Deep Learning and Acute Traumatic Rib Fractures: Does Diagnostic Performance Vary Based on Location and Displacement of Fracture?

Mohammad Alfaqih, MD, Assistant Professor, King Saud University, Lahey Hospital and Medical Center

Nathan Andrew, MD; Abdullah Alfaqih, MD; Jonathan Rassi, MD; Pauley Chea, MD; Jeffrey Hashim, MD

Introduction

Accurate detection of rib fractures in CT scans contributes to an appropriate patient care. Missed rib fractures in a whole-body CT in the sitting of poly trauma, are likely secondary to the time-pressure and the frequent associated multiple traumatic injuries.

Hypothesis

Compare performance of a convolutional neural network (CNN) to a radiologist in detecting acute rib fractures on computed tomography (CT). Investigate patterns of discrepancy between CNN output and radiologist detection.

Methods

Retrospective, IRB-approved search of institutional trauma registry from November 2018 to November 2019 was performed. Subjects with acute rib fracture on chest CT were included (N&#3f182). De-identified studies served as a validation data set for the model (Aidoc; Israel) and were unique from the original training data (N&#3f11000). CNN output was tabulated for ribs meeting inclusion and exclusion criteria (N&#3f4299). Radiologist interpretation served as the gold standard. Individual ribs were manually classified as positive or negative for fracture and compared to the CNN's classifications. Concordant results were identified as true positives or true negatives. Discordant results were identified as false positives or false negatives. Fractures were also classified by location and degree of displacement. CNN performance was analyzed for an association with these variables.

Results

CNN sensitivity for detecting at least one acute fracture in a CT exam and flagging the exam as positive was 91% (N&#3f182), whereas the sensitivity for accurately detecting acute fracture across all ribs included in the study was 48% (N&#3f4299). Specificity, PPV, and NPV were 99%, 13%, and 86% respectively. A relationship between CNN performance and degree of displacement was observed (nondisplaced=32%, comminuted=63%, partially displaced=70%, completely displaced=80%, p<.05). A relationship between CNN underperformance and anterior fractures was observed (38%; p<.05). When accounting for both fracture displacement and location, CNN performance was lowest for nondisplaced, anterior fractures (24%; p<.05) and highest for comminuted, anterior and completely displaced, posterior fractures (100%;p<.05).

Conclusion

The AI algorithm achieves its intended aim of triage at study level detection, and for rib level we found location & displacement to be important factors. The hardest cases for the AI algorithm are the anterior/non-displaced rib fractures.

Figures

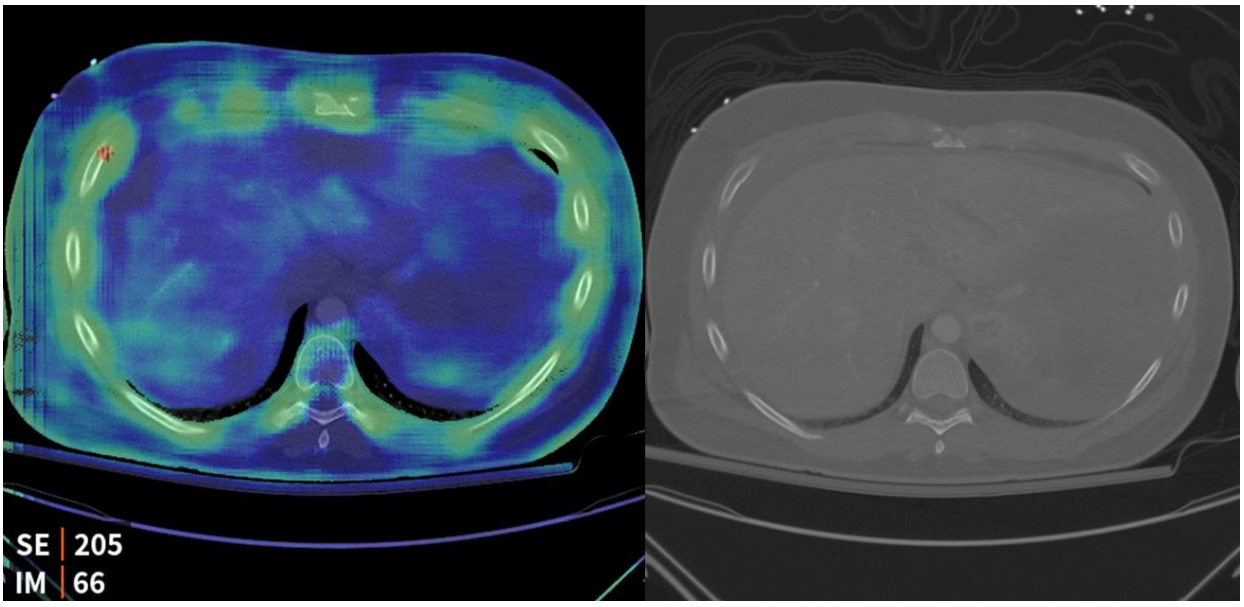

Figure 1. Nondisplaced right 6th rib fracture detected by deep learning algorithm.

Keywords

Applications; Artificial Intelligence; Clinical Workflow & Productivity

SIIM22 Presentation

Research Posters and Demonstrations | In-Person