ESC European Society of Cardiology

European Heart Journal - Digital Health
https://doi.org/10.1093/ehjdh/ztac001

ORIGINAL ARTICLE

Clinical validation of an artificial intelligence-assisted algorithm for automated quantification of left ventricular ejection fraction in real time by a novel handheld ultrasound device

Stella-Lida Papadopoulou [1*†], Vasileios Sachpekidis[1†], Vasiliki Kantartzi[1], Ioannis Styliadis[1], and Petros Nihoyannopoulos[2,3]

[1]Department of Cardiology, Papageorgiou General Hospital, Ring Road, Nea Efkarpia, Thessaloniki 56403, Greece; [2]Imperial College London, National Heart & Lung Institute, The Hammersmith Hospital, Du Cane Road, London W120NN, UK; and [3]First Cardiology Department, Medical School, University of Athens, Hippokration Hospital, 114 Vasilissis Sofias Avenue, 11527 Athens, Greece

Received 3 November 2021; revised 20 December 2021; accepted 10 January 2022

Aims We sought to evaluate the reliability and diagnostic accuracy of a novel handheld ultrasound device (HUD) with artificial intelligence (AI) assisted algorithm to automatically calculate ejection fraction (autoEF) in a real-world patient population.

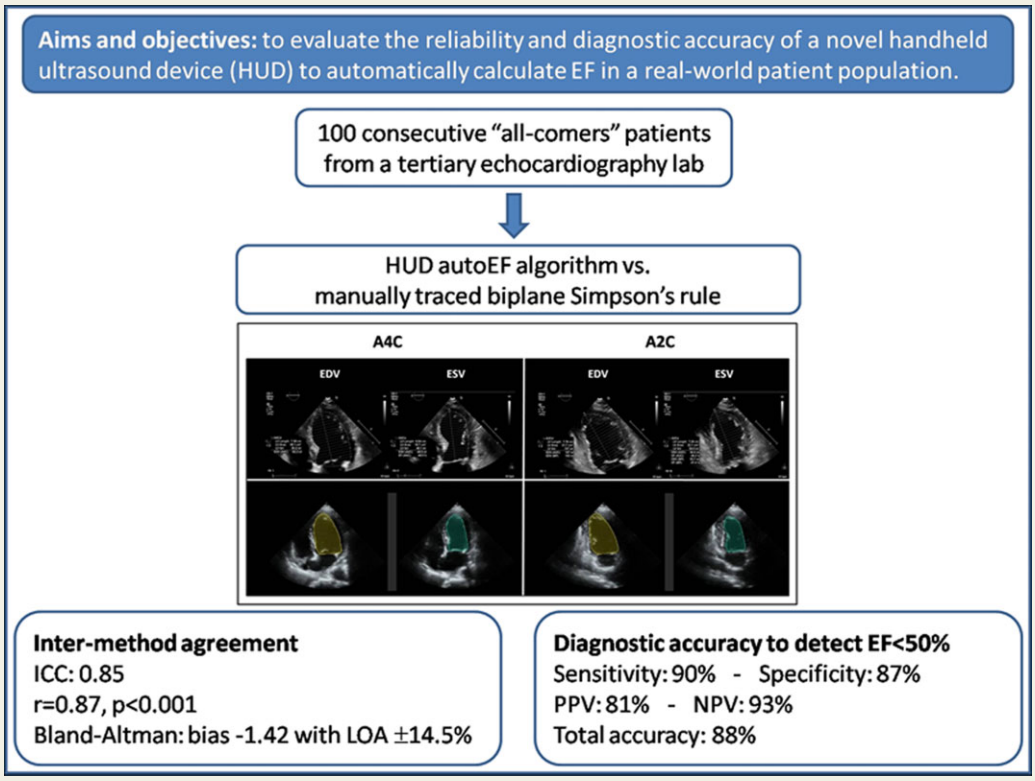
Methods and results We studied 100 consecutive patients (57 ± 15 years old, 61% male), including 38 with abnormal left ventricular (LV) function [LV ejection fraction (LVEF) < 50%]. The autoEF results acquired using the HUD were independently compared with manually traced biplane Simpson's rule measurements on cart-based systems to assess method agreement using intra-class correlation coefficient (ICC), linear regression analysis, and Bland–Altman analysis. The diagnostic accuracy for the detection of LVEF <50% was also calculated. Test–retest reliability of measured EF by the HUD was assessed by calculating the ICC and the minimal detectable change (MDC). The ICC, linear regression analysis, and Bland–Altman analysis revealed good agreement between autoEF and reference manual EF (ICC = 0.85; $r = 0.87$, $P < 0.001$; mean bias -1.42% with limits of agreement 14.5%, respectively). Detection of abnormal LV function (EF < 50%) by autoEF algorithm was feasible with sensitivity 90% (95% CI 75–97%), specificity 87% (95% CI 76–94%), PPV 81% (95% CI 66–91%), NPV 93% (95% CI 83–98%), and a total diagnostic accuracy of 88%. Test–retest reliability was excellent (ICC = 0.91, $P < 0.001$; $r = 0.91$, $P < 0.001$; mean difference ± SD: 0.54% ± 5.27%, $P = 0.308$) and MDC for LVEF measurement by autoEF was calculated at 4.38%.

Conclusion Use of a novel HUD with AI-enabled capabilities provided similar LVEF results with those derived by manual biplane Simpson's method on cart-based systems and shows clinical potential.

* Corresponding author. Tel: +30 231 332 3246, Email: elpa98@gmail.com

† The first two authors contributed equally to this article and are joint first authors.

Graphical Abstract



Keywords Left ventricular function • Handheld ultrasound device • Automated • Ejection fraction • Artificial intelligence • Echocardiography

Introduction

Transthoracic echocardiography is the most widely used cardiac imaging modality in clinical practice and plays a key role in the diagnosis, management, and follow-up of patients with cardiovascular disease.[1] Although far from perfect, the assessment of left ventricular ejection fraction (LVEF) as an index of LV systolic function has significant treatment and prognostic implications for heart failure patients.[2,3] Echocardiography due to its portability and ability for fast image acquisition and interpretation is the imaging modality of choice for assessment of LVEF in both elective and emergency settings.[1]

Most recently, machine learning techniques, commonly known as artificial intelligence (AI), have been employed to automatically identify LV endocardial boundaries and calculate LVEF. Several studies have evaluated AI-assisted echocardiography image analysis techniques, including automated contour-based segmentation, applied to routine two-dimensional (2D) images.[4,5] This approach, known as autoEF, has been developed using AI-learned pattern recognition programming trained on large databases of several thousand echocardiographic images of variable quality and diverse pathologies, in which endocardial boundaries have been traced by experts.

Technological advancements over the last two decades have enabled the development of miniaturized handheld ultrasound devices (HUDs) that are compact and battery operated and can provide echocardiographic images at the point of care with reasonable image quality. Their simplicity of use, availability, and relatively low cost compared to the high-end, fully equipped systems have made these devices attractive to use for rapid medical decision-making.[6] The widespread use of HUD has the potential to reform everyday practice of echocardiography, from the exclusive use of the technique by cardiologists in echocardiography laboratories, towards its use by other operators in the out-of-hospital setting.[7] Adequate training of operators in image acquisition, analysis, interpretation, and reporting is necessary to ensure quality in cardiac imaging. In addition, reliable quantification of imaging data is fundamental to reducing operator-dependent variability in echocardiography; however, until now, HUD has offered only limited quantification capabilities.[8] The application of AI algorithms (such as the autoEF) to point-of-care echocardiography may facilitate the optimization of acquisition and interpretation of a high volume of imaging data in real time, improve reproducibility, and diagnostic accuracy.

The objectives of this study were to test the ability of a novel HUD enabled with AI-assisted autoEF algorithm to (i) reliably

calculate LVEF compared to manually traced biplane Simpson's rule on cart-based machines and (ii) accurately identify impaired LV function in a real-world patient population.

Methods

Study population

Our study group comprised 100 consecutive 'all-comers' patients who were referred to our tertiary echocardiography laboratory over an approximate period of 6 weeks and agreed to participate in the study. All patients were >18 years old, haemodynamically stable, and underwent a clinically indicated transthoracic echocardiogram without contrast. Patients with atrial fibrillation or flutter and frequent atrial and ventricular ectopic beats were excluded from the study due to variation of EF between different cardiac cycles, often seen in this population. Patients with very bad image quality that precluded reliable assessment of LVEF with the biplane Simpson's method on the cart-based system were also excluded. All patients provided informed consent and were entered in our single-centre echocardiography database. The study conformed to the regulations of local ethics and the principles of the Declaration of Helsinki.

Standard echocardiography evaluation of left ventricular ejection fraction

All echocardiographic examinations were performed using a commercially available cart-based system (IE33, Affinity or Epic, Philips, Inc.). Images were acquired by an expert investigator (V.S.—level 3 training in echocardiography with 15 years of experience), following a standardized protocol. The 2D views used in our study were apical four-chamber (A4C) and apical two-chamber (A2C) views with the patient in left lateral decubitus position. Images were optimized to improve the signal-to-noise ratio and provide optimal endocardial definition. The LV endocardium was used as the boundary for volumetric measurements. Papillary muscles and visible trabeculae were part of the blood pool. If endocardial border was indistinguishable, non-visible parts were interpolated manually. The image quality for each examination was assessed and classified as good, moderate, and poor based on the number of LV walls (septal, anterior, lateral, and inferior) that endocardial borders were not clearly definable in end-diastole (0, 1, or ≥2, respectively). The modified biplane Simpson's method of discs was used to determine LV volumes and function. End-diastolic and end-systolic endocardial borders were traced manually on frozen 2D images obtained from the A4C and A2C views to derive end-diastolic volume (EDV) and end-systolic volume (ESV). End-diastole was defined as the peak of the electrocardiographic R wave and/or one frame before mitral valve closure. End-systole was defined as one frame before mitral valve opening or when end-systolic volume was deemed smallest by the operator. The LVEF was calculated according to the formula EF = (EDV - ESV)/EDV × 100% (*Figure 1A* and *B*). The time for obtaining the EF by manual tracing was calculated for the first 20 patients. The cart-based EF (CB-EF) values were considered the reference values for all comparisons. Furthermore, the CB-EF measurements were used to classify patients into three categories of LV function (reduced, mildly reduced, and preserved EF) as defined in the 2021 ESC Heart Failure guidelines[2] for LVEF ≤40%, 41–49%, and ≥50%, respectively. In addition, based on the CB-EF measurements, an LVEF <50% was considered a clinically relevant cut-off value to define abnormal LV function.

LVEF calculation by AI-assisted autoEF in handheld ultrasound device

All the study participants were subsequently scanned by the same cardiologist with a novel HUD (Kosmos, EchoNous, Inc.) equipped with a 2- to 5-MHz phased-array transducer (*Figure 2*), which allows calculation of LVEF with an AI-assisted autoEF algorithm. AutoEF applies the concept of learned pattern recognition from AI theory, which generically seeks to mimic human behaviour and learn from past experiences. The Kosmos AI platform (https://kosmosplatform.com/) is developed on convolution neural networks, which have been trained on expert annotated ultrasound clips. Of importance, no ECG tracing is required for the algorithm to operate. The Kosmos AI-assisted EF workflow follows the Simpson's Biplane LVEF as recommended by the American Society of Echocardiography.[8] The workflow begins with acquisition of A4C and A2C image sequences (clips) of the heart (5 s each). The A4C and A2C clips are then processed by a deep learning algorithm to compute the LVEF (processing time of ∼5 s). The algorithm first identifies the end-diastolic (ED) and end-systolic (ES) frames in both A4C and A2C clips and then segments the LV in all four frames (A4C ED and ES, A2C ED, and ES frames).

Fully automated estimation of the LVEF was possible after acquisition of these two views with the use of the AI-assisted autoEF algorithm, and the result was immediately saved without any correction by manual tracing (*Figure 1C* and *D*). After ∼1 h, a second independent acquisition with

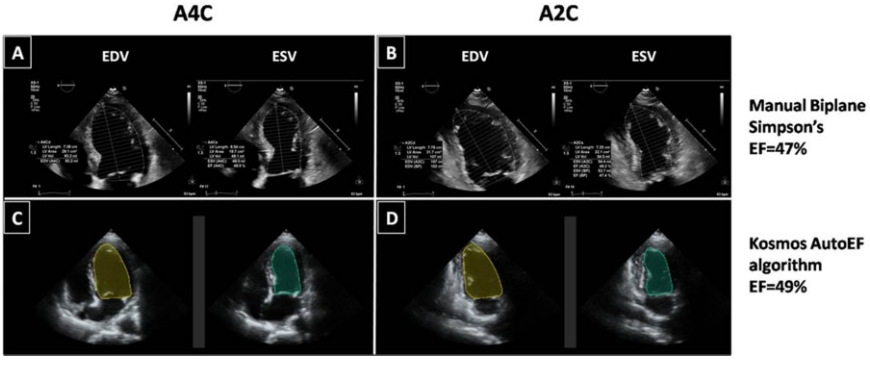


Figure 1 Endocardial border detection by manual biplane Simpson's method (*A* and *B*) and Kosmos HUD autoEF algorithm (*C* and *D*) in end-diastole and end-systole. All datasets are derived from the same patient, with a mildly reduced EF. A2C, apical 2-chamber; A4C, apical 4-chamber; EDV, end-diastolic volume; EF, ejection fraction; ESV, end-systolic volume; HUD, handheld ultrasound device.

Figure 2 The handheld ultrasound device used in the study (Kosmos, EchoNous Inc.), equipped with an artificial intelligence-assisted autoEF algorithm. The device consists of a tablet and probe, which connects to the tablet via a cable.

automated EF calculation was obtained for all subjects by the same observer in order to assess measurement reliability. The acquired Kosmos data (A4C and A2C views) for each patient were assessed for image quality and classified as good, moderate, and poor in a similar way as described above for the cart-based system images.

Reliability of autoEF acquired measurements

The reliability of the measured LVEF by the AI-assisted autoEF algorithm was assessed in two different ways, as intra- and inter-acquisition variability. First, the internal validity of the autoEF algorithm was tested on a subset of 10 randomly selected patient datasets, which were reanalysed offline by applying the AI-assisted algorithm 20 times on each original acquired dataset. Second, the test–retest reliability of the method was evaluated using the second independent autoEF measurement acquired using the HUD.

Statistical analysis

Normal distributions of variables were checked before analysis. Continuous variables were expressed as means ± standard deviation or medians with interquartile range (IQR) when not normally distributed; categorical variables were presented as counts and/or percentages. Agreement between CB-EF and HUD autoEF measurements was evaluated by intraclass correlation coefficient (ICC), linear regression analysis, and Bland–Altman analysis.[9] The 95% limits of agreement (LOA) were defined as the range of values between ±1.96 standard deviations from the mean difference. Weighted Cohen's kappa[10] was calculated to determine the inter-method agreement in classifying patients into different categories of LV function. It is usual to characterize the strength of agreement with kappa values as follows: <0 none, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, and 0.81–1.00 almost perfect. Comparison between continuous variables was performed using paired Student's *t*-test or analysis of variance (ANOVA) with Bonferroni's correction in post-hoc tests, whereas the variables not normally distributed were compared with the non-parametric Kruskal–Wallis test and the Wilcoxon-signed ranks test. Categorical variables were compared using the χ^2 test. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall diagnostic accuracy of Kosmos autoEF to detect LVEF <50% were calculated using 2 × 2 contingency tables and the corresponding 95% confidence intervals (CIs) were determined.

Test–retest reliability of measurements obtained by the autoEF was assessed using the ICC for single measures (two-way mixed model with interaction for the absolute agreement), linear regression analysis, Bland–Altman plot, and the minimal detectable change (MDC) which represents the minimal change required to ascertain that the differences observed reflect a real change rather than measurement error with at least 95% confidence.[11] The Standard Error of Measurement (SEM) was computed as $SD\sqrt{1-ICC}$ and subsequently, the MDC was calculated as $1.96 \times \sqrt{2} \times SEM$.

As a complementary analysis, a multiple linear regression model was constructed using the stepwise method in order to identify potential explanatory variables for the discrepancy between the CB-EF and the autoEF calculation. The absolute LVEF difference between the two methods was used as the dependent variable, whereas age, gender, image quality (as dichotomous variable poor vs. moderate/good), and body mass index were used as explanatory variables. Graphical inspection of scatter plots between predicted values of the dependent variable and regression standardized residuals excluded heteroscedasticity of distribution. The potential impact of image quality on measurement error was further investigated separately across the three categories (good, moderate, and poor).

For all statistical tests, a two-tailed *P* value <0.05 was considered statistically significant. Statistical analysis was performed using SPSS software, version 22.0 (IBM Inc., Chicago, IL, USA).

Table 1 Clinical and echocardiographic data of the study population (*n* = 100)

Patient characteristic	
Age (years)	57 ± 15
Male gender (%)	61
Height (cm)	171.4 ± 8.7
Weight (kg)	81.3 ± 13.9
BMI (kg/m^2)	27.8 ± 4.3
Main diagnosis	
Valvular heart disease	23
Cardiomyopathy	13
CAD	29
Cancer-related treatment	11
Post-COVID-19 infection	4
Hypertension	2
Ischaemic stroke	5
Palpitations/arrhythmias	4
Diabetes mellitus	2
Other[a]	7
Calculated LVEF	
Cart-based biplane Simpson's EF	51 ± 15
Kosmos AI-assisted autoEF	49 ± 12
LVEF category (reference method)	
Reduced (≤40%)	25
Mildly reduced (41–49%)	13
Preserved (≥50%)	62

Values are expressed as mean ± SD or percentage (%) of patients unless otherwise indicated.
BMI, body mass index; CAD, coronary artery disease; COVID, coronavirus disease; LVEF, left ventricular ejection fraction.
[a]Pericardial effusion, acute kidney failure, pre-eclampsia, pulmonary embolism, aortic aneurysm, and myositis.

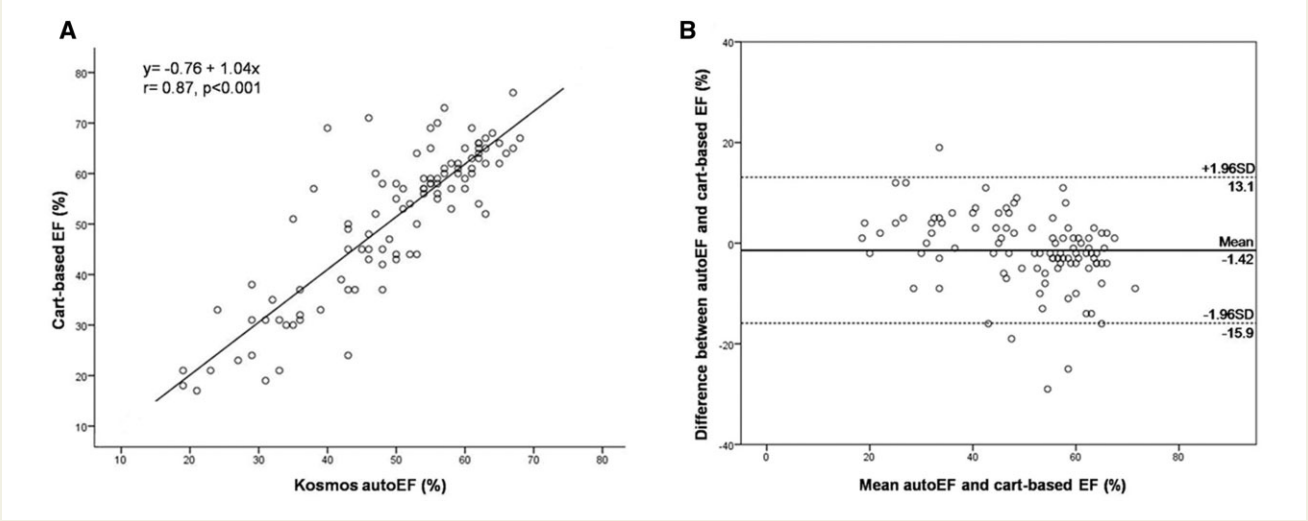


Figure 3 Linear regression analysis (*A*) and Bland–Altman plot (*B*) between Kosmos HUD autoEF algorithm and manual biplane Simpson's EF on cart-based system. EF, ejection fraction; HUD, handheld ultrasound device.

Table 2 Classification of patients into LV function categories with both methods of cart-based biplane Simpson's EF and Kosmos HUD autoEF

		Kosmos HUD autoEF			
Cart-based EF		**≤40%**	**41–49%**	**≥50%**	**Total**
	≤40%	20	5	0	25
	41–49%	0	9	4	13
	≥50%	3	5	54	62
	Total	23	19	58	100

EF, ejection fraction; HUD, handheld ultrasound device. The diagonal elements of the classification table (light blue shaded cells) represent correct classification into LV function categories.

Results

The study prospectively included 100 consecutive 'all-comers' patients (mean age 57 ± 15 years, 61% male); among them, 38% had an abnormal LVEF of <50%. Left ventricular ejection fraction measurements were completed for both cart-based systems and HUD in all patients included in the study. The clinical and echocardiographic characteristics of the study population are listed in *Table 1*. The image quality for the cart-based systems acquisition was assessed as good in 45%, moderate in 50%, and poor in 5% of cases, and for the HUD acquisition as good in 31%, moderate in 57%, and poor in 12% of cases. The average time for obtaining the EF by manual tracing was 84 ± 17 s, whereas the HUD autoEF algorithm provided EF calculation in ∼15 s for all patients.

Agreement between methods and diagnostic accuracy

There was good agreement between the calculated CB-EF and Kosmos autoEF (ICC = 0.85; 95% CI: 0.78–0.90). The results of linear regression analysis revealed a correlation coefficient $r = 0.87$, $P < 0.001$ (*Figure 3A*). The corresponding Bland-Altman plot shows a minimal non-significant bias of -1.42% ($P = 0.058$), with LOA of 14.5% for the auto-EF (*Figure 3B*). The paired comparison of the LVEF calculation by the two methods did not reveal a significant difference between CB-EF and HUD autoEF [56% (IQR 40–62%) vs. 53% (IQR 43–59%), respectively, $P = 0.106$].

Regarding the ability of Kosmos autoEF to correctly classify patients into three categories of LV function (reduced, mildly reduced, and preserved EF), the weighted κ-coefficient was 0.76 (judged as good). The detailed distribution of cases across LV function categories is presented in *Table 2* and *Figure 4*. The Kosmos autoEF AI-assisted algorithm was able to detect abnormal LV function (EF < 50%) with sensitivity 90% (95% CI: 75–97%), specificity 87% (95% CI: 76–94%), PPV 81% (95% CI: 66–91%), NPV 93% (95% CI: 83–98%), and total diagnostic accuracy of 88%.

Test–retest reliability analyses

The intra-acquisition variability for the automated EF measurements on each of the 10 randomly selected patients dataset was 0 (calculated within-subject SD = 0) since by default the AI algorithm followed the same pattern recognition on repeated analyses of the same acquisition (*Figure 5*). Finally, the inter-acquisition test–retest reliability for HUD measurements 1 and 2 was deemed as excellent (ICC = 0.91; 95% CI: 0.87–0.94, $P < 0.001$). The results of linear regression analysis revealed a correlation coefficient $r = 0.91$, $P < 0.001$ (*Figure 6A*). There was no significant difference in autoEF measurements between acquisition 1 and 2 (mean difference ± SD: 0.54% ± 5.27%, $P = 0.308$) and the corresponding Bland–Altman plot is presented in *Figure 6B*. The calculated MDC for the repeated LVEF measurements using autoEF algorithm was 4.38%.

The multiple linear regression model with the possible predictors of the absolute EF difference revealed that BMI was the only

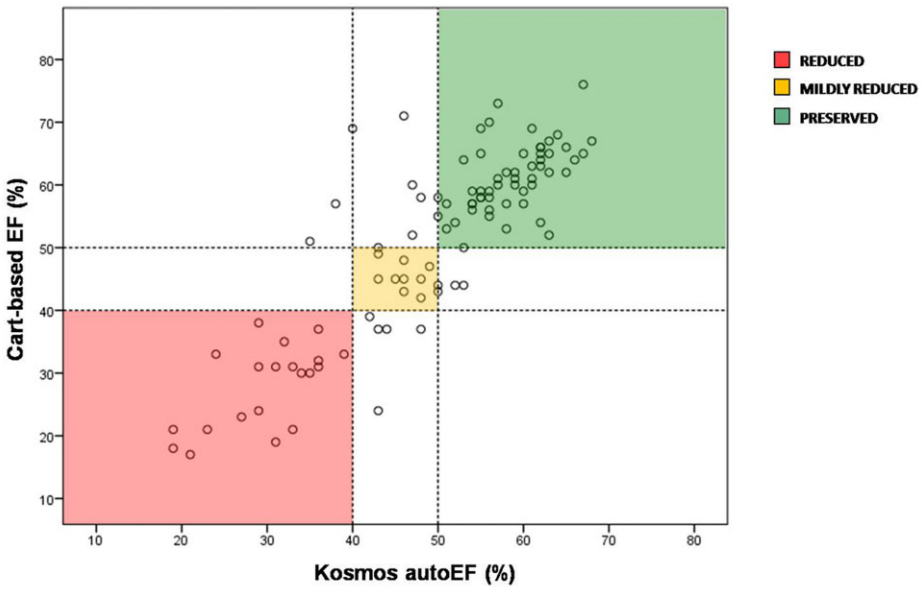


Figure 4 Scatter plot of EF measurements derived by Kosmos HUD autoEF algorithm and manual biplane Simpson's EF on cart-based system across different EF categories. The reduced, mildly reduced, and preserved EF groups are represented with red, yellow, and green colour, respectively. EF, ejection fraction; HUD, handheld ultrasound device.

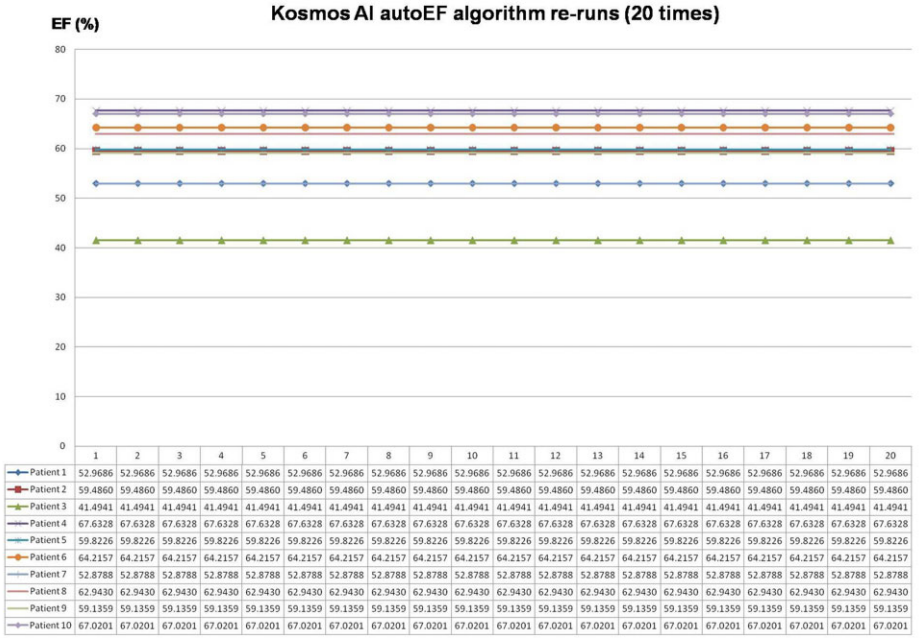


Figure 5 Results of the 20 re-runs of the autoEF algorithm on each of the 10 randomly selected patient datasets. AI, artificial intelligence; EF, ejection fraction.

statistically significant explanatory variable, which remained in the model ($B = 0.309$, 95% CI 0.067–0.551; $P = 0.013$). Furthermore, across the three categories of image quality (good, moderate, and poor), there was a non-significant trend for increasing absolute EF difference (4.6% ± 3.9%, 5.4% ± 5.0%, and 7.8% ± 8.3%, respectively, $P = 0.211$), as presented in *Figure 7*.

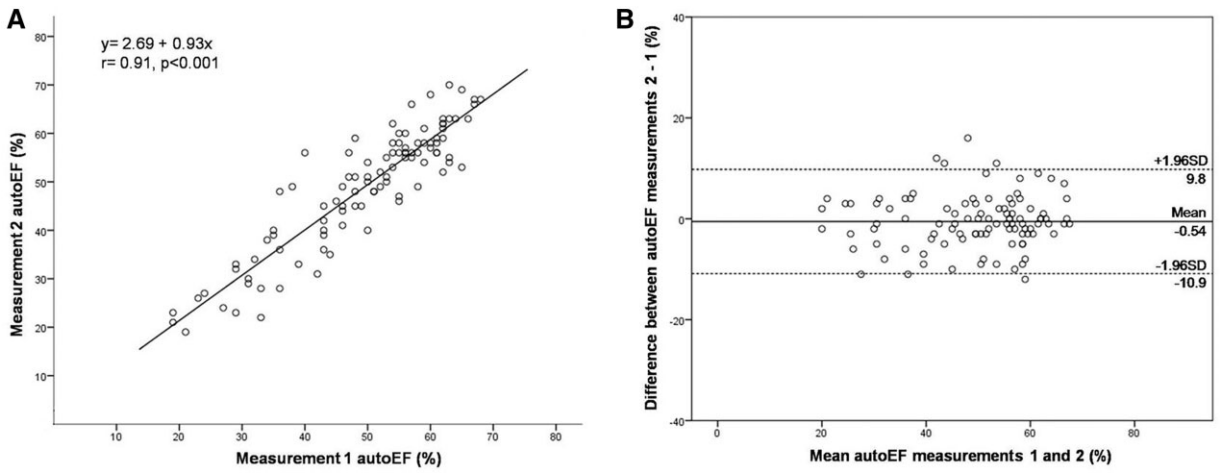


Figure 6 Linear regression analysis (*A*) and Bland–Altman plot (*B*) between the two separate measurements with the Kosmos device autoEF algorithm.

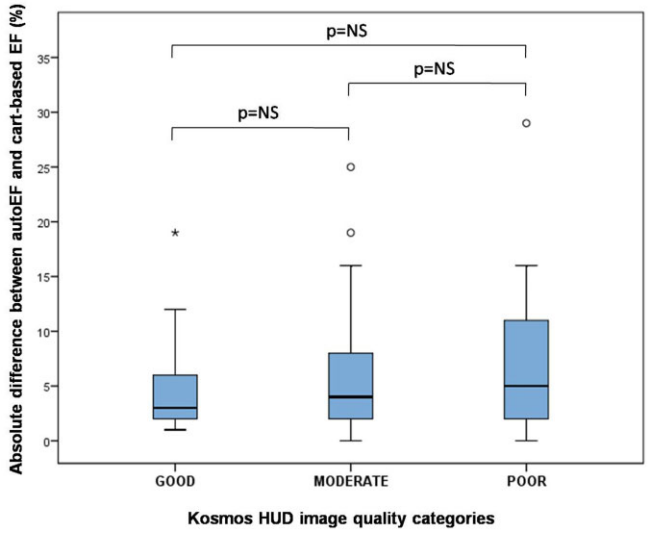


Figure 7 The mean absolute EF differences between Kosmos autoEF algorithm and manual biplane Simpson's EF on cart-based system across the three categories of image quality (good, moderate, and poor). EF, ejection fraction; NS, non-significant.

Discussion

The main finding of this study was that a fully automated measurement of EF using a novel HUD is feasible within a few seconds, and its results are reliable and comparable to the ones derived by standard method (biplane Simpson's method of discs). The AI-assisted autoEF was able to classify patients into different LV function categories and to identify LVEF <50% with good diagnostic accuracy compared to the cart-based echocardiography systems.

Over the last decades, referrals for echocardiographic examinations have increased substantially, occupying significant amount of time and resources of echocardiography laboratories.[12] Additionally, increased waiting times for service deliverance can lead to delays in diagnosis and treatment of cardiovascular disease with potential detrimental effects on the prognosis of patients. The use of HUD has the potential to influence bedside patient treatment decisions and expedite health care. A previous study by Gorcsan *et al.*[13] in 235 hospitalized patients showed that the HUD had an immediate effect on patient management in 149 patients (63%), i.e. either a change in medical therapy or a change in their diagnostic workup (most with changes in both). Furthermore, HUD could serve as a gatekeeper to standard echocardiography, especially in the setting of rarely appropriate indications. A recent study

demonstrated that a HUD echocardiography strategy led to a 59% decrease in the need for standard echocardiography and reduced considerably the total cost and time to decision making.[14] The use of HUD echocardiography as an initial screening tool prior to standard echocardiography is cost-effective, suggesting that the HUD is on the verge of reforming the current diagnostic strategy in clinical practice.[15] This has mainly resulted from technological advancements that have significantly improved the image quality of HUD, increasing doctors' confidence to use it as a reliable diagnostic bedside tool. In keeping with this, in only 12% of patients in our 'all-comers' population study, image quality of HUD was judged as poor.

Furthermore, it is widely accepted that the calculation of LVEF is pivotal in clinical decision-making. Many clinical trials investigating medical therapies for cardiology patients are based on LVEF estimation. More recent data continue to support the importance of LVEF to predict patient outcome and its impact on patient selection for novel therapies.[16] Nevertheless, a large variability in LVEF measurements has been observed between different centres and treatment strategy may be confounded in up to one-fifth of patients when decisions are based on LVEF.[17] Consequently, the clinical need for more reproducible ways of assessment of LV function has become apparent. The validity of autoEF algorithms implemented on standard echocardiography examinations has been previously investigated as a way to provide fast, reliable, and reproducible LVEF calculations. In early studies of computer-assisted EF analysis, correlation with 2D LVEF was suboptimal.[18] Nevertheless, numerous studies using autoEF software from different vendors have now shown high feasibility (83–100%) for biplane 2D LVEF measurements and excellent agreement with core laboratory measurements.[19–23]

Our study did not include a standard manual EF assessment on the HUD images because the device is devoid of the ability to complete manual LV endocardial border tracing; due to the size and touch screen functionality of these miniaturized devices, manual tracing for EF calculation is tedious and prone to error. There is a capability for the user to adjust the autoEF algorithm delineated borders (if deemed necessary); however, this would be a semi-automatic method, rather than standard manual EF assessment and would ultimately increase processing time. In our opinion, manual EF assessment using HUDs could not be implemented in clinical practice, because these devices are mainly used at the point-of-care, sometimes under emergency circumstances, which actually highlights the necessity for reliable autoEF algorithms on HUDs.

Notably, there are very limited data regarding the implementation of autoEF algorithms on HUD echocardiography. In an early study using images derived from HUD, the autoEF analysis was performed offline and the software required the examiner to define three regions of interest in the left ventricle, thus the method was not fully automated.[24] A recent study using an autoEF algorithm in HUD showed good correlation with the 3D measurement of LVEF on stationary echocardiography; however, the examinations with poor and moderate image quality were excluded from the analysis, which could affect the generalizability of the results.[25] Our investigation showed that the AI-assisted autoEF algorithm in a novel HUD can be used on a real-world patient population with results comparable to the cart-based echocardiography system and showed high sensitivity (90%) to detect impaired LVEF <50%. Importantly, patients with poor image quality on HUD were included in our method agreement analysis. In addition, the autoEF algorithm provided faster EF calculation in ~15 s for all patients than the manual tracing on cart-based system.

Besides the significance of accurate EF calculation in a timely manner, obtaining reliable measurements is also fundamental and gives confidence that the measured values can be used to make clinical decisions. An important factor for measurement variability is test–retest reliability (also called reproducibility or repeatability), which describes variability of separately acquired and interpreted echocardiographic measurements of the same patient. Early studies identified repeated acquisitions as the major component of variability of conventional echo parameters in a test–retest setting.[26] Our results demonstrated that the measurements provided by the autoEF algorithm are reliable; in addition, the ICC and MDC values are similar to the test–retest reliability reported in the literature for the biplane Simpson's method used in standard echocardiography for EF, with MDC ranging from 4.4% to 18.1%.[26–29] Importantly, the MDC of 4.38% we found for the calculation of LVEF by autoEF is below the 5% threshold that is often used in clinical practice to designate a meaningful change in LVEF in several clinical scenarios, such as the follow-up of oncology patients.[29]

Consequently, the reliable and rapid calculation of LVEF in real-time at the point-of-care can have significant clinical implications; apart from the use in hospitalized patients in need for a rapid assessment, it could be applied potentially as an initial screening tool on large scale healthy populations,[30] even by physicians who have received appropriate training but are non-experts in echocardiography. This is also supported by the high NPV of the AI algorithm performance in our population. Nevertheless, it has to be emphasized that the role of HUD is not to replace standard echocardiography but to act as a gatekeeper and facilitate workflow for echocardiography laboratories, whereas at the same time it should be able to identify individuals who must be further investigated by standard echocardiography. Of course, it is important to remember that when comparing HUD with fully equipped cart-based systems, one has to weight the advantages of portability and availability of echocardiography at the point of care with its current technological limitations.

Limitations

This is a single-centre and a single-operator study; however, we used the scanner exactly as we would use it in our clinical practice, which makes our findings relevant to the real-world patient population referred to a tertiary echocardiography laboratory. Patients with very bad image quality in whom the biplane Simpson's rule could not be applied on the cart-based system were excluded since no standard method to compare the performance of the autoEF algorithm would be present. However, such patients are uncommon in everyday clinical practice. Our analysis included tests with poor image quality, contrary to other studies. We did not acquire multiple datasets using the high-end systems to do a test–retest variability analysis. Nevertheless, it was outside the scope of our study, and the test–retest reliability for Simpson's EF using standard high-end systems has been previously reported in the literature.[26–29] Finally, given the inclusion of a relatively small number of patients, the possibility of a Type II error should be considered.

Conclusion

The AI-assisted autoEF algorithm in a novel HUD can automatically calculate LVEF in realtime in a fast and reliable way as compared to the recommended manual biplane Simpson's method on cart-based systems. These findings may have clinical potential.

Lead author biography

Stella-Lida Papadopoulou is a medical doctor and cardiologist with a special interest in multimodality cardiovascular imaging and digital applications. She obtained her MD from Aristotle University of Thessaloniki (GR). She was involved in several projects on the field of medical imaging at Erasmus Medical Center in Rotterdam (NL), as a clinical research fellow at the Departments of Cardiology and Radiology. Her work resulted in a PhD thesis 'Imaging of coronary atherosclerosis with CT coronary angiography' in 2016. After completing her cardiology training in 2020, she received a scholarship by the Hellenic Society of Cardiology for postgraduate fellowship in Digital Cardiology.

Funding

S.-L.P. was supported by a scholarship from the Hellenic Society of Cardiology.

Conflict of interest: none declared.

Data availability

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

References

1. Steeds RP, Garbi M, Cardim N, et al.; 2014–2016 EACVI Scientific Documents Committee. EACVI appropriateness criteria for the use of transthoracic echocardiography in adults: a report of literature and current practice review. *Eur Heart J Cardiovasc Imaging* 2017;**18**:1191–1204.
2. McDonagh TA, Metra M, Adamo M, *et al.*; ESC Scientific Document Group. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 2021;**42**:3599–3726.
3. Curtis JP, Sokol SI, Wang Y, *et al.* The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure. *J Am Coll Cardiol* 2003;**42**:736–742.
4. Nolan MT, Thavendiranathan P. Automated quantification in echocardiography. *JACC Cardiovasc Imaging* 2019;**12**:1073–1092.
5. Schuuring MJ, Isgum I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine echocardiography and artificial intelligence solutions. *Front Cardiovasc Med* 2021;**8**:648877.
6. Chamsi-Pasha MA, Sengupta PP, Zoghbi WA. Handheld echocardiography: current state and future perspectives. *Circulation* 2017;**136**:2178–2188.
7. Cardim N, Dalen H, Voigt JU, *et al.* The use of handheld ultrasound devices: a position statement of the European Association of Cardiovascular Imaging (2018 update). *Eur Heart J Cardiovasc Imaging* 2019;**20**:245–252.
8. Lang RM, Badano LP, Mor-Avi V, *et al.* Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2015;**16**:233–270.
9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**327**:307–310.
10. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;**70**:213–220.
11. Bunting KV, Steeds RP, Slater LT, Rogers JK, Gkoutos GV, Kotecha D. A practical guide to assess the reproducibility of echocardiographic measurements. *J Am Soc Echocardiogr* 2019;**32**:1505–1515.
12. Rahimi AR, York M, Gheewala N, Markson L, Hauser TH, Manning WJ. Trends in outpatient transthoracic echocardiography: impact of appropriateness criteria publication. *Am J Med* 2011;**124**:740–746.
13. Gorcsan J 3rd, Pandey P, Sade LE. Influence of hand-carried ultrasound on bedside patient treatment decisions for consultative cardiology. *J Am Soc Echocardiogr* 2004;**17**:50–55.
14. Pathan F, Fonseca R, Marwick TH. Usefulness of hand-held ultrasonography as a gatekeeper to standard echocardiography for "rarely appropriate" echocardiography requests. *Am J Cardiol* 2016;**118**:1588–1592.
15. Kitada R, Fukuda S, Watanabe H, *et al.* Diagnostic accuracy and cost-effectiveness of a pocket-sized transthoracic echocardiographic imaging device. *Clin Cardiol* 2013;**36**:603–610.
16. Zannad F, Ferreira JP, Pocock SJ, *et al.* SGLT2 inhibitors in patients with heart failure with reduced ejection fraction: a meta-analysis of the EMPEROR-Reduced and DAPA-HF trials. *Lancet* 2020;**396**:819–829.
17. Kaufmann BA, Min SY, Goetschalckx K, *et al.* How reliable are left ventricular ejection fraction cut offs assessed by echocardiography for clinical decision making in patients with heart failure? *Int J Cardiovasc Imaging* 2013;**29**:581–588.
18. Rahmouni HW, Ky B, Plappert T, *et al.* Clinical utility of automated assessment of left ventricular ejection fraction using artificial intelligence-assisted border detection. *Am Heart J* 2008;**155**:562–570.
19. Asch FM, Poilvert N, Abraham T, *et al.* Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. *Circ Cardiovasc Imaging* 2019;**12**:e009303.
20. Cannesson M, Tanabe M, Suffoletto MS, *et al.* A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction. *J Am Coll Cardiol* 2007;**49**:217–226.
21. Hovnanians N, Win T, Makkiya M, Zheng Q, Taub C. Validity of automated measurement of left ventricular ejection fraction and volume using the Philips EPIQ system. *Echocardiography* 2017;**34**:1575–1583.
22. Knackstedt C, Bekkers SC, Schummers G, *et al.* Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs multicenter study. *J Am Coll Cardiol* 2015;**66**:1456–1466.
23. Szulik M, Pappas CJ, Jurcut R, *et al.* Clinical validation of a novel speckle-tracking-based ejection fraction assessment method. *J Am Soc Echocardiogr* 2011;**24**:1092–1100.
24. Frederiksen CA, Juhl-Olsen P, Hermansen JF, Andersen NH, Sloth E. Clinical utility of semi-automated estimation of ejection fraction at the point-of-care. *Heart Lung Vessel* 2015;**7**:208–216.
25. Filipiak-Strzecka D, Kasprzak JD, Wejner-Mik P, Szymczyk E, Wdowiak-Okrojek K, Lipiec P. Artificial intelligence-powered measurement of left ventricular ejection fraction using a handheld ultrasound device. *Ultrasound Med Biol* 2021;**47**:1120–1125.
26. Otterstad JE, Froeland G, St John Sutton M, Holme I. Accuracy and reproducibility of biplane two-dimensional echocardiographic measurements of left ventricular dimensions and function. *Eur Heart J* 1997;**18**:507–513.
27. Baron T, Berglund L, Hedin EM, Flachskampf FA. Test-retest reliability of new and conventional echocardiographic parameters of left ventricular systolic function. *Clin Res Cardiol* 2019;**108**:355–365.
28. Houard L, Militaru S, Tanaka K, *et al.* Test-retest reliability of left and right ventricular systolic function by new and conventional echocardiographic and cardiac magnetic resonance parameters. *Eur Heart J Cardiovasc Imaging 2021;22:1157–1167.*
29. Thavendiranathan P, Grant AD, Negishi T, Plana JC, Popović ZB, Marwick TH. Reproducibility of echocardiographic techniques for sequential assessment of left ventricular ejection fraction and volumes: application to patients undergoing cancer chemotherapy. *J Am Coll Cardiol* 2013;**61**:77–84.
30. Tabib A, Samiei N, Peighambari MM, *et al.*; SHED LIGHT investigators. Pocket-sized echocardiography for screening structural heart disease: diagnostic accuracy and cost-effectiveness for population-based studies. *Cardiol Young* 2020;**30**:197–204.