G Model
ACVD-1810; No. of Pages 12
ARTICLE IN PRESS
Archives of Cardiovascular Disease xxx (xxxx) xxx–xxx
Disponible en ligne sur
ScienceDirect
www.sciencedirect.com
Elsevier Masson France
EM|consulte
www.em-consulte.com

Clinical Research

Integrating artificial intelligence into an echocardiography department: Feasibility and comparative study of automated versus human measurements in a high-volume clinical setting

Stéphane Lafitte[a,b,*], Louis Lafitte[a], Melchior Jonveaux[b], Zoe Pascual[b], Julien Ternacle[b], Marina Dijos[b], Guillaume Bonnet[a,b], Patricia Reant[a,b], Anne Bernard[c,d]

[a] Bordeaux University, 33000 Bordeaux, France
[b] Unité médicochirurgicale des valvulopathies, CHU de Bordeaux, 33000 Bordeaux, France
[c] Inserm U1327 ISCHEMIA 'Membrane signaling and inflammation in reperfusion injuries', Université de Tours, 37000 Tours, France
[d] Service de Cardiologie, CHU de Tours, 37000 Tours, France

ARTICLE INFO

Article history:
Received 6 January 2025
Received in revised form 25 March 2025
Accepted 5 April 2025
Available online xxx

Keywords:
Echocardiography
Artificial intelligence
Deep learning
Agreement analysis
Cardiac parameters

ABSTRACT

Background: Echocardiography is an important diagnostic tool in cardiology as it is essential for heart disease treatment. However, its time-consuming nature and reliance on user expertise constitutes a challenge for its use in high-volume clinics. Artificial intelligence (AI) offers the potential to automate tasks performed manually by echocardiographers and promises to improve efficiency and diagnostic consistency.
Aims: To evaluate the integration of AI-based tools in a high-volume echocardiography department and assess the concordance of AI-generated measurements with manually-performed measurements.
Methods: The study was conducted in the echocardiography department of Bordeaux University Hospital. Over 2 months, 894 echocardiograms were performed by operators with three experience levels (nurses, residents and experts), with measurements performed by AI and humans. The statistical analyses assessed measurement agreement between both.
Results: The AI system was successfully integrated into the hospital's infrastructure within 6 weeks. Concordance analysis revealed good to very good agreement between AI and human measurements for most parameters, especially for ejection fraction (intraclass correlation coefficient [ICC]: 0.81, 95% confidence interval [95% CI]: 0.78–0.85) and Doppler-based flow measurements (mitral E wave velocity: ICC 0.97, 95% CI 0.95–0.98). Bland-Altman analysis showed a global mean difference of −4% with a standard deviation of 15%. Subgroup analysis revealed higher concordance for experts and residents compared with nurses (mean ICCs: 0.78 and 0.79 vs. 0.72, respectively).
Conclusion: AI can be effectively integrated into clinical echocardiography practice, with high agreement between AI and human measurements. Further research is needed to investigate the long-term impact on clinical outcomes and efficiency.

1. Abbreviations

AI	artificial intelligence
CI	confidence interval
EF	ejection fraction
GDPR	General Data Protection Regulation
GLS	global longitudinal strain
ICC	intraclass correlation coefficient
IP	internet protocol
IT	information technology
LV	left ventricle/ventricular
LVEF	left ventricular ejection fraction
LVOT	left ventricular outflow tract
NUC	next unit of computing
PACS	Picture Archiving and Communication System
SD	standard deviation
Vmax	maximal velocity
VTI	velocity time integral

* Corresponding author at: Hôpital Cardiologique, avenue Magellan, 33600 Pessac, France.
E-mail address: stephane.lafitte@chu-bordeaux.fr (S. Lafitte).

https://doi.org/10.1016/j.acvd.2025.04.051
1875-2136/

Please cite this article as: S. Lafitte, L. Lafitte, M. Jonveaux et al., Integrating artificial intelligence into an echocardiography department: Feasibility and comparative study of automated versus human measurements in a high-volume clinical setting, Arch Cardiovasc Dis, https://doi.org/10.1016/j.acvd.2025.04.051

2. Background

Echocardiography is an important diagnostic tool in cardiology as it is essential for heart disease treatment. The growing demand for this examination has put increasing pressure on echocardiography laboratories to improve their efficiency without compromising diagnostic accuracy. The procedure is time-consuming (mean 30 minutes) because it requires multiple views, numerous measurements and the production of detailed reports. This time may vary depending on the complexity of the case and the experience level of the operator [1]. In addition, echocardiography requires a high operator expertise level, acquired through extensive training.

Artificial intelligence (AI) integration offers a promising solution by automating several steps of echocardiographic analysis that traditionally rely on manual input [2]. Several studies have shown that AI can facilitate various workflow aspects, including image acquisition [3,4], image recognition [5], parameter measurements [6–8] and diagnostic support [9–11]. Recently, commercial AI solutions have entered the market, offering 'all-in-one' tools to streamline department workflows.

In this study, we hypothesized that AI-based tools can be seamlessly integrated into a standard 'picture archiving and communication system' (PACS) environment and enable automated measurements comparable with those performed manually by echocardiographers.

The objectives of this study were: (1) to assess the ability of public and commercial institutions to integrate AI solutions while complying with regulatory patient data requirements and (2) to evaluate the consistency of AI-generated measurements with those performed manually in routine clinical practice over a 2-month period.

3. Methods

3.1. Study setting and integration

The study was conducted in the echocardiography department of Bordeaux University Hospital, where over 21,000 echocardiograms are performed annually. The global research comprised three phases, the first two of which are the focus of this manuscript. The third phase, which depends on the successful validation of the first two phases, will assess AI solution impact on the echocardiography department workflows. This project began with a contractual agreement with Us2.ai for the loan of AI-equipped echocardiography equipment specifically for this research.

3.2. Phase 1: technical integration and setup

The research project complied with the regulations on medical data use. To simplify the procedures, we opted for internal data processing within the institution, without transferring it from the secure environment. We therefore asked the AI service provider Us2.ai to provide an AI processing station for physical integration into the Bordeaux University Hospital's Department of Information Technology. The 'next unit of computing' (NUC) server was installed with Us2.ai software. This server was integrated into the hospital's internal network as a web internet protocol (IP) server configured with its own IP address and port number to operate within the hospital's data security framework, eliminating the need for a virtual private network. It was connected to two Vivid E95 echocardiography machines (General Electric Healthcare, Chicago, IL, US) in separate rooms. A dual data transmission system was set up so that each echocardiographic examination could be sent simultaneously to both the traditional PACS, including the reporting software (ComPACS, MediMatic, Italy), and to the NUC. This enabled real-time monitoring and data comparisons. The time required to achieve this goal was the primary quantitative evaluation parameter of Phase 1.

3.3. Phase 2: data collection and comparative analysis

All human echocardiographic measurements in this study were performed manually, without the use of automated calculation tools. This decision was made specifically to ensure a direct and unbiased comparison between AI-generated and human-performed measurements, avoiding any confounding effect from pre-existing automation. To compare measurements performed by the echocardiographers under real conditions with those generated by AI, all examinations from two specific rooms were systematically sent to both the PACS and the NUC over a 2-month period. Routine transthoracic echocardiography examinations were performed in these rooms, which were scheduled 2–6 months prior to the actual procedure. Room operation followed a schedule set 1 month in advance according to the department's protocol, which remained unchanged due to ongoing research.

The echocardiograms included in this study were systematically selected from two dedicated rooms where the AI system was implemented. These rooms were selected to ensure standardized AI integration and data collection and to represent a controlled subset of the total echocardiograms performed in our centre. No selection criteria were applied based on patient characteristics, case complexity or image quality. Thus, the discrepancy between the total number of echocardiograms performed in the department and the number of echocardiograms included in this study is due to this targeted room selection and not to any exclusion based on clinical factors. This approach ensured robust comparison between AI and human measurements under standardized conditions.

The echocardiographers were 'nurse' sonographers according to a competence protocol approved by the 'Agence Régionale pour la Santé' (ARS Nouvelle Aquitaine), 'residents' with < 12 months of experience and 'expert' echocardiographers with > 5 years of experience. Examinations carried out by nurses were systematically validated by an expert in echocardiography. Examinations were performed according to the traditional service protocol without modification. During each examination, 20–30 electrocardiogram synchronized loops and 10–20 images, mainly Doppler recordings, were systematically recorded. Measurements were made directly on the echocardiography machine, and standardized labels were applied to each 2D and Doppler echo image. At the end of each examination, the images and the Digital Imaging and Communications in Medicine Structured Report were automatically sent to both ComPACS and the NUC. The echocardiographer then generated the report in the ComPACS software and integrated the measurements into the report. The images received on the NUC were processed automatically and without human supervision, taking as many measurements as possible, mainly depending on image quality. The AI automatically identified the views and performed all possible measurements. The development method of this AI has been detailed in a previous study [12,13]. The Us2.ai solution used in this study has already received Conformité Européene and Food and Drug Administration marking, further emphasizing its reliability and safety for clinical use.

3.4. Statistical analysis

Statistical analyses were used to evaluate the agreement between automatic AI measurements and those performed manually by echocardiographers. Primary statistical methods included Pearson correlation coefficients to assess linear correlation between measurements, intraclass correlation coefficients (ICCs) to assess measurement reliability and agreement within and between observers, and Bland-Altman analyses to assess agreement between methods and to identify any bias. Limits of agreement were calculated to understand the variability and mean discrepancies between AI and manual measurements.

It must be emphasized that the above-mentioned Pearson correlation coefficients do not replace absolute agreement analysis (ICC and Bland-Altman) but rather complement the overall understanding of the results. It should also be noted that the AI measurements were obtained a posteriori, after the echocardiographic examinations and human measurements had been completed. In parallel, echocardiography data were automatically transmitted to two systems: the standard PACS used by echocardiographers and a separate server containing the AI tool. Data from AI and humans were compared only after automated extraction of anonymized data, with no prior intervention or visibility from human operators. Echocardiographers were unaware of the AI measurements during the manual performance or examination analysis.

Of note is that the acoustic quality of the windows was indirectly estimated from demographic characteristics (age, weight) in the absence of any specific subjective qualitative assessment.

Data from the PACS (ComPACS) and NUC were anonymized and extracted to a secure database for analysis.

4. Results

4.1. Phase 1: technical integration and implementation

The AI system was successfully integrated into the information technology (IT) infrastructure of Bordeaux University Hospital within the planned timeframe. The NUC server equipped with the Us2.ai software was put into operation 2 weeks after its arrival in the IT department. After that, network integration, including the accessible IP configuration, was completed within 10 days.

Within 4 weeks of project onset, the biomedical technicians had set up the echocardiography equipment for dual transmission to both ComPACS and the NUC. Functional testing over the following 5 days ensured system performance. This rapid integration met our primary goal of minimizing implementation time while complying with data security regulations. The setup enabled seamless data transfer between the echocardiography equipment and both the traditional PACS (ComPACS) and the NUC, ensuring that all examinations in the two designated rooms could be recorded without interruption or data loss. Real-time monitoring on the NUC confirmed that the AI processed the data as intended without external data transfer and strictly adhered to general data protection regulation (GDPR) requirements.

4.2. Phase 2: analysis of concordances, population of examinations, operators and indications

During this 2-month study, 894 echocardiographic examinations were performed in patients with a mean ± standard deviation (SD) age of 64.8 ± 16.3 years, ranging from young adults to the elderly, 57% were male, mean ± SD weight was 76.0 ± 19.2 kg and mean ± SD body surface area was 1.85 ± 0.09 m^2. Echocardiograms were performed by operators with three different expertise levels: nurses performed 28 examinations, residents 258 examinations and experts 569 examinations. The examinations were related to various clinical indications. Significant mitral regurgitation was present in 25% of cases and heart failure with preserved ejection fraction (EF) in 22%. Other common findings were left ventricle (LV) concentric remodelling (18%) and severe tricuspid regurgitation (15%). Overall, 14% of echocardiograms were classified as normal, while 86% showed pathological findings. Notable pathologies other than significant mitral regurgitation and heart failure with preserved EF included hypertrophic cardiomyopathy (12%), heart failure with reduced EF (9%) and pulmonary hypertension (8%).

4.2.1. Measurement description

A total of 31 pairs (paired measurements) were identified, where both AI and humans performed the same measurement. These paired measurements included key parameters such as left ventricular ejection fraction (LVEF), ventricular diameters and Doppler-based flow measurements. The most frequently encountered pairs were Doppler analyses of the LV outflow tract (LVOT), followed by measurements of the left atrial surface area, LV global longitudinal strain (GLS), and LV surface area. Each of these parameters had 700–800 corresponding values in both AI and operator datasets. In contrast, mitral inflow Doppler measurements were identified as pairs in only 650 cases, demonstrating a lower rate of concurrent measurement between AI and operators.

In addition to the paired measurements, there were 65 measurements identified exclusively by AI, without corresponding human measurements. Among these, several measurements were related to pairs where the echocardiographers did not complete the full analysis. Examples include indexing certain measurements to body surface area, when appropriate, or calculating derived values such as maximal or mean gradients (e.g. LVOT P and maximal velocity [Vmax]), as well as LV lengths in systole and diastole on apical four-chamber views and indexed LV volumes on the same view. Additionally, the AI identified structural measurements that were not exploited by echocardiographers, such as the right atrial surface area and various flow calculations. Interestingly, the AI also generated left atrial strain values and performed proximal isovelocity surface area calculations on cases of regurgitation, measurements that were often overlooked or underutilized by the echocardiographers.

4.2.2. Measurement comparisons and concordance

A total of 31 pairs were analysed in detail, allowing a targeted comparison between measurements generated by AI and those performed manually by echocardiographers, with a mean ± SD of 518 ± 160 measurements per pair. The mean Pearson correlation coefficient was 0.82 ($P < 0.001$), demonstrating a strong positive correlation between AI and human measurements. The Bland-Altman analysis showed a mean difference of −4% with a mean SD of 15%, indicating a high agreement between the two methods, while the mean ICC was 0.78, further highlighting the strong global agreement between pairs (Table 1).

Among the measurements with the highest concordance, the Doppler measurements of mitral and aortic flows (Vmax, mean velocity and velocity time integral [VTI]) stood out, along with tissue Doppler velocities of the right ventricle, which demonstrated the strongest agreement with an ICC of 0.90 (95% confidence interval [95% CI]: 0.87–0.94). Other notable concordances were observed for GLS (ICC: 0.82, 95% CI: 0.78–0.85) and the biplane LVEF (ICC: 0.81, 95% CI: 0.78–0.85) (Table 1 and Fig. 1), mostly driven by the end-systolic parameters, highlighting the fact that tracing systolic outline is easer that diastolic outlines.

Left atrial volume measurement was also consistent, with a correlation of 0.89 and an ICC of 0.71, 95% CI 0.67–0.76. However, less consistent results were found regarding the deceleration time of the mitral E wave, tricuspid regurgitation wave and wall measurements, with ICC values ranging from 0.35 to 0.60, indicating weaker agreement between AI and operators. Moreover, the Spearman correlation coefficient graphic of the tricuspid regurgitation wave showed that even if the correlation was not high ($r = 0.62$), the cloud plot seemed to correlate well up to 2.5–2.7 m/s, which corresponds to the limit of pulmonary hypertension definition (2.7 m/s).

Table 1
Results of AI and human[a] (H) comparisons integrating Student's *t*-test (*P*-value), correlation (Pearson), Bland & Altman data (mean differences, SD differences) and ICCs from the highest (top) to lowest (bottom) value parameters.

	n	AI mean	AI SD	H mean	H SD	Pearson	*P*	AI/H mean	AI/H mean diff	AI/H mean diff (%)	AI/H SD diff	AI/H SD diff (%)	ICC	95% CI
MV-E (cm/s)	312	79.7	25.5	79.9	26.0	0.97	< 0.001	79.8	−0.17	0	6.45	8	0.97	0.95–0.98
AoV VTI (cm)	602	32.7	16.4	34.7	17.0	0.97	< 0.001	33.7	−1.99	−6	4.23	13	0.96	0.94–0.97
MV-A (cm/s)	281	74.5	25.4	73.8	26.1	0.96	< 0.001	74.2	0.66	1	7.24	10	0.96	0.93–0.98
AoV Vmax (m/s)	604	1.6	0.7	1.7	0.7	0.96	< 0.001	1.7	−0.04	−2	0.20	12	0.96	0.93–0.98
AoV Vmean (m/s)	604	1.1	0.5	1.2	0.5	0.96	< 0.001	1.2	−0.06	−5	0.14	12	0.96	0.93–0.98
LVESV MOD A4C (mL)	584	50.3	34.5	54.3	40.9	0.94	< 0.001	52.3	−4.00	−8	14.12	27	0.93	0.90–0.95
LVESV MOD biplane (mL)	467	48.6	32.7	55.5	41.4	0.95	< 0.001	52.0	−6.89	−13	14.20	27	0.91	0.89–0.93
RV s' (cm/s)	667	11.7	3.1	12.3	3.1	0.93	< 0.001	12.0	−0.63	−5	1.21	10	0.91	0.89–0.93
LVESV MOD A2C (mL)	531	45.5	32.2	51.5	39.9	0.92	< 0.001	48.5	−6.01	−12	16.09	33	0.89	0.87–0.92
LVEDV MOD biplane (mL)	467	100.4	40.5	116.2	50.6	0.93	< 0.001	108.3	−15.75	−15	19.49	18	0.86	0.83–0.88
LVEDV MOD A4C (mL)	585	103.8	42.4	116.4	52.2	0.90	< 0.001	110.1	−12.57	−11	22.81	21	0.85	0.83–0.88
LVOT Vmean (m/s)	781	0.7	0.1	0.7	0.1	0.85	< 0.001	0.7	0.01	1	0.08	11	0.85	0.80–0.89
LVOT Vmax (m/s)	781	1.0	0.2	1.0	0.2	0.88	< 0.001	1.0	0.04	4	0.11	10	0.85	0.80–0.89
LVEDV MOD A2C (mL)	531	95.9	42.1	108.5	49.7	0.88	< 0.001	102.2	−12.67	−12	23.27	23	0.84	0.82–0.87
LV GLS (%)	238	−17.2	4.4	−15.8	4.2	0.87	< 0.001	−16.5	−1.37	8	2.22	−13	0.82	0.78–0.85
LVIDd (mm)	657	48.1	9.1	50.1	8.9	0.83	< 0.001	49.1	−1.96	−4	5.18	11	0.81	0.78–0.85
LVEF MOD biplane (%)	467	54.5	11.9	56.0	12.8	0.82	< 0.001	55.3	−1.49	−3	7.51	14	0.81	0.78–0.85
LVOT VTI (cm)	781	21.7	5.5	20.7	4.8	0.82	< 0.001	21.2	1.04	5	3.20	15	0.79	0.70–0.88
A2C LV GLS (%)	257	−16.8	4.7	−16.2	4.5	0.76	< 0.001	−16.5	−0.67	4	3.19	−19	0.75	0.67–0.82
LVEF MOD A4C (%)	584	54.3	13.1	56.6	13.4	0.75	< 0.001	55.4	−2.31	−4	9.29	17	0.74	0.70–0.78
LV mass (g)	330	158.9	57.2	163.6	57.3	0.73	< 0.001	161.3	−4.77	−3	41.98	26	0.73	0.64–0.80
A4C LV GLS (%)	257	−17.5	4.7	−15.8	4.5	0.79	< 0.001	−16.7	−1.70	10	3.01	−18	0.73	0.66–0.79
LAESV MOD A2C (mL)	480	60.3	30.9	76.1	39.3	0.83	< 0.001	68.2	−15.86	−23	22.13	32	0.72	0.68–0.76
LAESV MOD biplane (mL)	435	60.6	27.4	79.2	38.6	0.89	< 0.001	69.9	−18.64	−27	19.11	27	0.71	0.67–0.76
LAESV MOD A4C (mL)	640	61.9	29.6	78.9	42.3	0.84	< 0.001	70.4	−16.91	−24	23.69	34	0.70	0.68–0.76
A3C LV GLS (%)	255	−17.6	5.3	−15.7	4.3	0.73	< 0.001	−16.7	−1.86	11	3.65	−22	0.65	0.58–0.73
IVSd (mm)	661	9.8	2.4	10.0	3.2	0.64	< 0.001	9.9	−0.23	−2	2.49	25	0.61	0.47–0.74
TR Vmax (m/s)	496	2.6	0.7	2.7	0.6	0.62	< 0.001	2.7	−0.05	−2	0.56	21	0.61	0.47–0.73
LVEF MOD A2C (%)	531	55.3	12.4	56.1	15.8	0.61	< 0.001	55.7	−0.81	−1	12.75	23	0.60	0.55–0.64
LVPWd (mm)	636	9.2	1.9	8.6	2.2	0.44	< 0.001	8.9	0.61	7	2.17	24	0.41	0.33–0.51
DecT (ms)	562	208.2	38.3	212.2	67.3	0.40	< 0.001	210.2	−4.03	−2	62.56	30	0.35	0.26–0.43
Mean	518	–	–	–	–	0.82	< 0.001	46.8	−4.2	−4	11.4	15	0.78	–

A2C LV GLS: apical-2-chamber view left ventricular global longitudinal strain; A3C LV GLS: apical-3-chamber view left ventricular global longitudinal strain; A4C LV GLS: apical-4-chamber view left ventricular global longitudinal strain; AI: artificial intelligence; AoV Vmax: aortic valve transvalvular maximal velocity; AoV Vmean: aortic valve transvalvular mean velocity; AoV VTI: aortic valve transvalvular velocity time integral; CI: confidence interval; Dec T: deceleration time of the E wave; diff: difference; H: human; ICC: intraclass correlation coefficient; IVSd: interventricular septum thickness in diastole; LAESV MOD A2C: left atrium end-systolic volume modified apical-2-chamber view; LAESV MOD A4C: left atrium end-systolic volume modified apical-4-chamber view; LAESV MOD biplane: left atrium end-systolic volume modified biplane; LV GLS: left ventricular global longitudinal strain; LV mass: left ventricular mass; LVEDV MOD A2C: left ventricular end-diastolic volume modified Simpson apical-2-chamber view; LVEDV MOD A4C: left ventricular end-diastolic volume modified Simpson apical-4-chamber view; LVEDV MOD biplane: left ventricular end-diastolic volume modified Simpson biplane; LVEF MOD A2C: left ventricular ejection fraction modified Simpson apical-2-chamber view; LVEF MOD A4C: left ventricular ejection fraction modified Simpson apical-4-chamber view; LVEF MOD biplane: left ventricular ejection fraction modified Simpson biplane; LVESV MOD A2C: left ventricular end-systolic volume modified Simpson apical-2-chamber view; LVESV MOD A4C: left ventricular end-systolic volume modified Simpson apical-4-chamber view; LVESV MOD biplane: left ventricular end-systolic volume modified Simpson biplane; LVIDd: left ventricular internal diastolic diameter; LVOT Vmax: left ventricular outflow tract maximal velocity; LVOT Vmean: left ventricular outflow tract mean velocity; LVOT VTI: left ventricular outflow tract velocity time integral; LVPWd: left ventricular posterior wall diameter; MV-A: mitral valve A wave velocity; MV-E: mitral valve E wave velocity; RV s': right ventricular systolic velocity; SD: standard deviation; TR Vmax: tricuspid regurgitation maximal velocity.

[a] "Human" refers to measurements made by human operators in real clinical situations, thus integrating all operator categories (experts, residents and nurses). Our goal was to reflect current clinical practice, without favouring any specific category. Moreover, specific analyses by operator category (Table 2 in the manuscript) were carried out to provide additional nuance.

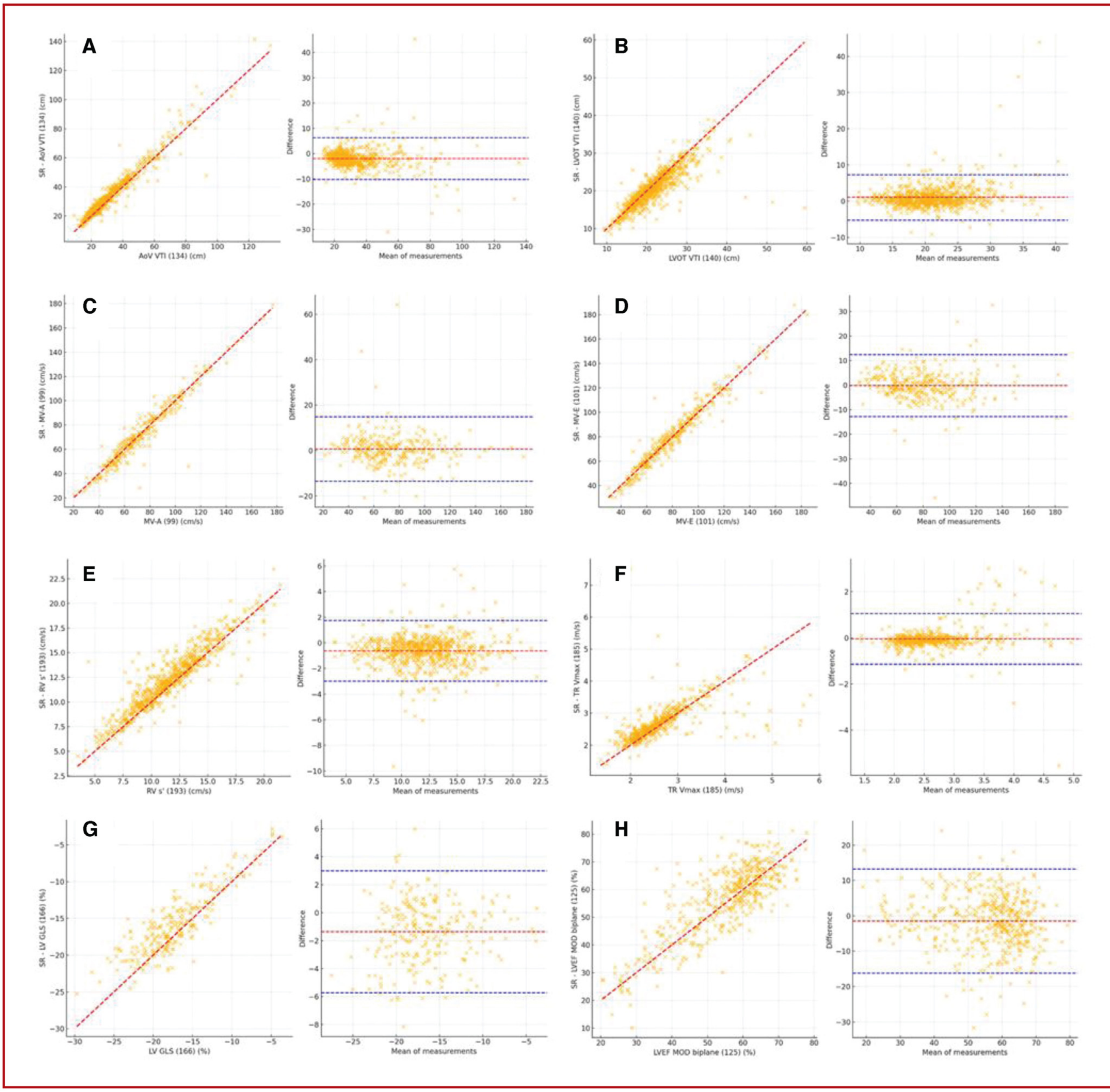


Fig. 1. Correlation and Bland-Altman plots for a representative measurement panel comparisons between AI and echocardiographers, featuring A. The aortic VTI. B. LVOT. C. Mitral A wave. D. Mitral E wave. E. RV S' wave. F. Tricuspid regurgitation Vmax. G. Global longitudinal strain. H. Biplane Simpson's method ejection fraction. AI: artificial intelligence; LVOT: left ventricular outflow tract; RV: right ventricle; VTI: velocity time integral.

These results illustrate AI's ability to be safe and similar to human quantification of important thresholds.

4.2.3. Subgroup analysis

A comparison between AI and human measurements was performed for the three skill levels. The analysis found that agreement varied with operator experience (Table 2). Nurses showed moderate agreement for most parameters (ICC: 0.72) compared to an ICC of 0.79 for experts. For example, LVEF had an ICC of 0.58 (95% CI: 0.27–0.78) for nurses, indicating reasonable agreement but considerable variability, versus an ICC of 0.84 (95% CI: 0.79–0.86) for experts. Similar discrepancies were observed for LV end-diastolic volume, interventricular septum thickness in diastole, LV mass and GLS. However, for simple measurements like Doppler flow measurements (mitral valve E wave velocity, mitral valve A wave velocity, right ventricle s', LVOT VTI, aortic valve transvalvular Vmax, aortic valve VTI), ICC values were comparable. Low differences were observed for all measurements between experts and residents (mean ICC: 0.78 versus 0.79), even though more marked differences were observed for sensitive parameters such as LVEF (residents: ICC: 0.79, 95% CI: 0.73–0.84 vs. experts: ICC: 0.84, 95% CI: 0.79–0.86).

Patient age-based analysis highlighted that the agreement between AI and human measurements was generally good in all age groups, with slight deviations in older patients. LVEF values were consistent: 55.8 ± 12.1% for AI and 55.7 ± 10.7% for humans (r = 0.98; P < 0.001). In patients older than 75 years, some Doppler measurements such as LVOT Vmax displayed a decrease in con-

Table 2
Results of skill subgroup comparison analysis between AI and humans (H) integrating Student's *t*-test (*P*-value), correlation (Pearson), Bland & Altman data (mean differences, SD differences) and ICCs. Three different levels of echocardiographers were compared: nurses, residents and experts.

	n	AI mean	AI SD	H mean	H SD	*P*	Pearson R	Pearson *P*	AI/H mean	AI/H mean diff	AI/H mean diff (%)	AI/H SD diff	AI/H SD diff (%)	ICC	95% CI
LVEF MOD biplane (%)															
Nurses	21	58.1	8.4	63.1	7.4	0.05	0.61	< 0.001	60.6	−5.00	−8	6.90	11	0.58	0.27–0.78
Residents	168	53.8	12.7	57.5	13.9	0.01	0.83	< 0.001	55.6	−3.65	−7	7.90	14	0.79	0.73–0.84
Experts	265	54.5	11.8	54.3	12.3	0.85	0.84	< 0.001	54.4	0.19	0	6.87	13	0.84	0.79–0.86
LVEDV MOD biplane (mL)															
Nurses	21	88.8	20.5	109.3	23.1	0.00	0.66	< 0.001	99.1	−20.50	−21	17.78	18	0.64	0.36–0.81
Residents	168	99.4	43.3	113.2	56.3	0.01	0.94	< 0.001	106.3	−13.82	−13	21.40	20	0.87	0.83–0.89
Experts	265	102.6	40.3	119.8	48.9	0.00	0.93	< 0.001	111.2	−17.14	−15	18.26	16	0.84	0.81–0.88
LVESV MOD biplane (mL)															
Nurses	21	37.5	11.6	40.4	11.5	0.41	0.78	< 0.001	38.9	−2.99	−8	7.40	19	0.88	0.76–0.94
Residents	168	49.4	36.3	53.5	46.6	0.37	0.96	< 0.001	51.4	−4.13	−8	15.89	31	0.90	0.89–0.94
Experts	265	49.4	31.9	58.7	39.9	0.00	0.96	< 0.001	54.1	−9.28	−17	13.25	25	0.90	0.88–0.92
IVSd (mm)															
Nurses	23	9.5	1.9	8.4	1.9	0.06	0.33	< 0.001	8.9	1.05	12	2.12	24	0.37	0.03–0.64
Residents	228	9.8	2.4	10.1	2.5	0.15	0.76	< 0.001	9.9	−0.32	−3	1.66	17	0.75	0.69–0.80
Experts	397	9.8	2.5	10.0	3.6	0.28	0.61	< 0.001	9.9	−0.24	−2	2.88	29	0.77	0.72–0.82
LVIDd (mm)															
Nurses	26	46.0	8.8	45.8	7.0	0.94	0.88	< 0.001	45.9	0.16	0	4.22	9	0.84	0.70–0.91
Residents	221	48.0	9.4	49.0	9.0	0.25	0.87	< 0.001	48.5	−1.02	−2	4.63	10	0.80	0.73–0.87
Experts	399	48.3	8.9	51.0	8.9	0.00	0.82	< 0.001	49.7	−2.70	−5	5.34	11	0.87	0.83–0.90
LVPWd (mm)															
Nurses	24	9.5	1.6	8.0	2.0	0.01	0.60	< 0.001	8.8	1.50	17	1.59	18	0.38	0.04–0.64
Residents	217	9.4	2.0	8.7	2.1	0.00	0.52	< 0.001	9.1	0.67	7	1.99	22	0.40	0.29–0.51
Experts	382	9.1	1.9	8.6	2.3	0.00	0.40	< 0.001	8.9	0.55	6	2.30	26	0.48	0.37–0.58
LV mass (g)															
Nurses	19	153.9	83.9	116.0	34.2	0.08	0.60	< 0.001	135.0	37.92	28	67.30	50	0.29	−0.19 to 0.66
Residents	66	148.6	48.1	153.4	52.4	0.59	0.74	< 0.001	151.0	−4.75	−3	36.51	24	0.73	0.59–0.82
Experts	244	162.1	57.1	170.2	58.2	0.12	0.77	< 0.001	166.2	−8.10	−5	38.96	23	0.77	0.70–0.82
MV-E (cm/s)															
Nurses	4	84.0	3.9	78.3	7.8	0.24	0.80	< 0.001	81.1	5.68	7	4.53	6	0.92	0.76–0.98
Residents	152	79.3	24.5	80.5	25.4	0.66	0.98	< 0.001	79.9	−1.25	−2	5.63	7	0.97	0.96–0.98
Experts	142	80.8	27.5	80.0	27.5	0.81	0.97	< 0.001	80.4	0.78	1	7.21	9	0.97	0.95–0.98
MV-A (cm/s)															
Nurses	3	97.2	45.3	97.6	47.8	0.99	1.00	< 0.001	97.4	−0.40	0	2.02	2	0.98	0.93–0.99
Residents	136	74.2	27.2	74.0	27.6	0.95	0.96	< 0.001	74.1	0.19	0	7.84	11	0.96	0.94–0.97
Experts	129	75.3	23.4	74.2	24.7	0.69	0.96	< 0.001	74.7	1.19	2	6.51	9	0.96	0.94–0.97
DecT (ms)															
Nurses	22	196.8	38.1	192.8	50.2	0.77	0.46	< 0.001	194.8	4.02	2	45.89	24	0.43	0.05–0.69
Residents	176	208.6	38.8	203.5	61.9	0.36	0.50	< 0.001	206.0	5.10	2	54.07	26	0.36	0.30–0.41
Experts	354	208.9	38.3	217.9	69.6	0.03	0.36	< 0.001	213.4	−9.07	−4	66.05	31	0.44	0.32–0.56
LAESV MOD biplane (mL)															
Nurses	23	65.8	22.5	72.8	26.9	0.34	0.90	< 0.001	69.3	−7.02	−10	11.52	17	0.76	0.55–0.88
Residents	165	62.0	26.2	79.3	33.1	0.00	0.89	< 0.001	70.6	−17.29	−24	15.33	22	0.72	0.64–0.79
Experts	236	59.1	28.7	79.6	43.2	0.00	0.89	< 0.001	69.3	−20.46	−30	21.70	31	0.69	0.62–0.75
RV s' (cm/s)															
Nurses	26	11.2	1.8	12.1	2.0	0.10	0.87	< 0.001	11.6	−0.88	−8	0.96	8	0.88	0.78–0.94
Residents	209	11.5	3.0	12.3	3.1	0.01	0.95	< 0.001	11.9	−0.78	−7	0.94	8	0.92	0.90–0.94
Experts	416	11.8	3.2	12.3	3.1	0.02	0.91	< 0.001	12.1	−0.53	−4	1.32	11	0.90	0.88–0.92
LVOT Vmax (m/s)															
Nurses	27	1.0	0.2	1.0	0.2	0.94	0.97	< 0.001	1.0	0.00	0	0.05	5	0.96	0.92–0.98
Residents	246	1.1	0.2	1.0	0.2	0.00	0.81	< 0.001	1.0	0.06	6	0.14	13	0.76	0.70–0.81
Experts	491	1.0	0.2	1.0	0.2	0.01	0.91	< 0.001	1.0	0.03	3	0.09	9	0.89	0.88–0.91

Table 2
(*Continued*)

	n	AI mean	AI SD	H mean	H SD	*P*	Pearson R	Pearson *P*	AI/H mean	AI/H mean diff	AI/H mean diff (%)	AI/H SD diff	AI/H SD diff (%)	ICC	95% CI
LVOT VTI (cm)															
Nurses	27	20.2	3.9	19.6	3.6	0.57	0.94	< 0.001	19.9	0.58	3	1.26	6	0.91	0.84–0.95
Residents	246	21.6	6.4	20.2	5.1	0.00	0.71	< 0.001	20.9	1.49	7	4.53	22	0.67	0.59–0.63
Experts	491	21.8	5.1	21.0	4.7	0.01	0.89	< 0.001	21.4	0.81	4	2.31	11	0.87	0.85–0.89
AoV Vmax (m/s)															
Nurses	26	1.3	0.6	1.3	0.6	0.91	0.99	< 0.001	1.3	0.02	1	0.09	7	0.99	0.98–0.99
Residents	191	1.6	0.6	1.6	0.7	0.39	0.98	< 0.001	1.6	−0.06	−4	0.14	9	0.97	0.96–0.98
Experts	370	1.7	0.8	1.7	0.8	0.56	0.95	< 0.001	1.7	−0.03	−2	0.23	14	0.95	0.94–0.96
AoV VTI (cm)															
Nurses	26	25.2	13.4	26.0	13.6	0.83	0.99	< 0.001	25.6	−0.80	−3	1.67	7	0.99	0.98–0.99
Residents	191	30.7	14.1	32.7	15.4	0.19	0.97	< 0.001	31.7	−1.98	−6	3.58	11	0.96	0.95–0.97
Experts	368	34.5	17.7	36.7	18.0	0.11	0.97	< 0.001	35.6	−2.13	−6	4.69	13	0.95	0.95–0.97
TR Vmax (m/s)															
Nurses	17	2.3	0.3	2.3	0.3	0.79	0.81	< 0.001	2.3	−0.03	−1	0.19	8	0.59	0.51–0.66
Residents	166	2.6	0.7	2.7	0.5	0.54	0.65	< 0.001	2.7	−0.04	−1	0.50	19	0.62	0.51–0.70
Experts	305	2.7	0.7	2.7	0.6	0.34	0.59	< 0.001	2.7	−0.05	−2	0.61	23	0.78	0.53–0.90
LV GLS (%)															
Nurses	18	−17.9	4.0	−17.0	3.6	0.46	0.73	< 0.001	−17.5	−0.96	5	2.78	−16	0.66	0.36–0.84
Residents	87	−16.4	4.5	−15.1	4.3	0.06	0.88	< 0.001	−15.8	−1.28	8	2.11	−13	0.84	0.77–0.89
Experts	126	−17.6	4.5	−16.1	4.3	0.01	0.87	< 0.001	−16.8	−1.52	9	2.18	−13	0.82	0.75–0.87
A4C LV GLS (%)															
Nurses	20	−17.9	3.6	−16.9	4.2	0.43	0.77	< 0.001	−17.4	−0.98	6	2.62	−15	0.83	0.66–0.92
Residents	92	−16.8	5.0	−15.2	4.4	0.03	0.79	< 0.001	−16.0	−1.55	10	3.08	−19	0.74	0.63–0.82
Experts	138	−18.0	4.8	−16.1	4.6	0.00	0.79	< 0.001	−17.0	−1.93	11	3.02	−18	0.70	0.61–0.78
A2C LV GLS (%)															
Nurses	20	−18.0	5.6	−16.8	3.9	0.46	0.47	< 0.001	−17.4	−1.14	7	5.01	−29	0.41	0.02–0.69
Residents	90	−15.9	4.6	−15.5	4.5	0.59	0.85	< 0.001	−15.7	−0.37	2	2.46	−16	0.85	0.78–0.90
Experts	140	−17.2	4.6	−16.4	4.5	0.15	0.74	< 0.001	−16.8	−0.78	5	3.27	−19	0.74	0.61–0.78
A3C LV GLS (%)															
Nurses	21	−18.2	4.6	−17.0	2.9	0.31	0.55	< 0.001	−17.6	−1.21	7	3.74	−21	0.45	0.09–0.71
Residents	89	−16.7	5.3	−14.9	4.5	0.01	0.72	< 0.001	−15.8	−1.82	12	3.73	−24	0.66	0.52–0.76
Experts	138	−18.2	5.4	−16.1	4.3	0.00	0.75	< 0.001	−17.1	−2.03	12	3.60	−21	0.66	0.55–0.74
LVEF MOD A4C (%)															
Nurses	24	57.2	8.4	63.8	9.2	0.01	0.49	< 0.001	60.5	−6.64	−11	8.77	14	0.37	0.04–0.63
Residents	191	53.3	14.1	57.9	14.3	0.00	0.78	< 0.001	55.6	−4.61	−8	9.39	17	0.73	0.66–0.79
Experts	354	54.5	12.9	55.2	13.0	0.45	0.76	< 0.001	54.8	−0.73	−1	8.87	16	0.76	0.71–0.80
LVEF MOD A2C (%)															

Table 2
(*Continued*)

	n	AI mean	AI SD	H mean	H SD	*P*	Pearson R	Pearson *P*	AI/H mean	AI/H mean diff	AI/H mean diff (%)	AI/H SD diff	AI/H SD diff (%)	ICC	95% CI
Nurses	23	58.6	9.5	61.8	9.9	0.28	0.52	< 0.001	60.2	−3.16	−5	9.27	15	0.52	0.32–0.71
Residents	187	54.3	13.2	57.2	14.9	0.05	0.73	< 0.001	55.8	−2.90	−5	10.47	19	0.70	0.62–0.77
Experts	308	55.4	12.2	54.8	16.7	0.59	0.56	< 0.001	55.1	0.63	1	14.18	26	0.53	0.45–0.61
LVEDV MOD A4C (mL)															
Nurses	24	97.2	24.4	120.3	35.9	0.01	0.65	< 0.001	108.8	−23.07	−21	26.83	25	0.57	0.28–0.76
Residents	192	101.4	44.1	112.0	55.2	0.04	0.90	< 0.001	106.7	−10.63	−10	24.19	23	0.86	0.82–0.89
Experts	354	106.3	42.9	119.7	52.0	0.00	0.91	< 0.001	113.0	−13.43	−12	21.71	19	0.86	0.83–0.88
LVEDV MOD A2C (mL)															
Nurses	23	76.7	21.3	90.5	21.2	0.03	0.80	< 0.001	83.6	−13.75	−16	13.16	16	0.79	0.60–0.89
Residents	187	95.5	43.4	107.3	55.5	0.02	0.89	< 0.001	101.4	−11.83	−12	25.52	25	0.84	0.79–0.88
Experts	308	97.9	42.8	111.1	47.9	0.00	0.88	< 0.001	104.5	−13.29	−13	22.53	22	0.83	0.80–0.87
LVESV MOD A4C (mL)															
Nurses	24	42.1	14.3	43.1	15.2	0.81	0.73	< 0.001	42.6	−1.02	−2	10.53	25	0.82	0.67–0.91
Residents	191	50.5	37.1	51.3	44.0	0.85	0.94	< 0.001	50.9	−0.77	−2	15.42	30	0.93	0.90–0.95
Experts	354	51.2	34.6	57.4	40.8	0.03	0.95	< 0.001	54.3	−6.20	−11	13.34	25	0.92	0.91–0.94
LVESV MOD A2C (mL)															
Nurses	23	31.8	11.1	34.1	9.7	0.44	0.76	< 0.001	32.9	−2.37	−7	7.19	22	0.89	0.78–0.94
Residents	187	46.7	34.9	51.0	45.2	0.30	0.92	< 0.001	48.9	−4.36	−9	18.63	38	0.89	0.85–0.91
Experts	308	46.1	31.9	53.5	38.2	0.01	0.92	< 0.001	49.8	−7.44	−15	14.96	30	0.89	0.86–0.91
LVOT Vmean (m/s)															
Nurses	27	0.7	0.1	0.7	0.1	0.13	0.91	< 0.001	0.7	−0.05	−8	0.06	9	0.88	0.78–0.94
Residents	246	0.7	0.2	0.7	0.1	0.08	0.82	< 0.001	0.7	0.02	3	0.09	13	0.80	0.75–0.84
Experts	491	0.7	0.1	0.7	0.1	0.91	0.88	< 0.001	0.7	0.00	0	0.07	10	0.88	0.86–0.90
LAESV MOD A4C (mL)															
Nurses	26	64.5	28.1	69.1	28.3	0.56	0.90	< 0.001	66.8	−4.63	−7	12.55	19	0.82	0.67–0.91
Residents	220	61.7	26.7	76.5	36.9	0.00	0.76	< 0.001	69.1	−14.78	−21	23.76	34	0.64	0.55–0.71
Experts	382	62.2	31.5	81.0	46.0	0.00	0.87	< 0.001	71.6	−18.83	−26	24.08	34	0.71	0.66–0.76
LAESV MOD A2C (mL)															
Nurses	24	66.0	21.7	68.0	26.8	0.77	0.77	< 0.001	67.0	−2.06	−3	16.74	25	0.67	0.42–0.82
Residents	177	63.5	30.8	79.3	36.9	0.00	0.84	< 0.001	71.4	−15.77	−22	19.85	28	0.73	0.66–0.80
Experts	265	57.5	31.4	74.6	41.8	0.00	0.82	< 0.001	66.1	−17.06	−26	23.94	36	0.71	0.64–0.76
AoV Vmean (m/s)															
Nurses	26	0.9	0.4	1.0	0.4	0.75	0.99	< 0.001	0.9	−0.04	−4	0.06	6	0.99	0.98–0.99
Residents	191	1.1	0.4	1.2	0.5	0.15	0.98	< 0.001	1.1	−0.07	−6	0.11	9	0.96	0.95–0.97
Experts	370	1.2	0.6	1.2	0.6	0.18	0.96	< 0.001	1.2	−0.06	−5	0.17	14	0.95	0.94–0.96

A2C LV GLS: apical-2-chamber view left ventricular global longitudinal strain; A3C LV GLS: apical-3-chamber view left ventricular global longitudinal strain; A4C LV GLS: apical-4-chamber view left ventricular global longitudinal strain; AI: artificial intelligence; AoV Vmax: aortic valve transvalvular maximal velocity; AoV Vmean: aortic valve transvalvular mean velocity; AoV VTI: aortic valve transvalvular velocity time integral; CI: confidence interval; Dec T: deceleration time of the E wave; diff: difference; H: human; ICC: intraclass correlation coefficient; IVSd: interventricular septum thickness in diastole; LAESV MOD A2C: left atrium end-systolic volume modified apical-2-chamber view; LAESV MOD A4C: left atrium end-systolic volume modified apical-4-chamber view; LAESV MOD biplane: left atrium end-systolic volume modified biplane; LV GLS: left ventricular global longitudinal strain; LV mass: left ventricular mass; LVEDV MOD A2C: left ventricular end-diastolic volume modified Simpson apical-2-chamber view; LVEDV MOD A4C: left ventricular end-diastolic volume modified Simpson apical-4-chamber view; LVEDV MOD biplane: left ventricular end-diastolic volume modified Simpson biplane; LVEF MOD A2C: left ventricular ejection fraction modified Simpson apical-2-chamber view; LVEF MOD A4C: left ventricular ejection fraction modified Simpson apical-4-chamber view; LVEF MOD biplane: left ventricular ejection fraction modified Simpson biplane; LVESV MOD A2C: left ventricular end-systolic volume modified Simpson apical-2-chamber view; LVESV MOD A4C: left ventricular end-systolic volume modified Simpson apical-4-chamber view; LVESV MOD biplane: left ventricular end-systolic volume modified Simpson biplane; LVIDd: left ventricular internal diastolic diameter; LVOT Vmax: left ventricular outflow tract maximal velocity; LVOT Vmean: left ventricular outflow tract mean velocity; LVOT VTI: left ventricular outflow tract velocity time integral; LVPWd: left ventricular posterior wall diameter; MV-A: mitral valve A wave velocity; MV-E: mitral valve E wave velocity; RV s': right ventricular systolic velocity; SD: standard deviation; TR Vmax: tricuspid regurgitation maximal velocity.

cordance, probably due to the usual poor echogenicity of older patients. However, most key parameters retained high ICC values (> 0.8).

Weight-based analysis showed that agreement was generally high in all weight groups, with variability increasing for patients with extreme weight. In the low-weight group, the LVEF displayed high agreement (ICC: 0.85), while in the high-weight group, concordance remained high (ICC: 0.77), but with higher variability. Some measurements, such as wall thickness and flow velocities, presented lower agreement in the extreme weight categories, suggesting that weight may influence the accuracy of certain parameters.

5. Discussion

This study investigated the integration capability of an AI system to automate measurements in a high-volume echocardiography department and concordance of the measurements generated by AI in 'real-world' conditions.

The main findings of our study were: (1) the feasibility analysis was positive, with a technical implementation time of < 4 weeks from receiving the AI server to capturing the first cases of echocardiographic modalities; (2) measurement comparison between AI and operators over the following 2 months (894 echocardiogram examinations) showed that overall agreement was high for most important parameters, but that disparities were present, thus inviting cautious clinical interpretability (of note is that in overweight and older patients, the agreement was indeed lower) and (3) some differences were observed depending on echocardiographer experience level: experts presented a higher agreement than residents and nurses.

5.1. Implementation feasibility

The successful integration of an AI system into an echocardiography department is associated with several critical challenges. The most important of these is compliance with data security and privacy regulations, particularly the GDPR, which governs the handling of sensitive patient data. The AI system must function within the facility's secure environment without transmitting data to external servers. Therefore, the AI solution used in the study was implemented internally in the IT department of the Bordeaux University Hospital, in compliance with strict security protocols. The system was connected to the hospital's internal network as a web IP server, so that no external virtual private network access was required. System integration into the existing echocardiography equipment was completed in < 4 weeks, underlining the simplicity of the technical setup. This rapid integration enabled seamless data flow between the conventional PACS (ComPACS) and the AI processing station, facilitating real-time analysis while ensuring full compliance with security and privacy standards. Moreover, this AI solution is vendor neutral and therefore broadly scalable.

5.2. Measurement concordance

The AI algorithm implemented by Us2.ai has been previously validated. Numerous publications have already demonstrated the reliability of AI in identifying echocardiographic views and performing automated measurements. For example, Tromp et al. [13] confirmed AI's ability to accurately recognize echocardiographic planes and extract important measurements. AI automated echocardiographic measurements [14] have also been validated, including accurate calculation of parameters such as EF [15–17] and GLS [18]. The ability of AI to continuously learn from large datasets has been shown to improve accuracy and minimize human error in measurement tasks. Moreover, Us2.ai's real-time measurements in < 2 minutes undoubtedly add great time value in terms of workflow integration.

The main objective of this study was to assess the agreement between human and AI measurements in a clinical setting. A wide user range with different expertise levels participated in this study, from nurses and residents to experienced echocardiographers, all working in the echocardiography department. Overall, the concordance analysis reported different statistical parameters, which showed heterogeneity in the results depending on measurements. Correlations were very strong, with a Pearson coefficient > 0.80 for most parameters. Nevertheless, it is the Bland-Altman and ICC analyses that provided the most authentic concordance picture. On average, differences between AI and human measurements were around 4%, with a standard deviation of 15%. In addition, the ICC exceeded 0.80 in 50% of the measurements, indicating true agreement between AI- and human-generated data, especially for critical parameters such as EF and strain, which were comparable to values obtained in validated studies.

The next question arising from this study could be: what is the true 'gold standard' for measurement – AI or human? [19]. Given the consistency and precision achieved by AI in various studies, AI may soon replace the human operator as the gold standard for echocardiographic measurements, at least before a full diagnosis is made [20,21].

In this context, AI use in routine practice increases measurement and improves both the result precision and echocardiography's overall diagnostic value. This progress will ultimately increase the expertise and diagnostic confidence resulting from these examinations and makes AI an important tool in modern cardiology [22].

5.3. Concordance results: subgroup analysis

The differences observed between qualification groups (nurses, residents and experts) are particularly interesting, even though case numbers were not homogeneous. When isolating the measurement comparisons of the experienced group, all statistical parameters increased significantly for the mean ICC and especially for parameters sensitive to reproducibility, such as EF, which increased from 0.58 for nurses to 0.84 for experts. This point is particularly noteworthy as it shows that concordance, and therefore reliability, depend on the user rather than AI. This supports the hypothesis that AI could become the gold standard, especially for operator-dependent parameters (EF, structural areas, volumes and tricuspid regurgitation) compared with simpler and more reproducible measurements (Vmax E and A, subaortic and aortic velocities, and the same for VTI).

Subgroup analyses based on patient age and weight were performed to test the hypothesis of the influence of window quality on concordance. The acoustic quality of the windows was indeed indirectly estimated from demographic characteristics (age, weight), in the absence of specific subjective qualitative assessment. No significant differences were found in the mean statistical assessment parameters. However, a parameter closely related to window quality, EF, showed an ICC decrease with age (from 0.97 to 0.73), as for parameters related to wall measurements. Similarly, extreme weight categories were associated with poorer concordance for the same measurements. While weight was selected as the primary variable due to its availability in our clinical database, body mass index may be a more accurate indicator of body composition and its impact on image quality. Future studies should consider incorporating body mass index to further refine the understanding of AI performance across different patient morphologies. Finally, due to the small case number in the nurse subgroup, it was unfortunately not possible to assess the relationship between ultrasound window quality and operator skill, although such a relationship is likely to exist.

One of the major potential benefits of AI in echocardiography is measurement standardization, thereby reducing inter-operator variability. Our results showed that concordance was higher for echocardiograms performed by experts than for those performed by residents and nurses. This may indicate that AI performs closer to expert-level standards, supporting the hypothesis that AI could improve reproducibility of measurements, particularly for less experienced operators. Future research should investigate the precise role of AI in minimizing inter-operator variability and its potential to harmonize echocardiographic assessment across different levels of expertise.

5.4. Study limitations

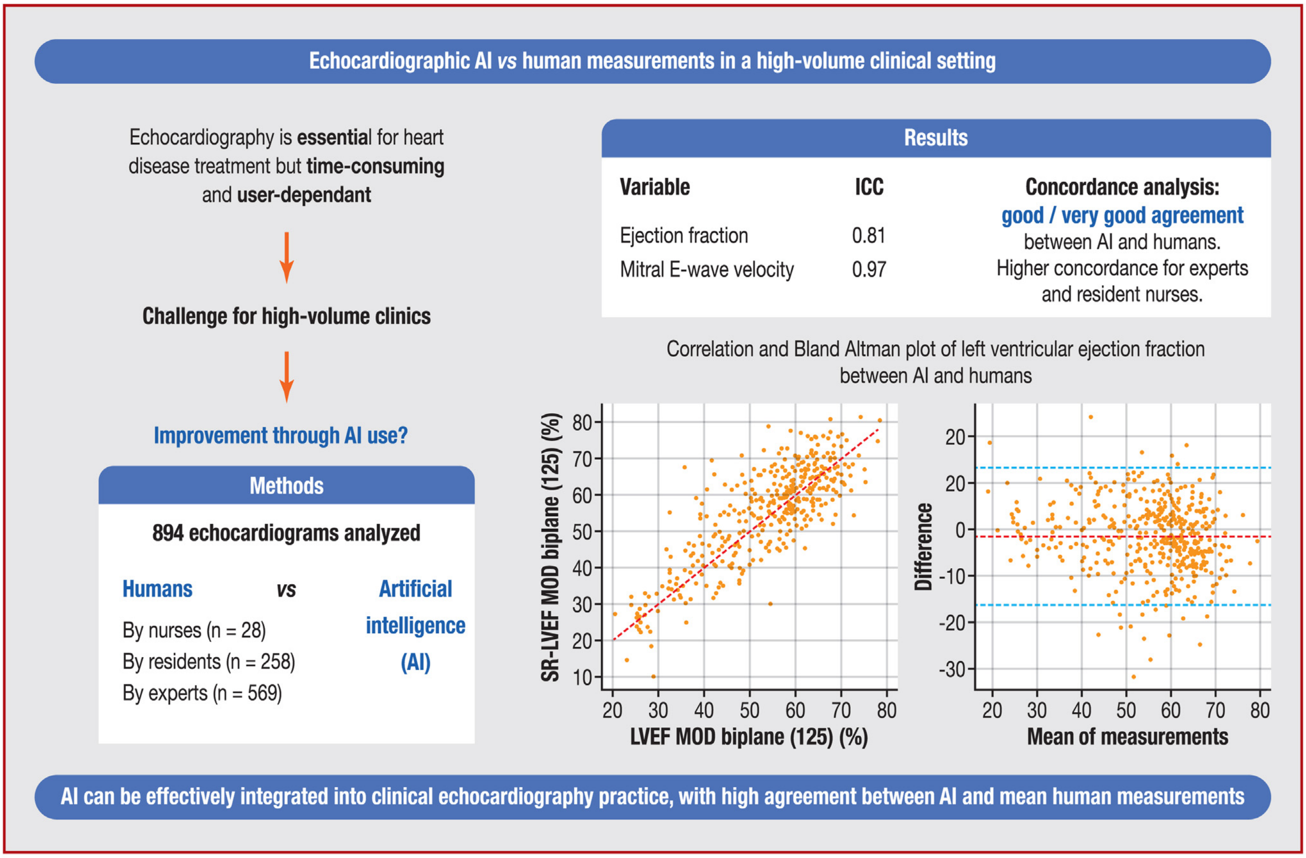
While this study emphasized a high agreement level between AI and human measurements, it has some limitations. Operators had varying experience levels, ranging from highly experienced seniors to residents in their first echocardiography training year, and nurses working under a cooperative protocol with measurements validated by a senior. Although this ensured measurement accuracy, the smaller number of examinations performed by nurses may limit the statistical power of subgroup analyses. This variability may have influenced the results, especially in the case of more complex measurements. However, as Us2.ai performs well in various demographic groups, its consistency could be valuable for less experienced operators.

The 2-month data collection period may not have captured the full range of clinical scenarios. The use of human-generated data as a reference carries the risk of bias, as AI may outperform human performance on certain measurements. Although our study demonstrated high concordance between AI and human measures, we did not assess the direct impact of AI on clinical decision-making, workflow efficiency or patient outcomes. These aspects are critical for understanding the true clinical value of AI integration. Therefore, future studies should focus on how AI affects diagnostic accuracy, decision confidence and patient management. In addition, longitudinal studies could help determine whether AI implementation translates into improved patient outcomes and streamlined echocardiography workflow.

Finally, this study was conducted in a scheduled setting, where all echocardiograms were scheduled at least 2 months in advance. The feasibility of AI in emergency echocardiograms, where image acquisition conditions are less controlled, remains an important concern for future investigation. Emergency echocardiograms often present additional challenges, including variable image quality, time constraints and less optimal patient positioning. Therefore, future studies should evaluate AI performance in these settings in order to determine its robustness under different clinical conditions.

6. Conclusions

This study showed that AI integration in a high-volume echocardiography department is feasible and yields a high degree of agreement with human measurements, even for operators with different experience levels. This broadly scalable AI system proved to be reliable, especially for standard echocardiographic measurements, and offers the potential to improve diagnostic accuracy. While further research is needed to assess the long-term impact on workflow and clinical outcomes, these results suggest that AI could play a key role in improving the reliability and efficiency of echocardiographic examinations in real-world clinical practice (Central Illustration).



Central Illustration. A total of 894 electrocardiograms were analysed by humans (nurses, residents and experts) and AI. Results showed that the agreement between AI and humans was good to very good, with a higher concordance for experts and residents versus nurses. AI: artificial intelligence; ICC: intraclass correlation coefficient.

Sources of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AI support

Some sections of this article were initially drafted with the assistance of AI.

Data availability statement

Data available upon reasonable request to the corresponding author.

Disclosure of interest

The authors declare that they have no competing interest.

Acknowledgements

We acknowledge the IT Department Management of Bordeaux University Hospital for their support in implementing the AI system and thank Us2.ai for making their resources available.

References

[1] Popescu BA, Stefanidis A, Nihoyannopoulos P, Fox KF, Ray S, Cardim N, et al. Updated standards and processes for accreditation of echocardiographic laboratories from The European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2014;15(7):717–27, http://dx.doi.org/10.1093/ehjci/jeu039.
[2] Gandhi S, Mosleh W, Shen J, Chow CM. Automation, machine learning, and artificial intelligence in echocardiography: a brave new world. Echocardiogr Mt Kisco N 2018;35(9):1402–18, http://dx.doi.org/10.1111/echo.14086.
[3] 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(11):e015569, http://dx.doi.org/10.1161/CIRCIMAGING.123.015569.
[4] Sabo S, Pasdeloup D, Pettersen HN, Smistad E, Østvik A, Olaisen SH, et al. Real-time guidance by deep learning of experienced operators to improve the standardization of echocardiographic acquisitions. Eur Heart J Imaging Methods Pract 2023;1(2):qyad040, http://dx.doi.org/10.1093/ehjimp/qyad040.
[5] Li X, Zhang H, Yue J, Yin L, Li W, Ding G, et al. A multi-task deep learning approach for real-time view classification and quality assessment of echocardiographic images. Sci Rep 2024;14(1):20484, http://dx.doi.org/10.1038/s41598-024-71530-z.
[6] Shiokawa N, Izumo M, Shimamura T, Kurosaka Y, Sato Y, Okamura T, et al. Accuracy and efficacy of artificial intelligence-derived automatic measurements of transthoracic echocardiography in routine clinical practice. J Clin Med 2024;13(7):1861, http://dx.doi.org/10.3390/jcm13071861.
[7] Louart B, Muller L, Emond B, Boulet N, Roger C. Agreement between manual and automatic ultrasound measurement of the velocity-time integral in the left ventricular outflow tract in intensive care patients: evaluation of the AUTO-VTI® tool. J Clin Monit Comput 2024;39(2):355–64, http://dx.doi.org/10.1007/s10877-024-01215-5.
[8] Jevsikov J, Ng T, Lane ES, Alajrami E, Naidoo P, Fernandes P, et al. Automated mitral inflow Doppler peak velocity measurement using deep learning. Comput Biol Med 2024;171:108192, http://dx.doi.org/10.1016/j.compbiomed.2024.
[9] Cai Q, Lin M, Zhang M, Qin Y, Meng Y, Wang J, et al. Automated echocardiographic diastolic function grading: a hybrid multi-task deep learning and machine learning approach. Int J Cardiol 2024;416:132504, http://dx.doi.org/10.1016/j.ijcard.2024.132504.
[10] Anand V, Weston AD, Scott CG, Kane GC, Pellikka PA, Carter RE. Machine learning for diagnosis of pulmonary hypertension by echocardiography. Mayo Clin Proc 2024;99(2):260–70, http://dx.doi.org/10.1016/j.mayocp.2023.05.006.
[11] Sadeghpour A, Jiang Z, Hummel YM, Frost M, Lam CSP, Shah SJ, et al. An automated machine learning-based quantitative multiparametric

approach for mitral regurgitation severity grading. JACC Cardiovasc Imaging 2024;18(1):1–12, http://dx.doi.org/10.1016/j.jcmg.2024.06.011 [S1936-878X(24)00247-X].

[12] 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(1):6776, http://dx.doi.org/10.1038/s41467-022-34245-1.

[13] 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(1):e46–54, http://dx.doi.org/10.1016/S2589-7500(21)00235-1.

[14] 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(5):1047–59, http://dx.doi.org/10.1007/s10554-021-02482-y.

[15] Moal O, Roger E, Lamouroux A, Younes C, Bonnet G, Moal B, et al. Explicit and automatic ejection fraction assessment on 2D cardiac ultrasound with a deep learning-based approach. Comput Biol Med 2022;146:105637, http://dx.doi.org/10.1016/j.compbiomed.2022.105637.

[16] Samtani R, Bienstock S, Lai AC, Liao S, Baber U, Croft L, et al. Assessment and validation of a novel fast fully automated artificial intelligence left ventricular ejection fraction quantification software. Echocardiogr Mt Kisco N 2022;39(3):473–82, http://dx.doi.org/10.1111/echo.15318.

[17] Liu X, Fan Y, Li S, Chen M, Li M, Hau WK, et al. Deep learning-based automated left ventricular ejection fraction assessment using 2-D echocardiography. Am J Physiol Heart Circ Physiol 2021;321(2):H390–9, http://dx.doi.org/10.1152/ajpheart.00416.2020.

[18] Kwan AC, Chang EW, Jain I, Theurer J, Tang X, Francisco N, et al. Deep learning-derived myocardial strain. JACC Cardiovasc Imaging 2024;17(7):715–25, http://dx.doi.org/10.1016/j.jcmg.2024.01.011.

[19] Galiuto L, Volpe M. Artificial intelligence in echocardiography: a better alternative to the human eye? Eur Heart J 2023;44(31):2891–2, http://dx.doi.org/10.1093/eurheartj/ehad401.

[20] Sengupta PP, Adjeroh DA. Will artificial intelligence replace the human echocardiographer? Circulation 2018;138(16):1639–42, http://dx.doi.org/10.1161/CIRCULATIONAHA.118.037095.

[21] 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(16):1623–35, http://dx.doi.org/10.1161/CIRCULATIONAHA.118.034338.

[22] Barone-Rochette G. Will artificial intelligence change the job of the cardiac imaging specialist? Arch Cardiovasc Dis 2020;113(1):1–4, http://dx.doi.org/10.1016/j.acvd.2019.11.002 [ISSN 1875-2136].