



Published in final edited form as:

*J Am Soc Echocardiogr.* 2021 April ; 34(4): 443–445. doi:10.1016/j.echo.2020.11.017.

## Use of Machine Learning to Improve Echocardiographic Image Interpretation Workflow: A Disruptive Paradigm Change?

**Roberto M. Lang, MD,**

University of Chicago Medical Center, Chicago, Illinois

**Karima Addetia, MD,**

University of Chicago Medical Center, Chicago, Illinois

**Tatsuya Miyoshi, MD,**

MedStar Heart and Vascular Institute/Health Research Institute, Washington, D.C.

**Kalie Kebed, MD,**

University of Chicago Medical Center, Chicago, Illinois

**Alexandra Blitz, MS,**

TOMTEC Imaging Systems, Unterschleissheim, Germany

**Marcus Schreckenber, MS,**

TOMTEC Imaging Systems, Unterschleissheim, Germany

**Niklas Hitschrich, BS,**

TOMTEC Imaging Systems, Unterschleissheim, Germany

**Victor Mor-Avi, PhD,**

University of Chicago Medical Center, Chicago, Illinois

**Federico M. Asch, MD**

MedStar Heart and Vascular Institute/Health Research Institute, Washington, D.C.

---

The typical reading of an echocardiographic study consists of reviewing multiple image sequences in the order they were acquired. The reader then integrates this information to make diagnoses on different aspects of cardiac anatomy and function. Additionally, readers need to perform measurements, the number of which has been steadily increasing under the pressure to comply with guideline-mandated quantification. Reading multiple echocardiograms is time consuming and tedious and may be overwhelming due to the large volume of information. Importantly, manual measurements are known to vary widely between readers.<sup>1–3</sup> Recent studies have shown that these tasks could be automated using machine learning (ML) techniques.<sup>4–6</sup> We hypothesized that ML could be used to improve the workflow efficiency of echocardiographic interpretation, while minimizing the interreader variability of common measurements. This approach would incorporate automated identification of image types and views, organizing images in thematic stacks,

fully automated measurements of standard parameters, and reading images organized in stacks while correcting the automated preliminary measurements as needed.

We used echocardiographic studies performed in 2,000 subjects from the World Alliance of Societies of Echocardiography Normal Values Study.<sup>7</sup> Images were labeled by an expert reader with respect to type and view. Common measurements were performed in the core laboratory according to the latest guidelines.<sup>8</sup> Data from 1,800 subjects were used for ML algorithm development purposes, and the remaining 200 subjects for testing of the newly developed algorithms.

Protocol 1 was designed to develop and test the accuracy of an ML approach for automated identification of image types and views, similar to recent studies,<sup>4-6</sup> and assigning them to “reading stacks,” in order to streamline review and interpretation. The convolutional neural network (CNN) was trained to identify image types and recognize 18 standard views from two-dimensional, tissue Doppler and pulsed-wave and continuous-wave Doppler images. Then the CNN was instructed to assign the images to eight thematic stacks, including left ventricular size and systolic function, diastolic function, right ventricular and right atrial dimensions/function, valves (mitral, aortic, tricuspid, and pulmonic), and pericardium. The results were compared to the “ground-truth” labels provided by an expert reader. Automated classification of views took <1 second per study and resulted in an overall accuracy of 90% for the two-dimensional and 94% for the Doppler images. While the agreement was excellent for most views, errors occurred mostly in labeling nonstandard, suboptimal, accidentally saved views. The CNN was able to sort the images into stacks with an accuracy of 91%. When counting views required for each stack to be complete, stack composition was 99% accurate.

Protocol 2 was designed to test the accuracy of automated measurements of standard echocardiographic parameters. The ML algorithm was trained to measure 16 commonly used parameters, the accuracy of which was tested in 200 subjects. We found excellent agreement with manual measurements by an expert reader, as reflected by high correlations, small biases, and narrow limits of agreement for most parameters (Table 1). The largest relative biases were noted for left atrial volumes in both apical views (18% and 25% of the measured values), followed by left ventricular volumes (6.5% and 7.8%), while biases were minimal for all other parameters (#3%). In a subset of 30 studies, the differences between the automated ML measurements and reference values (Table 2, fourth column) were found to be comparable to human interobserver variability between two independent expert readers who used conventional methodology (Table 2, third 3).

Protocol 3 was designed to evaluate the effectiveness of using the combination of these ML techniques in terms of efficiency and reproducibility, when compared with the current reading methodology. The two readers repeated their interpretation two weeks later utilizing the ML-assisted mode by using stacks and automated premeasurements, which they adjusted as needed. Conventional image interpretation took an average of 11’33” per study, while ML-assisted image interpretation took 6’48” on the average, namely, 41% less time. Also, with the ML-assisted interpretation, interreader variability was lower for 15/16 parameters (Table 2, fifth vs third columns).

The hypothesis that drove this study was that ML techniques could result in a disruptive change in the manner echocardiographic studies are read, while simultaneously leading to improved reproducibility of clinical measurements. Indeed, the computer was able to quickly and accurately identify the majority of image types and views, accurately organize them in thematic “stacks” designed to help answer clinically relevant questions, identify structures of interest, and perform accurate measurements, other than left atrial volume, which needs further improvement. Moreover, the use of the ML-assisted interpretation with manual adjustments saved almost half of the analysis time and reduced the variability of most parameters to below 10%, which is generally considered as optimal in the assessment of cardiac function. It is likely that, pending additional algorithm training on images with pathological findings, widespread implementation of this ML-assisted approach in clinical practice will result in significant cost savings driven by improved efficiency, physician satisfaction, and improved diagnostic performance.

## Acknowledgments

This work was supported by the American Society of Echocardiography (ASE) Foundation, MedStar Health Research Institute, and the University of Chicago, with in-kind support from TomTec.

Conflicts of Interest: A.B., M.S. and N.H. are full-time employees of TOMTEC. K.K. was funded by a T32 Cardiovascular Sciences Training Grant (5T32HL7381). No other authors have any additional disclosures.

## REFERENCES

1. 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: 507–13. [PubMed: 9076390]
2. Thomson HL, Basmadjian AJ, Rainbird AJ, Razavi M, Avierinos JF, Pellikka PA, et al. Contrast echocardiography improves the accuracy and reproducibility of left ventricular remodeling measurements: a prospective, randomly assigned, blinded study. *J Am Coll Cardiol* 2001; 38:867–75. [PubMed: 11527647]
3. Colan SD, Shirali G, Margossian R, Gallagher D, Altmann K, Canter C, et al. The ventricular volume variability study of the pediatric heart network: study design and impact of beat averaging and variable type on the reproducibility of echocardiographic measurements in children with chronic dilated cardiomyopathy. *J Am Soc Echocardiogr* 2012;25:842–8546. [PubMed: 22677278]
4. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med* 2018;1.
5. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation* 2018;138:1623–35. [PubMed: 30354459]
6. Howard JP, Tan J, Shun-Shin MJ, Mahdi D, Nowbar AN, Arnold AD, et al. Improving ultrasound video classification: an evaluation of novel deep learning methods in echocardiography. *J Med Artif Intell* 2020;3.
7. Asch FM, Banchs J, Price R, Rigolin V, Thomas JD, Weissman NJ, et al. Need for a global definition of normative echo values-rationale and design of the World Alliance of Societies of Echocardiography Normal Values Study (WASE). *J Am Soc Echocardiogr* 2019;32:157–1622. [PubMed: 30459124]
8. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, 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;28:1–3914. 10.1016/j.echo.2020.11.017 [PubMed: 25559473]

**Table 1**

Agreement between automated ML-derived measurements and expert reader's reference values in 200 studies, including linear regression with Pearson's correlation coefficients (R values) and Bland-Altman analyses of biases and limits of agreement

	<u>Linear regression</u>	<u>Bland-Altman analysis</u>			
	<b>R value</b>	<b>Bias ± SD</b>	<b>% Bias</b>	<b>Lower limit</b>	<b>Upper limit</b>
IVS thickness, mm	0.65	-0.11 ± 1.3	-1.5	-2.7	2.4
LVPW thickness, mm	0.64	-0.04 ± 1.1	-0.5	-2.2	2.1
LVIDs, mm	0.78	0.85 ± 2.5	3.0	-4.2	5.9
LVIDd, mm	0.82	0.65 ± 3.1	1.5	-5.6	6.9
LVOT diameter, mm	0.82	0.85 ± 1.5	4.1	-2.1	3.8
LV EDV (A2C), mL	0.91	6.6 ± 12.5	6.5	-18.4	31.6
LV EDV (A4C), mL	0.94	7.4 ± 9.5	7.8	-11.5	26.3
LV ESV (A2C), mL	0.87	-0.5 ± 6.2	-1.3	-12.9	11.9
LV ESV (A4C), mL	0.89	0.6 ± 5.6	1.6	-10.7	11.8
LA Vol (A2C), mL	0.87	11.8 ± 10.3	25	-8.9	32.5
LA Vol (A4C), mL	0.89	8.6 ± 8.5	18	-8.4	25.6
LVOT VTI, cm	0.91	0.46 ± 1.7	2.2	-2.8	3.8
MV E Vel, cm/sec	0.96	-0.01 ± 0.05	-1.1	-0.11	0.09
MV A Vel, cm/sec	0.95	-0.01 ± 0.05	-1.1	-0.11	0.10
LV E'(l), cm/sec	0.96	-0.03 ± 1.30	-0.2	-2.62	2.56
LV E'(s), cm/sec	0.90	-0.03 ± 1.21	-0.3	-2.45	2.39

A2C, Apical two chamber; A4C, apical four chamber; EDV, end-diastolic LV volume; E'(l), lateral mitral annular peak early; E'(s), septal mitral annular peak early; ESV, end-systolic LV volume; IVS, interventricular septum; LA, left atrial; LV, left ventricular; LVIDd, LV internal dimensions at end-diastole; LVIDs, LV internal dimensions at end-systole; LVPW, LV posterior wall; LVOT, LV outflow tract; MV, mitral valve; Vel, velocity; Vol, volume; VTI, velocity-time integral.

Positive biases represent overestimation by the ML technique, while conversely, negative biases reflect underestimation. IVS and LVPW thicknesses are measured at end diastole; LVIDs and LVIDd, LVOT diameter in midsystole, and LV ESV and LV EDV are measured in the A2C and A4C views; maximum LA volume is measured in the A2C and A4C views.

**Table 2**

Comparison between automated ML-derived measurements against expert reader's reference values obtained using the conventional methodology, side by side with intra- and interobserver variability of the conventional methodology

	Conventional interpretation		Automated ML interpretation vs conventional interpretation	ML-assisted interpretation interobserver variability
	Intraobserver variability	Interobserver variability		
IVS thickness, mm	7 ± 5	11 ± 8	14 ± 10	0 ± 1*
LVPW thickness, mm	8 ± 7	15 ± 13	17 ± 15	1 ± 3*
LVIDs, mm	3 ± 2	8 ± 6	10 ± 10	3 ± 5*
LVIDd, mm	2 ± 2	4 ± 4	6 ± 5	0 ± 1*
LVOT diameter, mm	2 ± 3	4 ± 3	5 ± 4	6 ± 14
LV EDV (A2C), mL	10 ± 9	20 ± 13	14 ± 10	6 ± 8*
LV EDV (A4C), mL	7 ± 5	22 ± 7	16 ± 8	4 ± 5*
LV ESV (A2C), mL	11 ± 9	23 ± 14	27 ± 19	3 ± 4*
LV ESV (A4C), mL	9 ± 7	32 ± 13	35 ± 16	4 ± 5*
LA Vol (A2C), mL	14 ± 9	17 ± 22	14 ± 10	9 ± 9
LA Vol (A4C), mL	13 ± 13	18 ± 13	16 ± 8	9 ± 8*
LVOT VTI, cm	5 ± 4	7 ± 5	8 ± 7	1 ± 4*
MV E Vel, cm/sec	4 ± 4	8 ± 7	6 ± 5	3 ± 16
MV A Vel, cm/sec	3 ± 3	14 ± 11	14 ± 11	3 ± 16*
LV E'(l), cm/sec	7 ± 9	10 ± 20	11 ± 17	2 ± 8
LV E'(s), cm/sec	4 ± 4	6 ± 8	8 ± 7	0 ± 0*

Values represent absolute difference in percentage of the mean. Abbreviations are as in Table 1.

\*  $P < .05$  for ML-assisted vs conventional interpretation.