Am J Transl Res 2026;18(10):9618-9630
www.ajtr.org /ISSN:1943-8141/AJTR0174669

Original Article

AI-assisted reporting improves diagnostic concordance and reduces turnaround time in CT evaluation of rib fractures

Chun Xiong[1], Yiting Hu[2], Xinmeng Yu[2], Chen Yan[3]

[1]Department of Medical Imaging, Affiliated Hospital of Jinggangshan University, Ji'an 343400, Jiangxi, China; [2]Jinggangshan University, Ji'an 343400, Jiangxi, China; [3]Department of Radiology, Ji'an County People's Hospital, Ji'an 343100, Jiangxi, China

Received June 11, 2026; Accepted September 2, 2026; Epub October 15, 2026; Published October 30, 2026

Abstract: Aims: To evaluate whether Artificial Intelligence (AI)-integrated reporting improves diagnostic concordance, fracture detection performance, and report turnaround time compared with conventional interpretation and standalone AI in patients undergoing Computed Tomography (CT) for rib fracture assessment, thereby clarifying the clinical value of AI-assisted workflow integration in routine trauma radiology practice. Methods: This retrospective study included 410 consecutive chest CT examinations from patients with rib fractures confirmed by an expert-consensus reference standard. Pre-AI cases were assigned to the conventional reporting group (Group III, n = 150), post-implementation cases to the AI-assisted group (Group I, n = 260), and all examinations were additionally processed by the AI model for standalone evaluation (Group II). Diagnostic performance and reporting efficiency were assessed against a senior-radiologist consensus reference standard. Primary outcomes focused on concordance and turnaround time; secondary outcomes included fracture detection yield, morphology-specific sensitivity, localization concordance, report quality, and subgroup performance. Results: Compared with conventional reporting, AI-assisted reporting detected a significantly higher mean number of fractures per patient ($P = 0.029$). Reporting efficiency was significantly improved, with shorter mean and median turnaround times (both $P < 0.001$), a higher proportion of reports completed within 30 minutes ($P < 0.001$), and shorter time to first fracture identification ($P < 0.001$). AI-assisted reporting also demonstrated better concordance with expert consensus ($P < 0.001$), with a higher rate of complete concordance (84.6% vs. 66.0%) and fewer major discrepancies (3.8% vs. 12.0%), together with superior laterality-, rib-, and segment-level concordance ($P = 0.007$, $P = 0.001$, and $P = 0.001$, respectively). Report completeness and linguistic quality were also significantly improved with AI integration (all $P < 0.001$). Across fracture types, AI-assisted interpretation consistently showed the highest agreement with the reference standard ($\kappa = 0.89$-0.94), outperforming both standalone AI ($\kappa = 0.76$-0.83) and manual interpretation ($\kappa = 0.45$-0.71). Conclusion: AI-assisted rib fracture reporting improved diagnostic agreement and reduced turnaround time, supporting the clinical value of AI-assisted reporting in routine trauma CT interpretation.

Keywords: Rib fractures, artificial intelligence, diagnostic concordance, turnaround time, radiology reporting

Introduction

Rib fractures are among the most common findings in patients with blunt thoracic trauma and are an important source of morbidity, especially in older patients and in those with multiple injuries [1-3]. In addition to being an indicator of the severity of the chest trauma, rib fractures are a strong predictor of pain, respiratory compromise, prolonged hospitalizations, and are associated with poor outcomes when complicated by pneumothorax, hemothorax, or pulmonary contusion [4-6]. Thus, the early and accurate diagnosis of rib fractures is critical, not only for diagnosis, but also for risk stratification, treatment planning, and prompt escalation of care in the trauma setting.

Computed tomography is considered the gold standard imaging technique for the assessment of thoracic trauma as it provides high-resolution imaging and enables identification of both bony injury and underlying intrathoracic abnormalities [7, 8]. Despite this, rib fractures are still at risk of being missed on routine chest radiographs, particularly when fractures

https://doi.org/10.62347/PUOD9220

are nondisplaced, subtle, or obscured by complex anatomy, especially in the context of a busy emergency worklist. Double reading has been shown to increase detection; however, it is resource intensive and not necessarily feasible in real-world settings, where imaging volumes are increasing, the number of radiologists is limited, and turnaround time expectations are stringent. This means that there remain clinical issues of missed fractures and delayed reporting in trauma imaging.

Recent developments in artificial intelligence, especially on the basis of deep learning for image analysis, have opened up new possibilities for complementing the radiologic detection of fractures [9-12]. Previous research has demonstrated that AI systems can achieve high sensitivity in identifying rib fractures and have the potential to enhance radiologist performance and reduce interpretation time [9, 13, 14]. Yet, most of the literature to date has been focused on algorithm development or reader-study environments, rather than on AI deployed in routine clinical reporting. As a result, the practical effects of AI on diagnostic agreement, the quality of reporting, and efficiency of turnaround are not well established. Specifically, it remains unclear whether AI provides greater value as a standalone detector or as an integrated decision-support system to complement radiologist judgment.

Against this background, the current study assessed the impact of AI-assisted reporting for rib fracture evaluation in the clinical setting, comparing traditional radiologist interpretation, AI alone, and AI-assisted interpretation. We investigated diagnostic agreement with expert consensus, patient- and case-level fracture detection performance, morphology-specific sensitivity, report completeness, and turnaround-related workflow parameters. Based on this, we hypothesized that AI-assisted reporting would not only promote diagnostic consistency but also improve operational efficiency, representing a clinically scalable approach to enhance trauma CT interpretation in daily radiology practice.

Materials and methods

Study design

This single-center retrospective cohort study used a natural time-window design and included 410 consecutive chest CT examinations from 410 patients who underwent CT evaluation for suspected rib fractures following chest trauma between January and December 2025. Patients were initially screened based on chest trauma and clinical suspicion of rib fracture; final inclusion required at least one rib fracture confirmed by the expert-consensus reference standard. Accordingly, all 410 patients included in the final cohort were classified as rib-fracture positive, and no reference-standard-negative cases were included. The study comprised two independent clinical cohorts defined according to the timing of artificial intelligence (AI) implementation. The conventional reporting cohort included 150 examinations performed between January and June 2025, before implementation of the AI system, and these examinations were interpreted by radiologists using the standard clinical workflow without AI support. The AI-assisted reporting cohort included 260 examinations performed between July and December 2025, after implementation of the AI system. In this cohort, radiologists reviewed the fracture locations and segmentation outputs generated by the AI system and incorporated these findings into the final radiology report when they considered it clinically appropriate. In addition to the two independent clinical cohorts, all 410 computed tomography examinations were retrospectively processed using the same fixed AI model to generate standalone AI interpretation results. These standalone AI results did not constitute a third independent patient cohort but represented an additional interpretation condition paired with the corresponding clinical interpretation for each examination. Accordingly, comparisons between the conventional and AI-assisted cohorts were conducted as independent-cohort analyses, whereas analyses involving standalone AI accounted for the paired structure of the data. The expert-consensus reference standard was established through independent review of all examinations by two senior thoracic radiologists who had not participated in the original clinical interpretations. Both reviewers were blinded to the original clinical reports, standalone AI outputs, and cohort allocation. Disagreements were resolved through joint review and consensus, with adjudication by a third senior radiologist when necessary. The study was conducted in accordance with the Declaration of Helsinki

and was approved by the Ethics Committee of the Affiliated Hospital of Jinggangshan University (approval No. JDFY2025YXK0064). The requirement for written informed consent was waived because of the retrospective design and the use of anonymized clinical and imaging data. This design enabled comparison of the two clinical reporting workflows and paired evaluation of standalone AI performance against the same expert-consensus reference standard.

Patient selection criteria

A total of 850 patients who underwent chest CT following thoracic trauma between January and December 2025 were retrospectively identified from the hospital Picture Archiving and Communication System (PACS). Consecutive patients undergoing CT evaluation for suspected rib fractures were screened for eligibility. The inclusion criteria were as follows: (1) age ≥ 8 years; (2) a clinical history of thoracic trauma for which chest CT was performed as part of the routine diagnostic work-up; (3) complete coverage of the thoracic cage with image quality sufficient for rib fracture assessment, including axial images and coronal and sagittal reconstructions; and (4) availability of complete clinical and imaging data, including demographic characteristics, mechanism of injury, and history of previous thoracic interventions. The exclusion criteria were as follows: (1) severe motion or metal artifacts, or incomplete image acquisition, that could compromise fracture assessment (n = 150); (2) previous thoracic surgery or implanted hardware within the region of interest that could obscure or mimic rib fractures (n = 120); (3) incomplete clinical or imaging records that precluded reliable evaluation against the expert-consensus reference standard (n = 110); and (4) ambiguous imaging findings or unresolved diagnostic discrepancies that could not be adjudicated through the expert-consensus process (n = 60). Accordingly, 440 patients were excluded, leaving 410 patients for the final analysis. Of the included patients, 41% were female, and the mean age was 55.8 years. Rib fractures identified by the expert-consensus reference standard were classified according to their temporal and morphological characteristics as follows: Type I, chronic occult fracture with callus formation; Type II, chronic displaced fracture with callus formation; Type III, acute occult fracture; and Type IV, acute displaced fracture. All imaging and clinical data were anonymized before analysis to protect patient confidentiality and support blinded assessment. The detailed patient-selection process is presented in **Figure 1**.

The diagnostic reference standard

To address the absence of a predefined diagnostic gold standard, an expert consensus reference standard was established for all included CT examinations. All anonymized chest CT images were independently reviewed by two senior thoracic radiologists who were blinded to the original clinical reports, AI outputs, and study group allocation. The reference diagnosis included the presence or absence of rib fracture, total fracture number, laterality, rib level, anatomical segment, fracture morphology, displacement status, temporal pattern, and associated thoracic injuries such as pneumothorax, hemothorax, and pulmonary contusion. When discrepancies occurred between the two reviewers, a consensus decision was reached through joint review and discussion; unresolved cases were adjudicated by a third senior radiologist. This consensus-based reference standard was subsequently used as the diagnostic gold standard for evaluating the performance and concordance of conventional interpretation, standalone AI assessment, and AI-assisted reporting.

Data collection and outcome measures

A standardized case-report form was used to extract demographic, clinical, imaging, and reporting data from the hospital Picture Archiving and Communication System (PACS) and radiology information system. Because each patient contributed only one CT examination, patient-level and examination-level analyses were equivalent unless otherwise specified. Baseline variables, including age, sex, body mass index (BMI), Injury Severity Score (ISS), mechanism of injury (MOI), time from trauma to CT, chronic obstructive pulmonary disease, osteoporosis/osteopenia, and anticoagulant use, were recorded on a per-patient basis. Imaging characteristics, including fracture positivity, bilateral involvement, the presence of displaced or comminuted fractures, and associated thoracic injuries such as pneu-

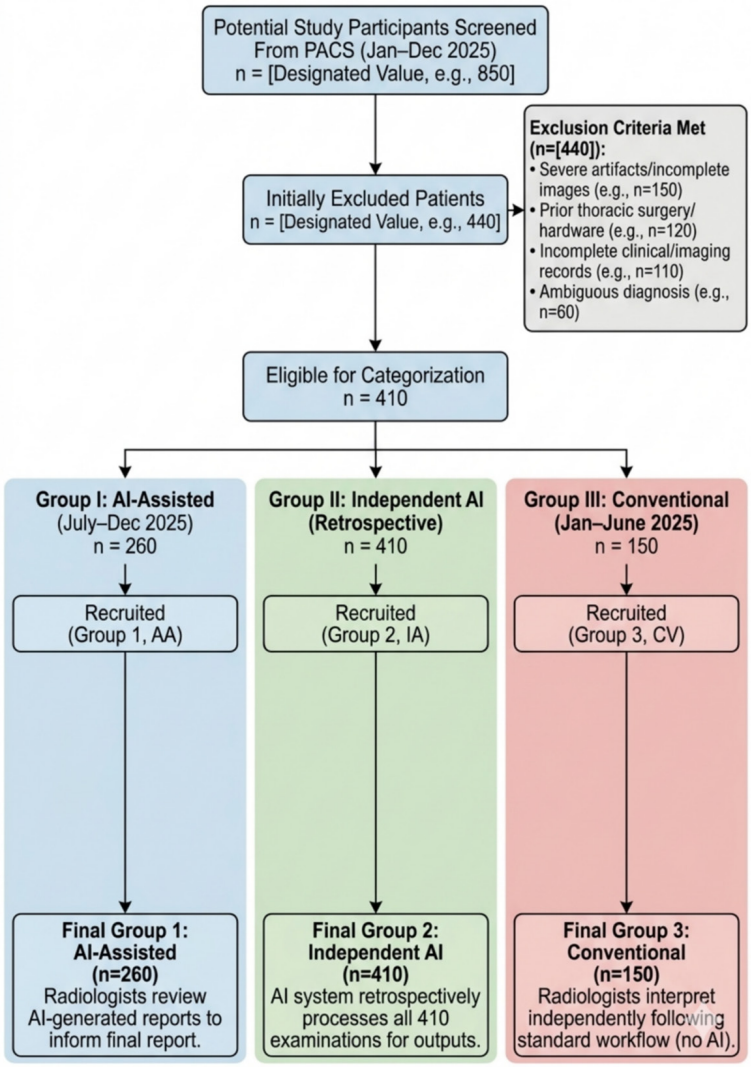


Figure 1. Participant enrollment and group assignment flowchart. Notes: PACS: Picture Archiving and Communication System; AI: Artificial Intelligence; AA: AI-Assisted; IA: Independent AI; CV: Conventional interpretation; Jan: January; Dec: December.

mothorax, hemothorax, and pulmonary contusion, were also evaluated per patient. The total number of fractures, missed fractures, and false-positive findings was calculated as the number of fracture sites per patient. Individual fracture sites were further classified according to rib level, anatomical segment, upper/middle/lower rib location, displacement status, temporal pattern, and morphology, including nondisplaced, displaced, buckle/greenstick, comminuted, segmental, acute, and acute-on-healed fractures. Reporting and workflow variables were assessed per CT examination or corresponding final radiology report and included the initial interpretation, AI output where applicable, final signed report, time to first fracture identification, preliminary-to-final sign-off interval, overall report turnaround time, after-hours reporting delay, requirement for senior revision, and 24-hour report addendum status. The primary outcomes were diagnostic concordance with the senior-radiologist consensus reference standard and reporting efficiency. Patient-level diagnostic outcomes included fracture detection status, complete or partial concordance, discrepancy category, and per-patient fracture burden agreement. Fracture-site-level outcomes included morphology-specific sensitivity and concordance for laterality, rib level, and anatomical segment, with the corresponding reference-standard fracture sites used as the denominator. Secondary outcomes included the number of detected, missed, and overcalled fracture sites per patient; inter-method agreement indices; report completeness; linguistic quality; compliance with structured terminology; recommendation consistency; and prespecified subgroup performance according to age, sex, trauma pattern, mechanism of injury, and fracture burden. Report-quality outcomes were calculated per finalized report. All diagnostic and localization outcomes were evaluated against the expert-consensus reference standard.

AI system used

The AI system used in this study was a deep learning-based computer-aided detection (CAD) tool specifically developed for automated rib fracture detection on chest CT. The algorithm was built on a three-dimensional convolutional neural network (3D-CNN) framework with a two-stage detection strategy. In the first stage, candidate rib lesions were automatically local-

ized from volumetric CT images, and in the second stage, the detected candidates were further classified as fracture or non-fracture using multi-scale image features extracted from adjacent slices. The model was trained using a large retrospective dataset of chest CT examinations collected from multiple medical centers. All training cases were independently annotated by at least two board-certified thoracic radiologists, and discrepancies were resolved by consensus among senior musculoskeletal radiologists to establish the reference standard. During model development, an independent validation cohort was used for parameter tuning and model selection, while a separate internal test cohort was used to assess algorithm performance before clinical deployment.

For the present study, the AI system was applied in a fixed, inference-only mode without any retraining, fine-tuning, or threshold adjustment. The software automatically analyzed thin-section chest CT images and generated suspected rib fracture locations together with confidence scores, which were subsequently reviewed by the interpreting radiologists as part of the AI-assisted reading workflow. Importantly, all 410 CT examinations included in this study were completely independent of the AI model development process. None of these cases had been used for model training, validation, parameter optimization, or internal testing. Therefore, the diagnostic performance reported in this study represents an independent evaluation of the AI system in a real-world clinical setting, ensuring that no data leakage occurred between the development and evaluation cohorts.

Statistical analysis

IBM SPSS Statistics version 27.0 (IBM Corp., Armonk, NY, USA) was used for all statistical analyses. Continuous variables were summarized as mean ± standard deviation (SD) or median (interquartile range [IQR] or range), as appropriate, whereas categorical variables were summarized as counts and percentages. The conventional reporting cohort (Group III) and the AI-assisted reporting cohort (Group I) were treated as two independent clinical cohorts. For comparisons between these two cohorts, normally distributed continuous variables were analyzed using the independent-samples t test, non-normally distributed continuous variables using the Mann-Whitney U test, and categorical variables using the chi-square test or Fisher's exact test, as appropriate. The standalone AI interpretation (Group II) was not treated as a third independent group, as the AI outputs were retrospectively generated from the same CT examinations in Groups III and I. Accordingly, comparisons involving standalone AI were performed separately within the corresponding conventional and AI-assisted cohorts as paired analyses, thereby preserving the one-to-one correspondence between each examination and its clinical and standalone AI interpretations. For such paired comparisons, continuous variables were analyzed using the paired-samples t test or Wilcoxon signed-rank test, as appropriate, and paired categorical variables were analyzed using McNemar's test or other appropriate methods for paired categorical data, as determined by the outcome structure. No three-group omnibus tests treating Groups III, I, and standalone AI as independent samples were performed. Diagnostic outcomes were evaluated at the patient/examination level, fracture-site level, or report level, according to the predefined outcome. Agreement with the expert-consensus reference standard was assessed using Cohen's κ for categorical agreement, weighted κ for ordinal agreement, and the intraclass correlation coefficient (ICC) for fracture-count agreement, as appropriate. Reporting-efficiency outcomes, including turnaround time and other workflow-related time intervals, were compared between the two independent clinical cohorts using parametric or nonparametric tests according to the distribution of the data, whereas categorical workflow outcomes were analyzed using the chi-square test or Fisher's exact test, as appropriate. Prespecified subgroup analyses according to age, sex, trauma pattern, mechanism of injury, and fracture burden were performed to evaluate the consistency of diagnostic performance across clinically relevant subgroups. All statistical tests were two-sided, and a P value < 0.05 was considered statistically significant.

Results

Patient demographics and baseline characteristics

The baseline demographic and clinical characteristics were well balanced between the two

Table 1. Patient demographics and baseline characteristics

Parameter	Conventional reporting (Group III), n = 150	AI-assisted reporting (Group I), n = 260	Statistic	*P* value
Age, years	56.63 ± 13.14	55.20 ± 13.63	t = 1.039	0.299
Male sex, n (%)	92 (61.3)	154 (59.2)	χ^2 = 0.175	0.676
BMI, kg/m^2	24.56 ± 3.97	25.38 ± 3.71	t = -2.098	0.057
Injury Severity Score	13.33 ± 4.54	13.79 ± 5.10	t = -0.925	0.355
Polytrauma, n (%)	42 (28.0)	80 (30.8)	χ^2 = 0.349	0.555
Osteoporosis/osteopenia, n (%)	19 (12.7)	37 (14.2)	χ^2 = 0.197	0.657
COPD, n (%)	14 (9.3)	28 (10.8)	χ^2 = 0.213	0.644
Anticoagulant use, n (%)	17 (11.3)	33 (12.7)	χ^2 = 0.164	0.685
Time from trauma to CT, h	5.87 ± 2.54	5.59 ± 2.28	t = 1.142	0.254
Fall-related injury, n (%)	73 (48.7)	128 (49.2)	χ^2 = 0.012	0.912

Note: BMI: body mass index; COPD: chronic obstructive pulmonary disease; CT: computed tomography; Group III: conventional reporting group; Group I: AI-assisted reporting group.

study groups, with no statistically significant differences in any of the pre-specified variables. Specifically, age (P = 0.299), sex distribution (P = 0.676), BMI (P = 0.057), and Injury Severity Score (P = 0.355) were comparable between the conventional reporting and AI-assisted reporting groups. Similarly, the prevalence of polytrauma (P = 0.555), osteoporosis/osteopenia (P = 0.657), COPD (P = 0.644), and anticoagulant use (P = 0.685) did not differ significantly between the groups. Time from trauma to CT acquisition was also comparable (P = 0.254), as was the proportion of fall-related injuries (P = 0.912) (**Table 1**).

CT acquisition and reference-standard injury burden

The mean number of confirmed fractures per patient (P = 0.340) and the proportion of patients with a higher fracture burden (≥ 3 fractures, P = 0.343) were comparable between the two cohorts. Likewise, the distributions of more complex injury patterns, including bilateral rib fractures (P = 0.504), at least one displaced fracture (P = 0.691), and at least one comminuted fracture (P = 0.606), did not differ significantly between the conventional and AI-assisted cohorts. Associated thoracic injuries were also comparable, with no significant between-cohort differences in pneumothorax (P = 0.546), hemothorax (P = 0.495), or pulmonary contusion (P = 0.623). In addition, the use of thin-slice CT reconstruction (≤ 1.0 mm), an important technical factor affecting fracture conspicuity, was similar between the two cohorts (P = 0.559) (**Table 2**).

Rib fracture detection per case

The AI-assisted group demonstrated a significantly higher mean number of detected fractures per patient than the conventional reporting group (2.48 ± 1.38 vs. 2.17 ± 1.31; P = 0.029). The median number of detected fractures was 2 (IQR: 2-4) in the AI-assisted group and 2 (IQR: 1-3) in the conventional group. Although a greater proportion of patients in the AI-assisted group had multiple fractures (≥ 3 fractures) detected (42.3% vs. 34.0%), the difference did not reach statistical significance (P = 0.097). Similarly, bilateral rib fractures were detected in 15.0% of patients in the AI-assisted group compared with 12.0% in the conventional group, with no significant between-group difference (P = 0.398). The AI-assisted group also showed numerically fewer missed fractures per patient (0.18 vs. 0.42) and fewer overcalled fractures per patient (0.09 vs. 0.15). The anatomical distribution of detected fractures was comparable between the two groups, with no significant differences in upper rib involvement (21% vs. 19%; P = 0.131), mid-rib involvement (36% vs. 34%; P = 0.179), or lower rib involvement (27% vs. 23%; P = 0.140) (**Table 3**).

Concordance with expert consensus reporting

AI-assisted reporting showed significantly better overall concordance with the consensus interpretation than conventional reporting (P < 0.001). The complete concordance rate was higher (84.6% [220/260] vs. 66.0% [99/150]), while minor (11.5% [30/260] vs. 22.0% [33/

Table 2. CT acquisition characteristics and reference-standard injury burden in the two independent clinical cohorts

Parameter	Conventional reporting, Group III (n = 150)	AI-assisted reporting, Group I (n = 260)	Overall cohort (n = 410)	Statistic	*P* value
Rib-fracture-positive cases, n (%)	150 (100)	260 (100)	410 (100)	-	-
Confirmed fractures per case	2.31 ± 1.44	2.46 ± 1.58	2.41 ± 1.53	F = 1.08	0.340
≥ 3 fractures per case, n (%)	58 (38.7)	113 (43.5)	171 (41.7)	χ^2 = 0.900	0.343
Bilateral rib fractures, n (%)	20 (13.3)	41 (15.8)	61 (14.9)	χ^2 = 0.446	0.504
At least one displaced fracture, n (%)	36 (24.0)	67 (25.8)	103 (25.1)	χ^2 = 0.158	0.691
At least one comminuted fracture, n (%)	17 (11.3)	34 (13.1)	51 (12.4)	χ^2 = 0.266	0.606
Associated pneumothorax, n (%)	28 (18.7)	55 (21.2)	83 (20.2)	χ^2 = 0.364	0.546
Associated hemothorax, n (%)	21 (14.0)	43 (16.5)	64 (15.6)	χ^2 = 0.465	0.495
Pulmonary contusion, n (%)	41 (27.3)	77 (29.6)	118 (28.8)	χ^2 = 0.242	0.623
Thin-slice CT (≤ 1.0 mm), n (%)	123 (82.0)	219 (84.2)	342 (83.4)	χ^2 = 0.342	0.559

Note: Data are presented as mean ± standard deviation or number (percentage). Statistical tests compare only the two independent clinical cohorts: the conventional reporting cohort and the artificial intelligence-assisted reporting cohort. The overall cohort is shown for descriptive purposes and was not included as an additional comparison group. Standalone AI interpretation (Group II) is not presented here, as it represents a paired interpretation condition rather than an independent patient cohort; therefore, it has no independent baseline or reference-standard characteristics. No statistical test was performed for rib-fracture positivity because all included patients were positive according to the expert-consensus reference standard. CT: computed tomography; Group III: conventional reporting group; Group I: AI-assisted reporting group; Group II: standalone AI interpretation.

Table 3. Per-patient rib fracture detection performance according to interpretation strategy

Parameter	Conventional reporting, Group III (n = 150)	AI-assisted reporting, Group I (n = 260)	Statistic	*P* value
Detected fractures per case, mean ± SD	2.17 ± 1.31	2.48 ± 1.38	t = 2.194	0.029
Detected fractures per case, median (IQR)	2 (1-3)	2 (2-4)	-	-
Range	0-7	0-8	-	-
Multi-fracture cases (≥ 3 fractures), n (%)	51 (34.0)	110 (42.3)	χ^2 = 2.753	0.097
Missed fractures per patient	0.42	0.18		
Overcalled fractures per patient	0.15	0.09		
Bilateral fractures detected, n (%)	18 (12.0)	39 (15.0)	χ^2 = 0.715	0.398
Upper rib involvement, %	19	21	χ^2 = 0.276	0.131
Mid rib involvement, %	34	36	χ^2 = 1.809	0.179
Lower rib involvement, %	23	27	χ^2 = 2.176	0.140

Note: All outcomes were calculated at the patient level. Each patient contributed one chest CT examination. Detected, missed, and false-positive fracture counts represent the number of fracture sites per patient. Categorical outcomes represent the number or percentage of patients meeting the specified criterion. Group III: conventional reporting group; Group I: AI-assisted reporting group.

150]) and major discrepancies (3.8% [10/260] vs. 12.0% [18/150]) were lower. Agreement was also significantly greater with AI-assisted reporting at progressively finer levels of anatomic localization, including exact laterality concordance (97.3% vs. 91.3%; P = 0.007), exact rib-level concordance (90.8% vs. 79.3%; P = 0.001), and exact segment concordance (87.3% vs. 74.7%; P = 0.001). In addition, fewer cases in the AI-assisted group required adjudication (6.5% vs. 14.7%; P = 0.007), and the 24-hour addendum rate after adjudication was also lower (3.1% vs. 9.3%; P = 0.007) (**Table 4**), indicating improved reporting concordance and greater downstream report stability with AI assistance.

Reporting efficiency and turnaround metrics

Compared with the conventional reporting group (Group III), the AI-assisted reporting group (Group I) demonstrated significantly

Table 4. Concordance with expert consensus reporting

Parameter	Conventional reporting, Group III (n = 150)	AI-assisted reporting, Group I (n = 260)	Statistic	*P* value
Overall concordance category			χ^2 (2) = 20.272	< 0.001
Complete concordance, n/N (%)	99/150 (66.0)	220/260 (84.6)		
Minor discrepancy, n/N (%)	33/150 (22.0)	30/260 (11.5)		
Major discrepancy, n/N (%)	18/150 (12.0)	10/260 (3.8)		
Exact laterality concordance, n/N (%)	137/150 (91.3)	253/260 (97.3)	χ^2 (1) = 7.317	0.007
Exact rib-level concordance, n/N (%)	119/150 (79.3)	236/260 (90.8)	χ^2 (1) = 10.710	0.001
Exact segment concordance, n/N (%)	112/150 (74.7)	227/260 (87.3)	χ^2 (1) = 10.616	0.001
Cases requiring adjudication, n/N (%)	22/150 (14.7)	17/260 (6.5)	χ^2 (1) = 7.301	0.007
24-h addendum after adjudication, n/N (%)	14/150 (9.3)	8/260 (3.1)	χ^2 (1) = 7.332	0.007

Note: AI: artificial intelligence; ICC: intraclass correlation coefficient; Group III: conventional reporting group; Group I: AI-assisted reporting group.

Table 5. Reporting efficiency and turnaround metrics

Parameter	Conventional reporting, Group III (n = 150)	AI-assisted reporting, Group I (n = 260)	Statistic	*P* value
Mean report TAT, min, mean ± SD	40.19 ± 11.32	30.46 ± 8.46	t = 9.881	< 0.001
Median report TAT, min	40	28	H = 11.74	< 0.001
Report TAT range, min	25-78	18-55	-	-
Reports completed within 30 min, %	32 (21.3%)	161 (61.9%)	χ^2 = 62.902	< 0.001
Reports completed within 31-45 min, %	51 (34.0%)	26 (10.0%)	χ^2 = 35.920	< 0.001
Reports completed within 46-60 min, %	44 (29.3%)	26 (10.0%)	χ^2 = 25.112	< 0.001
Reports completed after 60 min, %	32 (21.3%)	8 (3.1%)	χ^2 = 12.770	< 0.001
Mean time to first fracture identification, min	35.2	22.1	t = 13.44	< 0.001
Radiologist time saved per examination, min	-	12.8	t = 12.98	< 0.001
Examinations with overlapping findings, %	18 (12.0%)	30 (11.5%)	χ^2 = 0.020	0.889
Preliminary-to-final sign-off interval, min, mean ± SD	17.80 ± 6.38	10.80 ± 4.97	t = 12.350	< 0.001
After-hours finalization delay, min, mean ± SD	24.80 ± 10.15	19.25 ± 8.33	t = 5.991	< 0.001
Senior revision required, n (%)	27 (18.0%)	24 (9.2%)	χ^2 = 6.716	0.010
24-h report addendum, n (%)	13 (8.7%)	9 (3.5%)	χ^2 = 5.075	0.024
Examinations interpreted per radiologist-hour, mean ± SD	1.40 ± 0.32	2.03 ± 0.41	t = 16.282	< 0.001

Note: Group III: conventional reporting group; Group I: AI-assisted reporting group.

improved reporting efficiency. The mean report turnaround time (TAT) decreased from 40.19 ± 11.32 min to 30.46 ± 8.46 min (P < 0.001), while the median TAT decreased from 40 to 28 min (P < 0.001). The interval from preliminary interpretation to final sign-off was also significantly shortened from 17.80 ± 6.38 min to 10.80 ± 4.97 min (P < 0.001). Time to first fracture identification decreased from 35.2 to 22.1 min with AI assistance (P < 0.001), indicating earlier recognition of actionable findings. AI integration markedly increased the proportion of reports completed within 30 min from 21.3% to 61.9% (P < 0.001), while significantly reducing the proportions completed within 46-60 min and after 60 min (both P < 0.001). Additional workflow improvements included a shorter after-hours finalization delay (24.80 ± 10.15 vs. 19.25 ± 8.33 min, P < 0.001), a lower proportion of cases requiring senior revision (18.0% vs. 9.2%, P = 0.010), and a lower 24-hour report addendum rate (8.7% vs. 3.5%, P = 0.024). Radiologist throughput also increased significantly from 1.40 ± 0.32 to 2.03 ± 0.41 examinations per radiologist-hour (P < 0.001) (**Table 5**).

Table 6. Per-report completeness and linguistic-quality outcomes according to interpretation strategy

Parameter	Conventional reporting, Group III (n = 150)	AI-assisted reporting, Group I (n = 260)	Statistic	*P* value
Explicit fracture count stated, n/N (%)	124/150 (82.7)	250/260 (96.2)	$\chi^2 = 21.603$	< 0.001
Laterality explicitly stated, n/N (%)	132/150 (88.0)	255/260 (98.1)	$\chi^2 = 18.242$	< 0.001
Rib number explicitly stated, n/N (%)	121/150 (80.7)	249/260 (95.8)	$\chi^2 = 24.643$	< 0.001
Segment/anatomical arc stated, n/N (%)	100/150 (66.7)	238/260 (91.5)	$\chi^2 = 40.646$	< 0.001
Displacement status stated, n/N (%)	87/150 (58.0)	224/260 (86.2)	$\chi^2 = 41.165$	< 0.001
Associated thoracic findings included, n/N (%)	107/150 (71.3)	236/260 (90.8)	$\chi^2 = 26.284$	< 0.001
Impression specificity score (0-10), mean ± SD	7.05 ± 1.46	8.94 ± 0.82	t = 16.767	< 0.001
Structured terminology compliance, n/N (%)	104/150 (69.3)	246/260 (94.6)	$\chi^2 = 48.669$	< 0.001
Ambiguous wording present, n/N (%)	29/150 (19.3)	18/260 (6.9)	$\chi^2 = 14.435$	< 0.001
Recommendation concordance with trauma team, n/N (%)	112/150 (74.7)	232/260 (89.2)	$\chi^2 = 14.939$	< 0.001

Note: All outcomes were assessed at the report level. Denominators correspond to finalized radiology reports for Groups III and I. Values are presented as percentages unless otherwise specified. The impression-specificity score ranges from 0 to 10, with higher scores indicating greater specificity. Group III: conventional reporting group; Group I: AI-assisted reporting group; Group II: standalone AI interpretation.

Report completeness and linguistic quality outcomes according to interpretation strategy

AI integration was associated with a substantial improvement in both report completeness and linguistic quality. Compared with conventional reporting, the AI-assisted workflow achieved significantly higher rates of explicit documentation of fracture count, laterality, rib number, segmental/anatomical location, displacement status, and associated thoracic findings (all P < 0.001), together with a higher impression specificity score (P < 0.001). Recommendation concordance with the trauma team was also significantly improved (P < 0.001), suggesting that the gains in reporting quality were clinically meaningful rather than purely stylistic. Notably, structured terminology compliance differed significantly across groups (P < 0.001), with both AI-assisted and AI-alone reporting outperforming conventional interpretation, while ambiguous wording was markedly reduced after AI integration (P < 0.001), reaching elimination of ambiguous wording in the AI-alone output (**Table 6**).

Diagnostic concordance across rib fracture types according to interpretation strategy

For Type I fractures, κ was 0.91 for AI-assisted interpretation, compared with 0.76 for AI-alone and 0.63 for manual reading. For Type II fractures, AI-assisted interpretation achieved the highest agreement (κ = 0.94), exceeding both AI-alone (κ = 0.78) and conventional interpretation (κ = 0.71). For Type III fractures, which represented the most challenging category, AI-assisted reading maintained high concordance (κ = 0.89), whereas agreement declined to 0.83 for AI-alone and dropped markedly to 0.45 for manual interpretation, representing the largest performance gap across all fracture types. In Type IV fractures, concordance remained highest with AI assistance (κ = 0.90), followed by AI-alone (κ = 0.81) and manual interpretation (κ = 0.69) (**Figure 2**).

Discussion

This study highlights the significant benefits of AI integration into the clinical reporting workflow for rib fractures, in terms of both accuracy and efficiency. The AI-assisted strategy was superior to both conventional interpretation and standalone AI across all measured outcomes, including patient-level detection, fracture-level identification, expert consensus agreement, report completeness, and turnaround time. Key improvements were particularly notable for occult and morphologically complex fractures, with increased agreement at the rib- and segment-level localization, and significantly shortened turnaround time. The current study conceptualizes AI not merely as an independent detector, but as an integrated reporting support system embedded in the real-world radiology workflow. Accordingly, we evaluated not only algorithm performance but also its association with clinically meaningful endpoints, including downstream report stability, linguistic quality, and radiologist productivity.

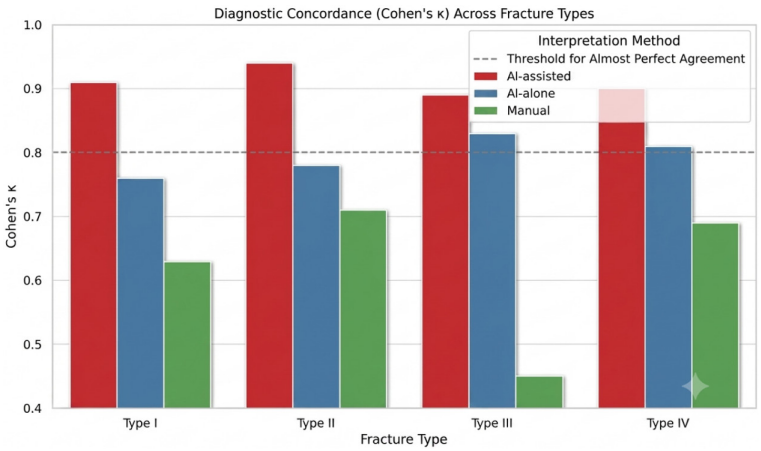


Figure 2. Diagnostic concordance between each interpretation strategy and the reference standard for different rib fracture types. Cohen's kappa (κ) coefficients were calculated to assess agreement between each interpretation strategy and the expert consensus reference standard. Note: Group I: AI-assisted radiologist interpretation; Group II: standalone AI interpretation; Group III: conventional radiologist interpretation. Higher κ values indicate better agreement with the reference standard.

AI support was associated with improved diagnostic performance across several clinically relevant measures, including per-patient fracture detection, reduced missed fractures and overcalls, and greater concordance with the expert-consensus reference standard. This finding is consistent with previous research demonstrating the potential of deep learning systems to improve rib fracture detection on CT and enhance radiologists' sensitivity to subtle or easily overlooked fractures [15-17]. Our data extend this literature by showing that the AI-assisted workflow achieved the greatest fracture detection yield per patient, together with the fewest missed fractures and overcalls and the strongest agreement with the reference standard. One possible explanation is that AI functions as a highly sensitive second reader, systematically highlighting suspicious sites and reducing perceptual errors, while subsequent radiologist review filters out false-positive findings. These complementary roles suggest that the most promising clinical application of AI may be not to replace expert interpretation, but rather to augment radiologist performance by improving fracture detection and diagnostic concordance. Nevertheless, AI performance in isolation does not necessarily translate into optimal real-world performance, where final interpretation depends on radiologist judgment and integration with clinical context.

AI integration provided the greatest benefit for morphologically challenging fracture subtypes, including nondisplaced, buckle/greenstick, acute-on-healed, and chest-wall instability fractures. This finding is consistent with previous reports suggesting that AI systems may be especially valuable for identifying rib abnormalities that are difficult to recognize on a routine examination [18]. We found that AI-assisted interpretation was consistently superior to conventional reading across all prespecified morphology- and severity-specific categories, suggesting that the incremental value of AI was not uniform but was concentrated in high-difficulty lesion phenotypes in our cohort. Mechanistically, this is likely attributable to the ability of convolutional models to recognize subtle spatial patterns and cortical irregularities on multiplanar CT images - findings that may be masked by reader fatigue, distractions, or competing priorities in trauma imaging. In practice, this is an important consideration for emergency and out-of-hours radiology. Although AI may facilitate prompt identification of occult fractures-thereby reducing delays in risk evaluation and pain management-it remains crucial for radiologists to exercise contextual adjudication rather than blindly accepting AI prompts, particularly in complex post-traumatic anatomy.

AI integration improved diagnostic concordance and interpretive reliability overall, across fracture subtypes, and at finer levels of anatomic granularity. While some evidence supports AI's ability to reduce inter-reader variance and increase consistency [19, 20], evidence for its impact on finer levels of localization agreement in routine reporting has been limited. In this study, the AI-assisted approach achieved the highest complete concordance rates, the lowest discrepancy burden, and best reliability metrics for fracture count, laterality, rib-level assignment, and segment-level localization. It also yielded near-perfect κ values across all fracture categories, with the greatest

benefit observed for recent occult fractures. The most likely explanation is that the localization cues provided by AI offer a shared visual and semantic reference point for readers, facilitating alignment in the identification, characterization, and reporting of abnormalities. This is clinically significant because greater diagnostic agreement is likely to translate into more consistent management decisions, fewer adjudications, and lower addendum rates. However, it also underscores the importance of scrutinizing AI localization outputs, as systematic localization errors could propagate into systematic errors in management decision-making across readers.

Integrating AI into reporting can significantly improve workflow without compromising report quality. Previous studies have shown that AI can reduce interpretation times [21, 22]; however, most have focused only on reading time rather than the entire reporting pathway. In this study, AI assistance was correlated with shorter overall turnaround time, reduced preliminary-to-final sign-off intervals, fewer delayed reports, lower after-hours finalization burden, and increased radiologist throughput. These improvements are likely attributable to a several factors: earlier algorithmic triage of lesions, reduced time spent on exhaustive manual search, and more efficient radiologist confirmation and synthesis of candidate lesions identified by the algorithm. Such gains may be especially valuable in high-volume trauma settings from a systems perspective, where reporting delays can negatively impact downstream clinical coordination. However, these efficiency gains should not be interpreted as a justification for reducing expert supervision. Rather, they underscore that AI can liberate radiologist time from lesion search, allowing greater focus on more valuable, integrative judgment.

AI-assisted reporting not only improved diagnostic detection but also enhanced the comprehensiveness, clarity, and consistency of radiology reports. However, although most AI studies in rib fracture imaging have focused on accuracy metrics, few have examined whether AI adds value to the content and linguistic quality of final report [23, 24]. In our study, the AI-assisted workflow increased explicit documentation of fracture number, laterality, segmental location, displacement status, associated thoracic findings, and recommendations concordant with the trauma team, while also improving impression specificity and reducing ambiguous language. A possible explanation is that AI-generated outputs promote more structured thinking during report composition, as readers are prompted to systematically verify and explicitly document salient lesion features. These findings suggest that AI can serve as a quality-enhancement tool for reporting, not merely a detection aid, which is an important consideration for multidisciplinary trauma care. However, caution is warranted, as templated or AI-influenced language may compromise nuance and detail in complex or equivocal cases.

The advantages of AI-assisted reporting persisted across all age groups, both sexes, isolated chest trauma and polytrauma, various injury mechanisms, and both low- and high-burden fracture presentations. This trend aligns with previous studies indicating the wide applicability of AI-based fracture detection across heterogeneous patient population [25]. The improvement was largest in patients with single fractures, who are typically most prone to under-detection and may be viewed as lower-risk cases. One explanation is that AI support may be most beneficial when lesion conspicuity is low, or when the number of lesions is insufficient to raise clinical suspicion. While these subgroup findings support broad clinical application rather than selective use, they also highlight that robustness observed in one center's subgroup does not guarantee generalizability across different centers, scanner platforms, clinical workflows, or disease prevalence settings.

The observed improvement in report quality should not be interpreted as equivalent to an enhancement in radiologists' diagnostic reasoning. In the AI-assisted workflow, the system generated structured information on fracture number, laterality, rib level, anatomical segment, and displacement status, thereby prompting radiologists to document these elements more consistently and reducing omissions, transcription burden, and variability in terminology. Accordingly, the advantages shown in **Table 6** are more appropriately regarded as workflow- and documentation-level benefits of

structured reporting rather than direct evidence of cognitive enhancement. By contrast, the potential diagnostic value of AI is better reflected by outcomes that require image interpretation and clinical adjudication, including fewer missed and false-positive fractures, improved detection of morphologically challenging lesions, and stronger concordance with the expert-consensus reference standard. Nevertheless, because the present retrospective pre-post design did not isolate the effects of lesion detection, structured template generation, and radiologist decision-making, their independent contributions cannot be determined. Future prospective studies should compare detection-only assistance, structured-report assistance, and their combined use, while assessing reader confidence, correction behavior, cognitive workload, and automation bias. Thus, structured AI output should be viewed as a complementary tool for improving reporting completeness, consistency, and clinical communication, whereas final diagnostic interpretation and responsibility remain with the radiologist.

Several limitations of this work should be acknowledged. First, it was a single center study, and the results may not be extrapolatable to other institutions, scanner platforms, environments for reporting, or patient populations. Second, although the study included real-world workflow results, the temporal grouping of conventional and AI-assisted cases introduces the potential for confounding by secular trends in workflow, staffing, or case handling. Third, the analysis was confined to rib fracture diagnosis and does not fully capture potential interactions between AI integration and other competing thoracic findings or priorities in trauma reporting. Fourth, although expert consensus was used as the reference standard, consensus adjudication is not entirely free from bias and does not fully replace outcome-based truth adjudication. Fifth, we did not directly evaluate radiologist trust calibration, cognitive workload, burnout, or potential automation bias, all of which are likely important factors in long-term adoption. Finally, external validation, multicenter implementation studies and prospective evaluation of patient-centered outcomes will be required before the results of this study can be generalized to large-scale implementation.

In conclusion, AI-assisted reporting for rib fracture assessment was associated with higher diagnostic accuracy, greater concordance with expert consensus, improved report quality, and significantly increased reporting efficiency compared with conventional interpretation. This benefit was seen in all fracture phenotypes, across all anatomic levels of agreement, workflow parameters, and clinically relevant subgroups, and in this regard, AI is best suited to be a radiologist-facing decision-support tool rather than a standalone tool. Overall, these findings support the use of AI models that are integrated with care to improve both the accuracy and efficiency of trauma imaging practice, and underscore the importance of further external validation, workflow adaptation and human oversight in the transition of these systems to clinical use.

Disclosure of conflict of interest

None.

Address correspondence to: Chen Yan, Department of Radiology, Ji'an County People's Hospital, No. 20, Luling Avenue, Dunhou Town, Ji'an 343100, Jiangxi, China. Tel: +86-0796-8442158; E-mail: Tarryxx@163.com

References

[1] Peek J, Ochen Y, Saillant N, Groenwold RHH, Leenen LPH, Uribe-Leitz T, Houwert RM and Heng M. Traumatic rib fractures: a marker of severe injury. A nationwide study using the National Trauma Data Bank. Trauma Surg Acute Care Open 2020; 5: e000441.

[2] O'Donovan S, van den Heuvel C, Baldock M, Humphries MA and Byard RW. Fatal blunt chest trauma: an evaluation of rib fracture patterns and age. Int J Legal Med 2022; 136: 1351-1357.

[3] Hoepelman RJ, Beeres FJP, Heng M, Knobe M, Link BC, Minervini F, Babst R, Houwert RM and van de Wall BJM. Rib fractures in the elderly population: a systematic review. Arch Orthop Trauma Surg 2023; 143: 887-893.

[4] Rogers FB, Larson NJ, Rhone A, Amaya D, Olson-Bullis BA and Blondeau BX. Comprehensive review of current pain management in rib fractures with practical guidelines for clinicians. J Intensive Care Med 2023; 38: 327-339.

[5] Peek J, Kremo V, Beks R, van Veelen N, Leiser A, Link BC, Houwert RM, Minervini F, Knobe M, Babst RH and Beeres FJP. Long-term quality of

life and functional outcome after rib fracture fixation. Eur J Trauma Emerg Surg 2022; 48: 255-264.
[6] Feng LR, Lilienthal M, Galet C and Skeete DA. Frailty as a predictor of negative outcomes in trauma patients with rib fractures. Surgery 2023; 173: 812-820.
[7] Polireddy K, Hoff C, Kinger NP, Tran A and Maddu K. Blunt thoracic trauma: role of chest radiography and comparison with CT - findings and literature review. Emerg Radiol 2022; 29: 743-755.
[8] Chen-Yoshikawa TF. Evolution of three-dimensional computed tomography imaging in thoracic surgery. Cancers (Basel) 2024; 16: 2161.
[9] Kuo RYL, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, Stewart M, Collins GS and Furniss D. Artificial intelligence in fracture detection: a systematic review and meta-analysis. Radiology 2022; 304: 50-62.
[10] Lex JR, Di Michele J, Koucheki R, Pincus D, Whyne C and Ravi B. Artificial intelligence for hip fracture detection and outcome prediction: a systematic review and meta-analysis. JAMA Netw Open 2023; 6: e233391.
[11] Calleja J, Muscat K, Calleja J and Firth G. Artificial intelligence in paediatric and adolescent fracture detection: a systematic review and meta-analysis. Cureus 2025; 17: e92199.
[12] Kutbi M. Artificial intelligence-based applications for bone fracture detection using medical images: a systematic review. Diagnostics (Basel) 2024; 14: 1879.
[13] Tieu A, Kroen E, Kadish Y, Liu Z, Patel N, Zhou A, Yilmaz A, Lee S and Deyer T. The role of artificial intelligence in the identification and evaluation of bone fractures. Bioengineering (Basel) 2024; 11: 338.
[14] Niiya A, Murakami K, Kobayashi R, Sekimoto A, Saeki M, Toyofuku K, Kato M, Shinjo H, Ito Y, Takei M, Murata C and Ohgiya Y. Development of an artificial intelligence-assisted computed tomography diagnosis technology for rib fracture and evaluation of its clinical usefulness. Sci Rep 2022; 12: 8363.
[15] Yao L, Guan X, Song X, Tan Y, Wang C, Jin C, Chen M, Wang H and Zhang M. Rib fracture detection system based on deep learning. Sci Rep 2021; 11: 23513.
[16] Wu M, Chai Z, Qian G, Lin H, Wang Q, Wang L and Chen H. Development and evaluation of a deep learning algorithm for rib segmentation and fracture detection from multicenter chest CT images. Radiol Artif Intell 2021; 3: e200248.
[17] Zhang B, Jia C, Wu R, Lv B, Li B, Li F, Du G, Sun Z and Li X. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation. Br J Radiol 2021; 94: 20200870.
[18] Huang ST, Liu LR, Tsai MF, Huang MY and 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.
[19] Iuga AI, Carolus H, Caldeira LL, Kottlors J, Rinneburger M, Weisthoff M, Fervers P, Rauen P, Fichter F, Goertz L, Niederau P, Siedek F, Heneweer C, Gietzen C, Pennig L, Dobrostal A, Laqua F, Woznicki P, Maintz D, Baessler B and Persigehl T. Fully automated AI-based lymph node measurements in chest CT: accuracy and reproducibility compared with multi-reader assessment. Diagnostics (Basel) 2026; 16: 967.
[20] Messina E, Borrelli A, Mezzapesa F, Laschena L, Lucciola S, Novelli S, Landini N, Pecoraro M, Sciarra A, Santini D and Panebianco V. Artificial intelligence-assisted reading of non-contrast prostate MRI: application and concordance with expert interpretation in a screening population within the PROSA trial. Eur Radiol 2026; 36: 5690-5700.
[21] Yacoub B, Varga-Szemes A, Schoepf UJ, Kabakus IM, Baruah D, Burt JR, Aquino GJ, Sullivan AK, Doherty JO, Hoelzer P, Sperl J and Emrich T. Impact of artificial intelligence assistance on chest CT interpretation times: a prospective randomized study. AJR Am J Roentgenol 2022; 219: 743-751.
[22] van Winkel SL, Rodríguez-Ruiz A, Appelman L, Gubern-Mérida A, Karssemeijer N, Teuwen J, Wanders AJT, Sechopoulos I and Mann RM. Impact of artificial intelligence support on accuracy and reading time in breast tomosynthesis image interpretation: a multi-reader multicase study. Eur Radiol 2021; 31: 8682-8691.
[23] Angeletti C, Zaccagna G, Vaccarili M, Salve G, De Vico A, Ciccozzi A and Divisi D. Contemporary and future perspectives on thoracic trauma care: surgical stabilization, multidisciplinary approaches, and the role of artificial intelligence. J Clin Med 2025; 14: 8041.
[24] Sun L, Fan Y, Shi S, Sun M, Ma Y, Zhang K, Zhang F, Liu H, Yu T, Tong H and Yang X. AI-assisted radiologists vs. standard double reading for rib fracture detection on CT images: a real-world clinical study. PLoS One 2025; 20: e0316732.
[25] Jung J, Dai J, Liu B and Wu Q. Artificial intelligence in fracture detection with different image modalities and data types: a systematic review and meta-analysis. PLOS Digit Health 2024; 3: e0000438.