Title Page

Title:
Cardiac Measurement Calculation on Point-of-Care Ultrasonography with Artificial Intelligence

Authors:
Sarah F Mercaldo, PhD[1,2,3]
Bernardo C Bizzo, MD PhD[1,2,3]
Tsion Sadore, MD[1]
Madeleine A Halle, BSc[1]
Ashley L MacDonald, BSc[1]
Isabella Newbury-Chaet, BSc[1]
Eric L'Italien[1]
Alex S Schultz[1]
Victor Tam[1]
Sheila M Hegde, MD[3,4]
Judy R Mangion, MD[3,4]
Praveen Mehrotra, MD[5]
Qiong Zhao, MD PhD[6]
Justina Wu, MD PhD*[3,4]
James M Hillis, MBBS DPhil*[1,3,7]
* These authors contributed equally to the work and share senior authorship.

Author Affiliations:
[1]Digital Clinical Research Organization, Mass General Brigham AI, Boston, MA, USA
[2]Department of Radiology, Massachusetts General Hospital, Boston, MA, USA
[3]Harvard Medical School, Boston, MA, USA
[4]Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
[5]Division of Cardiology, Sidney Kimmel Medical College, Thomas Jefferson University, PA
[6]Inova Schar Heart and Vascular, Inova Health System, VA
[7]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

Corresponding Author:
Dr. James Hillis
Digital Clinical Research Organization, Mass General Brigham AI
399 Revolution Drive, Somerville, MA 02145
james.hillis@mgh.harvard.edu

617-726-2000

Acknowledgements
The authors thank the broader team members at the hospitals that acquired cases (Baruch Padeh Medical Center, Inova Health, Sheba Medical Center and Thomas Jefferson University Hospital), at Mass General Brigham AI and at AISAP LTD.

Funding / Role of Funder:
SFM, BCB, TS, MAH, ALM, INC, EL, ASS, VT, SMH, JRM, JW and JH were employees of Mass General Brigham and/or affiliated hospitals at the time of the study, which had received institutional funding from the sponsor (AISAP LTD) to conduct this project. PM and QZ were employees of Thomas Jefferson University and Inova Health System respectively, which had received institutional funding from the sponsor to conduct this project.

The sponsor was involved in the design and conduct of the study; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The sponsor was involved in the collection of POCUS cases, which were transferred to Mass General Brigham; the sponsor was otherwise not involved in the collection, management, analysis, and interpretation of the data.

Conflicts of Interest:
SMH received fees paid to her institution for core lab services from Cytokinetics and BMS; she is on the advisory board for Cytokinetics.

Data Accountability:
SFM had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Date of Revision:
June 27 2025

Abstract

Introduction:

Point-of-care ultrasonography (POCUS) enables clinicians to obtain critical diagnostic information at the bedside especially in resource limited settings. This information may include 2D cardiac quantitative data, although measuring the data manually can be time-consuming and subject to user experience. Artificial intelligence (AI) can potentially automate this quantification. This study assessed the interpretation of key cardiac measurements on POCUS images by an AI-enabled device (AISAP Cardio V1.0).

Methods:

This retrospective diagnostic accuracy study included 200 POCUS cases from four hospitals (two in Israel and two in the United States). Each case was independently interpreted by three cardiologists and the device for seven measurements (left ventricular (LV) ejection fraction, inferior vena cava (IVC) maximal diameter, left atrial (LA) area, right atrial (RA) area, LV end diastolic diameter, right ventricular (RV) fractional area change and aortic root diameter). The endpoints were the root mean square error (RMSE) of the device compared to the average cardiologist measurement (LV ejection fraction and IVC maximal diameter were primary endpoints; the other measurements were secondary endpoints). Predefined passing criteria were based on the upper bounds of the RMSE 95% confidence intervals (CIs). The inter-cardiologist RMSE was also calculated for reference.

Results:

The device achieved the passing criteria for six of the seven measurements. While not achieving the passing criterion for RV fractional area change, it still achieved a better RMSE than the inter-cardiologist RMSE. The RMSE was 6.20% (95% CI: 5.57 to 6.83; inter-cardiologist RMSE of 8.23%) for LV ejection fraction, 0.25cm (95% CI: 0.20 to 0.29; 0.36cm) for IVC maximal diameter, 2.39cm^2 (95% CI: 1.96 to 2.82; 4.39cm^2) for LA area, 2.11cm^2 (95% CI: 1.75 to 2.47; 3.49cm^2) for RA area, 5.06mm (95% CI: 4.58 to 5.55; 4.67mm) for LV end diastolic diameter, 10.17% (95% CI: 9.01 to 11.33; 14.12%) for RV fractional area change and 0.19cm (95% CI: 0.16 to 0.21; 0.24cm) for aortic root diameter.

Discussion:

The device accurately calculated these cardiac measurements especially when benchmarked against inter-cardiologist variability. Its use could assist clinicians who utilize POCUS and better enable their clinical decision-making.

Introduction

Point-of-care ultrasonography (POCUS) has revolutionized many aspects of acute clinical care over recent decades.[1-3] It enables clinicians to obtain more diagnostic information at the bedside, and has been adopted by multiple specialties especially emergency medicine and hospital internal medicine.[4-6] With some handheld probes connecting directly to a smartphone or tablet, it can be particularly beneficial in resource limited settings.[7-10] Its use includes the quantification of key cardiac measurements such as left ventricular (LV) ejection fraction and inferior vena cava (IVC) size.[11]

Artificial intelligence (AI) has similarly started to revolutionize clinical care with many of its current applications relating to medical image interpretation.[12,13] One of its core uses is the automatic quantification of measurements, especially those that otherwise require time-consuming manual tracings. It has previously been used to quantify cardiac measurements using conventional echocardiograms.[14-16] It has also been shown to improve interobserver variability and workflow efficiency for echocardiogram interpretation.[17,18] Echocardiography, however, typically provides more comprehensive evaluations with a higher image resolution than POCUS.[19]

This study evaluated the ability of an AI-enabled device to calculate seven key cardiac measurements based on POCUS images including LV ejection fraction, IVC maximal diameter, left atrial (LA) area, right atrial (RA) area, LV end diastolic diameter, right ventricular (RV) fractional area change and aortic root diameter.

Methods

Study design

This diagnostic study was conducted using POCUS cases acquired prospectively from four hospitals in Israel and the United States. The acquisition of cases was approved by local Institutional Review Boards at each of Baruch Padeh Medical Center, Inova Health, Sheba Medical Center and Thomas Jefferson University Hospital. All patients provided informed consent. The subsequent diagnostic study involved transfer of de-identified cases to Mass General Brigham and was approved by the Mass General Brigham Institutional Review Board. The study was conducted in accordance with relevant guidelines and regulations. This manuscript followed the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline.[20]

Inclusion criteria

Eligible participants were identified in a consecutive manner in both the inpatient and outpatient settings. They were patients with suspected or established cardiac pathology, and with an indication for cardiac ultrasound. They were initially identified by treating physicians including cardiologists, internal medicine physicians and emergency department physicians, and were consented by local research personnel. The inclusion criteria included male or female individuals aged 21 or older, the provision of a signed and dated informed consent form, a stated willingness to comply with all study procedures and availability for the study duration. The exclusion criteria included unstable clinical condition as determined by the principal investigator (where ultrasonography was considered detrimental to the participant), clinical conditions that

prevented an ultrasonography study according to the investigators or treating physicians, presence of a prosthetic cardiac valve of any type or position, pregnancy or lactation, contact isolation due to infectious disease, presence of a left ventricular assist device, known congenital heart disease, body mass index (BMI) of 40 or greater, history of lung resection, known very poor cardiac ultrasonography image quality based on prior comprehensive echocardiography examinations, or mechanical ventilation or other situations where the required ultrasonography views could not be adequately obtained.

POCUS

The POCUS cases were acquired by physicians and sonographers at the bedside using mostly handheld devices. They contained only four cardiac views according to the American Society of Echocardiography guidelines.[21] These views included parasternal long axis with and without color Doppler, apical four chamber with and without color Doppler, parasternal short axis at the aortic valve level, and subcostal IVC view.

Cohort selection

Once cases were acquired at individual sites, they underwent de-identification prior to transfer to the sponsor. These cases were maintained separately from the cases used for device development. Only one of the four sites had contributed cases for device development, which had all been acquired prior to the cases used for the validation study. The deidentified cases were subsequently transferred to Mass General Brigham for the diagnostic study. The final cohort required only a subset of cases based on powering calculations, which was selected in a randomized manner.

Ground truth process

The three ground truth cardiologists were US board-certified with a minimum of 5 years of experience in echocardiography. They received training for this project and performed their annotations using a validated custom interface in Siemens syngo.via (version VA20). They reviewed the available loops of the relevant view(s) for each measurement (Supplementary Table 1) and chose the most adequate loop to perform the measurement. They did not receive additional clinical information about the cases. For LV ejection fraction, they visually estimated the measurement; other measurements were calculated based on the American Society of Echocardiography guidelines.[21] RV fractional area change was calculated from manually traced end-diastolic and end-systolic RV areas. The component measurements for RV fractional area change were measured 3 times on different beats and the average fractional area change was calculated; each cardiologist obtained the other measurements once per case. Separately they indicated when a measurement could not be obtained and should therefore be excluded; if at least two of the cardiologists indicated that the measurement could not be obtained then the measurement was excluded for that case; if only one of them indicated that the measurement could be obtained then that cardiologist was asked to review the case and make an actual best determination of the measurement.

Device development and inference

The device evaluated in this study was trained using labeled echocardiography data acquired at Sheba Medical Center between 2009 and 2020. The development set

included 120,127 adult patient examinations, which were split into a training set (76,342 cases, 63.6%), validation set (22,825 cases, 19.0%), and a hold-out test set (20,960 cases, 17.4%; which was separate to the cases evaluated in the current study). This development set represented a mix of inpatient and outpatient studies on a range of echocardiography machines. Each examination comprised 20-80 ultrasound images and loops, capturing various cardiac views according to echocardiography protocols.

In brief, the device architecture consists of 2 types of models. The first is a model which has a combination of convolutional neural networks (CNNs) and transformers. Feature vectors are extracted using designated trained EfficientNetV2[22] like architecture and fed into transformer blocks with connected CNNs to highlight relevant features across time. The final output vectors then underwent max-pooling and were processed by a regressor head. The second type of model is an instance segmentation CNN model trained for segmentation of relevant anatomical parts within an ultrasound image.

The device analyzed parasternal, apical and subcostal views; for measurements requiring multiple views, frames from both views were concatenated. The presence of multiple loops of a view during inference resulted in averaged predictions in an effort to increase the robustness of measurements. The device performed inference independently of the ground truth cardiologists and did not receive additional clinical information about the cases.

The device provided a single value for each of the 7 measurements. It could also provide an output of “NULL” that reflected it could not interpret that measurement for the case. The reasons included insufficient quality of images (as the device sees them) or inability to calculate the measurement due to the absence of key points in the image used by the device.

Statistical Analysis

The statistical analysis was conducted using software R version 4.2.0 (R Foundation for Statistical Computing, Vienna, Austria, 2019). The predefined primary and secondary endpoints were the diagnostic accuracy of the device as determined by the root mean square error (RMSE) when compared with the average of the three ground truth cardiologists. LV ejection fraction and IVC maximal diameter were considered primary endpoints given these metrics often impact patient care more directly; the other measurements were considered secondary endpoints. The RMSE was calculated as the square root of the mean of the squared differences between the device measurements and the average cardiologist measurements, providing a measure of overall deviation.[23] The inter-cardiologist RMSE was also calculated for reference, which was the average of the pairwise RMSEs between each combination of cardiologists. The passing criteria for the primary endpoints were RMSEs of less than 7.0% for LV ejection fraction and 0.3cm for IVC maximal diameter, which were to be met by the upper bounds of the 95% confidence intervals (CIs). The passing criteria for the secondary endpoints were RMSEs less than 20% of the average of the absolute

measurement from the three cardiologists, which similarly was to be met by the upper bounds of the 95% CIs.

To further evaluate the interchangeability and accuracy of the device measurements compared to cardiologist measurements, additional predefined secondary assessments were calculated including the individual equivalence coefficient (IEC; to compare variability of the relative differences between the device and the cardiologists with the variability of the relative differences between cardiologists),[24] mean absolute error (MAE; to assess the average magnitude of errors),[23] limits of agreement (LOA) using Bland-Altman analysis (to evaluate the presence of systematic errors between the device and cardiologists),[25] and intraclass correlation coefficient (ICC; to measure the consistency and reproducibility of the measurements).[26] MAE was computed as the mean of the absolute differences between the device and cardiologist measurements, providing a metric of average error magnitude without being influenced by the direction of errors. The Bland-Altman analysis involved plotting the mean of each device-cardiologist measurement pair against their difference, allowing for visual assessment of bias and agreement. The 95% LOAs were defined as the mean difference ± 1.96 × standard deviation of the differences, indicating the range within which most differences are expected to lie. For the IEC and ICC calculations, the measurements from each of the cardiologists were considered. For MAE and LOA, the device was compared with the average measurement from the three cardiologists.

Subgroup analyses were calculated for demographic (sex, age, race and ethnicity, BMI), disease (LV ejection fraction) and image (ultrasound probe manufacturer, image quality score, role of person acquiring scan) parameters. The subgroups were predefined except for race and role of person acquiring scan (which were added as post-hoc analyses), and BMI and LV ejection fraction (for which the subgroups were updated to have more subgroups following regulatory feedback). Each ultrasound probe manufacturer only had a single model so both the manufacturers and models are specified in the results.

All metrics were accompanied by bootstrapped 95% CIs with 2,000 resamples. An emphasis was placed on the interpretation of CIs to infer the reliability of calculations. No adjustments for multiple comparisons were made. Sample size calculations were performed a priori for the primary endpoints and powered based on preliminary device results.

Results:

Participant characteristics

The overall cohort included cases from 200 patients (Figure 1). The mean age (± standard deviation) was 65.4 (± 16.0) years and 74 (37.0%) were female patients (Table 1). A case was excluded from the analysis for a given measurement if the cardiologists or the device could not obtain the measurement (Supplementary Table 2).

Primary and secondary endpoint analysis

The RMSE was calculated between the device and the three cardiologists (Table 2). Both primary endpoint passing criteria were met, with the RMSEs for LV ejection fraction and IVC maximal diameter having a 95% CI upper bound less than 7.0% and 0.3cm respectively. Four out of the five secondary endpoint passing criteria were met, with the RMSEs achieving a 95% CI upper bound less than 20% of the mean cardiologist measurement. RV fractional area change was the one secondary endpoint that did not achieve its passing criterion, although the RMSE was still better between the device and cardiologists than between the cardiologists. The other measurements, except for LV end diastolic diameter, also had better RMSEs between the device and the cardiologists than between the cardiologists.

Secondary assessment analysis

The IEC, MAE, LOA and ICC were calculated between the device and the three cardiologists (Table 3). The IECs were mostly less than 0, showing better variability of the relative differences between the device and cardiologists compared with the

variability of the relative differences between the cardiologists. The MAEs, which provided an additional statistical method to RMSEs for measuring the error of the device, were less than the RMSEs. The LOAs, which can show the presence of systematic bias in a device, had mean differences that were unlikely to be clinically meaningful. The ICCs were mostly above 0.75, reflecting the strength of the correlation between the measurements from the device and the measurements from the cardiologists.

Subgroup analyses

The subgroup analyses evaluated the RMSEs across demographic, disease and image parameter subgroups (Table 4). The vast majority of RMSEs were similar across subgroups as suggested by overlapping ranges of 95% CIs: there were overlapping ranges for all measurements in the subgroups for age, sex, race and ethnicity, image quality score and role of person acquiring scan. The measurements with non-overlapping ranges included RA area for BMI subgroups (the <25 subgroup had lower RMSE than the [30,35) subgroup), LV ejection fraction for LV ejection fraction subgroups (the <30% subgroup had lower RMSE than the ≥30% to <53% and ≥53% subgroups), LV end diastolic diameter for LV ejection fraction subgroups (the <30% subgroup had lower RMSE than the ≥53% subgroup), and LV ejection fraction for manufacturer subgroups (the GE Healthcare subgroup had lower RMSE than the Philips or Wisonic subgroups). In addition, the BMI unknown subgroup had only 2 cases and often had non-overlapping 95% CIs.

Discussion:

This study assessed the performance of an AI-enabled device at interpreting seven key cardiac measurements on images acquired using POCUS devices. It achieved both of its primary endpoint passing criteria and all but one of its secondary endpoint passing criteria, which were based on the RMSE between the device and the average of three cardiologists for the measurements. It also demonstrated better performance compared to the inter-cardiologist RMSE for all but one of the measurements. Importantly, this study was based mostly on images acquired from inexpensive, predominantly handheld devices that could perform a short POCUS scan at the bedside unlike many other studies that have used images from comprehensive echocardiography examinations obtained in echocardiography labs.[14-16]

Overall, this performance suggests that the device provides accurate and reliable measurements. The improved RMSEs compared with cardiologists may reflect the ability of the device to average across multiple beats and loops, with the cardiologists having been asked to quantify most measurements only once. This situation, however, mirrors the real-world in that cardiologists and other clinicians are often constrained by the need to manually obtain measurements while automated measurements can leverage a greater ability to average. A similar improvement in interobserver variability has previously been shown for conventional echocardiograms.[17] The ability to average may also reflect the ability of the device to learn from the measurement styles of multiple cardiologists; in the context of electroencephalogram interpretation, this use of AI to average has been described as simulating a committee of independent experts.[27]

The one measurement where the device did not achieve the passing criteria was for RV fractional area change. For this measurement, however, the RMSE between the device and cardiologists was still better than between the cardiologists, reflecting the large variability that can be seen with this measurement and reinforcing the benefit that the device can provide through averaging. LV end diastolic diameter was the one measurement where the RMSE was greater between the device and cardiologists compared to between the cardiologists, although the RMSE between the device and cardiologists still achieved the secondary endpoint passing criterion.

POCUS provides clinicians with the ability to obtain critical diagnostic information at the bedside. For example, the knowledge of a patient's LV ejection fraction or IVC maximal diameter can guide next management steps in a critical ill patient including whether to administer fluids.[28,29] The Lumify (Philips) ultrasound probe, which was one of four assessed here, can connect directly to smart phones and tablets,[30] while the Kosmos (EchoNous) ultrasound probe can connect directly to tablets.[31] The use of the device with these ultrasound probes may be particularly beneficial in resource limited settings.

A key challenge with POCUS acquisition is that the image acquisition is often not as standardized as a conventional echocardiogram performed in an echocardiography laboratory. This study assessed how the device handles this challenge in multiple ways. Firstly, the device has quality control checks to ensure sufficient image quality and the presence of key points within images; the device did not provide a quantification for

between 0.5% (for LV ejection fraction) and 13.5% (for LV end diastolic diameter) of cases for each measurement. Secondly, an image quality score was provided for each case; the RMSEs for image quality scores of ≤4 and >4 were similar for all measurements. Thirdly, the RMSEs were also similar for all measurements when the person acquiring the scan was a physician or a sonographer.

Limitations

A key limitation of this study is that it was performed outside the clinical environment and therefore did not assess the impact of the device use on patient care and outcomes. Further to this point, the study did not assess the ability to generate the measurements in real-time with POCUS acquisition, where the device may provide the greatest clinical utility. Nevertheless, the images in this study were prospectively acquired at the bedside by mostly handheld devices and not as part of formal echocardiography studies obtained in an echocardiography lab; they therefore represented the real-world scenario where this device would be used. This study was also limited to four sites and the device will inevitably encounter new scenarios with clinical use, including further ultrasound probe manufacturers and models. For these reasons, the ongoing assessment of the device will be critical as it is deployed in the clinical environment.

Conclusion

This AI-enabled device accurately calculated seven cardiac measurements on POCUS especially when benchmarked against inter-cardiologist variability. It could facilitate better clinical decision-making including in the acute care environment.

References

1. Bernstein E, Wang TY. Point-of-Care Ultrasonography: Visually Satisfying Medicine or Evidence-Based Medicine? *JAMA Intern Med*. Dec 1 2021;181(12):1558-1559. doi:10.1001/jamainternmed.2021.5831

2. Moore CL, Copel JA. Point-of-care ultrasonography. *N Engl J Med*. Feb 24 2011;364(8):749-57. doi:10.1056/NEJMra0909487

3. Narula J, Chandrashekhar Y, Braunwald E. Time to Add a Fifth Pillar to Bedside Physical Examination: Inspection, Palpation, Percussion, Auscultation, and Insonation. *JAMA Cardiol*. Apr 1 2018;3(4):346-350. doi:10.1001/jamacardio.2018.0001

4. Soni NJ, Schnobrich D, Mathews BK, et al. Point-of-Care Ultrasound for Hospitalists: A Position Statement of the Society of Hospital Medicine. *J Hosp Med*. Jan 2 2019;14:E1-E6. doi:10.12788/jhm.3079

5. Durston W, Carl ML, Guerra W. Patient satisfaction and diagnostic accuracy with ultrasound by emergency physicians. *Am J Emerg Med*. Nov 1999;17(7):642-6. doi:10.1016/s0735-6757(99)90150-x

6. Ultrasound Guidelines: Emergency, Point-of-Care, and Clinical Ultrasound Guidelines in Medicine. *Ann Emerg Med*. Sep 2023;82(3):e115-e155. doi:10.1016/j.annemergmed.2023.06.005

7. Le MT, Voigt L, Nathanson R, et al. Comparison of four handheld point-of-care ultrasound devices by expert users. *Ultrasound J*. Jul 7 2022;14(1):27. doi:10.1186/s13089-022-00274-6

8. Rykkje A, Carlsen JF, Nielsen MB. Hand-Held Ultrasound Devices Compared with High-End Ultrasound Systems: A Systematic Review. *Diagnostics (Basel)*. Jun 15 2019;9(2)doi:10.3390/diagnostics9020061

9. Stringer JSA, Pokaprakarn T, Prieto JC, et al. Diagnostic Accuracy of an Integrated AI Tool to Estimate Gestational Age From Blind Ultrasound Sweeps. *JAMA*. Aug 27 2024;332(8):649-657. doi:10.1001/jama.2024.10770

10. Thavanathan RS, Woo MY, Hall G. The future is in your hands - Handheld ultrasound in the emergency department. *CJEM*. Nov 2020;22(6):742-744. doi:10.1017/cem.2020.449

11. Kirkpatrick JN, Grimm R, Johri AM, et al. Recommendations for Echocardiography Laboratories Participating in Cardiac Point of Care Cardiac Ultrasound (POCUS) and Critical Care Echocardiography Training: Report from the American Society of Echocardiography. *J Am Soc Echocardiogr*. Apr 2020;33(4):409-422 e4. doi:10.1016/j.echo.2020.01.008

12. Rajpurkar P, Lungren MP. The Current and Future State of AI Interpretation of Medical Images. *N Engl J Med*. May 25 2023;388(21):1981-1990. doi:10.1056/NEJMra2301725

13. Thrall JH, Li X, Li Q, et al. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success. *J Am Coll Radiol*. Mar 2018;15(3 Pt B):504-508. doi:10.1016/j.jacr.2017.12.026

14. Tromp J, Bauer D, Claggett BL, et al. A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *Nat Commun*. Nov 9 2022;13(1):6776. doi:10.1038/s41467-022-34245-1

15. Tromp J, Seekings PJ, Hung CL, et al. Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *Lancet Digit Health*. Jan 2022;4(1):e46-e54. doi:10.1016/S2589-7500(21)00235-1

16. Zhang J, Gajjala S, Agrawal P, et al. Fully Automated Echocardiogram Interpretation in Clinical Practice. *Circulation*. Oct 16 2018;138(16):1623-1635. doi:10.1161/CIRCULATIONAHA.118.034338

17. Mor-Avi V, Blitz A, Schreckenberg M, et al. Deep learning assisted measurement of echocardiographic left heart parameters: improvement in interobserver variability and workflow efficiency. *Int J Cardiovasc Imaging*. Dec 2023;39(12):2507-2516. doi:10.1007/s10554-023-02960-5

18. Hirata Y, Nomura Y, Saijo Y, Sata M, Kusunose K. Reducing echocardiographic examination time through routine use of fully automated software: a comparative study of measurement and report creation time. *J Echocardiogr*. Sep 2024;22(3):162-170. doi:10.1007/s12574-023-00636-6

19. Lee L, DeCara JM. Point-of-Care Ultrasound. *Curr Cardiol Rep*. Sep 17 2020;22(11):149. doi:10.1007/s11886-020-01394-y

20. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology*. Dec 2015;277(3):826-32. doi:10.1148/radiol.2015151516

21. Mitchell C, Rahko PS, Blauwet LA, et al. Guidelines for Performing a Comprehensive Transthoracic Echocardiographic Examination in Adults: Recommendations from the American Society of Echocardiography. *J Am Soc Echocardiogr*. Jan 2019;32(1):1-64. doi:10.1016/j.echo.2018.06.004

22. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. PMLR; 2021:10096-10106.

23. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci Model Dev*. 2014;7(3):1247-1250. doi:10.5194/gmd-7-1247-2014

24. Barnhart HX, Kosinski AS, Haber MJ. Assessing individual agreement. *J Biopharm Stat*. 2007;17(4):697-719. doi:10.1080/10543400701329489

25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. Feb 8 1986;1(8476):307-10.

26. Bartko JJ. The Intraclass Correlation Coefficient as a Measure of Reliability. *Psychological Reports*. 1966;19(1):3-11. doi:10.2466/pr0.1966.19.1.3

27. Jing J, Sun H, Kim JA, et al. Development of Expert-Level Automated Detection of Epileptiform Discharges During Electroencephalogram Interpretation. *JAMA Neurol*. Jan 1 2020;77(1):103-108. doi:10.1001/jamaneurol.2019.3485

28. Furtado S, Reis L. Inferior vena cava evaluation in fluid therapy decision making in intensive care: practical implications. *Rev Bras Ter Intensiva*. Jun 27 2019;31(2):240-247. Avaliacao da veia cava inferior na decisao de fluidoterapia em cuidados intensivos: implicacoes praticas. doi:10.5935/0103-507X.20190039

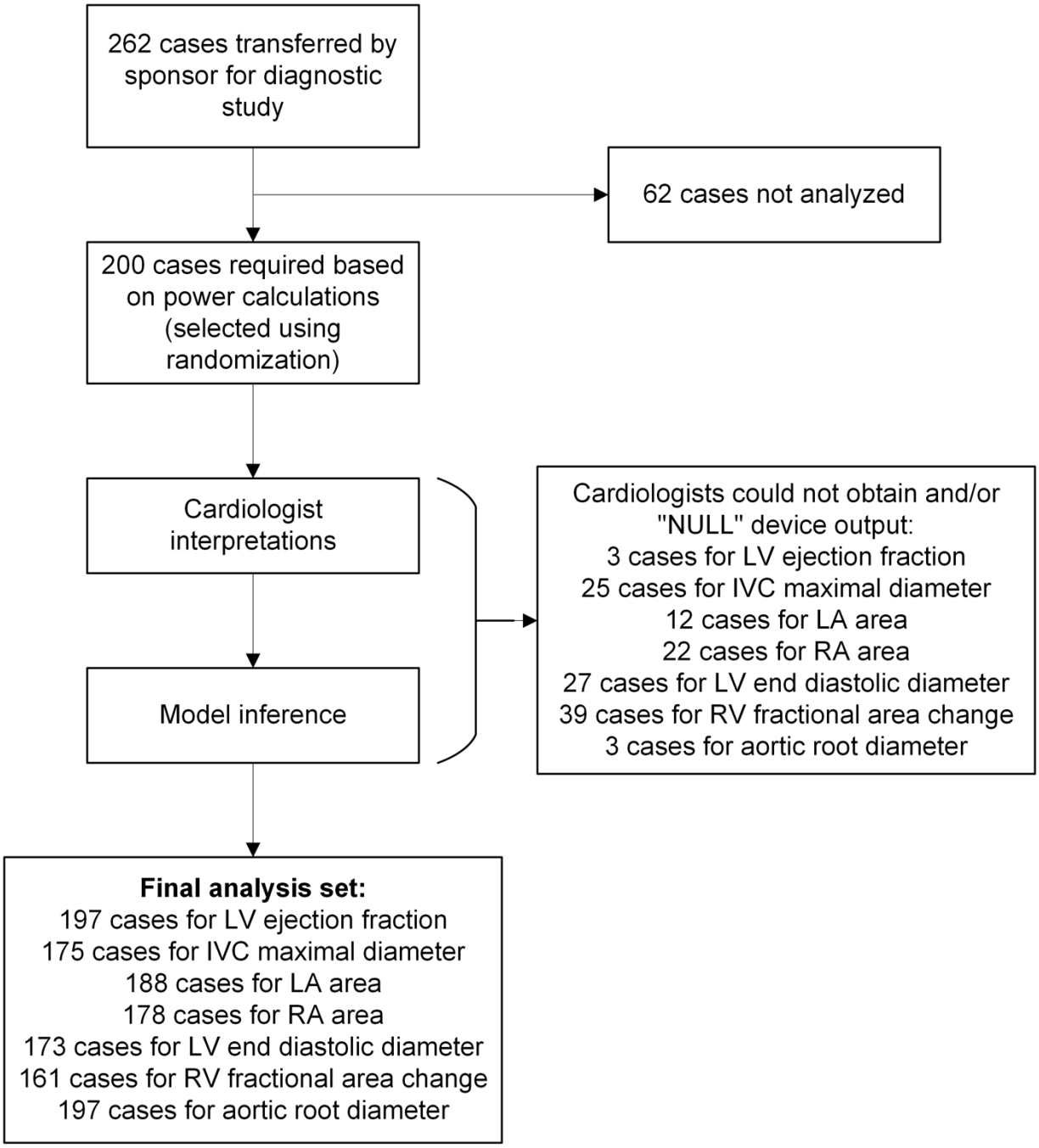
29. Miller A, Mandeville J. Predicting and measuring fluid responsiveness with echocardiography. *Echo Res Pract*. Jun 2016;3(2):G1-G12. doi:10.1530/ERP-16-0008

30. Philips - Lumify. Accessed February 3, 2025. https://www.usa.philips.com/healthcare/product/HC795005/lumify

31. Kosmos AI-driven POCUS device. Accessed February 3, 2025. https://echonous.com/product/kosmos-ultraportable-ultrasound/

Figure

Figure 1: Cohort selection. Abbreviations: IVC, inferior vena cava; LA, left atrial; LV, left ventricular; RA, right atrial; RV, right ventricular.



Tables

Table 1: Baseline demographic and technical information. Abbreviations: BMI, body mass index; LV, left ventricular; SD, standard deviation.

	Number of cases (%)
Total cohort	200
Age	
≤65 years	84 (42.0%)
>65 years	116 (58.0%)
Mean (± SD)	65.4 (± 16.0)
Sex	
Female	74 (37.0%)
Male	126 (63.0%)
Race and Ethnicity	
American Indian or Alaska Native	2 (1.0%)
Asian	11 (5.5%)
Black or African American	26 (13.0%)
Caucasian	151 (75.5%)
Hispanic or Latino	8 (4.0%)
Mixed	1 (0.5%)
Other	1 (0.5%)
BMI	
<25	68 (34.0%)
25 to <30	82 (41.0%)
30 to <35	37 (18.5%)
≥35	11 (5.5%)
Unknown	2 (1.0%)
Mean (± SD)	27.2 (± 4.7)
LV ejection fraction	
<30%	19 (9.6%)
≥30% to <53%	49 (24.7%)

≥53%	130 (65.7%)
Ultrasound Probe Manufacturer (Model)	
EchoNous (Kosmos)	37 (18.5%)
Philips (Lumify)	82 (41.0%)
Wisonic (Navi)	25 (12.5%)
GE Healthcare (Venue)	56 (28.0%)
Image quality score	
1	1 (0.5%)
2	2 (1.0%)
3	11 (5.5%)
4	36 (18.0%)
5	99 (49.5%)
Unknown	51 (25.5%)
Role of person acquiring scan	
Physician	37 (18.5%)
Sonographer	183 (81.5%)

Table 2: Primary and secondary endpoint results. The RMSE between the device and the average of the three cardiologists for the seven measurements (LV ejection fraction and IVC maximal diameter were primary endpoints; the others were secondary endpoints). The inter-cardiologist RMSE is provided for reference. This table also incorporates the number of cases and mean measurements for each measurement. Abbreviations: CI, confidence interval; IVC, inferior vena cava; LA, left atrial; LV, left ventricular; N, number of cases; RA, right atrial; RMSE, root mean square error; RV, right ventricular; SD, standard deviation.

	N	Mean measurement (SD)		RMSE (95% CI)	
		Cardiologists	Device	Device versus cardiologists	Inter-cardiologist
LV ejection fraction	197	53.15% (14.62)	49.92% (12.84)	6.20% (5.57, 6.83)	8.23% (7.52, 8.94)
IVC maximal diameter	175	1.83cm (0.57)	1.85cm (0.50)	0.25cm (0.20, 0.29)	0.36cm (0.32, 0.40)
LA area	188	21.40cm^2 (7.49)	21.63cm^2 (6.19)	2.39cm^2 (1.96, 2.82)	4.39cm^2 (2.88, 5.91)
RA area	178	16.78cm^2 (6.74)	16.49cm^2 (5.63)	2.11cm^2 (1.75, 2.47)	3.49cm^2 (3.17, 3.82)
LV end diastolic diameter	173	45.76mm (7.71)	49.20mm (6.98)	5.06mm (4.58, 5.55)	4.67mm (4.26, 5.07)
RV fractional area change	161	45.61% (13.12)	41.05% (9.43)	10.17% (9.01, 11.33)	14.12% (13.10, 15.13)
Aortic root diameter	197	3.17cm (0.47)	3.23cm (0.40)	0.19cm (0.16, 0.21)	0.24cm (0.21, 0.28)

Table 3: Secondary assessment results. The IEC, MAE, LOA and ICC between the device and the three cardiologists for the seven measurements. The LOA is expressed as the mean difference ± 1.96 × standard deviation of the differences. This table also incorporates the number of cases for each measurement. Abbreviations: CI, confidence interval; ICC, intraclass coefficient; IEC, individual equivalence coefficient; IVC, inferior vena cava; LA, left atrial; LOA, limits of agreement; LV, left ventricular; MAE, mean absolute error; N, number of cases; RA, right atrial; RV, right ventricular.

	N	IEC (95% CI)	MAE (95% CI)	LOA	ICC (95% CI)
LV ejection fraction	197	-0.20 (-0.49, 0.09)	5.01% (4.48, 5.53)	3.18 ± 10.47%	0.84 (0.81, 0.88)
IVC maximal diameter	175	-0.42 (-0.82, -0.03)	0.17cm (0.14, 0.20)	-0.03 ± 0.48cm	0.78 (0.72, 0.84)
LA area	188	-0.74 (-1.05, -0.43)	1.67cm^2 (1.42, 1.91)	-0.39 ± 4.64cm^2	0.84 (0.75, 0.93)
RA area	178	-0.69 (-0.99, -0.40)	1.48cm^2 (1.27, 1.70)	-0.03 ± 4.15 cm^2	0.85 (0.80, 0.90)
LV end diastolic diameter	173	0.93 (0.35, 1.51)	4.09mm (3.65, 4.54)	-3.46 ± 7.28mm	0.77 (0.73, 0.82)
RV fractional area change	161	-0.33 (-0.59, -0.06)	7.91% (6.93, 8.89)	4.77 ± 17.67%	0.42 (0.33, 0.51)
Aortic root diameter	197	-0.16 (-0.61, 0.30)	0.14cm (0.13, 0.16)	-0.07 ± 0.34cm	0.86 (0.82, 0.90)

Table 4: Subgroup analysis results. The RMSE between the device and the average of the three cardiologists for the seven measurements for demographic, disease and image parameter subgroups. Abbreviations: BMI, body mass index; CI, confidence interval; IVC, inferior vena cava; LA, left atrial; LV, left ventricular; N, number of cases; RA, right atrial; RMSE, root mean square error; RV, right ventricular.

	RMSE (95% CI; N)						
	LV ejection fraction	IVC maximal diameter	LA area	RA area	LV end diastolic diameter	RV fractional area change	Aortic root diameter
Age							
≤65 years	5.61% (4.50, 6.73; 82)	0.26cm (0.17, 0.34; 74)	2.35cm^2 (1.42, 3.29; 79)	2.13cm^2 (1.50, 2.76; 78)	4.76mm (4.04, 5.48; 77)	11.03% (9.22, 12.85; 67)	0.17cm (0.12, 0.21; 84)
>65 years	6.59% (5.86, 7.32; 115)	0.24cm (0.20, 0.28; 101)	2.42cm^2 (2.05, 2.79; 109)	2.10cm^2 (1.68, 2.52; 100)	5.29mm (4.64, 5.95; 96)	9.51% (8.03, 11.00; 94)	0.20cm (0.17, 0.23; 113)
Sex							
Female	6.45% (5.44, 7.46; 73)	0.23cm (0.17, 0.28; 66)	2.46cm^2 (1.61, 3.31; 69)	1.78cm^2 (1.01, 2.54; 66)	4.60mm (3.79, 5.40; 65)	9.29% (7.71, 10.86; 59)	0.19cm (0.16, 0.22; 72)
Male	6.05% (5.26, 6.84; 124)	0.26cm (0.20, 0.32; 109)	2.35cm^2 (1.91, 2.79; 119)	2.29cm^2 (1.88, 2.69; 112)	5.33mm (4.73, 5.92; 108)	10.65% (9.06, 12.25; 102)	0.19cm (0.15, 0.22; 125)
Race and Ethnicity							
Black or African American	7.77% (5.24,10.30; 25)	0.20cm (0.08, 0.32; 22)	2.05cm^2 (1.49, 2.62; 23)	2.19cm^2 (1.17, 3.21; 22)	4.37mm (2.95, 5.78; 25)	9.71% (7.43, 11.98; 19)	0.20cm (0.15, 0.25; 26)
Caucasian	6.14% (5.54, 6.74; 150)	0.25cm (0.20, 0.31; 132)	2.36cm^2 (1.86, 2.87; 145)	2.07cm^2 (1.65, 2.50; 135)	5.34mm (4.80, 5.87; 126)	9.64% (8.35, 10.93; 124)	0.19cm (0.16, 0.22; 148)
Other races (combined)	4.38% (3.11, 5.65; 22)	0.24cm (0.15, 0.33; 21)	2.90cm^2 (1.50, 4.30; 20)	2.28cm^2 (1.13, 3.44; 21)	4.10mm (2.53, 5.67; 22)	13.66% (9.33, 17.98; 18)	0.17cm (0.12, 0.23; 23)
BMI							
<25	6.49% (5.42, 7.57; 67)	0.21cm (0.16, 0.27; 60)	2.32cm^2 (1.78, 2.86; 64)	1.43cm^2 (1.09, 1.77; 58)	4.56mm (3.84, 5.28; 58)	9.15% (7.02, 11.28; 57)	0.19cm (0.15, 0.22; 67)
[25,30)	6.11% (5.05, 7.17; 81)	0.28cm (0.20, 0.36; 75)	2.62cm^2 (1.79, 3.44; 78)	1.92cm^2 (1.47, 2.37; 74)	5.33mm (4.58, 6.07; 73)	11.39% (9.58, 13.20; 68)	0.15cm (0.12, 0.18; 81)
[30,35)	6.29% (5.00, 7.57; 36)	0.24cm (0.17, 0.30; 30)	2.12cm^2 (1.61, 2.63; 33)	3.03cm^2 (2.05, 4.02; 34)	5.22mm (3.92, 6.52; 29)	9.37% (6.96, 11.78; 27)	0.24cm (0.17, 0.31; 36)
≥35	5.01% (3.65, 6.37; 11)	0.16cm (0.05, 0.27; 8)	2.10cm^2 (1.17, 3.04; 11)	2.92cm^2 (1.21, 4.64; 10)	5.59mm (3.44, 7.74; 11)	8.60% (3.03, 14.18; 7)	0.23cm (0.14, 0.32; 11)
Unknown	4.16%	0.24cm	0.43cm^2	1.18cm^2	3.51mm	9.53%	0.03cm

	(4.13, 4.20; 2)	(0.00, 0.47; 2)	(0.08, 0.79; 2)	(0.55, 1.82; 2)	(2.13, 4.89; 2)	(2.25, 16.81; 2)	(0.03, 0.03; 2)
LV ejection fraction							
<30%	4.29% (3.24, 5.33; 19)	0.27cm (0.17, 0.38; 16)	3.45cm^2 (1.83, 5.06; 18)	1.78cm^2 (1.25, 2.31; 17)	3.57mm (2.28, 4.85; 14)	10.74% (7.42, 14.06; 14)	0.14cm (0.08, 0.19; 19)
≥30% to <53%	7.04% (5.79, 8.29; 49)	0.18cm (0.14, 0.22; 41)	1.63cm^2 (1.26, 2.00; 46)	1.79cm^2 (1.11, 2.47; 45)	4.44mm (3.53, 5.34; 46)	9.75% (7.35, 12.15; 38)	0.19cm (0.14, 0.24; 49)
≥53%	6.10% (5.33, 6.87; 129)	0.26cm (0.20, 0.32; 117)	2.45cm^2 (1.92, 2.98; 123)	2.27cm^2 (1.79, 2.75; 115)	5.46mm (4.85, 6.07; 111)	10.24% (8.78, 11.71; 109)	0.19cm (0.16, 0.22; 127)
Ultrasound Probe Manufacturer (Model)							
EchoNous (Kosmos)	5.20% (4.20, 6.21; 37)	0.30cm (0.15, 0.45; 34)	2.54cm^2 (0.99, 4.08; 36)	1.76cm^2 (1.40, 2.13; 34)	5.18mm (4.02, 6.35; 32)	8.50% (6.29, 10.71; 25)	0.15cm (0.12, 0.19; 37)
Philips (Lumify)	7.12% (6.01, 8.23; 80)	0.23cm (0.17, 0.28; 70)	1.96cm^2 (1.50, 2.43; 74)	1.86cm^2 (1.42, 2.31; 71)	5.10mm (4.32, 5.88; 70)	9.71% (8.27, 11.15; 66)	0.21cm (0.16, 0.25; 80)
Wisonic (Navi)	7.28% (5.93, 8.63; 25)	0.19cm (0.13, 0.25; 23)	2.62cm^2 (2.00, 3.24; 25)	1.90cm^2 (1.22, 2.59; 24)	5.09mm (3.54, 6.64; 18)	8.77% (5.72, 11.82; 22)	0.18cm (0.13, 0.23; 25)
GE Healthcare (Venue)	4.67% (3.75, 5.59; 55)	0.25cm (0.19, 0.31; 48)	2.70cm^2 (1.97, 3.43; 53)	2.69cm^2 (1.78, 3.60; 49)	4.94mm (4.11, 5.76; 53)	12.03% (9.51, 14.54; 48)	0.18cm (0.14, 0.22; 55)
Image quality score							
≤4	5.68% (4.49, 6.87; 48)	0.26cm (0.18, 0.33; 41)	2.32cm^2 (1.79, 2.85; 45)	2.27cm^2 (1.76, 2.78; 41)	4.51mm (3.83, 5.18; 44)	9.07% (6.56, 11.58; 39)	0.19cm (0.15, 0.23; 48)
>4	5.70% (5.00, 6.40; 99)	0.27cm (0.19, 0.34; 89)	2.62cm^2 (1.90, 3.35; 94)	2.15cm^2 (1.53, 2.76; 92)	5.02mm (4.31, 5.73; 86)	10.88% (9.14, 12.63; 82)	0.18cm (0.14, 0.22; 98)
Unknown	7.51% (6.02, 8.99; 50)	0.19cm (0.14, 0.23; 45)	1.95cm^2 (1.51, 2.39; 49)	1.88cm^2 (1.41, 2.35; 45)	5.66mm (4.59, 6.73; 43)	9.68% (7.90, 11.46; 40)	0.20cm (0.16, 0.23; 51)
Role of person acquiring scan							
Physician	5.90% (4.77, 7.03; 37)	0.19cm (0.14, 0.23; 35)	1.83cm^2 (1.38, 2.28; 36)	1.79cm^2 (1.33, 2.24; 36)	5.12mm (3.91, 6.33; 29)	9.48% (6.93, 12.03; 29)	0.20cm (0.12, 0.28; 36)
Sonographer	6.27% (5.58, 6.97; 160)	0.26cm (0.20, 0.31; 140)	2.51cm^2 (2.02, 3.00; 152)	2.19cm^2 (1.74, 2.63; 142)	5.05mm (4.51, 5.60; 144)	10.32% (9.03, 11.61; 132)	0.18cm (0.16, 0.21; 161)

Supplementary Table 1: Views used to obtain each measurement.

Abbreviations: A4, apical four chamber; PLAX, parasternal long axis.

Measurement	**View**
Left ventricular ejection fraction	A4 / PLAX
Inferior vena cava maximal diameter	Subcostal
Left atrial area	A4
Right atrial area	A4
Left ventricular end diastolic diameter	PLAX
Right ventricular area (end of diastole and end of systole); to calculate right ventricular fractional area change	A4
Aortic root diameter	PLAX

Supplementary Table 2: Excluded cases. The cases were excluded when the cardiologists could not obtain the measurement or the device outputted "NULL" (see Methods for explanation). If a row exceeds a total of 200 cases, it indicates that a case was both excluded by the cardiologists and received a "NULL" device output. Abbreviations: IVC, inferior vena cava; LA, left atrial; LV, left ventricular; RA, right atrial; RV, right ventricular.

	Cardiologists could not obtain	"NULL" device output	Final cohort
LV ejection fraction	2 (1.0%)	1 (0.5%)	197
IVC maximal diameter	16 (8.0%)	25 (12.5%)	175
LA area	3 (1.5%)	10 (5.0%)	188
RA area	5 (2.5%)	18 (9.0%)	178
LV end diastolic diameter	1 (0.5%)	27 (13.5%)	173
RV fractional area change	25 (12.5%)	18 (9.0%)	161
Aortic root diameter	1 (0.5%)	2 (1.0%)	197