

Original Paper

Prospective Diagnostic Accuracy and Technical Feasibility of Artificial Intelligence-Assisted Rib Fracture Detection on Chest Radiographs: Observational Study

Shu-Tien Huang^{1,2,3}, MSc, MD; Liong-Rung Liu^{1,2,3}, MSc, MD; Ming-Feng Tsai^{2,3,4}, MD; Ming-Yuan Huang^{1,3*}, MSc, MD; Hung-Wen Chiu^{2,5*}, PhD

¹Department of Emergency Medicine, Mackay Memorial Hospital, Taipei, Taiwan

²Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, New Taipei City, Taiwan

³College of Medicine, Mackay Medical University, New Taipei City, Taiwan

⁴Division of Plastic Surgery, Mackay Memorial Hospital, Taipei, Taiwan

⁵Clinical Big Data Research, Taipei Medical University Hospital, Taipei City, Taiwan

*these authors contributed equally

Corresponding Author:

Hung-Wen Chiu, PhD

Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University
9F, Education & Research Building, Shuang-Ho Campus, No. 301, Yuantong Rd., Zhonghe Dist.
New Taipei City 235603

Taiwan

Phone: 886 266202589 ext 10929

Email: hwchiu@tmu.edu.tw

Abstract

Background: Rib fractures are present in 10%-15% of thoracic trauma cases but are often missed on chest radiographs, delaying diagnosis and treatment. Artificial intelligence (AI) may improve detection and triage in emergency settings.

Objective: This study aims to evaluate diagnostic accuracy, processing speed, and technical feasibility of an artificial intelligence-assisted rib fracture detection system using prospectively collected data within a real-world, high-volume emergency department workflow.

Methods: We conducted an observational feasibility study with prospective data collection of a faster region-based convolutional neural network-based AI model deployed in the emergency department to analyze 23,251 real-world chest radiographs (22,946 anteroposterior; 305 oblique) from April 1 to July 2, 2023. This study was approved by the Institutional Review Board of MacKay Memorial Hospital (IRB No. 20MMHIS483e). AI operated passively, without influencing clinical decision-making. The reference standard was the final report issued by board-certified radiologists. A subset of discordant cases underwent post hoc computed tomography review for exploratory analysis.

Results: AI achieved 74.5% sensitivity (95% CI 0.708-0.780), 93.3% specificity (95% CI 0.930-0.937), 24.2% positive predictive value, and 99.2% negative predictive value. Median inference time was 10.6 seconds versus 3.3 hours for radiologist reports (paired Wilcoxon signed-rank test $W=112\ 987.5$, $P<.001$). The analysis revealed peak imaging demand between 08:00 and 16:00 and Thursday-Saturday evenings. A 14-day graphics processing unit outage underscored the importance of infrastructure resilience.

Conclusions: The AI system demonstrated strong technical feasibility for real-time rib fracture detection in a high-volume emergency department setting, with rapid inference and stable performance during prospective deployment. Although the system showed high negative predictive value, the observed false-positive and false-negative rates indicate that it should be considered a supportive screening tool rather than a stand-alone diagnostic solution or a replacement for clinical judgment. These findings support further clinician-in-the-loop studies to evaluate clinical feasibility, workflow integration, and impact on diagnostic decision-making. However, interpretation is limited by reliance on radiology reports as the reference standard and the system's passive, non-interventional deployment.

JMIR Med Inform 2026;14:e77965; doi: [10.2196/77965](https://doi.org/10.2196/77965)

Keywords: artificial intelligence; chest radiograph; clinical decision support; digital health; emergency radiology; feasibility study; faster R-CNN; real-world data; rib fracture

Introduction

Digital health technologies, particularly artificial intelligence (AI), are increasingly used to address diagnostic delays in high-acuity clinical settings. In emergency departments, timely identification of injuries is essential, yet radiographic interpretation remains constrained by heavy workloads and the inherent complexity of imaging—especially for subtle findings such as rib fractures.

Rib fractures are a frequent consequence of thoracic trauma, occurring in 10%-15% of trauma patients and often indicating more serious underlying injuries [1-3]. When missed, they may lead to inadequate pain management, delayed respiratory support, pneumonia, or even preventable intensive care unit admissions. Beyond clinical harm, undetected fractures also carry medicolegal implications and increase health care costs.

Despite their significance, rib fractures are notoriously difficult to detect on chest radiographs (CXR)—the first-line imaging modality in most emergency departments—due to overlapping anatomical structures and subtle fracture lines. Reported sensitivities for radiologist detection can be as low as 15%, with up to half of fractures potentially missed in high-volume settings [4,5]. Although computed tomography (CT) and ultrasound can improve accuracy, they are resource-intensive and not always feasible for frontline triage [6-8]. These limitations highlight an urgent need for AI-driven tools that can assist clinicians by rapidly identifying suspected rib fractures in routine CXRs, enabling more effective prioritization and timely intervention.

Recent advances in AI, particularly deep learning, have demonstrated strong potential in automating image analysis tasks across medical domains, including dermatology, ophthalmology, and pulmonary imaging [9-11]. Deep learning models, especially convolutional neural networks, can automatically extract complex image features and have shown superior performance compared to traditional machine learning methods in various image classification tasks [11-13]. Transfer learning further enables the adaptation of pretrained convolutional neural networks—originally developed for natural images—for medical image classification tasks, including bone fracture detection [14,15].

Although prior studies have applied deep learning to rib fracture detection with promising results, most were retrospective, limited in scale, and did not assess feasibility in operational emergency department workflows [7,16,17].

These proof-of-concept efforts did not address the practical barriers to integrating AI into emergency radiology workflows, such as inference latency, system interoperability, or artifact handling.

To address this gap, we conducted an observational feasibility study with prospective data collection, evaluating an AI model for rib-fracture detection on CXRs. The system was passively deployed in parallel with routine emergency department imaging workflows using real-world data, without influencing clinical decisions. This design allowed the assessment of diagnostic performance, processing speed, and operational characteristics within standard clinical workflows.

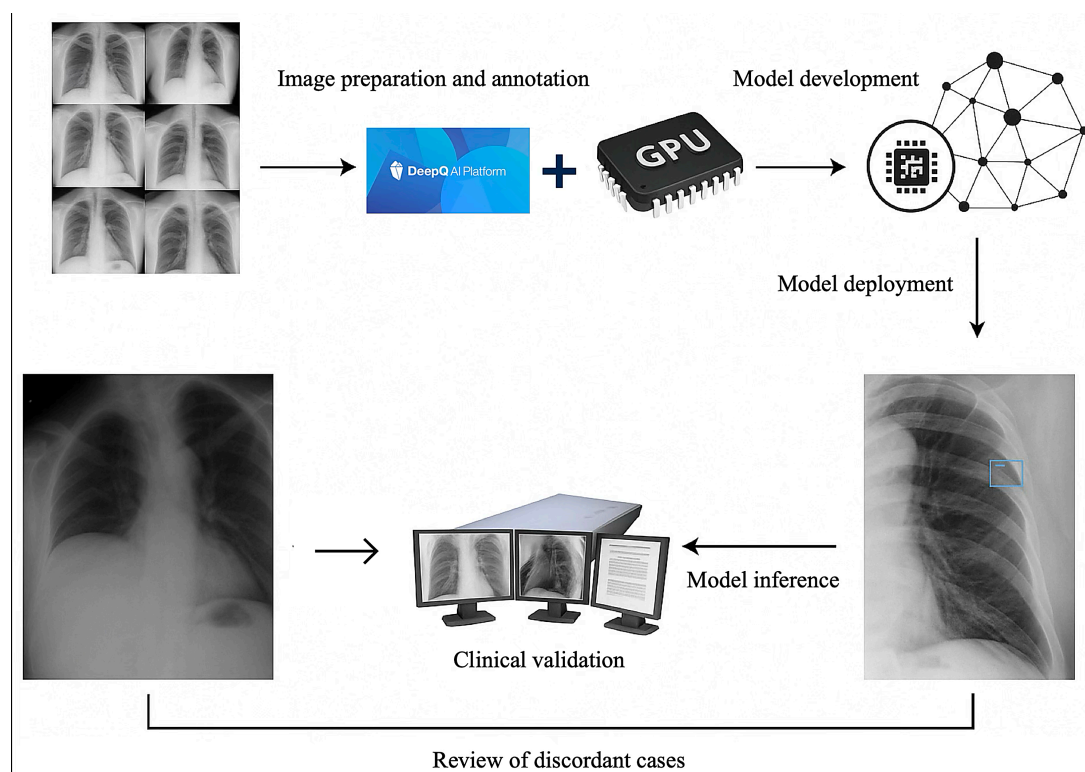
Methods

Study Design

The observational feasibility study protocol was reviewed and approved by the Institutional Review Board of MacKay Memorial Hospital (IRB No. 20MMHIS483e) prior to the initiation of data collection. MacKay Memorial Hospital is a tertiary referral and level 1 trauma center in northern Taiwan. The AI system functioned passively in real time without influencing clinical decisions or patient management. As the system functioned in a noninterventional, observational manner, prospective trial registration was not required.

From April 1 to July 2, 2023, all chest and rib radiographs acquired in the emergency department were automatically processed by the AI system in near-real-time. Both standard CXRs and rib-only views acquired during the study period were automatically analyzed, as both modalities are routinely used for suspected thoracic trauma. During the study period, a temporary 14-day graphics processing unit (GPU) hardware outage occurred, during which radiographs were not processed in real time; these examinations were excluded from turnaround time analysis but retained for diagnostic accuracy, as formal radiology reports were available. No additional exclusion criteria were applied beyond the 14-day system outage; all eligible emergency radiographs during the study period were included in the analysis. The system operated passively alongside routine clinical workflows, without influencing clinical decisions. AI-identified suspected rib fractures were highlighted using bounding boxes on a backend interface, which was accessible only for research evaluation and remained hidden from the clinical care team (Figure 1).

Figure 1. Study workflow of a prospective observational feasibility study evaluating an artificial intelligence–assisted rib fracture detection system using chest radiographs in patients admitted to the emergency department at a high-volume tertiary medical center (April 1–July 2, 2023). GPU: graphics processing unit.



Ethical Considerations

This study was approved by the Institutional Review Board of MacKay Memorial Hospital (IRB No. 20MMHIS483e) and conducted in accordance with the Declaration of Helsinki. Informed consent was waived because the study involved secondary analysis of routinely acquired, deidentified clinical imaging data, and the AI system operated passively without influencing patient management.

All data were deidentified prior to analysis and processed on secure institutional servers with access limited to authorized research personnel. No compensation was provided to participants. All images included in the manuscript were fully anonymized, and no identifiable patient information is disclosed.

Consistent with the approved observational study design, all AI outputs—including discordant cases—were withheld from treating clinicians and did not influence patient management.

AI Model Development

CXRs were retrospectively collected from the hospital picture archiving and communication system (PACS) for model training. All images were deidentified and preprocessed using histogram equalization and image inversion to improve fracture conspicuity. Fracture locations were annotated using bounding boxes by a board-certified emergency physician with 18 years of clinical experience via the DeepQ AI platform [18].

A deep learning model was developed using PyTorch (v1.13) with GPU acceleration. The architecture was based on faster region-based convolutional neural network (R-CNN) [19], incorporating a ResNet-50 backbone for feature extraction, a region proposal network for candidate region generation, and a classification head for fracture detection.

The dataset comprised 2079 CXRs (1065 fracture-positive and 1014 normal) collected between 2010 and 2020. Images were randomly divided into training (80%) and validation (20%) sets at the image level, as each radiograph represented an independent study. When multiple images were obtained from the same patient encounter, each radiograph was treated as an independent sample. Data augmentation—including random rotation, flipping, brightness, and contrast adjustment—was applied to improve generalization. To address the inherent class imbalance given the low fracture prevalence, class-weighted loss and oversampling of fracture-positive images were employed.

Model Validation

Model performance was assessed on a hold-out test set of 262 CXRs containing 724 expert-annotated rib fractures. Evaluation metrics were reported at both the case and object levels.

At the case level, the unit of analysis was the radiographic study. A study was considered positive if at least 1 rib fracture was detected, regardless of the number of fractures present. The model correctly identified fractures in 230 of 257 fracture-positive studies, achieving a sensitivity of 89.5%.

With only 8 false-positive cases, precision reached 96.6%, yielding an overall F_1 -score of 0.93.

At the object level, performance reflected per-lesion detection accuracy. The model correctly localized 680 of 724 annotated rib fractures (recall=94.0%) and generated 55 false-positive boxes. The mean average precision at an intersection-over-union threshold of ≥ 0.5 was 0.65, indicating robust lesion-level localization (Table 1).

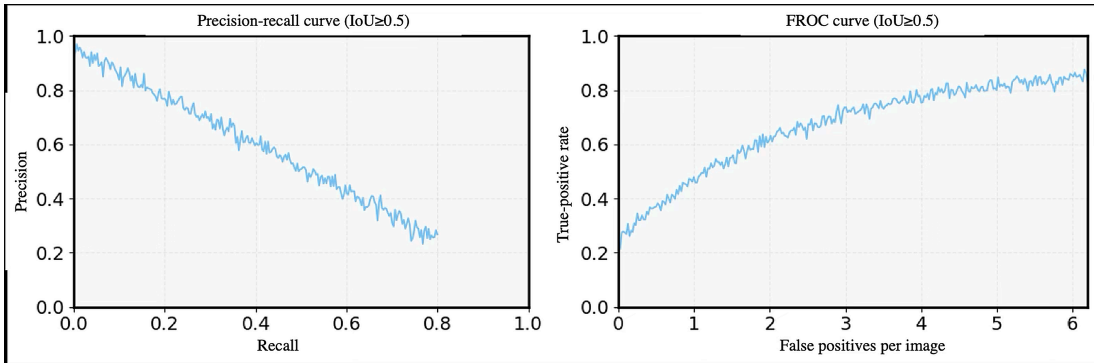
Table 1. Dataset composition and performance of the artificial intelligence (AI) model using retrospective emergency department chest radiographs, including case-level detection and object-level localization (intersection-over-union [IoU] ≥ 0.50).

Category and metric	Value
Dataset	
Total images	262
Ground-truth boxes	724
Case-level detection (%)	
Sensitivity (recall)	89.6
Precision	96.6
F_1 -score	0.93
Object-level localization	
Recall	94.0%
mAP ^a (IoU ≥ 0.5)	0.65

^amAP: mean average precision.

Curve-based analyses further characterized the model’s detection behavior (Figure 2). The precision-recall curve (Figure 2A) maintained precision ≥ 0.90 until recall fell below 0.55 (precision-recall area under the curve=0.65), demonstrating high reliability across a broad sensitivity range. The free-response receiver operating characteristic (FROC) curve (Figure 2B) showed true-positive rates of 0.77 at 1 false positive per image and 0.88 at three, representing practical trade-offs between sensitivity and alert frequency in potential clinical deployment.

Figure 2. The performance of the artificial intelligence (AI)-assisted rib fracture detection model was evaluated in the retrospective model development and validation dataset using emergency department chest radiographs. (A) Precision-recall curve demonstrating case-level detection performance at an intersection-over-union (IoU) threshold of ≥ 0.50 (area under the curve=0.65). (B) Free-response receiver operating characteristic (FROC) curve showing lesion-level sensitivity as a function of false positives per image, with localization performance assessed at the same IoU threshold (mean average precision=0.65; recall=94%).



Prospective Evaluation of AI Model in Emergency Department Workflow

The trained model was prospectively evaluated in parallel with clinical workflow, performing automated inference on incoming emergency radiographs. During the automated inference process, all incoming radiographs were standardized and resized to a fixed resolution of 512×512 pixels. The prospective evaluation cohort (April-July 2023) was temporally and operationally independent from the retrospective training and validation dataset (2010-2020), and no patient overlap existed between the 2 cohorts. During the prospective evaluation phase, all chest and rib radiographs from emergency department encounters were automatically

processed by the AI system without disrupting clinical workflows. The bounding box outputs were logged for research analysis but were not disclosed to radiologists or used for patient management.

Performance Assessment Using NLP-Derived Labels

To evaluate AI performance in the real-world setting, output was compared to formal radiology reports issued by board-certified radiologists. A rule-based natural language processing (NLP) pipeline was developed to extract structured rib fracture labels (positive, negative, or ambiguous) from free-text reports. The algorithm combined keyword

detection (eg, “rib fracture,” “fx”) and negation handling (eg, “no evidence of,” “no definite fracture”).

To validate NLP accuracy, a random sample of 200 radiology reports was manually reviewed by 2 emergency physicians blinded to both NLP and AI results. The NLP

classification achieved 96.5% agreement (193/200) with manual review, with a Cohen κ of 0.91, indicating excellent concordance. Most discrepancies were due to ambiguous language or complex negation structures (Table 2).

Table 2. Confusion matrix comparing natural language processing (NLP)–extracted radiology report labels with manual expert review in a randomly selected subset of 200 emergency department chest radiographs, used to assess labeling accuracy in the retrospective dataset.

Manual review	Fracture present	Fracture absent	Ambiguous	Total (NLP prediction)
NLP: fracture present	95	2	1	98
NLP: fracture absent	3	90	2	95
NLP: ambiguous	1	2	4	7
Total (manual review)	99	94	7	200

Formal radiology reports served as the primary reference standard for AI evaluation. This approach may have underestimated AI sensitivity because CT confirmation was not performed systematically. In select discordant cases—where the AI flagged fractures not documented in the reports—subsequent CT scans confirmed some of these findings. These discrepancies were retrospectively reviewed to investigate potential underreporting by radiologists. While informative, these exploratory adjudications were not used as a universal reference standard due to inconsistent CT availability. Nonetheless, radiology reports remained the definitive benchmark for all performance metrics. Additionally, the selected misclassified cases were examined to identify recurring patterns of diagnostic oversight among frontline physicians.

Targeted Adjudication of Discordant Cases

To further explore potential underreporting within the report-based reference standard, we performed a focused review of discordant cases in which the AI system flagged suspected rib fractures not documented in the corresponding formal radiology reports. Because not all discordant cases had confirmatory CT imaging, an illustrative subset of 11 cases was selected based on the availability of same-encounter chest CT and clinical relevance for qualitative adjudication. Each case was reviewed to determine whether the AI-predicted fractures corresponded to true fractures confirmed on CT. These adjudications were exploratory and intended to contextualize the potential clinical value of AI detection beyond the report-based benchmarking.

Data Analysis

Case-level performance was assessed at the radiographic study (accession) level by comparing AI outputs to NLP-derived labels from radiology reports. A study was

considered positive if at least 1 image within the same examination was flagged as having a rib fracture by the AI system; otherwise, the study was classified as negative. Key metrics included sensitivity, specificity, accuracy, positive predictive value, negative predictive value (NPV), and F_1 -score. Ninety-five percent CIs were calculated using nonparametric bootstrap resampling (1000 iterations). The results were summarized in confusion matrices and diagnostic performance plots.

Statistical Analysis

All statistical analyses were performed using Python 3.11 (pandas v2.2, scikit-learn v1.4) and R 4.3.2. Continuous variables were reported as mean (SD) or median with IQR. Categorical variables were summarized as counts and percentages. A 2-tailed $P<.001$ was considered statistically significant.

This study was reported in accordance with the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) guideline, with the completed checklist provided as Checklist 1.

Results

Study Cohort

From April 1 to July 2, 2023, all chest and dedicated rib radiographs acquired in the emergency department were automatically processed by the AI system in a parallel workflow, yielding 23,251 imaging studies from 20,908 unique patient visits. Population demographics are summarized in Table 3. The mean age was 55.9 years (SD 22.3; range 0–106), with 10,770 (51.5%) male and 10,138 (48.5%) female patients. A radiologist review identified 589 rib-fracture cases (prevalence 2.8%).

Table 3. Demographic and clinical characteristics of patients in the emergency department included in a prospective observational study of artificial intelligence (AI)–assisted rib fracture detection (April 1–July 2, 2023).

Characteristic	Value
Total cases	20,908
Age (y), mean (SD; range)	55.9 (22.3; 0–106)

Characteristic	Value
Sex, n (%)	
Male	10,770 (51.5)
Female	10,138 (48.5)
Radiologist-confirmed rib fractures, n (%)	589 (2.8)

AI Model Performance

AI model outputs were compared on a per-case basis against structured rib-fracture labels derived from board-certified radiology reports. At the selected operating point—corresponding to approximately one false-positive per image on the FROC curve—the system achieved a sensitivity of 0.745

(95% CI 0.708-0.780) and specificity of 0.933 (95% CI 0.930-0.937). Positive predictive value was 0.242 (95% CI 0.223-0.262), and negative predictive value was 0.992 (95% CI 0.991-0.994). The overall F_1 -score was 0.365 (95% CI 0.340-0.390) with an accuracy of 0.928 (Table 4).

Table 4. Case-based diagnostic performance of an artificial intelligence–assisted rib fracture detection system in a prospective observational emergency department study^a.

Metric	Estimate (95% CI)
Sensitivity	0.745 (0.708-0.780)
Specificity	0.933 (0.930-0.937)
PPV ^b	0.242 (0.223-0.262)
NPV ^c	0.992 (0.991-0.994)
F_1 -score	0.365 (0.340-0.390)
Accuracy	0.928 (N/A ^d)

^aPerformance metrics are reported with 95% CIs using final radiologist reports as the reference standard.

^bPPV: positive predictive value.

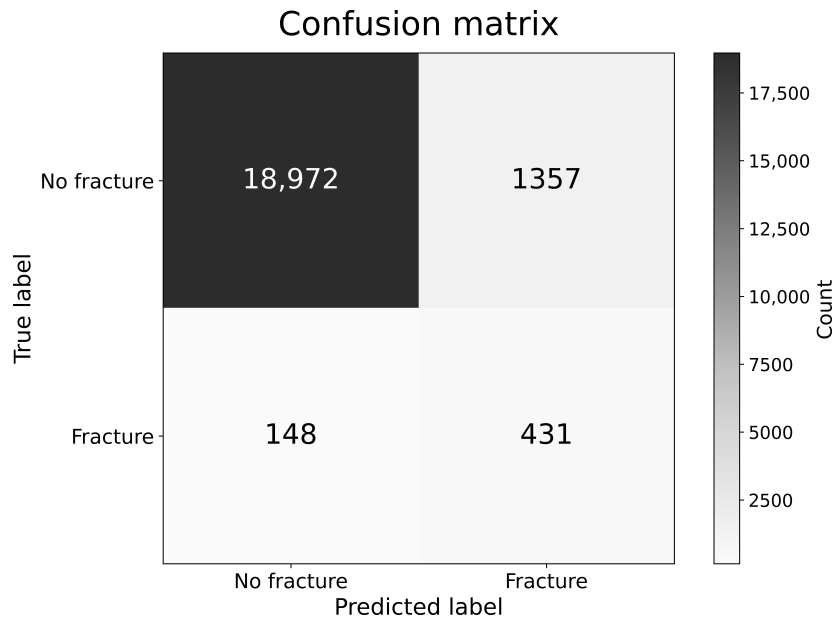
^cNPV: negative predictive value.

^dN/A: not available.

As shown in Figure 3, the AI system correctly identified 431 (74.5%) fracture-positive cases while producing 1357 (6.1%) false positives and 148 (0.7%) false negatives across 23,251 studies. This distribution demonstrates the model’s high true-negative count (n=18,972, 93.3%) and its strong

negative predictive value during deployment in the emergency department. No temporal drift or learning curve effects were observed, as the deployed model remained fixed throughout the study period.

Figure 3. Case-level confusion matrix of the artificial intelligence (AI)–assisted rib fracture detection system during prospective emergency department deployment, using final radiologist reports as the reference standard. Darker blue indicates a higher number of cases (count), as shown in the color bar.

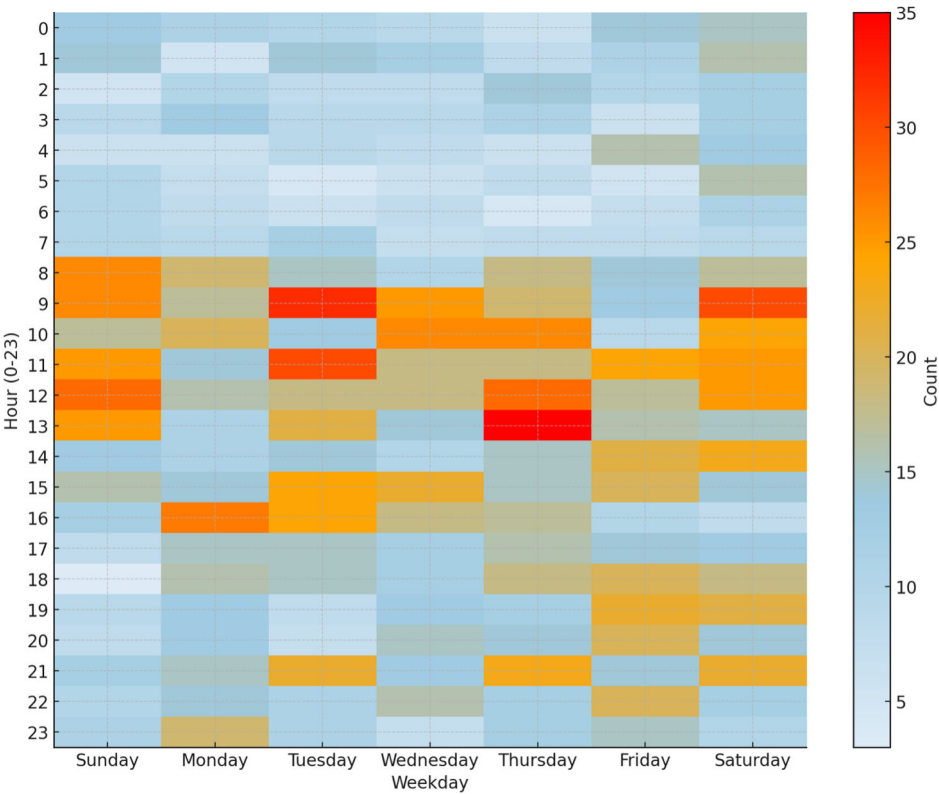


Imaging Workload Patterns

Analysis of imaging demand revealed predictable diurnal and weekly patterns, with peak volumes between 08:00 and

16:00 daily and secondary surges on Thursday to Saturday evenings. Demand was the lowest between 00:00 and 07:00 across all days of the week (Figure 4).

Figure 4. Heatmap illustrating the temporal distribution of chest radiograph imaging workload by hour of day and weekday during prospective emergency department deployment. Color intensity represents the number of chest radiographs acquired per hour, highlighting peak imaging periods across weekdays and weekends.



Inference Turnaround Time

A total of 19,641 paired cases were included to compare AI inference and radiologist report turnaround times. As shown in Table 5, the AI system achieved a median processing time of 10.6 seconds per image (IQR 9.0-14.0; range 3-35 s), compared with a median of 3.3 hours (IQR 1.31-4.80; range 0.08-72 h) for radiologist reports. This represents a more than

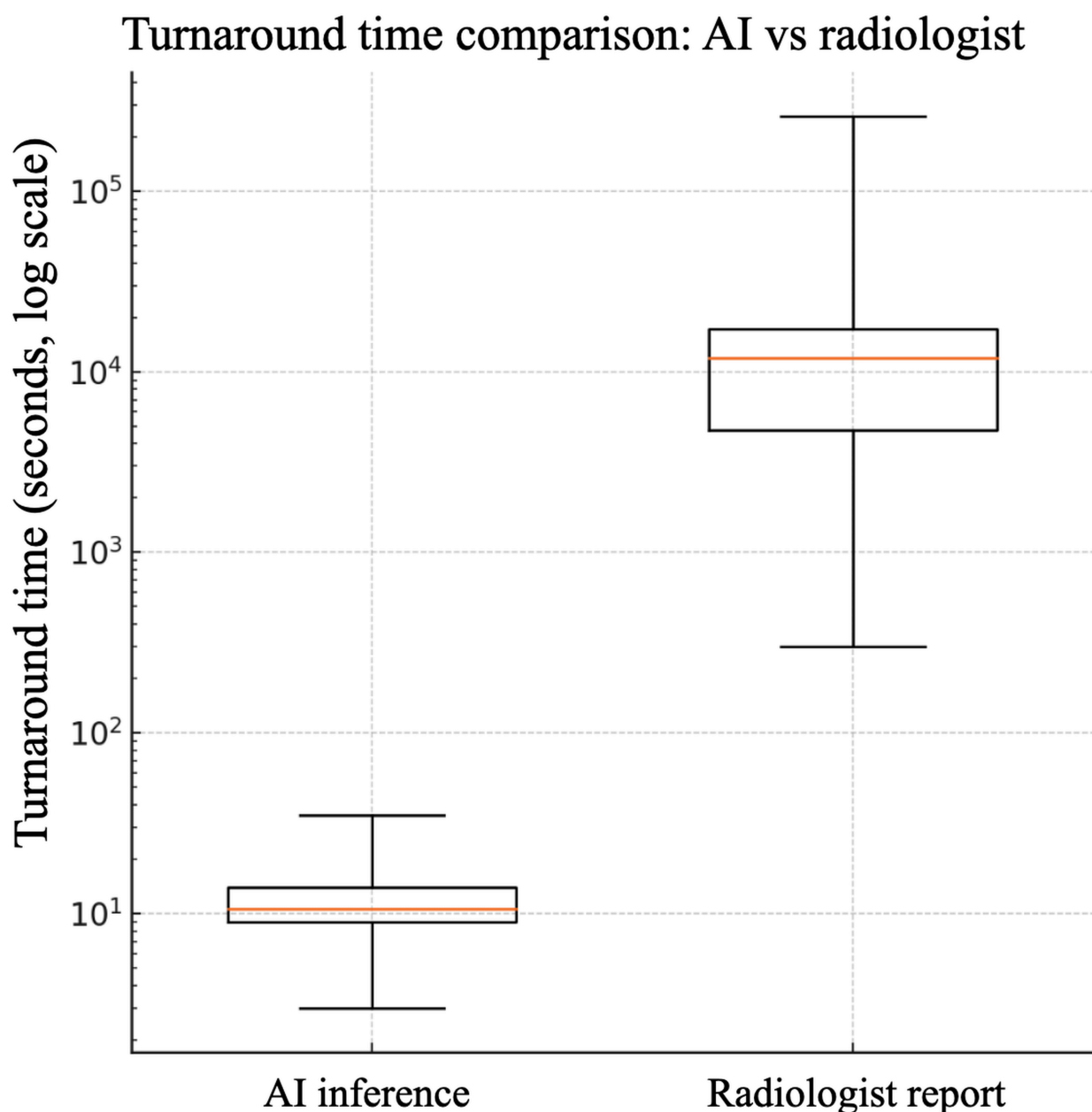
1000-fold reduction in turnaround time. Figure 5 illustrates this disparity using boxplots on a logarithmic scale. A paired Wilcoxon signed-rank test confirmed that AI inference was significantly faster than radiologist reporting ($W=112,987.5$; $P<.001$). This median reporting time reflects the full clinical workflow, including overnight and backlog delays typical of high-volume emergency radiology settings.

Table 5. Turnaround times for artificial intelligence (AI) inference versus radiologist reporting during prospective emergency department deployment (n=19,641)^a.

Metric	AI inference time (s)	Radiologist report time (s)
Mean	10.9	10,877
Standard deviation	3.0	4008
Median (50%)	10.6	11,880
IQR (25%-75%)	9.0-14.0	4728-17,280
Min	3.0	300
Max	35.0	259,200

^aTimes are summarized using descriptive statistics, including median and interquartile range.

Figure 5. Comparison of turnaround times between artificial intelligence (AI) inference and radiologist reporting during prospective emergency department deployment (April 1-July 2, 2023). Boxplots on a logarithmic scale illustrate differences in processing time distributions between real-time AI inference and routine clinical reporting.



This processing speed highlights the potential of AI-assisted triage systems to complement radiology workflows by rapidly identifying cases for prioritized review, especially in high-volume emergency settings.

System Reliability

During the prospective evaluation of the AI system operating in a parallel clinical workflow, the AI platform experienced 1 service interruption—from 27 April to 10 May 2023—caused by GPU overload that halted all inference operations. A total of 3610 studies acquired during this 14-day outage were excluded from the turnaround time analysis but retained in the diagnostic accuracy evaluation (their formal radiology reports remained available). Following hardware replacement, full functionality was restored on 11 May, and no further outages occurred over the remainder of the study period.

Illustrative Review of Discordant Cases

To further evaluate the AI system's potential diagnostic value beyond report-based benchmarking, a targeted adjudication was conducted on 11 representative discordant cases in which the AI system flagged suspected rib fractures not described in the corresponding radiology reports (Table 6). Among these, 7 cases (cases 1-7) were subsequently confirmed as true fractures on CT ("AI-CT concordant"), indicating that several AI-labeled false positives in the report-based analysis represented true fractures missed in the reference standard. The remaining 4 cases (cases 8-11) were confirmed negative on CT, primarily attributable to nonfracture anatomical structures or imaging artifacts. Including these CT-confirmed fractures as true positives would modestly increase the model's effective sensitivity and positive predictive value,

highlighting the underestimation inherent in report-based benchmarking.

Table 6. Targeted post hoc computed tomography (CT) adjudication of representative discordant cases in which artificial intelligence (AI) flagged suspected rib fractures not described in the corresponding radiology reports during prospective emergency department deployment.

Case	AI output	Radiologist report	Emergency physician	Outcome	Note (AI significance)
1	Flagged right rib fracture	No fracture	Noted with POCUS ^a	CT-confirmed	Triage value, prompting clinicians to perform US ^b
2	Flagged right rib fracture	No fracture	Missed	CT-confirmed	AI-CT concordance
3	Flagged left fifth rib fracture	No fracture	Missed	CT-confirmed	AI-CT concordance
4	Flagged rib fracture post chest tube	No fracture	Missed	CT-confirmed	Chest tube artifact did not impair detection
5	Flagged lower-rib fracture	No fracture	Missed	CT-confirmed	AI-CT concordance
6	Flagged fracture near hardware	No fracture	Missed	CT-confirmed	Detected fracture adjacent to surgical hardware
7	Flagged fracture under scapular shadow	No fracture	Missed	CT-confirmed	AI-CT concordance
8	Flagged fracture at scapula border	No fracture	No fracture	False positive	Scapular margin misidentified
9	Flagged rib fracture at bra clasp	No fracture	No fracture	False positive	Bra hardware artifact
10	Flagged fracture at chest tube marker	No fracture	No fracture	False positive	Chest tube marker misinterpreted
11	Flagged fracture (image noise)	No fracture	No fracture	False positive	Image noise

^aPOCUS: point-of-care ultrasound.

^bUS: ultrasound.

In one representative case (Case 3), the AI system correctly identified a subtle nondisplaced fracture of the left fifth rib that was not documented in the radiology report but later verified on 3D CT reconstruction (Figure

6). In contrast, Figure 7 illustrates the main sources of false positives, including scapular margin misinterpretation, chest-tube hardware artifacts, and motion-induced noise.

Figure 6. Representative true-positive rib fracture detected by artificial intelligence (AI) and confirmed by computed tomography (CT) during prospective emergency department deployment. (A) Chest radiograph showing an AI-flagged fracture of the left fifth rib that was not described in the initial radiology report. (B) Corresponding CT image confirming the fracture at the same anatomical location.

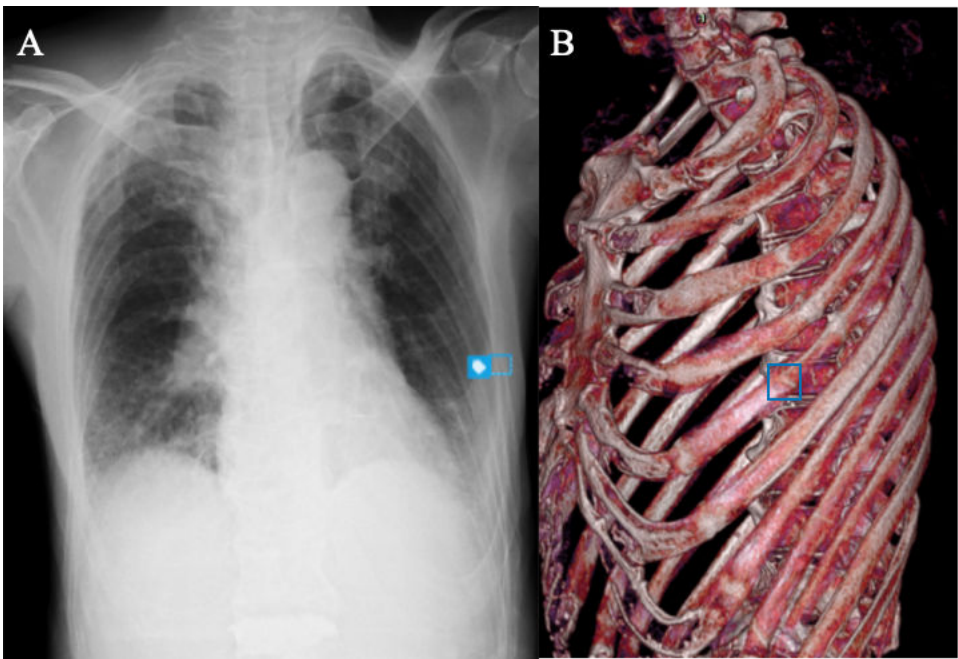
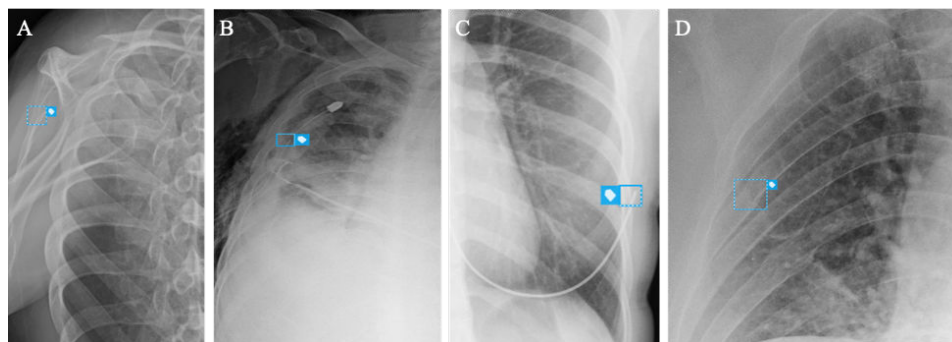


Figure 7. Representative false-positive detections generated by an artificial intelligence (AI)-assisted rib fracture detection system during prospective emergency department deployment. The examples illustrate common sources of false-positive signals on chest radiographs. (A) Scapular margin overlap misinterpreted as a rib fracture. (B) Chest tube marker misidentified as a rib discontinuity. (C) Bra hardware producing a linear opacity mimicking a fracture. (D) Image noise and low-contrast regions leading to spurious detection.



This targeted CT adjudication underscores the potential of AI-assisted screening to augment clinical vigilance by identifying subtle or overlooked fractures, while also emphasizing the need to improve artifact robustness and optimize false-positive suppression for practical clinical integration.

Common Pitfalls in Frontline Rib Fracture Detection

Our review of discordant cases identified 3 principal drivers of missed rib fractures by emergency physicians: first, non-thoracic presenting complaints (eg, catheter malfunction or abdominal pain) led interpreters to focus on unrelated findings and overlook subtle rib breaks; second, the absence of classic chest pain—patients describing only mild discomfort or a vague “pop”—lowered clinical vigilance for nondisplaced fractures; and third, competing urgent injuries (facial, limb, or soft-tissue trauma) diverted attention from the chest, resulting in underappreciated fractures.

Discussion

Principal Findings

In this observational feasibility study using real-world emergency department imaging data, we demonstrated that a faster R-CNN-based AI system can operate in parallel with routine clinical workflows to provide near-instantaneous rib fracture triage without influencing patient care. During a 3-month evaluation period, the model automatically processed 23,251 CXRs with a median inference time of 10.6 seconds per image, achieving a turnaround time reduction exceeding 3 orders of magnitude compared with formal radiologist reporting, while maintaining 74.5% sensitivity and 93.3% specificity. These findings position AI as a potential automated screening aid capable of rapidly identifying low-risk examinations and generating signals that could inform future prioritization strategies. The observed discrepancy between the 10.6-second AI inference time and the 3.3-hour radiologist turnaround time reflects a critical clinical bottleneck in busy emergency departments. Although these metrics represent different stages—technical processing versus final clinical documentation—the delay in official reporting highlights the diagnostic gap AI aims to address.

In this context, near-instantaneous AI alerts may support case prioritization before formal reporting, although no clinician-facing alerts were implemented in this study.

However, the current findings primarily demonstrate technical feasibility rather than full clinical feasibility, as the AI system operated passively without direct clinician interaction. Future clinician-in-the-loop evaluations will be necessary to assess workflow integration, usability, and impact on diagnostic behavior or patient outcomes.

Although the system's positive predictive value was relatively low (24.2%), this trade-off aligns with its design as a triage support tool rather than a stand-alone diagnostic system. In high-volume emergency departments, the ability to rapidly identify examinations with a low likelihood of fracture is crucial. The model's high NPV of 99.2% allows clinicians to focus on a smaller, higher-risk subset of cases, thereby improving efficiency and reducing cognitive load. Given the observed sensitivity of 74.5%, false-negative cases remain possible, and the system should not be used as a stand-alone rule-out mechanism or as a substitute for clinical judgment.

Comparison to Prior Work

Similar findings have been reported by Yao et al [20], who demonstrated that deep learning systems with high NPV on chest CT can reduce radiologists' workload by effectively identifying nonfracture cases. A recent systematic review by van den Broek et al [21] further emphasized the triage value of AI in fracture detection, underscoring its potential across multiple imaging modalities.

Our AI system also demonstrated robust performance across circadian and weekly imaging surges, with peak volumes observed between 08:00 and 16:00 and during Thursday to Saturday evenings. However, a 14-day GPU hardware outage during the study period highlighted a real-world challenge of maintaining the AI system's reliability in clinical environments. This incident underscores the need for infrastructure redundancy, real-time monitoring, and failover protocols—key considerations for sustainable AI deployment. These practical aspects of AI deployment remain underreported in most published studies [22].

Compared with prior approaches such as the PACS-AI platform [23], our system offered full automation, operating continuously in real time and without the need for manual image selection. This better reflects the demands of frontline emergency radiology. Herpe et al [24] demonstrated improved diagnostic accuracy with PACS-integrated AI for limb fractures; however, their study did not evaluate scalability or autonomous triage capability under high-throughput conditions. In contrast, our study incorporated prospective data collection within a real-world emergency department workflow, allowing the assessment of AI performance, reliability, and operational feasibility under authentic clinical conditions.

Focused adjudication of discordant cases revealed that the AI system correctly identified rib fractures that were missed by both radiologists and emergency physicians in 7 cases, all subsequently confirmed on follow-up CT ("AI-CT concordant"). These findings highlight the potential of AI to strengthen diagnostic vigilance in complex clinical scenarios. While prior studies—such as Zhou et al [25]—have shown that AI can detect rib fractures overlooked in initial CT interpretations, with confirmation on follow-up imaging, these investigations have largely centered on CT-based workflows. Although CT is highly sensitive, its routine use is limited by concerns over radiation, cost, and logistics. In contrast, our CXR-focused approach targets the most widely used imaging modality in acute care, offering a more scalable and practical solution for real-world emergency triage. Notably, Brady et al [26] have emphasized that diagnostic errors and discrepancies are not uncommon in radiology, with daily error rates estimated at 3%-5%, reinforcing the importance of AI as a complementary tool to enhance diagnostic accuracy.

Four false-positive cases revealed predictable pitfalls, including the misinterpretation of scapular margins, chest-tube hardware, and motion artifacts. Similar findings were reported by Sun et al [27], who noted frequent false positives in an AI model for rib fracture detection on CXR, often due to anatomical overlap and imaging artifacts. These results support the need for artifact-aware retraining and preprocessing optimization to reduce false alerts and improve clinical integration.

This targeted CT adjudication further highlights the complementary role of AI in identifying subtle or overlooked fractures and underscores the inherent limitation of using report-based labels as the reference standard in real-world studies.

Future Directions

In addition, integrating AI-generated alerts into emergency radiology workflows will require careful calibration of alert thresholds to minimize false positives and prevent alert fatigue among clinicians. Human-centered design, interface refinement, and iterative feedback from end users will be critical to achieving effective and sustainable adoption.

While most prior prospective studies have emphasized diagnostic performance or radiologist feedback, our findings

extend beyond these metrics to include diagnostic efficacy, operational resilience, and system failure contingencies. These real-world insights support the feasibility and clinical value of embedding AI into routine emergency department workflows. Recent work has highlighted the importance of not only measuring accuracy but also assessing robustness across patient and workflow variability [28]. Furthermore, the need for deployment frameworks that address hardware resilience, continuous quality monitoring, and interpretability safeguards is increasingly recognized as essential for sustainable AI adoption in high-acuity settings [29].

Recent reports and position statements have highlighted a persistent gap between the promising diagnostic performance of AI systems and their limited demonstrated clinical benefit. Robust, prospective, and randomized clinical studies remain urgently needed to justify large-scale implementation [30,31]. Even high-performing AI models (area under the curve \approx 0.85) have failed to surpass standard clinical practice in improving patient outcomes [32,33]. These findings reinforce ongoing concerns that most AI or machine learning devices, despite regulatory authorization, are primarily validated using retrospective data and therefore remain susceptible to selection bias, distributional shift, and overestimation of generalizability [34,35].

Clinical decision-making in emergency care is inherently multimodal: physicians integrate imaging findings with the mechanism of injury, examination, and vital signs to guide judgment. In contrast, this AI system analyzes images in isolation and is designed not to replace but to support clinicians as a rapid screening aid—enhancing vigilance in high-volume, high-pressure environments where missed fractures may occur. Incorporating multimodal clinical data in future models could further improve diagnostic relevance and workflow integration.

Although this was an observational feasibility study, it represents one of the largest evaluations of an AI-assisted rib fracture detection system in real-world emergency radiology. The findings demonstrate that such a system can provide meaningful diagnostic support, maintain consistent performance at scale, and potentially enhance patient safety. Future implementation should therefore shift from technical to clinical feasibility, focusing on clinician-in-the-loop impact studies, PACS-integrated trials, and workflow efficiency assessments. Although user perception was not formally assessed, informal feedback from emergency physicians indicated strong interest in AI-supported flagging—particularly for subtle fractures and during periods of high patient volume.

Future research should prioritize prospective, multicenter studies to validate generalizability and quantify AI impact on workflow, resource utilization, and patient outcomes. Model improvements—including artifact-aware retraining, expanded fracture coverage in challenging scenarios such as subtle or anatomically obscured fractures, and continuous learning—will be critical to enhance diagnostic precision. Finally, building infrastructure resilience and integrating effective

alert management into radiologist workflows are essential for sustainable clinical adoption.

Limitations

First, this single-center observational study may limit generalizability to other institutions with different imaging protocols, patient populations, or workflow environments. Second, during retrospective model development, training and validation were performed at the image level rather than the patient level. This may have resulted in optimistic internal validation estimates due to potential within-patient similarity, which likely contributes to the observed performance gap between retrospective validation and prospective real-world deployment. Additionally, stratified performance analyses by age group and sex were not performed due to the low prevalence of rib fractures in certain subgroups, particularly pediatric patients. Similarly, the small fraction of oblique views (approximately 1.3%) prevented a dedicated analysis by imaging view, as the limited sample size would yield unstable estimates for these specific cohorts.

The prospective evaluation period also did not include winter months. Seasonal variation in trauma mechanisms or imaging artifacts may influence fracture detectability in certain settings; therefore, caution is warranted when generalizing these findings across different seasonal contexts.

Additionally, the use of a 512×512 resolution for model inference represents a technical trade-off; while it facilitates rapid processing, the associated downsampling may limit the detection of very subtle cortical disruptions.

Third, using radiology reports as the reference standard—while pragmatic—may underestimate the AI system's true performance, as subtle or occult fractures can be underreported in clinical practice. A focused CT review of representative discordant cases further supported this concern, revealing instances where AI-predicted fractures were subsequently confirmed as true fractures on CT. This approach likely yielded conservative performance estimates, since NLP-derived labels may not capture subtle fractures identified by AI or CT.

Regarding the study design, although post hoc adjudication is generally more appropriate for hypothesis generation

than for definitive performance reassessment, modifying the reference standard after study completion may introduce bias. Accordingly, in this study, performance evaluation was anchored to the contemporaneous clinical reference standard used in routine practice. More comprehensive adjudication strategies—such as consensus radiologist review of AI-positive, report-negative cases—may provide additional insights when implemented within a separately designed study.

Finally, because AI predictions were not disclosed to clinicians, we did not assess downstream clinical outcomes, including changes in diagnostic behavior, time to intervention, or patient management. As for the data pipeline, although we used NLP to extract rib fracture labels from radiology reports, which may introduce misclassification in ambiguous cases, the pipeline demonstrated high agreement with manual review ($\kappa=0.91$). Given the observed 3.5% discrepancy rate, any residual label noise propagating through the large-scale dataset may introduce modest uncertainty into performance estimates. However, the substantial sample size ($n=23,251$) is expected to attenuate the impact of such noise, supporting the stability of the resulting confidence intervals for large-scale clinical benchmarking.

Conclusions

In this observational feasibility study, we evaluated a faster R-CNN-based AI system deployed in parallel with clinical workflows to automatically detect rib fractures on CXRs using real-world emergency department data. Although AI outputs were not visible to clinicians, the system processed over 23,000 studies with high throughput, achieving 74.5% sensitivity and 93.3% specificity and delivering results within seconds—over 1000 times faster than formal radiologist reports.

These findings demonstrate strong technical feasibility of real-time AI-assisted rib fracture detection in emergency radiology. While clinical decisions remained unaffected during this observational phase, future studies should validate clinical feasibility through clinician-in-the-loop evaluation, PACS integration, and workflow optimization to address potential alert fatigue and false-positive management.

Acknowledgments

We thank the emergency department, radiology department, and information technology teams for their support during system deployment and data extraction. Generative artificial intelligence tools were used solely for language editing and grammar checking. All scientific content, analyses, interpretations, and conclusions were developed by the authors. The artificial intelligence system evaluated in this study was deployed on the DeepQ AI platform as a technical deployment environment. The authors have no financial, employment, or equity relationship with DeepQ Technology or HTC, and the platform was used solely for noncommercial research purposes.

Funding

This research was financially supported by National Science and Technology Council, Taiwan, under Grant No. NSTC113-2221-E-038-006

Data Availability

The imaging data analyzed in this study are not publicly available due to institutional regulations and patient privacy protections. Deidentified data may be made available from the corresponding author upon reasonable request and subject to institutional review board approval.

Authors' Contributions

Conceptualization: H-WC, S-TH

Methodology: H-WC, S-TH

Data curation: M-YH, S-TH

Formal analysis: M-YH, S-TH

Investigation: M-YH, S-TH

Writing – original draft: M-YH, S-TH

Writing – review & editing: H-WC, L-RL, M-FT, M-YH, S-TH

Supervision: CH-W

Conflicts of Interest

None declared.

Checklist 1

CLAIM (Checklist for Artificial Intelligence in Medical Imaging).

[\[PDF File \(Adobe File\), 4 KB-Checklist 1\]](#)

References

1. Tignanelli CJ, Rix A, Napolitano LM, Hemmila MR, Ma S, Kummerfeld E. Association between adherence to evidence-based practices for treatment of patients with traumatic rib fractures and mortality rates among US trauma centers. *JAMA Netw Open*. Mar 2, 2020;3(3):e201316. [doi: [10.1001/jamanetworkopen.2020.1316](https://doi.org/10.1001/jamanetworkopen.2020.1316)] [Medline: [32215632](https://pubmed.ncbi.nlm.nih.gov/32215632/)]
2. Colling KP, Goettl T, Harry ML. Outcomes after rib fractures: more complex than a single number. *J Trauma Inj*. Dec 2022;35(4):268-276. [doi: [10.20408/jti.2021.0096](https://doi.org/10.20408/jti.2021.0096)] [Medline: [39380934](https://pubmed.ncbi.nlm.nih.gov/39380934/)]
3. Sharma OP, Oswanski MF, Jolly S, Lauer SK, Dressel R, Stombaugh HA. Perils of rib fractures. *Am Surg*. Apr 2008;74(4):310-314. [doi: [10.1177/000313480807400406](https://doi.org/10.1177/000313480807400406)] [Medline: [18453294](https://pubmed.ncbi.nlm.nih.gov/18453294/)]
4. Griffith JF, Rainer TH, Ching AS, Law KL, Cocks RA, Metreweli C. Sonography compared with radiography in revealing acute rib fracture. *AJR Am J Roentgenol*. Dec 1999;173(6):1603-1609. [doi: [10.2214/ajr.173.6.10584808](https://doi.org/10.2214/ajr.173.6.10584808)] [Medline: [10584808](https://pubmed.ncbi.nlm.nih.gov/10584808/)]
5. Tomas X, Facenda C, Vaz N, et al. Thoracic wall trauma-misdiagnosed lesions on radiographs and usefulness of ultrasound, multidetector computed tomography and magnetic resonance imaging. *Quant Imaging Med Surg*. Aug 2017;7(4):384-397. [doi: [10.21037/qims.2017.08.02](https://doi.org/10.21037/qims.2017.08.02)] [Medline: [28932697](https://pubmed.ncbi.nlm.nih.gov/28932697/)]
6. Chapman BC, Overbey DM, Tesfalidet F, et al. Clinical utility of chest computed tomography in patients with rib fractures CT chest and rib fractures. *Arch Trauma Res*. Dec 2016;5(4):e37070. [doi: [10.5812/atr.37070](https://doi.org/10.5812/atr.37070)] [Medline: [28144607](https://pubmed.ncbi.nlm.nih.gov/28144607/)]
7. Huang ST, Liu LR, Chiu HW, Huang MY, Tsai MF. Deep convolutional neural network for rib fracture recognition on chest radiographs. *Front Med*. 2023;10. [doi: [10.3389/fmed.2023.1178798](https://doi.org/10.3389/fmed.2023.1178798)]
8. Pishbin E, Ahmadi K, Foogardi M, Salehi M, Seilanian Toosi F, Rahimi-Movaghar V. Comparison of ultrasonography and radiography in diagnosis of rib fractures. *Chin J Traumatol*. Aug 2017;20(4):226-228. [doi: [10.1016/j.cjte.2016.04.010](https://doi.org/10.1016/j.cjte.2016.04.010)] [Medline: [28687342](https://pubmed.ncbi.nlm.nih.gov/28687342/)]
9. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. Dec 13, 2016;316(22):2402-2410. [doi: [10.1001/jama.2016.17216](https://doi.org/10.1001/jama.2016.17216)] [Medline: [27898976](https://pubmed.ncbi.nlm.nih.gov/27898976/)]
10. Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv*. Preprint posted online on Nov 14, 2017. [doi: [10.48550/arXiv.1711.05225](https://doi.org/10.48550/arXiv.1711.05225)]
11. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. Feb 2, 2017;542(7639):115-118. [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)] [Medline: [28117445](https://pubmed.ncbi.nlm.nih.gov/28117445/)]
12. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. Dec 2017;42:60-88. [doi: [10.1016/j.media.2017.07.005](https://doi.org/10.1016/j.media.2017.07.005)] [Medline: [28778026](https://pubmed.ncbi.nlm.nih.gov/28778026/)]
13. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. May 2019;29(2):102-127. [doi: [10.1016/j.zemedi.2018.11.002](https://doi.org/10.1016/j.zemedi.2018.11.002)] [Medline: [30553609](https://pubmed.ncbi.nlm.nih.gov/30553609/)]
14. Kora P, Ooi CP, Faust O, et al. Transfer learning techniques for medical image analysis: a review. *Biocybern Biomed Eng*. Jan 2022;42(1):79-107. [doi: [10.1016/j.bbe.2021.11.004](https://doi.org/10.1016/j.bbe.2021.11.004)]
15. Olczak J, Fahlberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop*. Dec 2017;88(6):581-586. [doi: [10.1080/17453674.2017.1344459](https://doi.org/10.1080/17453674.2017.1344459)] [Medline: [28681679](https://pubmed.ncbi.nlm.nih.gov/28681679/)]

16. Wu J, Liu N, Li X, et al. Convolutional neural network for detecting rib fractures on chest radiographs: a feasibility study. *BMC Med Imaging*. Jan 30, 2023;23(1):18. [doi: [10.1186/s12880-023-00975-x](https://doi.org/10.1186/s12880-023-00975-x)] [Medline: [36717773](https://pubmed.ncbi.nlm.nih.gov/36717773/)]
17. Lee K, Lee S, Kwak JS, Park H, Oh H, Koh JC. Development and validation of an artificial intelligence model for detecting rib fractures on chest radiographs. *J Clin Med*. Jun 30, 2024;13(13):3850. [doi: [10.3390/jcm13133850](https://doi.org/10.3390/jcm13133850)] [Medline: [38999416](https://pubmed.ncbi.nlm.nih.gov/38999416/)]
18. Chang EY. DeepQ: advancing healthcare through artificial intelligence and virtual reality. Presented at: MM '17: Proceedings of the 25th ACM international conference on Multimedia New York; Oct 23-27, 2017; Mountain View, CA, USA. [doi: [10.1145/3123266.3130875](https://doi.org/10.1145/3123266.3130875)]
19. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *arXiv*. Preprint posted online on Jun 4, 2015. [doi: [10.48550/arXiv.1506.01497](https://doi.org/10.48550/arXiv.1506.01497)]
20. Yao L, Guan X, Song X, et al. Rib fracture detection system based on deep learning. *Sci Rep*. Dec 6, 2021;11(1):23513. [doi: [10.1038/s41598-021-03002-7](https://doi.org/10.1038/s41598-021-03002-7)] [Medline: [34873241](https://pubmed.ncbi.nlm.nih.gov/34873241/)]
21. van den Broek MCL, Buijs JH, Schmitz LFM, Wijffels MME. Diagnostic performance of artificial intelligence in rib fracture detection: systematic review and meta-analysis. *Surgeries*. 2024;5(1):24-36. [doi: [10.3390/surgeries5010005](https://doi.org/10.3390/surgeries5010005)]
22. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. Oct 29, 2019;17(1):195. [doi: [10.1186/s12916-019-1426-2](https://doi.org/10.1186/s12916-019-1426-2)] [Medline: [31665002](https://pubmed.ncbi.nlm.nih.gov/31665002/)]
23. Theriault-Lauzier P, Cobin D, Tastet O, et al. A responsible framework for applying artificial intelligence on medical images and signals at the point of care: the PACS-AI platform. *Can J Cardiol*. Oct 2024;40(10):1828-1840. [doi: [10.1016/j.cjca.2024.05.025](https://doi.org/10.1016/j.cjca.2024.05.025)] [Medline: [38885787](https://pubmed.ncbi.nlm.nih.gov/38885787/)]
24. Herpe G, Nelken H, Vendevure T, et al. Effectiveness of an artificial intelligence software for limb radiographic fracture recognition in an emergency department. *J Clin Med*. Sep 20, 2024;13(18):5575. [doi: [10.3390/jcm13185575](https://doi.org/10.3390/jcm13185575)] [Medline: [39337062](https://pubmed.ncbi.nlm.nih.gov/39337062/)]
25. Zhou Q, Qin P, Luo J, et al. Evaluating AI rib fracture detections using follow-up CT scans. *Am J Emerg Med*. Oct 2023;72:34-38. [doi: [10.1016/j.ajem.2023.07.018](https://doi.org/10.1016/j.ajem.2023.07.018)] [Medline: [37478635](https://pubmed.ncbi.nlm.nih.gov/37478635/)]
26. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging*. Feb 2017;8(1):171-182. [doi: [10.1007/s13244-016-0534-1](https://doi.org/10.1007/s13244-016-0534-1)] [Medline: [27928712](https://pubmed.ncbi.nlm.nih.gov/27928712/)]
27. Sun H, Wang X, Li Z, et al. Automated rib fracture detection on chest X-ray using contrastive learning. *J Digit Imaging*. Oct 2023;36(5):2138-2147. [doi: [10.1007/s10278-023-00868-z](https://doi.org/10.1007/s10278-023-00868-z)] [Medline: [37407842](https://pubmed.ncbi.nlm.nih.gov/37407842/)]
28. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS ONE*. 2018;13(7):e0201016. [doi: [10.1371/journal.pone.0201016](https://doi.org/10.1371/journal.pone.0201016)] [Medline: [30028888](https://pubmed.ncbi.nlm.nih.gov/30028888/)]
29. Moskalenko V, Kharchenko V. Resilience-aware MLOps for AI-based medical diagnostic system. *Front Public Health*. 2024;12:1342937. [doi: [10.3389/fpubh.2024.1342937](https://doi.org/10.3389/fpubh.2024.1342937)] [Medline: [38601490](https://pubmed.ncbi.nlm.nih.gov/38601490/)]
30. Khara R, Butte AJ, Berkwits M, et al. AI in medicine-JAMA's focus on clinical outcomes, patient-centered care, quality, and equity. *JAMA*. Sep 5, 2023;330(9):818-820. [doi: [10.1001/jama.2023.15481](https://doi.org/10.1001/jama.2023.15481)] [Medline: [37566406](https://pubmed.ncbi.nlm.nih.gov/37566406/)]
31. Armoundas AA, Narayan SM, Arnett DK, et al. Use of artificial intelligence in improving outcomes in heart disease: a scientific statement from the American Heart Association. *Circulation*. Apr 2, 2024;149(14):e1028-e1050. [doi: [10.1161/CIR.0000000000001201](https://doi.org/10.1161/CIR.0000000000001201)] [Medline: [38415358](https://pubmed.ncbi.nlm.nih.gov/38415358/)]
32. Mazor T, Farhat KS, Trukhanov P, et al. Clinical trial notifications triggered by artificial intelligence-detected cancer progression: a randomized trial. *JAMA Netw Open*. Apr 1, 2025;8(4):e252013. [doi: [10.1001/jamanetworkopen.2025.2013](https://doi.org/10.1001/jamanetworkopen.2025.2013)] [Medline: [40257799](https://pubmed.ncbi.nlm.nih.gov/40257799/)]
33. Mandair D, Elia MV, Hong JC. Considerations in translating AI to improve care. *JAMA Netw Open*. Apr 1, 2025;8(4):e252023. [doi: [10.1001/jamanetworkopen.2025.2023](https://doi.org/10.1001/jamanetworkopen.2025.2023)] [Medline: [40257804](https://pubmed.ncbi.nlm.nih.gov/40257804/)]
34. Habib AR, Gross CP. FDA regulations of AI-driven clinical decision support devices fall short. *JAMA Intern Med*. Dec 1, 2023;183(12):1401-1402. [doi: [10.1001/jamainternmed.2023.5006](https://doi.org/10.1001/jamainternmed.2023.5006)] [Medline: [37812411](https://pubmed.ncbi.nlm.nih.gov/37812411/)]
35. Chouffani El Fassi S, Abdullah A, Fang Y, et al. Not all AI health tools with regulatory authorization are clinically validated. *Nat Med*. Oct 2024;30(10):2718-2720. [doi: [10.1038/s41591-024-03203-3](https://doi.org/10.1038/s41591-024-03203-3)] [Medline: [39187696](https://pubmed.ncbi.nlm.nih.gov/39187696/)]

Abbreviations

AI: artificial intelligence
AP: anteroposterior
CLAIM: Checklist for Artificial Intelligence in Medical Imaging
CT: computed tomography
CXR: chest radiograph
FROC: free-response receiver operating characteristic
GPU: graphics processing unit
NLP: natural language processing

NPV: negative predictive value

PACS: picture archiving and communication system

R-CNN: region-based convolutional neural network

Edited by Andrew Coristine, Arriel Benis; peer-reviewed by Dongjoon Yoo, Enze Bai, Jun Zhang; submitted 22.May.2025; final revised version received 30.Dec.2025; accepted 01.Jan.2026; published 29.Jan.2026

Please cite as:

Huang ST, Liu LR, Tsai MF, Huang MY, Chiu HW

Prospective Diagnostic Accuracy and Technical Feasibility of Artificial Intelligence-Assisted Rib Fracture Detection on Chest Radiographs: Observational Study

JMIR Med Inform 2026;14:e77965

URL: <https://medinform.jmir.org/2026/1/e77965>

doi: [10.2196/77965](https://doi.org/10.2196/77965)

© Shu-Tien Huang, Liong-Rung Liu, Ming-Feng Tsai, Ming-Yuan Huang, Hung-Wen Chiu. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 29.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.