ESC European Society of Cardiology

European Heart Journal - Digital Health (2026) **7**, ztaf121
https://doi.org/10.1093/ehjdh/ztaf121

ORIGINAL ARTICLE

Artificial intelligence implementation in automated heart chambers quantification during pharmacological stress echocardiography

Arnas Karuzas [1,*], Quirino Ciampi [2], Ieva Kazukauskiene[3,4], Laurynas Miscikas[1], Karolis Sablauskas[4], Antanas Kiziela[4], Dovydas Verikas[1], Jurgita Plisiene[5], Vaiva Lesauskaite[1], Lauro Cortigiani [6], Karina Wierzbowska-Drabik[7], Jaroslaw D. Kasprzak[8], Jorge Lowenstein[9], Costantina Prota [10], Nicola Gaibazzi [11], Domenico Tuttolomondo [11], Attilio Lepone[12], Sofia Marconi[9], Rosina Arbucci[9], and Eugenio Picano[13]; Stress Echo 2030 Study Group of the Italian Society of Echocardiography and Cardiovascular Imaging

[1]Institute of Cardiology, Lithuanian University of Health Sciences, Sukileliu pr. 15, Kaunas, Lithuania; [2]Cardiology Division, Fatebenefratelli Hospital of Benevento, Benevento, Italy; [3]Faculty of Medicine, Department of Pathology and Forensic Medicine, Vilnius University, Vilnius, Lithuania; [4]Department of Research and Development, Ligence Ltd, Vilnius, Lithuania; [5]Clinical Department of Cardiology, Lithuanian University of Health Sciences, Kaunas, Lithuania; [6]Cardiology Department, San Luca Hospital, Lucca, Italy; [7]Department of Internal Disease and Clinical Pharmacology, Medical University of Lodz, Lodz, Poland; [8]Department of Cardiology, Medical University of Lodz/Bieganski Hospital, Lodz, Poland; [9]Cardiodiagnosticos, INVESTIGACIONES MEDICAS, Buenos Aires, Argentina; [10]Department of Cardiology, University Hospital San Giovanni di Dio e Ruggi dAragona, Salerno, Italy; [11]Cardiology Department, University Hospital of Parma, Parma, Italy; [12]Cardiology Division, Pisa University Hospital, Pisa, Italy; and [13]University of Belgrade, University Clinical Center Serbia, Medical School, University Center Serbia, Cardiology Clinic, Belgrade, Serbia

Received 11 June 2025; revised 1 August 2025; accepted 26 September 2025; online publish-ahead-of-print 23 October 2025

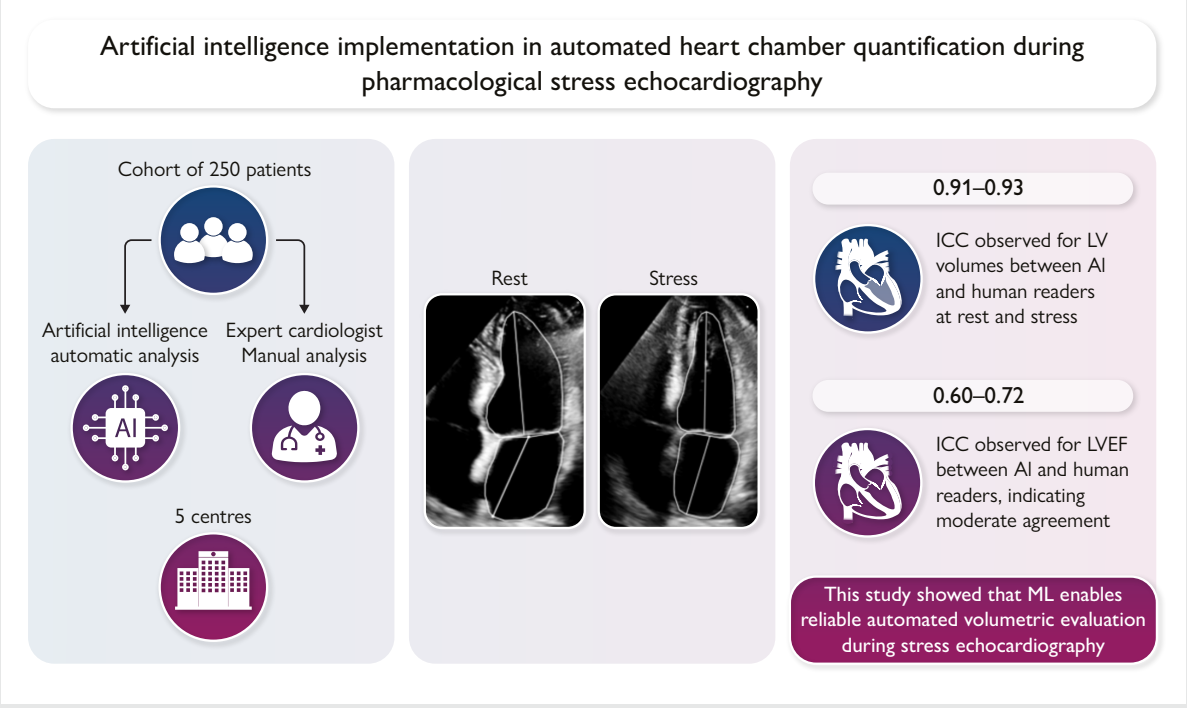
Aims Stress echocardiography (SE) is widely used for assessing coronary artery disease, but volumetric chamber analysis during SE is limited by time-consuming manual tracings and operator-dependent variability. Automated evaluation may overcome these barriers and enhance efficiency.

Methods and results This multi-centre study included 240 participants undergoing pharmacological SE for ischaemic heart disease evaluation from five sites in four countries. SE imaging data from apical four-chamber and two-chamber views were acquired during rest and stress phases. Expert cardiologists manually traced endocardial borders for left ventricular (LV), left atrial (LA) and right ventricular (RV), right atrial (RA) areas, which were compared to machine learning (ML) derived measurements. Image quality was categorized as optimal, good, fair, or poor, and its influence on ML performance was analysed. Statistical methods included Intraclass Correlation Coefficients (ICCs), Bland–Altman testing, and within-patient coefficient of variation. The yield of the ML algorithm demonstrated consistency across rest and stress phases. It demonstrated strong agreement with cardiologists for LV and LA volumes, with ICCs ranging from 0.84 to 0.93 across rest and stress conditions. RA and RV areas measurements showed moderate correlations, with better agreement at rest than during stress phases. Image quality significantly influenced ML performance, as poor-quality images reduced diagnostic yield.

* Corresponding author. Tel: +37067926921, Email: arnas.karuzas@lsmu.lt

Conclusion AI-driven volumetric analysis is a reliable method for quantifying left-sided heart chambers during pharmacological SE, with results closely matching expert measurements. Moderate reliability for right-sided chambers highlights the need for high-quality imaging and standardized protocols. AI integration may streamline SE workflows and support improved clinical decision-making.

Graphical Abstract



Keywords Artificial intelligence • stress echocardiography • cardiac chambers • volumes • image quality

Introduction

Coronary artery disease (CAD) remains one of the leading causes of disability and premature mortality worldwide, accounting for 30–40% of all cardiovascular deaths.[1,2] Early identification of high-risk patients and timely diagnosis are essential for improving outcomes through preventive or interventional strategies. Stress echocardiography (SE) is a widely utilized non-invasive imaging technique for evaluating suspected CAD.[3] SE can provoke ischaemia in myocardial regions with insufficient blood supply due to flow-limiting coronary artery stenosis, with the mechanisms of increased myocardial oxygen demand (exercise or dobutamine) or reduced myocardial oxygen supply through steal phenomena (vasodilators).[4] This leads to regional wall motion abnormalities (RWMAs), a highly sensitive and specific indicator of obstructive CAD.[5–7] Furthermore, SE enables the assessment of various haemodynamic and functional parameters that change dynamically during stress, providing additional diagnostic and prognostic insights.[8,9]

According to the 2024 ESC Guidelines for the Management of Chronic Coronary Syndromes, SE is advised for patients with a moderate to high pre-test probability (15–85%) of obstructive CAD.[3] However, recent studies suggest that incorporating additional echocardiographic parameters, such as left ventricular ejection fraction (LVEF), global longitudinal strain (GLS), and left atrial volume (LAV), along with other functional and hemodynamic measures, could provide valuable prognostic information for detecting CAD and identification of functional vulnerabilities beyond CAD.[10] Reflecting these findings, multi-parametric SE protocols, such as the ABCDE-SE protocol, have emerged, integrating a spectrum of parameters beyond RWMA, such as pulmonary congestion (B-lines), coronary flow velocity reserve (CFVR), left ventricular contractile reserve, heart rate reserve, LAV, and GLS.[8,10–13]

Despite its diagnostic advantages, SE remains operator-dependent and time-consuming, limiting reproducibility and efficiency.[3,14] In particular, the measurement by hand of planimetric areas is required to derive LAV, LV end-diastolic volume (EDV), and end-systolic volumes (ESV). EDV is needed to assess the pre-load reserve, the ESV for contractile reserve, and LAV for assessing diastolic function during stress.[15] However, these measurements typically require several minutes to perform, often exceeding 5–10 min in routine practice, and are subject to intra- and interobserver variability. An operator-independent approach would improve the workflow and increase the accuracy of these important measurements of recognized clinical and prognostic importance.[15]

Artificial intelligence (AI) has emerged as a promising solution to address these challenges. Convolutional neural networks (CNNs) have demonstrated significant potential in echocardiography. CNNs can perform tasks such as volumetric chamber analysis, LVEF and GLS quantification and even RWMA detection.[16–18] Recent advancements have enabled AI solutions to identify RWMAs in transthoracic echocardiography, achieving a diagnostic accuracy comparable to expert consensus assessments.[17,19] These capabilities not only reduce operator bias but also streamline workflows by automating time-consuming processes. While the application of AI in SE is still in early stages, it holds promise for improving diagnostic precision and enabling wider adoption of multi-parametric SE.

This study aimed to evaluate the validity and reliability of an automated machine learning (ML) system for volumetric assessment of heart chambers during pharmacological SE and to compare its performance with that of expert cardiologists. Additionally, the study aimed to investigate the impact of varying echocardiographic image quality on the ML system's performance in quantifying heart volumes.

Methods

The multi-centre study enrolled 240 participants who underwent pharmacological SE testing between 1 November 2023 and 21 June 2024 across five accredited centres, which participated in the Stress Echo 2030 trial. The study duration refers to the overall multi-centre enrolment window. At each participating centre, inclusion was limited to shorter pre-defined periods (up to 1 month per centre). Patients were enrolled consecutively at each centre according to the Stress Echo 2030 inclusion protocol. Inclusion criteria included patients with clinical indications for pharmacological SE to evaluate inducible ischaemia in known or suspected ischaemic heart disease. Exclusion criteria were participants with a poor acoustic window at rest, clinically significant valvular or congenital heart disease, and prognostically relevant non-cardiac diseases such as advanced cancer, end-stage renal disease, or severe obstructive pulmonary disease.

Pharmacological SE was performed using commercially available ultrasound machines. In total, 228 patients (95%) underwent dipyridamole SE, while 12 patients (5%) underwent dobutamine SE. No exercise protocols were performed. Contrast echocardiography was not used in any patient; all analyses were performed on native images. The test was terminated according to the guidelines criteria.[9]

The imaging data were acquired in DICOM format and anonymized for subsequent analysis. Then, an expert cardiologist trained in SE reviewed the studies, selecting apical four-chamber (4Ch) and two-chamber (2Ch) view images during stress and rest phases. The cardiologists identified and marked end-systolic (ES) and end-diastolic (ED) frames. Then they traced the endocardial borders for the left atrium (LA) and the right atrium (RA) in end-systole, for the left ventricle (LV) and the right ventricle (RV) in end-systole and end-diastole. More detailed methodology of the measurements is available in the Supplementary material online, *Supplement data*.

The same ED and ES frames of the duplicate images were then processed by an ML algorithm trained on a separate set of images. Each evaluator (cardiologist and ML algorithm) was blinded to each other's measurements. All tracings and automated measurements were performed using the same software. The ML algorithm was trained on an independent, third-party multi-centre dataset that included 21 539 patient studies with 24 632 labelled apical four- and two-chamber echocardiographic images from centres in Germany, France, Lithuania, and the USA. None of the patients from the present study cohort were included in the training dataset, ensuring external validation.

The study protocol was reviewed and approved by the institutional ethics committees in its latest versions as a part of the more comprehensive SE 2030 study 291/294/295 Comitato Etico Lazio-1, 8 March 2021; https://clinicaltrials.gov/ Identifier NCT NCT050.81115. All patients gave their informed consent to enter the study.

Image quality assessment

The duplicate dataset was created for image quality assessment. Following the European Association of Cardiovascular Imaging (EACVI) guidelines,[20] image quality was categorized into one of four levels: optimal, good, fair, or poor (*Figure 1*). An expert cardiologist, blinded to the measurements, evaluated the quality of all visible cardiac chambers (LV, RV, LA, and RA) in the previously selected ES and ED frames. The cardiologist's task was solely to assess image quality based on the pre-marked frames. Subsequently, an expert cardiologist, blinded to the previous measurements, assessed the image quality of LA and LV. The image quality for each chamber in the ES or ED frame was categorized into one of four grades: optimal, good, fair, or poor.

Statistical analysis

Data analysis involved descriptive statistics to summarize participant characteristics and paired *t*-tests or Wilcoxon signed-rank tests to compare continuous variables between rest and stress phases, with statistical significance set at $P < 0.05$. Some clinical variables had missing entries; therefore, percentages were calculated per available characteristic.

The secondary parameter for evaluation is the Intraclass Correlation Coefficient (ICC) using a two-way mixed model with fixed raters. The ICC will quantify the degree of agreement between the device's measurements and those of the human raters. The Within-Patient Coefficient of Variation will be used to evaluate the variability of measurements within individual patients.

Furthermore, Bland–Altman testing will be conducted to assess the presence of bias and determine the limits of agreement. This testing will facilitate the identification of systematic differences (bias) and provide an overall assessment of the agreement between the Ligence Heart device and the human raters. It will also establish the range within which most differences are likely to occur, thus offering a comprehensive evaluation of the device's performance.

Results

The study enrolled 240 patients from five different centres. One hundred and thirty nine (57) of the patients were male, the average age was 68 years. All patients had a sinus rhythm. The main indication for pharmacological SE was the suspicion of ischaemic heart disease. The main characteristics of the study patients are summarized in *Table 1*.

Yield

The yield of the ML algorithm was consistent across both rest and stress phases. The highest yield was observed for LVEDV4A and LVEDV2A, with 98% and 94% at rest and 95% during stress (*Table 2*). Other measurements, including LV systolic volumes (LVESV4A and LVESV2A) and RVEDA, also showed high yields ranging from 83% to 93%. However, yields for measurements like RVESA, LAV4A, LAV2A, and RAA were lower, ranging from 75% to 83%.

Measurement agreement between cardiologists and ML

We compared chamber measurements performed by expert cardiologists and AI. The results demonstrated strong agreement between human and AI measurements for LV volumes in both rest and stress phases, as observed in apical four-chamber view (LVEDV4 and LVESV4) and apical two-chamber view (LVEDV2 and LVESV2), with correlation coefficient ranging between 0.92–0.95 and 0.89–0.91, respectively. However, there were slight positive biases (*Table 3*), suggesting a tendency for AI to underestimate LV volumes (except LVESV4) compared with human raters.

The measurements of LVEF demonstrated moderate agreement between the cardiologists and AI in both the rest and stress phases. *Figure 2* shows scatter and Bland–Altman plots for LVEF derived from apical four chamber and two chamber views at rest and stress. The ICCs ranged from 0.69 to 0.77 and 0.6–0.72. respectively, indicating moderate agreement.

Left atrial volume measurements (LAV4 and LAV2) showed high ICC values (>0.95), indicating excellent agreement between AI and human

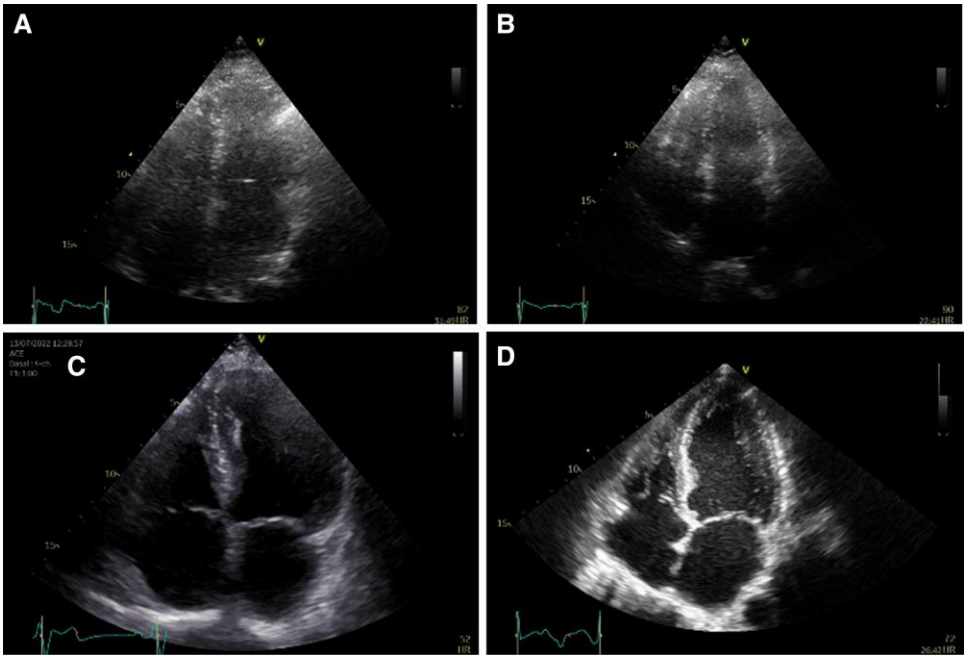


Figure 1 Quality grades of left ventricle apical four-chamber view: (*A*) poor, (*B*) fair, (*C*) good, (*D*) optimal.

Table 1 Clinical characteristics of patients

Characteristics	Value
Age (years)[a]	67.5 ± 10.4
Sex, male, *n* (%)	139 (57)
Height (m)	166 (13)
Weight (kg)	76 (17)
BMI[a]	27.8 ± 4.2
BSA[a] (m^2)	1.89 ± 0.19
Dyslipidaemia, *n* (%)	159 (65.2)
Diabetes mellitus, *n* (%)	68 (27.9)
Ischaemic heart disease, *n* (%)	77 (31.6)
Previous MI, *n* (%)	60 (24.6)
Hypertension, *n* (%)	177 (72.5)

Percentages are calculated based on available data for each variable; not all variables were available for all patients.
[a]Mean ± SD.

Table 2 Comparison of measurement yields between cardiologist and machine learning in stress echocardiography

Measurement	Stress phase	*n*, human	*n*, ML	Yield, %
LVEDV4A	Rest	235	230	98
	Stress	226	214	95
LVESV4A	Rest	234	209	89
	Stress	221	206	93
LVEDV2A	Rest	230	216	94
	Stress	225	214	95
LVESV2A	Rest	234	199	85
	Stress	222	185	83
LAV4A	Rest	185	146	79
	Stress	176	142	81
LAV2A	Rest	175	145	83
	Stress	168	126	75
RVEDA	Rest	176	153	87
	Stress	174	156	90
RVESA	Rest	179	140	78
	Stress	170	131	77
RAA	Rest	174	144	83
	Stress	174	136	78

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LAV, left atrial volume; RAA, right atrial area; RVEDA, right ventricular end-diastolic area; RVESA, right ventricular end-systolic area.

measurements. Additionally, the bias approached negligible levels, with root mean square error (RMSE) values ranging between 4 and 7 implying minimal deviation or discrepancy between the AI and human measurements.

Our findings indicated good agreement in the right atrial area (RAA) measurements in the rest phase but moderate in the stress phase, ICC 0.81 and 0.69. respectively. Additionally, a moderate agreement was found between the cardiologists and AI for RV area measurements (RVEDA and RVESA), with a higher agreement in rest and lower agreement in stress phases; ICCs ranged between 0.59–0.74 and 0.52–0.62, respectively.

However, the assessment of fractional area change (FAC) indicated low agreement between human and AI measurements, with ICCs of

Table 3 Agreement and Bland–Altman bias between cardiologist and machine learning in stress echocardiography

Measurement	Stress phase	Pearson R	ICC	Bias	LOA	MAE	RMSE
LVEDV4A	Rest	0.9	0.89	8.15	±31.81	13.48	18.16
	Stress	0.84	0.83	7.62	±38.08	15.4	20.87
LVESV4A	Rest	0.86	0.84	−0.1	±25.62	9.3	13.07
	Stress	0.87	0.86	−0.85	±25.01	9.25	12.79
LVEDV2A	Rest	0.85	0.84	7.97	±39.7	15.55	21.76
	Stress	0.84	0.83	7.29	±38.7	16.18	21.04
LVESV2A	Rest	0.88	0.87	1.34	±22.8	8.04	11.71
	Stress	0.85	0.83	0.42	±23.59	8.56	12.04
LAV4A	Rest	0.93	0.93	3.35	±17.86	7.11	9.71
	Stress	0.93	0.93	0.97	±15.8	6.07	8.12
LAV2A	Rest	0.85	0.83	3.73	±26.24	9.52	13.9
	Stress	0.88	0.88	3.73	±19.6	8.09	10.67
RAA	Rest	0.72	0.71	−0.15	±6.11	2.15	3.12
	Stress	0.56	0.56	0.19	±7.15	2.45	3.65
RVEDA	Rest	0.71	0.71	1.56	±7.73	3.11	4.24
	Stress	0.58	0.57	1.32	±8.75	3.17	4.65
RVESA	Rest	0.6	0.6	1.05	±5.45	2.18	2.97
	Stress	0.52	0.52	1.54	±6.13	2.52	3.48

ICC, intraclass correlation coefficient; LOA, limits of agreement; MAE, mean absolute error; RMSE, root mean square error.

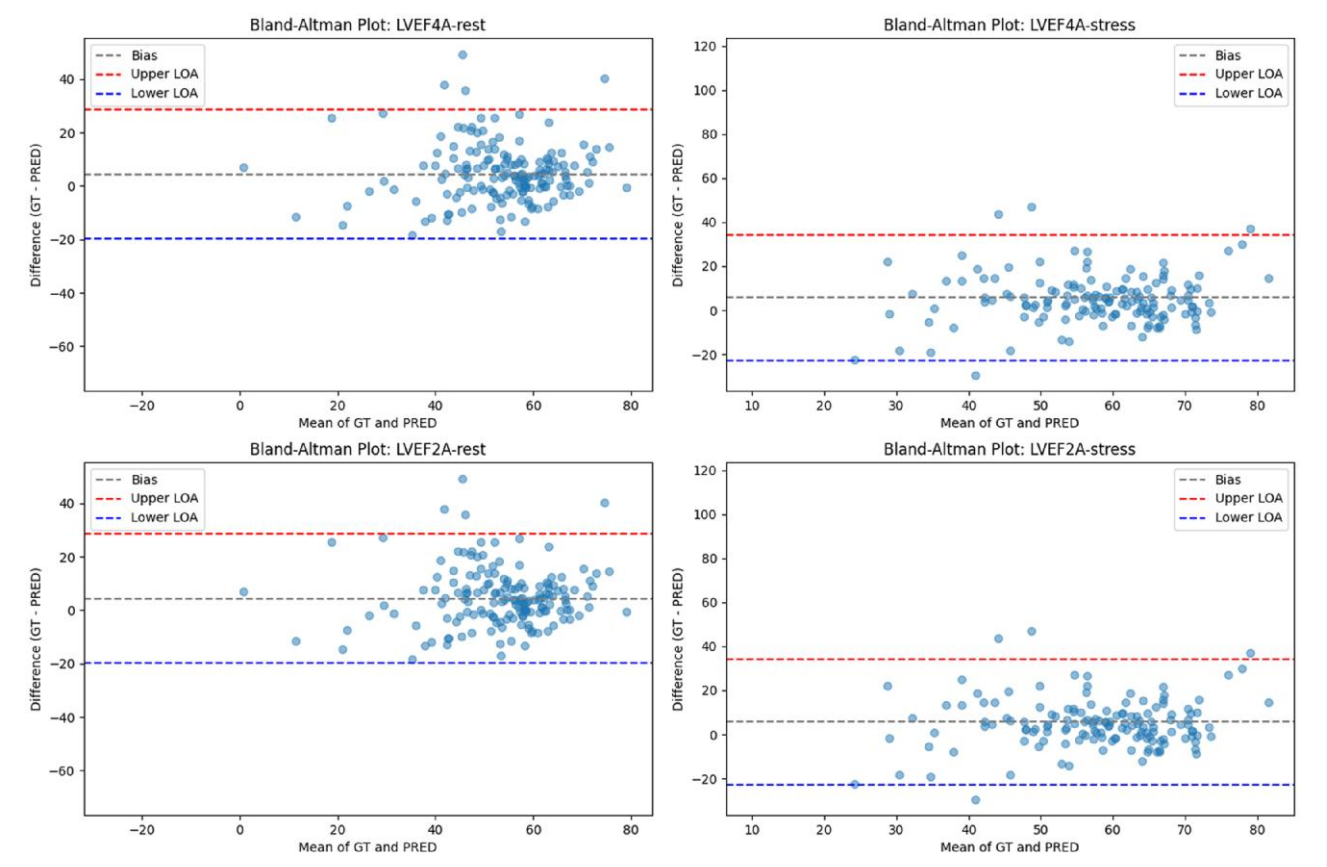


Figure 2 Agreement between machine learning and cardiologist derived left ventricular ejection fraction.

Table 4 Impact of image quality on measurement yields between cardiologist and machine learning in stress echocardiography

Measurement	Image quality	Rest phase			Stress phase		
		Human, *N*	ML, *N*	Yield	Human, *N*	ML, *N*	Yield
LVEDV4A	Poor	1	0	—	4	1	—
	Fair	37	35	94.6	47	41	87.2
	Good	96	95	98.9	94	92	97.9
	Optimal	99	99	100	80	79	98.8
LVESV4A	Poor	6	0	—	12	9	75
	Fair	38	27	71.1	49	43	87.8
	Good	91	86	94.5	80	75	93.8
	Optimal	99	96	97	80	79	98.8
LVEDV2A	Poor	11	5	45.5	15	12	80
	Fair	86	78	90.7	82	77	93.9
	Good	82	82	100	77	74	96.1
	Optimal	51	51	100	51	51	100
LVESV2A	Poor	8	4	50	16	6	37.5
	Fair	63	44	69.8	66	51	77.3
	Good	83	76	91.6	76	69	90.8
	Optimal	80	75	93.8	64	59	92.2
LAV4A	Poor	4	2	—	7	0	—
	Fair	28	15	53.6	28	18	64.3
	Good	64	47	73.4	54	46	85.2
	Optimal	83	81	97.6	80	78	97.5
LAV2A	Poor	4	2	—	7	1	—
	Fair	49	37	75.5	59	36	61
	Good	61	56	91.8	46	42	91.3
	Optimal	54	48	88.9	51	47	92.2
RAA	Poor	35	18	51.4	32	10	31.3
	Fair	68	58	85.3	76	64	84.2
	Good	36	36	100	30	29	96.7
	Optimal	33	32	97	28	28	100
RVEDA	Poor	26	15	57.7	32	23	71.9
	Fair	63	57	90.5	52	48	92.3
	Good	42	41	97.6	42	42	100
	Optimal	13	13	100	10	10	100
RVESA	Poor	57	30	52.6	60	36	60
	Fair	54	47	87	59	47	79.7
	Good	39	37	94.9	29	29	100
	Optimal	11	11	100	7	7	100

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LAV, left atrial volume; RAA, right atrial area; RVEDA, right ventricular end-diastolic area; RVESA, right ventricular end-systolic area.

0.27 in the rest phase and 0.33 in the stress phase. Furthermore, a notable negative bias was observed, indicating a tendency for AI to overestimate FAC.

Agreement between cardiologists and ML within different centres

We performed the analysis on the whole cohort and then separately on the data of each centre. The agreement between AI and cardiologists was consistent across the five centres, with Center 3 showing the strongest agreement (see Supplementary material online, *Supplement data, Tables S1–S5*). Rest-phase agreement was higher overall, while stress conditions led to a modest decline, particularly for right-sided parameters.

Image quality

We evaluated the image quality of each chamber within the frame, in which measurements were performed, into four quality categories: optimal, good, fair, or poor. Although during the stress phase quality of the images tended to decrease, the differences were not statistically significant ($P > 0.05$) (*Table 4*). However, we observed that the left heart chambers were more frequently rated as good or optimal compared with the right heart chambers. Specifically, 84% of LV images and

Table 5 Agreement between cardiologists and machine learning within image quality groups

Measurement	Image quality	Rest phase							Stress phase						
		N	*r*	ICC	Bias	LOA	MAE	RMSE	*N*	*r*	ICC	Bias	LOA	MAE	RMSE
LVEDV4A	Poor	0	—	—	—	—	—	—	1	—	—	—	—	—	—
	Fair	35	0.91	0.91	10.8	±38.1	16.6	22.2	41	0.83	0.82	6.8	±53.6	20.9	28.2
	Good	95	0.91	0.89	6.0	±30.6	12.0	16.7	92	0.85	0.83	6.7	±38.4	15.6	20.7
	Optimal	99	0.88	0.86	9.6	±29.6	13.7	17.9	79	0.84	0.84	9.1	±26.5	12.5	16.3
LVESV4A	Poor	0	—	—	—	—	—	—	9	0.79	0.79	−4.2	±23.3	9.2	12.6
	Fair	27	0.80	0.78	−5.7	±30.3	11.7	16.5	43	0.92	0.90	0.6	±31.6	11.8	16.1
	Good	86	0.88	0.87	−0.5	±23.6	8.9	12.0	75	0.84	0.83	−2.1	±24.9	9.3	12.9
	Optimal	96	0.87	0.83	1.8	±24.0	8.9	12.9	79	0.84	0.84	−0.1	±20.5	7.8	10.5
LVEDV2A	Poor	5	0.66	0.64	−7.7	±50.0	22.0	26.6	12	0.67	0.67	9.1	±46.2	20.2	25.3
	Fair	78	0.87	0.85	7.1	±41.8	15.6	22.5	77	0.86	0.85	5.1	±44.3	17.7	23.2
	Good	82	0.85	0.85	6.0	±35.1	14.5	18.9	74	0.84	0.83	8.5	±33.9	15.3	19.3
	Optimal	51	0.92	0.82	14.1	±38.8	16.6	24.3	51	0.80	0.78	8.4	±33.3	14.2	19.0
LVESV2A	Poor	4	0.76	0.65	−12.4	±20.4	13.0	16.2	6	0.65	0.61	−14.1	±24.8	14.4	19.0
	Fair	44	0.78	0.74	−0.1	±30.8	9.2	15.7	51	0.88	0.85	−1.0	±29.0	11.2	14.8
	Good	76	0.92	0.92	1.0	±20.1	7.9	10.3	69	0.88	0.87	0.7	±18.4	6.9	9.4
	Optimal	75	0.92	0.89	3.3	±18.3	7.2	9.9	59	0.81	0.76	2.8	±21.3	7.7	11.2
LAV4A	Poor	2	—	—	—	—	—	—	0	—	—	—	—	—	—
	Fair	15	0.76	0.75	4.6	±22.9	8.9	12.6	18	0.97	0.96	1.1	±11.6	5.1	6.0
	Good	47	0.96	0.96	3.6	±16.0	7.0	8.9	46	0.91	0.91	0.2	±20.8	7.7	10.6
	Optimal	81	0.92	0.92	2.4	±16.3	6.4	8.7	78	0.93	0.93	1.4	±12.9	5.3	6.7
LAV2A	Poor	2	—	—	—	—	—	—	1	—	—	—	—	—	—
	Fair	37	0.83	0.82	2.7	±28.5	10.6	14.8	36	0.86	0.86	2.2	±23.7	9.4	12.3
	Good	56	0.88	0.86	2.8	±23.0	9.2	12.1	42	0.90	0.89	3.9	±16.3	6.9	9.2
	Optimal	48	0.86	0.84	3.8	±17.9	7.5	9.9	47	0.90	0.90	5.1	±18.0	8.1	10.5
RAA	Poor	18	0.27	0.25	1.4	±11.3	4.9	6.0	10	0.37	0.37	4.8	±11.8	5.9	7.7
	Fair	58	0.72	0.72	−0.1	±6.1	2.2	3.1	64	0.50	0.50	0.3	±7.6	2.8	3.9
	Good	36	0.91	0.91	−0.3	±3.4	1.4	1.8	29	0.69	0.68	−0.4	±4.6	1.9	2.4
	Optimal	32	0.92	0.91	−0.9	±3.0	1.4	1.8	28	0.95	0.95	−1.2	±2.3	1.3	1.7
RVEDA	Poor	15	0.04	0.04	3.4	±10.7	4.2	6.4	23	0.47	0.46	4.9	±13	5.8	8.2
	Fair	57	0.76	0.75	1.9	±7.9	3.5	4.5	48	0.52	0.50	1.7	±8.1	3.2	4.5
	Good	41	0.80	0.80	0.7	±5.3	2.2	2.8	42	0.88	0.88	0.4	±4.8	2.0	2.5
	Optimal	13	0.73	0.72	0.8	±6.6	2.6	3.5	10	0.63	0.63	−0.1	±7.1	3.0	3.6
RVESA	Poor	30	0.37	0.37	2.1	±5.2	2.5	3.3	36	0.55	0.55	2.8	±7.1	3.4	4.6
	Fair	47	0.62	0.60	1.0	±5.4	2.4	3.0	47	0.39	0.38	1.5	±5.8	2.3	3.3
	Good	37	0.68	0.66	0.2	±5.3	1.7	2.7	29	0.66	0.66	0.3	±4.8	1.9	2.5
	Optimal	11	0.94	0.94	0.8	±2.0	1.1	1.3	7	0.89	0.66	0.4	±2.2	1.0	1.2

LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; LAV, left atrial volume; RAA, right atrial area; RVEDA, right ventricular end-diastolic area; RVESA, right ventricular end-systolic area; ICC, intraclass correlation coefficient; LOA, limits of agreement; MAE, mean absolute error; RMSE, root mean square error.

54% of LA images achieved good or optimal ratings, whereas only 38% of RV images and 35% of RA images reached these quality levels. In some patients, individual chamber measurements could not be obtained due to insufficient image quality in the selected views, even though the patients were included in the overall cohort.

Yield depending on the quality

The study results varied depending on the quality of the images. The lowest yield was observed for LV measurements in poor-quality images (45.5–57.7% at rest and 37.5–80% at stress). Right-sided chambers showed somewhat different yield patterns, as detailed in *Table 4*. As image quality improved, the results also improved, with the highest yield obtained from optimal images, reaching above 89% for all measurements and even reaching 100% in measurements such as LVEDV4A, LVEDV2A, RAA, RVEDA, RVESA. This trend was consistent in both stress test phases.

Agreement between cardiologists and ML within image quality groups

We compared the agreement between ML and cardiologists in assessing image quality (*Table 5*). Our analysis showed that there was higher accuracy and less variability in measurements for images rated as good or optimal quality compared with those rated as poor or fair. Specifically, when looking at left heart volume measurements, there was excellent agreement between ML and cardiologists for images of optimal and good

quality, with ICC ranging from 0.91 to 1.00. For right heart measurements, there was higher variability, but still good to excellent agreement (ICC ranged from 0.72 to 0.97). On the other hand, measurements in poor or fair quality images showed lower agreement compared with good or optimal quality images, but still fell within the good to excellent range (ICC ranged from 0.83 to 0.96). It is worth noting that right heart measurements in poor or fair quality images showed moderate agreement (ICC ranged from 0.45 to 0.71), indicating more significant variability. This pattern was consistent across both rest and stress phases.

Furthermore, we found that the impact of image quality on measurement accuracy was more noticeable for right heart measurements compared with left heart measurements. This was evident in a two- to three-fold lower measurement error and at least a two-fold reduction in variability range (lower and upper limits of agreement) in good and optimal quality images compared with poor or fair images.

Discussion

This study represents one of the first multi-centre trials evaluating the performance of an AI-based ML system for volumetric assessment of heart chambers during pharmacological SE. By including data from five centres, our findings demonstrated that AI-enabled volumetric analysis achieves high reproducibility of AI vs. expert cardiologists' measurements, particularly for LV and LA volumes. The study highlights AI's ability to automate chamber quantification during pharmacological SE, providing a potential pathway for addressing current challenges associated with traditional SE practices.

Volumetric assessment of heart chambers during SE is gaining recognition as a valuable tool, complementing existing diagnostic algorithms. By capturing changes in LV and LA volumes, SE can provide detailed insights into global systolic and diastolic function.[10,11] Elevated LV volumes under stress may indicate impaired contractile reserve, while increased LA volumes are markers of chronically elevated filling pressures, often associated with diastolic dysfunction and adverse outcomes.[10,21] Despite its potential applications of AI-driven volumetric analysis in pharmacological SE remain limited. Previous trials have demonstrated the feasibility of CNNs for volumetric analysis in resting echocardiography. Zhang *et al.* introduced a fully automated system for the precise measurement of LV and LA volumes, representing a significant advancement in cardiac imaging automation.[18] Tromp *et al.* conducted multiple studies, showing strong correlations and yields exceeding 87% for LV and LA volumetric measurements, findings consistent with the robust agreement observed in our study across rest and stress phases.[22,23] Another study focusing on STEMI patients introduced an AI-based system for evaluating LV volumes, GLS, and LAV, demonstrating strong agreement with manual measurements, with correlations ranging from 0.81 to 0.92.[24] Myhre *et al.* demonstrated that AI-derived measurements of LV volumes and LAV significantly reduced inter-operator variability while maintaining performance comparable to human operators, with correlation coefficients of 0.89 for LVEDV and 0.92 for LVESV.[25] Similarly, our study confirmed high agreement for LV volumes and LAV, highlighting AI's reliability across both rest and stress phases. While prior volumetric studies have primarily focused on resting echocardiography, our findings emphasize the complexities and advantages of assessing these metrics under stress conditions. Maintaining high image quality during stress is crucial, as it ensures that volumetric measurements remain reliable and accurate, comparable with the precision achieved under resting conditions. While ICCs indicated excellent agreement, LOA analysis revealed that in some instances discrepancies between AI and human LV volume measurements exceeded 35 mL. Such differences may have clinical relevance, especially when assessing borderline values for decision-making. Even though AI showed strong agreement with humans for LV volumes, LVEF demonstrated only moderate agreement. This discrepancy likely reflects error reproduction, as LVEF is a derived parameter highly sensitive to small differences in EDV and ESV. Even minimal tracing variability can disproportionately affect LVEF. This underscores the importance of expert oversight in unclear cases.

Despite its diagnostic potential, the evaluation of RV and RA volumes during SE remains underexplored, primarily due to the inherent complexities of right heart imaging. Intricate geometry and anatomical positioning of RV present challenges for accurate assessment.[26] However, right heart function during SE is increasingly recognized as a critical component in assessing cardiovascular conditions, including pulmonary hypertension and advanced CAD, often involving right coronary artery damage. Our study observed moderate to good agreement for RA and RV area measurements during pharmacological SE, with higher agreement at rest compared to stress conditions, especially as image quality declines under stress. In a study focusing on paediatric patients, AI-derived measurements of FAC showed moderate correlations with cardiologists' assessments, which aligns with the ongoing challenges observed in our trial, where lower agreement was noted during both phases.[27] RA area measurements in Tromp *et al.*'s study showed slightly better outcomes, with strong correlations and yields exceeding 90% under resting conditions.[23] In comparison, our findings demonstrated moderate correlations, with yields at 83% at rest and declining to 78% during stress. This discrepancy can be attributed to Tromp *et al.*'s exclusion of low-quality images and differences in patient positioning during SE, which likely contributed to their stronger performance.

Challenges in SE include operator dependence, poor acoustic windows, suboptimal image quality, and time-consuming workflows. Olaisen *et al.* showed that real-time AI integration in resting echocardiography could reduce acquisition and processing times by up to 77%, streamlining workflows without compromising accuracy.[28] Moreover, Mor-Avi *et al.* reported that novice operators using AI-assisted systems were able to acquire diagnostic-quality images comparable to expert sonographers, suggesting AI's potential to reduce reliance on specialized SE expertise.[29] Additionally, poor image quality is linked to higher variability and lower diagnostic accuracy. Our study, similar to others in the field, emphasized the influence of suboptimal and poor image quality on diagnostic yield and concordance with experts.[30] Notably, our findings revealed more significant variations with changes in image quality, prompting questions about whether these differences arise from variations in baseline image quality or the use of more advanced neural networks in other studies. Interestingly, optimal image quality did not consistently enhance accuracy, possibly due to training datasets being predominantly composed of good and fair-quality images. The high prevalence of good/optimal LV images in our dataset partly reflects expert frame selection, which may not fully mirror real-world practice where fair image quality is frequent. This could have positively influenced AI performance. Developing AI algorithms to handle suboptimal imaging conditions remains a critical area for future research.

AI applications in CAD detection using SE have shown promising results in recent trials, particularly in disease detection and risk prediction.[31] Upton *et al.* demonstrated that AI models significantly increased inter-reader agreement as well as sensitivity for detection of disease by 10%, compared to traditional SE interpretations.[4] Similarly, O'Driscoll *et al.* reported that AI-derived peak LVEF and GLS significantly enhanced CAD detection when combined with traditional wall motion scoring.[32] The PROTEUS trial, though not meeting non-inferiority criteria against expert assessments, indicated AI's potential utility in low-volume centres with limited expertise.[32,33] These findings highlight the potential of AI-driven SE to enhance diagnostic workflows, especially in resource-limited settings.

Our findings demonstrate that AI-based volumetric quantification during pharmacological SE is ready for clinical implementation, providing reliable measurements that closely match expert interpretation.

The next step should be the application of this technology to exercise SE, which represents the guideline-recommended first-line modality due to its physiological relevance and superior prognostic value. Given the additional technical challenges of exercise testing, including motion artefacts and accelerated heart rates, validating AI performance in this setting will be essential for ensuring broader clinical adoption.

Limitations

The limitations of our study should be acknowledged. First, although the multi-centre design enhances generalizability, the sample size remains limited, and further studies with larger cohorts are needed to validate these findings. Second, AI analysis was limited to image frames selected by human experts, which may have introduced bias and overestimated the agreement between AI and human measurements. Real-world validation with video-based AI analysis is essential for evaluating the feasibility of fully automated pharmacological SE workflows. Third, each study was analysed by a single rater, and results might differ if multiple raters were involved. However, all raters were experienced, certified, and underwent additional training, ensuring consistency across centres. In majority of cases, we used dipyridamole, but it is the one recommended to combine RWMA and CFVR assessment, and by far the stress with less degradation of image quality, due to limited tachycardia and lack of hyperventilation and excessive chest wall movement.[34–36]

Finally, this study focused exclusively on volumetric parameters, excluding potentially valuable metrics such as GLS and coronary flow reserve, which have shown promise in CAD diagnosis. Future studies should explore integrating these parameters into AI-based SE analysis.

Conclusions

This multi-centre study demonstrated the strong potential of ML for automated volumetric evaluation during pharmacological SE, achieving high yields and strong correlations with expert cardiologists' measurements for LV volumes, LVEF, and atrial areas. While the CNNs performed well for most parameters, their accuracy for certain right-sided measurements and FAC was more limited. These findings underscore the promise of ML to enhance diagnostic workflows and support clinical decision-making in SE.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Funding

This research was co-funded by the European Union under HORIZON EIC Accelerator program grant agreement no. 190190768.

Conflict of interest: The following participants of the study are employees of Ligence, Ltd: A.K., A.K., K.S., I.K. and D.V.

Data availability

The data underlying this article could be shared on reasonable request to the corresponding author.

References

1. Timmis A, Aboyans V, Vardas P, Townsend N, Torbica A, Kavousi M, *et al.* European society of cardiology: the 2023 atlas of cardiovascular disease statistics. *Eur Heart J* 2024;**45**:4019–4062.
2. Martin SS, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, *et al.* 2024 Heart disease and stroke statistics: a report of US and global data from the American Heart Association [published correction appears in circulation. 2024 May 7;149(19): e1164. Doi: 10.1161/CIR.0000000000001247.]. *Circulation*. 2024;**149**:e347–e913
3. Vrints C, Andreotti F, Koskinas KC, Rossello X, Adamo M, Ainslie J, *et al.* 2024 ESC guidelines for the management of chronic coronary syndromes [published correction appears in Eur Heart J. 2025 Apr 22;46(16):1565]. *Eur Heart J* 2024;**45**:3415–3537.
4. Upton R, Mumith A, Beqiri A, Parker A, Hawkes W, Gao S, *et al.* Automated echocardiographic detection of severe coronary artery disease using artificial intelligence. *JACC Cardiovasc Imaging* 2022;**15**:715–727.
5. Medina R, Panidis IP, Morganroth J, Kotler MN, Mintz GS. The value of echocardiographic regional wall motion abnormalities in detecting coronary artery disease in patients with or without a dilated left ventricle. *Am Heart J* 1985;**109**:799–803.
6. Roggel A, Jehn S, Dykun I, Balcer B, Al-Rashid F, Totzeck M, *et al.* Regional wall motion abnormalities on focused transthoracic echocardiography in patients presenting with acute chest pain: a predefined post hoc analysis of the prospective single-centre observational EPIC-ACS study. *BMJ Open* 2024;**14**:e085677.
7. Ihekwaba U, Johnson N, Choi JS, Savarese G, Orsini N, Khoo J, *et al.* Long-term prognostic value of contemporary stress echocardiography in patients with suspected or known coronary artery disease: systematic review and meta-analysis. *Heart* 2024; **110**:1349–1356.
8. Picano E, Ciampi Q, Arbucci R, Cortigiani L, Zagatina A, Celutkiene J, *et al.* Stress Echo 2030: the new ABCDE protocol defining the future of cardiac imaging. *Eur Heart J Suppl* 2023;**25**:C63–C67.
9. Picano E, Pierard L, Peteiro J, Djordjevic-Dikic A, Sade LE, Cortigiani L, *et al.* The clinical use of stress echocardiography in chronic coronary syndromes and beyond coronary artery disease: a clinical consensus statement from the European association of cardiovascular imaging of the ESC. *Eur Heart J Cardiovasc Imaging* 2024;**25**:e65–e90.
10. Prota C, Cortigiani L, Campagnano E, Wierzbowska-Drabik K, Kasprzak J, Colonna P, *et al.* Left atrium stress echocardiography: correlation between left atrial volume, function, and B-lines at rest and during stress. *Explor Cardiol* 2024;**2**:19–30.
11. Pellikka PA, Arruda-Olson A, Chaudhry FA, Chen MH, Marshall JE, Porter TR, *et al.* Guidelines for performance, interpretation, and application of stress echocardiography in ischemic heart disease: from the American society of echocardiography. *J Am Soc Echocardiogr* 2020;**33**:1–41.e8.
12. Gupta K, Kakar TS, Gupta A, Singh A, Gharpure N, Aryal S, *et al.* Role of left ventricle deformation analysis in stress echocardiography for significant coronary artery disease detection: a diagnostic study meta-analysis. *Echocardiography* 2019;**36**:1084–1094.
13. Frydas A, Morris DA, Belyavskiy E, Radhakrishnan AK, Kropf M, Tadic M, *et al.* Left atrial strain as sensitive marker of left ventricular diastolic dysfunction in heart failure. *ESC Heart Fail* 2020;**7**:1956–1965.
14. Pellikka PA. Artificially intelligent interpretation of stress echocardiography: the future is now. *JACC Cardiovasc Imaging* 2022;**15**:728–730.
15. Sveric KM, Ulbrich S, Dindane Z, Winkler A, Botan R, Mierke J, *et al.* Improved assessment of left ventricular ejection fraction using artificial intelligence in echocardiography: a comparative analysis with cardiac magnetic resonance imaging. *Int J Cardiol* 2024;**394**: 131383.
16. Kshatri SS, Singh D. Convolutional neural network in medical image analysis: a review. *Arch Comput Methods* 2023;**30**:2793–2810.
17. Huang MS, Wang CS, Chiang JH, Liu PY, Tsai WC. Automated recognition of regional wall motion abnormalities through deep neural network interpretation of transthoracic echocardiography. *Circulation* 2020;**142**:1510–1520.
18. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, *et al.* Fully automated echocardiogram interpretation in clinical practice. *Circulation* 2018;**138**: 1623–1635.
19. Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, *et al.* A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC Cardiovasc Imaging* 2020;**13**:374–381.
20. Galderisi M, Cosyns B, Edvardsen T, Cardim N, Delgado V, Di Salvo G, *et al.* Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European association of cardiovascular imaging. *Eur Heart J Cardiovasc Imaging* 2017;**18**:1301–1310.
21. Wang Y, Yin L. Noninvasive identification and therapeutic implications of supernormal left ventricular contractile phenotype. *Explor Cardiol* 2024;**2**:97–113.
22. Tromp J, Seekings PJ, Hung CL, Iversen MB, Frost MJ, Ouwerkerk W, *et al.* Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *Lancet Digit Health* 2022;**4**:e46–e54.
23. Tromp J, Bauer D, Claggett BL, Frost M, Iversen MB, Prasad N, *et al.* A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *Nat Commun* 2022;**13**:6776.
24. Jang Y, Choi H, Yoon YE, Jeon J, Kim H, Kim J, *et al.* An artificial intelligence-based automated echocardiographic analysis: enhancing efficiency and prognostic evaluation in patients with revascularized STEMI. *Korean Circ J* 2024;**54**:743–756.
25. Myhre PL, Gaibazzi N, Tuttolomondo D, Sartorio D, Ugolotti PT, Covani M, *et al.* Concordance of left ventricular volumes and function measurements between two

human readers, a fully automated AI algorithm, and the 3D heart model. *Front Cardiovasc Med* 2024;**11**:1400333.
26. DiLorenzo MP, Bhatt SM, Mercer-Rosa L. How best to assess right ventricular function by echocardiography. *Cardiol Young* 2015;**25**:1473–1481.
27. Vasile CM, Bouteiller XP, Avesani M, Velly C, Chan C, Jalal Z, *et al.* Exploring the potential of artificial intelligence in pediatric echocardiography-preliminary results from the first pediatric study using AI software developed for adults. *J Clin Med* 2023;**12**: 3209.
28. Olaisen S, Smistad E, Espeland T, Hu J, Pasdeloup D, Østvik A, *et al.* Automatic measurements of left ventricular volumes and ejection fraction by artificial intelligence: clinical validation in real time and large databases. *Eur Heart J Cardiovasc Imaging* 2024;**25**: 383–395.
29. Mor-Avi V, Khandheria B, Klempfner R, Cotella JI, Moreno M, Ignatowski D, *et al.* Real-time artificial intelligence-based guidance of echocardiographic imaging by novices: image quality and suitability for diagnostic interpretation and quantitative analysis. *Circ Cardiovasc Imaging* 2023;**16**:e015569.
30. Kim S, Park HB, Jeon J, Arsanjani R, Heo R, Lee SE, *et al.* Fully automated quantification of cardiac chamber and function assessment in 2-D echocardiography: clinical feasibility of deep learning-based algorithms. *Int J Cardiovasc Imaging* 2022;**38**:1047–1059.
31. Hadida Barzilai D, Cohen-Shelly M, Sorin V, Zimlichman E, Massalha E, Allison TG, *et al.* Machine learning in cardiac stress test interpretation: a systematic review. *Eur Heart J Digit Health* 2024;**5**:401–408.
32. O'Driscoll JM, Hawkes W, Beqiri A, Mumith A, Parker A, Upton R, *et al.* Left ventricular assessment with artificial intelligence increases the diagnostic accuracy of stress echocardiography. *Eur Heart J Open* 2022;**2**:oeac059.
33. Upton R, Akerman AP, Marwick TH, Johnson CL, Piotrowska H, Bajre M, *et al.* PROTEUS: a prospective RCT evaluating use of AI in stress echocardiography. *NEJM AI* 2024;**1**:00865.
34. Martin TW, Seaworth JF, Johns JP, Pupa LE, Condos WR. Comparison of adenosine, dipyridamole, and dobutamine in stress echocardiography. *Ann Intern Med* 1992;**116**: 190–196.
35. Sochowski RA, Yvorchuk KJ, Yang Y, Rattes MF, Chan KL. Dobutamine and dipyridamole stress echocardiography in patients with a low incidence of severe coronary artery disease. *J Am Soc Echocardiogr* 1995;**8**:482–487.
36. Beleslin BD, Ostojic M, Stepanovic J, Djordjevic-Dikic A, Stojkovic S, Nedeljkovic M, *et al.* Stress echocardiography in the detection of myocardial ischemia. Head-to-head comparison of exercise, dobutamine, and dipyridamole tests. *Circulation* 1994;**90**: 1168–1176.