BRIGHAM HEALTH
BWH BRIGHAM AND WOMEN'S HOSPITAL

A Prospective Validation of a Deep Learning-Based Automated Workflow for the Interpretation of The Echocardiogram.

Jasper Tromp (1), David Bauer (2), Brian L Claggett (2), Matthew J Frost (3), Mathias B Iversen (3), Narayana Prasad (2), Mark Petrie (4), Martin G Larson (5), Justin A. Ezekowitz (6), Scott D Solomon (2)

BACKGROUND

Echocardiography is the test of choice to assess cardiac systolic and diastolic function to diagnose and manage heart failure (HF). However, manual interpretation of the echocardiogram can be time-consuming and subject to human error.

OBJECTIVES

This study aimed to prospectively assess the interchangeability of deep learning algorithms with expert human measurements for interpreting echocardiographic studies, the primary method for assessing cardiac structure and function.

METHODS

We compared a deep learning interpretation of 23 echocardiographic parameters—including cardiac volumes, ejection fraction, and Doppler measurements—with three repeated measurements by core lab human experts in a prospective study for submission to the United States Food and Drug Administration (FDA). The primary outcome metric was the individual equivalence coefficient (IEC), which compares the disagreement between deep learning and human readers relative to the disagreement among human readers. The pre-determined non-inferiority criterion was 0.25 for the upper bound of the 95% confidence interval (CI). Secondary outcomes included measures of agreement, including the mean absolute deviation.

RESULTS

- We included 602 anonymized echocardiographic studies from 600 patients (421 with heart failure, 179 controls, 69% women) with a mean age of 57 ± 16 years.
- **The point estimates of IEC were all <0**, indicating that the disagreement between the deep learning and human measures were lower than the disagreement among three core lab readers, and the upper bound of the 95% CI of IECs fell below the prespecified success criterion of 0.25.
- **Figure 1** shows the relative absolute difference among humans (dark blue) and between automated measurements and humans (light blue) for key measurements

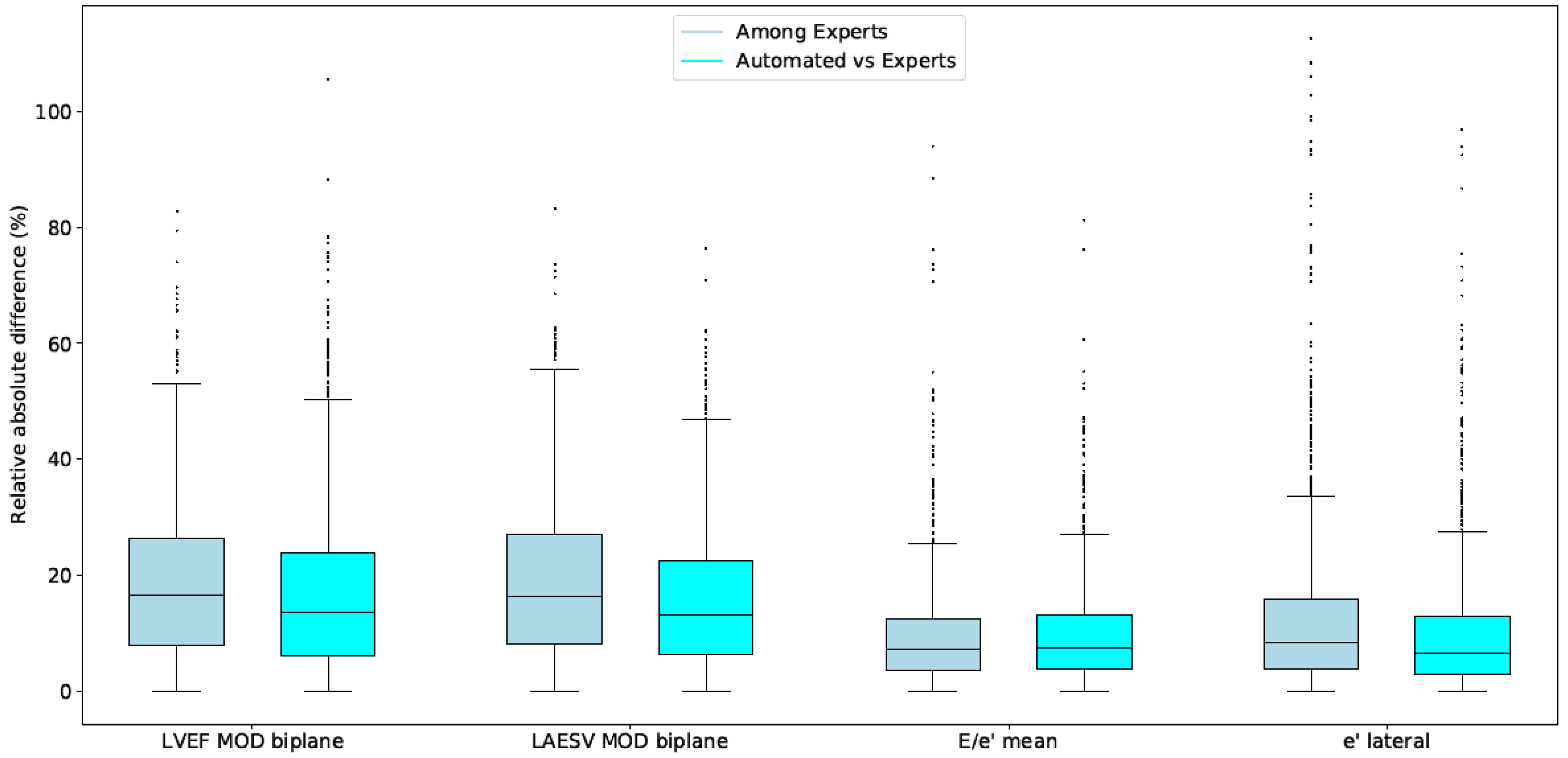

Figure 1: Relative absolute difference among humans (dark blue) and between automated measurements and humans (light blue) for LVEF, LAESV, E/e' mean and e' lateral

RESULTS

	Mean absolute deviation		Interclass correlation	
Measurement	**Automated + human readers**	Human readers	**Automated + humans readers**	**Human readers**
IVSd (mm)	**1.15**	1.20	**0.63**	0.61
LVIDd (mm)	**2.77**	2.97	**0.89**	0.88
LVIDs (mm)	**2.95**	3.60	**0.91**	0.89
LVPWd (mm)	**1.16**	1.16	**0.62**	0.63
LVEDV (mL)	**21.17**	27.6	**0.83**	0.79
LVESV (mL)	**15.6**	19.81	**0.86**	0.82
LVEF (%)	**6.7**	7.62	**0.77**	0.76
LAESV (mL)	**9.2**	11.44	**0.85**	0.82
RA (cm^2)	**1.8**	1.86	**0.89**	0.89
MV-E (cm/s)	**4.5**	4.62	**0.96**	0.96
MV-A (cm/s)	**3.9**	4.27	**0.97**	0.97
e' lateral (cm/s)	**0.8**	1.01	**0.93**	0.92
E/e' mean	**1.26**	1.28	**0.94**	0.93

Table 1: Mean Absolute Deviation (MAD) and interclass correlation coefficient (ICC) among human readers and among human + automated measurements for key measurements.

- **Table 1** shows that the mean absolute deviation decreased when automated measurements were added to the measurements by three human readers.
- The interclass correlation between automated and the average human measurements ranged from 0.61 for IVSD to 0.97 for mitral valve A.

CONCLUSIONS

This prospective validation study demonstrated excellent agreement between deep learning and expert human interpretation for a wide range of echocardiographic measurements. These results highlight the potential of deep learning algorithms to improve efficiency and reduce costs of echocardiography.

Affiliations:
(1) Saw Swee Hock School of Public Health, Singapore, Duke-NUS Medical School Singapore; (2) Cardiovascular Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; (3) Us2.ai, Singapore; (4) British Heart Foundation Cardiovascular Research Centre, University of Glasgow, Glasgow, United Kingdom & Golden Jubilee National Hospital, Clydebank, United Kingdom; (5) Department of Biostatistics, School of Public Health, Boston University, Boston, MA, USA; (6) Canadian VIGOUR Centre, University of Alberta, Edmonton, AB, Canada

Disclosures: JT has received consultancy fees from Us2.ai; MBI, MJF are employees of Us2.ai, JAE and JT serve as a scientific advisors to Us2.ai; JT holds a patent related to this work.

US2.AI