openheart

Artificial intelligence-assisted interpretation of systolic function by echocardiogram

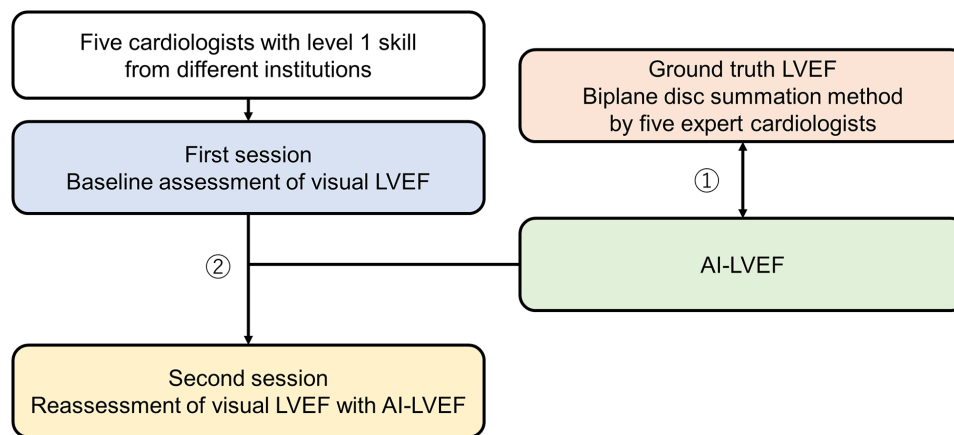Natsumi Yamaguchi,[1] Yoshitaka Kosaka,[1] Akihiko Haga,[2] Masataka Sata,[1] Kenya Kusunose [ORCID] [3]

[1]Department of Cardiovascular Medicine, Tokushima University Hospital, Tokushima, Japan
[2]Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan
[3]Department of Cardiovascular Medicine, Nephrology, and Neurology, University of the Ryukyus, Okinawa, Japan

Correspondence to
Dr Kenya Kusunose; kusunok@med.u-ryukyu.ac.jp

ABSTRACT

Objective Precise and reliable echocardiographic assessment of left ventricular ejection fraction (LVEF) is needed for clinical decision-making. Recently, artificial intelligence (AI) models have been developed to estimate LVEF accurately. The aim of this study was to evaluate whether an AI model could estimate an expert read of LVEF and reduce the interinstitutional variability of level 1 readers with the AI-LVEF displayed on the echocardiographic screen.

Methods This prospective, multicentre echocardiographic study was conducted by five cardiologists of level 1 echocardiographic skill (minimum level of competency to interpret images) from different hospitals. Protocol 1: Visual LVEFs for the 48 cases were measured without input from the AI-LVEF. Protocol 2: the 48 cases were again shown to all readers with inclusion of AI-LVEF data. To assess the concordance and accuracy with or without AI-LVEF, each visual LVEF measurement was compared with an average of the estimates by five expert readers as a reference.

Results A good correlation was found between AI-LVEF and reference LVEF (r=0.90, p<0.001) from the expert readers. For the classification LVEF, the area under the curve was 0.95 on heart failure with preserved EF and 0.96 on heart failure reduced EF. For the precision, the SD was reduced from 6.1±2.3 to 2.5±0.9 (p<0.001) with AI-LVEF. For the accuracy, the root-mean squared error was improved from 7.5±3.1 to 5.6±3.2 (p=0.004) with AI-LVEF.

Conclusions AI can assist with the interpretation of systolic function on an echocardiogram for level 1 readers from different institutions.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Precise and reliable echocardiographic assessment of left ventricular ejection fraction (LVEF) is needed for clinical decision-making.

WHAT THIS STUDY ADDS

⇒ Assessment of LVEF using the artificial intelligence (AI) algorithm is an objective method with no intraobserver error, and its accuracy was equal to that of assessment by expert reader consensus. Moreover, AI algorithms can reduce interobserver and intraobserver variability.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study is a part of a broader paradigm shift in echocardiography, which can possibly augment or replace experts' tasks.

Left ventricular ejection fraction (LVEF) is widely used and is an important parameter to assess LV systolic function, as well as to help guide the management of various cardiac diseases, including heart failure (HF).[1] Precise and reliable echocardiographic assessment of LVEF is required for clinical decision-making. Echocardiographic guidelines recommend that EF should be assessed by the biplane method of disks, and then the measurement should be confirmed by visual estimation.[2 3] Alternatively, a visual estimation of LVEF is widely used to confirm the quantitative EF values, particularly in emergency department settings. The visual estimation of LVEF is an important component to determine LV function in all institutions. The visual assessment is subjective, and variability can be influenced by reader experience. Several institutions have readers with various experience levels, and because there is a large variability in LVEF measurements within different centres, therapies may be confounded when the decisions are made on the basis of LVEF.[4 5] An effective method to reduce variability in LVEF assessment is needed.[6–8]

Artificial intelligence (AI) has been developed as state-of-the-art applications for the detection and classification of diseases in various medical fields.[9–14] It has been shown to be a useful tool for assessing cardiovascular diseases.[15–18] Recently, we reported that an AI model based on echocardiographic images can predict LVEF in patients with HF.[19] The estimated LVEF by AI (AI-LVEF) may be a reliable and precise method to use in a clinical setting. However, the optimal way to integrate AI into the clinical process is

Figure 1 Study design flowchart: Protocol 1: the correlation between AI-LVEF and the ground truth LVEF. Protocol 2: the AI-assisted accuracy and precision for level 1 cardiologists. AI, artificial intelligence; LVEF, left ventricular ejection fraction.

under debate. We hypothesised that AI-LVEF would help reduce the interinstitutional variability of level 1 readers and improve performance to that of an expert reader. This study aimed to evaluate if the AI model could estimate LVEF similar to expert readers, and reduce the interinstitutional variability among level 1 readers with the AI-LVEF displayed on the echocardiographic screen.

METHODS
Design
We designed a prospective, multicentre echocardiographic study with first and second session assessments and analyses (figure 1). A total of five cardiologists with level 1 echocardiographic skills as defined by a training statement[20] from different tertiary care centres participated Kurashiki Central Hospital. Japan Red Cross Wakayama Medical Center, Japan Red Cross Society Tokushima Hospital, HITO Medical Center and Hyogo Prefectural Amagasaki General Medical Center. All participants were blinded to readers' interpretations. The echocardiography was performed using a commercially available ultrasound machine (Aplio i900; Canon Medical Systems, Odawara, Japan). All echocardiographic measurements were obtained according to the American Society of Echocardiography recommendations.[18] The apical 2-chamber (AP2), apical 4-chamber (AP4), apical 3-chamber (AP3), parasternal long axis (PLAX) and parasternal short axis (PSAX) views were stored digitally for playback and analysis.

We prospectively enrolled 48 patients who were diagnosed with HF. To overcome the small size of the dataset, we sampled the patients so that their EF was homogeneously distributed over the full EF range (from 10% to 80%). In this cohort, 4 patients (8%) had LVEF=10%–20%, 5 patients (10%) had LVEF=21%–30%, 10 patients (21%) had LVEF=31%–40%, 10 patients (21%) had LVEF=41%–50%, 8 patients (17%) had LVEF=51%–60% and 11 patients (23%) had LVEF=over 60%. No patients had atrial fibrillation or severe valvular disease.

All selected images had good or adequate acoustic quality based on the visualisation of the LV walls and endocardium.

Level 1 skill in echocardiography refers to the minimum level of competency required to perform and interpret basic echocardiographic examinations for diagnostic purposes. This level of competency is achieved after completing a dedicated period of training, typically lasting 3 months. During this training period, the trainee is expected to develop a thorough understanding of functional anatomy and physiology in relation to the echocardiographic examination. In addition to theoretical training, the trainee is required to participate in the interpretation of a minimum of 150 complete echocardiographic examinations, including M-mode, 2D and Doppler studies.[20]

Protocol 1
To assess the accuracy of AI-estimated LVEF for each case, we compared the AI-LVEF to the reference LVEF value, which was calculated from an average of the assessments by the five expert readers as a ground truth. We used our previously developed AI model to estimate LVEF for this study.[19] To obtain the reference LVEF values, all studies were independently analysed by five expert readers with more than 10 years' experience with echocardiography as well as certification as Registered Medical Sonographers or Board Certificated Fellows by The Japan Society of Ultrasonics in Medicine. LVEF was calculated by the biplane method of disks using the AP2 and AP4 views, and then the measurement was confirmed on the other echocardiographic views (AP3, PLAX and PSAX). The HF with reduced EF (HFrEF) was defined as the clinical diagnosis of HF with LVEF<50%, whereas an HF with preserved EF (HFpEF) was the clinical diagnosis of HF with LVEF≥50%, as based on the current American Society of Echocardiography (ASE)/European Association of Cardiovascular Imaging (EACVI) guidelines.[21]

Protocol 2

To assess the interinstitutional variability, LVEFs for the 48 cases were assessed visually by 5 readers with level 1 echocardiographic skill from different tertiary care centres. Readers were required to provide visual estimates of LVEF as single integers and were blinded to other readers' interpretations. To avoid bias, no clinical data about the cases were provided. All data were collected on an answer sheet with each case coded separately.

To assess changes in the variability of LVEF, all 48 cases were shown to the same readers with the AI-LVEF displayed on the echocardiographic screen at 1 month after the end of the first reading session. The display of AI-LVEF was automatically generated through the execution of the AI algorithm within the prototype echocardiographic software in the commercially available machine, which was executing in the background, by acquiring and analysing five cross-sectional images (AP2, AP4, AP3, PLAX and PSAX) from one cardiac cycle. For each case, the individual visual estimates of LVEF were again compared with the reference values. The changes in accuracy and variability after the second session were assessed.

Statistical analysis

The data were presented as mean±SD if the Kolmogorov-Smirnov test showed a normal distribution. Otherwise, the median and interquartile ranges were calculated. We used Pearson's correlation coefficients. A Bland-Altman analysis was used to determine the bias and 95% limits of agreement (LOA) between the AI-LVEF and the reference LVEF values. The diagnostic performance of the AI algorithm was evaluated using receiver operating characteristic (ROC) analysis and pairwise comparisons of the area under the ROC curve (AUC) according to the DeLong method.[22] An SD was calculated to assess the variability of the LVEF assessment among readers. A root mean square error (RMSE) calculation was performed to assess the accuracy of LVEF by the five readers after the AI assistance. The statistical analysis was performed using standard statistical software packages (SPSS software V.21.0 and MedCalc Software V.18; Mariakerke, Belgium). The threshold for statistical significance was set to $p<0.05$.

Patient and public involvement

This study did not involve direct patient participation.

RESULTS

The subject demographics for this study are shown in table 1. In this cohort, 50% of the patients had ischaemic cardiomyopathy.

Estimation of LVEF by the AI model

The comparison between AI-LVEF estimates and the reference LVEF values in this cohort is shown in figure 2. An excellent correlation was found between AI-LVEF and reference LVEF values ($r=0.91$, $p<0.001$). A comparison

Table 1 Baseline demographics of the study participants

	Test cohort
Number	48
Age, years	66±12
Male, %	75
Ischaemic cardiomyopathy, %	50
Heart rate, bpm	71±15
Blood pressure, mm Hg	125±24
LVEDVi, mL/m^2	73 (48–109)
LVESVi, mL/m^2	46 (25–62)
WMSI	1.6±0.5
LVEF, %	45 (35–58)

Data are presented as number of patients (percentage), mean±SD or median (IQR).
LVEDVi, Left Ventricular End Diastolic Volume Index; LVEF, left ventricular ejection fraction; LVESVi, Left Ventricular End Systolic Volume Index; WMSI, Wall Motion Score Index.
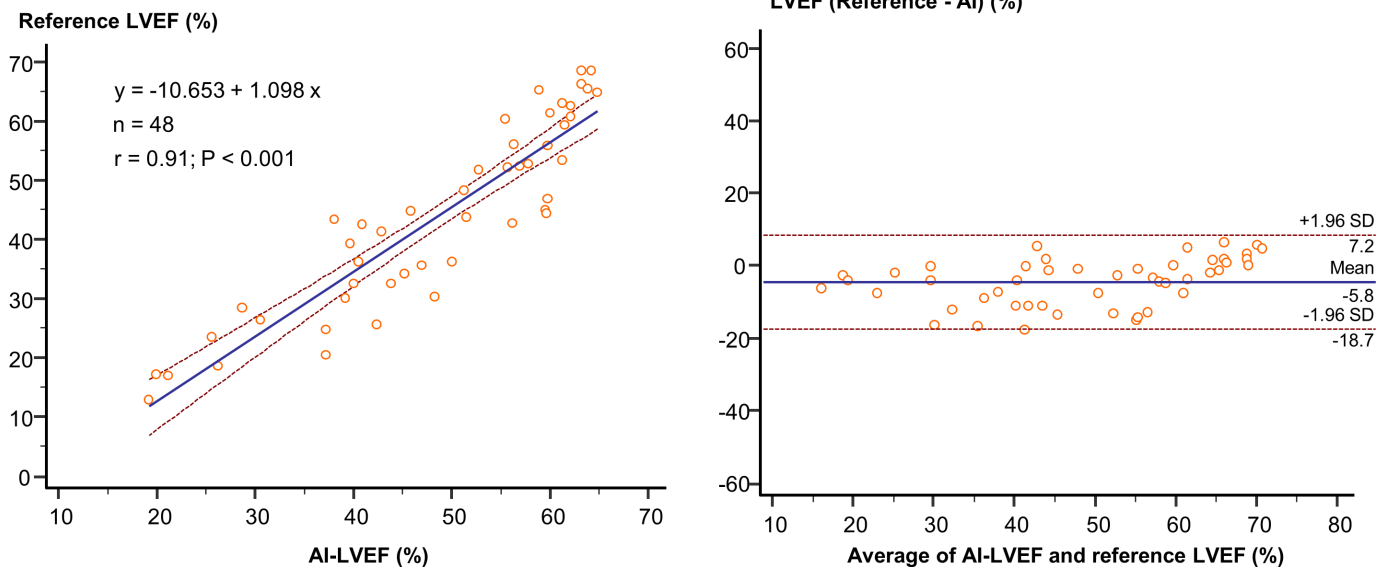
between AI-LVEF and reference LVEF by a Bland-Altman analysis showed a mean difference of −5.8, with an LOA of ±12.9. The results of the ROC analysis used to assess the diagnostic ability for the classification of HF types are shown in figure 3. For the classification of HF types based on LVEF, we assessed the AUCs. The AUCs by the AI were 0.96 for both HFpEF and HFrEF.

The reliability and accuracy after AI processing

Figure 4 shows the reliability and accuracy of LVEF as assessed by five level 1 echocardiographer readers from the first and second sessions. For the first session, the SD for the reliability of LVEF by the five readers was 6.1±2.3, and the RSME for the accuracy was 7.5±3.1. With the AI-LVEF included for assistance with the read, the SD and RSME were significantly improved in the second session. The SD improved from 6.1±2.3 to 2.5±0.9, ($p<0.001$), and the RSME improved from 7.5±3.1 to 5.6±3.2, ($p=0.004$). These results indicate that displaying the assessment by AI-LVEF on the screen improved the concordance of level 1 readers from different institutions. Interestingly, the SD of LVEF assessed by the five expert readers was 3.1±1.4 and similar to that by level 1 readers in the second session.

DISCUSSION

The LVEF estimate is key for HF management in a clinical setting. However, the measurement of LVEF is time-consuming with high interobserver and intraobserver variability.[23] The AI algorithm is an objective method with no intraobserver error, and its accuracy is similar to the assessment by expert readers. Importantly, we showed that the variability of the assessment by five readers of level 1 skill with AI assistance was similar to that by expert reader assessment. This diagnostic system may be a useful tool to estimate LVEF and classify HF for clinical evaluation.

Figure 2 The correlation between AI-LVEF and reference LVEF: an excellent correlation was found between AI-LVEF and reference LVEF (r=0.91, p<0.001). AI, artificial intelligence; LVEF, left ventricular ejection fraction.

Deep learning in echocardiography

The use of quantitative assessment is thought to improve the accuracy and objectivity of echocardiography. Recently, several groups have developed automated algorithms for the analysis of left ventricular function and endocardial border detection.[24 25] However, most methods remain semiautomatic where observer input is initially needed to manually annotate important landmarks (eg, mitral plane, apex). A fully automated assessment is needed to obtain quantitative results without any user interaction including marker positioning, contour

drawings and modification. Our results demonstrate that a 3D-convolutional neural network can be trained to estimate LVEF on echocardiographic images. We believe this study supports the use of AI algorithms for echocardiographic images in future applications.

Improvement of visual LVEF

Several papers have reported the use of quality assessment programmes in clinical settings.[26 27] The investigators used a learning session for reference LVEF to reduce inter-reader variability. Reference LVEFs were provided
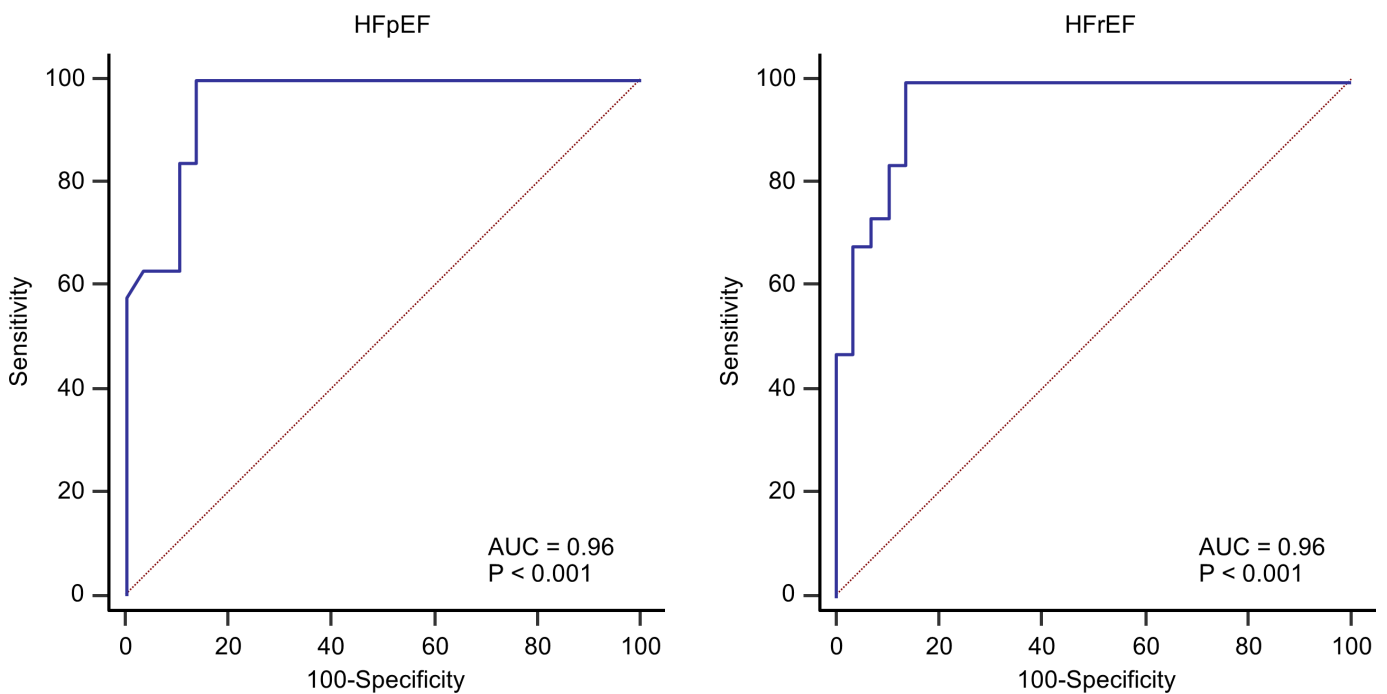
Figure 3 The diagnostic ability for the classification of HF subtype (HFpEF or HFrEF): the areas under the curve were good (both AUCs: 0.96). AUC, area under the curve; HF, heart failure; HFpEF, HF with preserved ejection fraction; HFrEF, HF with reduced ejection fraction.
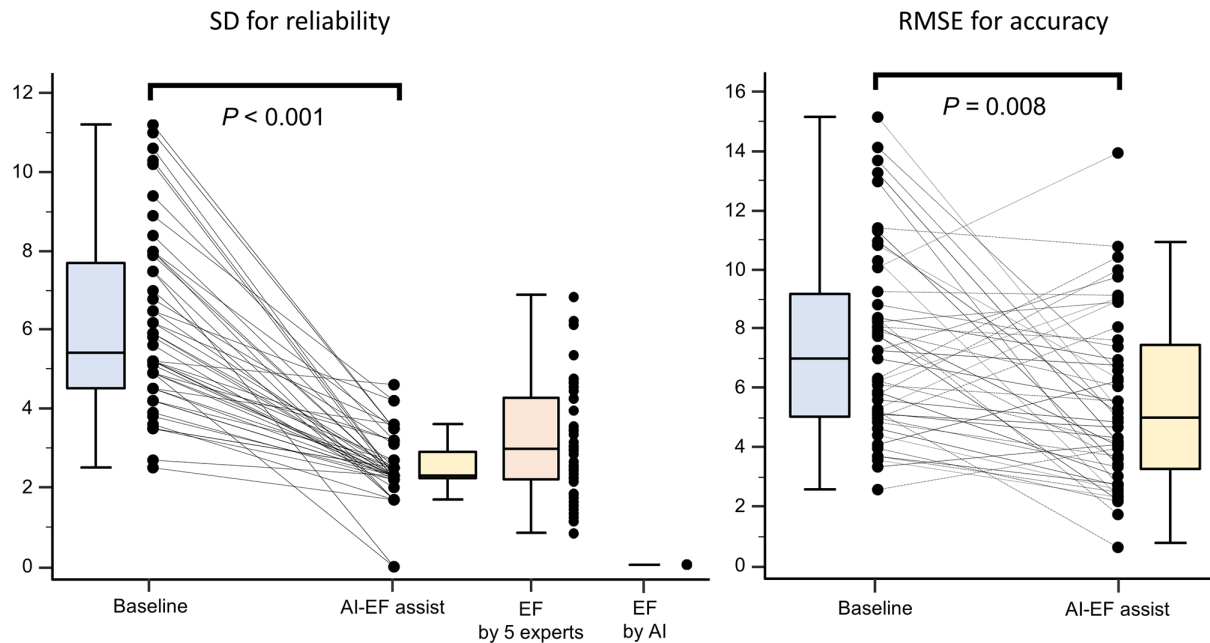
Figure 4 An improvement of the precise and accuracy using AI-assisted LVEF: the SD and RMSE were improved by AI-LVEF assistance. AI, artificial intelligence; LVEF, left ventricular ejection fraction; RMSE, root mean square error.

by radionuclide imaging, cardiac MRI and/or echocardiographic expert reads. In these previous studies, the learning session using reference images improved the reproducibility of visually estimated LVEF.[26 28 29] However, these methods require a certain amount of training time and have not been generalised to broader applications.

Our paper is the first study to demonstrate the utility of AI-LVEF estimates displayed on the echocardiographic screen. The reliability and accuracy of the LVEF estimation by level 1 readers was improved by showing AI-LVEF estimates on the screen during the read. In our prospective study, the variability in level 1 reader assessments with AI assistance was equal to or less than the variability of expert reader assessments. The AI-assisted LVEF assessment may be useful to standardise the actual read of visually estimated LVEF. We believe that this is a unique and important contribution to the field, as it addresses a practical issue in the clinical setting that has not been thoroughly investigated before.

Clinical implications
The measurement of LVEF with echocardiography is observer-dependent and requires experience.[30] The assessment of LVEF using the AI algorithm is an objective method with no intraobserver error, and its accuracy was equal to that of assessment by expert reader consensus. Moreover, AI algorithms can reduce interobserver and intraobserver variability. This study is a part of a broader paradigm shift in echocardiography, which can augment or replace experts' tasks. Combined with the development of handheld echocardiographic devices, AI software support for echocardiographic interpretation may increase access to cardiac imaging in settings where clinical expertise and resources are lacking. As echocardiography is a commonly used diagnostic tool in clinical practice, the ability of an AI model to improve the accuracy and precision of LVEF measurements has the potential to improve patient outcomes by enabling more informed clinical decision-making especially in HF.

Limitations
First, the LVEF assessment is based on echocardiographic results by expert readers' assessment as a ground truth. Second, echocardiographic images do not consist of structured data and cannot be reconfigured. Thus, the accuracy of diagnosis may be influenced by the image quality. Third, we included only patients with HF in this study. We may be unable to apply this algorithm in patients without HF. Fourth, we did not apply the AI algorithms to estimate LV volumes, and applied the AI algorithms to directly estimate LVEF, since a deviation in the volume estimation can influence the estimation of LVEF. Finally, there might be an anchoring bias in this study. Awareness of a preliminary assessment influenced clinicians to be inclined towards that LVEF assessment. To mitigate this bias, the protocol was not disclosed to the evaluators and the evaluation was carried out on the actual machine in a manner as close to daily operations as possible.

CONCLUSIONS
AI can assist in the interpretation of systolic function on echocardiograms by level 1 readers from different institutions. These results represent an important improvement for the assessment of LVEF in HF, and highlight the possibility of AI to provide assistance for the interpretation of echocardiograms, which can support clinicians and augment clinical care.

ORCID iD
Kenya Kusunose http://orcid.org/0000-0002-4909-754X

REFERENCES

1. Lyle MA, Brozovich FV. Hfpef, a disease of the vasculature: a closer look at the other half. *Mayo Clin Proc* 2018;93:1305–14.
2. 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;16:233–70.
3. Enomoto M, Ishizu T, Seo Y, *et al*. Myocardial dysfunction identified by three-dimensional speckle tracking echocardiography in type 2 diabetes patients relates to complications of microangiopathy. *J Cardiol* 2016;68:282–7.
4. 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;29:581–8.
5. Okuma H, Noto N, Tanikawa S, *et al*. Impact of persistent left ventricular regional wall motion abnormalities in childhood cancer survivors after anthracycline therapy: assessment of global left ventricular myocardial performance by 3d speckle-tracking echocardiography. *J Cardiol* 2017;70:396–401.
6. Gottdiener JS, Bednarz J, Devereux R, *et al*. American society of echocardiography recommendations for use of echocardiography in clinical trials. *J Am Soc Echocardiogr* 2004;17:1086–119.
7. Bhattacharyya S, Lloyd G. Improving appropriateness and quality in cardiovascular imaging: a review of the evidence. *Circ Cardiovasc Imaging* 2015;8:e003988.
8. Kusunose K, Shibayama K, Iwano H, *et al*. Reduced variability of visual left ventricular ejection fraction assessment with reference images: the Japanese Association of young echocardiography fellows multicenter study. *J Cardiol* 2018;72:74–80.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
10. Esteva A, Kuprel B, Novoa RA, *et al*. Corrigendum: dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;546:686.
11. Gulshan V, Peng L, Coram M, *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402–10.
12. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017;19:221–48.
13. Suzuki K. Machine learning in medical imaging before and after introduction of deep learning. *Medical Imaging and Information Sciences* 2017;34:14–24.
14. Kida S, Nakamoto T, Nakano M, *et al*. Cone beam computed tomography image quality improvement using a deep convolutional neural network. *Cureus* 2018;10:e2548.
15. Morita SX, Kusunose K, Haga A, *et al*. Deep learning analysis of echocardiographic images to predict positive genotype in patients with hypertrophic cardiomyopathy. *Front Cardiovasc Med* 2021;8:669860.
16. Hirata Y, Kusunose K, Tsuji T, *et al*. Deep learning for detection of elevated pulmonary artery wedge pressure using standard chest X-ray. *Can J Cardiol* 2021;37:1198–206.
17. Kusunose K, Hirata Y, Yamaguchi N, *et al*. Deep learning for detection of exercise-induced pulmonary hypertension using chest X-ray images. *Front Cardiovasc Med* 2022;9:891703.
18. Kusunose K, Abe T, Haga A, *et al*. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC Cardiovasc Imaging* 2020;13(2 Pt 1):374–81.
19. Kusunose K, Haga A, Yamaguchi N, *et al*. Deep learning for assessment of left ventricular ejection fraction from echocardiographic images. *J Am Soc Echocardiogr* 2020;33:632–5.
20. Ryan T, Berlacher K, Lindner JR, *et al*. COCATS 4 task force 5: training in echocardiography. *J Am Coll Cardiol* 2015;65:1786–99.
21. 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. *J Am Soc Echocardiogr* 2015;28:1–39.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
23. Dorosz JL, Lezotte DC, Weitzenkamp DA, *et al*. Performance of 3-dimensional echocardiography in measuring left ventricular volumes and ejection fraction: a systematic review and meta-analysis. *J Am Coll Cardiol* 2012;59:1799–808.
24. Leung KYE, Bosch JG. Automated border detection in three-dimensional echocardiography: principles and promises. *Eur J Echocardiogr* 2010;11:97–108.
25. Yang L, Georgescu B, Zheng Y, *et al*. A fast and accurate tracking algorithm of left ventricles in 3d echocardiography. Proceedings/IEEE International Symposium on Biomedical Imaging: from nano to macro IEEE International Symposium on Biomedical Imaging: NIH Public Access; 2008:221
26. Johri AM, Picard MH, Newell J, *et al*. Can a teaching intervention reduce Interobserver variability in LVEF assessment? *JACC: Cardiovascular Imaging* 2011;4:821–9.
27. Frikha Z, Girerd N, Huttin O, *et al*. Reproducibility in echocardiographic assessment of diastolic function in a population based study (the STANISLAS cohort study). *PLoS ONE* 2015;10:e0122336.
28. Akinboboye O, Sumner J, Gopal A, *et al*. Visual estimation of ejection fraction by two-dimensional echocardiography: the learning curve. *Clin Cardiol* 1995;18:726–9.
29. Thavendiranathan P, Popović ZB, Flamm SD, *et al*. Improved Interobserver variability and accuracy of echocardiographic visual left ventricular ejection fraction assessment through a self-directed learning program using cardiac magnetic resonance images. *J Am Soc Echocardiogr* 2013;26:1267–73.
30. Torii Y, Kusunose K, Yamada H, *et al*. Updated left ventricular diastolic function recommendations and cardiovascular events in patients with heart failure hospitalization. *J Am Soc Echocardiogr* 2019;32:1286–97.