



Retrospective Comparison of Fracture Detection Performance Before and After Implementation of an AI Fracture Detection Tool

Charit Tippareddy, MD, Radiology Resident (PGY4), Department of Radiology, University Hospitals Cleveland Medical Center

Sarang Ingole, MD; Sameed Khan; Kacey Pagano; Amy Zhou; Mia Zivkovic; Orlando Martinez; Navid Faraji, MD

Introduction

Al tools have been widely implemented in radiology departments promising improved efficiency, accuracy, and workload management of growing imaging volumes. This study evaluated a fracture detection tool's effect on emergency worklist prioritization, resident sensitivity, specificity, and concordance of fracture detection.

Hypothesis

AI tool implementation will decrease time-to-first-read and increase fracture detection sensitivity, specificity and resident concordance with attending read.

Methods

The study analyzed 2159 patients with extremity radiographs, 1516 of which contained both resident and attendingauthored reports. Time from exam completion to interpretation was collected. Studies with time-to-final-report greater than 120 minutes from exam completion were excluded as statistical outliers. Final attending report was used as ground truth. Resident concordance, specificity, sensitivity, and time-to-first-read were assessed before and after implementation of the AI fracture detector. Statistical significance was determined using a generalized linear mixed effects model that adjusted for scan anatomy (shoulder, humerus, elbow, wrist, hand, femur, ankle, or foot) and resident experience as fixed effects and the resident identity as a random effect. Bonferroni multiple comparisons adjustment was performed separately for each model across all coefficients.

Results

The average time from scan completion to initial interpretation was 38.0 minutes, non-significantly changed from 38.3 minutes before tool implementation (p = 0.10). Resident concordance did not differ, increasing from 94.1 percent before implementation to 95.2 percent (p = 0.36). However, resident fracture detection sensitivity, before and after adjusting for resident experience, increased from 83.7 percent to 93.1 percent ($p = 4.76 \times 10$ -4). Resident fracture specificity decreased from 98.5 percent to 96.0 percent (p = 0.06).

Conclusion

Software implementation did not affect time to first or final interpretation, likely due to de-prioritization of radiographs compared to other modalities and case types. However, the tool augmented resident sensitivity, indicating that it aids residents in identifying subtler fracture findings.

Figure(s)



Time to Read Completion by Timepoint

Figure 1. Violin plot to compare time-to-first-read before and after implementation of a fracture detection tool demonstrating no significant difference.

Keywords

Applications; Artificial Intelligence/Machine Learning; Imaging Research