Critical Care

BRIEF REPORT

Machine learning for the real-time assessment of left ventricular ejection fraction in critically ill patients: a bedside evaluation by novices and experts in echocardiography

Rita Varudo[1], Filipe A. Gonzalez[1,2], João Leote[1], Cristina Martins[1], Jacobo Bacariza[1], Antero Fernandes[1,2,3] and Frederic Michard[4*]

Abstract

Background: Machine learning algorithms have recently been developed to enable the automatic and real-time echocardiographic assessment of left ventricular ejection fraction (LVEF) and have not been evaluated in critically ill patients.

Methods: Real-time LVEF was prospectively measured in 95 ICU patients with a machine learning algorithm installed on a cart-based ultrasound system. Real-time measurements taken by novices ($LVEF_{Nov}$) and by experts ($LVEF_{Exp}$) were compared with LVEF reference measurements ($LVEF_{Ref}$) taken manually by echo experts.

Results: $LVEF_{Ref}$ ranged from 26 to 80% (mean $54 \pm 12\%$), and the reproducibility of measurements was $9 \pm 6\%$. Thirty patients (32%) had a $LVEF_{Ref} < 50\%$ (left ventricular systolic dysfunction). Real-time $LVEF_{Exp}$ and $LVEF_{Nov}$ measurements ranged from 31 to 68% (mean $54 \pm 10\%$) and from 28 to 70% (mean $54 \pm 9\%$), respectively. The reproducibility of measurements was comparable for $LVEF_{Exp}$ ($5 \pm 4\%$) and for $LVEF_{Nov}$ ($6 \pm 5\%$) and significantly better than for reference measurements ($p < 0.001$). We observed a strong relationship between $LVEF_{Ref}$ and both real-time $LVEF_{Exp}$ ($r = 0.86$, $p < 0.001$) and $LVEF_{Nov}$ ($r = 0.81$, $p < 0.001$). The average difference (bias) between real time and reference measurements was $0 \pm 6\%$ for $LVEF_{Exp}$ and $0 \pm 7\%$ for $LVEF_{Nov}$. The sensitivity to detect systolic dysfunction was 70% for real-time $LVEF_{Exp}$ and 73% for $LVEF_{Nov}$. The specificity to detect systolic dysfunction was 98% both for $LVEF_{Exp}$ and $LVEF_{Nov}$.

Conclusion: Machine learning-enabled real-time measurements of LVEF were strongly correlated with manual measurements obtained by experts. The accuracy of real-time LVEF measurements was excellent, and the precision was fair. The reproducibility of LVEF measurements was better with the machine learning system. The specificity to detect left ventricular dysfunction was excellent both for experts and for novices, whereas the sensitivity could be improved.

Trial registration: NCT05336448. Retrospectively registered on April 19, 2022.

Keywords: Echocardiography, Left ventricular ejection fraction, Artificial intelligence, Machine learning, Point of care ultrasound, Novice

Introduction

The assessment of left ventricular ejection fraction (LVEF) is part of the point of care echocardiographic evaluation of critically ill patients [1–3]. It has the

*Correspondence: frederic.michard@bluewin.ch

[4] MiCo, Vallamand, Switzerland
Full list of author information is available at the end of the article

Varudo *et al. Critical Care*   (2022) 26:386

Page 2 of 5

disadvantage of being time-consuming and operator dependent. Machine learning algorithms have recently been developed to facilitate, automate, and decrease the variability of echocardiographic measurements [4–7]. Several algorithms have been designed specifically for the real-time assessment of LVEF [8–10]. They have been trained to recognize specific ultrasound images, enable instantaneous image quality control, and measure LVEF automatically in just a few seconds. However, clinical validation studies remain scarce and have been done in ambulatory cardiac patients [8–10].

In critically ill patients, we compared real-time LVEF measurements taken with a new machine learning algorithm to reference manual measurements taken by experts in echocardiography.

Methods

We prospectively studied critically ill patients who required an echocardiographic evaluation during their ICU stay and in whom it was possible to obtain transthoracic images enabling a manual and quantitative evaluation of left ventricular systolic function. Real-time LVEF measurements were taken with a machine learning algorithm (Real-Time EF, GE Healthcare, Chicago, USA) installed on a cart-based ultrasound system (Venue, GE Healthcare). The real-time LVEF software is a neural network algorithm which has been trained with thousands of cardiac images to automatically detect the 4-chamber view of the heart, locate landmarks on the left ventricular wall and detect end-diastolic and end-systolic times from the mitral valve motion. Once the endocardial border is detected, the algorithm provides immediate user feedback regarding image quality using color-coding. When image quality is considered acceptable (green or yellow endocardial border displayed on screen), left ventricular volumes are automatically estimated from the single-plane Simpson disk method, enabling LVEF calculation from real-time end-diastolic and end-systolic volumes.

Real-time LVEF measurements obtained by a novice ($LVEF_{Nov}$) and by an expert ($LVEF_{Exp}$) were compared with LVEF measurements taken manually by an expert in critical care echocardiography ($LVEF_{Ref}$). Seven novices (all residents in our department and beginners in echocardiography) and two experts (senior intensivists with the European Diploma in Advanced Critical Care Echocardiography) participated in data collection. Measurements taken in triplicate were averaged for comparisons, and the intra-operator reproducibility was assessed by calculating the coefficient of variation (standard deviation divided by the mean) expressed as a percentage.

The quality of echo images was classified as good, fair, or poor by the experts, and as green (optimal), yellow (acceptable), or red (not acceptable for real-time LVEF measurements) by the machine learning algorithm.

Results are expressed as mean ± standard deviation (SD). Agreement between real-time and reference LVEF measurements was tested using the Bland–Altman method. Statistical comparisons were made with a *t*-test. A *p* value < 0.05 was considered statistically significant.

Results

We prospectively enrolled 95 patients (mean age 60 ± 17 yr) over a 9-month period. Most patients were admitted for medical reasons and 32 (34%) were mechanically ventilated at the time of the ultrasound evaluation (Additional file 1: Table S1). Reference LVEF ranged from 26 to 80% (mean 54 ± 12%) and the reproducibility of manual measurements was 9 ± 6%. Thirty patients (32%) had a $LVEF_{Ref} < 50\%$ (left ventricular systolic dysfunction).

Real-time $LVEF_{Exp}$ ranged from 31 to 68% (mean 54 ± 10%). We observed a strong relationship ($r = 0.86$, $p < 0.001$) between reference and real-time $LVEF_{Exp}$ (Fig. 1). The average difference (bias) between real-time $LVEF_{Exp}$ and reference LVEF was 0 ± 6% with 95% limits of agreement of −12 to +11% (Fig. 1). The intra-operator reproducibility of measurements was better for real-time $LVEF_{Exp}$ than for reference manual measurements (5 ± 4% vs. 9 ± 6%, $p < 0.001$). The sensitivity and specificity of real-time $LVEF_{Exp}$ to detect systolic dysfunction were 70% and 98%, respectively.

Real-time $LVEF_{Nov}$ ranged from 28 to 70% (mean 54 ± 9%). We observed a strong relationship ($r = 0.81$, $p < 0.001$) between $LVEF_{Ref}$ and real-time $LVEF_{Nov}$ (Fig. 1). The average difference (bias) between real-time $LVEF_{Nov}$ and $LVEF_{Ref}$ was 0 ± 7% with 95% limits of agreement of −14 to +13% (Fig. 1). The intra-operator reproducibility of measurements was better for real-time $LVEF_{Nov}$ than for reference manual measurements (6 ± 5% vs. 9 ± 6%, $p < 0.001$). The sensitivity and specificity of real-time $LVEF_{Nov}$ to detect systolic dysfunction were 73% and 98%, respectively.

According to experts' judgement, the quality of echo images was good, fair, and poor in 41, 43, and 11 patients, respectively. The average difference (bias) between real-time and reference LVEF measurements was comparable when images were of good quality ($n = 41$) and of fair or poor quality ($n = 54$), both for experts and novices (Table 1). And results did not change significantly after excluding the 11 patients with poor image quality (Table 1).

According to the machine learning algorithm, the quality of echo images was green, yellow, and red flagged in 80, 15 and 0 patients, respectively. Results did not change significantly after excluding the 15 patients in whom images were non-optimal/yellow flagged (Table 1).
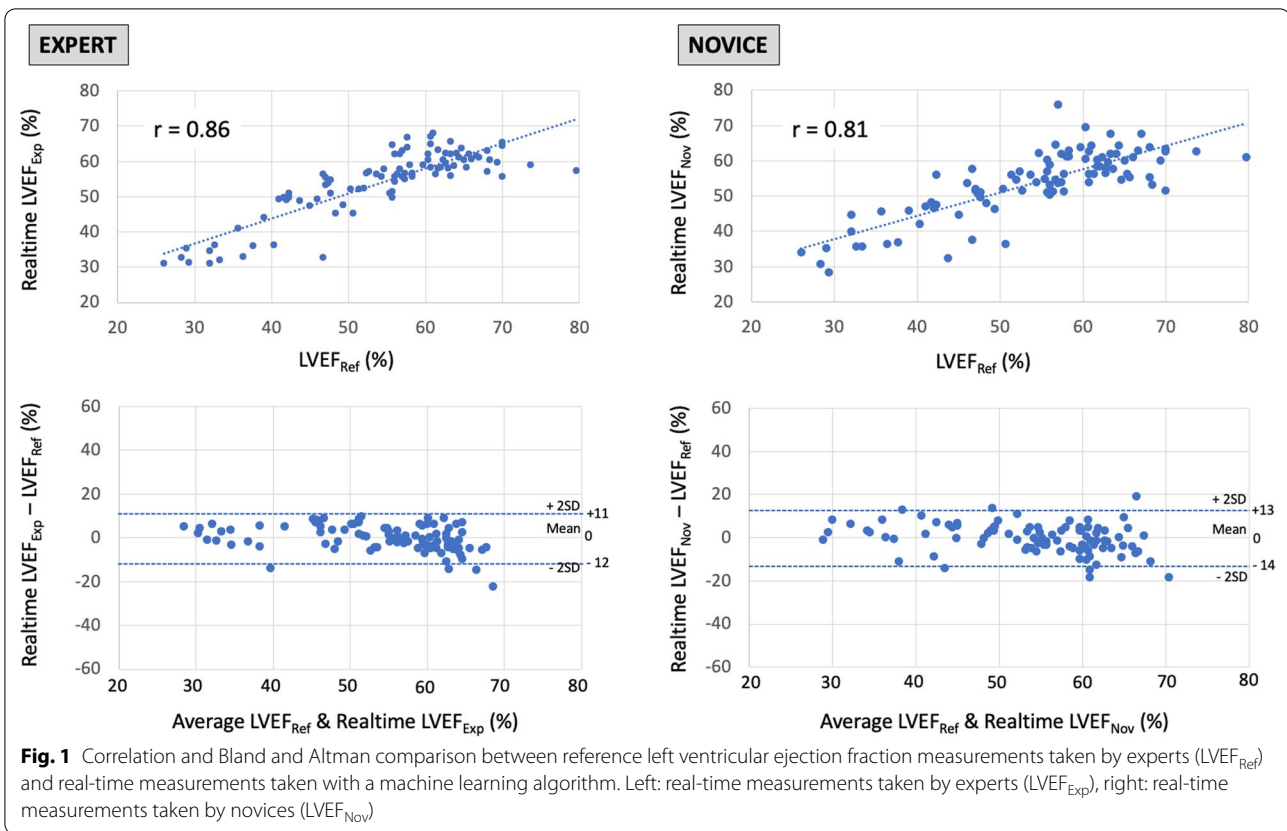
Varudo *et al. Critical Care* (2022) 26:386

Page 3 of 5

Fig. 1 Correlation and Bland and Altman comparison between reference left ventricular ejection fraction measurements taken by experts (LVEF$_{Ref}$) and real-time measurements taken with a machine learning algorithm. Left: real-time measurements taken by experts (LVEF$_{Exp}$), right: real-time measurements taken by novices (LVEF$_{Nov}$)

Table 1 Main results in subgroups based on image quality

Image quality	Good	Good	Good and fair	Good and fair	Fair and poor	Fair and poor	Green-flagged	Green-flagged
Measurement	Manual	Manual	Manual	Manual	Manual	Manual	ML	ML
Operator	Expert	Novice	Expert	Novice	Expert	Novice	Expert	Novice
Patients, n	41	41	84	84	54	54	80	80
Reproducibility	5±4%	6±4%	5±4%	6±5%	5±4%	5±5%	5±4%	6±5%
Correlation coef.	r=0.86	r=0.78	r=0.85	r=0.81	r=0.86	r=0.83	r=0.86	r=0.82
Bias (mean±SD)	0±6%	0±7%	0±6%	0±7%	−1±6%	−1±6%	0±6%	0±7%
Sensitivity	–	–	65%	73%	–	–	65%	69%
Specificity	–	–	100%	100%	–	–	100%	100%

ML = Machine learning, coef. = coefficient. Sensitivity and specificity were not calculated for small subgroups

The average difference (bias) between real-time and reference LVEF measurements was slightly higher in mechanically ventilated ($n = 32$) than in non-mechanically ventilated patients ($n = 63$), both for experts ($-2 \pm 7\%$ vs. $0 \pm 5\%$) and novices ($-1 \pm 8\%$ vs. $0 \pm 6\%$). However, observed differences did not reach statistical significance.

Discussion

An increasing number of anesthesiologists and intensivists have been trained to perform qualitative echocardiographic assessments [1–3]. However, quantitative evaluations remain challenging for many, particularly for novices. In the present study, we tested an artificial intelligence-enabled tool specifically designed to facilitate and

Varudo *et al. Critical Care*   (2022) 26:386

Page 4 of 5

automatize the bedside measurements of LVEF. Our findings suggest that this tool enables a clinically acceptable estimation of LVEF when compared to manual measurements. They also suggest that the real-time LVEF tool enables novices to assess LVEF with a better reproducibility than what experts can achieve manually.

Several machine learning algorithms have been designed to assess LVEF from a parasternal long axis view or from an apical 2 or 4-chamber view [8–10]. Comparison studies published so far yielded promising results. Indeed, close correlations and good agreements have been reported between LVEF measurements taken by skilled operators and by machine learning algorithms, particularly when the algorithm detects and analyze the apical 4-chamber view [9, 10]. However, clinical validation studies remain scarce and have been done in ambulatory cardiac patients. Our study appears to be the first evaluation done in critically ill patients in whom transthoracic echocardiography is often challenging, in particular when patients are mechanically ventilated. Our findings suggest that the real-time LVEF algorithm may help clinicians, including beginners in echocardiography, to accurately measure LVEF in just a few seconds. Such a tool may contribute to further increase the adoption of point of care echocardiographic evaluations in critically ill patients.

Our study has limitations. Because ultrasound evaluations are time-consuming, we studied hemodynamically stable patients to ensure comparability between measurements taken at each step of the evaluation (LVEF measurements were first taken by a trainee, then by an expert both manually and with the automatic method). Also, we did not assess the ability of the new real-time LVEF method to track changes in LVEF. A small number of patients had a severely impaired left ventricular systolic function ($LVEF_{Ref} < 30\%$, $n = 4$) or a hyperkinetic ventricle ($LVEF_{Ref} > 70\%$, $n = 2$). Therefore, future studies will need to assess the clinical value of the real-time LVEF algorithm during hemodynamic instability, in patients with a very low or supranormal LVEF, and during therapeutic interventions (e.g., inotropic stimulation) known to induce significant changes in systolic function.

Conclusion

Machine learning-enabled real-time measurements of LVEF were strongly correlated with manual measurements obtained by experts. The accuracy of real-time LVEF measurements was excellent, and the precision was fair. The reproducibility of LVEF measurements was better with the machine learning system, including for novices. The specificity to detect left ventricular systolic dysfunction was excellent both for experts and novices, whereas the sensitivity could be improved. Studies are needed to confirm our findings in mechanically ventilated patients with cardiogenic shock or hyperdynamic states.

Abbreviations
LVEF: Left ventricular ejection fraction; $LVEF_{Ref}$: Reference left ventricular ejection fraction; $LVEF_{Exp}$: Real-time left ventricular ejection fraction measured by experts; $LVEF_{Nov}$: Real-time left ventricular ejection fraction measured by novices.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13054-022-04269-6.

Additional file 1: Table S1 Main characteristics of the study population.

Author contributions
FAG, AF and FM designed the study. FAG, JB and RV, JL, CM were involved in study supervision and data collection as experts and trainees, respectively. FAG and FM were involved in data analysis and interpretation. FM drafted the manuscript, and all authors revised the manuscript and approved the final version.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The study has been conducted in accordance with the Declaration of Helsinki principles and was approved by the ethical committee of hospital Garcia de Orta, Almada (# TI 71/2021) on July 27, 2021, and written informed consent was obtained for all patients.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Intensive Care Department, Hospital Garcia de Orta, Almada, Portugal. [2]Faculdade de Medicina da Universidade de Lisboa, Lisbon, Portugal. [3]Faculdade de Ciencias da Saude da Universidade da Beira Interior, Covilha, Portugal. [4]MiCo, Vallamand, Switzerland.

References
1. Orde S, Slama M, Hilton A, et al. Pearls and pitfalls in comprehensive critical care echocardiography. Crit Care. 2017;21:279.
2. Vieillard-Baron A, Millington SJ, Sanfilippo F, et al. A decade of progress in critical care echocardiography: a narrative review. Intensive Care Med. 2019;45:770–88.

Varudo *et al. Critical Care*   (2022) 26:386

Page 5 of 5

3. Marbach JA, Almufleh A, Di Santo P, et al. A shifting paradigm: the role of focused cardiac ultrasound in bedside patient assessment. Chest. 2020;158:2107–18.
4. Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging. J Am Coll Cardiol. 2019;73:1317–35.
5. Narang A, Bae R, Hong H, et al. Using a deep-learning algorithm to guide novice to acquire echocardiograms for limited diagnostic use. JAMA Cardiol. 2021;6:624–32.
6. Nabi W, Bansal A, Xu B. Applications of artificial intelligence and machine learning approaches in echocardiography. Echocardiography. 2021;38:982–92.
7. Gonzalez FA, Varudo R, Leote J, et al. The automation of sub-aortic velocity time integral measurements by transthoracic echocardiography: clinical evaluation of an artificial intelligence-enabled tool in critically ill patients. Br J Anaesth. 2022;129:e116–9. https://doi.org/10.1016/j.bja.2022.07.037.
8. 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.
9. Schneider M, Bartko P, Geller W, et al. A machine-learning algorithm supports ultrasound-naïve novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. Int J Cardiovasc Imaging. 2021;37:577–86.
10. Asch FM, Mor-Avi V, Rubenson D, et al. Deep-learning based automated echocardiographic quantification of left ventricular ejection fraction: a point-of-care solution. Circ Cardiovasc Imaging. 2021;14:e012293.

Publisher's Note