Novel Deep Learning Model for the Detection of Cardiac Amyloidosis: A Pilot Reader Study

Jeremy A. Slivnick[1], Ashley Ackerman[2], Jorge Oliveira[2], Will Hawkes[2], Agis Chartsias[2], Ross Upton[2], Federico M. Asch[3], Juan I. Cotella[1], Linda Lee[1], Giancarlo Saldana[1], Christopher Fernandez[1], Jia Guo[1], Karima Addetia[1], Matthew Maurer[4], Steven Helmke[4], Vidhushei Yogeswaran[5], Richard Cheng[5], James N. Kirkpatrick[5], Karolina Zareba[6], Masaaki Takeuchi[7], Tetsuji Kitano[7], Marcelo Luz[8], Viviane Hotta[8], Aldo Prado[9], Pablo Ellisamburu[10], Marielle Scherrer-Crosbie[11], Marwa Soltani[11], Akhil Narang[12], Gary Woodward[2], Roberto M. Lang[1]

[1]University of Chicago, Chicago IL; [2]Ultromics Ltd., Oxford, UK; [3]MedStar Health, Washington DC; [4]Columbia University, New York NY; [5]University of Washington, Seattle WA; [6]Ohio State University, Columbus, OH, [7]University of Occupational and Environmental Health, Kitakyushu, Japan; [8]INCOR, Saõ Paolo, Brazil; [9]Instituto Privado Tucuman, Tucuman, Argentina; [10]ICBA, Buenos Aires, Argentina; [11]University of Pennsylvania, Philadelphia PA; [12]Northwestern University, Chicago IL

Background

- While TTE remains the frontline imaging modality for patients with cardiac amyloidosis (CA), classical TTE findings often lack sensitivity, resulting in delayed diagnosis and treatment.
- The aim of this study was to develop a deep learning (DL) model to detect CA and perform a preliminary assessment of its capabilities to assist readers.

Methods

- We trained and validated (75%/25% split) a 3D convolutional neural network to detect CA using 2757 apical 4-chamber (A4C) images derived from confirmed CA patients and controls.
- Utilizing a separate test dataset of 60 CA (30 AL, 30 ATTR) and 60 clinically relevant controls (**Figure 1A**), we performed a pilot study to assess the accuracy of 2 expert and 3 non-expert readers for the detection of CA using only A4C images.
- Readers assessed all images in a fully-crossed design (**Figure 1A**) with and without the aid of the DL model output.
- The reads consisted of binary interpretations indicating presence or absence of CA and high/low confidence in the interpretation.
- Accuracy, sensitivity, and specificity were compared between aided and non-aided reads for statistical difference (paired t-test) and statistical equivalence (two paired one-sided t-tests).

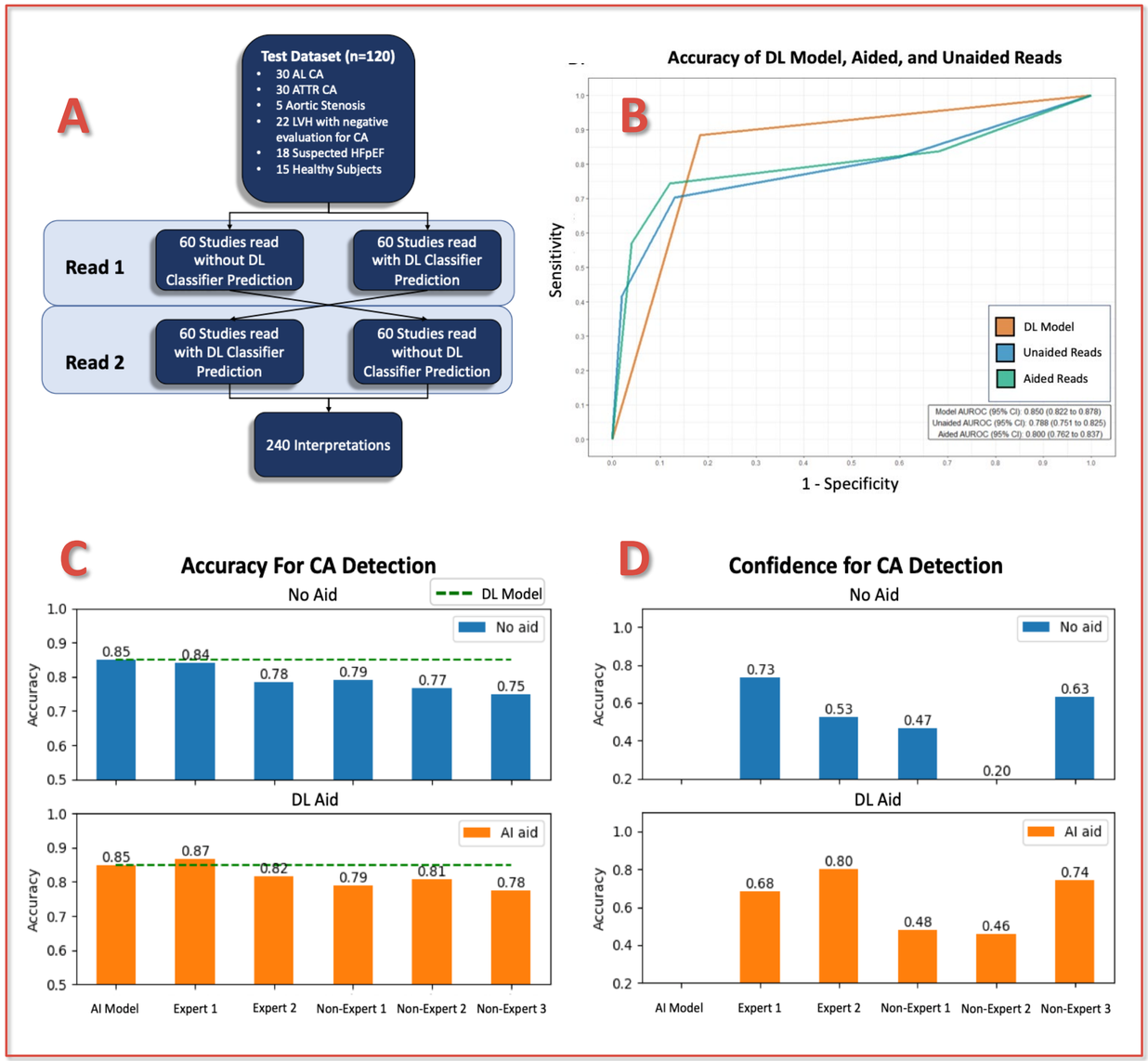

Figure 1. (A) Schematic of reader study design; (B) ROC curves for DL model, aided, and unaided readers in the test cohort. Comparison of reader (C) overall accuracy and (D) proportion of studies in which readers provided a confident interpretation for detection of CA with and without the aid of the DL model.

Results

- The DL model demonstrated an accuracy, sensitivity and specificity in the test dataset of 85% (95% CI: 81.6%, 88.2%), 88.3% (95% CI: 84.1, 92.3%), and 81.7% (95% CI 76.6, 80.3%), respectively (**Figures 1B/1C**).
- Aided by the DL model, readers demonstrated the potential for small improvements in performance (**Figure 1C/1D**), with the suggestion of larger benefits to be observed in non-experts.
- However, comparisons between aided and unaided reads were not significantly different, nor statistically equivalent, for any performance metric (all $p \geq 0.176$).

Conclusions

- In this multicenter, multi-vendor study, we developed a novel DL model which demonstrated excellent performance for differentiating CA from clinically relevant controls.
- Although the study demonstrates the potential for DL model to improve reader accuracy, particularly in non-experts, our findings indicate the need for a larger reader study to better understand if DL can augment clinical decision making to promote earlier CA diagnosis and treatment.

Disclosures: Several co-authors are employees of Ultromics Ltd.