

Received:
16 July 2020

Revised:
05 November 2020

Accepted:
07 December 2020

<https://doi.org/10.1259/bjr.20200870>

Cite this article as:

Zhang B, Jia C, Wu R, Lv B, Li B, Li F, et al. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation. *Br J Radiol* 2020; **94**: 20200870.

FULL PAPER

Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation

¹BIN ZHANG, MD, ²CHUNXUE JIA, MD, ³RUNZE WU, PhD, ⁴BAOTAO LV, MD, ⁵BEIBEI LI, ⁴FUZHOU LI, ⁴GUIJIN DU, MD, ⁴ZHENCHAO SUN, MD and ⁴XIAODONG LI, MD

¹Department of Radiology, Linyi Cancer Hospital, Shandong, China

²Clinical Research Center, Linyi People's Hospital, Shandong, China

³United Imaging Research, Shanghai, China

⁴Department of Radiology, Linyi People's Hospital, Shandong, China

⁵Department of Radiology, Linyi Central Hospital, Shandong, China

Address correspondence to:

Dr. Xiaodong Li

E-mail: lx8199819@126.com

Dr. Runze Wu

E-mail: runze.wu@gmail.com

The authors Bin Zhang and Chunxue Jia contributed equally to the work.

Objectives: To investigate the impact of deep learning (DL) on radiologists' detection accuracy and reading efficiency of rib fractures on CT.

Methods: Blunt chest trauma patients ($n = 198$) undergoing thin-slice CT were enrolled. Images were read by two radiologists (R1, R2) in three sessions: S1, unassisted reading; S2, assisted by DL as the concurrent reader; S3, DL as the second reader. The fractures detected by the readers and total reading time were documented. The reference standard for rib fractures was established by an expert panel. The sensitivity and false-positives per scan were calculated and compared among S1, S2, and S3.

Results: The reference standard identified 865 fractures on 713 ribs (102 patients). The sensitivity of S1, S2, and

S3 was 82.8, 88.9, and 88.7% for R1, and 83.9, 88.7, and 88.8% for R2, respectively. The sensitivity of S2 and S3 was significantly higher compared to S1 for both readers (all $p < 0.05$). The sensitivity between S2 and S3 did not differ significantly (both $p > 0.9$). The false-positive per scan had no difference between sessions for R1 ($p = 0.24$) but was lower for S2 and S3 than S1 for R2 (both $p < 0.05$). Reading time decreased by 36% (R1) and 34% (R2) in S2 compared to S1.

Conclusions: Using DL as a concurrent reader can improve the detection accuracy and reading efficiency for rib fracture.

Advances in knowledge: DL can be integrated into the radiology workflow to improve the accuracy and reading efficiency of CT rib fracture detection.

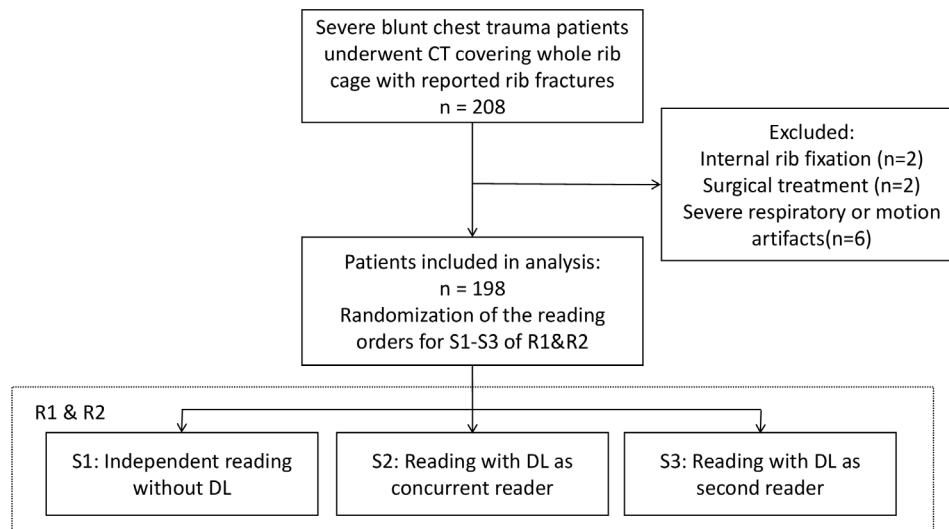
INTRODUCTION

Rib fractures are the most common type of chest injury in blunt chest trauma patients, with an estimated incidence of 10–38.7%.^{1–3} The number and pattern of rib fractures are indicative for trauma severity and can predict complications and mortality rates.^{4,5} CT is the main imaging modality for polytrauma patients and is considered appropriate for observing rib fractures caused by high-energy impact, suspected rib fractures after cardiopulmonary resuscitation,⁶ or circumstances requiring legal documentation of the injury.⁵ CT not only provides a detailed assessment of rib fractures but also a comprehensive and accurate assessment of thoracic and

abdominal injuries in patients with multiple trauma.^{7,8} Ribs are long flat bones with curved shapes, and rib fractures can present with various patterns, some of which may not be apparent on axial CT views.⁹ Furthermore, numerous CT images must be evaluated sequentially to count the ribs and detect rib fractures, which is meticulous and time-consuming work.

Deep learning (DL) is a powerful technology that can learn multiple-level representations of medical images¹⁰ that can overcome several limitations of traditional machine learning.¹¹ Several studies demonstrated great

Figure 1. The summary of the study. DL, deep learning



potential for DL-based solutions in the detection of a variety of lesions, such as diabetic retinopathy,¹² lung cancer,¹³ and colon polyps.¹⁴

A DL-based automatic rib fracture detection algorithm was developed recently.¹⁵ The integration of such a system into the workflow to assist human readers in detecting rib fractures may have a direct impact on the detection accuracy and reading efficiency of CT images. However, this effect has not been fully investigated. Therefore, we examined the rib fracture detection accuracy and reading efficiency of human readers on chest CT without DL and with DL as a concurrent or second reader.

METHODS AND MATERIALS

Patients

From January 2019 to June 2019, 208 blunt chest trauma patients who underwent CT examinations were enrolled in this retrospective study. The inclusion criterion was: patients undergoing CT examination for a suspected severe chest injury that occurred within the last month. Patients were excluded if they had internal rib fixation ($n = 2$) or had received surgical treatment ($n = 2$), or if the image had severe respiratory or motion artifacts ($n = 6$). The study design is summarised in Figure 1. The final analysis included 198 patients (115 males and 83 females; mean age 52.7 ± 13.2 years; range, 18–81 years). The enrollment was retrospective in nature, without any change in the patient management or imaging protocols. The institutional review board approved the study, and the need for informed consent was waived.

CT and image processing

All patients underwent CT scans based on our department's standard-of-care protocol using one of three multidetector CT scanners with the following parameters: Force (Siemens Healthineers, Forheim, Germany) with rotation time 0.5 s, pitch 1.0, tube voltage 120 kVp, tube current 80–350 mAs with automatic modulation; Definition AS+ (Siemens Healthineers, Forheim, Germany) with rotation time 0.5 s, pitch 1.0, tube voltage 120 kVp, tube current 100–350 mAs with automatic modulation;

Optima660 (GE Healthcare, Milwaukee) rotation time 0.5 s, pitch 0.992, tube voltage 120 kVp, tube current 100–350 mAs with automatic modulation. Scanning was performed from the level of the thoracic inlet to the level of the upper portion of the kidneys. All images were reconstructed with a slice thickness of 0.625 mm and I50/B50f or bone kernel.

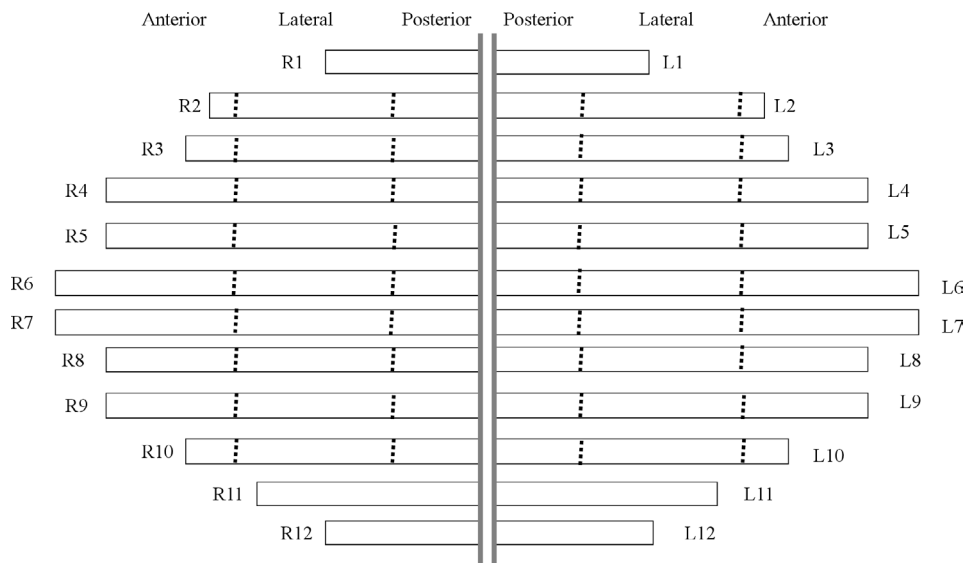
All images were transferred to an independent workstation installed with the prototype DL-based rib fracture evaluation software (Release 2019.01, United Imaging Healthcare, Shanghai, China). For simplicity, we will refer to the detection software as DL for the rest of the manuscript. The calculation time was about 2 min per patient using the current hardware configuration (CPU: Intel Xeon E5 with 24 GB RAM, GPU: Nvidia GTX1060 with 6 GB memory). The results of the rib fractures were presented in the axial view, with the automatically generated curved-planar reformation (CPR) and volume rendering (VR) views, and rib labeling to aid the human reader who checked the rib fractures detected by DL (Supplementary Material 1).

The architecture of automatic rib fracture detection is composed of two cascaded convolutional neural networks (CNN) models that are based on the Foveal network¹⁶ and Faster R-CNN.¹⁷ The first model acquires a segmentation mask for the ribs, which feed the mask into the second model in which the candidates of rib fractures are proposed. The details of the model development and training are provided in Supplementary Material 1.

Rib fracture evaluation by human readers

Two radiologists (R1 and R2) with 6 and 7 years of experience, respectively, on thoracic CT evaluated the rib fractures independently. The readers went through each case over three reading sessions (S1, S2, and S3): S1 included reading the CT images to identify the rib fractures in routine manner without DL assistance using a commercial medical image workstation (uWS-CT R004, United Imaging Healthcare, Shanghai, China); S2 included reading the CT images with rib fractures marked by DL, *i.e.* using DL as a concurrent reader; and S3 included reading

Figure 2. The readers drew a mark on the schematic diagram to show the location of the rib fracture. The dot lines divided the 2nd to 10th rib into anterior, lateral, and posterior arcs. The fracture category was documented by writing the acronym for dislocated, non-displaced, buckle fracture, and callus next to the fracture mark.

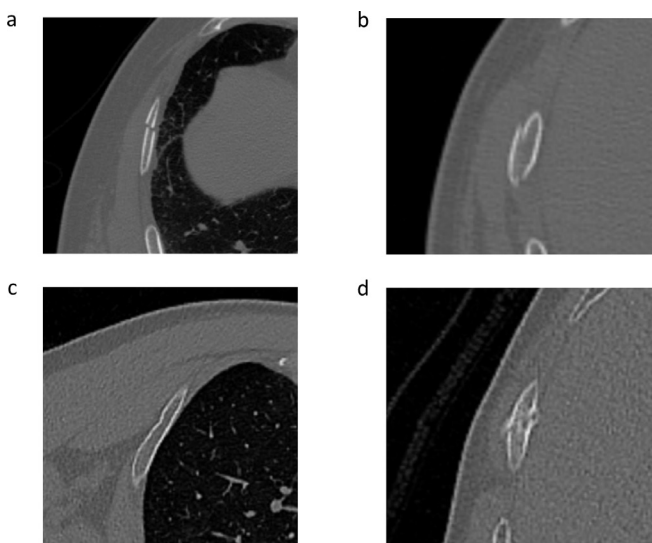


all slices initially without DL, and subsequently checking the DL marked rib fractures to make adjustments to their initial detection results, *i.e.* using DL as a second reader. The readers used the axial, coronal, and sagittal views to assess the rib fracture in all sessions. The multiplanar reconstruction (MPR), CPR, and VR views were also used to confirm the fracture whenever needed. The readers were unaware of the performance characteristics of DL but were informed that DL may or may not identify the fractures correctly, and the final diagnosis should be based on their own discretion.

The location and category of each fracture were documented on a schematic diagram of the ribs for each patient (Figure 2). Each

reader marked the fractured ribs on the diagram. To facilitate the marking process, two ancillary lines were drawn at the level of the anterior axillary line and the inferior scapula point to divide rib 2 to 10 into the anterior, lateral, and posterior regions.¹⁸ Rib fractures were classified into four categories: dislocated, non-displaced, buckle fracture, and callus formation (Figure 3). Fractures with a displacement greater than 2mm were classified as displaced fractures. A non-displaced fracture was characterized by the disruption of the cortex without significant displacement, including simple, transverse, oblique, or butterfly fractures.¹⁸ A fracture with partial disruption or “wrinkles” was classified as a buckle fracture.⁹ The fracture was classified as a callus if it was in the stage of callus formation and fracture healing, irrespective of cortical displacement or disruption.

Figure 3. Example images of four fracture categories: (a) Dislocated fracture (b) non-displaced fracture (c) buckle fracture (d) callus.



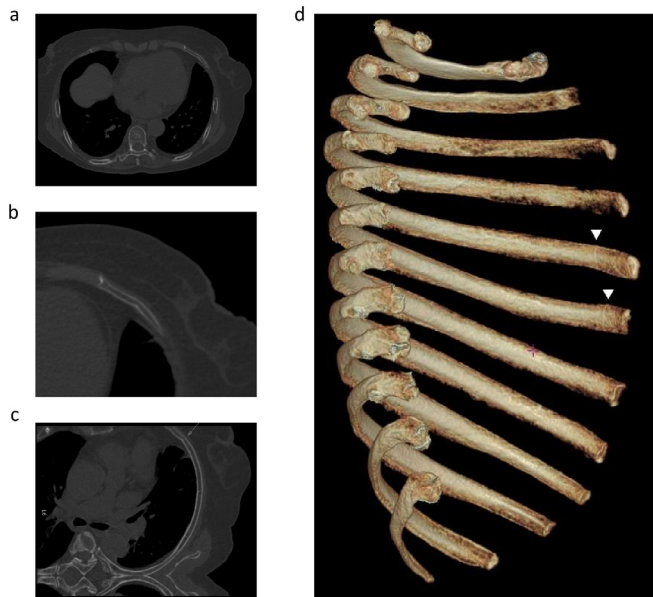
We implemented a minimum 1 month interval between each reading session to reduce recall bias. During the reading, the readers were blinded to patient information, and the order of the data was randomized in each session. A digital timer was used to measure the image evaluation time for each session. The reader started and stopped the timer at the beginning and end of the case interpretation session. All readers underwent training to familiarize themselves with the presentation and operation of the DL software before the reading session.

The fractures detected by the DL software were reviewed by a study coordinator who did not take part as a reader, and the location of the fracture was documented using the same above-mentioned schematic diagram of the ribs.

Standard of reference for rib fractures

Two radiologists with 12 and 15 years of experience in thoracic CT established the reference standard that included the exact location and type of rib fracture in a panel reading session by first reviewing all CT images to determine the rib fractures. They

Figure 4. An example of using the DL model to assist radiologists in detecting rib fractures that may be missed. (a) A CT scan of a 78-year-old patient with multiple rib fractures showing a cortical disruption on the anterior segment of the sixth left rib in the axial view, and (b) a magnified image suggesting a buckle fracture. (c) The curved-planar reformation view prepared using the DL model that also revealed a dent on the inner cortex (arrow). (d) The volume rendering view of the isolated left rib cage observed from right side further depicted notches in the vicinity of the costochondral junctions on the fifth and sixth rib (triangles). DL, deep learning.



subsequently reviewed the schematic diagrams that documented the findings of DL, R1, and R2 to reach the final decision. The location and the type of fracture of the panel review were documented with the same schematic chart used by R1 and R2.

Data analysis

All rib fractures detected by the readers and the DL were compared to the reference standard by an independent data scientist who did not act as a reader. A fracture was counted as a true positive if a fracture was documented at the same location by the reference standard. Otherwise, the fracture was counted as a false positive.

The sensitivity was calculated by dividing the number of true positives by the total number of fractures in the reference standard. The false-positive per scan (FPS) was calculated by dividing the number of false positives by the total number of the cases. Since the DL did not provide the information on the fracture types, the calculation of overall sensitivity and FPS for DL, R1, and R2 only included the location irrespective of the fracture classification. However, in the analysis of the sensitivity for each fracture type for R1 and R2, the rib fracture had to match both location and fracture type to be counted as a true positive. If the location or the type did not match, the fracture was counted as a false positive for that fracture category.

Data were processed using Microsoft Excel 2013 (v. 15.0) and R Statistical Package (R v. 3.5.2, R Foundation). The sensitivity and FPS of the fracture detection by R1 and R2 were compared between S1, S2, and S3 sessions using the χ^2 test. If a significant difference was found among the three sessions, a pairwise χ^2 test with Holm correction was performed to test the difference between each pair. The 95% confidence intervals were calculated using Wilson procedure with continuity correction. A p -value < 0.05 was considered statistically significant.

RESULTS

The reference standard analysis found 865 true fractures in 713 ribs (102 patients) among a total of 4752 ribs, including 262 (30.3%) dislocated fractures, 254 (29.4%) non-displaced fractures, 236 (27.3%) buckle fractures, and 113 (13.1%) calluses. The remaining 96 patients did not have fractures.

Sensitivity and FPS of DL

DL detected 687 (79.4%) of the 865 true fractures with 86 false-positive findings, *i.e.* 0.43 FPS. According to the fracture type assigned by the reference standard, DL detected 92.4% dislocated fractures, 78.3% non-displaced fractures, 58.1% buckle fractures, and 96.5% calluses. Moreover, DL found 75 and 66 fractures that were unidentified by R1 and R2 during S1, respectively, without DL assistance but were confirmed as true fractures by the reference standard (Figure 4). The FPS of DL was higher than that of R1 and R2 (0.43 vs 0.16 vs 0.19, $p < 0.001$). The misinterpretation of normal rib morphology as fracture (68 out of 86) was the major cause of false-positive findings.

Sensitivity and FPS of human readers

The sensitivity and FPS for R1 and R2 in S1, S2, and S3 are listed in Table 1. The sensitivity of R1 and R2 significantly increased by 4.8–6.1% in S2 and S3 compared to S1 (all $p < 0.05$). The difference in the sensitivity between S2 and S3 was not significant for both readers (both $p > 0.93$). The FPS was 0.16, 0.21, and 0.22 for R1 and 0.19, 0.13, and 0.091 for R2 in S1, S2, and S3 respectively. The FPS of R1 increased over the three sessions but was not statistically significant ($p = 0.24$). In contrast, the FPS of R2 was significantly lower in S3 than S1 ($p < 0.05$), but not different between S1 and S2 ($p = 0.13$), and between S2 and S3 ($p = 0.26$). Moreover, R1 and R2 scored 3 and 2 false positives in S1, respectively, in 96 patients without the fractures. The false-positive rates remained unchanged for S2 and S3, where DL suggested 18 candidates.

Sensitivity of four fracture categories for human readers

We found various degrees of improvement in the sensitivity for detecting four fracture categories for R1 and R2 in S2 or S3 compared to S1 (Table 2). The sensitivity of R1/R2 for detecting dislocated fracture ranged from 92.4–94.3% in S1, which increased to 95.8–98.1% in S2/S3. The sensitivity of callus detection increased from 93.8–94.7% to 94.7–98.2%. The sensitivity for detecting non-displaced fractures, however, increased from 75.2–85.4% to 86.6–89.0%. Similarly, the detection of buckle fractures increased from 67.8–72.9% to 77.1–78.0%. Statistical analysis revealed that the sensitivity of R1 in detecting non-displaced

Table 1. Sensitivity and FPS for detecting rib fractures in the three sessions

	S1	S2	S3
R1			
Sensitivity	82.8% (716/865) [80.1%, 85.2%]	88.9% (769/865) [86.6%, 90.9%]	88.7% (767/865) [86.3%, 90.7%]
FPS	0.16 (31/198) [0.11, 0.22]	0.21 (42/198) [0.16, 0.28]	0.22 (43/198) [0.16, 0.28]
R2			
Sensitivity	83.9% (726/865) [81.3%, 86.3%]	88.7% (767/865) [86.3%, 90.7%]	88.8% (768/865) [86.5%, 90.8%]
FPS	0.19 (38/198) [0.14, 0.26]	0.13 (26/198) [0.089, 0.19]	0.091 (18/198) [0.056, 0.14]

DL, deep learning; FPS, false-positive per scan.

The data in parentheses represent the detected fractures over the total number of fractures or false-positive over total scans for the sensitivity and FPS, respectively.

The data in brackets represent the 95% confidence intervals.

S1, S2, and S3 represent the reading session without DL, DL as a concurrent reader, and DL as a second reader, respectively.

R1 and R2 indicate reader 1 and reader 2, respectively.

fractures was significantly greater in S2 and S3 than in S1 (both $p < 0.01$). The sensitivity of R2 was significantly higher for the detection of dislocated fracture in S2 than in S1 ($p < 0.05$) and nearly significant for buckle fractures between S3 and S1 ($p = 0.052$). No other improvements over the three sessions were significant (all $p > 0.25$).

Reading efficiency

R1 spent 1193, 765, and 1326 min (6.0, 3.9, and 6.7 min per case) for reviewing all CT images during S1, S2, and S3, respectively. R2 spent 1069, 701 and 1198 min in S1, S2, and S3, respectively (5.4, 3.5, and 6.1 min per case). The evaluation time was 36% (R1) and 34% (R2) shorter for the two readers in S2 compared to S1 ($p < 0.001$). The evaluation time was noticeably longer in S3 compared to S1.

DISCUSSION

In our study, we investigated the detection accuracy and reading efficiency of rib fractures on CT images by radiologists with and without DL as an assistant. The results showed that 4.8–6.1% more rib fractures were found by the readers with the assistance of DL than without DL. Furthermore, the false-positive rate remained the same or decreased. Moreover, compared to the session without DL, the reading time was significantly shorter in the session using DL as a concurrent reader, while the sensitivity was improved without affecting the FPS.

The rib fracture is considered an indicator of severe trauma.^{4,5} Hospitalization is suggested for patients with three or more isolated fractured ribs, and intensive care should be given to elderly patients with six or more fractured ribs due to the elevated risks of complications and mortality.² The importance of identifying flail chest and non-flail fractures involving more than three ribs was highlighted by a recent epidemiology study that found these conditions were associated with 15- to 107-fold higher risk of surgical fixation in the rib fracture patients.¹⁹ Furthermore, the number of rib fractures and the magnitude of displacement can predict opioid requirements, providing information for pain management.²⁰ The increased sensitivity of fracture detection with DL-assisted reading may influence other algorithms of evidence-based patient stratification, such as “RibScore.”²¹ Thus, it may be used as an indicator for further studies, such as contrast-enhanced CT.⁶ Therefore, although our study focused on the development and application of DL on rib fracture detection by CT, it may also potentially influence clinical decisions for patients with severe trauma.

To detect all rib fractures, including minor fractures, consecutive slices of CT images over the entire rib cage should be carefully evaluated, which is not only time consuming but also prone to mistakes.²² Several approaches have been proposed to improve the performance of rib fracture detection and reading efficiencies, such as VR,²³ MPR, CPR, and unfolding views for the ribs.²⁴

Table 2. Sensitivity for detecting the four fracture categories

	R1			R2		
	S1	S2	S3	S1	S2	S3
Dislocated	94.3%	96.9%	96.6%	92.4%	98.1%	95.8%
Non-displaced	75.2%	87.8%	87.8%	85.4%	86.6%	89.0%
Buckle	72.9%	77.5%	77.1%	67.8%	75.8%	78.0%
Callus	93.8%	96.5%	96.5%	94.7%	98.2%	94.7%

DL, deep learning.

S1, S2, and S3 represent the reading session without DL, DL as a concurrent reader, and DL as a second reader, respectively.

R1 and R2 represent Reader 1 and Reader 2, respectively.

These approaches function on the rationale of providing comprehensive views of the rib fractures to optimize the reading process. In contrast, the DL-based solution provides a list of candidates of the fractures. The CPR view, in addition to the axial view for the rib fractures, was automatically generated to facilitate checking the results of the DL. Therefore, the DL-based solution improved the detection rate of rib fractures by providing a second opinion and reduced the time required for checking DL findings by providing an advanced visualization tool.

Our results showed that DL found rib fractures that were not initially detected by human readers, explaining the improved sensitivity of detecting rib fractures in DL-assisted sessions. Using DL as a concurrent reader also improved the time efficiency for detecting rib fractures, suggesting that the automatic CPR saved the radiologist time that would have been spent from the manual MPR preparation, because the axial view may not always be sufficient to detect a rib fracture.⁹

The adoption of deep learning may add value and efficiency to several components of the skeletal imaging workflow, including diagnostic accuracy and efficiency.²⁵ The CNNs have been applied to plain radiographs,²⁶ CT,²⁷ and PET/CT²⁸ to detect hip fractures, orbital blowout fractures, and multiple myeloma. Recently, Weikert et al²⁹ found a ResNet-based algorithm achieve good performance on per-examination level, but with lower sensitivity than our study. Furthermore, Zhou et al³⁰ also suggested that DL could support human readers and improve the sensitivity. Our study demonstrated a higher sensitivity with DL for the detection of rib fractures, consistent with previous findings. More importantly, we investigated the potential application of inserting the DL algorithm into various points of the radiology workflow and the resultant impact on diagnostic performance. When using DL as a second reader in the S3 session, the risk of potential interference of DL with the human reader's judgment was minimized, which was the chief distinction from the S2 session using DL as a concurrent reader. However, we found that rib fracture detection was equivalent between the two reading sessions, and the reading time was significantly shorter when using DL as a concurrent reader, suggesting a potential role of DL in routine practice, especially in time-sensitive settings, such as emergency rooms.

Detection performance improved in various ways for both readers and categories of fractures. The difference in the sensitivity between the two readers was greater when unassisted compared to the DL-assisted reading sessions. As expected, the sensitivity of non-displaced and buckle fractures benefited from DL assistance, as these fractures prove challenging to identify due to mild unilateral cortical lesions or bilateral cortical ruptures without significant displacement. The largest difference in sensitivity between the two readers was found for non-displaced fractures (10.2%), followed by buckle fractures (5.1%). After DL was applied, the difference was reduced to <1.9% in all fracture categories. The readers demonstrated high performance for dislocated fractures and calluses (92.4%–94.7%), but less sensitivity to non-displaced and buckle fractures (67.8%–85.4%). With DL, the sensitivity increased most significantly

for non-displaced fractures in one reader (12.6%) and buckle fracture in the other reader (10.2%). Meanwhile, the sensitivity improved to a lesser degree (0%–5.7%) for dislocated fractures and calluses when the reader performed well without assistance. Nevertheless, the overall sensitivity significantly improved with DL for both readers for four fracture categories. Therefore, DL can provide supplemental rib fracture information to human readers, improve detection performance, and reduce fracture-dependent reader variance.

Two complementary measures, sensitivity and FPS, were used to evaluate the true fracture detection efficiency and the likelihood of generating false positives. FPS could also serve as an indicator of the radiologists' additional workload involving the cross-examination of the false-positive candidates suggested by DL or colleagues in a double-reading setting. Therefore, we opined that FPS is the price paid for the gain of sensitivity, which might be a more useful measure than specificity in our study with an emphasis on the working efficiency (refer to the [Supplementary Material 2](#) for the analysis of the specificity between the sessions with and without using DL).

Our result also revealed that DL influenced the FPS in a reader-dependent manner. While not significant, the FPS of R1 increased after using DL, and that of R2 decreased. This discrepancy may be caused by the reader's experience and individual tendency in response to the DL suggestions and the different presentation sequence of DL's findings in the S2 and S3 sessions. Furthermore, the readers had a higher successful rate of rejecting false positives from DL in patients without fractures compared to patients with fractures.

Our results showed that a large proportion of false positives from DL were located on normal morphological variants. It remains a challenge to explain the features learned by the deep neural networks, and hence to infer the cause of the misclassification. From the human perspective, many false positives might be associated with the complex structure of the costal tubercle and groove of the rib. This complexity may be further enhanced by the non-uniform bone loss near cortical surface due to osteoporosis. The tubercle structure and osteoporosis could be patient dependent and require an experience radiologist to discriminate between the true- and false-positive findings. The reader should be aware of the performance of DL; a training session on DL-assisted reading with the essential nature of rib fractures could further improve the readers' ability to reject false-positive findings before adopting the DL algorithm, whose benefits should be studied in the future.

Our study had several limitations. First, the incidence of rib fractures was high. This reflected the status quo in Linyi People's Hospital that CT was often ordered for patients with severe injuries, an appropriate approach based on the Expert Panel on Thoracic Imaging recommendations.⁶ The influence of the fracture incidence on DL-assisted reading should be studied using subgroup analyses with a larger patient cohort. Second, the reference standard was established by an expert panel, which may vary among different groups. The reference standard could be further

improved by adding the follow-up CT images, which would enable serial observation of the fractures and callus formation. In our practice, however, we found a few cases of buckle fractures in which the callus formation was not observed on follow-up CT. An agreement should be made before the expert reading session to handle such cases. Third, only two readers participated in this study. Although the sensitivity improved consistently for both readers with DL-assisted reading for the detection of four types of the fractures, the FPS increased insignificantly in one reader and decreased for the other reader. A study with more readers would enable a more realistic performance evaluation, and training programs preparing radiologists to work with DL assistance may be implemented to abate the discrepancies.

The pressure to report radiological findings quickly was lower than that in real-time clinical situations, owing to the retrospective nature of our study. This could have resulted in overestimation of sensitivity. Nevertheless, our results found an increase in the sensitivity of rib fracture detection with DL compared to that without DL. Furthermore, our study simulated the rib fracture detection process under the bone window that is a part of the radiological interpretation process for chest CT of trauma.

The influence of DL-assisted rib fracture detection on the other results of trauma CT should be studied in the future.

In conclusion, the application of DL as a concurrent reader or second reader improved the performance of radiologists in detecting rib fractures in patients with blunt chest trauma. Furthermore, the reading time was significantly shorter with DL as a concurrent reader compared to without DL.

ACKNOWLEDGEMENTS

We would like to thank Dr. Dijia Wu for preparing the supplementary document.

COMPETING INTERESTS

Runze Wu is an employee of United Imaging Research. The other authors declare no conflict of interests.

FUNDING

This work was supported by the Science and Technology Department of Shandong Province (General Program) [2014GSF118132] in the collection, analysis and interpretation of data.

REFERENCES

- Ziegler DW, Agarwal NN. The morbidity and mortality of rib fractures. *J Trauma* 1994; **37**: 975–9. doi: <https://doi.org/10.1097/00005373-199412000-00018>
- Sirmali M, Türüt H, Topçu S, Gülhan E, Yazıcı U, Kaya S, et al. A comprehensive analysis of traumatic rib fractures: morbidity, mortality and management. *Eur J Cardiothorac Surg* 2003; **24**: 133–8. doi: [https://doi.org/10.1016/S1010-7940\(03\)00256-2](https://doi.org/10.1016/S1010-7940(03)00256-2)
- Marini CP, Petrone P, Soto-Sánchez A, García-Santos E, Stoller C, Verde J. Predictors of mortality in patients with rib fractures. *Eur J Trauma Emerg Surg* 2019; **37**. doi: <https://doi.org/10.1007/s00068-019-01183-5>
- Talbot BS, Gange CP, Chaturvedi A, Klionsky N, Hobbs SK, Chaturvedi A. Traumatic rib injury: patterns, imaging pitfalls, complications, and treatment. *Radiographics* 2017; **37**: 628–51. doi: <https://doi.org/10.1148/rg.2017160100>
- Murphy CE, Raja AS, Baumann BM, Medak AJ, Langdorf MI, Nishijima DK, et al. Rib fracture diagnosis in the Panscan era. *Ann Emerg Med* 2017; **70**: 904–9. doi: <https://doi.org/10.1016/j.annemergmed.2017.04.011>
- Henry TS, Donnelly EF, Boiselle PM, Crabtree TD, Iannettoni MD, Johnson GB, et al. ACR appropriateness Criteria® rib fractures. *Journal of the American College of Radiology* 2019; **16**(5S): S227–34. doi: <https://doi.org/10.1016/j.jacr.2019.02.019>
- Huber-Wagner S, Lefering R, Qvick L-M, Körner M, Kay MV, Pfeifer K-J, et al. Effect of whole-body CT during trauma resuscitation on survival: a retrospective, multicentre study. *Lancet* 2009; **373**: 1455–61. doi: [https://doi.org/10.1016/S0140-6736\(09\)60232-4](https://doi.org/10.1016/S0140-6736(09)60232-4)
- Hilbert P, zur Nieden K, Hofmann GO, Hoeller I, Koch R, Stuttmann R. New aspects in the emergency room management of critically injured patients: a multi-slice CT-oriented care algorithm. *Injury* 2007; **38**: 552–8. doi: <https://doi.org/10.1016/j.injury.2006.12.023>
- Cho SH, Sung YM, Kim MS. Missed rib fractures on evaluation of initial chest CT for trauma patients: pattern analysis and diagnostic value of coronal multiplanar reconstruction images with multidetector row CT. *Br J Radiol* 2012; **85**: e845–50. doi: <https://doi.org/10.1259/bjr/28575455>
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44. doi: <https://doi.org/10.1038/nature14539>
- Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017; **19**: 221–48. doi: <https://doi.org/10.1146/annurev-bioeng-071516-044442>
- Gulshan V, Peng L, Coram N, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus Photographs. *JAMA* 2016; **316**: 2402–10. doi: <https://doi.org/10.1001/jama.2016.17216>
- Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-To-End lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; **25**: 954–61. doi: <https://doi.org/10.1038/s41591-019-0447-x>
- Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology* 2018; **155**: 1069–78. doi: <https://doi.org/10.1053/j.gastro.2018.06.037>
- Li X. A deep-learning method for fast detection of rib fracture in ct images: effect of computer-aided diagnosis to radiologists, radiological society of north america 2018 scientific assembly and annual meeting, 2018, Chicago IL. 2018. Available from: <http://archive.rsna.org/2018/18012506.html>.
- Brosch T, Saalbach A. Foveal fully convolutional nets for multi-organ segmentation. In *medical imaging 2018: image processing*. 10574. Houston, Texas, United States: International Society for Optics and Photonics; 2018. pp. 10574.

17. Ren S, He K, Girshick R, Sun J, Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017; **39**: 1137–49. doi: <https://doi.org/10.1109/TPAMI.2016.2577031>
18. Bemelman M, Baal Mvan, Raaijmakers C, Lansink K, Leenen L, Long W. An interobserver agreement study with a new classification for rib fractures. *Chirurgia* 2019; **114**: 352–8. doi: <https://doi.org/10.21614/chirurgia.114.3.352>
19. Ingoe HM, Eardley W, McDaid C, Rangan A, Lawrence T, Hewitt C. Epidemiology of adult rib fracture and factors associated with surgical fixation: analysis of a chest wall injury dataset from England and Wales. *Injury* 2020; **51**: 218–23. doi: <https://doi.org/10.1016/j.injury.2019.10.030>
20. Bugaev N, Breeze JL, Alhazmi M, Anbari HS, Arabian SS, Holewinski S, et al. Magnitude of rib fracture displacement predicts opioid requirements. *J Trauma Acute Care Surg* 2016; **81**: 699–704. doi: <https://doi.org/10.1097/TA.0000000000001169>
21. Chapman BC, Herbert B, Rodil M, Salotto J, Stovall RT, Biffl W, et al. RibScore: a novel radiographic score based on fracture pattern that predicts pneumonia, respiratory failure, and tracheostomy. *J Trauma Acute Care Surg* 2016; **80**: 95–101. doi: <https://doi.org/10.1097/TA.0000000000000867>
22. Omert L, Yeane WW, Protetch J. Efficacy of thoracic computerized tomography in blunt chest trauma. *Am Surg* 2001; **67**: 660–4.
23. Alkadhi H, Wildermuth S, Marincek B, Boehm T. Accuracy and time efficiency for the detection of thoracic cage fractures: volume rendering compared with transverse computed tomography images. *J Comput Assist Tomogr* 2004; **28**: 378–85. doi: <https://doi.org/10.1097/00004728-200405000-00013>
24. Urbaneja A, De Verbizier J, Formery A-S, Tobon-Gomez C, Nace L, Blum A, et al. Automatic rib cage unfolding with CT cylindrical projection reformat in polytraumatized patients for rib fracture detection and characterization: feasibility and clinical application. *Eur J Radiol* 2019; **110**: 121–7. doi: <https://doi.org/10.1016/j.ejrad.2018.11.011>
25. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *AJR Am J Roentgenol* 2019; **213**: 506–13. doi: <https://doi.org/10.2214/AJR.19.21117>
26. Cheng C-T, Ho T-Y, Lee T-Y, Chang C-C, Chou C-C, Chen C-C, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol* 2019; **29**: 5469–77. doi: <https://doi.org/10.1007/s00330-019-06167-y>
27. Li L, Song X, Guo Y, Liu Y, Sun R, Zou H, et al. Deep convolutional neural networks for automatic detection of orbital blowout fractures. *J Craniofac Surg* 2020; **31**: 400–3. doi: <https://doi.org/10.1097/SCS.00000000000006069>
28. Xu L, Tetteh G, Lipkova J, Zhao Y, Li H, Christ P, et al. Automated whole-body bone lesion detection for multiple myeloma on ⁶⁸Ga-Pentixafor PET/CT imaging using deep learning methods. *Contrast Media Mol Imaging* 2018; **2018**: 2391925. doi: <https://doi.org/10.1155/2018/2391925>
29. Weikert T, Noordtzi LA, Bremerich J, Stieltjes B, Parmar V, Cyriac J, et al. Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *Korean J Radiol* 2020; **21**: 891–9. doi: <https://doi.org/10.3348/kjr.2019.0653>
30. Zhou QQ, Wang J, Tang W, Hu ZC, Xia ZY, Li XS, ZC H, XS L, et al. Automatic detection and classification of rib fractures on thoracic CT using Convolutional neural network: accuracy and feasibility. *Korean J Radiol* 2020; **21**: 869–79. doi: <https://doi.org/10.3348/kjr.2019.0651>