



Automated echocardiographic measurements for longitudinal monitoring of ATTR cardiomyopathy: agreement and repeatability analysis

Amely Walser¹ · Olivier F. Clerc^{2,3} · Constantin Mork¹ · Andreas J. Flammer¹ · Peder L. Myhre⁴ · Rahel Schwotzer⁵ · Christoph Gräni⁶ · Frank Ruschitzka¹ · Felix C. Tanner¹ · Dominik C. Benz^{1,7,8}

Received: 17 June 2026 / Accepted: 20 September 2026

© The Author(s) 2026

Abstract

Detection of disease progression is key to personalize treatment strategies in transthyretin cardiomyopathy (ATTR-CM), particularly with emerging therapies. Echocardiography can detect subtle longitudinal changes but is limited by operator dependence. This study evaluates agreement and reproducibility of fully automated, AI-assisted echocardiographic measurements under real-world conditions. This retrospective study included 62 patients with ATTR-CM undergoing 178 serial annual echocardiograms assessed by a reference human reader, a second cardiologist (cardiologist 2), a novice reader, and a fully automated AI algorithm (Us2.ai). Interrater agreement was assessed using Bland–Altman analysis and intraclass correlation coefficients (ICCs). Intrarater variability for human readers was derived from repeated blinded measurements, with limits of agreement (LoA = mean difference $\pm$ 1.96 $\times$ SD) defining the minimal detectable change. AI repeatability was assessed using within-study automated pipeline variability across available acquisitions. AI showed moderate agreement with the reference human reader for IVSd and LVEDV (ICC 0.65 and 0.51), with biases of -1.9 mm and -39 mL, respectively. Interrater agreement between cardiologists was good (ICC 0.79 and 0.84) with minimal bias (-0.2 mm and $+3$ mL). Intrarater variability was moderate to excellent for both cardiologists (LoA 3.0 mm and 43 mL for the reference human reader; 2.7 mm and 44 mL for cardiologist 2). AI demonstrated comparable repeatability (LoA 3.6 mm and 37 mL), while the novice showed higher variability (5.1 mm and 61 mL). Fully automated measurements showed within-study variability of similar magnitude to the intrarater variability of experienced readers, although obtained with different experimental designs, supporting further prospective evaluation for serial assessment. Agreement in absolute terms was moderate for wall thickness and LV volumes, indicating expert oversight remains necessary in advanced disease.

✉ Dominik C. Benz
dominik.benz@usz.ch

¹ Department of Cardiology, University Hospital Zurich, Zurich, Switzerland

² Amyloidosis Program, Division of Cardiology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

³ CV Imaging Program, Cardiovascular Division, Department of Medicine and Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

⁴ Department of Cardiology, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway

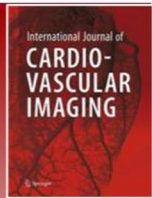
⁵ Department of Medical Oncology and Hematology, University Hospital Zurich, Zurich, Switzerland

⁶ Department of Cardiology, Bern University Hospital, Inselspital, University of Bern, Bern, Switzerland

⁷ Advanced Clinician Scientist Program, University Medicine Zurich (UMZH), Zurich, Switzerland

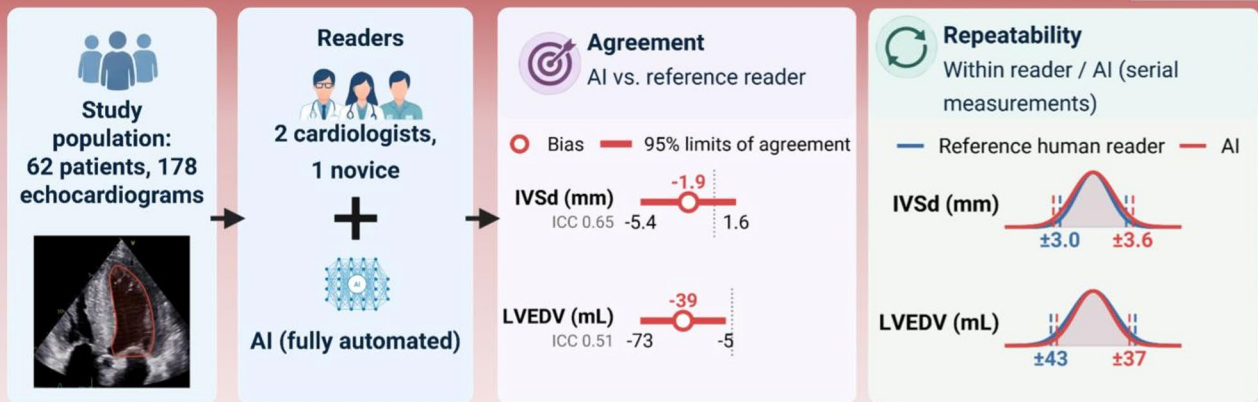
⁸ Departments of Nuclear Medicine and Cardiology, University Hospital of Zurich, Ramistrasse 100, Zurich 8091, CH, Switzerland

Graphical abstract



Automated echocardiographic measurements for longitudinal monitoring of ATTR cardiomyopathy

Agreement and repeatability analysis



Key Findings

- **Comparable repeatability:** within-study variability of AI similar to intrarater variability of expert readers
- **Moderate agreement:** systematic underestimation by AI, expert oversight remains necessary
- **Implication:** Supports prospective evaluation of automated measurements for serial monitoring

Keywords Cardiac amyloidosis · Transthyretin · Echocardiography · Artificial intelligence · Reproducibility · Disease monitoring

Abbreviations

AI	artificial intelligence
ATTR-CM	transthyretin cardiomyopathy
CMR	cardiac magnetic resonance
EDV	end-diastolic volume
ICC	intraclass correlation coefficient
IVSd	interventricular septal thickness in diastole
LoA	limits of agreement
LV	left ventricular
LVEF	left ventricular ejection fraction
PWd	posterior wall thickness in diastole

Introduction

Transthyretin cardiomyopathy (ATTR-CM) is caused by the accumulation of amyloid fibrils in the myocardium. Disease-modifying therapies, including TTR stabilizers such as tafamidis and acoramidis, as well as TTR silencers such as vutrisiran, significantly reduce morbidity and mortality in

most patients with ATTR-CM [1–3]. However, around 30% of patients still experience disease progression [1]. With new therapies emerging, understanding the disease course and identifying treatment response early are crucial to address this risk and transform patient management and outcome. Currently, disease progression is monitored through functional capacity, quality of life, or N-terminal pro brain natriuretic peptide (NT-proBNP) [4]. However, these markers change slowly over time and lack specificity for cardiac involvement and disease progression, as they are influenced by various extra-cardiac factors [4]. Therefore, more precise biomarkers are needed to monitor disease progression, leading to a shift towards cardiac imaging for monitoring disease progression [5]. Serial echocardiographic assessments, due to their accessibility and cost-effectiveness, are a promising tool for monitoring disease progression, especially when assessed in a core lab [6]. Even small changes in echocardiographic parameters are shown to have prognostic significance [7]. However, manual echocardiographic measurements, often by various operators, may lack the precision in daily practice to reliably detect such subtle changes.

Artificial intelligence (AI), with its capacity to minimize measurement variability [8], presents a unique opportunity to enhance the precision of serial echocardiographic evaluations. By improving precision, AI could enable detection of subtle longitudinal changes. In this study, we evaluated the potential of automated AI-based echocardiographic measurements for longitudinal assessment in ATTR cardiomyopathy by comparing AI-derived measurements with those obtained by a reference human reader, a second cardiologist (cardiologist 2) and a novice reader. Because routine echocardiographic studies typically include multiple acquisitions of the same view, automated analysis may be performed on different images than those used for human measurements, leading to slightly different values across these acquisitions. In this context, variability in AI-derived measurements primarily reflects differences between image acquisitions and selection of image rather than inconsistencies in the measurement algorithm itself. This study therefore deliberately accepted this source of variability to mirror a real-world approach, where repeated measurements are typically performed without systematic documentation of the specific acquisitions used.

We hypothesized that AI-based measurements would show good agreement with experienced human readers and that their repeatability would be comparable to expert measurements while being less variable than the interrater differences observed between human readers. We therefore assessed both agreement between raters and repeatability of measurements to better understand the role of fully automated analysis for real-world longitudinal monitoring.

Materials and methods

Participant selection criteria

This longitudinal cohort study, approved by the Zurich Cantonal Ethics Committee (KEK-ZH 2014–0490), included 62 patients with ATTR-CM from the University Hospital Zurich Amyloidosis Registry. Inclusion criteria were a confirmed diagnosis of ATTR-CM and availability of at least two echocardiograms performed 12 months apart. ATTR-CM was diagnosed according to established international criteria, requiring grade 2 or 3 myocardial uptake on bone scintigraphy together with exclusion of monoclonal protein. No further exclusion criteria were applied. All participants provided written informed consent. Patients underwent serial echocardiograms between 2013 and 2024, performed at the University Hospital Zurich or other Swiss hospitals or private practices. Each patient had two echocardiograms 12 months apart (baseline and follow-up), with 25 patients undergoing additional annual follow-up echos, resulting in

a total of 178 studies. Agreement analyses were performed on cases with complete data for each respective parameter. New York Heart Association (NYHA) functional class and NT-proBNP were documented within 3 months of the baseline echocardiography.

Image acquisition and analysis

Comprehensive echocardiograms were acquired according to the recommendations of the American Society of Echocardiography [9]. Echocardiograms were analysed by four raters: (i) an AI-based algorithm, (ii) a board-certified cardiologist with several years of experience, who served as the reference human reader for all agreement analyses, (iii) a recently board-certified cardiologist (referred to as “cardiologist 2” hereafter) and (iv) a novice medical student without prior echocardiography experience (referred to as “novice” hereafter). Several parameters, including interventricular septal thickness in diastole (IVSd), posterior wall thickness in diastole (PWd), left ventricular (LV) end-diastolic volume (EDV), LV end-systolic volume (ESV), LV ejection fraction (LVEF), and E/e’ ratio were measured. The novice received about 12 h of training, including initial instruction, self-study, and review of 5 cases with the reference human reader.

The AI-based algorithm (Us2.ai) algorithm performs fully automated echocardiographic analysis, calculating parameters without manual input [10, 11]. It uses deep learning to classify video clips into A4C, A3C, or A2C views and filters out low-quality images. A convolutional neural network contours the endocardial border for each frame, automatically detecting end-diastolic and end-systolic frames based on video volume curves, with support from an electrocardiogram when available [12]. The algorithm withholds measurements that do not meet its internal confidence and quality criteria rather than reporting them. The software provides an interface for reviewing and correcting contours and measurements, but this was deliberately not used and all values were analysed exactly as returned by the fully automated algorithm.

For human raters, LV volumes and ejection fraction (EF) were calculated using the modified Simpson’s rule following the 2015 American Society of Echocardiography (ASE)/European Association of Cardiovascular Imaging (EACVI) recommendations for cardiac chamber quantification [13]. Unless stated otherwise, all LV volumes and LVEF reported in this study are biplane measurements, for the human raters and for the AI algorithm alike. Single-plane four-chamber volumes were used exclusively for the repeatability analyses. Echocardiographic measurements by the cardiologists and novice were performed using dedicated software (TOMTEC Image Arena, TOMTEC Imaging Systems GmbH).

For AI-human agreement, the single, automatically selected optimal value per parameter from AI were compared to the reference human reader. For the AI algorithm, within-study automated pipeline variability across available acquisitions was quantified: for each study, the software returns a measurement for every acquisition of a suitable view that passes its internal quality criteria, and we computed the differences between these measurements within the same study. This quantity therefore reflects the combined effect of acquisition-to-acquisition differences and automated loop selection within a single examination, and it is not a direct measure of algorithmic repeatability. To evaluate human intrarater variability, 20 echocardiograms were re-measured by all raters after a 3-month interval. For intrarater variability assessment, original loops were not retained. Instead, the most suitable loop for measurements was selected at baseline and after the 3-month interval separately to ensure conditions comparable to the AI algorithm, which in real-world selects the best echocardiographic loop of each study for measurement. In addition, echocardiographic measurements were compared against cardiac magnetic resonance imaging (CMR). A total of 20 CMR examinations, performed and reported by cardiologists independent of the echocardiographic raters, were available. These were clinically indicated studies acquired according to the standard departmental protocol rather than a study-specific standardised protocol. The median interval between CMR and the corresponding echocardiogram was 2.5 months (interquartile range (IQR): 0.8, 4.8; range 2–176 days). IVSd, PWD, LV EDV and LVEF were compared.

Statistical Analysis

Continuous variables are summarized as median (IQR), and categorical variables as frequency (percentage).

(1) **Interrater agreement** between each rater and the reference human reader was assessed using Bland–Altman analysis, reporting the mean bias and 95% limits of agreement ($\text{LoA} = \text{mean difference} \pm 1.96 \times \text{SD of differences}$). Intraclass correlation coefficients (ICCs) were calculated using a two-way random-effects model for single measures, and relative absolute differences were reported as an additional measure of agreement. ICC values were interpreted according to established thresholds: $\text{ICC} < 0.50$ was considered poor, $0.50\text{--}0.74$ moderate, $0.75\text{--}0.89$ good, and ≥ 0.90 excellent agreement.

(2) **Agreement with CMR** was assessed for a subset of 20 patients with available CMR performed within a median of 2.5 months of the corresponding echocardiogram, using Bland–Altman analysis and ICCs (two-way random-effects model for single measures).

(3) **Intrarater variability** of human raters was assessed by repeated blinded measurements of 20 echocardiograms

performed after a 3-month interval by each rater independently, without access to prior measurements. Variability was expressed as limits of agreement ($\text{LoA} = \text{mean difference} \pm 1.96 \times \text{SD of differences}$), and ICC was calculated using a two-way mixed-effects model for agreement. Based on the principle that LoA encompass 95% of differences between repeated measurements in a stable patient, differences beyond this range are unlikely to reflect measurement variability alone and were therefore defined as the minimal detectable change (MDC) [14]. Within-study repeatability for the AI algorithm was derived from pairwise differences across multiple automated measurements per study as described above, with equivalent LoA calculated from the weighted mean difference and pooled SD. Because pairwise differences derived from the same examination are not independent, these estimates were additionally verified using a variance-component model with automated measurements nested within examinations and examinations nested within patients. Thresholds were derived separately for each rater, as they reflect the precision of that individual reader and the readers differed substantially in experience. They are not intended as thresholds for clinical use.

(4) **Changes exceeding the minimal detectable change** were assessed by comparing, for each rater, the absolute difference between consecutive 12-month echocardiograms with the rater-specific LoA (i.e. the MDC) derived under (3). The proportion of 12-month intervals with a change exceeding the MDC is reported descriptively for each rater. Because MDC thresholds are rater-specific, proportions were not formally compared between raters. For volume parameters, repeatability data were not available from biplane views; therefore 4-chamber views were used consistently across all raters to ensure comparability.

(5) **Longitudinal repeatability** was assessed exploratively using annual changes in echocardiographic parameters over consecutive 12-month intervals. All patients contributed at least one 12-month interval (baseline to first follow-up), and the 25 patients with additional annual echocardiograms contributed several consecutive intervals. For each interval, the annual change was calculated per parameter and rater as the difference between the two echocardiograms acquired 12 months apart. To assess whether treatment with TTR-modifying therapy influenced observed changes, intervals were classified as ‘with TTR-modifying treatment’ or ‘without TTR-modifying treatment’ based on treatment status at the interval midpoint, and mean annual changes were compared between groups using Wilcoxon rank-sum tests. Sensitivity analyses restricted to one examination and one interval per patient were performed to assess the robustness of the estimates.

All analyses were performed in R version 4.5.2 using the packages tidyverse, lubridate, stringr, skimr, janitor,

Table 1 Patient characteristics at baseline

Metric	Value
Demographics	
Age, years	73 (67–78)
Male sex	56 (90%)
Body mass index, kg/m ²	26 (24–29)
ATTRv	7 (11%)
ATTRwt	55 (89%)
NYHA class	2 (2–2)
Laboratory Findings	
NT-proBNP, pg/mL	1349 (562–2476)
Medication	
TTR-modifying treatment* at Baseline	21 (34%)
TTR-modifying treatment During Follow-Up	29 (47%)
TTR-modifying treatment Tafamidis	12 (20%)
Comorbidities	
Atrial Fibrillation	18 (29%)
CAD	21 (34%)
Carpal Tunnel Syndrome	12 (19%)
Diabetes	3 (5%)
Hypertension	21 (40%)
Neuropathy	12 (19%)

Continuous variables are presented as median (interquartile range) and categorical variables as frequency (percentage). ATTRv, variant transthyretin amyloidosis; ATTRwt, wild-type transthyretin amyloidosis; CAD coronary artery disease; NYHA, New York Heart Association; NT-proBNP, N-terminal pro-B-type natriuretic peptide. Missing data: For all clinical variables except hypertension, missingness was <25%

*TTR-modifying treatment included TTR stabilizers and silencers

Table 2 Baseline echocardiographic parameters

Metric	Reference Human Reader	AI	Cardiologist 2	Novice
IVSd (mm)	16 (15–18)	14 (12–16)*	16 (14–18)	17 (15–19)*
PWd (mm)	14 (12–16)	15 (11–17)*	15 (12–17)*	14 (12–15)
LV EDV biplane (ml)	124 (110–154)	88 (75–110)*	121 (106–144)*	122 (103–146)*
LV ESV biplane (ml)	64 (52–88)	37 (29–54)*	57 (48–78)*	60 (48–80)*
LVEF biplane (%)	49 (40–54)	56 (50–61)*	51 (44–57)*	50 (44–57)*
E/e' ratio	17 (12–21)	14 (11–17)*	15 (11–19)*	14 (11–18)*

Variables are presented as median (interquartile range). *p<0.05 versus the reference human reader (Wilcoxon signed-rank test on paired observations, Holm-adjusted within each parameter). Numbers of studies contributing to each measurement and exact p-values are given in Supplemental Table S1

EDV, end-diastolic volume; ESV, end-systolic volume; EF, ejection fraction; IVSd, interventricular septal thickness at diastole; LV, left ventricular; PWd, posterior wall thickness at diastole

irr, DescTools, lme4 and rstatix. Two-sided p-values<0.05 were considered statistically significant.

Results

Participant characteristics at baseline

A total of 62 participants with ATTR-CM (90% male) were included in the study, with a median age of 73 years (IQR: 67–78). ATTRv was present in 7 patients, with Val50Met being the most common variant. Atrial fibrillation was present in 29% of patients, 34% had coronary artery disease and 34% were treated with TTR-modifying treatment (mostly tafamidis) at baseline. (Table 1). Echocardiographic parameters by the reference human reader at baseline included median IVSd 16 (15–18) mm, LVEDV 124 (110–154) ml and LVEF 49% (40–54) (Table 2; full distributions in Supplemental Table S1).

Agreement of AI and human raters with the reference human reader

Interrater agreement metrics are summarized in Fig. 1 (Bland–Altman analysis for IVSd and LVEDV as representative parameters), Table 3 (interrater ICC values for all parameters), and Supplemental Figure S1 (relative absolute differences). AI showed the weakest overall agreement with the reference human reader. ICC was moderate for IVSd (0.65), LV volumes (0.51 for LVEDV, 0.49 for LVESV), LVEF (0.62), and SV (0.61), and good for PWd (0.81) and E/e' (0.76). AI showed mean biases of −1.9 mm for IVSd, −39 mL for LVEDV, and +6% for LVEF. Cardiologist 2 showed good agreement with the reference human reader across most parameters, with ICCs ranging from 0.70 to 0.88, and minimal biases (−0.2 mm for IVSd, +3 mL for LVEDV, 0% for LVEF). The novice showed good ICC for LV volumes (0.87) and E/e' (0.82), but only moderate agreement for wall thickness and LVEF, with a systematic overestimation of IVSd (+1.1 mm). Estimates were unchanged in sensitivity analyses restricted to the earliest examination per patient (for IVSd: ICC 0.64 versus 0.65 for AI, 0.84 versus 0.79 for cardiologist 2 and 0.77 versus 0.71 for the novice) and were similarly unchanged for all other parameters.

Agreement of AI and human raters with cardiac magnetic resonance imaging

Agreement metrics between echocardiographic measurements and CMR measurements are summarized in Fig. 2 (Bland–Altman analysis for IVSd and LVEDV), and Table 4 (ICC values for all parameters). For wall thickness, ICCs

Fig. 1 Bland-Altman plots for agreement of AI and human readers with the reference human reader. The solid line represents the mean difference (bias), while the dashed lines indicate the 95% limits of agreement (LOA), i.e. ± 1.96 standard deviations. Legend entries display bias (lower LOA, upper LOA) for each rater compared to the reference human reader. **(A)** IVSd, interventricular septal thickness at diastole; **(B)** LVEDV, left ventricular end-diastolic volume

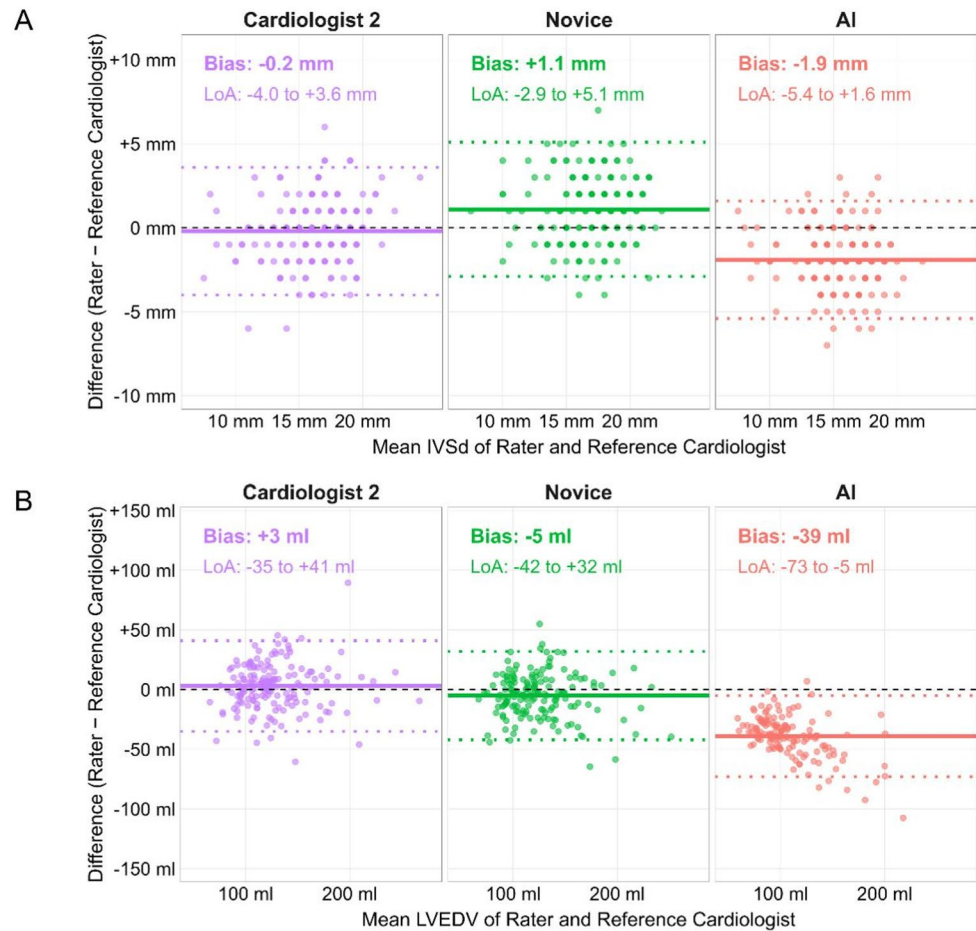


Table 3 Intraclass correlation coefficients (ICC) for agreement of echocardiographic metrics from AI, cardiologist 2 and novice compared to the reference human reader

Parameter	AI (ICC 95% CI)	Cardiologist 2 (ICC 95% CI)	Novice (ICC 95% CI)
IVSd	0.65 (0.07, 0.85)	0.79 (0.73, 0.85)	0.71 (0.48, 0.82)
PWd	0.81 (0.74, 0.86)	0.71 (0.61, 0.78)	0.62 (0.51, 0.71)
LV EDV biplane	0.51 (-0.07, 0.82)	0.84 (0.78, 0.89)	0.87 (0.81, 0.90)
LV ESV biplane	0.49 (-0.08, 0.81)	0.83 (0.75, 0.88)	0.89 (0.82, 0.93)
SV biplane	0.61 (-0.03, 0.84)	0.75 (0.65, 0.82)	0.71 (0.60, 0.79)
LVEF biplane	0.62 (0.12, 0.81)	0.70 (0.61, 0.78)	0.70 (0.61, 0.78)
E/e' ratio	0.76 (0.38, 0.89)	0.88 (0.78, 0.93)	0.82 (0.70, 0.89)

Red = poor ICC (ICC < 0.50). Orange = moderate ICC (0.50 < ICC < 0.75). green = good ICC (0.75 < ICC < 0.90).

CI, confidence interval; EDV, end-diastolic volume; ESV, end-systolic volume; EF, ejection fraction; ICC, intraclass correlation coefficient; IVSd, interventricular septal thickness at diastole; LV, left ventricular; PWd, posterior wall thickness at diastole; SV, stroke volume. Numbers of paired observations (n) per parameter and rater: IVSd – AI 166, Cardiologist 2 171, Novice 171; PWd – AI 163, Cardiologist 2 171, Novice 171; LV EDV biplane – AI 140, Cardiologist 2 166, Novice 165; LV ESV biplane – AI 140, Cardiologist 2 166, Novice 165; SV biplane – AI 140, Cardiologist 2 166, Novice 165; LVEF biplane – AI 141, Cardiologist 2 173, Novice 173; E/e' ratio – AI 131, Cardiologist 2 146, Novice 153.

Fig. 2 Agreement between echocardiographic measurements and CMR. The solid line represents the mean difference (bias), while the dashed lines indicate the 95% limits of agreement (LOA), i.e. ± 1.96 standard deviations. Legend entries display bias (lower LOA, upper LOA) for each rater compared to CMR. **(A)** IVSd, interventricular septal thickness at diastole; **(B)** LVEDV, left ventricular end-diastolic volume

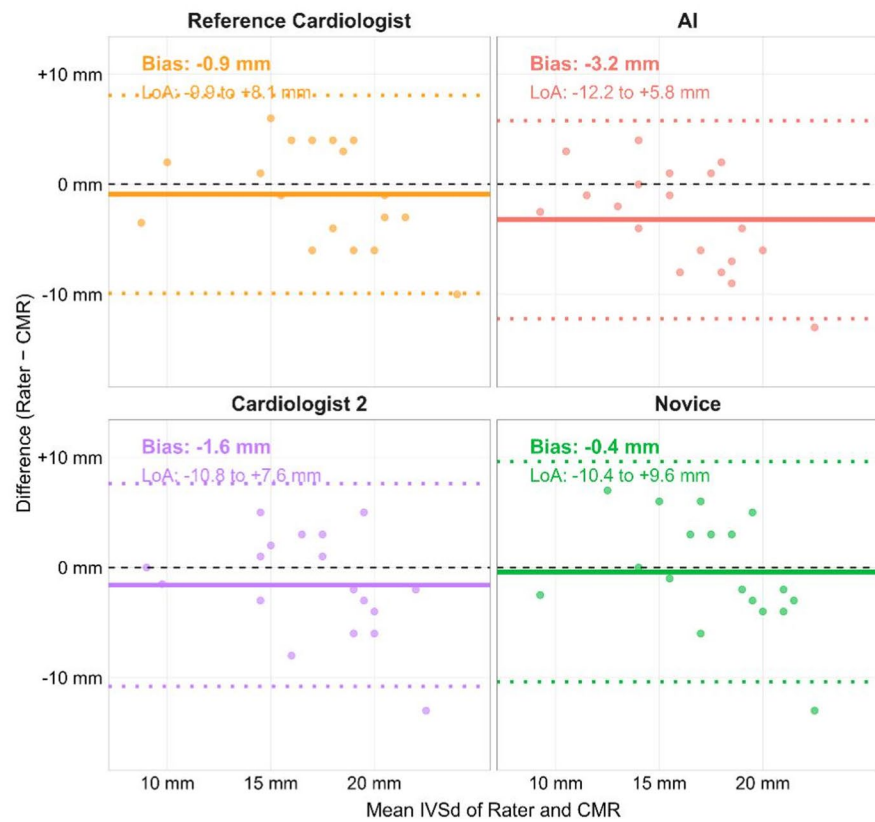
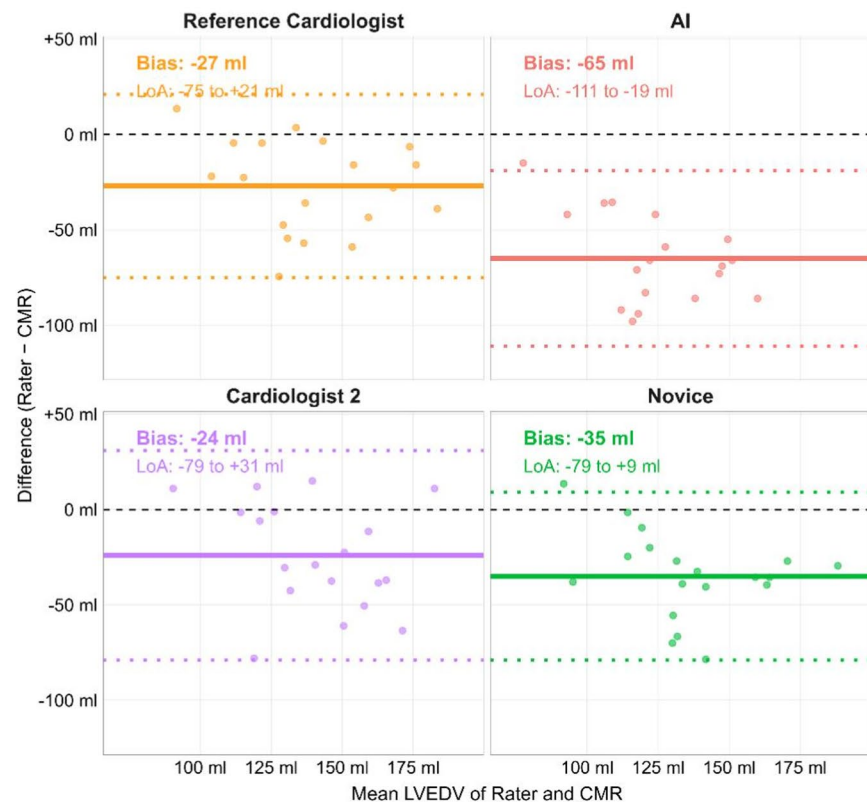
A**B**

Table 4 Intraclass correlation coefficients (ICC) for agreement of echocardiographic metrics with MR measurements

Parameter	Reference Human Reader (ICC 95% CI)	AI (ICC 95% CI)	Cardiologist 2 (ICC 95% CI)	Novice (ICC 95% CI)
IVSd	0.46 (0.03, 0.75)	0.32 (-0.07, 0.65)	0.39 (-0.03, 0.70)	0.29 (-0.19, 0.66)
LVEDV	0.43 (-0.09, 0.76)	0.12 (-0.05, 0.44)	0.29 (-0.09, 0.63)	0.37 (-0.11, 0.74)
LVEF	0.77 (0.40, 0.91)	0.81 (0.57, 0.93)	0.63 (0.28, 0.83)	0.63 (0.24, 0.84)
PWd	0.37 (-0.11, 0.74)	0.39 (-0.07, 0.74)	0.54 (0.02, 0.83)	0.42 (-0.09, 0.77)

Red = poor ICC (ICC < 0.50), Orange = moderate ICC (0.50 < ICC < 0.75), green = good ICC (0.75 < ICC < 0.90).

CI, confidence interval; EDV, end-diastolic volume; EF, ejection fraction; ICC, intraclass correlation coefficient; IVSd, interventricular septal thickness at diastole; LV, left ventricular; PWd, posterior wall thickness at diastole

Table 5 Intraclass correlations coefficients (ICC) for intrarater variability

Parameter	Reference Human Reader (ICC 95% CI)	Cardiologist 2 (ICC 95% CI)	Novice (ICC 95% CI)
IVSd	0.80 (0.58, 0.91)	0.90 (0.77, 0.95)	0.47 (-0.03, 0.76)
PWd	0.74 (0.48, 0.88)	0.80 (0.58, 0.91)	0.50 (0.07, 0.76)
LV EDV biplane	0.83 (0.55, 0.87)	0.87 (0.36, 0.96)	0.71 (0.44, 0.87)
LV ESV biplane	0.80 (0.51, 0.92)	0.91 (0.75, 0.97)	0.72 (0.45, 0.87)
LVEF biplane	0.84 (0.66, 0.93)	0.80 (0.59, 0.91)	0.86 (0.70, 0.94)
E/e' ratio	0.97 (0.93, 0.99)	0.95 (0.88, 0.98)	0.85 (0.67, 0.94)

Red = poor ICC (ICC < 0.50), Orange = moderate ICC (0.50 < ICC < 0.75), green = good ICC (0.75 < ICC < 0.90), dark green = excellent ICC (ICC > 0.90).

CI, confidence interval; EDV, end-diastolic volume; EF, ejection fraction; ESV, end-systolic volume; ICC, intraclass correlations coefficients; IVSd, interventricular septal thickness at diastole; LV, left ventricular; PWd, posterior wall thickness at diastole

indicated poor agreement with CMR for all raters for both IVSd and PWd. All raters slightly underestimated IVSd relative to CMR (biases ranging from -0.4 mm to -3.2 mm), while all raters slightly overestimated PWd (+0.9 mm to +2.3 mm). For LV volumes, ICCs indicated poor agreement with CMR across all raters, with all raters systematically underestimating CMR-derived LVEDV. AI showed the largest bias (-65 ml), followed by the reference human reader (-27 ml), and cardiologist 2 (-24 ml). For LVEF, ICCs indicated moderate to good agreement with CMR across all raters. Human raters showed small negative biases (-3% to -5%), consistent with the known tendency of echocardiography to underestimate CMR-derived LVEF, while AI showed minimal positive bias (+1%).

Intrarater variability

The repeatability for human measurements was assessed by ICCs (Table 5). Mean differences and 95% limits of agreement for repeated measurements are summarized in Supplemental Table S2. Intrarater variability was moderate to excellent in the reference human reader and cardiologist 2, with lower variability than the novice, where intrarater variability was poor for IVSd and moderate for PWd as well as volumes (Table 5).

Histograms depict differences for repeated measurements by human raters and pairwise differences across multiple automated measurements within the same echocardiographic study by AI (Fig. 3). For the AI-algorithm, within-study repeatability showed LoA of 3–4 mm for IVSd, 4–5 mm for PWd, and 30–40 mL for LVEDV, which were numerically comparable to those of the expert cardiologists.

Repeatability coefficients from the variance-component model were essentially identical to these limits. The novice demonstrated higher variability, with LoA of 5 mm for IVSd, 7 mm for PWD, and 60 mL for LVEDV.

Longitudinal repeatability

Annual changes over consecutive 12-month intervals are summarized in Fig. 4 and Supplemental Table 3. Overall, temporal changes in the echocardiographic parameters were minimal across all raters, with median changes close to zero for wall thickness parameters and small volume changes for most raters. There were no significant differences in temporal trends for patients treated with vs. without TTR-modifying treatment, with consistent findings across all raters (Supplemental Table 3).

The SD of annual changes was comparable across all raters, with signs of lower SD for AI for volume parameters (Fig. 3C). This suggests comparable longitudinal repeatability across raters, with a trend towards higher repeatability for AI in volume parameters. Bias and limits of agreement for the results were similar whether all annual intervals were included or restricted to one interval per patient and the comparison of annual change between treated and untreated intervals remained non-significant in both datasets.

Agreement of annual changes

Agreement of annual change scores against the reference human reader is shown in Supplemental Table S4. Bias of the change scores was small for all raters, and for LV volumes the systematic underestimation seen for absolute AI measurements was largely absent. Directional agreement was highest for E/e' and LV volumes and near chance for wall thickness, where annual changes are small relative to measurement variability.

Changes exceeding the minimal detectable change

By definition, changes exceeding the MDC occurred more often in raters with smaller intrarater variability. The MDC thresholds and the proportion of 12-month intervals with a change exceeding the MDC are listed and visualized in Fig. 5 and Supplemental Table S2. The novice showed the lowest proportion for all parameters, below those of AI, the reference human reader and cardiologist 2, and identified no change exceeding the MDC for IVSd or PWD. Because the thresholds are rater-specific, these proportions describe each rater's own precision and were not compared statistically.

Discussion

Automated echocardiographic measurements using artificial intelligence have the potential to enhance longitudinal assessment of patients with ATTR cardiomyopathy by reducing measurement variability and improving efficiency of serial examinations performed by different operators. In this retrospective analysis, two main findings emerged. First, fully automated measurements showed within-study variability of similar magnitude to the intrarater variability observed in experienced readers, although these quantities were obtained using different experimental designs. This supports further prospective evaluation of automated analysis for serial echocardiographic assessment, but does not establish its ability to detect true individual disease progression. Second, moderate agreement between AI and expert measurements and systematic differences in some volumetric measures indicate that human oversight may still be required when applying fully automated measurements in real-world settings with advanced cardiac disease such as ATTR cardiomyopathy. Nevertheless, the low within-study variability of AI-derived measurements suggests that automated analysis could offer a consistent basis for serial measurements, which warrants prospective evaluation.

Agreement of automated measurements

AI showed moderate-to-good agreement with the reference human reader's measurements for most parameters, consistent with previous studies on the agreement between AI and human evaluations [12, 15, 16]. In particular, AI achieved good agreement with the reference human reader's values for PWD, for which agreement of the novice with the reference human reader was also very good. In contrast, for IVSd, both AI and the novice exhibited discrepancies compared with the reference human reader's assessments: while AI consistently underestimated IVSd, the novice tended to overestimate it. This aligns with previous research [16] and is likely due to challenges in accurately delineating the septal borders from the adjacent right ventricular structures, particularly in ATTR-CM patients with increased septal thickness. Regarding volumetric parameters, both end-diastolic and end-systolic volumes were underestimated by AI, with the underestimation being more pronounced in LVESV. This finding aligns with previous research, which indicated that AI-based measurements exhibited lower reliability and higher bias in LVESV than LVEDV [12, 17]. This significant underestimation in volumes may be attributed to difficulties in tracing the endocardial border and differentiating between compact myocardium and trabeculations in thicker ventricles. Figure 6 illustrates a representative case, showing automated and manual measurements

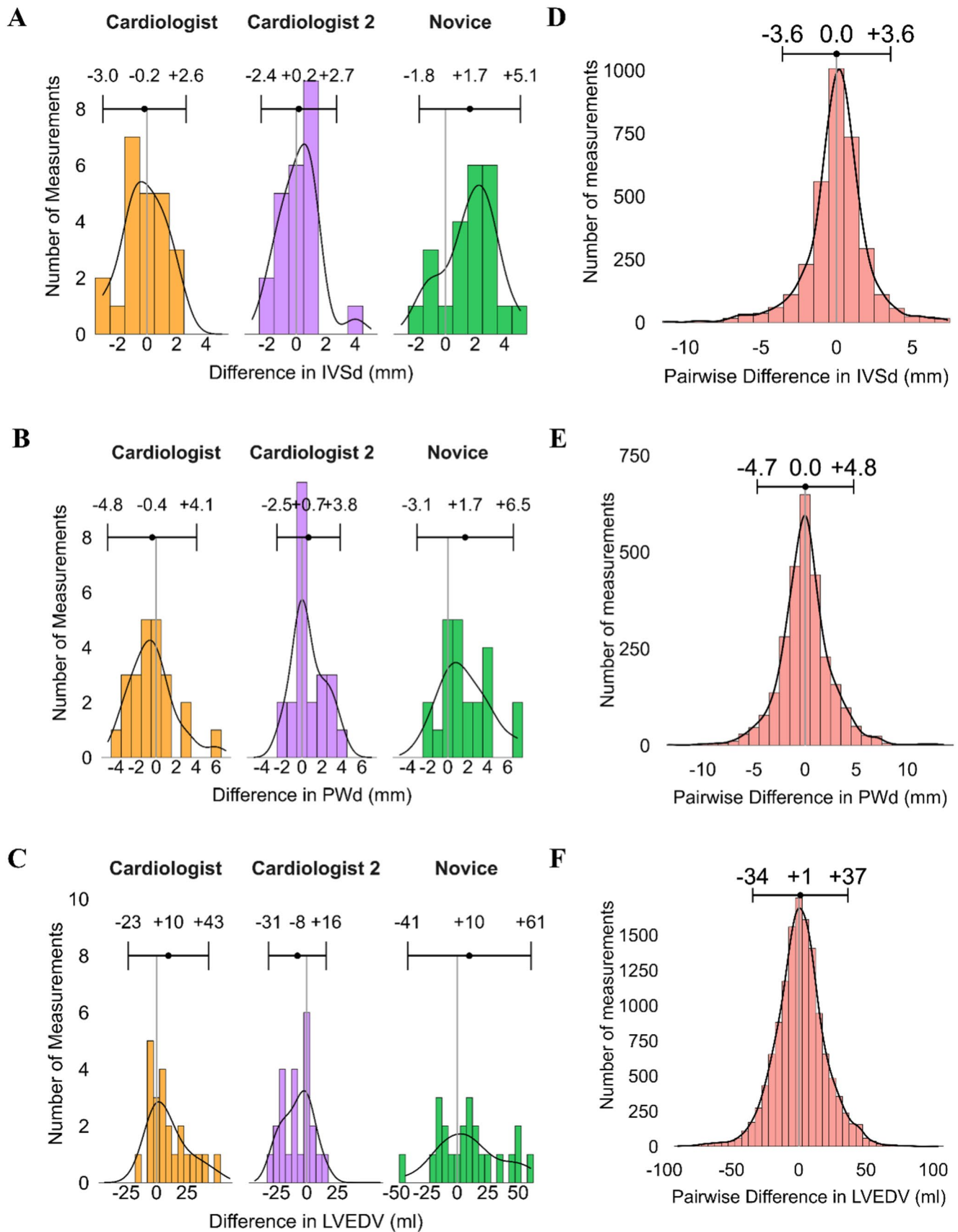


Fig. 3 Histogram with differences between two rounds of measurements for human raters (A–C) and with pairwise differences between measurements for AI (D–F). (A, D) IVSd, interventricular septal thickness at diastole; (B, E) PWd, posterior wall thickness at diastole; (C, F) LVEDV, left ventricular end-diastolic volume. Values shown represent mean $\pm 1.96 \times \text{SD}$ of the differences. (A–C) $n=20$ examinations from 20 patients, each measured twice per rater. (D–F) The y-axis denotes the number of pairwise differences, not of individual measurements. Pairwise differences are correlated observations derived from a much smaller number of automated measurements and are shown descriptively; underlying data: IVSd, 61 patients, 155 examinations, 947 automated measurements; PWd, 60 patients, 153 examinations, 877 automated measurements; LVEDV, 62 patients, 170 examinations, 2012 automated measurements

of LV end-diastolic and end-systolic volume alongside the corresponding echocardiographic images.

The agreement we observed was lower than that reported in previous studies involving non-amyloid patients or cohorts with higher LVEF, which generally reported higher AI–human agreement [10, 15, 17, 18]. Many of these previous studies compared AI-based measurements to human measurements for each individual image or video loop. This contrasts the current study where the AI algorithm performed on multiple images even of the same view. Among studies that did include broader populations, Tromp et al. demonstrated that AI–human differences were similar to or smaller than inter-human variability in a mixed cohort with preserved and reduced LVEF [10]. The lower agreement in our cohort is likely attributable to several factors specific to our study: first, the real-world setting with unselected patients and multiple sites entails heterogeneous image quality; second, the particular challenges posed by cardiac amyloidosis, where markedly increased wall thickness complicates automated border delineation; third, the wide range of ejection fractions in our cohort, including patients with reduced LVEF and advanced remodelling; fourth, the possibility that the AI-based and human measurements were performed on different images within the same echocardiographic exam.

Several recent studies have examined AI-based echocardiographic tools in cardiac amyloidosis, demonstrating their potential for diagnosis, screening, and disease monitoring. AI-derived measurements have been shown to accurately identify CA in globally diverse cohorts using both multiparametric scoring and deep-learning approaches [19], and in a real-world unselected population of over 11,000 patients, AI-flagged individuals were 50 times more likely to have confirmed CA [20]. Furthermore, longitudinal changes in AI-derived LVOT-VTI have been shown to independently predict all-cause mortality in ATTR-CM [21]. While these studies establish the diagnostic and prognostic potential of AI echocardiography, none specifically examined the agreement and reproducibility of AI-derived measurements against human raters in established ATTR-CM. The only

study to specifically examine AI performance in ATTR-CM evaluated fully automated AI-derived LVEF and GLS and found these to be comparable to manual measurements in patients at preclinical stages and at the time of diagnosis [22]. However, direct comparison with our findings is limited, as that study focused exclusively on LVEF and GLS in patients at earlier disease stages with higher LVEF than ours. The more advanced cardiac involvement in our cohort, with greater wall thickening and more pronounced systolic dysfunction, likely explains the lower AI agreement we observed, consistent with the known reduction in AI accuracy in populations with underlying structural heart disease [23].

The comparison with CMR should be interpreted with caution, given the small subset, the clinically indicated rather than protocol-driven acquisition and the variable interval to the echocardiogram. Within these constraints, all raters underestimated wall thickness and volumes relative to CMR. The CMR examinations were single clinical reads, and the reproducibility of

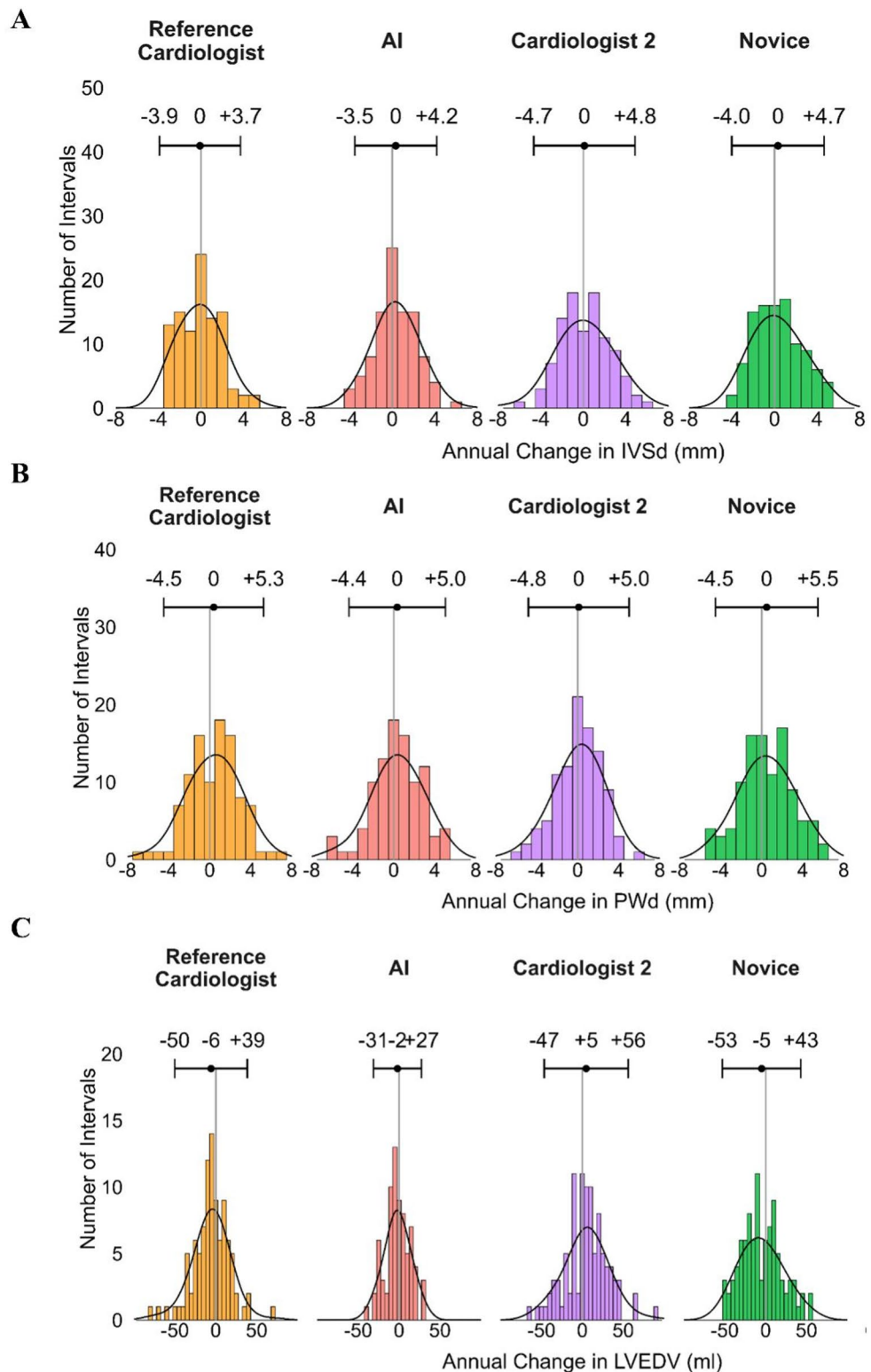
the CMR measurements themselves was not assessed. For LVEF, where echocardiography is known to consistently underestimate CMR-derived values [24], human raters followed this pattern with small negative biases, while AI showed minimal bias, suggesting that despite its underestimation of volumes, AI-derived LVEF agreed closely with CMR.

Repeatability of measurements

Beyond agreement at a single point in time, high repeatability is critical for long-term monitoring, especially in ATTR-CM, where changes are often subtle and low precision causes missing associations between prognosis and echocardiographic parameters [25]. Previous work has shown that AI is non-inferior to human intra-observer variability for left ventricular volumes and ejection fraction [18] and AI models assessing EF by leveraging information across multiple cardiac cycles, were found to be more reproducible than human evaluation [16, 26].

In our study, the two cardiologists consistently achieved moderate-to-excellent repeatability, while the novice exhibited lower repeatability. AI's within-study repeatability across various acquisitions of a single study, was comparable to that of the expert cardiologists. Notably, large deviations observed in histograms are attributable to image acquisition variability rather than the measurement algorithm itself. Although variability between acquisitions is common in routine clinical imaging, automated analysis may mitigate the impact of extreme outliers by prioritizing measurements derived from the highest-confidence image acquisitions. Using the mean $\pm 1.96 \times \text{SD}$ of repeated measurements as

Fig. 4 Annual changes. **(A)** IVSd, interventricular septal thickness at diastole; **(B)** PWd, posterior wall thickness at diastole; **(C)** LVEDV, left ventricular end-diastolic volume. Histograms show the distribution of annual changes between consecutive 12-month intervals for each rater. Values above each panel indicate the mean and 95% limits of agreement ($\pm 1.96 \times \text{SD}$). EDV, end-diastolic volume; ESV, end-systolic volume; IVSd, interventricular septum thickness in diastole; LV, left ventricular; PWd, posterior wall thickness in diastole



the minimal detectable change (MDC), the MDC for AI was numerically smaller than that of the novice and of similar magnitude to those of the cardiologists; these values were not formally compared. These MDC thresholds were applied for illustrative purposes to demonstrate the implications of

differing measurement precision and should not be interpreted as clinically validated criteria for disease progression. Nevertheless, using these illustrative thresholds, the reference human reader identified the highest proportion of

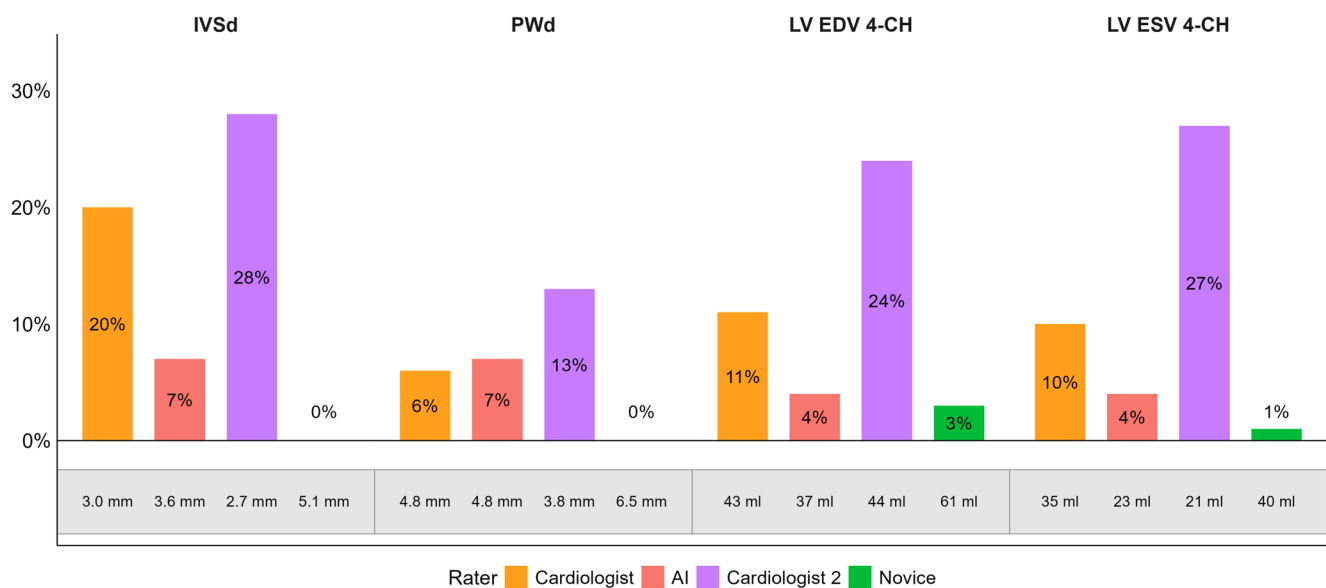


Fig. 5 Proportion of 12-month intervals with a change exceeding the minimal detectable change (MDC). EDV, end-diastolic volume; EF, ejection fraction; IVSd, interventricular septal thickness at diastole; LV, left ventricular; PWd, posterior wall thickness at diastole. Values in grey box indicate rater-specific minimal detectable change (MDC)

thresholds (i.e. $1.96 \times \text{SD}$). Proportions are based on $n=98-102$ twelve-month intervals from 58 patients and are shown descriptively; because thresholds are rater-specific, proportions were not compared statistically

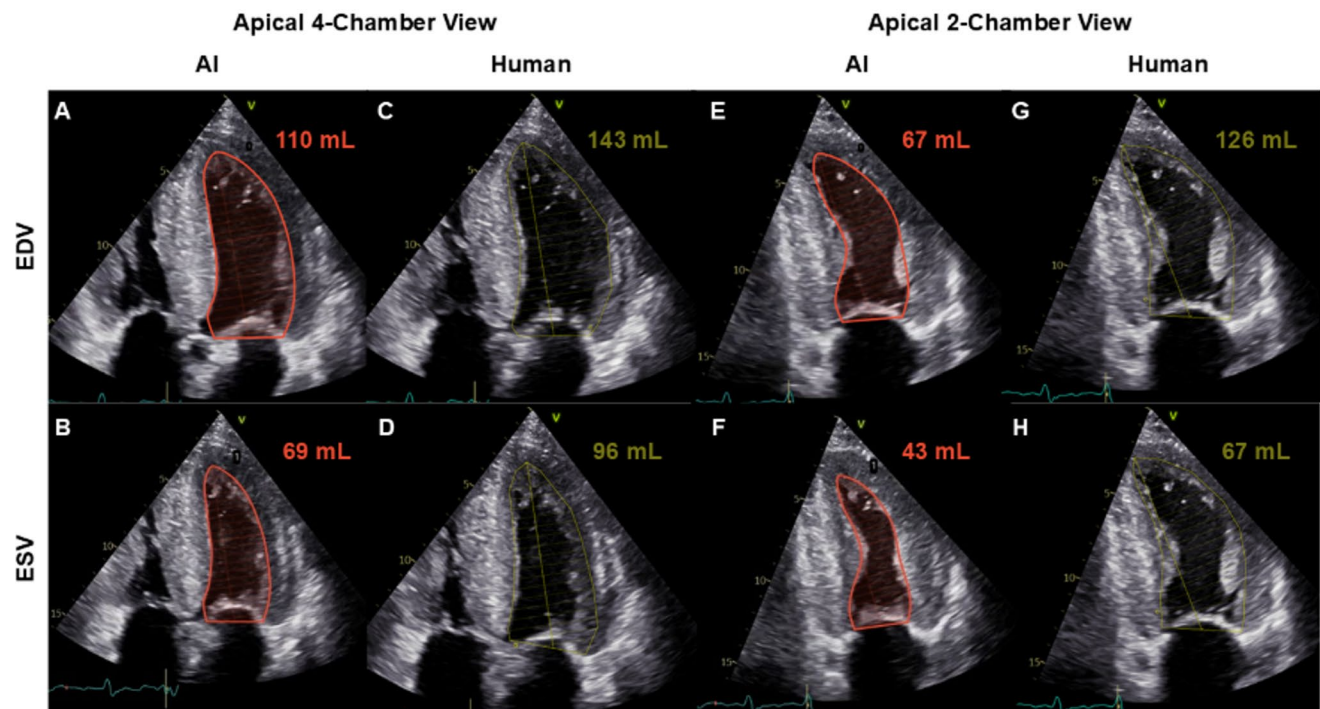


Fig. 6 Example of automated and manual measurements of LV end-diastolic and end-systolic volume. AI, artificial intelligence; EDV, end-diastolic volume; ESV, end-systolic volume

changes, and the proportion for AI was higher than that of the novice.

Apart from one isolated finding for one rater, no differences were detected between intervals with and without treatment. Because ATTR-CM is progressive and treatment

exposure in our cohort was heterogeneous, the dispersion of annual changes reflects a mixture of true biological change and measurement variability that cannot be separated with this design. The SD of annual changes was comparable across all raters for wall thickness parameters. For

LV volumes, AI showed the lowest SD of annual changes. Because it also measured a narrower range of volumes, this lower dispersion is consistent with, but does not prove, higher longitudinal precision, and range compression is a plausible alternative explanation. Taken together, it can be concluded that the majority of the measurement variability we observed is attributable more to the image acquisition than to the measurement step itself, although the AI's repeatability nearly matches that of experienced cardiologists. This has two important implications: For imaging-based surrogate endpoints in clinical trials, larger sample sizes are required to detect true treatment effects when variability attributable to image acquisition dominates the measurement error and in routine clinical practice, only the largest annual changes can be reliably distinguished from measurement noise at annual intervals.

Clinical Implications

Current guidelines suggest that ≥ 2 mm increase in LV wall thickness, $\geq 5\%$ decreases in LV ejection fraction or ≥ 5 mL decrease in stroke volume are indicative of disease progression [4]. However, in everyday clinical practice, it is challenging to reliably detect these subtle changes because inter- and intrarater variability limits the precision of echocardiographic assessments. AI offers a promising solution by standardizing measurements and improving consistency in longitudinal evaluations. Standardized assessments of this kind are essential for detecting subtle treatment-related changes. In clinical trials, such core labs standardizing echocardiographic data were critical to assess response to disease-modifying treatment [6]. If the within-study variability observed here is confirmed prospectively, automated analysis could reduce variability in routine clinical practice and enable more reliable monitoring of treatment responses. In the context of the numerous ongoing therapeutic studies on ATTR amyloidosis, AI-based analysis represents an attractive, cost-effective solution for echocardiographic examinations. In fact, in a randomized controlled trial, integration of AI-based analysis of echocardiography to the clinical workflow enabled sonographers to complete more echocardiograms with less mental fatigue, supporting its role in sustaining high-volume laboratories while maintaining diagnostic quality [27].

For such fully automated measurements in patients with ATTR-CM, differences smaller than 3.6 mm for septal wall thickness, 4.8 mm for posterior wall thickness, 37 mL for LVEDV and 23 mL for LVESV could not be distinguished from measurement noise in our cohort. For septal wall thickness, this exceeds the previously proposed 2 mm progression criterion not only for AI but also for both cardiologists, whose smallest detectable changes were 3.0 mm

and 2.7 mm. These are descriptive, study-specific estimates derived from a single cohort, a single algorithm and a limited number of repeat measurements and should not be applied as validated thresholds for disease progression in clinical practice.

Study limitations

This study has several limitations. First, our assessment of human raters was based on a small number of raters, limiting the generalizability of our findings. Second, within-study repeatability were available only for certain parameters from AI measurements, specifically wall thickness parameters and 4-chamber volume measurements, because the AI software provides multiple repeated measurements for these parameters, whereas LVEF is reported as a single calculated value without individual measurement repetitions. Further, the observed variability, both in human intrarater measurements and AI within-study measurements, may be due to differences in image acquisition and choice of measured image loop rather than the measurement process itself. However, this was deliberately chosen to reflect real clinical practice, where repeated measurements are usually performed without systematic documentation of the specific acquisitions used. Third, the longitudinal analysis of annual changes should be regarded as exploratory: because ATTR-CM is a progressive disease, true change cannot be fully separated from measurement variability over each 12-month interval. The absence of a difference between intervals with and without TTR-modifying treatment argues against a major treatment effect. Fourth, serial scans were performed at multiple institutions using equipment from different vendors, and the scanner used for a given patient also changed over time. This heterogeneity is inherent to the specific research question addressed by the study but cannot be disentangled from assessor-related variability within the framework of the present study design. Relatedly, our argument that a stable systematic bias does not compromise longitudinal monitoring assumes that this bias remains constant across time, disease severity, scanner type, image quality and loading conditions. Our retrospective dataset does not allow this assumption to be formally tested. Fifth, a single reader served as the reference, whereas a consensus reached by several experienced readers would have constituted the most appropriate reference standard. This study should therefore be considered a single-center pilot evaluation. Lastly, all automated measurements were used exactly as returned by the algorithm, without operator review or editing, whereas routine clinical practice would likely involve quality control and editing of individual measurements by an experienced reader. Such editing may itself introduce additional inter- and intra-individual variability that our

fully automated design cannot capture. Nevertheless, our study also has notable strengths. It expands the comparison of AI-based measurements with human raters to a cardiac amyloidosis cohort at advanced disease stages with a broad range of ejection fractions, thereby highlighting challenges of applying AI in patients with advanced cardiac disease. We also compared echocardiographic measurements against cardiac MR in a subset of patients.

Conclusion

Fully automated measurements showed within-study variability of similar magnitude to the intrarater variability observed in experienced readers, although these quantities were obtained using different experimental designs. These findings support further prospective evaluation of automated analysis for serial echocardiographic assessment in ATTR cardiomyopathy but do not establish its ability to detect true individual disease progression. Because agreement in absolute terms was moderate, AI-assisted monitoring requires expert oversight, particularly in advanced ATTR-CM, where reduced image quality, marked hypertrophy, trabeculations and altered ventricular geometry challenge automated segmentation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10554-026-03835-1>.

Author contributions A.W.: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. O.F.C.: Writing – review & editing. C.M.: Methodology, Investigation, Writing – review & editing. A.J.F.: Writing – review & editing. P.L.M.: Writing – review & editing. R.S.: Writing – review & editing. C.G.: Writing – review & editing. F.R.: Writing – review & editing. F.C.T.: Writing – review & editing. D.C.B.: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing.

Funding The study was supported by an investigator-initiated grant from AstraZeneca to fund measurements by Us2.ai. The company had no involvement in the design, conduct, analysis, or reporting of the study.

Data availability Due to regulatory restrictions by the hospital, data cannot be shared.

Declarations

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. Approval of this study was granted by the Cantonal Ethics Committee Zurich (KEK-ZH 2014–0490). Written informed consent was obtained from all individual participants included in the study.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Maurer MS, Schwartz JH, Gundapaneni B, Elliott PM, Merlini G, Waddington-Cruz M et al (2018) Tafamidis Treatment for Patients with Transthyretin Amyloid Cardiomyopathy. *N Engl J Med* 379(11):1007–1016. <https://doi.org/10.1056/NEJMoa1805689>
2. Gillmore JD, Judge DP, Cappelli F, Fontana M, Garcia-Pavia P, Gibbs S et al (2024) Efficacy and Safety of Acoramidis in Transthyretin Amyloid Cardiomyopathy. *N Engl J Med* 390(2):132–142. <https://doi.org/10.1056/NEJMoa2305434>
3. Fontana M, Berk JL, Gillmore JD, Witteles RM, Grogan M, Drachman B et al (2024) Vutrisiran in Patients with Transthyretin Amyloidosis with Cardiomyopathy. *N Engl J Med*. <https://doi.org/10.1056/NEJMoa2409134>
4. Garcia-Pavia P, Bengel F, Brito D, Damy T, Duca F, Dorbala S et al (2021) Expert consensus on the monitoring of transthyretin amyloid cardiomyopathy. *Eur J Heart Fail* 23(6):895–905. <https://doi.org/10.1002/ejhf.2198>
5. Benz DC, Grani C, Antiochos P, Heydari B, Gissler MC, Ge Y et al (2023) Cardiac magnetic resonance biomarkers as surrogate endpoints in cardiovascular trials for myocardial diseases. *Eur Heart J* 44(45):4738–4747. <https://doi.org/10.1093/eurheartj/eha510>
6. Shah SJ, Fine N, Garcia-Pavia P, Klein AL, Fernandes F, Weissman NJ et al (2024) Effect of Tafamidis on Cardiac Function in Patients With Transthyretin Amyloid Cardiomyopathy: A Post Hoc Analysis of the ATTR-ACT Randomized Clinical Trial. *JAMA Cardiol* 9(1):25–34. <https://doi.org/10.1001/jamacardio.2023.4147>
7. Jering KS, Fontana M, Lairez O, Longhi S, Azevedo O, Morbach C et al (2025) Effects of vutrisiran on cardiac structure and function in patients with transthyretin amyloidosis with cardiomyopathy: secondary outcomes of the HELIOS-B trial. *Nat Med*. <https://doi.org/10.1038/s41591-025-03851-z>
8. Duffy G, Cheng PP, Yuan N, He B, Kwan AC, Shun-Shin MJ et al (2022) High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy With Cardiovascular Deep Learning. *JAMA Cardiol* 7(4):386–395. <https://doi.org/10.1001/jamacardio.2021.6059>
9. Mitchell C, Rahko PS, Blauwet LA, Canaday B, Finstuen JA, Foster MC et al (2019) Guidelines for Performing a Comprehensive Transthoracic Echocardiographic Examination in Adults: Recommendations from the American Society of Echocardiography. *J Am Soc Echocardiogr* 32(1):1–64. <https://doi.org/10.1016/j.echo.2018.06.004>
10. Tromp J, Bauer D, Claggett BL, Frost M, Iversen MB, Prasad N et al (2022) A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *Nat Commun* 13(1):6776. <https://doi.org/10.1038/s41467-022-34245-1>
11. Myhre PL, Hung CL, Frost MJ, Jiang Z, Ouwerkerk W, Teramoto K et al (2024) External validation of a deep learning algorithm

- for automated echocardiographic strain measurements. *Eur Heart J Digit Health* 5(1):60–68. <https://doi.org/10.1093/ehjdh/zta072>
12. Myhre PL, Gaibazzi N, Tuttolomondo D, Sartorio D, Ugolotti PT, Covani M et al (2024) 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* 11:1400333. <https://doi.org/10.3389/fcvm.2024.1400333>
 13. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L et al (2015) Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr* 28(1):1–39e14. <https://doi.org/10.1016/j.echo.2014.10.003>
 14. Martinez-Naharro A, Patel R, Kotecha T, Karia N, Ioannou A, Petrie A et al (2022) Cardiovascular magnetic resonance in light-chain amyloidosis to guide treatment. *Eur Heart J* 43(45):4722–4735. <https://doi.org/10.1093/eurheartj/ehac363>
 15. Tromp J, Seekings PJ, Hung CL, Iversen MB, Frost MJ, Ouwerkerk W et al (2022) Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *Lancet Digit Health* 4(1):e46–e54. [https://doi.org/10.1016/S2589-7500\(21\)00235-1](https://doi.org/10.1016/S2589-7500(21)00235-1)
 16. Shiokawa N, Izumo M, Shimamura T, Kurosaka Y, Sato Y, Okamura T et al (2024) Accuracy and Efficacy of Artificial Intelligence-Derived Automatic Measurements of Transthoracic Echocardiography in Routine Clinical Practice. *J Clin Med* 13(7). <https://doi.org/10.3390/jcm13071861>
 17. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L et al (2018) Fully Automated Echocardiogram Interpretation in Clinical Practice. *Circulation* 138(16):1623–1635. <https://doi.org/10.1161/CIRCULATIONAHA.118.034338>
 18. Olaisen S, Smistad E, Espeland T, Hu J, Pasdeloup D, Ostvik A et al (2024) 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* 25(3):383–395. <https://doi.org/10.1093/ehjci/jead280>
 19. Ioannou A, Khouri MG, Kitai T, Vemulapalli S, Hung CL, Lim SC et al (2026) Diagnosis of Cardiac Amyloidosis on Echocardiography Using Artificial Intelligence. *Circ Cardiovasc Imaging* e018991. <https://doi.org/10.1161/CIRCIMAGING.125.018991>
 20. Sakamoto AKN, Nakamura Y, Fujita W, Sato E, Kaneko T, Miyazaki S, Minamino T (2025) Artificial Intelligence-based Software for Checking REal-world Echocardiography to ideNtify hidden Cardiac Amyloidosis: AI-SCREEN-CA. American Heart Association Scientific Sessions
 21. Venneri L, Aimo A, Porcari A, Sezer I, Ioannou A, Sheikh A et al (2025) Artificial intelligence-based echocardiographic assessment for monitoring disease progression in transthyretin cardiac amyloidosis. *Eur J Heart Fail* 27(12):3392–3400. <https://doi.org/10.1002/ehj.70073>
 22. Cotella JI, Slivnick JA, Sanderson E, Singulane C, O'Driscoll J, Asch FM et al (2023) Artificial intelligence based left ventricular ejection fraction and global longitudinal strain in cardiac amyloidosis. *Echocardiography* 40(3):188–195. <https://doi.org/10.1111/echo.15516>
 23. Wegner FK, Benesch Vidal ML, Niehues P, Willy K, Radke RM, Garthe PD et al (2022) Accuracy of Deep Learning Echocardiographic View Classification in Patients with Congenital or Structural Heart Disease: Importance of Specific Datasets. *J Clin Med* 11(3). <https://doi.org/10.3390/jcm11030690>
 24. Simpson R, Bromage D, Dancy L, McDiarmid A, Monaghan M, McDonagh T et al (2018) 6 Comparing echocardiography and cardiac magnetic resonance measures of ejection fraction: implications for HFMRf research. *Heart* 104(Suppl 5):A3–A. <https://doi.org/10.1136/heartjnl-2018-BCVI.6>
 25. Chacko L, Karia N, Venneri L, Bandera F, Passo BD, Buonamici L et al (2022) Progression of echocardiographic parameters and prognosis in transthyretin cardiac amyloidosis. *Eur J Heart Fail* 24(9):1700–1712. <https://doi.org/10.1002/ehj.2606>
 26. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP et al (2020) Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 580(7802):252–256. <https://doi.org/10.1038/s41586-020-2145-8>
 27. Sakamoto A, Kagiya N, Sato E, Nakamura Y, Murata A, Kaneko T et al (2026) Artificial Intelligence-Based Automated Echocardiographic Analysis and the Workflow of Sonographers: A Randomized Crossover Trial (AI-Echo RCT). *J Am Heart Assoc* 15(1):e045637. <https://doi.org/10.1161/JAHA.125.045637>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.