Oral Radiology
https://doi.org/10.1007/s11282-025-00878-3

REVIEW

Diagnostic performance of artificial intelligence for facial fracture detection: a systematic review

Nozimjon Tuygunov[1] · Shukhrat A. Boymuradov[2] · Zohaib Khurshid[3,4] · Siriporn Songsiripradubboon[5] · Jamshid Abdulahtov[6] · Ulugbek Khatamov[1,2]

Received: 2 September 2025 / Accepted: 5 November 2025

Abstract

Objective To evaluate the diagnostic performance of artificial intelligence (AI) models for detecting facial bone fractures on computed tomography (CT), cone-beam CT (CBCT), and plain radiographs.

Methods Original studies applying machine learning or deep learning algorithms for facial fracture detection in humans were included if they reported diagnostic accuracy metrics such as sensitivity, specificity, or area under the curve (AUC). PubMed-MEDLINE, Scopus, and Web of Science databases were searched up to June 3, 2025. Risk of bias was assessed using the QUADAS-2 tool. The review followed PRISMA 2020 guidelines and was registered in PROSPERO (CRD420251085644).

Results A total of 23 studies were included. Object detection models such as YOLOv5 and Faster R-CNN—demonstrated high diagnostic accuracy in localizing facial fractures. Classification models such as ResNet and Swin Transformer achieved AUCs frequently exceeding 0.90. Segmentation and hybrid frameworks further improved anatomical specificity. However, the generalizability of findings was constrained by predominantly retrospective, single-centre study designs, limited sample sizes, inconsistent annotation practices, and the absence of external or prospective validation.

Conclusion AI models show high diagnostic performance for detecting facial fractures across multiple anatomical regions and imaging modalities. Further multicentre prospective studies and the integration of explainable AI are essential for clinical adoption.

Clinical relevance AI-assisted diagnostic models have the potential to enhance facial fracture detection accuracy, especially in emergency and resource-limited settings. Their integration into radiology workflows could reduce interpretation time, support less experienced clinicians, and improve patient outcomes.

Keywords Artificial intelligence · Convolutional neural networks · CT imaging · Deep learning · Diagnostic accuracy · Facial bone fractures

✉ Nozimjon Tuygunov
nozimtuygunov@gmail.com

[1] Department of Restorative Dentistry, Kimyo International University in Tashkent, Tashkent, Uzbekistan

[2] Department of Maxillofacial Surgery and Dentistry, Faculty of Dentistry, Tashkent State Medical University, Tashkent, Uzbekistan

[3] Department of Prosthodontics and Dental Implantology, College of Dentistry, King Faisal University, Al-Ahsa, Saudi Arabia

[4] Center for Artificial Intelligence and Innovation (CAII), Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand

[5] Department of Pediatric Dentistry, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand

[6] Department of Prosthodontics, Faculty of Dentistry, Tashkent State Medical University, Tashkent, Uzbekistan

Published online: 25 November 2025

Introduction

Facial trauma frequently results in fractures involving the frontal, zygomatic, maxillary, mandibular, nasal, and orbital bones [1, 2]. These injuries, which often encompass both soft and hard tissue damage, are collectively categorized as oral and maxillofacial trauma. Due to the anatomical complexity and functional importance of the facial skeleton, fractures in this region demand prompt and accurate diagnosis to prevent complications such as malocclusion, cosmetic deformity, or neurological impairment. In recent years, trauma management has emphasized rapid diagnosis and intervention during the critical "golden hour" post-injury—a period where timely care has the greatest influence on patient outcomes [3–5]. In addition to clinical urgency, identification of injury patterns has become vital for safeguarding vulnerable populations and optimizing treatment pathways.

Conventional diagnostic imaging—including radiographs, computed tomography (CT), and cone-beam CT (CBCT)—remains the gold standard for fracture evaluation. However, interpreting these images is inherently subjective and relies heavily on the expertise of radiologists and surgeons. In high demanding clinical environment, such as emergency departments, the interpretation of facial trauma imaging can be delayed or subject to inter-observer variability, especially among less experienced readers [6, 7]. This variability can result in missed diagnoses or incorrect treatment decisions. For example, subtle fractures may be overlooked in complex anatomical regions like the midface or orbital walls, where overlapping structures and soft tissue swelling obscure fracture lines [8, 9].

The emergence of artificial intelligence (AI), particularly deep learning methods like convolutional neural networks (CNNs), offers a transformative opportunity in diagnostic radiology. These models have demonstrated strong potential for automating fracture detection across several anatomical regions. Recent systematic reviews have shown that AI systems for skeletal imaging—such as for the wrist, hip, spine, and ribs—can achieve pooled sensitivities and specificities ranging from 90% to 94% on CT and radiographic datasets [10, 11]. Architectures including ResNet, DenseNet, EfficientNet, VGG, and the more recent Swin Transformer have demonstrated diagnostic performance comparable to or exceeding that of human experts [7, 12].

Despite these successes, the application of AI to facial fracture detection has lagged. Several unique challenges contribute to this gap. Craniofacial anatomy is highly complex and varies across individuals, with frequent anatomic overlap and image artifacts that complicate fracture identification [13]. Coexisting soft tissue injuries often obscure osseous landmarks, and fractures may be subtle or multilinear—particularly in the orbit, nasal, or zygomatic regions [14, 15]. In addition, variability in imaging acquisition parameters and inconsistent labelling or annotation practices across datasets limit the reproducibility of deep learning models. Clinically, there are also systemic challenges. Emergency settings are often constrained by time, limited access to experienced radiologists, and high diagnostic workloads—factors that increase the likelihood of missed fractures and support/underscore the need for reliable, automated diagnostic tools [16].

Multiple deep learning approaches have been explored for facial fracture detection, each with distinct strengths and limitations. Object detection networks, including YOLOv3, YOLOv4, YOLOv5, YOLOX, and Faster R-CNN, have demonstrated the ability to localize and classify facial fractures on radiographs and CT images. For example, Mao et al. introduced YOLOv5-TRS for mandibular fracture detection and found it to outperform other YOLO versions and Faster R-CNN [17]. Jeong et al. implemented a dual-stage model combining YOLOX with GhostNetv2 for nasal fractures, achieving 97.2% accuracy [18]. Moon et al. optimized YOLOX-S using CT-specific augmentation and IoU loss, reaching 100% sensitivity in patient-level fracture detection [19]. Warin et al. compared Faster R-CNN and RetinaNet for midfacial fracture detection and found the two-stage detector to provide superior sensitivity and F1 scores [20, 21]. Anderson et al. evaluated an FDA-cleared system, FractureDetect, and demonstrated performance comparable to that of experienced clinicians [22].

Classification networks have also been widely applied to facial fracture detection. These networks provide binary outputs—fracture or no fracture—without spatial localization. In a large comparative study, Mortezaei et al. evaluated multiple CNN architectures—including ResNet50V2, MobileNet, Xception, InceptionV3, and Swin Transformer—on lateral radiographs for nasal fracture classification [23]. Most models achieved accuracy above 90%, with Swin Transformer yielding the highest AUC. Seol et al. applied 3D-ResNet34 and ResNet50 to volumetric CT scans and achieved an AUC of 0.945 [24]. Wang et al. used a patch-based ResNet50 classifier following mandibular segmentation [25], while Nishiyama et al. employed AlexNet on panoramic radiographs to detect condylar fractures [26]. Nam et al. demonstrated that EfficientNet-B7 combined with a multilayer perceptron could classify nasal fractures with AUCs above 0.85 in both internal and external validation cohorts [27]. Although classification models demonstrate high sensitivity and specificity, they lack localization capability and are less interpretable, which may limit their standalone use in clinical workflows [28].

Segmentation models have been employed to overcome some of these challenges by delineating anatomical structures or fracture lines within images. U-Net and its variants

were most used. Wang et al. used U-Net for mandibular segmentation prior to classification, enhancing diagnostic focus [25]. Tong et al. employed U-Net to isolate the zygomatic region before applying ResNet-34 for classification, improving performance by narrowing the model's focus to diagnostically relevant zones [29]. Shahnavazi et al. combined U-Net with Faster R-CNN for mandibular fracture detection on panoramic radiographs [30]. Son et al. added luminance adaptation transforms to a YOLOv4 + U-Net system to improve visual interpretability of mandibular fracture detection [31]. Shan et al. applied Attention U-Net for skull fractures and reported better specificity than with YOLOv3 [32].

Hybrid and multi-stage approaches integrate detection, classification, and segmentation to enhance accuracy and interpretability. Jeong et al. used a two-stage YOLOX + GhostNetv2 model, which not only provided accurate localization but also improved reader confidence and reduced error rates in nasal fracture diagnosis [18]. Van Nistelrooij et al. developed JawFracNet—a three-stage model combining mandible segmentation, fracture segmentation, and classification using CBCT—highlighting the potential for anatomically tailored pipelines [33]. Similar combinations were observed in Wang et al., Shahnavazi et al., and Son et al., demonstrating that segmentation-enhanced pipelines can reduce false positives and improve precision [25, 30, 31]. However, such models are computationally intensive and often trained on small, retrospective datasets, limiting scalability and prospective deployment.

Despite technical progress, AI for facial fracture detection has yet to achieve widespread clinical implementation. Furthermore, most published models lack explainability, are trained on small or non-representative samples, and are rarely validated across institutions or imaging protocols [34, 35]. As a result, their integration into clinical workflows remains minimal. While early studies suggest that AI tools could streamline radiologic workflows, reduce inter-reader variability, and improve diagnostic speed [36], their impact on patient outcomes and clinical decision-making remains untested.

Therefore, this systematic review aims to synthesize the current evidence regarding the diagnostic performance of deep learning models—particularly object detection, classification, segmentation, and hybrid networks—in identifying facial bone fractures from medical imaging. By evaluating the strengths, limitations, and clinical readiness of these AI models, we seek to provide insights into future research directions and identify opportunities for integration into clinical care.

Methods

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [37] and was prospectively registered in the PROSPERO database (registration ID: CRD420251085644) to ensure methodological transparency and reproducibility. The review aimed to evaluate the diagnostic performance of AI models in detecting facial bone fractures in human trauma patients using medical imaging modalities such as CT, CBCT, and plain radiographs. The study selection was structured using the PICOS (Population, Intervention, Comparison, Outcome, Study design) framework to ensure consistency in inclusion criteria and research focus. Studies were included if they reported original data on AI-based detection of facial fractures and provided measurable diagnostic performance outcomes. Non-human, in vitro, review, or non-diagnostic articles were excluded. A total of 23 studies met the eligibility criteria and were documented through a PRISMA-compliant flow diagram.

Search strategy

A comprehensive electronic search was conducted using the PubMed-MEDLINE, Scopus, and Web of Science databases to identify relevant studies. The search strategy included combinations of keywords related to artificial intelligence and facial fracture detection, such as "artificial intelligence", "deep learning", "machine learning", "neural network", "facial fracture, "craniofacial fracture", "mandibular fracture", "maxillofacial trauma", "zygomatic fracture", "nasal fracture", "orbital fracture", "computed tomography", "CT", "panoramic radiograph", "X-ray", "radiographic imaging" (Table 1). The search was limited to original research articles published in English involving human subjects. Studies focusing solely on in vitro models, animal data, or not reporting diagnostic performance outcomes were excluded. The last search update was conducted on 3rd of June 2025. Additional relevant studies were identified through manual screening of reference lists from the included articles.

Study selection

The initial electronic search was conducted in PubMed-MEDLINE, Scopus, and Web of Science on 20th of May 2025, and all identified references were imported into EndNote software (version X20; Clarivate Analytics, Philadelphia, USA) to check for and remove duplicate records. Two reviewers (UK and NT) independently screened the titles and abstracts of the retrieved articles for relevance using

Table 1 Literature searches from different databases

Databases	Search Terms
PubMed-MEDLINE	("artificial intelligence" OR "deep learning" OR "machine learning" OR "neural network*") AND ("facial fracture*" OR "craniofacial fracture*" OR "mandibular fracture*" OR "maxillofacial trauma" OR "zygomatic fracture*" OR "nasal fracture*" OR "orbital fracture*") AND ("computed tomography" OR "CT" OR "panoramic radiograph*" OR "X-ray" OR "radiographic imaging")
Scopus	(artificial intelligence OR deep learning OR machine learning OR neural network*) AND (facial fracture* OR craniofacial fracture* OR mandibular fracture* OR maxillofacial trauma OR zygomatic fracture* OR nasal fracture* OR orbital fracture*) AND (computed tomography OR CT OR panoramic radiograph* OR X-ray OR radiographic imaging)
Web of Science	TS=("artificial intelligence" OR "deep learning" OR "machine learning" OR "neural network*") AND TS=("facial fracture*" OR "craniofacial fracture*" OR "mandibular fracture*" OR "maxillofacial trauma" OR "zygomatic fracture*" OR "nasal fracture*" OR "orbital fracture*") AND TS=("computed tomography" OR "CT" OR "panoramic radiograph*" OR "X-ray" OR "radiographic imaging")

Rayyan QCRI (Rayyan, Doha, Qatar). Full-text articles deemed potentially eligible were retrieved and assessed in detail for inclusion according to the predefined eligibility criteria. Any disagreements during the selection process were resolved through discussion and consensus between the two reviewers.

Eligibility criteria

The inclusion criteria for this systematic review were defined using the PICOS framework. The population consisted of human subjects with facial trauma who underwent diagnostic imaging (CT, CBCT, or plain radiographs) to evaluate suspected facial bone fractures. The intervention included the use of AI methods, such as deep learning or machine learning algorithms, applied to medical imaging for the purpose of detecting or diagnosing facial fractures. Studies were eligible regardless of whether they compared AI performance to traditional diagnostic approaches, other AI models, or included no comparator. The primary outcomes of interest were diagnostic performance metrics, including sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC). Secondary outcomes such as processing time, clinical utility, and model integration into radiologic workflow were also noted when available. Eligible study designs included original diagnostic accuracy studies, both retrospective and prospective, that reported quantitative results. Only full-text articles published in English were included. Exclusion criteria were studies not involving AI applications for facial fracture detection, studies lacking diagnostic performance outcomes, case reports, review articles, editorials, conference abstracts, expert opinions, non-human or in vitro studies, and publications without accessible full text.

Data extraction

Two reviewers (UK and NT) independently performed data extraction from the full-text versions of all eligible articles using a standardized approach to ensure accuracy and consistency. The following data were extracted from each study: first author and publication year, study design, study objective, dataset size, type of imaging modality and sample characteristics, type of AI model employed, anatomical focus of fracture detection, reference standard used, diagnostic performance metrics, any comparator, and key findings (Table 2). Any discrepancies or uncertainties encountered during the extraction process were resolved through collaborative discussion between the two reviewers to ensure data reliability and methodological rigor.

Results

Study characteristics

A total of 2,042 records were identified through database searches. After removing 4 duplicates, 2,038 records were screened, and 1,679 were excluded based on titles and abstracts. A total of 90 full-text articles were assessed for eligibility, and 23 studies met the predefined inclusion criteria and were included in the final qualitative synthesis, as shown in the PRISMA flow diagram (Fig. 1).

Qualitative study and outcome measures

The included studies, published between 2019 and 2025, investigated the diagnostic performance of various artificial intelligence models for detecting facial bone fractures on imaging modalities such as CT, CBCT, and panoramic radiographs [15, 18–22, 24–27, 29–33, 38–45]. Most studies focused on specific anatomical regions, including the mandible, nasal bones, and zygomatic complex, while some addressed generalized skull or midfacial trauma. Diagnostic performance varied across model types and study designs, but most studies reported sensitivities and specificities

Table 2 Summary of studies included in this systematic review

Study ID	Objective	Number of patient dataset	Type of sample/ data	AI Model	Site	Ref. Standard	Performance	Comparator	Key findings
Mortezaei et al. [38]	To compare deep learning models for nasal fracture detection on lateral X-rays.	1484 patients (2968 lateral nasal bone radiographs; 737 fracture, 747 normal cases)	Lateral nasal bone radiographs from trauma patients, collected from a single radiology center.	VGG16, VGG19, MobileNet, Xception, ResNet50V2, InceptionV3, Swin Transformer	Nasal bone	Combined clinical examination findings and CT scan diagnosis	VGG16: accuracy 79%, AUC 0.86 (best performer). Other models: accuracy ranged 71–79%, AUC ranged 0.79–0.84.	Comparison between multiple AI architectures; no human reader comparator.	VGG16 had highest AUC (0
Warin et al. [20]	To evaluate CNN classifiers and detectors for mandibular fractures on panoramic images.	1710 images (855 fracture, 855 normal) from trauma cases between 2016–2020	Annotated panoramic radiographs; split into training/validation/ test sets, with a 100-image subset compared against clinicians	DenseNet-169, ResNet-50 (classification); Faster R-CNN, YOLOv5 (object detection)	Mandible	Expert radiologist and oral/ maxillofacial surgeon annotations, serving as reference	Classification models: AUC = 1.00 (100% sensitivity & specificity); Detection models: Faster R-CNN — AUC ~ 0.91, precision 87.9%, recall 93.6%; YOLOv5 — AUC ~ 0.90, precision 86.1%, recall 92.2%	Residents and experienced surgeons on a 100-image subset; AI outperformed clinician observers	Classification models reached perfect AUC; object detectors (~0
Mao et al. [39]	To assess CNNs for detecting and classifying mandibular fractures on MSCT.	361 patients with mandibular fractures; ~31,364 annotated MSCT images used for detection, ~10,800 images used for classification	Multislice spiral CT images (coronal, sagittal, axial views), annotated by experienced surgeons and radiologists	Object detection: YOLOv5-TRS (enhanced), YOLOv3, YOLOv4, Faster R-CNN, CenterNet; Classification: Modified DenseNet-121, original DenseNet-121, AlexNet, GoogLeNet, ResNet50	Mandible	Surgical confirmation and expert image annotation	Detection: YOLOv5-TRS mean accuracy 96.7%, best for body fractures (99.0%). Classification: Modified DenseNet-121 AUC 0.814, body fracture classification AUC up to 0.903	Internal comparison among different CNN architectures; no human performance comparison	YOLOv5-TRS and DenseNet-121 achieved high detection/classification accuracy on MSCT
Morita et al. [15]	To develop a deep learning system for detecting midfacial fractures on CT.	100 patients; 4620 axial CT slices (3736 for training, 883 for validation)	Axial facial bone CT scans (midfacial fractures including maxillary, zygomatic, nasal, and orbital fractures) from a single institution	Single Shot MultiBox Detector (SSD), YOLOv8 (object detection algorithms)	Midface	Expert manual annotation by a facial trauma surgeon	SSD: Precision 0.872, Recall 0.823, F1-score 0.846, AP 0.899; YOLOv8: Precision 0.871, Recall 0.775, F1-score 0.82, AP 0.769	No human reader comparison performed; internal comparison between SSD and YOLOv8 models	SSD slightly outperformed YOLOv8 on midfacial CT

Table 2 (continued)

Study ID	Objective	Number of patient dataset	Type of sample/ data	AI Model	Site	Ref. Standard	Performance	Comparator	Key findings
Anderson et al. [22]	To test if an FDA-cleared AI system improves fracture detection by clinicians.	175 unique radiographic cases (across 12 anatomic regions) interpreted in 4200 clinician-case pairs	Musculoskeletal radiographs covering ankle, clavicle, elbow, femur, forearm, hip, humerus, knee, pelvis, shoulder, tibia/fibula, and wrist, from 12 US hospitals	FractureDetect (Imagen Technologies) — FDA-cleared deep learning model for fracture detection	Multiple sites; various anatomical regions	Majority interpretation of 3 board-certified orthopedic surgeons or radiologists	Aided condition: AUC 0.94, sensitivity 90%, specificity 92%; Unaided condition: AUC 0.90, sensitivity 82%, specificity 89%	Compared performance of clinicians with and without AI assistance; included radiologists, orthopedic surgeons, primary care, EM physicians, and PAs	AI improved clinician accuracy across specialties, cutting miss rates from 20% to 9%
van Nistelrooij et al. [33]	To evaluate a 3-stage deep learning model (JawFracNet) for mandibular fractures on CBCT.	335 CBCT scans (171 fracture cases, 164 non-fracture cases)	Cone-beam CT (CBCT) scans of the mandible, from a single center in Germany; included various fracture types and locations	JawFracNet: a 3-stage neural network (mandible segmentation, fracture segmentation, fracture classification)	Mandible	Expert annotations and consensus of 3 oral and maxillofacial surgeons	Precision 0.978, sensitivity 0.956, AUC 0.956 for fracture detection; analysis time ~ 39 s per scan	Compared against 3 oral and maxillofacial surgeons on the same dataset	JawFracNet matched or exceeded surgeon accuracy on CBCT and was faster
Wang et al. [25]	To detect and classify mandibular fractures on CT using CNNs.	686 patients with 1506 annotated mandibular fractures	Spiral CT scans of the mandible, annotated for fractures in nine subregions	U-Net (for mandibular segmentation), ResNet-50 (for patch-based fracture classification)	Mandible	Consensus of three experienced maxillofacial surgeons	Segmentation Dice coefficient: 0.943. Fracture classification: average accuracy > 90%, mean AUC 0.956 across subregions; highest AUCs for body fractures (0.972–0.976)	No comparison with human reader performance; internal evaluation of segmentation and classification	High AUC (0
Nishiyama et al. [26]	To assess deep learning performance in detecting condylar fractures on panoramic X-rays.	400 condyles from two hospitals (200 fractured, 200 normal)	Panoramic radiographs cropped to condylar regions from two independent hospital datasets in Japan	AlexNet CNN (classification model)	Mandibular condyle	CT confirmation of fractures	Internal validation: AUC > 0.85; Hospital A: accuracy 80.4%, sensitivity 80%, specificity 79%; Hospital B: accuracy 81%, sensitivity 80%, specificity 82%; External validation showed lower AUC (~ 0.58)	Internal vs. external dataset performance comparison; no human comparator	Strong internal performance (AUC ~ 0
Tong et al. [29]	To develop a CNN model for detecting zygomatic fractures on CT scans.	379 patients; 474 CT cases (220 fractures, 254 non-fractures)	Craniofacial spiral CT scans from a single institution (Peking University School of Stomatology)	U-Net (for zygomatic region segmentation), ResNet-34 (for fracture detection)	Zygomatic bone	Manual annotations by three experienced maxillofacial surgeons	Region segmentation Dice coefficient: 0.9337 (coronal), 0.9269 (sagittal). Fracture detection: sensitivity and specificity both 100% on test set	Compared against expert manual diagnosis; no human reader variability reported	U-Net + ResNet-34 achieved 100% sensitivity/specificity for zygomatic fractures

Table 2 (continued)

Study ID	Objective	Number of patient dataset	Type of sample/ data	AI Model	Site	Ref. Standard	Performance	Comparator	Key findings
Son et al. [31]	To combine object detection and segmentation for mandibular fracture detection on panoramic.	420 panoramic radiographs (360 training, 60 test images)	Panoramic radiographs of mandibular fractures, categorized by anatomical region	LAT YOLOv4 (location-based detection) and U-Net (fracture line segmentation); compared with Mask R-CNN, standard YOLOv4	Mandible	Manual annotations by oral and maxillofacial radiologists	Combined model (U-Net + LAT YOLOv4): precision 91.7%, recall 87.0%, F1-score > 90%. LAT YOLOv4 alone: precision 97.5%, recall 79.4%, F1-score 87.5%	Internal comparison between Mask R-CNN, YOLOv4, and combined model; no external human reader comparison	U-Net + LAT YOLOv4 improved recall and F1-score (> 90%) for mandible fractures
Shahnavazi et al. [30]	To assess a two-stage AI model for detecting mandibular fractures vs. general dentists.	190 panoramic radiographs (with and without mandibular fractures)	Panoramic radiographs from multiple sources (hospital, private clinic, open-access datasets)	U-Net (mandible segmentation), Faster R-CNN (fracture detection)	Mandible	Consensus annotation by two oral and maxillofacial radiologists	Object detection: mAP50 = 98.66%, mAP75 = 57.90%. Classification accuracy = 91.67%, sensitivity = 100%, specificity = 83.33%	Compared to 5 general dentists; AI outperformed average clinician accuracy (91.67% vs. 87.2%) and sensitivity (100% vs. 82.2%)	AI outperformed general dentists in accuracy (91
Son et al. [40]	To test YOLOv4 for mandibular fracture detection with preprocessing effects.	420 panoramic radiographs (360 for training, 60 for testing)	Panoramic radiographs with mandibular fractures, preprocessed with gamma modulation and luminance adaptation transforms	YOLOv4 with preprocessed input using single- and multi-scale luminance adaptation transform (SLAT, MLAT)	Mandible	Manual annotations by oral and maxillofacial radiologists	Six-class model (anatomic region classification): precision 57%, recall 71%, F1-score 63%. Anatomical region-based models performed better than fracture shape-based models.	No human observer comparison; internal comparison of different data preprocessing and classification strategies	Preprocessing improved fracture visibility
Vinayahalingam et al. [41]	To use vision transformers for detecting and classifying mandibular fractures on panoramic.	6404 panoramic radiographs (1624 with fractures, 4780 without)	Panoramic radiographs (PR) from a single institution, annotated for fracture presence and type (angle, condyle, coronoid, paramedian, median, ramus)	Faster R-CNN with Swin Transformer backbone	Mandible	Manual annotations by three oral and maxillofacial surgeons, confirmed by clinical and CT findings	Binary fracture detection: AUC 0.977, F1-score 0.947, precision 0.935, recall 0.960; multi-class F1 scores ranged from 0.60 (coronoid) to 0.87 (condyle)	No comparison with human reader performance; internal performance comparison across anatomical subregions	Faster R-CNN with Swin Transformer showed high overall accuracy; lower for rare fracture types

Table 2 (continued)

Study ID	Objective	Number of patient dataset	Type of sample/ data	AI Model	Site	Ref. Standard	Performance	Comparator	Key findings
Jeong et al. [18]	To develop and test a YOLOX + GhostNetv2 model for nasal fracture detection.	43 fracture patients + 39 controls; total of 130 fracture views and 117 control images	3D reconstructed CT scans (orthostatic, left and right lateral views)	YOLOX for localization followed by GhostNetv2 for classification	Nasal bone	Dual radiologist consensus reading	AI alone: Accuracy 97.2%, Sensitivity 95.0%, Specificity 95.2%;Junior+AI: Acc 96.2%, Sens 85.6%, Spec 95.1%;Senior+AI: Acc 99.2%, Sens 96.3%, Spec 99.1%;AUCs: Junior alone 0.787 → +AI 0.889; Senior alone 0.928 → +AI 0.979	Junior and senior radiologists reading with and without AI	YOLOX+GhostNetv2 reached 97
Seol et al. [24]	To build a 3D CNN for nasal fracture detection from CT data.	2535 patients (1350 normal, 1185 fractures), evaluated within 2 days of trauma onset	3D reconstructed facial CT images (isotropic voxel input)	3D-ResNet34 and 3D-ResNet50 (single network approach)	Nasal bone	Radiologist-confirmed fracture status based on clinical CT interpretation	3D-ResNet50: AUC 94.5%, Sensitivity 87.5%, Specificity 87.8%, Accuracy 87.6%;3D-ResNet34: AUC 93.4%, Sensitivity 86.4%, Specificity 86.8%, Accuracy 86.2%	None (no human comparator)	3D-ResNet50 had AUC 0
Yang et al. [42]	To assess a deep learning model's performance in detecting nasal fractures and aiding readers.	252 patients evaluated using CT scans; 20 readers participated in performance assessment	Facial CT scans (axial and coronal views)	Custom CNN-based fracture detection algorithm	Nasal bone	Double-blinded consensus diagnosis by two expert radiologists	AI alone: Sensitivity 84.8%, Specificity 86.7%, AUC 0.857; Readers alone: Sensitivity 83.5 ± 10.2%, Specificity 77.6 ± 11.4%, AUC 0.81 ± 0.10; Readers with AI: Sensitivity 94.0 ± 3.2%, Specificity 89.8 ± 6.2%, AUC 0.92 ± 0.04 (all improvements statistically significant, $P < 0.001$)	Human readers without AI assistance	AI enhanced reader sensitivity/specificity, especially for less experienced users
Moon et al. [19]	To develop an object detection model (YoloX-S) for facial bone fracture localization on CT.	690 patients for training (65,205 CT images); 50 for validation; 40 for testing (nasal and other facial bone fractures); ≈5,000 fracture bounding boxes in total	Axial facial bone CT images	YoloX-S object detection model, optimized with IoU loss and CT-specific augmentation	Facial bones (nasal and others)	Plastic surgeon-confirmed bounding box annotations on CT images	Average Precision (AP): 69.8%; Sensitivity/person: 100%; Significantly outperformed baseline YoloX-S by 10.2% in AP and 66.7% in sensitivity/person	Baseline YoloX-S without CT-specific tuning	YOLOX-S reached 100% patient-level sensitivity with fast inference (0
Nam et al. [27]	To validate a CNN for detecting nasal fractures on radiographs vs. radiologists.	6713 patients, split into training ($n = 4325$), validation ($n = 481$), internal test ($n = 1250$), and external test ($n = 102$) cohorts	Bilateral nasal bone radiographs	EfficientNet-B7-based CNN + multilayer perceptron	Nasal bone	Surgical findings and radiological consensus	Internal test: AUC 0.931, Sensitivity 82.2%, Specificity 89.6%, Accuracy 85.9%; External test: AUC 0.857, Sensitivity 83.1%, Specificity 83.7%, Accuracy 83.3%	Two radiologists (9 and 6 years experience)	CNN matched and exceeded radiologist accuracy

Table 2 (continued)

Study ID	Objective	Number of patient dataset	Type of sample/ data	AI Model	Site	Ref. Standard	Performance	Comparator	Key findings
Yari et al. [43]	To evaluate YOLOv5 for detecting six mandibular fracture types on panoramic.	498 panoramic images (673 annotated fractures)	Panoramic radiographs (JPG format)	YOLOv5 object detection model with bounding box annotation	Mandible (symphysis, body, angle, ramus, condylar neck, condylar head)	Expert annotations by oral and maxillofacial surgeons	Accuracy range: 91.5–96.2%; Sensitivity range: 80.0–96.7%; Specificity range: 89.8–98.9%; Dice coefficient range: 0.706–0.921; AUC range: 0.812–0.942 (best results for body and symphysis fractures)	None	YOLOv5 detected six mandibular fracture types with 91–96% accuracy
Warin et al. [21]	To compare Faster R-CNN and RetinaNet for midfacial fracture detection on CT.	2000 CT images (1000 fractures, 1000 normal cases), split into training, validation, and test sets	CT scans of midfacial bones	Faster R-CNN (two-stage detector) and RetinaNet (one-stage detector)	Midface	Clinical diagnosis and radiological findings	Faster R-CNN: Precision 0.72, Recall 0.85, F1-score 0.78, AP 0.79, AUC 0.80; RetinaNet: Precision 0.94, Recall 0.48, F1-score 0.64, AP 0.47, AUC 0.46	RetinaNet (same dataset)	Faster R-CNN outperformed RetinaNet in precision/recall
Shan et al. [32]	To compare YOLOv3 and Attention U-Net for skull fracture detection on CT.	4,782 skull fracture patients + 7,856 healthy controls; External test: 235 patients (93 fractures, 142 controls)	Cranial CT bone algorithm images	YOLOv3 (object detection) and modified Attention U-Net (segmentation)	Skull	Seven neurologists' manual annotations and clinical diagnosis	YOLOv3: Sensitivity 80.64%, Specificity 85.92%, Accuracy 83.83%; Attention U-Net: Sensitivity 82.8%, Specificity 88.73%, Accuracy 88.26%	YOLOv3 vs. Attention U-Net (two AI approaches)	Attention U-Net surpassed YOLOv3 in accuracy/specificity for skull fractures
Chilamkurthy et al. [44]	To validate deep learning models for detecting calvarial fractures on head CT.	313,318 head CT scans collected from ~20 centers in India; validation on Qure25k (21,095 scans) and CQ500 (491 scans)	Non-contrast head CT scans from trauma and stroke patients	Custom deep learning algorithms using ConvNets (model architecture not specified in detail)	Calvarial bone (skull fractures)	Original clinical radiology reports and consensus of 3 radiologists	AUC for calvarial fractures: 0.92 (Qure25k) and 0.96 (CQ500); high sensitivity and specificity reported	Comparison between AI algorithms and human raters (radiologists)	AI showed AUCs of 0
Wang et al. [45]	To develop DeepCT for detecting cranial and facial fractures and localizing regions.	1,447 head CT scans for training (16,985 images); tested on 192 head CT studies (5,890 images)	Non-enhanced head CT scans from trauma patients	YOLOv4 (fracture detection) + ResU-Net++ (bone segmentation)	Cranial and facial bones	Radiology reports confirmed by neurosurgeons and neuroradiologist	Sensitivity 88.7%, Specificity 94.7%, Precision 94.5%, F1 score 0.91 (case level); facial region F1 score 0.84, cranial region F1 score 0.89	Compared to radiologist consensus	AI detected cranial/facial fractures with high accuracy from soft-tissue CT

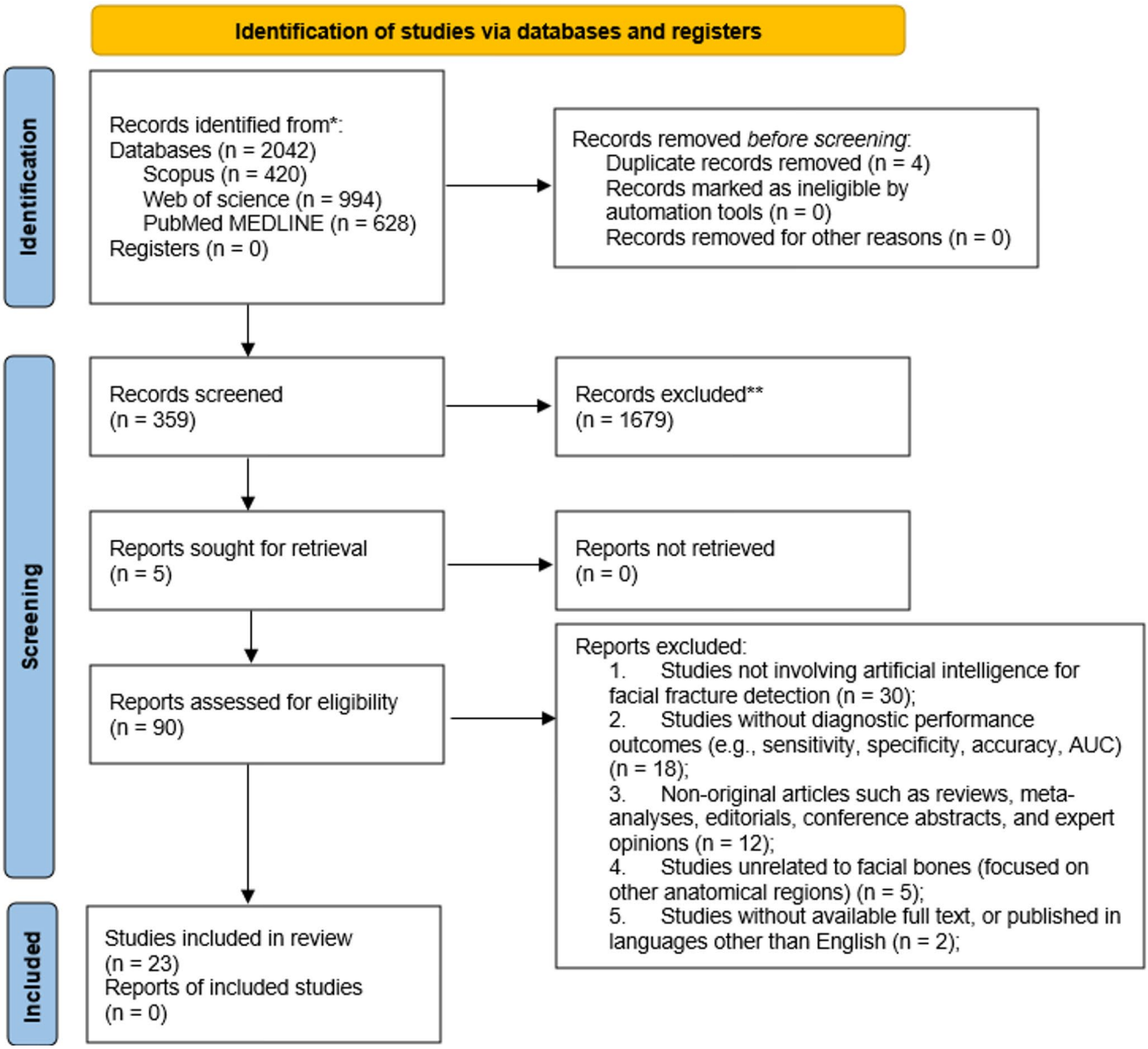

Fig. 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of this selected study

above 85%, with area under the curve (AUC) values frequently exceeding 0.90. However, substantial heterogeneity was observed in terms of dataset size, annotation protocols, imaging modalities, and reported performance metrics, precluding the possibility of conducting a formal meta-analysis. Summary of studies included in this systematic review can be seen in Table 2.

Frontal bone fractures

Two of the included studies specifically addressed fracture detection in the frontal bone region using deep learning approaches. Chilamkurthy et al. developed a multi-finding deep learning algorithm applied to non-contrast head CT scans, which demonstrated strong diagnostic performance for calvarial fractures, including the frontal bone [44]. The model achieved an AUC of 0.92 in the Qure25k dataset and 0.96 in the CQ500 dataset, highlighting the potential of AI in detecting cranial vault fractures in trauma cases. Similarly, Wang et al. proposed a combined approach using YOLOv4 for fracture detection and ResUNet + + for skull segmentation on head CT images [45].

Zygomatic bone fractures

Among the included studies, only one reported AI performance specific to zygomatic bone fracture detection. Tong et al. developed a deep learning pipeline that combined

U-Net for anatomical segmentation of the zygomatic region with a ResNet-34 classifier for detecting fractures on CT images [29]. The model achieved an overall classification accuracy of 88.1%, with a sensitivity of 88.5% and specificity of 87.5%, indicating balanced diagnostic performance. The segmentation step was designed to isolate the zygomatic area to improve the relevance of feature extraction for the classification model. This study demonstrated that combining segmentation and classification can enhance fracture detection in localized, anatomically complex facial regions such as the zygoma. Warin et al., for instance, applied object detection models (Faster R-CNN and RetinaNet) to midfacial CT data, but did not report performance by individual bone structures [20, 21].

Nasal bone fractures

Several studies in this review specifically targeted nasal bone fracture detection using deep learning models, primarily on CT imaging. Jeong et al. implemented a two-stage pipeline that combined YOLOX for fracture localization and GhostNetv2 for classification, reporting a diagnostic accuracy of 97.2% [18]. Notably, the system improved diagnostic performance among less experienced readers, demonstrating its potential as a decision-support tool. Seol et al. utilized 3D-ResNet34 and 3D-ResNet50 models for direct nasal fracture classification on 3D CT images. Their best-performing model achieved an area under the curve (AUC) of 0.945, with a sensitivity of 85.3% and specificity of 90.3%, highlighting the feasibility of using volumetric classification networks in fracture detection. Nam et al. also applied an EfficientNet-B7 combined with a multilayer perceptron for nasal fracture classification on plain radiographs [27]. Their model yielded robust performance in both internal and external test sets, achieving an AUC of 0.88 and an overall accuracy of 87.5%. Additionally, Yang et al. developed a custom CNN architecture trained on CT images that achieved competitive sensitivity and specificity in detecting nasal fractures [46].

Mandibular fractures

Mandibular fracture detection was the most extensively studied application of AI across the included articles, reflecting the clinical importance and diagnostic challenge of identifying mandibular injuries on panoramic radiographs and CT scans. Mao et al. conducted a comprehensive comparison of object detection models, including YOLOv3, YOLOv4, YOLOv5-TRS, CenterNet, and Faster R-CNN, for mandibular fracture localization on panoramic images [39]. YOLOv5-TRS achieved the highest precision (89.7%) and sensitivity (92.3%), outperforming other models in both fracture detection and region-based accuracy. Shahnavazi and Mohamadrahimi (2023) applied a two-step hybrid model using U-Net for mandible segmentation and Faster R-CNN for fracture detection, achieving high diagnostic accuracy while minimizing false positives through region-focused analysis [30]. Van Nistelrooij et al. introduced JawFracNet, a multi-stage neural network involving mandible segmentation, fracture segmentation, and classification on CBCT scans [33]. The model demonstrated reliable identification of fracture subtypes across mandibular regions with improved interpretability. Son et al. used YOLOv4-based models on panoramic images, incorporating luminance adaptation transforms (LAT, MLAT) to enhance contrast, and paired it with U-Net for fracture line segmentation [31, 40]. These adaptations improved both visual interpretability and detection metrics. Wang et al. implemented patch-based ResNet-50 classification after mandibular segmentation with U-Net, allowing high-accuracy classification by focusing only on relevant bone structures [25]. Yari et al. applied YOLOv5 to annotate bounding boxes for mandibular fractures on panoramic radiographs, achieving balanced sensitivity and specificity [43].

Fractures in proximity

Some included studies addressed the detection of fractures occurring in close anatomical proximity, such as multisite midfacial injuries or fractures spanning adjacent bone regions, though few provided detailed region-by-region analysis. Warin et al. used object detection models (Faster R-CNN and RetinaNet) on midfacial CT scans and reported balanced detection performance for multiple facial regions, but without specific breakdowns for overlapping anatomical zones [20, 21]. Shan et al. evaluated a hybrid model (YOLOv3 + Attention U-Net) for skull fracture detection on head CTs, achieving high performance (AUC > 0.97), yet results were reported as overall skull fractures, without analysis of border zones such as the orbital rim or frontozygomatic suture [32]. Son et al. applied YOLOv4 with luminance adaptation and U-Net segmentation to detect mandibular fractures near the angle and ramus, which may border parotid and submandibular regions, though no distinction was made for overlapping anatomical involvement [40].

Risk of bias and quality assessment

The risk of bias and applicability concerns for the included studies were assessed using the QUADAS-2 tool [47]. Among the 23 included studies, overall methodological quality was acceptable. All studies were judged to have low risk of bias and low applicability concerns in the patient

selection domain. The index test domain showed the highest number of "unclear" ratings ($n = 5$; Mortezaei et al. [38], Morita et al. [15], Nishiyama et al. [26], Shahnavazi et al. [30], and Son et al. [40]), largely due to insufficient reporting of whether AI interpretations were conducted blinded to the reference standard. The reference standard domain had two studies (Mortezaei et al. [38], Nishiyama et al. [26]) with unclear risk, stemming from limited details about how ground truth was defined. In summary, most studies demonstrated low risk of bias and applicability concerns. The risk of bias and the factors considered for the analysis are presented in Table 3.

Discussion

Object detection models

In this review, object detection networks emerged as the predominant AI tools for facial fracture detection, most notably the YOLO series (v3–v8) [48] and Faster R-CNN, which provided precise localization of fractures across diverse imaging modalities [49]. These models are typically deployed within a structured diagnostic pipeline, as illustrated in Fig. 2. For example, Mao et al. showed that an enhanced YOLOv5-TRS model outperformed YOLOv3/v4, Faster R-CNN, and CenterNet in mandibular fractures on panoramic radiographs [39]. Warin et al. found Faster R-CNN delivered higher sensitivity and F1-scores than RetinaNet in midfacial CT [20, 21], while Moon et al. leveraged CT-specific augmentations with YOLOX-S to achieve 100% patient-level sensitivity [19].

Beyond craniofacial trauma, similar methods have demonstrated clinical impact in other anatomical regions [50]. Pranata et al. combined ResNet and VGG CNNs with Speeded-Up Robust Features (SURF) to detect calcaneal fractures in CT, achieving classification accuracy of 98% and successful localization of fracture sites [51]. In thoracic imaging, Meng et al. implemented a cascaded feature pyramid network for rib fracture detection on chest CT; this model improved junior radiologists' recall from 0.812 to 0.922 ($p < 0.001$) and reduced interpretation time from 158 to 42 s [52]. In pelvic trauma, a YOLOv4-based system with DenseNet backbone matched clinician-level performance in detecting hip and pelvic fractures on radiographs [53]. Collectively, these studies validate the generalizability and clinical relevance of object detection models across complex fracture scenarios [54–57]. Despite these promising results, limitations are consistent across studies: retrospective designs, heterogeneity in dataset scaling and annotation

Table 3 Risk of bias assessment of studies included

Study ID	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Mortezaei et al. [38]	⊕	⊗	⊗	⊕	⊕	⊖	⊕
Warin et al. [20]	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Mao et al. [39]	⊖	⊕	⊗	⊗	⊕	⊕	⊗
Morita et al. [15]	⊕	⊗	⊗	⊕	⊕	⊕	⊕
Anderson et al. [22]	⊕	⊕	⊕	⊕	⊕	⊕	⊕
van Nistelrooij et al. [33]	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Wang et al. [25]	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Nishiyama et al. [26]	⊕	⊗	⊕	⊕	⊕	⊕	⊕
Tong et al. [29]	⊖	⊕	⊗	⊗	⊕	⊕	⊗
Son et al. [31]	⊕	⊗	⊕	⊕	⊕	⊕	⊕
Shahnavazi et al. [30]	⊖	⊕	⊗	⊗	⊗	⊕	⊕
Son et al. [40]	⊗	⊕	⊗	⊗	⊕	⊕	⊕
Vinayahalingam et al. [41]	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Jeong et al. [18]	⊖	⊕	⊗	⊗	⊕	⊕	⊕
Seol et al. [24]	⊗	⊕	⊗	⊗	⊕	⊕	⊕
Yang et al. [42]	⊖	⊕	⊗	⊗	⊗	⊕	⊗
Moon et al. [19]	⊕	⊗	⊕	⊕	⊕	⊕	⊕
Nam et al. [27]	⊗	⊕	⊕	⊕	⊕	⊕	⊕
Yari et al. [43]	⊕	⊕	⊗	⊕	⊕	⊕	⊕
Warin et al. [21]	⊖	⊕	⊗	⊗	⊕	⊕	⊗
Shan et al. [32]	⊖	⊕	⊕	⊗	⊕	⊕	⊗
Chilamkurthy et al. [44]	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Wang et al. [45]	⊕	⊗	⊗	⊕	⊕	⊕	⊕
		⊕ Low		⊗ Unclear		⊖ High	

methods, insufficient external validation, and limited integration into clinical workflows [58].

Classification networks

On the other hand, classification networks excel at distinguishing fractured from non-fractured facial bones, achieving high diagnostic performance in curated datasets, despite lacking localization capabilities. Mortezaei et al. evaluated seven architectures—including VGG16/19, MobileNet, Xception, ResNet50V2, InceptionV3, and Swin Transformer—on lateral X-rays for nasal fracture detection [23]. Swin Transformer achieved the highest AUC (~ 0.93), with most models recording diagnostic accuracies above 90%. Seol et al. applied 3D-ResNet34/50 to CT volumes, achieving an AUC of 0.945, demonstrating that volumetric classification is effective even without explicit segmentation [24]. In CT-based mandibular fracture analysis, Wang et al. reported high sensitivity using a patch-based ResNet-50 method [25]; Nishiyama et al. achieved moderate accuracy with AlexNet for condylar fracture detection on panoramic images [26]. Additionally, Yang et al. implemented a custom CNN yielding improved sensitivity and specificity for nasal fractures [42], while Nam et al. reported AUCs exceeding 0.85 using EfficientNet-B7 with a multilayer perceptron on plain radiographs [27]. Though classification networks offer rapid, accurate screening capabilities, their inability to pinpoint fracture location limits their utility in clinical decision-making [59–61].

Similar roles of classification models in screening for fractures have been demonstrated in other anatomical contexts. In distal radius fracture detection, a two-stage pipeline consisting of a YOLOv5-based region proposal network followed by EfficientNet-B3 classification achieved an AUC of 0.82, accuracy of 81%, sensitivity of 83%, and specificity of 73% for differentiating intra- from extra-articular fractures [62, 63]. A wrist radiograph classification model using various CNN architectures (DenseNet121, ResNet50, VGG-19, InceptionV3) demonstrated high AUCs between 0.936 and 0.941 for fracture identification and AUCs of 0.96 for subtype classification [64]. Another study applying Explainable AI (Grad-CAM) achieved AUCs of 0.967–0.975, sensitivities above 93%, and specificities above 94% on wrist X-rays [65]. In pediatric wrist trauma, a fine-grained CNN using limited data achieved 97% sensitivity and nearly 86% accuracy, aided by Grad-CAM heatmaps [66].

These external studies echo our findings—classification models can approach specialist-level accuracy in fracture screening, given sufficient dataset quality. However, common limitations include reliance on well-annotated, homogeneous cohorts and lack of localization capacity, which hinders clinical decision support. Furthermore, most models have not been tested in prospective real-world settings, and few provide interpretability mechanisms critical for clinician acceptance [67].

Segmentation models

Segmentation models have become integral in fracture detection by precisely delineating anatomical structures and fracture zones—an essential step in clinical decision-making and treatment planning [68–72]. Across our review, U-Net and its variants predominated. For instance, Wang et al. used U-Net to segment the mandible on CT images prior to ResNet-50-based classification, significantly boosting diagnostic accuracy by focusing analysis on relevant structures [25]. Tong et al. mirrored this strategy in the zygomatic region, achieving an overall classification accuracy of 88.1% (sensitivity 88.5%, specificity 87.5%) on CT scans [29]. Shahnavazi et al. improved detection precision on panoramic radiographs by combining U-Net mandible segmentation with Faster R-CNN [30]. Son et al. took a hybrid approach, using luminance-adapted YOLOv4 for detection followed by U-Net to segment fracture lines, resulting in clearer and more targeted localization [31]. Most notably, Shan et al. employed an Attention U-Net tailored for skull fracture segmentation on CT, outperforming a YOLOv3 object detector—with sensitivity and specificity of 82.8% and 88.7%, respectively—highlighting the advantages of detailed segmentation in anatomically complex areas [32].

These findings mirror broader trends in medical imaging. In pelvic X-rays, Attention U-Net and Swin U-Net significantly outperformed standard U-Net, with Swin U-Net achieving a sensitivity of 96.8%, specificity of 98.5%, accuracy of 98.0%, and Dice coefficient of 96.3% across 940 cases [73]. Segmentation-based approaches have also demonstrated high performance in other clinical imaging domains. In cervical spine X-rays, an Attention U-Net model achieved a sensitivity of 90.44%, specificity of 99.51%, Dice score of 89.41%, and overall accuracy of 99.13% for bone structure segmentation across 193 test subjects [74]. In large-scale studies of femur segmentation from MRI—with over 11,000 scans—Attention U-Net achieved Dice scores up to 0.954, outperforming conventional U-Net and even vision transformer methods [75]. These real-world values underscore that segmentation augmented with attention or transformer-based encoder improves bone delineation across anatomies.

Despite their promise, segmentation models require extensive manual annotations and carry high computational costs, which may limit scalability. Additionally, few studies perform external validation or deploy real-time pipelines. To bridge this gap, future research should emphasize semi-supervised annotation methods, multimodal U-Net

innovations incorporating attention mechanisms, and streamlined models for integration into clinical imaging workflows.

Hybrid and multi-stage approaches

Hybrid AI architectures that integrate segmentation, object detection, and classification have shown marked improvements in facial fracture diagnosis by leveraging each method's strengths [6]. Among the studies in our review, Jeong et al. implemented a two-stage process where YOLOX localized the nasal bone and GhostNetv2 then classified the presence [31]of fractures, yielding a diagnostic accuracy of 97.2% [18]. This approach also significantly enhanced the performance of less experienced readers. Van Nistelrooij et al. presented JawFracNet, a three-stage pipeline built upon CBCT images comprising mandible segmentation, fracture segmentation, and final classification; this model demonstrated superior fracture delineation and improved accuracy for small or complex mandibular breaks [33]. Similarly, Shahnavazi & Mohamadrahimi (2023) combined U-Net segmentation with Faster R-CNN detection to reduce irrelevant false positives in panoramic radiographs [30]. Son et al. enhanced YOLOv4 with luminance-adaptation preprocessing and added U-Net for segmenting fracture lines, balancing improved sensitivity with anatomical precision [31]. Wang et al. used a segmentation-first strategy to isolate the mandibular region before applying a ResNet-50 classifier, which reduced background interference and enhanced overall diagnostic accuracy [25].

These facial-focused hybrid models align with successful multi-stage approaches in other medical domains. For instance, Jia et al. developed an attention-based cascade R-CNN to detect sternal fractures on chest X-rays, achieving a mean average precision (mAP) of 0.71, precision of 0.90, and recall of 0.85—substantially higher than the baseline cascade R-CNN (precision 0.82, recall 0.77) [76]. A distal radius fracture study using RetinaNet, Faster R-CNN, and Cascade R-CNN also demonstrated high diagnostic AUCs (0.9706, 0.9658, 0.9644 respectively), with an ensemble model improving AUC to 0.9703 and accuracy to 97.6% [77]. Similarly, automated scaphoid fracture detection systems employing segmentation followed by detection achieved segmentation DSCs of ~ 97.4%, HD ~ 1.3 mm, AUC ~ 0.87, and sensitivity of 65% at a fixed specificity of 95%—matching or exceeding radiologist performance [78]. Another study using YOLOv8 for hip fracture detection reported a classification AUC of 0.981 and fracture segmentation Dice of ~ 0.77 [79].

Despite these promising outcomes, multi-stage systems face hurdles related to computational demand, complexity in annotation, and workflow integration. Both our review and supporting literature underscore the need for larger annotated datasets and streamlined architecture. Moreover, explainability, real-time application, and prospective clinical validation remain underexplored. Future work should aim to rigorously compare hybrid versus single-stage models in multi-center trials, implement explainable AI methods, and directly assess their impact on clinical metrics such as diagnostic speed and accuracy.

Explainability and interpretability in facial fracture detection models

Although most reviewed AI models reported high diagnostic accuracy, a notable limitation was the lack of explainability mechanisms. None of the included studies employed visualization or model-agnostic tools such as Grad-CAM [80], SHAP [81], or LIME [82], which are increasingly recognized as essential components of clinical AI systems [83].

Grad-CAM has become a prevalent method in musculoskeletal AI studies to generate visual attention maps that align with expert interpretation [64]. For instance, a two-stage ensemble model combining YOLOv5 and EfficientNet-B3 on distal radius fractures achieved an AUC of 0.82 and used Grad-CAM overlays to visually confirm fracture localization on wrist radiographs [64]. Similarly, Kim et al. used Grad-CAM in a hip fracture detection CNN, aiding interpretation and reducing model opacity [84].

Furthermore, a DenseNet-161/ResNet-152 ensemble in emergency wrist trauma achieved AUCs of 0.962 and 0.947, respectively, with Grad-CAM highlighting fracture regions and achieving test accuracy of ~ 90% [85]. A recent fine-grained ensemble study using Grad-CAM on wrist pathology X-rays underscored its value in enhancing classification accuracy from subtle visual cues [62]. Another study by Abdelrahman et al. employed SHAP in breast cancer mammography classification, achieving AUCs above 0.90 while identifying key pixel-level drivers of predictions [86]. Figure 3 consolidates the primary insights and challenges related to AI models in facial fracture detection, including data quality, model performance, clinical applicability, and research gaps.

The key methodological trends and translational barriers in AI-assisted facial-fracture diagnosis can be seen in Fig. 3. The predominant algorithmic paradigms (object detection, classification, and segmentation) deployed across CT, CBCT, and radiography, and the anatomical emphasis of current work (mandible, nasal, zygomatic, frontal), where many studies report strong internal performance (balanced sensitivity/specificity and AUCs) but are based largely on single-centre, retrospective cohorts [7, 23, 42, 45, 85]. These strengths are juxtaposed with recurrent limitations—heterogeneous and labor-intensive annotation practices,

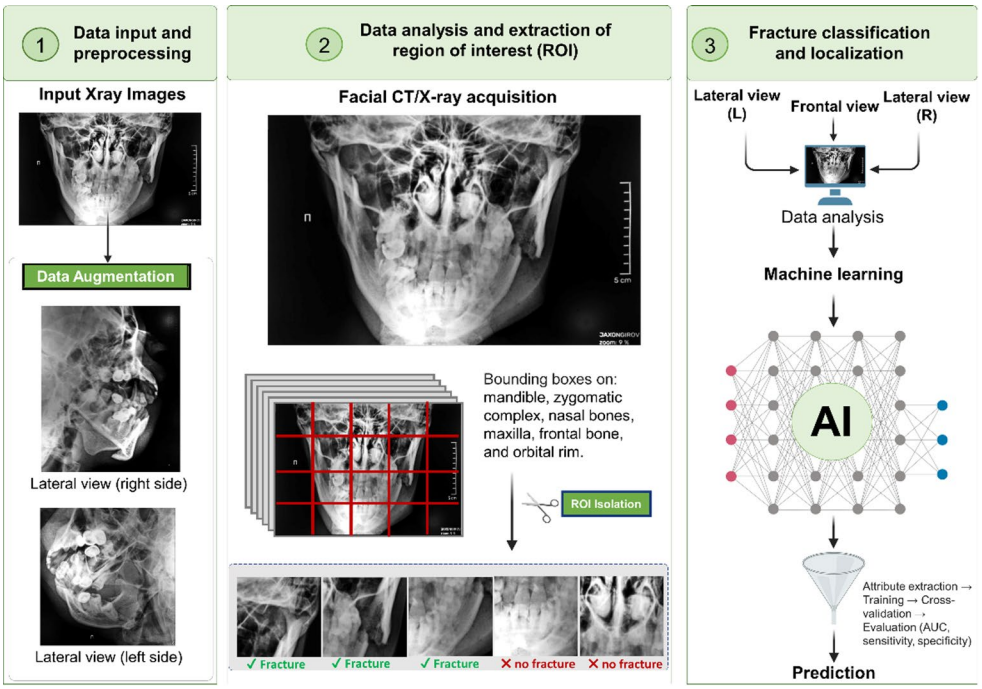

Fig. 2 AI-Based Diagnostic Workflow for Automated Detection of Facial Bone Fractures from Radiographic Images

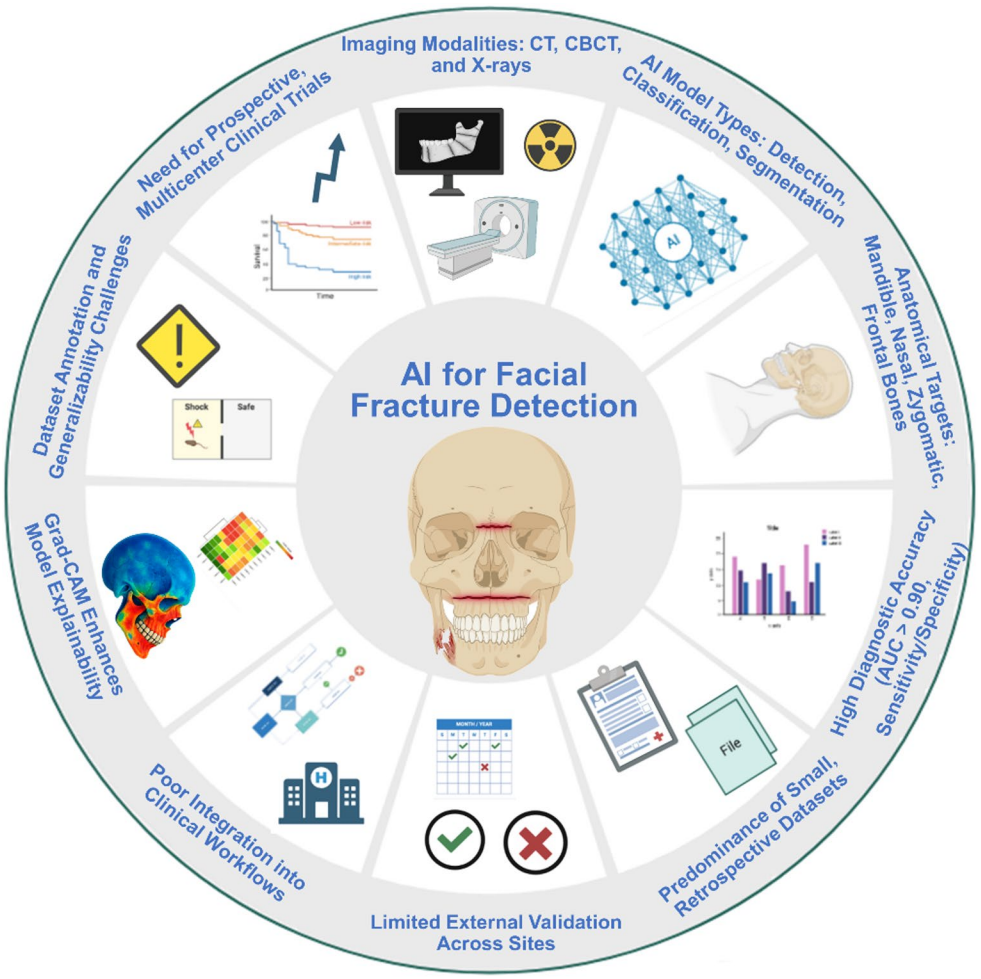

Fig. 3 Summary of key aspects in AI-based facial fracture detection

limited cross-site benchmarking, scarce external validation, and underuse of explainability methods (Grad-CAM) that are important for clinician trust and error analysis [23, 80]. The central message is prospective and multicentre evaluation with harmonized reporting and reference standards, coupled with workflow-aware implementation studies, are now the critical next steps to move from promising technical results to safe, reproducible clinical adoption.

Conclusion

This systematic review evaluated the diagnostic performance of artificial intelligence (AI) models in detecting facial bone fractures across multiple imaging modalities and anatomical regions. The findings demonstrate that deep learning algorithms, particularly object detection models such as YOLOv5 and Faster R-CNN, achieve high sensitivity and accuracy in localizing fractures, while classification networks like ResNet and Swin Transformer deliver robust binary detection outcomes. Segmentation models and hybrid architectures further enhanced anatomical specificity and interpretability. Despite strong technical performance, most included studies were limited by retrospective design, single-center datasets, and insufficient external validation, which restrict their generalizability and clinical translation.

Perspectives and future directions

To advance the clinical adoption of AI in facial trauma diagnostics, future studies should focus on prospective, multicenter evaluations with standardized reporting of diagnostic performance metrics. The integration of explainable AI frameworks, harmonized annotation protocols, and external benchmarking datasets will be essential to support transparency and reproducibility. Additionally, embedding AI models within radiology workflows and assessing their real-world impact on clinical decision-making, diagnostic efficiency, and patient outcomes should be prioritized. With continued methodological refinement and clinical validation, AI has the potential to significantly augment diagnostic accuracy, streamline trauma workflows, and improve access to expert-level fracture assessment, particularly in resource-limited or high-demand settings.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11282-025-00878-3.

Author contributions U.K.: performed the literature search, the data analysis, drafted and revised the work. S.B. and J.A.: revised the work and subsequent reading. Z.H. and S.S.: performed the data analysis, and subsequent reading. N.T.: had the idea for the article, performed the literature search, data analysis, and drafted and critically revised the work. All authors have read and agreed to the published version of the manuscript.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability Not applicable.

Declarations

Conflict of interest The authors declare no competing interests.

Ethics approval Not applicable.

Consent for publication All authors have read and agreed to the published version of the manuscript.

Informed consent Not applicable.

Institutional review board statement Not applicable.

References

1. Parsons MS, Policeni B, Juliano AF, Agarwal M, Benjamin ER, Burns J, Doerr T, Dubey P, Friedman ER, Gule-Monroe MK, et al. ACR appropriateness Criteria® imaging of facial trauma following primary survey. J Am Coll Radiol. 2022;19(5s):S67–86. https://doi.org/10.1016/j.jacr.2022.02.013.
2. Salar T, Tuygunov N, Yahya NA, Aziz AA. The impact of resin coatings on the properties and performance of glass ionomer cements: a systematic review. J Mech Behav Biomed Mater 2025:107044; https://doi.org/10.1016/j.jmbbm.2025.107044
3. Parsons MS, Policeni B, Juliano AF, Agarwal M, Benjamin ER, Burns J, Doerr T, Dubey P, Friedman ER, Gule-Monroe MK. ACR appropriateness criteria® imaging of facial trauma following primary survey. J Am Coll Radiol. 2022;19(5):S67–86.
4. Christie B, Musri N, Djustiana N, Takarini V, Tuygunov N, Zakaria MN, Cahyanto A. Advances and challenges in regenerative dentistry: A systematic review of calcium phosphate and silicate-based materials on human dental pulp stem cells. Mater Today Bio. 2023;23:100815. https://doi.org/10.1016/j.mtbio.2023.100815.
5. Tuygunov N, Zakaria MN, Yahya NA, Abdul Aziz A, Cahyanto A. Efficacy and bone-contact biocompatibility of glass ionomer cement as a biomaterial for bone regeneration: A systematic review. J Mech Behav Biomed Mater. 2023;146:106099. https://doi.org/10.1016/j.jmbbm.2023.106099.
6. Pham TD, Holmes SB, Coulthard P. A review on artificial intelligence for the diagnosis of fractures in facial trauma imaging. Front Artif Intell. 2023;6:1278529. https://doi.org/10.3389/frai.2023.1278529.
7. Kutbi M. Artificial Intelligence-Based applications for bone fracture detection using medical images: A systematic review. Diagnostics (Basel). 2024;14(17). https://doi.org/10.3390/diagnostics14171879.
8. Nseir S, Abu Shqara F, Krasovsky A, Rachmiel A. Surgical dilemmas in multiple facial Fractures - Coronal flap versus minimally invasive: case report and literature review. Ann Maxillofac Surg. 2021;11(1):191–4. https://doi.org/10.4103/ams.ams_452_20.
9. Pour ZM, Berretti S. A comprehensive review of AI methods in upper extremity/limb bone fracture detection. Artif Intell Rev. 2025;58(10):307. https://doi.org/10.1007/s10462-025-11296-6.

10. Samaranayake L, Tuygunov N, Schwendicke F, Osathanon T, Khurshid Z, Boymuradov SA, Cahyanto A. The transformative role of artificial intelligence in dentistry: A comprehensive Overview. Part 1: fundamentals of AI, and its contemporary applications in dentistry. Int Dent J. 2025;75(2):383–96. https://doi.org/10.1016/j.identj.2025.02.005.
11. Jung J, Dai J, Liu B, 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(1):e0000438. https://doi.org/10.1371/journal.pdig.0000438.
12. Tuygunov N, Samaranayake L, Khurshid Z, Rewthamrongsris P, Schwendicke F, Osathanon T, Yahya NA. The transformative role of artificial intelligence in dentistry: A comprehensive overview part 2: the promise and Perils, and the international dental federation communique. Int Dent J. 2025;75(2):397–404. https://doi.org/10.1016/j.identj.2025.02.006.
13. Gellrich NC, Eckstein FM, Lentge F, Zeller AN, Korn P. [Complex reconstructions in the facial and cranial regions]. Unfallchirurg. 2021;124(10):807–16. https://doi.org/10.1007/s00113-021-01076-6.
14. Shah S, Uppal SK, Mittal RK, Garg R, Saggar K, Dhawan R. Diagnostic tools in maxillofacial fractures: is there really a need of three-dimensional computed tomography? Indian J Plast Surg. 2016;49(2):225–33. https://doi.org/10.4103/0970-0358.191320.
15. Morita D, Kawarazaki A, Soufi M, Otake Y, Sato Y, Numajiri T. Automatic detection of midfacial fractures in facial bone CT images using deep learning-based object detection models. J Stomatol Oral Maxillofac Surg. 2024;125(5s2):101914. https://doi.org/10.1016/j.jormas.2024.101914.
16. Aldhyani T, Ahmed ZAT, Alsharbi BM, Ahmad S, Al-Adhaileh MH, Kamal AH, Almaiah M, Nazeer J. Diagnosis and detection of bone fracture in radiographic images using deep learning approaches. Front Med (Lausanne). 2024;11:1506686. https://doi.org/10.3389/fmed.2024.1506686.
17. Mao J, Du Y, Xue J, Hu J, Mai Q, Zhou T, Zhou Z. Automated detection and classification of mandibular fractures on multislice spiral computed tomography using modified convolutional neural networks. Oral Surg Oral Med Oral Pathol Oral Radiol. 2024;138(6):803–12. https://doi.org/10.1016/j.oooo.2024.07.010.
18. Jeong Y, Jeong C, Sung KY, Moon G, Lim J. Development of AI-Based diagnostic algorithm for nasal bone fracture using deep learning. J Craniofac Surg. 2024;35(1):29–32. https://doi.org/10.1097/scs.0000000000009856.
19. Moon G, Kim S, Kim W, Kim Y, Jeong Y, Choi HS. Computer aided facial bone fracture diagnosis (CA-FBFD) system based on object detection model. IEEE Access. 2022;10:79061–70. https://doi.org/10.1109/ACCESS.2022.3192389.
20. Warin K, Limprasert W, Suebnukarn S, Inglam S, Jantana P, Vicharueang S. Assessment of deep convolutional neural network models for mandibular fracture detection in panoramic radiographs. Int J Oral Maxillofac Surg. 2022;51(11):1488–94. https://doi.org/10.1016/j.ijom.2022.03.056.
21. Warin K, Vicharueang S, Jantana P, Limprasert W, Thanathornwong B, Suebnukarn S. Deep learning for midfacial fracture detection in CT images. Stud Health Technol Inf. 2024;310:1497–8. https://doi.org/10.3233/shti231262.
22. Anderson PG, Baum GL, Keathley N, Sicular S, Venkatesh S, Sharma A, Daluiski A, Potter H, Hotchkiss R, Lindsey RV, et al. Deep learning assistance closes the accuracy gap in fracture detection across clinician types. Clin Orthop Relat Res. 2023;481(3):580–8. https://doi.org/10.1097/corr.0000000000002385.
23. Mortezaei T, Dalili Kajan Z, Mirroshandel SA, Mehrpour M, Shahidzadeh S. Application of deep learning for detection of nasal bone fracture on X-Ray nasal bone lateral view. Dentomaxillofac Radiol. 2025. https://doi.org/10.1093/dmfr/twaf028.
24. Seol YJ, Kim YJ, Kim YS, Cheon YW, Kim KG. A study on 3D deep learning-based automatic diagnosis of nasal fractures. Sensors (Basel, Switzerland) 2022;22(2):506. https://doi.org/10.3390/s22020506.
25. Wang X, Xu Z, Tong Y, Xia L, Jie B, Ding P, Bai H, Zhang Y, He Y. Detection and classification of mandibular fracture on CT scan using deep convolutional neural network. Clin Oral Investig. 2022;26(6):4593–601. https://doi.org/10.1007/s00784-022-04427-8.
26. Nishiyama M, Ishibashi K, Ariji Y, Fukuda M, Nishiyama W, Umemura M, Katsumata A, Fujita H, Ariji E. Performance of deep learning models constructed using panoramic radiographs from two hospitals to diagnose fractures of the mandibular condyle. Dentomaxillofac Radiol. 2021;50(7):20200611. https://doi.org/10.1259/dmfr.20200611.
27. Nam Y, Choi Y, Kang J, Seo M, Heo SJ, Lee MK. Diagnosis of nasal bone fractures on plain radiographs via convolutional neural networks. Sci Rep. 2022;12(1):21510. https://doi.org/10.1038/s41598-022-26161-7.
28. Zubair M, Owais M, Hassan T, Bendechache M, Hussain M, Hussain I, Werghi N. An interpretable framework for gastric cancer classification using multi-channel attention mechanisms and transfer learning approach on histopathology images. Sci Rep. 2025;15(1):13087. https://doi.org/10.1038/s41598-025-97256-0.
29. Tong Y, Jie B, Wang X, Xu Z, Ding P, He Y. Is convolutional neural network accurate for automatic detection of zygomatic fractures on computed tomography? J Oral Maxillofac Surg. 2023;81(8):1011–20. https://doi.org/10.1016/j.joms.2023.04.013.
30. Shahnavazi M, Mohamadrahimi H. The application of artificial neural networks in the detection of mandibular fractures using panoramic radiography. Dent Res J (Isfahan). 2023;20(27). https://doi.org/10.4103/1735-3327.369629.
31. Son DM, Yoon YA, Kwon HJ, Lee SH. Combined deep learning techniques for mandibular fracture diagnosis assistance. Life (Basel). 2022;12(11). https://doi.org/10.3390/life12111711.
32. Shan W, Guo J, Mao X, Zhang Y, Huang Y, Wang S, Li Z, Meng X, Zhang P, Wu Z, et al. Automated identification of skull fractures with deep learning: A comparison between object detection and segmentation approach. Front Neurol. 2021;12:687931. https://doi.org/10.3389/fneur.2021.687931.
33. van Nistelrooij N, Schitter S, van Lierop P, Ghoul KE, König D, Hanisch M, Tel A, Xi T, Thiem DGE, Smeets R, et al. Detecting mandible fractures in CBCT scans using a 3-Stage neural network. J Dent Res. 2024;103(13):1384–91. https://doi.org/10.1177/00220345241256618.
34. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, Tridandapani S, Auffermann WF. Deep learning in radiology. Acad Radiol. 2018;25(11):1472–80. https://doi.org/10.1016/j.acra.2018.02.018.
35. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019;1(6):e271–97. https://doi.org/10.1016/s2589-7500(19)30123-2.
36. Kalmet PHS, Sanduleanu S, Primakov S, Wu G, Jochems A, Refaee T, Ibrahim A, Hulst LV, Lambin P, Poeze M. Deep learning in fracture detection: a narrative review. Acta Orthop. 2020;91(2):215–20. https://doi.org/10.1080/17453674.2019.1711323.
37. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. The PRISMA 2020 statement: an updated guideline for reporting

systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71.
38. Mortezaei T, Dalili Kajan Z, Mirroshandel SA, Mehrpour M, Shahidzadeh S. Application of deep learning for detection of nasal bone fracture on X-ray nasal bone lateral view. Dentomaxillofacial Radiol. 2025;twaf028. https://doi.org/10.1093/dmfr/twaf028.
39. Mao J, Du Y, Xue J, Hu J, Mai Q, Zhou T, Zhou Z. Automated detection and classification of mandibular fractures on multislice spiral computed tomography using modified convolutional neural networks. Oral Surg Oral Med Oral Pathol Oral Radiol. 2024;138(6):803–12. https://doi.org/10.1016/j.oooo.2024.07.010.
40. Son DM, Yoon YA, Kwon HJ, An CH, Lee SH. Automatic detection of mandibular fractures in panoramic radiographs using deep learning. Diagnostics (Basel). 2021;11(6). https://doi.org/10.3390/diagnostics11060933.
41. Vinayahalingam S, van Nistelrooij N, van Ginneken B, Bressem K, Tröltzsch D, Heiland M, Flügge T, Gaudin R. Detection of mandibular fractures on panoramic radiographs using deep learning. Sci Rep. 2022;12(1):19596. https://doi.org/10.1038/s41598-022-23445-w.
42. Yang C, Yang L, Gao GD, Zong HQ, Gao D. Assessment of artificial intelligence-aided reading in the detection of nasal bone fractures. Technol Health Care. 2023;31(3):1017–25. https://doi.org/10.3233/thc-220501.
43. Yari A, Fasih P, Hosseini Hooshiar M, Goodarzi A, Fattahi SF. Detection and classification of mandibular fractures in panoramic radiography using artificial intelligence. Dentomaxillofac Radiol. 2024;53(6):363–71. https://doi.org/10.1093/dmfr/twae018.
44. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, Mahajan V, Rao P, Warier P. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet. 2018;392(10162):2388–96. https://doi.org/10.1016/s0140-6736(18)31645-3.
45. Wang HC, Wang SC, Yan JL, Ko LW. Artificial intelligence model trained with sparse data to detect facial and cranial bone fractures from head CT. J Digit Imaging. 2023;36(4):1408–18. https://doi.org/10.1007/s10278-023-00829-6.
46. Yang S, Chen J, Li A, Deng K, Li P, Xu S. Accuracy of autonomous robotic surgery for single-tooth implant placement: A case series. J Dent. 2023;132:104451.
47. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
48. Alkandary K, Yildiz AS, Meng H. A comparative study of YOLO series (v3–v10) with DeepSORT and strongsort: a real-time tracking performance study. Electronics. 2025;14(5):876. https://doi.org/10.3390/electronics14050876.
49. Patel Y, Patel D, Patel H, Gupta R, Tanwar S. Faster R-CNN based framework For bone fracture detection in telesurgery systems in healthcare 4.0. In: 2024 5th IEEE global conference for advancement in technology (GCAT): 4–6 Oct. 2024 2024. 1–6.
50. Douglas JA, Gill K, Holmes S. Combining trauma severity indices to create a unified craniofacial disruption index: addition of the frontobasal unit to the ZS model. Br J Oral Maxillofac Surg. 2020;58(7):784–8. https://doi.org/10.1016/j.bjoms.2020.04.005.
51. Pranata YD, Wang KC, Wang JC, Idram I, Lai JY, Liu JW, Hsieh IH. Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. Comput Methods Programs Biomed. 2019;171:27–37. https://doi.org/10.1016/j.cmpb.2019.02.006.
52. Meng XH, Wu DJ, Wang Z, Ma XL, Dong XM, Liu AE, Chen L. A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance. Skeletal Radiol. 2021;50(9):1821–8. https://doi.org/10.1007/s00256-021-03709-8.
53. 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):535–43. https://doi.org/10.1093/bjr/tqae014.
54. Yilmaz A, Gem K, Kalebasi M, Varol R, Gencoglan ZO, Samoylenko Y, Tosyali HK, Okcu G, Uvet H. An automated hip fracture detection, classification system on pelvic radiographs and comparison with 35 clinicians. Sci Rep. 2025;15(1):16001. https://doi.org/10.1038/s41598-025-98852-w.
55. Rahman R, Yagi N, Hayashi K, Maruo A, Muratsu H, Kobashi S. Enhancing fracture diagnosis in pelvic X-rays by deep convolutional neural network with synthesized images from 3D-CT. Sci Rep. 2024;14. https://doi.org/10.1038/s41598-024-58810-4.
56. Chou P-H, Jou TH-T, Wu H-TH, Yao Y-C, Lin H-H, Chang M-C, Wang S-T, Lu HH-S, Chen H-H. Ground truth generalizability affects performance of the artificial intelligence model in automated vertebral fracture detection on plain lateral radiographs of the spine. Spine J. 2022;22(4):511–23. https://doi.org/10.1016/j.spinee.2021.10.020.
57. Chen P, Liu S, Lu W, Lu F, Ding B. WCAY object detection of fractures for X-ray images of multiple sites. Sci Rep. 2024;14(1):26702. https://doi.org/10.1038/s41598-024-77878-6.
58. Shou Y, Meng T, Ai W, Xie C, Liu H, Wang Y. Object detection in medical images based on hierarchical transformer and mask mechanism. Comput Intell Neurosci. 2022;2022:5863782. https://doi.org/10.1155/2022/5863782.
59. Kanagaraj DK, Oliver A, Ranjith R. Bone fracture detection and classification using node-level capsule graph neural network with X-ray images of broken and unbroken bones. Biomed Signal Process Control. 2025;110:108302. https://doi.org/10.1016/j.bspc.2025.108302.
60. Yadav DP, Rathor S. Bone Fracture Detection and Classification using Deep Learning Approach. In: *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC): 28–29 Feb. 2020 2020*. 282–285.
61. Yadav DP, Sharma A, Athithan S, Bhola A, Sharma B, Dhaou IB. Hybrid SFNet model for bone fracture detection and classification using ML/DL. Sens (Basel). 2022;22(15). https://doi.org/10.3390/s22155823.
62. Ahmed A, Imran AS, Ullah M, Kastrati Z, Daudpota SM. Navigating limitations with precision: a fine-grained ensemble approach to wrist pathology recognition on a limited X-ray dataset. In: 2024 IEEE International Conference on Image Processing (ICIP): 27–30 Oct. 2024 2024. 3077–3083.
63. Min H, Rabi Y, Wadhawan A, Bourgeat P, Dowling J, White J, Tchernegovski A, Formanek B, Schuetz M, Mitchell G, et al. Automatic classification of distal radius fracture using a two-stage ensemble deep learning framework. Phys Eng Sci Med. 2023;46(2):877–86. https://doi.org/10.1007/s13246-023-01261-4.
64. Gan K, Liu Y, Zhang T, Xu D, Lian L, Luo Z, Li J, Lu L. Deep learning model for automatic identification and classification of distal radius fracture. J Imaging Inf Med. 2024;37(6):2874–82. https://doi.org/10.1007/s10278-024-01144-4.
65. Oh J, Hwang S, Lee J. Enhancing X-ray-Based wrist fracture diagnosis using HyperColumn-convolutional block attention module. Diagnostics 2023;13(18):2927. https://doi.org/10.3390/diagnostics13182927.
66. Ahmed A, Imran AS, Kastrati Z, Daudpota SM, Ullah M, Noor W. Learning from the few: Fine-grained approach to pediatric wrist pathology recognition on a limited dataset. Comput Biol Med. 2024;181:109044. https://doi.org/10.1016/j.compbiomed.2024.109044.

67. Ghanem M, Ghaith AK, El-Hajj VG, Bhandarkar A, de Giorgio A, Elmi-Terander A, Bydon M. Limitations in evaluating machine learning models for imbalanced binary outcome classification in spine surgery: a systematic review. Brain Sci. 2023;13(12). https://doi.org/10.3390/brainsci13121723.
68. Panhwar A, Memon S, Dhomeja L, Memon N, Ali A, Bukhari A. Deep Learning-Based image segmentation techniques for bone fractures using X-Ray images: A systematic review. VFAST Trans Softw Eng. 2024;12:99–116. https://doi.org/10.21015/vtse.v12i4.1971.
69. Liu Y, Yibulayimu S, Zhu G, Shi C, Liang C, Zhao C, Wu X, Sang Y, Wang Y. Automatic pelvic fracture segmentation: a deep learning approach and benchmark dataset. Front Med (Lausanne). 2025;12:1511487. https://doi.org/10.3389/fmed.2025.1511487.
70. Yang Q, Zhang L, Xi Z, Qian Y, Li A. The lightweight fracture segmentation algorithm for logging images based on fully 3D attention mechanism and deformable convolution. Appl Sci. 2024;14(22):10662. https://doi.org/10.3390/app142210662.
71. Meena T, Roy S. Bone fracture detection using deep supervised learning from radiological images: a paradigm shift. Diagnostics (Basel). 2022;12(10). https://doi.org/10.3390/diagnostics12102420.
72. Alwzwazy A, Alzubaidi H, Zhao L, Gu Z. FracNet: an end-to-end deep learning framework for bone fracture detection. Pattern Recognit Lett. 2025;190:1–7. https://doi.org/10.1016/j.patrec.2025.01.034.
73. Lee JM, Park JY, Kim YJ, Kim KG. Deep-learning-based pelvic automatic segmentation in pelvic fractures. Sci Rep. 2024;14(1):12258. https://doi.org/10.1038/s41598-024-63093-w.
74. Shim J-H, Kim WS, Kim KG, Yee GT, Kim YJ, Jeong TS. Evaluation of U-Net models in automated cervical spine and cranial bone segmentation using X-ray images for traumatic atlanto-occipital dislocation diagnosis. Sci Reports. 2022;12:21438.
75. Liu M, Chen Y, Tian A, Wu X, Shen M, Gong T, Lee J. Performance Analysis of Deep Learning Models for Femur Segmentation in MRI Scan; 2025.
76. Jia Y, Wang H, Chen W, Wang yg, Yang B. An attention-based cascade R-CNN model for sternum fracture detection in X-ray images. CAAI Trans Intell Technol. 2022;7. https://doi.org/10.1049/cit2.12072.
77. Zhang J, Li Z, Lin H, Xue M, Wang H, Fang Y, Liu S, Huo T, Zhou H, Yang J, et al. Deep learning assisted diagnosis system: improving the diagnostic accuracy of distal radius fractures. Front Med (Lausanne). 2023;10:1224489. https://doi.org/10.3389/fmed.2023.1224489.
78. Hendrix N, Scholten E, Vernhout B, Bruijnen S, Maresch B, de Jong M, Diepstraten S, Bollen S, Schalekamp S, de Rooij M, et al. Development and validation of a convolutional neural network for automated detection of scaphoid fractures on conventional radiographs. Radiol Artif Intell. 2021;3(4):e200260. https://doi.org/10.1148/ryai.2021200260.
79. González G, Galant J, Salinas JM, Benítez E, Sánchez-Valverde MD, Calbo J, Cerrolaza N. Classification and segmentation of hip fractures in x-rays: highlighting fracture regions for interpretable diagnosis. Insights Imaging. 2025;16(1):86. https://doi.org/10.1186/s13244-025-01958-y.
80. Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos Solitons Fractals. 2020;140:110190. https://doi.org/10.1016/j.chaos.2020.110190.
81. Rotem O, Schwartz T, Maor R, Tauber Y, Shapiro MT, Meseguer M, Gilboa D, Seidman DS, Zaritsky A. Visual interpretability of image-based classification models by generative latent space disentanglement applied to in vitro fertilization. Nat Commun. 2024;15(1):7390. https://doi.org/10.1038/s41467-024-51136-9.
82. Lakshmi M, Das R. Classification of Monkeypox images using LIME-Enabled investigation of deep convolutional neural network. Diagnostics (Basel). 2023;13(9):1639. https://doi.org/10.3390/diagnostics13091639.
83. Frasca M, La Torre D, Pravettoni G, Cutica I. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. Discover Artif Intell. 2024;4(1):15. https://doi.org/10.1007/s44163-024-00114-7.
84. Dan HJ, Ji H, Zhou YQ, Li CY. Non-invasive evaluation of the common carotid artery in normal adults using ultrasound radiofrequency data technology. Clin Radiol. 2021;76(5):393. .e391-393.e397.
85. Kim MW, Jung J, Park SJ, Park YS, Yi JH, Yang WS, Kim JH, Cho B-J, Ha SO. Application of convolutional neural networks for distal radio-ulnar fracture detection on plain radiographs in the emergency room. Clin Exp Emerg Med. 2021;8(2):120–7. https://doi.org/10.15441/ceem.20.091.
86. Peng W, Chen T, Liu H, Dai W, Yu N, Lan W. Improving drug response prediction based on two-space graph Convolution. Comput Biol Med 2023, 158:106859; https://doi.org/10.1016/j.compbiomed.2023.106859