Comparative effectiveness of standard versus AI-assisted PET/CT reading workflow for pre-treatment lymphoma staging – Multi-institutional reader study evaluation

RSNA 2022 abstract - Oral presentation M3-SSNMMI03-2

Russell Frood[1,2], Julien Willaime[3], Brad Miles[4], Greg Chambers[1], H'ssein Al-Chalabi[1,5], Tamir Ali[6], Natasha Hougham[7], Naomi Brooks[4], George Petrides[6], Matthew Naylor[6], Daniel Ward[6], Tom Sulkin[7], Richard Chaytor[7], Peter Strouhal[4], Chirag Patel[1], Andrew Scarsbrook[1,2].

1. Department of Radiology, Leeds Teaching Hospitals NHS Trust, Leeds, UK
2. Leeds Institute of Health Research, University of Leeds, Leeds, UK
3. Mirada Medical Ltd., Oxford, UK
4. Alliance Medical Ltd., Warwick, UK
5. Department of Radiology, York and Scarborough Teaching Hospitals NHS Foundation Trust, York, UK
6. Department of Radiology, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK
7. Department of Radiology, Royal Cornwall Hospitals NHS Trust, Truro, UK

Purpose

FDG PET/CT is widely used for staging of high-grade lymphoma. Evaluation of scans can be time consuming depending on complexity and disease extent, taking 15-60 minutes to analyse/report. Use of artificial intelligence (AI) could improve reporting efficiency. The study aim was to evaluate an integrated prototype segmentation tool implemented within diagnostic reading software (XD, Mirada Medical Ltd., UK) on quality and speed.

Materials and Methods

9 blinded reporters (3 trainees, 3 junior and 3 senior Consultants) from 3 imaging centres participated in a 2-part reader study. 15 lymphoma staging PET/CTs were evaluated twice. Initially, using a standard PET/CT reporting workflow, then, after a 6-week gap, with an AI-assisted workflow incorporating pre-segmentation of disease sites within the reading software. An even split of PET/CTs with ground truth (GT) only, with additional false positives (FP), and with false negative (FN) segmentations were provided. Read duration for each case was calculated using screen recording software and file logs. Report quality was independently assessed by 2 radiologists with >15 years experience using a 5-point audit score used in clinical practice. Significance was calculated using McNemar-Bowker or Wilcoxon ranked signed tests.

Results

Overall, there was a significant decrease in time between non-AI and the AI-assisted reads (median 15 vs 13.3 min, $p<0.001$). Sub-analyses by reporter experience showed this held true for junior (14.5 vs 12.7 min, $p=0.03$) and senior Consultants (15.1 vs 12.2 min, $p=0.03$) with no significant improvement in trainees (18.1 vs 18.0 min, $p=0.2$). There was no significant difference in report quality with the mode score being 5 (perfect) for both AI and non-AI assisted reads ($p=0.07$). This held true when splitting data into trainee ($p=0.7$), junior ($p=0.7$), and senior reporters ($p=0.06$), and by segmentation type, GT ($p=0.54$), FP ($p=0.18$) and FN ($p=0.56$). Post-hoc analysis showed a subset of lower quality scores for senior reporters for session 2 were not attributable to the tool.

Conclusion

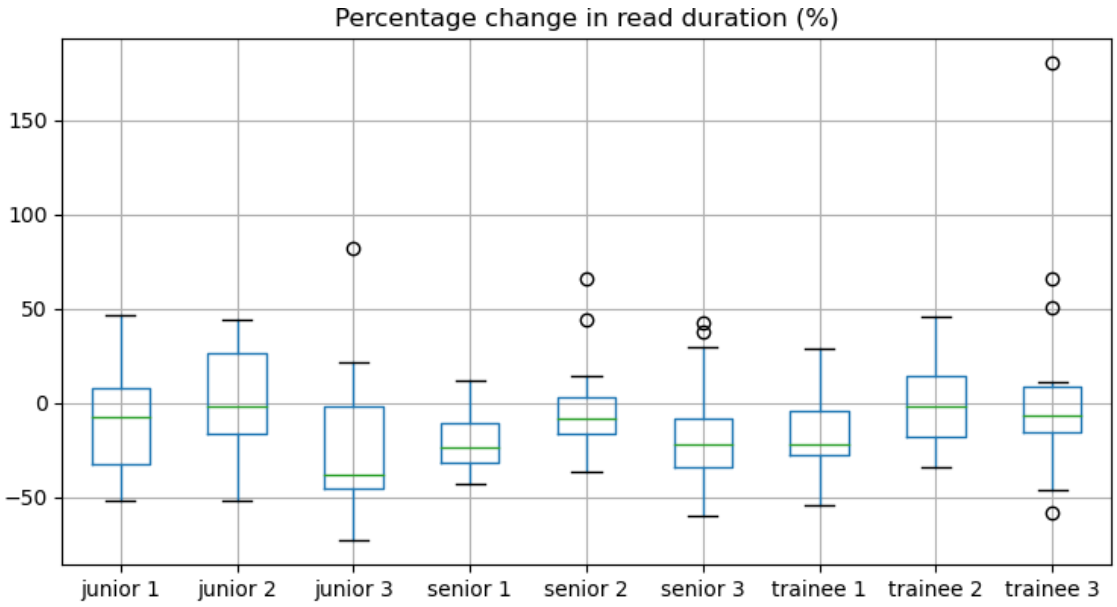
An AI-assisted tool has the potential to decrease reporting time without adversely affecting report quality.

Clinical Relevance Statement

An integrated AI-assisted reporting workflow has the potential to increase PET/CT reporting efficiency which could reduce costs and report turnaround times.

Figure: Read duration and quality of reporting

(a) Boxplots of percentage change in read duration per participant radiologist across the 15 PET/CTs. *Calculated as (session 2 - session 1) x 100 / session 1*



(b) Stacked Histogram of absolute change in report quality score per participant experience. *Calculated as (session 2 – session 1)*