



Deep learning techniques for rib fractures diagnosis using computed tomography images: a state-of-the-art review

Samuel Chikasha¹

Serestina Viriri²

¹Wits University, School of Public Health,
Johannesburg, South Africa

²University of KwaZulu-Natal, Computer Science,
Durban, South Africa

PURPOSE

Quantitative analysis of rib fractures using computed tomography (CT) is critical for accurate diagnosis, severity assessment, and treatment planning in trauma care. Recent advances in deep learning (DL) have markedly improved automated rib fracture analysis, occasionally achieving performance comparable with or exceeding that of radiologists. However, the translation of these advances into routine clinical workflows remains limited.

METHODS

As this study is a state-of-the-art with a structured search strategy, elements of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 guidelines were adopted to enhance transparency in the literature search and selection process. A comprehensive literature search was performed across the Web of Science, Scopus, PubMed, Institute of Electrical and Electronics Engineers Xplore databases, as well as the Google Scholar search engine, to identify peer-reviewed English-language studies published between January 2020 and March 2025. Studies applying DL techniques for rib fracture detection, classification, or segmentation using CT imaging were included. Following screening and eligibility assessment, 38 studies were retained for qualitative synthesis.

RESULTS

Across the reviewed studies, DL-assisted diagnostic approaches demonstrated favorable performance relative to conventional radiological assessment, although the magnitude of benefit varied across study designs and evaluation settings. Detection dominated the research landscape, accounting for > 80% of included studies, whereas classification and segmentation remained comparatively underexplored. Reported performance metrics frequently demonstrated > 90% sensitivity under experimental conditions, and artificial intelligence (AI) assistance reduced interpretation time by up to 40% in clinical simulations. Geographically, research contributions were concentrated in a limited number of countries, with China accounting for approximately 57% of publications. Despite promising performance, real world clinical integration was uncommon.

CONCLUSION

DL has reached a high level of technical maturity for rib fracture detection; however, its potential clinical impact is constrained by imbalanced research focus, limited external validation, and insufficient workflow integration. Greater emphasis on fracture classification and segmentation is necessary to support comprehensive clinical decision-making.

CLINICAL SIGNIFICANCE

This review was limited by heterogeneity across study designs, datasets, and evaluation metrics, which precluded meta-analysis. Most studies relied on retrospective, single-center datasets, thereby restricting generalizability and real-world applicability. Future research should prioritize multi-task learning frameworks capable of simultaneously detecting, classifying, and segmenting fractures, alongside well-annotated multicenter datasets to enhance robustness and generalizability. Incorporating ensemble learning and explainable AI techniques is essential to improve interpretability, reduce false positives, and foster clinician trust, thereby supporting regulatory approval and sustainable integration into clinical workflows.

KEYWORDS

Computed tomography, rib fractures, deep learning, artificial intelligence, fracture diagnosis, medical image analysis, clinical decision support, explainable artificial intelligence

Handling editor: Furkan Ufuk

Corresponding author: Samuel Chikasha

E-mail: samuel.chikasha@wits.ac.za

Received 17 March 2026; revision requested 06 April 2026; last revision requested 04 June 2026; accepted 08 June 2026.



Epub: 29.09.2026

DOI: 10.4274/dir.2026.263967

Rib fractures are among the most frequently encountered injuries in patients with thoracic trauma,¹ typically resulting from high-impact mechanisms, such as motor vehicle collisions, falls, and blunt force trauma.² Although often perceived as minor injuries, rib fractures are strongly associated with high morbidity and mortality, particularly among elderly patients and those with multiple fractures.³ Complications, including impaired ventilation, pneumonia, pneumothorax, and hemothorax, substantially worsen clinical outcomes and prolong hospital stay.⁴ Consequently, early and accurate detection of rib fractures is essential for timely intervention, appropriate pain management, and the prevention of secondary complications.

Computed tomography (CT) has become the imaging modality of choice for rib fracture assessment due to its superior sensitivity compared with conventional radiography, particularly for non-displaced and subtle fractures.⁵ However, CT-based interpretation remains labor-intensive and time-consuming, requiring careful multiplanar evaluation of large image volumes. Missed fractures remain a persistent challenge, especially in busy trauma settings and in cases involving complex anatomy or subtle fracture lines.⁵ These limitations have driven increasing interest in automated and semi-automated diagnostic solutions aimed at improving accuracy, consistency, and efficiency.

Main points

- Deep learning (DL) methods have the potential to substantially advance automated rib fracture diagnosis using computed tomography imaging, although most of the research has focused on fracture detection.
- Current literature shows limited research on rib fracture classification and segmentation compared with detection tasks.
- Although many DL models report high diagnostic performance, most systems remain at the experimental stage and are rarely integrated into clinical radiology workflows.
- Future research should prioritize balanced development of detection, classification, and segmentation models, alongside clinically validated systems that can be integrated into routine trauma imaging workflows.
- There is a persistent lack of reporting transparency in artificial intelligence-based rib fracture research, particularly with respect to external validation and annotation practices, which limits reliable comparison across studies and hampers clinical translation.

Recent advances in artificial intelligence (AI), particularly deep learning (DL), have transformed medical image analysis. DL models have demonstrated strong performance across a wide range of diagnostic imaging tasks, including fracture detection, classification, and segmentation.^{6,7} In the context of rib fracture diagnosis, CT imaging provides high-resolution, three-dimensional (3D) representations of the rib cage, enabling DL models to learn fracture-related spatial and textural features that may be difficult to perceive visually.^{3,4,6} Consequently, a growing body of research has explored convolutional neural networks (CNNs) and related architectures for automated rib fracture detection and characterization.⁷⁻¹¹

Despite this rapid progress and promising performance in experimental settings, translation into routine clinical workflows remains limited. Many proposed models are evaluated using retrospective, single-center datasets, with limited external validation and inconsistent reporting of clinical integration, interpretability, and human-AI interaction. Furthermore, existing research has disproportionately focused on fracture detection, with comparatively fewer studies addressing fracture classification (e.g., displaced vs. non-displaced, acute vs. healing) or precise segmentation, despite their importance for treatment planning and prognostication.

Types of rib fractures

Understanding the various types and classifications of rib fractures is essential for developing effective AI-based diagnostic systems. Rib fractures can be categorized according to displacement, fracture age, complexity, and associated complications. Figure 1 illustrates these categories, including normal rib anatomy on axial CT, simple non-displaced fractures, displaced fractures, comminuted fractures, healing fractures with early callus formation, and old healed fractures with mature callus.

Displacement-based classification includes non-displaced fractures, in which bone fragments remain in anatomical alignment, and displaced fractures, where bone fragments have shifted from their original position, potentially leading to complications. Angulated fractures are characterized by angular deformity at the fracture site.

Temporal classification distinguishes acute fractures (0–2 weeks), which typically present with sharp fracture lines and no callus formation, from healing fractures (2–8

weeks), where early callus formation becomes visible, and the fracture line becomes less distinct. Chronic fractures (> 8 weeks) demonstrate mature callus formation and near-complete healing.

Complexity-based classification includes simple fractures, involving a single fracture line, and comminuted fractures, which consist of multiple bone fragments at the fracture site.

The complexity and variability of rib fracture presentations highlight the diagnostic challenges that motivate AI-assisted systems. Each fracture type presents distinct imaging characteristics that DL models must recognize, ranging from subtle hairline fractures that are easily missed to complex comminuted patterns requiring urgent intervention.

Overview of rib fracture diagnosis

Rib fracture diagnosis comprises three core tasks: detection, classification, and segmentation, each representing increasing levels of diagnostic detail and clinical utility. Image analysis approaches range from manual interpretation to computer-aided detection and fully automated DL-based systems. Manual diagnosis relies on expert visual assessment of multiplanar CT reconstructions and remains time-consuming and prone to error, particularly for non-displaced fractures. Computer-aided detection systems provide algorithmic assistance to radiologists, whereas fully automated systems aim to independently identify and characterize fractures with minimal human intervention.

Rib fracture detection

Detection focuses on determining the presence of a fracture. CT imaging offers superior sensitivity compared with radiography, particularly for subtle injuries. DL-based detection systems, typically implemented using object detection frameworks, have demonstrated improved sensitivity and diagnostic efficiency, occasionally outperforming junior radiologists.¹²

Rib fracture classification

Classification involves characterizing fracture type, severity, and anatomical location, thereby informing treatment decisions, such as surgical vs. conservative management. DL models can learn discriminative imaging features and may incorporate clinical information to enhance performance, providing a more comprehensive representation of the injury.¹³

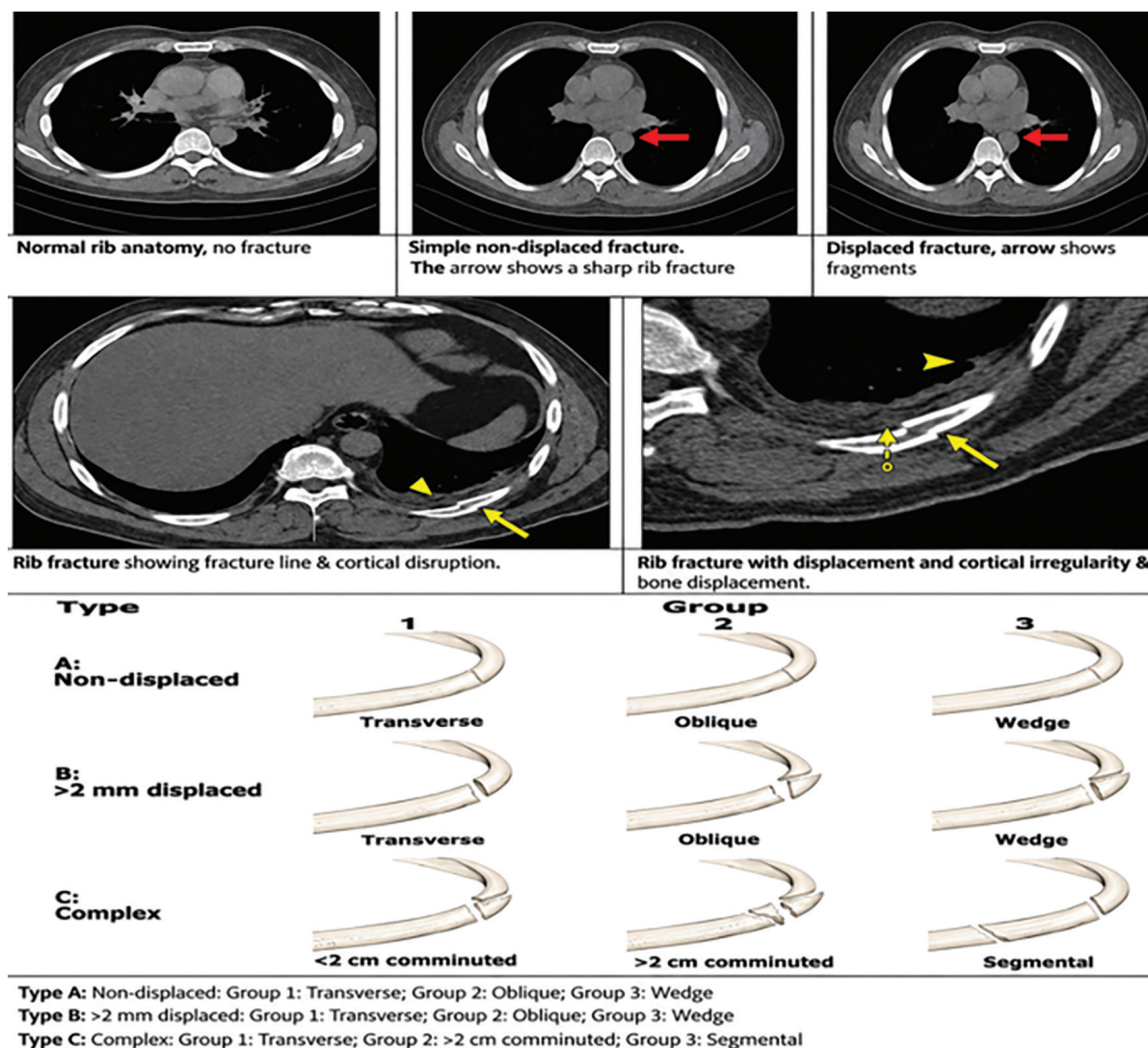


Figure 1. Rib fracture types and classification on computed tomography images (Karlson and French⁴³, Prins et al.⁴⁴).

Rib fracture segmentation

Segmentation provides voxel-level delineation of fracture regions, enabling precise localization, 3D visualization, and longitudinal monitoring. U-Net-based architectures and their variants are widely used due to their effectiveness in capturing fine anatomical detail.^{14,15} Accurate segmentation supports surgical planning, fracture quantification, and clinician–patient communication.

Purpose and contribution of this review

The objective of this state-of-the-art review is to critically synthesize and evaluate studies published between 2020 and 2025 that apply DL techniques to rib fracture detection, classification, or segmentation using CT imaging. The aim is to assess the current state of AI in this domain.

Specifically, this review addresses the following research questions:

1. What is the status of DL models for rib fracture diagnosis across detection, classification, and segmentation tasks, and how mature is their clinical applicability?
2. Which model architectures and methodological approaches are most used, and how does their performance compare with radiologist assessment?
3. What barriers limit clinical workflow integration, and what methodological, technical, and clinical gaps must be addressed to enable real-world deployment?

By systematically analyzing model performance, study design, data characteristics, evaluation metrics, and reported clinical implementation, this review moves beyond descriptive reporting to provide a structured synthesis of the field. It identifies key limitations in current research and outlines evidence-based directions for developing robust, interpretable, and clinically deployable DL frameworks for rib fracture diagnosis.

Methods

This state-of-the-art review adopted selected elements of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to enhance transparency in the identification and selection of relevant studies. A structured search strategy was implemented across multiple databases, and the screening process is illustrated using a PRISMA-style flow diagram. However, this study does not meet the criteria of a full systematic review, as no formal protocol registration, comprehensive risk-of-bias assessment, or quantitative meta-analysis was conducted.

Two independent reviewers screened titles, abstracts, and full-text articles, with disagreements resolved through consensus. Data extraction included study characteristics, dataset size, validation strategy, and reported performance metrics. A structured risk-of-bias assessment was conducted using

predefined criteria, including dataset diversity, external validation, and annotation quality.

The objective of this review was to systematically identify, evaluate, and synthesize published evidence on the application of DL models for rib fracture detection, classification, and segmentation using CT imaging. The narrative synthesis and critical discussion were conducted to ensure clarity, balanced interpretation of findings, and overall methodological rigor and reproducibility.

Information sources and search strategy

Although this review does not constitute a fully systematic review, a comprehensive literature search was performed across five electronic databases: Clarivate Analytics' Web of Science (WoS), Elsevier's Scopus, PubMed, Institute of Electrical and Electronics Engineers (IEEE) Xplore, and the search engine Google Scholar. These databases were selected to ensure broad coverage of both clinical and technical literature.

PubMed was included due to its focus on biomedical and radiological research, where clinical validation of machine learning approaches is commonly reported. The IEEE Xplore database was selected to capture technically oriented conference proceedings and methodological advances in DL. WoS and Scopus were included for their comprehensive indexing and citation tracking capabilities, supporting cross-disciplinary coverage and systematic search practices.

The search was conducted between 25 March 2025 and 2 May 2025. Studies were restricted to those published between January 2020 and March 2025, reflecting the period during which DL approaches became widely adopted in medical image analysis and diagnostic imaging.^{16,17}

The search strategy combined controlled vocabulary terms with free-text keywords related to rib fractures, DL, and diagnostic tasks. A representative PubMed search query is provided below:

((rib fracture[Title] OR rib fractures[Title]) AND ("deep learning"[MeSH Terms] OR ("deep"[All Fields] AND "learning"[All Fields]) OR "deep learning"[All Fields])) AND (detection[Title] OR classification[Title] OR segmentation[Title])) AND ("2020/01/01"[PubDate]: "2025/03/31"[PubDate]) NOT retracted[All Fields] NOT preprints[All Fields] NOT editorials[All Fields].

An equivalent Boolean logic was adapted for the stated databases to ensure consistency across platforms. For Google Scholar, ("rib fracture" OR "rib fractures") AND "deep learning" AND (detection OR classification OR segmentation) were used, then filtered by year. Further analysis was conducted to remove conference papers and articles studies that were not compliant with the inclusion criteria.

Study eligibility was determined using predefined inclusion and exclusion criteria consistent with the objectives of a state-of-the-art review. Eligible studies included peer-reviewed journal articles and conference proceedings published in English that applied DL techniques to rib fracture detection, classification, or segmentation using CT as the primary imaging modality. Studies were required to report quantitative performance outcomes to enable a comparative methodological appraisal.

Studies were excluded if they were non-peer-reviewed publications, editorials, opinion pieces, letters, case reports, or state-of-the-arts. Articles that did not focus on rib

fracture diagnosis and did not employ any DL methods were also excluded.

Study selection process and Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow description

A PRISMA-style flow diagram was used to illustrate the study selection process, although a full systematic review protocol was not followed. The study selection process followed the PRISMA framework (see checklist Appendix A). The article selection process is summarized visually in Figure 2 and is described below to ensure transparency and reproducibility.

A comprehensive database search across WoS, Scopus, PubMed, IEEE Xplore, and Google Scholar identified 179 records published between January 2020 and March 2025. After removal of 10 duplicate records, 169 unique studies remained for screening.

Title and abstract screening excluded studies that were not related to rib fractures, did not employ DL techniques, were not CT-based, or were non-peer-reviewed, resulting

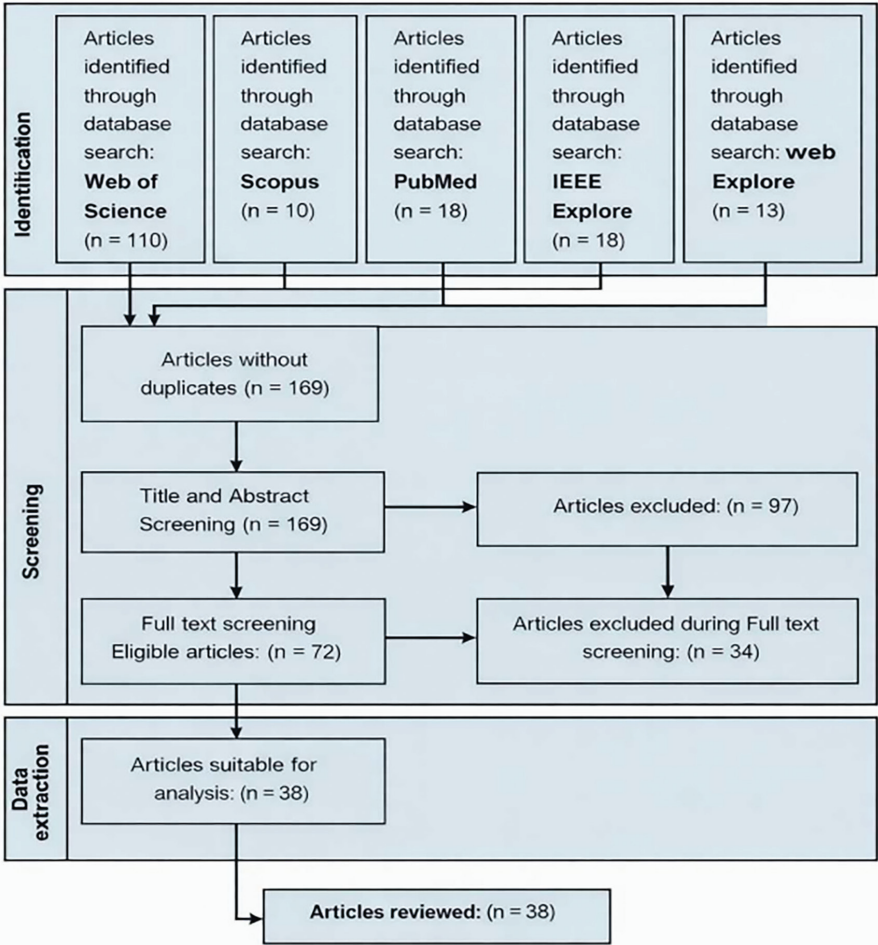


Figure 2. The article selection process. IEEE, Institute of Electrical and Electronics Engineers.

in 72 articles retained for fulltext evaluation. These fulltext articles were assessed against predefined inclusion and exclusion criteria. Following fulltext review, 34 articles were excluded due to reasons such as lack of CT imaging, absence of DL methods, insufficient performance reporting, or focus on unrelated anatomical structures. Ultimately, 38 studies met all eligibility criteria and were included in the qualitative synthesis.

Data extraction

Data extraction was conducted using a standardized data collection framework to ensure consistency across all included studies. For each article, bibliographic information, including authorship and year of publication, was recorded alongside the country of origin and study setting. Study design characteristics, dataset composition, sample size, and imaging modality were extracted to contextualize methodological variability. Details of the DL architecture, training strategy, and diagnostic task addressed (i.e., fracture detection, classification, or segmentation) were documented. In addition, information on input data characteristics, preprocessing methods, evaluation metrics, and reported performance was collected. Where available, comparisons with radiologists or conventional diagnostic approaches were noted, together with evidence of clinical validation or workflow integration. Finally, the extraction process captured reported model interpretability or explainability techniques, as well as stated challenges, limitations, and proposed directions for future research.

This structured extraction approach aligns with recent guidance on transparency and reporting in AI-based medical imaging research.¹⁷

Quality assessment and risk of bias

Given the heterogeneity of study designs, datasets, and evaluation protocols, a formal meta-analysis was not feasible. Methodological quality and potential sources of bias were assessed qualitatively using criteria informed by PRISMA guidance and AI-specific reporting recommendations. This assessment considered dataset size and diversity, the use of external or multicenter validation strategies, and the clarity with which model architectures and training procedures were described. In addition, the appropriateness of evaluation metrics and the transparency of reporting regarding study limitations and clinical applicability were examined to identify factors that may influence the ro-

bustness, generalizability, and translational relevance of the reported findings.

Each study was evaluated across these dimensions to identify strengths, weaknesses, and common sources of bias that may influence reported performance.

Risk of bias considerations using the Prediction Model Risk of Bias Assessment Tool for Artificial Intelligence

The risk of bias consideration was explored using the Prediction Model Risk of Bias Assessment Tool for Artificial Intelligence (PROBAST AI). The tool was applied descriptively to evaluate methodological transparency and potential sources of bias, rather than as a formal inclusion or exclusion criterion.

All included studies were examined for reporting related to the four PROBAST AI domains: participants, predictors, outcome, and analysis. In accordance with PROBAST AI guidance, risk of bias judgment was based strictly on information explicitly reported by the study authors, and no assumptions were made where reporting was incomplete.

Only studies that clearly reported dataset size, annotation procedures, and validation strategy were considered eligible for structured PROBAST AI domain-level assessment. Studies lacking one or more of these elements were not assigned risk of bias judgments and were retained solely for narrative synthesis.

Data synthesis

Quantitative meta-analysis was not performed due to substantial heterogeneity in datasets, architectures, outcome measures, and reporting standards. However, findings were synthesized using a structured qualitative approach, grouping studies according to diagnostic task (detection, classification, or segmentation) and model architecture. Comparative analyses focused on performance trends, clinical relevance, and barriers to real-world implementation rather than pooled accuracy estimates.

Results

Overview of included studies

The studies collectively explore the application of DL to rib fracture detection, classification, and segmentation using CT imaging. Considerable heterogeneity was observed across study design, datasets, model architectures, evaluation metrics, and validation strategies.

Risk of bias assessment

Of the 38 studies included in this state-of-the-art review, only 10 (26%) reported sufficient methodological detail to permit structured PROBAST AI assessment. The remaining 28 studies (74%) did not report one or more key elements required for PROBAST AI evaluation, most commonly external validation strategy and annotation quality, and were therefore not assigned risk of bias categories.

Among the 10 assessable studies, 3 were judged to have low overall risk of bias, and 7 were judged to have moderate overall risk of bias. No study demonstrated high-risk of bias in any PROBAST AI domain. Low-risk studies were characterized by multicenter or independent external validation and expert radiologist annotation with consensus or senior adjudication. Moderate risk ratings were primarily driven by limitations in external validation rather than deficiencies in outcome definition.

Table 1 summarizes dataset size, external validation strategy, annotation quality, and overall risk level for the subset of studies that reported sufficient methodological detail to permit a structured assessment.

Green shading indicates low-risk of bias, and yellow shading indicates moderate risk of bias. Overall risk ratings reflect domain-level judgments based strictly on explicitly reported methodological information.

A detailed PROBAST-AI traffic-light risk-of-bias assessment for studies with sufficient methodological reporting is presented in Table 2 below.

Quantitative synthesis

Due to heterogeneity in datasets, tasks, and evaluation metrics, statistical meta-analysis was not feasible. Table 3 provides a structured, meta-analysis-like synthesis summarizing task type, dataset characteristics, validation strategy, and reported performance ranges.

To improve clarity and analytical rigor, results are presented below according to diagnostic task, followed by geographical distribution, performance metrics, and comparison with human readers. The general trend of DL in rib fracture diagnosis in terms of the study focus during the period under review is as shown in Figure 3 below.

Study	Dataset size	External validation	Annotation quality	Overall risk level
Wu et al., ³⁶ (2021)	10,943 computed tomography (CT) scans (6 hospitals)	Yes–multicenter	1–3 radiologists + senior consensus	Low
Li et al., ³⁷ (2023)	18,172 CT scans (8 hospitals)	Yes–independent test set	3 senior physicians (consensus)	Low
Zhou et al., ¹ (2021)	1,020 CT scans (26,580 annotations)	Yes–multicenter datasets	2 radiologists + senior + surgeon consensus	Low
Jin et al., ⁴⁰ (2020)	900 CT scans (7,473 fractures)	No	5 radiologists + senior confirmation	Low–moderate
Chai et al., ²¹ (2024)	Multi-dataset (thousands of scans)	Yes–cross-dataset validation	Multiple radiologists (consensus)	Low–moderate
Kaiume et al., ³⁰ (2021)	3,644 CT scans + 39 external patients	Yes–small external test set	≥ 3 radiologists training; 2 for testing	Moderate
Burkow et al., ³⁹ (2024)	1,109 radiographs	No	7 pediatric radiologists	Moderate
Meng et al., ³⁵ (2021)	8,529 CT scans	Partial	3–4 senior radiologists	Moderate
Weikert et al., ³³ (2020)	11,455 CT scans + 511 tests	Partial	Board-certified radiologist; subset double-read	Moderate
Sun et al., ¹⁸ (2025)	243 patients (826 lesions)	Yes–independent dataset	2 musculoskeletal radiologists	Moderate

Study	Participants	Predictors	Outcome	Analysis	Overall risk
Wu et al., ³⁶ (2021)	Low	Low	Low	Low	Low
Li et al., ³⁷ (2023)	Low	Low	Low	Low	Low
Zhou et al., ¹ (2021)	Low	Low	Low	Low	Low
Jin et al., ⁴⁰ (2021)	Low	Low	Low	Moderate	Low–moderate
Chai et al., ²¹ (2024)	Low	Low	Low	Moderate	Low–moderate
Kaiume et al., ³⁰ (2021)	Moderate	Low	Moderate	Moderate	Moderate
Burkow et al., ³⁹ (2024)	Moderate	Low	Moderate	Moderate	Moderate
Meng et al., ³⁵ (2021)	Moderate	Low	Moderate	Moderate	Moderate
Weikert et al., ³³ (2020)	Moderate	Low	Moderate	Moderate	Moderate
Sun et al., ¹⁸ (2025)	Moderate	Low	Moderate	Moderate	Moderate

Task	No. of studies	Dataset type	Validation	Performance metrics	Clinical evaluation
Detection	34	Mostly single-center computed tomography (CT)	Internal ± limited external	Sensitivity 82%–94%	Occasional reader studies
Classification	5	Single-center CT	Internal only	Accuracy 80%–92%	Rare
Segmentation	5	Annotated CT volumes	Internal only	Dice 0.65–0.80	None

Deep learning tasks in rib fracture diagnosis

Rib fracture detection

Rib fracture detection was the most extensively researched task, accounting for 34 of 38 studies (89.5%). Most detection approaches framed the problem as an object detection or slice-level classification task, aiming to identify the presence and location of fractures within CT volumes.

Across studies, DL-based detection consistently demonstrated strong diagnostic performance. Reported sensitivity values

frequently exceeded 90% under controlled experimental conditions, with false positive rates varying depending on dataset complexity and annotation quality. Models such as Faster R-CNN, You Only Look Once variants, RetinaNet, and FracNet-based architectures were most commonly employed. Several studies reported that AI-assisted detection improved radiologist sensitivity by 10%–25% and reduced reporting time by approximately 30%–40%, particularly benefiting less experienced readers. Detection therefore represents the most mature and clinically evaluated application of DL in rib fracture diagnosis.

Rib fracture classification

Only 5 studies (13.2%) explicitly focused on fracture classification. Classification tasks included discrimination between displaced and non-displaced fractures, acute vs. healing fractures, and categorization based on fracture morphology or rib level. Classification models typically used CNNs trained on localized rib regions or fracture-centered patches. Performance was generally lower and more variable than in detection models, reflecting greater task complexity and limited annotated data. Few studies integrated classification outputs into downstream clinical decision-making workflows.

The relative scarcity of classification-focused research represents a critical gap, given the importance of fracture type and chronicity for treatment planning and prognostic assessment.

Rib fracture segmentation

Segmentation was addressed in 5 studies (13.2%), primarily using U-Net-based architectures and their 3D or attention-enhanced variants. Segmentation models aimed to delineate fracture boundaries at the voxel or pixel level, enabling precise localization and volumetric analysis. Reported Dice similarity coefficients ranged widely, reflecting differences in ground truth annotation quality, fracture subtlety, and imaging protocols. Although segmentation offers high clinical value for surgical planning and longitudinal monitoring, few studies evaluated these models in real-world clinical environments.

The included studies were distributed across the three diagnosis tasks, as shown in Table 4 below.

Global distribution of research

The geographical analysis revealed a substantial concentration of these research activities in specific regions. China contributed approximately 57.7% of included studies, followed by Japan (10.5%), the United States (7.9%), and Switzerland (7.9%). The UK, France, the Netherlands, Taiwan, and the UAE contributed the remaining 16.0%. No eligible studies originated from Africa. This geographic distribution of studies is shown in Figure 4, in the form of a Sankey diagram showing the country of study, DL architecture used, and the study focus. The distribution raises concerns regarding dataset diversity, population bias, and generalizability of reported model performance.

Performance metrics and reporting practices

The most frequently reported evaluation metrics were as follows: a) sensitivity/recall, which was reported in ~92% of studies; b) specificity (~84%); c) F1 score (~76%); d) precision (~71%); e) area under the curve (AUC) (~45%); and f) Dice coefficient (~32%; primarily segmentation studies). Metric reporting lacked standardization, limiting direct comparison across studies and precluding quantitative meta-analysis.

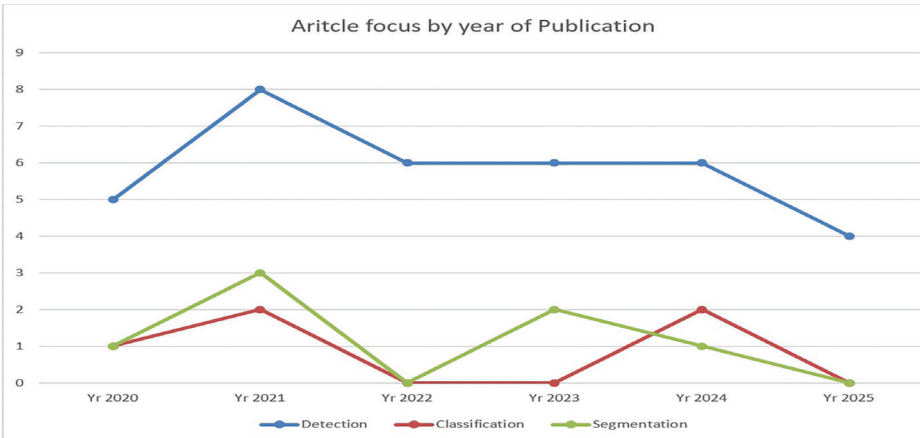


Figure 3. Deep learning in rib fracture diagnosis.

Table 4. Overview of studies' focus on diagnosis tasks		
Reference	Author (year)	Focus
(18)	Sun et al., 2025	Detection
(8)	Kayo et al., 2025	
(19)	Collins et al., 2025	
(20)	Chai et al., 2025	
(21)	Chai et al., 2024	
(22)	van den Broek et al., 2024	
(23)	Ye et al., 2023	
(24)	Tan et al., 2023	
(9)	Kuo et al., 2022	
(25)	Yu et al., 2023	
(26)	Azuma et al., 2022	
(27)	Blum et al., 2021	
(14)	Zhang et al., 2021	
(5)	Cho et al., 2012	
(28)	Niiya et al., 2022	
(29)	Chai, Z. et al., 2022	
(11)	Zhou, Z. et al., 2022	
(30)	Kaiume et al., 2021	
(31)	Yao et al., 2021	
(7)	Ibanez et al., 2022	
(32)	Alzahrani et al., 2021	
(33)	Weikert et al., 2020	
(6)	Lopez-Melia et al., 2024	
(3)	Yang et al., 2024	
(34)	Cao et al., 2023	
(35)	Meng et al., 2021	
(15)	Haitaamar and Abdulaziz, 2021	
(36)	Wu et al., 2021	
(37)	Li et al., 2023	
(38)	Chan et al., 2023	
(12)	Jin et al., 2024	
(39)	Burkow et al., 2024	
(40)	Jin et al., 2020	Classification
(13)	Wang and Wang, 2023	
(6)	Lopez-Melia et al., 2024	
(3)	Yang et al., 2024	
(14)	Zhang et al., 2021	Segmentation
(1)	Zhou, et al., 2021	
(34)	Cao et al., 2023	
(11)	Zhou et al., 2022	
(35)	Meng et al., 2021	
(14)	Zhang et al., 2021	
(15)	Haitaamar and Abdulaziz, 2021	
(36)	Wu et al., 2021	

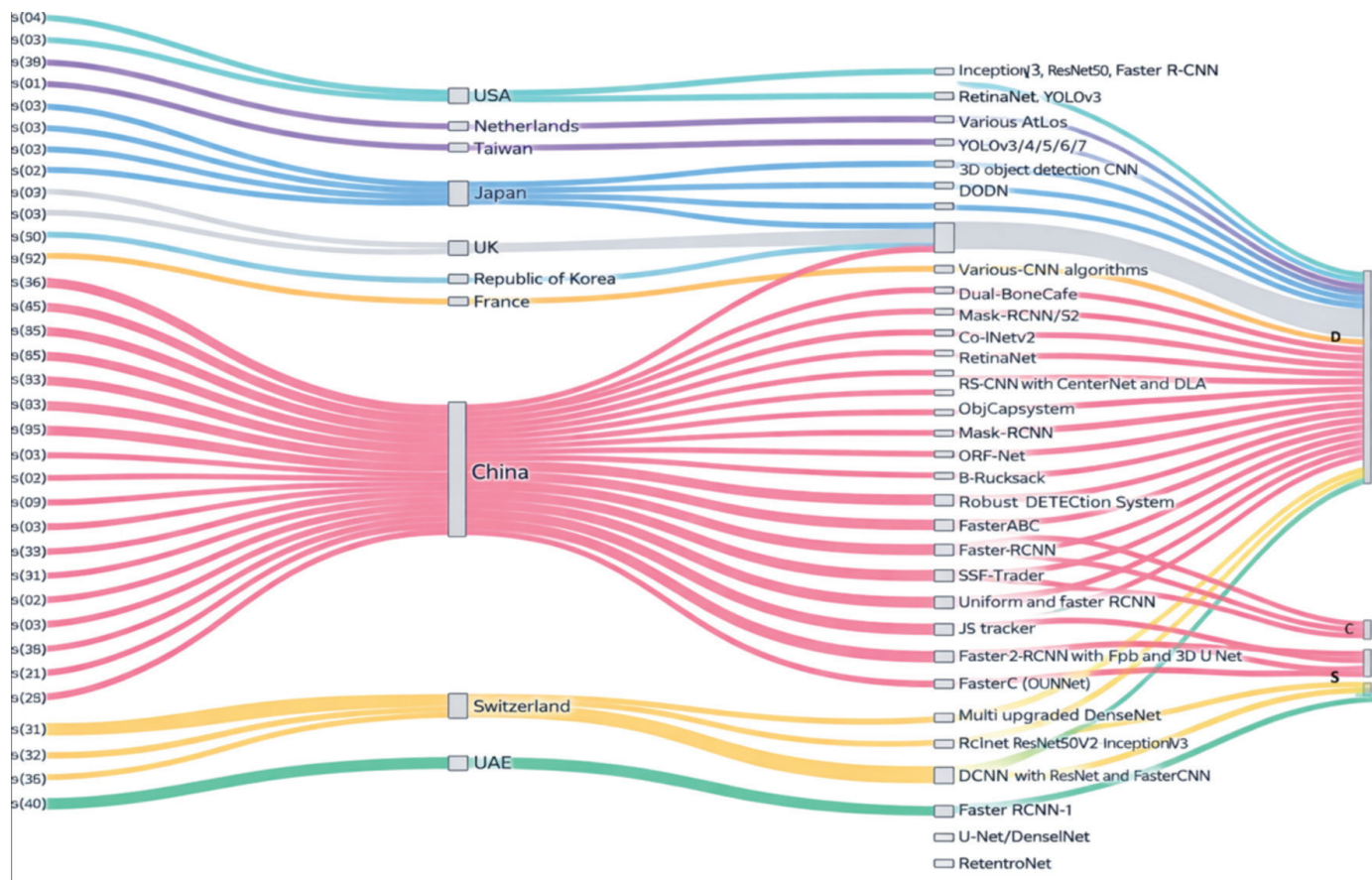


Figure 4. Sankey diagram of source, country, deep learning architecture, and focus: D = Detection, C = Classification, and S = Segmentation. CNN, convolutional neural network; R-CNN, region-based convolutional neural network; YOLO, You Only Look Once; DLA, deep layer aggregation; DCNN, deep convolutional neural network; ResNet, residual network; DenseNet, densely connected convolutional network; FPN, feature pyramid network; 3D, three-dimensional.

Comparison with human readers

Twelve studies explicitly compared DL performance with radiologists. The comparisons revealed that 1) AI-assisted diagnosis consistently outperformed traditional double reading; 2) sensitivity improved, with studies reporting 10%–25% improvement in sensitivity when AI assistance was used; 3) reading time was reduced, with an average reduction of 30%–40% in diagnostic time; and 4) AI assistance was particularly beneficial for less experienced clinicians. The results showed that, in terms of sensitivity, AI-assisted diagnosis achieved 94.2% against 69.2% for double reading. Internal sensitivity increased from 68.82% to 91.76% with AI assistance. These findings support the role of DL as an augmentative tool rather than a replacement for clinicians.

Clinical workflow integration

Only a minority of studies evaluated workflow integration. Reported challenges included false-positive burden, reader fatigue, lack of picture archiving and communication

systems (PACSs) integration, medico-legal responsibility, and regulatory uncertainty. AI systems consistently performed better as second readers rather than autonomous tools.

Deep learning architectures usage in rib fracture diagnosis

Recent breakthrough approaches were associated with FracNet variants, in which FracNet+ achieved a free-response receiver operating characteristic score of 84.29% and an F1-score of 0.3038, and SA-FracNet incorporated self-attention mechanisms for improved detection. Three-dimensional FracNet was extended to volumetric processing with a Dice coefficient of 71.5%. The ORF-Net Series ORF-Netv2 network achieved a mean average precision of 34.7% on the RibFrac dataset, with an average precision at 50% intersection over union of 48.1%. Deep Omni-Supervised Learning was used to address limited annotation challenges. Several models have been used in isolation in rib fracture diagnosis, as shown in Figure 5 below.

Multi-modal approaches

Some studies propose exploring a combination of CT imaging with clinical data and patient history for comprehensive assessment.

Ensemble methods

Ensemble learning and multimodal fusion (combining CT with clinical variables and history) are increasingly used to improve robustness and accuracy. Multiple studies report > 90% accuracy with ensembles. For example, Wu et al.⁴¹ integrated five CNN architectures, achieving 96% accuracy, 97% AUC, and 97% recall. Den Hengst et al.⁴² showed that DL systems can exceed clinician performance, with higher sensitivity for detection (86.7% vs. 75.4%) and classification (97.3% vs. 88.2%). Most current ensembles are static; there is an opportunity to develop adaptive/dynamic ensembles and to integrate multimodal inputs for real-time clinical deployment.

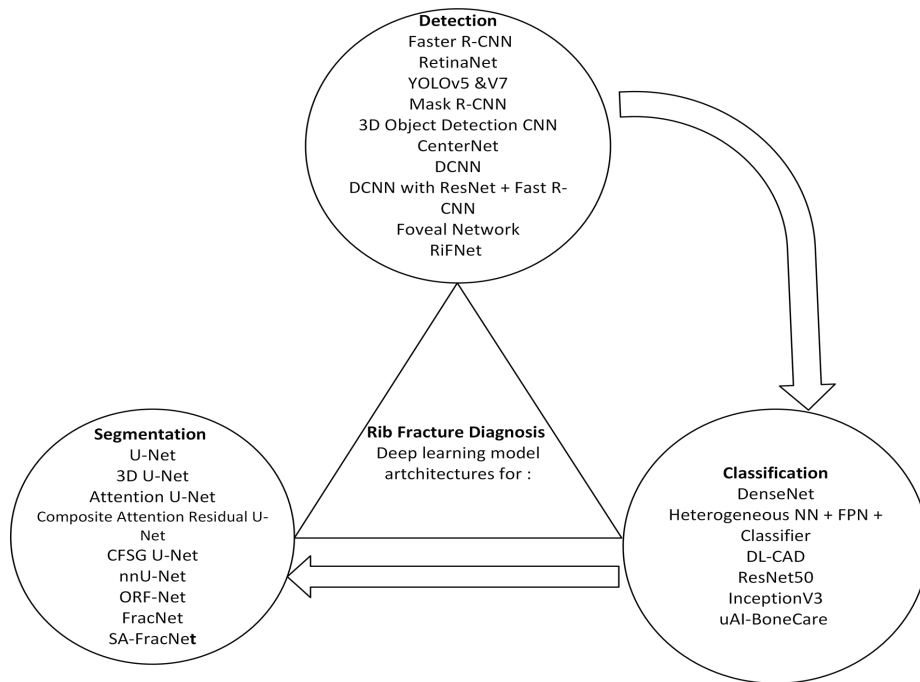


Figure 5. Deep learning architecture usage in rib fracture diagnosis. R-CNN, region-based convolutional neural network; YOLO, You Only Look Once; 3D, three-dimensional; CNN, convolutional neural network; DCNN, deep convolutional neural network; ResNet, residual network; U-Net, U-shaped network; nnU-Net, no-new-Net; FPN, feature pyramid network; NN, neural network; DL-CAD, deep learning-based computer-aided detection/diagnosis; AI, artificial intelligence.

Technical achievements

AI-assisted diagnosis consistently outperformed traditional approaches. Detection studies still dominate research focus, with 89.5% of the studies focusing on detection, whereas classification and segmentation remain underexplored. Furthermore, diverse architectural approaches, such as Faster R-CNN and U-Net variants, are the most adopted. Performance metrics showed a high sensitivity of > 90% and specificity of > 85% in optimal conditions.

Research landscape

The research landscape was found to be concentrated in very few countries. Hence, there is limited global diversity. There is also a predominant use of data augmentation as a preprocessing technique, as well as limited real-world clinical validation despite promising laboratory results.

Critical gaps

The survey revealed some critical gaps, including (i) insufficient focus on classification and segmentation tasks, (ii) lack of multicenter, diverse datasets, which limits generalizability, (iii) limited clinical workflow

integration despite experimental success, and (iv) absence of standardized evaluation protocols and benchmarks.

Discussion

This state-of-the-art review synthesizes recent advances in DL-based rib fracture diagnosis using CT imaging, covering detection, classification, and segmentation tasks. The findings demonstrate substantial progress in automated rib fracture diagnosis. The review demonstrates that DL models show high diagnostic performance for rib fracture detection, particularly in controlled settings. However, performance varies considerably depending on dataset characteristics and validation strategies. Studies using external validation reported lower but more realistic performance metrics, highlighting concerns regarding generalizability. These findings emphasize the need for larger, multicenter datasets and standardized evaluation frameworks to facilitate clinical translation.

Compared with prior systematic reviews, this work assesses state-of-the-art techniques and expands the scope to classification, segmentation, workflow integration, and explainability considerations, as it ques-

tions the reasons for the lack of workflow integration.

Maturity of deep learning for rib fracture detection

Among the three diagnostic tasks, fracture detection has reached the highest level of technical maturity. Most included studies reported sensitivity of > 90%, often with meaningful reductions in diagnostic reading time when AI systems were used as decision support tools. These improvements are clinically significant, particularly in trauma settings where rapid and accurate interpretation is critical. Importantly, comparative studies consistently showed that AI-assisted reading improves sensitivity relative to unaided double reading, especially for subtle or non-displaced fractures and for less-experienced radiologists.

Nevertheless, detection-focused research has largely emphasized algorithmic performance under controlled conditions, often using retrospective, single-center datasets. Although these studies establish proof of feasibility, they do not fully reflect the complexity of real-world radiology workflows, where variation in scanner protocols, patient populations, and imaging artifacts can substantially affect model performance.

Underrepresentation of classification and segmentation

In contrast to detection, rib fracture classification and segmentation remain comparatively underdeveloped. Only a small proportion of studies addressed fracture characterization (e.g., displaced vs. non-displaced, acute vs. healing), despite its direct relevance to treatment planning, surgical decision-making, and prognostication. Segmentation studies revealed that 3D visualization and longitudinal monitoring were limited in number and often lacked external validation, although clinically valuable for precise localization.

This imbalance reflects a broader trend in medical AI research, where binary detection tasks dominate due to simpler annotation requirements and more straightforward evaluation. However, clinical utility depends not only on identifying whether a fracture exists but also on understanding its type, extent, and anatomical context. Addressing this gap is therefore essential for the development of clinically meaningful AI systems.

Generalizability, dataset bias, and evaluation heterogeneity

A recurring limitation across the reviewed literature is the reliance on small or geographically restricted datasets. The observed concentration of studies from a limited number of countries raises concerns regarding dataset bias and generalizability. Models trained on homogeneous populations or institution-specific imaging protocols may fail to maintain performance when deployed in different clinical environments.

In addition, substantial heterogeneity in evaluation metrics and reporting practices hinders cross-study comparison and precludes quantitative meta-analysis. Although sensitivity and specificity were frequently reported, standardized benchmarks, external validation cohorts, and clinically meaningful outcome measures were inconsistently applied. These issues underscore the need for harmonized evaluation protocols and shared, multicenter datasets.

Clinical workflow integration and human-artificial intelligence collaboration

Despite promising accuracy metrics, real-world clinical integration remains rare. Few studies described prospective evaluation, seamless integration into PACS, or sustained use by clinicians. Workflow incompatibility, increased cognitive load due to false positives, and lack of trust in opaque model outputs were commonly cited barriers. The reviewed evidence thus supports a human-AI collaboration paradigm, in which AI functions as a second reader rather than an autonomous diagnostic agent. Such collaborative models consistently improved sensitivity while preserving clinical oversight, aligning with regulatory and ethical expectations. However, effective collaboration requires AI outputs that are interpretable, reliable, and presented in a manner that complements radiologist workflows.

Role of explainability and trust

Lack of interpretability emerged as a critical obstacle to clinician adoption. Most models operate as black boxes, providing predictions without transparent justification. Studies that incorporated explainability techniques, such as attention maps or saliency visualizations, were more likely to discuss clinician trust and usability, although systematic evaluation of explainability was uncommon.

Explainable AI is particularly important in rib fracture diagnosis, where false positives can result from overlapping anatomical structures, respiratory motion, or degenerative changes. Transparent visualization of model attention may help clinicians rapidly verify findings and mitigate over-reliance on automated outputs.

Implications for future system development

Collectively, the findings indicate that future research must shift from isolated proof-of-concept studies toward clinically oriented system development. Emphasis should be placed on multi-task learning architectures capable of simultaneous detection, classification, and segmentation, supported by robust external validation. Integration of ensemble learning approaches may further improve robustness and uncertainty estimation, addressing variability in imaging conditions and fracture presentation.

Equally important is the alignment of technical innovation with clinical and regulatory requirements. Prospective validation studies, standardized reporting, and explicit consideration of workflow integration are essential to transition DL systems from research prototypes to routinely deployed clinical tools.

There are several limitations associated with this study, namely (i) not following a fully systematic methodology, (ii) the reliance on a limited number of databases, which may have excluded other relevant literature, (iii) not including protocol registration, and (iv) the diversity among the included studies in terms of methodologies, datasets, and evaluation metrics, which limits comparability. Lack of quantitative synthesis or meta-analysis leads the study findings to be interpreted as descriptive rather than definitive. Despite these limitations, this state-of-the-art review provides a comprehensive and structured overview of the current state of research on DL in rib fracture diagnosis. The study paves the way for improved approaches that can see incorporation of the findings in clinical workflows.

Implications for future research

A key limitation of this review is that fewer than one third of the included studies reported sufficient methodological detail to support formal PROBAST AI assessment. Importantly, exclusion from risk of bias classification reflects incomplete reporting rather

than confirmed methodological weakness. This finding highlights a persistent lack of reporting transparency in AI-based rib fracture research, particularly with respect to external validation and annotation practices, which limits reliable comparison across studies and hampers clinical translation.

The findings of this semi-systematic review underscore several research priorities that must be addressed to advance DL-based rib fracture diagnosis from experimental success toward reliable clinical deployment. Future investigations, as described below, should place greater emphasis on methodological rigor, translational design, and clinical relevance rather than isolated performance gains.

First, model development strategies must evolve toward unified, multi-task frameworks. Most existing studies address detection, classification, and segmentation as independent tasks, which limits clinical applicability and workflow efficiency. Future research should prioritize architectures capable of jointly detecting fractures, characterizing fracture type and age, and delineating precise anatomical boundaries within a single model. Such unified approaches better reflect real clinical decision-making processes and may reduce cumulative error propagation across sequential systems.

Second, there is a pressing need for large-scale, multicenter, and demographically diverse datasets. Current research is dominated by single-institution, geographically localized datasets, which increases the risk of dataset bias and limits external validity. Collaborative data-sharing initiatives and federated learning frameworks offer promising avenues to increase dataset diversity while addressing privacy and regulatory constraints. Future studies should explicitly report dataset composition and conduct robust external validation.

Third, learning paradigms that reduce reliance on exhaustive manual annotation warrant further investigation. Approaches such as self-supervised, semi-supervised, and omni-supervised learning have demonstrated potential in related imaging tasks and may significantly improve scalability in rib fracture research. These methods could mitigate annotation cost while maintaining or improving diagnostic sensitivity, particularly for subtle or rare fracture patterns.

Fourth, explainability and uncertainty quantification should be treated as core

design requirements rather than optional enhancements. Although some studies have explored saliency maps or attention mechanisms, systematic evaluation of explainability methods remains limited. Future research should investigate how interpretable outputs influence clinician confidence, diagnostic accuracy, and error correction in real-world settings and how uncertainty estimates can be incorporated into clinical reporting.

Fifth, human–AI interaction models require dedicated empirical study. Most existing evaluations focus on standalone model performance rather than the dynamics of clinician–AI collaboration. Future work should explore optimal interaction paradigms, including confidence-adaptive assistance, second-reader workflows, and user-centered interface design, using prospective and observational study designs.

Finally, clinical impact studies must extend beyond diagnostic accuracy. Prospective trials assessing workflow efficiency, patient outcomes, downstream clinical decisions, and cost-effectiveness are essential for regulatory approval and adoption. Standardized evaluation frameworks and consensus benchmarks should be adopted to facilitate comparison across studies and support evidence-based translation.

Collectively, these research directions highlight the need for a shift from model-centric experimentation toward clinically grounded, explainable, and generalizable AI systems that address real-world diagnostic challenges in rib fracture care.

This semi-systematic state-of-the-art review demonstrates that DL has achieved substantial technical maturity in rib fracture detection using CT imaging, with consistent improvements in diagnostic sensitivity and reporting efficiency compared with conventional radiological workflows. Detection remains the most developed application, benefiting from robust model architectures and extensive experimental validation. In contrast, rib fracture classification and segmentation remain comparatively underexplored and insufficiently validated despite their critical importance for treatment planning, severity assessment, and longitudinal monitoring.

The review highlights that model performance is highly context-dependent, influenced by dataset quality, population diversity, imaging protocols, and annotation

strategies. The predominance of retrospective, single-center studies and the absence of standardized evaluation protocols limit generalizability and complicate comparison across studies. As a result, high reported performance metrics alone are insufficient to justify routine clinical deployment.

Reviewed studies showed human–AI collaboration to be the most effective deployment paradigm. When used as a second reader, DL systems enhance sensitivity and reduce reading time without replacing clinical judgment. However, widespread adoption is constrained by challenges related to workflow integration, interpretability, regulatory uncertainty, and clinician trust.

Explainability emerges as a central requirement for clinical translation. Models that provide transparent visual or quantitative justification for their predictions are more likely to be trusted, audited, and integrated into real-world practice. Addressing false positives, improving uncertainty estimation, and aligning system outputs with radiologists' decision-making processes are therefore essential.

In conclusion, although DL for rib fracture diagnosis shows clear promise, realizing its clinical impact requires a shift from isolated proof-of-concept studies toward clinically validated, explainable, and workflow-ready solutions. Future efforts should prioritize multi-task learning frameworks that unify detection, classification, and segmentation, leverage multicenter datasets to ensure generalizability, and adopt standardized reporting and evaluation practices. Success should ultimately be measured not only by algorithmic accuracy but also by the ability of these systems to integrate seamlessly into clinical workflows, support clinician decision-making, and meaningfully improve patient outcomes.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Appendix: <https://d2v96fxpocvxx.cloudfront.net/477dd3ac-0088-448a-a453-fd2929e03f96/content-images/dfa37602-efc0-4ee7-bcef-1c0ead532ecc.pdf>

References

1. Zhou QQ, Tang W, Wang J, et al. Automatic detection and classification of rib fractures

based on patients' CT images and clinical information via convolutional neural network. *Eur Radiol.* 2021;31(6):3815-3825. [\[Crossref\]](#)

2. Brasel KJ, Guse CE, Layde P, Weigelt JA. Rib fractures: relationship with pneumonia and mortality. *Crit Care Med.* 2006;34(6):1642-1646. [\[Crossref\]](#)
3. Yang J, Shi R, Jin L, et al. Deep rib fracture instance segmentation and classification from CT on the RibFrac Challenge [Preprint]. *arXiv.* 2024. [\[Crossref\]](#)
4. Ibanez V, Jucker D, Ebert LC, Franckenberg S, Dobay A. Classification of rib fracture types from postmortem computed tomography images using deep learning. *Forensic Sci Med Pathol.* 2024;20(4):1208-1214. [\[Crossref\]](#)
5. Cho SH, Sung YM, Kim MS. Missed rib fractures on evaluation of initial chest CT for trauma patients: pattern analysis and diagnostic value of coronal multiplanar reconstruction images with multidetector row CT. *Br J Radiol.* 2012;85(1018):e845-e850. [\[Crossref\]](#)
6. Lopez-Melia M, Magnin V, Marchand-Maillet S, Grabherr S. Deep learning for acute rib fracture detection in CT data: a systematic review and meta-analysis. *Br J Radiol.* 2024;97(1155):535-543. [\[Crossref\]](#)
7. Ibanez V, Gunz S, Erne S, et al. RiFNet: Automated rib fracture detection in postmortem computed tomography. *Forensic Sci Med Pathol.* 2022;18(1):20-29. [\[Crossref\]](#)
8. Kayo A, Tsuchiya N, Yonemoto K, et al. Detection of costal cartilage fractures on CT images with computer-aided detection system for rib fractures. *In Vivo.* 2025;39(1):390-395. [\[Crossref\]](#)
9. Kuo RYL, Harrison C, Curran TA, et al. Artificial intelligence in fracture detection: a systematic review and meta-analysis. *Radiology.* 2022;304(1):50-62. [\[Crossref\]](#)
10. Zhang J, Li Z, Yan S, Cao H, Liu J, Wei D. An algorithm for automatic rib fracture recognition combined with nnU-Net and DenseNet. *Evid Based Complement Alternat Med.* 2022;2022:5841451. [\[Crossref\]](#)
11. Zhou Z, Fu Z, Jia J, Lv J. Rib fracture detection with dual-attention enhanced U-Net. *Comput Math Methods Med.* 2022;2022:8945423. [\[Crossref\]](#)
12. Jin L, Youjun E, Ye Z, et al. Feasibility of rib fracture detection in low-dose computed tomography images with a large, multicenter datasets-based model. *Heliyon.* 2024;10(10):e31010. doi: [\[Crossref\]](#)
13. Wang X, Wang Y. Composite attention residual U-Net for rib fracture detection. *Entropy (Basel).* 2023;25(3):466. [\[Crossref\]](#)
14. Zhang B, Jia C, Wu R, et al. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation. *Br J Radiol.* 2021;94(1118):20200870. [\[Crossref\]](#)

15. Haitaamar ZN, Abdulaziz N. Detection and semantic segmentation of rib fractures using a convolutional neural network approach. In: 2021 IEEE Region 10 Symposium (TENSYP); 2021. p. 1-4. [\[Crossref\]](#)
16. Ahsan MM, Siddique Z. Industry 4.0 in healthcare: a systematic review. *Int J Inf Manag Data Insights*. 2022;2(1):100079. [\[Crossref\]](#)
17. Tunc MA, Gures E, Shaye I. A survey on IoT smart healthcare: emerging technologies, applications, challenges, and future trends [Preprint]. arXiv. 2021. [\[Crossref\]](#)
18. Sun L, Fan Y, Shi S, et al. AI-assisted radiologists vs. standard double reading for rib fracture detection on CT images: a real-world clinical study. *PLoS One*. 2025;20(1):e0316732. [\[Crossref\]](#)
19. Collins CE, Giammanco PA, Trivedi SM, Sarsour RO, Kricfalusi M, Elsisy JG. Diagnostic accuracy of artificial intelligence for detection of rib fracture on X-ray and computed tomography imaging: a systematic review. *J Imaging Inform Med*. 2025;38(5):2973-2982. [\[Crossref\]](#)
20. Chai S, Gao P, Li M, Jin L. Mask R-CNN for predicting rib fractures on CT images with interpretability and ChatGPT-based structured outcomes. *Expert Syst Appl*. 2025;274:127047. [\[Crossref\]](#)
21. Chai Z, Luo L, Lin H, Heng PA, Chen H. Deep omni-supervised learning for rib fracture detection from chest radiology images. *IEEE Trans Med Imaging*. 2024;43(5):1972-1982. [\[Crossref\]](#)
22. 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. [\[Crossref\]](#)
23. Ye J, Li H, Zhang M, et al. Oblique axis rib stretch and curved planar reformats in patients for rib fracture detection and characterization: feasibility and clinical application. *Emerg Med Int*. 2023;2023:4904844. [\[Crossref\]](#)
24. Tan H, Xu H, Yu N, et al. The value of deep learning-based computer aided diagnostic system in improving diagnostic performance of rib fractures in acute blunt trauma. *BMC Med Imaging*. 2023;23(1):55. [\[Crossref\]](#)
25. Yu H, Ding Y, Sun J, et al. Deep-learning algorithm for rib fracture detection in low-dose chest CT images. In: Proceedings of the 2022 9th International Conference on Biomedical and Bioinformatics Engineering (ICBBE '22). New York (NY): ACM; 2023. p. 55-60. [\[Crossref\]](#)
26. Azuma M, Nakada H, Takei M, et al. Detection of acute rib fractures on CT images with convolutional neural networks: effect of location and type of fracture and reader's experience. *Emerg Radiol*. 2022;29(2):317-328. [\[Crossref\]](#) Erratum in: *Emerg Radiol*. 2022;29(2):425. [\[Crossref\]](#)
27. Blum A, Gillet R, Urbaneja A, Gondim Teixeira P. Automatic detection of rib fractures: are we there yet? *EBioMedicine*. 2021;63:103158. [\[Crossref\]](#)
28. Niiya A, Murakami K, Kobayashi R, et al. Development of an artificial intelligence-assisted computed tomography diagnosis technology for rib fracture and evaluation of its clinical usefulness. *Sci Rep*. 2022;12:8363. [\[Crossref\]](#)
29. Chai Z, Lin H, Luo L, Heng A, Chen H. ORF-Net: deep omni-supervised rib fracture detection from chest CT scans. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Lecture Notes in Computer Science. Vol. 13433. Cham: Springer; 2022. [\[Crossref\]](#)
30. Kaiume M, Suzuki S, Yasaka K, et al. Rib fracture detection in computed tomography images using deep convolutional neural networks. *Medicine (Baltimore)*. 2021;100(20):e26024. [\[Crossref\]](#)
31. Yao L, Guan X, Song X, et al. Rib fracture detection system based on deep learning. *Sci Rep*. 2021;11(1):23513. [\[Crossref\]](#)
32. Alzahrani NM, Jeanes A, Paddock M, Shuweihdi F, Offiah AC. The diagnostic performance of chest computed tomography in the detection of rib fractures in children investigated for suspected physical abuse: a systematic review and meta-analysis. *Eur Radiol*. 2021;31(9):7088-7097. [\[Crossref\]](#)
33. Weikert T, Noordtzi LA, Bremerich J, et al. Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *Korean J Radiol*. 2020;21(7):891-899. [\[Crossref\]](#)
34. Cao Z, Xu L, Chen DZ, Gao H, Wu J. A robust shape-aware rib fracture detection and segmentation framework with contrastive learning. *IEEE Trans Multimedia*. 2023;25:1584-1591. [\[Crossref\]](#)
35. Meng XH, Wu DJ, Wang Z, et al. A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance. *Skeletal Radiol*. 2021;50(9):1821-1828. [\[Crossref\]](#)
36. Wu M, Chai Z, Qian G, et al. Development and evaluation of a deep learning algorithm for rib segmentation and fracture detection from multicenter chest CT images. *Radiol Artif Intell*. 2021;3(5):e210024. [\[Crossref\]](#)
37. Li N, Wu Z, Jiang C, et al. An automatic fresh rib fracture detection and positioning system using deep learning. *Br J Radiol*. 2023;96(1146):20221006. [\[Crossref\]](#)
38. Chan HT, Liu YH, Tarng YW, Lin CH, Wu ML, Chen YY. An image enhancement method for deep learning-based rib fracture detection. In: 2023 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan); 2023. p. 417-418. [\[Crossref\]](#)
39. Burkow J, Holste G, Otjen J, et al. High sensitivity methods for automated rib fracture detection in pediatric radiographs. *Sci Rep*. 2024;14:8372. [\[Crossref\]](#)
40. Jin L, Yang J, Kuang K, et al. Deep-learning-assisted detection and segmentation of rib fractures from CT scans: development and validation of FracNet. *EBioMedicine*. 2020;62:103106. [\[Crossref\]](#)
41. Wu L, Chen H, Li P, Yang K. A novel ensemble approach for rib fracture detection and visualization using CNNs and Grad-CAM. *Ann Ital Chir*. 2025;96(1):86-97. [\[Crossref\]](#)
42. Den Hengst S, Borren N, Van Lieshout EMM, et al. Detection, classification, and segmentation of rib fractures from CT data using deep learning models: a review of literature and pooled analysis. *J Thorac Imaging*. 2025;40(5):e0833. [\[Crossref\]](#)
43. Karlson KA, French A. Initial evaluation and management of rib fractures. UpToDate. Updated Jul 7, 2026. [\[Crossref\]](#)
44. Prins JTH, van Lieshout EMM, Reijnders MRL, Verhofstad MHJ, Wijffels MME. Rib fractures after blunt thoracic trauma in patients with normal versus diminished bone mineral density: a retrospective cohort study. *Osteoporos Int*. 2020;31(2):225-231. [\[Crossref\]](#)