis repository and written after the primary estimates had been locked; they are therefore post-lock, not pre-registered, use the same out-of-fold predictions, are descriptive, and do not alter any primary figure. For the referral rule the positive state was a measured AV Vmax of 3.0 m/s or more or a measured PASP of 40 mmHg or more, among recordings with both measurements; the rule fired when the predicted AV Vmax or the predicted PASP (averaged over the five seeds) exceeded an operating point set, for each fold, at the 10th percentile of the predicted values among the positive recordings of the other four folds, so that each indicator term targeted 90% sensitivity without using the evaluated fold. Sensitivity, specificity, predictive values and referral rate are reported with patient-clustered bootstrap intervals, and net benefit was computed as the true-positive rate minus the false-positive rate weighted by the odds of the threshold probability. Subgroup accuracy was computed within tertiles of age and body-mass index (cut points from the first recording of each patient), by sex, and within tertiles of the number of usable cycles, as the mean over the reported panel of the seed-averaged correlation. As a proxy for rhythm irregularity, the coefficient of variation of the pulse-wave cycle length was computed over the usable cycles of each recording, so that beats rejected by the quality selection are not counted; a value above 0.15 was pre-specified and 0.25 added post hoc. A cleaner proxy would require the pulse-peak series before quality selection, which the current pipeline does not retain. Availability of each discretionary Doppler indicator was tabulated against the grade of aortic stenosis (aortic-flow indices) or the highest grade among aortic, mitral and tricuspid regurgitation (regurgitant indices). For the modality knockout, each of the 25 trained networks was applied to its own validation fold with the acoustic embedding (256 dimensions), the pulse-wave embedding (128 dimensions) or both replaced, for every cycle of every validation patient, by the mean embedding computed over the training patients of that fold; replacement by the mean rather than by zeros keeps the input to the fusion trunk within its training distribution. Per-cycle predictions were averaged and the demographic offset added exactly as in the primary evaluation, and the unmodified condition was required to reproduce the primary out-of-fold predictions to within 10^{-3} (observed maximum difference 3×10^{-6}). The loss attributable to each channel is the correlation of the unmodified condition minus that of the knockout, with patient-clustered bootstrap intervals (500

resamples); the share attributable to the pulse wave is reported only where the two losses are both positive.

Statistical analysis

Differences between the development and held-out cohorts (Table 1) are summarized as standardized mean differences and tested with Welch's t-test for continuous variables, the χ^2 test for sex and the Mann–Whitney U test on the ordinal grade for the four lesions, all two-sided at $\alpha = 0.05$ with the actual P values reported; these comparisons describe the cohorts and play no part in model selection. The agreement between measured and predicted indicators is summarized by the Pearson correlation of out-of-fold predictions pooled within each seed and averaged across the five seeds. For the abnormality analysis, the label is the measured value of the indicator dichotomized at the conventional threshold and the score is the predicted continuous value of the same indicator, so no threshold is applied to a prediction; discrimination is summarized by the AUROC. Tolerance hit rates and mean absolute errors are computed in the physical unit of each indicator after inverting the target normalization, with tolerances fixed a priori at the order of a clinical decision step. Bland–Altman agreement (Table 2; Fig. 5) uses the out-of-fold predictions averaged over seeds in the same physical units: bias is predicted minus measured, the limits of agreement are the bias ± 1.96 SD of the differences, the compression slope is that of the regression of predicted on measured values (1 = no compression), and the 90% prediction-interval half-width is the 90th percentile of the absolute out-of-fold residual, on which the split-conformal intervals are calibrated; the tolerances are those of Fig. 4b. Confidence intervals are two-sided 95% intervals from a patient-clustered paired bootstrap (2,000 resamples; 1,000 for the Bland–Altman bias and slope; 500 for the knockout and subgroup analyses)⁵⁰. Because all indicators are regressed jointly by one network, their errors are correlated and the per-indicator intervals are not independent; no adjustment for multiplicity was made, and comparisons between indicators are descriptive. Model comparisons require sign agreement across seeds and a decision criterion fixed a priori in a dated pre-registration document held in the analysis repository; interventions whose bootstrap interval includes zero are recorded as unadopted regardless of the point estimate. Analyses used Python with PyTorch and scikit-learn.

Abbreviations

Abbreviation	Definition
AI	artificial intelligence
AUROC	area under the receiver operating characteristic curve
AVA	aortic valve area
AV Vmax	aortic valve peak velocity
CI	confidence interval
CV	coefficient of variation
dP/dt	rate of rise of ventricular pressure
DVI	dimensionless velocity index
EACTS	European Association for Cardio-Thoracic Surgery
EACVI	European Association of Cardiovascular Imaging
E/e'	early diastolic mitral inflow to annular velocity ratio
ESC	European Society of Cardiology
GLS	global longitudinal strain
ICC	intraclass correlation coefficient
IQR	interquartile range
IRB	institutional review board
LAD	left atrial diameter
LAVI	left atrial volume index
LoA	limits of agreement
LV	left ventricle
LVDD	left ventricular end-diastolic diameter
LVDs	left ventricular end-systolic diameter
LVEF	left ventricular ejection fraction
LVOT	left ventricular outflow tract
NPV	negative predictive value
PASP	pulmonary artery systolic pressure
PCG	phonocardiogram
PG	pressure gradient

Abbreviation	Definition
PI	prediction interval
PISA	proximal isovelocity surface area
PPG	photoplethysmogram
PPV	positive predictive value
RegVol	regurgitant volume
SD	standard deviation
S1	first heart sound
S2	second heart sound
SMD	standardized mean difference
TTE	transthoracic echocardiography
TRIPOD+AI	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis, artificial-intelligence extension
VC	vena contracta
VTI	velocity–time integral

Declarations

Use of AI-assisted tools

A large-language-model assistant (Claude, Anthropic) was used under author direction for language editing, for writing and checking the analysis and figure-plotting scripts held in the analysis repository, and for drafting the schematic in Fig. 1. All data figures were generated by those scripts from the locked prediction files; no image was generated or altered by a generative image model. The authors reviewed and verified all analyses, figures and text and take full responsibility for the content.

Code availability

The signal-processing pipeline, the model architecture and the training code are proprietary to Clairaudience Co., Ltd. and are not publicly available; they will be made available to the editors and reviewers for the purposes of peer review under a confidentiality agreement, and to qualified researchers from the corresponding author on reasonable request under a material-transfer or licensing agreement. What is released is what the reported results rest on rather than the model itself: the evaluation scripts that regenerate every figure and number in this report from the released prediction files; the two scripts that depend on the withheld pipeline, the hand-crafted-descriptor baseline (scripts/paper_ridge_baseline.py) and the modality knockout (scripts/modality_knockout_armD.py), are

included for inspection and their outputs are released as prediction files; the pre-registration documents (indicator panel, target-definition module `focus_targets.py` and model-comparison criteria), the dated post-lock analysis plan of 13 September 2026, and the de-identified prediction files described under Data availability. These will be deposited, the prediction files subject to hospital approval, in a public repository with a digital object identifier at publication; until then they are available from the corresponding author. The a priori definitions on which the reported results depend – the panel coverage criterion, the reported panel and the decision criteria of the model comparisons – are held in dated files of that release whose file dates and internal dates precede the runs they govern. A TRIPOD+AI checklist is provided as Supplementary Table S1.

Data availability

The echocardiographic reference data are held by Keimyung University Dongsan Hospital, contain protected health information and cannot be provided. Requests for the de-identified heart-sound and pulse-wave recordings will be considered by the corresponding author on a case-by-case basis, subject to institutional review board approval and a data-use agreement. De-identified per-recording out-of-fold and held-out predictions, the corresponding reference indicator values (numerical values without identifiers) and observation masks underlying Figs. 2–5 and Tables 2 and 3 will be deposited, subject to approval by the hospital, in a public repository with a digital object identifier at publication, together with the evaluation scripts that regenerate the reported figures and numbers from them (Code availability).

Acknowledgements

We thank the echocardiography laboratory staff of Keimyung University Dongsan Hospital.

Funding

This work received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author contributions

Q.W. conceived the study, supervised the project, developed the algorithms, designed the experiments and wrote the manuscript. J.-S.H. built the database. J.-H.K. designed the circuitry. J.-W.J., H.-J.K. and S.-Y.Y. collected the clinical data. M.K. labeled the data. S.-W.L. performed the signal processing. J.C. performed the medical statistics. H.K. supervised the clinical management of the patients and the labeling of the patient data. C.Y. supervised the development of the AI algorithms. Y.-S.P. developed the firmware and the application. M.-S.K. developed the application. I.-C.K. supervised the clinical and medical aspects of the study. G.-O.N. designed the circuitry and developed the device hardware. G.K. developed the data server. Q.W. and I.-C.K. contributed equally as first authors. All authors reviewed and approved the final manuscript.

Competing interests

Q.W. and I.-C.K. are co-founders and shareholders of Clairaudience Co., Ltd., which develops the MUST device. The clinical data are held by Keimyung University Dongsan Hospital, and the analyses were performed jointly by the hospital, the university and the company; the prediction files were locked before the held-out cohort was evaluated (Methods). J.-S.H., J.-H.K., J.-W.J., S.-W.L. and Y.-S.P. are affiliated with Clairaudience Co., Ltd. The other authors declare no competing interests.

References

1. Praz, F. et al. 2025 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur. Heart J.* 46, 4635–4736 (2025).
2. Baumgartner, H. et al. Recommendations on the echocardiographic assessment of aortic valve stenosis: a focused update from the EACVI and the ASE. *Eur. Heart J. Cardiovasc. Imaging* 18, 254–275 (2017).
3. Zoghbi, W. A. et al. Recommendations for noninvasive evaluation of native valvular regurgitation. *J. Am. Soc. Echocardiogr.* 30, 303–371 (2017).
4. Mitchell, C. et al. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the American Society of Echocardiography. *J. Am. Soc. Echocardiogr.* 32, 1–64 (2019).
5. Chamsi-Pasha, M. A., Sengupta, P. P. & Zoghbi, W. A. Handheld echocardiography: current state and future perspectives. *Circulation* 136, 2178–2188 (2017).
6. Nkomo, V. T. et al. Burden of valvular heart diseases: a population-based study. *Lancet* 368, 1005–1011 (2006).
7. Coffey, S. et al. Global epidemiology of valvular heart disease. *Nat. Rev. Cardiol.* 18, 853–864 (2021).
8. d'Arcy, J. L. et al. Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: the OxVALVE Population Cohort Study. *Eur. Heart J.* 37, 3515–3522 (2016).
9. d'Arcy, J. L., Prendergast, B. D., Chambers, J. B., Ray, S. G. & Bridgewater, B. Valvular heart disease: the next cardiac epidemic. *Heart* 97, 91–93 (2011).
10. Iung, B. et al. Contemporary presentation and management of valvular heart disease: the EURObservational Research Programme Valvular Heart Disease II Survey. *Circulation* 140, 1156–1169 (2019).
11. Johnson, G. R., Myers, G. S. & Lees, R. S. Evaluation of aortic stenosis by spectral analysis of the murmur. *J. Am. Coll. Cardiol.* 6, 55–63 (1985).
12. Nygaard, H., Thuesen, L., Hasenkam, J. M., Pedersen, E. M. & Paulsen, P. K. Assessing the severity of aortic valve stenosis by spectral analysis of cardiac murmurs (spectral vibrocardiography). Part I: technical aspects. *J. Heart Valve Dis.* 2, 454–467 (1993).
13. Donnerstein, R. L. Continuous spectral analysis of heart murmurs for evaluating stenotic cardiac lesions. *Am. J. Cardiol.* 64, 625–630 (1989).

14. Kim, D. & Tavel, M. E. Assessment of severity of aortic stenosis through time-frequency analysis of murmur. *Chest* 124, 1638–1644 (2003).
15. Etchells, E., Bell, C. & Robb, K. Does this patient have an abnormal systolic murmur? *JAMA* 277, 564–571 (1997).
16. Mangione, S. & Nieman, L. Z. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. *JAMA* 278, 717–722 (1997).
17. Gardezi, S. K. M. et al. Cardiac auscultation poorly predicts the presence of valvular heart disease in asymptomatic primary care patients. *Heart* 104, 1832–1835 (2018).
18. Chorba, J. S. et al. Deep learning algorithm for automated cardiac murmur detection via a digital stethoscope platform. *J. Am. Heart Assoc.* 10, e019905 (2021).
19. Guo, L. et al. Development and evaluation of a deep learning–based pulmonary hypertension screening algorithm using a digital stethoscope. *J. Am. Heart Assoc.* 14, e036882 (2025).
20. Liu, C. et al. An open access database for the evaluation of heart sound algorithms. *Physiol. Meas.* 37, 2181–2213 (2016).
21. Clifford, G. D. et al. Classification of normal/abnormal heart sound recordings: the PhysioNet/Computing in Cardiology Challenge 2016. *Comput. Cardiol.* 43, 609–612 (2016).
22. Reyna, M. A. et al. Heart murmur detection from phonocardiogram recordings: the George B. Moody PhysioNet Challenge 2022. *PLOS Digit. Health* 2, e0000324 (2023).
23. Oliveira, J. et al. The CirCor DigiScope dataset: from murmur detection to murmur classification. *IEEE J. Biomed. Health Inform.* 26, 2524–2535 (2022).
24. Schmidt, S. E., Holst-Hansen, C., Graff, C., Toft, E. & Struijk, J. J. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiol. Meas.* 31, 513–529 (2010).
25. Springer, D. B., Tarassenko, L. & Clifford, G. D. Logistic regression-HSMM-based heart sound segmentation. *IEEE Trans. Biomed. Eng.* 63, 822–832 (2016).
26. Babu, K. A., Ramkumar, B. & Manikandan, M. S. Automatic identification of S1 and S2 heart sounds using simultaneous PCG and PPG recordings. *IEEE Sens. J.* 18, 9430–9440 (2018).
27. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* 28, R1–R39 (2007).
28. Elgendi, M. et al. The use of photoplethysmography for assessing hypertension. *npj Digit. Med.* 2, 60 (2019).
29. Weissler, A. M., Harris, W. S. & Schoenfeld, C. D. Systolic time intervals in heart failure in man. *Circulation* 37, 149–159 (1968).
30. Acosta, J. N., Falcone, G. J., Rajpurkar, P. & Topol, E. J. Multimodal biomedical AI. *Nat. Med.* 28, 1773–1784 (2022).
31. Lai, C. et al. Multimodal AI to forecast arrhythmic death in hypertrophic cardiomyopathy. *Nat. Cardiovasc. Res.* 4, 891–903 (2025).

32. Kannan, A., Saikia, M. J., Kumar, S. & Datta, S. Detection of valvular heart diseases from PCG signals using machine and deep learning models: a review. *IEEE Access* 13, 110344–110360 (2025).
33. Finlayson, S. G. et al. The clinician and dataset shift in artificial intelligence. *N. Engl. J. Med.* 385, 283–286 (2021).
34. Reyes-Toledo, R. E. et al. Routine echocardiographic measurements in clinical practice: reproducibility of left and right ventricular indices and agreement with cardiac magnetic resonance. *Front. Cardiovasc. Med.* 13, 1843290 (2026).
35. Manna, D., Eliasson, M., Bech-Hanssen, O. & Lindow, T. Reproducibility of echocardiographic measures of aortic stenosis severity and its impact on grading of severity. *J. Am. Soc. Echocardiogr.* 37, 370–372.e2 (2024).
36. Sacchi, S. et al. Doppler assessment of aortic stenosis: a 25-operator study demonstrating why reading the peak velocity is superior to velocity time integral. *Eur. Heart J. Cardiovasc. Imaging* 19, 1380–1389 (2018).
37. Leeflang, M. M. G., Bossuyt, P. M. M. & Irwig, L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J. Clin. Epidemiol.* 62, 5–12 (2009).
38. Lijmer, J. G. et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 282, 1061–1066 (1999).
39. Sakamoto, T., Kusukawa, R., MacCanon, D. M. & Luisada, A. A. Hemodynamic determinants of the amplitude of the first heart sound. *Circ. Res.* 16, 45–57 (1965).
40. Leatham, A. Splitting of the first and second heart sounds. *Lancet* 264, 607–614 (1954).
41. O'Rourke, M. F. The arterial pulse in health and disease. *Am. Heart J.* 82, 687–702 (1971).
42. Van de Werf, F., Minten, J., Carmeliet, P., De Geest, H. & Kesteloot, H. The genesis of the third and fourth heart sounds: a pressure–flow study in dogs. *J. Clin. Invest.* 73, 1400–1407 (1984).
43. Lang, R. M. et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J. Am. Soc. Echocardiogr.* 28, 1–39.e14 (2015).
44. Nagueh, S. F. et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J. Am. Soc. Echocardiogr.* 29, 277–314 (2016).
45. Rudski, L. G. et al. Guidelines for the echocardiographic assessment of the right heart in adults: a report from the American Society of Echocardiography. *J. Am. Soc. Echocardiogr.* 23, 685–713 (2010).
46. Voigt, J.-U. et al. Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging. *Eur. Heart J. Cardiovasc. Imaging* 16, 1–11 (2015).
47. Al Tawil, A., Bin Sheeha, B. & Aburub, A. An empirical comparison of deep CNN architectures for heart sound classification from phonocardiogram signals. *Front. Physiol.* 17, 1927399 (2026).

48. Vovk, V., Gammerman, A. & Shafer, G. Algorithmic Learning in a Random World (Springer, 2005).

49. Angelopoulos, A. N. & Bates, S. Conformal prediction: a gentle introduction. Found. Trends Mach. Learn. 16, 494–591 (2023).

50. Efron, B. & Tibshirani, R. J. An Introduction to the Bootstrap (Chapman & Hall/CRC, 1994).

Figures

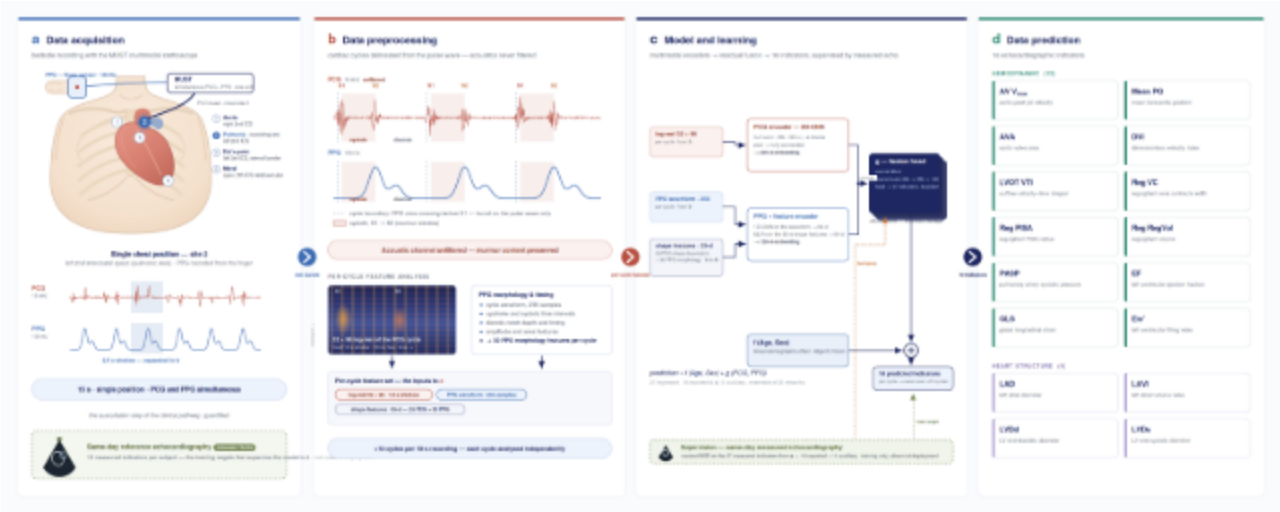


Figure 1

System overview. a) Data acquisition with the multimodal stethoscope at a single chest site, with the pulse wave from the fingertip. b) Data preprocessing: pulse-wave-anchored cycle delineation and per-cycle inputs. c) Model and learning: modality encoders, shared trunk and indicator head, with the demographic offset added; dashed elements exist only during training. d) Data prediction: the panel of continuous echocardiographic indicators in two tiers.

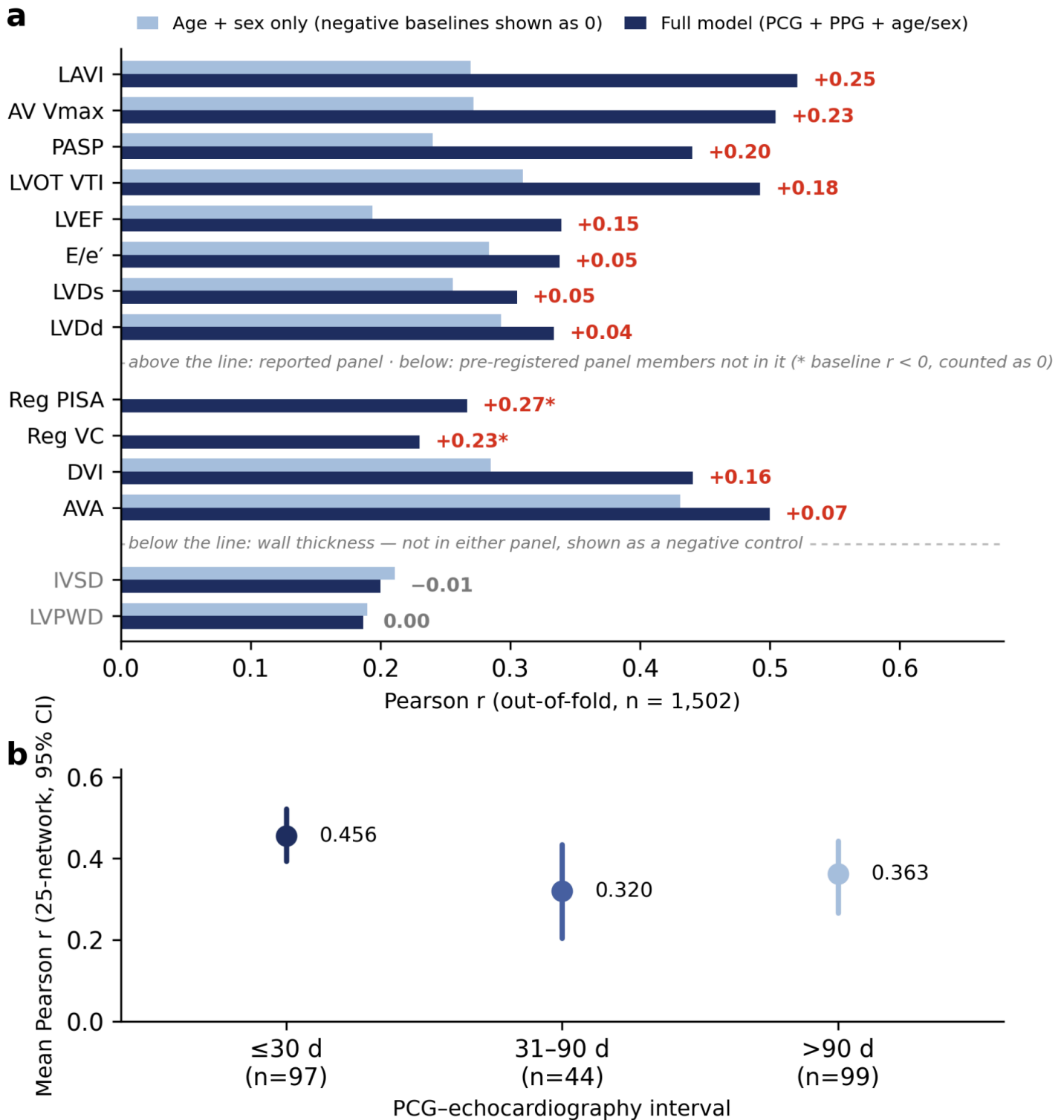


Figure 2

Indicator prediction and the net contribution of auscultation. a) Out-of-fold correlation, demographic baseline versus full model: the reported panel (top), the pre-registered panel members not in it (middle), and wall thickness as a negative control (bottom). b) Mean correlation in the loosely paired held-out set by recording-echocardiography interval.

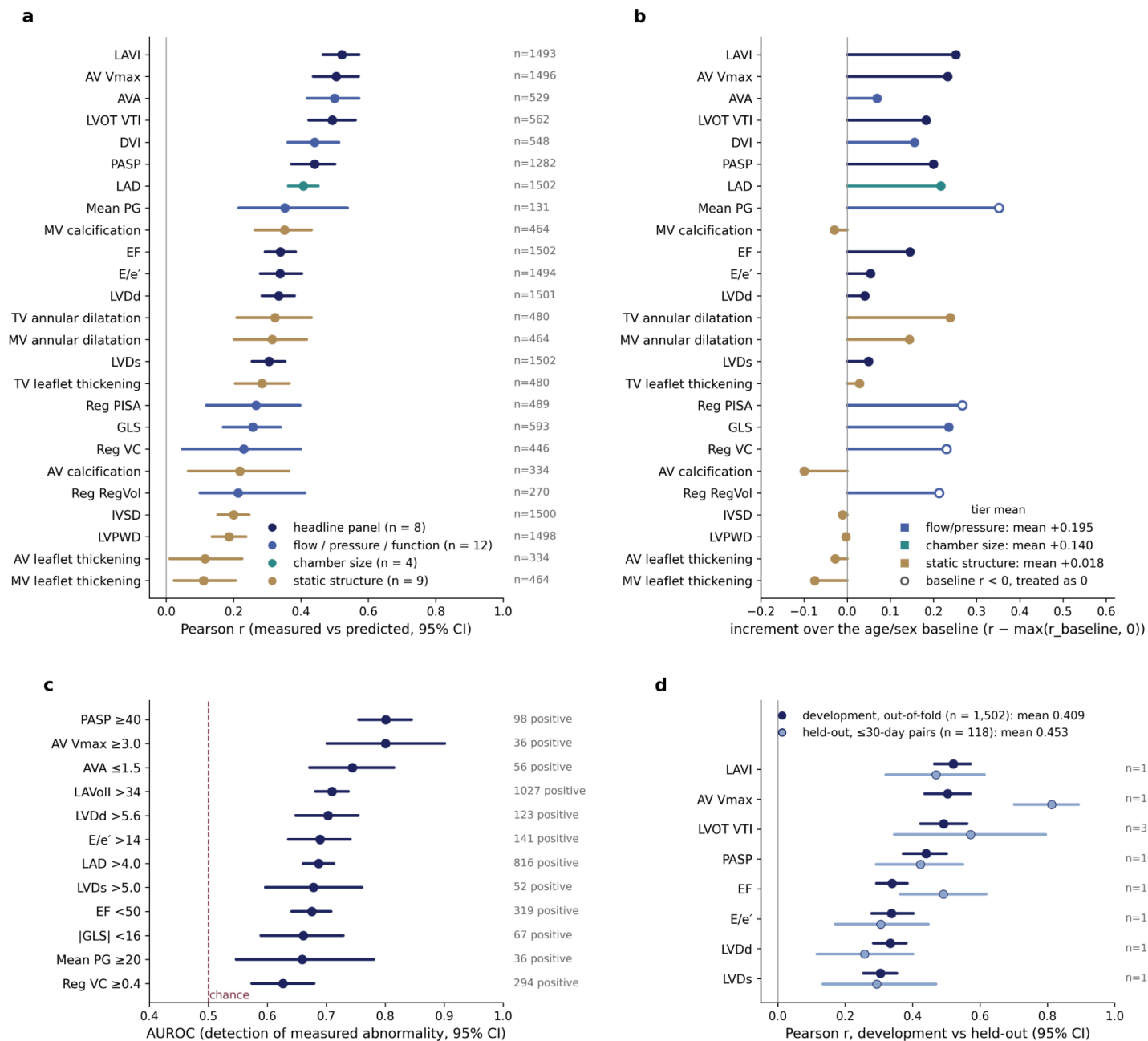


Figure 3

Per-indicator prediction accuracy across the target set. a) Continuous agreement (Pearson r). b) Increment over the age–sex baseline (open markers: baseline counted as zero). c) Detection of measured abnormality (AUROC). d) Development versus held-out correlation for the reported indicators.

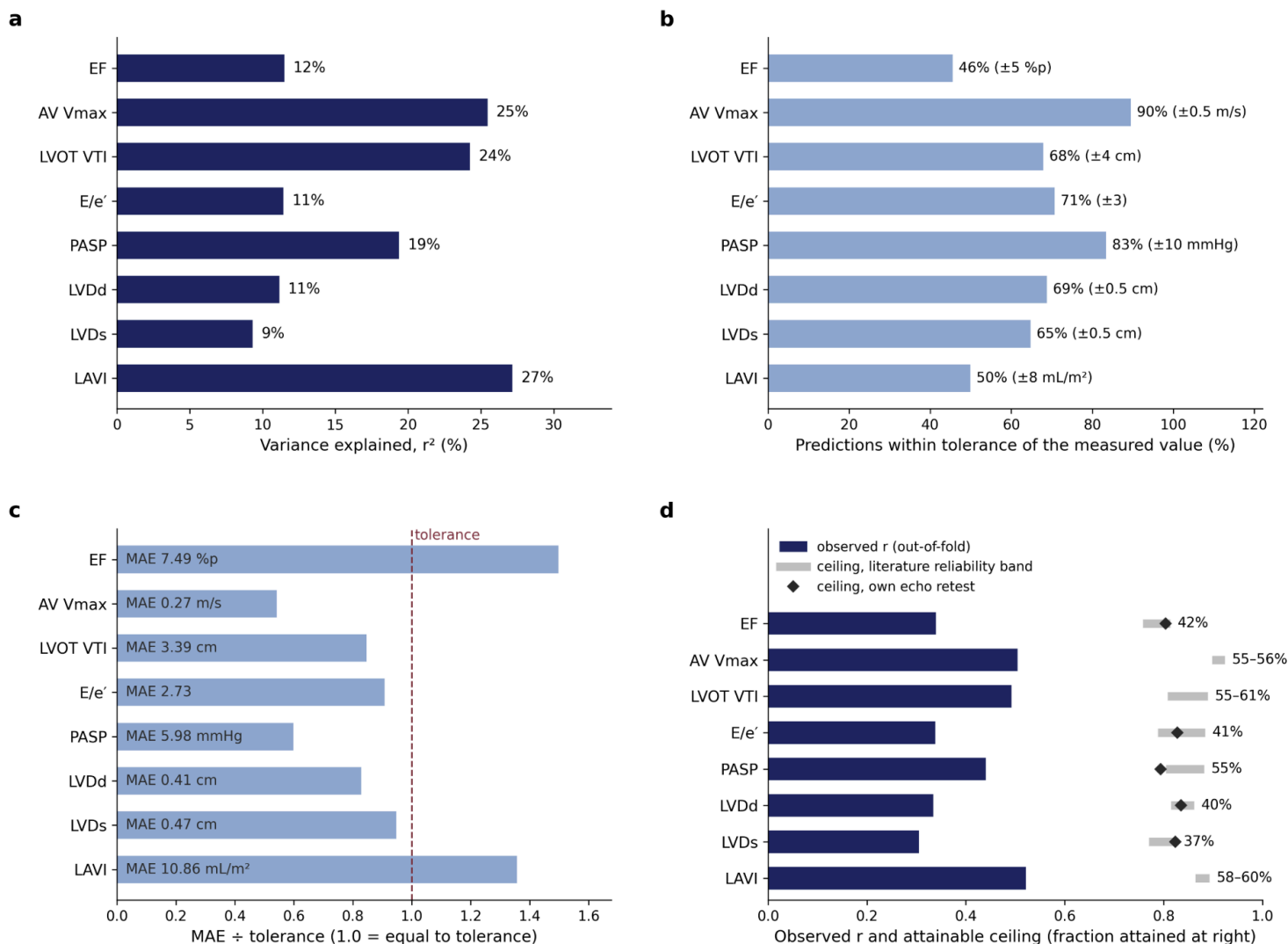


Figure 4

Four views of the same output for the eight reported indicators (protocol-acquired panel, post hoc, chosen for complete denominators). a) Variance explained. b) Predictions within clinical tolerance. c) Mean absolute error relative to tolerance. d) Observed correlation against the attainable ceiling.

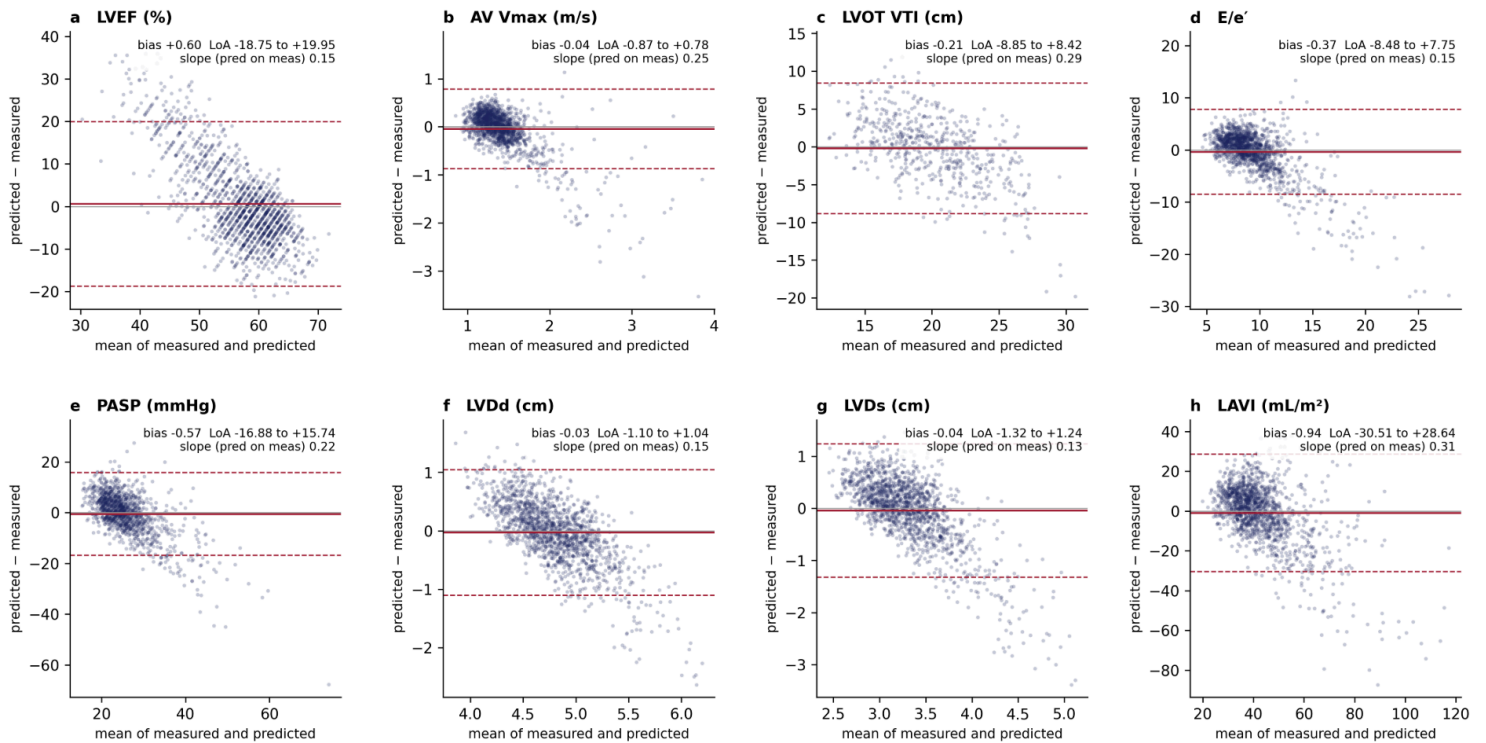


Figure 5

Bland–Altman agreement in physical units for the eight reported indicators (protocol-acquired panel, post hoc, chosen for complete denominators). a) LVEF. b) AV Vmax. c) LVOT VTI. d) E/e'. e) PASP. f) LVDd. g) LVDs. h) LAVI. Solid line, mean bias; dashed lines, 95% limits of agreement.

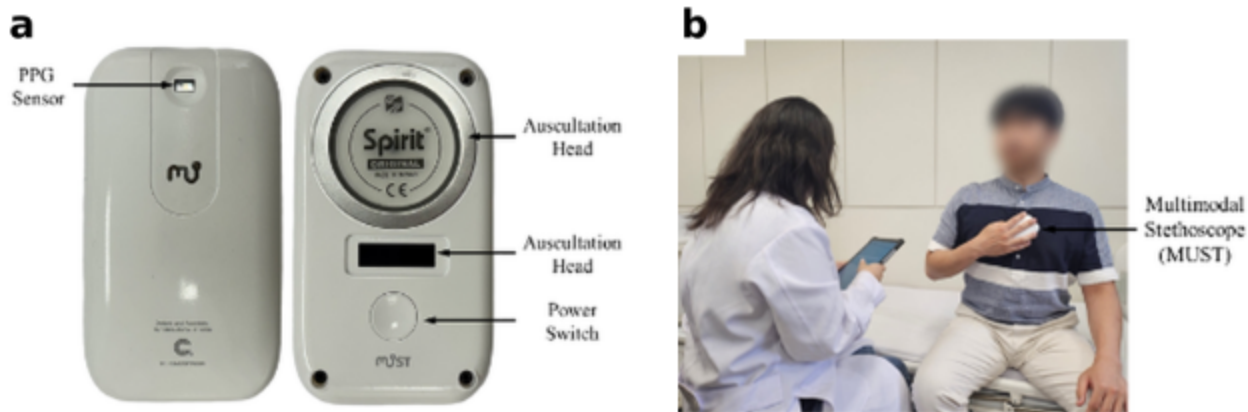
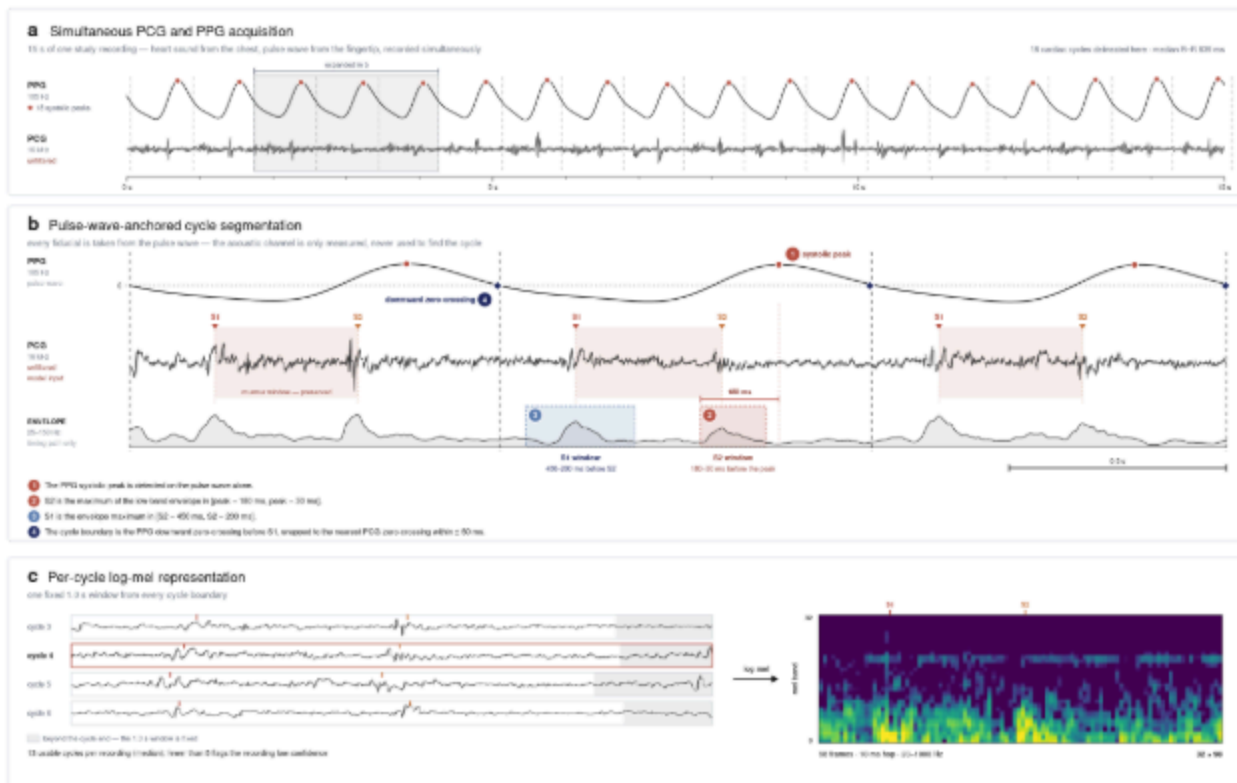


Figure 6

The measurement device and measurement photographs. a) Front and back of the device: the auscultation head acquires the PCG from the chest, and the PPG sensor on the opposite face simultaneously records the pulse wave from the fingertip. b) Physiological signal measurement in progress.



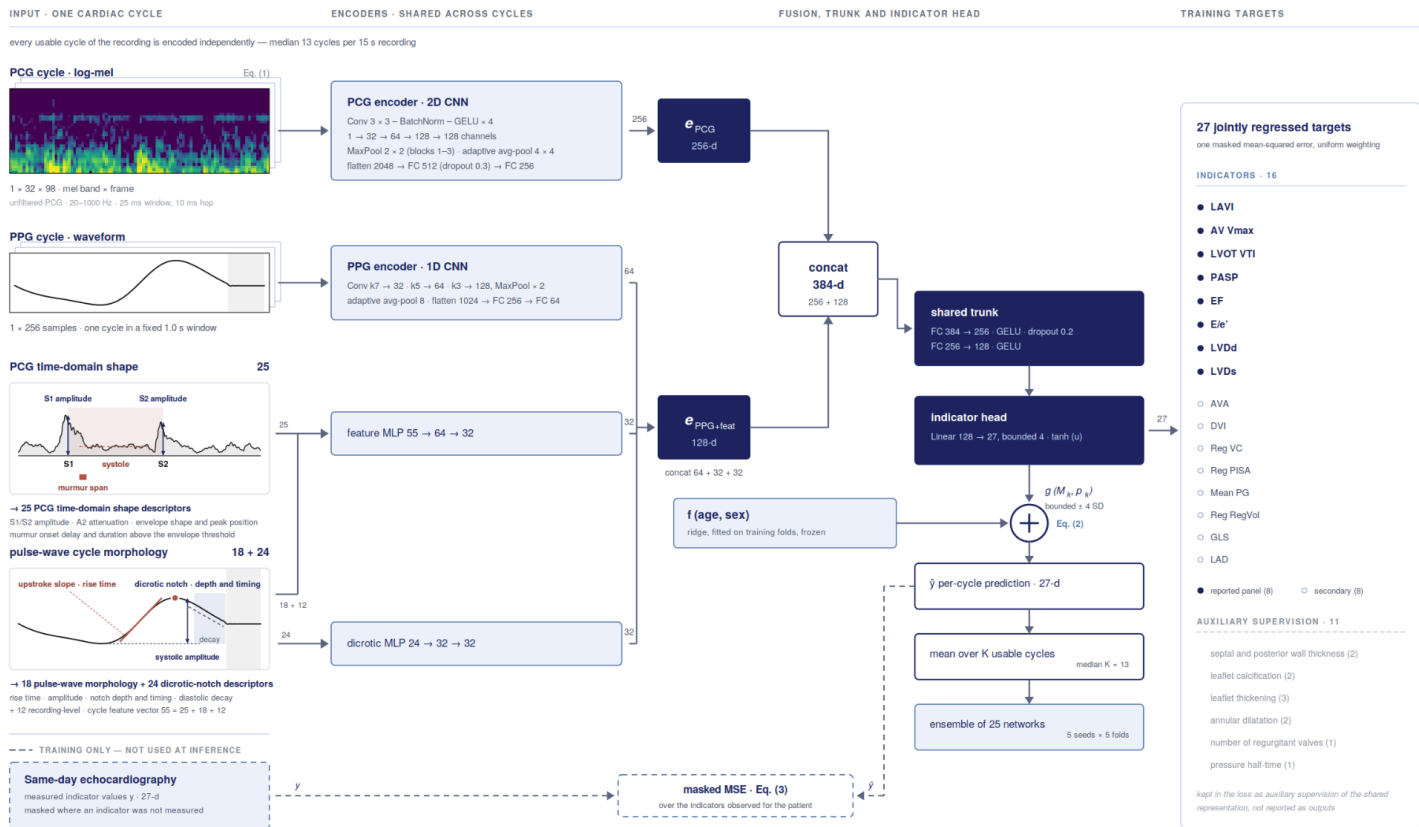


Figure 8

Network architecture of the echocardiographic indicator prediction model. Left, per-cycle inputs; center, encoders; right, fusion trunk, indicator head and demographic offset; far right, training targets. Dashed elements exist only during training.

Supplementary Files

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

- [Supplementaryinformation.docx](#)