

UWL REPOSITORY

repository.uwl.ac.uk

Performance of human expert and AI measurements in conditions of progressive image degradation

Jevsikov, J, Stowell, C, Ng, T, Unsworth, B, Zolgharni, Massoud ORCID: https://orcid.org/0000-0003-0904-2904, Francis, D P and Shun-Shin, M J (2025) Performance of human expert and AI measurements in conditions of progressive image degradation. European Heart Journal -Cardiovascular Imaging, 26 (Supp.1). ISSN 2047-2404

http://dx.doi.org/10.1093/ehjci/jeae333.025

This is the Published Version of the final output.

UWL repository link: https://repository.uwl.ac.uk/id/eprint/13247/

Alternative formats: If you require this document in an alternative format, please contact: <u>open.research@uwl.ac.uk</u>

Copyright:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at <u>open.research@uwl.ac.uk</u> providing details, and we will remove access to the work immediately and investigate your claim.

Imaging – Echocardiography, Technology

Performance of human expert and AI measurements in conditions of progressive image degradation

J. Jevsikov¹, C. Stowell¹, T. Ng², B. Unsworth³, M. Zolgharni², D.P. Francis¹, M.J. Shun-Shin¹

¹National Heart and Lung Institute Imperial College, London, United Kingdom of Great Britain & Northern Ireland ²Imperial College Healthcare NHS Trust, London, United Kingdom of Great Britain & Northern Ireland ³Royal Devon and Exeter NHS Foundation Trust, Exeter, United Kingdom of Great Britain & Northern Ireland

Funding Acknowledgements: Type of funding sources: Foundation. Main funding source(s): British Heart Foundation National Institute for Health and Research

Background: A major challenge in real-world echocardiography is the difficulty in obtaining high quality images in some patients or some clinical settings. Is AI only useful when image quality is good?

Purpose: To artificially degrade adequate-quality images, and compare the ability of human experts and AI, to make measurements correctly, as image degradation worsens.

Methods: PLAX dimension measurements were made on videos of 30 patients with a range LV dimensions (mean 138mm, SD 37mm). To set the gold standard, 9 experts measured each image, blinded to the measurements made by others.

For each original image, 5 degraded versions were then made, each progressively more degraded. The degradation was designed to be maximally confusing for this measurement, namely the addition of faint ghost images of random PLAX views of other patients. This process is automatable, reproducible, and not easy to undo by conventional image processing techniques.

The 30 cases, each in original form plus 5 degraded versions, totalled 180 images, They were presented in random order for labelling by a pool of experts. They were also analysed by the Unity UK Echocardiography AI collaboratives.

Results: An example degradation sequence is shown in Figure 1 (upper panel), with images cropped for this abstract.

Across all measurements, the expert and AI suffered progressively greater measurement error as the level of image degradation increased. On average, the AI error was N smaller than the expert error (p<0.05).

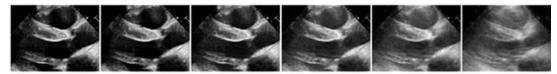
For example, for LV internal dimension, the progressive rise in human error was 2.2mm, 2.5mm, 3.1mm, 3.6mm, 5.3mm, 9.6 mm, p<0.001 for trend, Figure 1 lower panel grey bars.

Meanwhile for the AI, the corresponding errors were: 2.5mm, 2.5 mm, 2.7 mm, 3 mm, 3.9 mm, 8 mm. (p<0.001 for trend, Figure 1 lower panel grey bars).

The Minimum Heatmap Amplitude (MHA), an automatic index of AI confidence in its measurement, also declined progressively (p<0.001 for trend).

Conclusion: Both humans and AI alike make less accurate measurements as image quality degrades, although the deterioration in accuracy is more predictable for AI measurements. Importantly, deterioration in image quality (and therefore doubtfulness of the measurement) can be automatically quantified through the MHA to flag measurements needing special attention.

Figure 1: Progressive degradation in images (upper panel) and corresponding increase in measurement error by both human experts and AI (lower panel)



Average Errors

