Supplementary Table 1: Image processing studies								
Article title	**Author**	**Year**	**Type of image**	**Aim**	**AI Methods**	**Performance metrics**	**Opportunities**	**Limitations**
Care to Explain? AI Explanation Types Differentially Impact Chest Radiograph Diagnostic Performance and Physician Trust in AI	Prinster et al.	2024	Chest Radiographs	To evaluate the effects of AI explanation types and confidence on diagnostic performance and trust in chest radiograph diagnosis.	ChestAId (AI tool) with explanation mechanisms	Local: Accuracy 92.8%, Global: Accuracy 85.3%	Enhances trust and diagnostic accuracy with appropriate explanation design, aiding adoption of AI in clinical practice.	Risk of overreliance on incorrect AI advice with local explanations; balancing trust and critical evaluation by clinicians.
Deep learning approach based on a patch residual for pediatric supracondylar subtle fracture detection	Ye et al.	2025	Pediatric Elbow X-Ray	To develop and validate a multiscale patch residual (MPR) network for subtle fracture detection and localization in pediatric elbow radiographs.	MPR network (CNN + GAN hybrid)	Accuracy: 90.5%, F1 Score: 0.906	Addresses challenges in subtle pediatric fracture diagnosis; rapid inference enhances emergency care efficiency.	Limited generalization to diverse imaging conditions; dependency on high-quality X-rays for optimal performance.
Deep Learning-Based Diagnosis of Pulmonary Tuberculosis on Chest X-ray in the Emergency Department: A Retrospective Study	Wang et al.	2024	Chest X-ray	To develop and validate a deep learning CAD algorithm for detecting pulmonary tuberculosis in ED settings.	EfficientNetV2, SegResNet for segmentation, GradCAM for explainability	Testing AUC: 0.878; External Validation: Montgomery AUC: 0.838, Shenzhen AUC: 0.806	Enhances diagnostic speed and accuracy in EDs; aids in early isolation and treatment of PTB patients.	Suboptimal performance on AP views; single-center retrospective design limits generalizability.
Deploying artificial intelligence in the detection of adult appendicular and pelvic fractures in the Singapore emergency department after hours: efficacy, cost savings and non-monetary benefits	Quek et al.	2024	X-ray	To evaluate the efficacy of AI in fracture detection during after-hours in the ED and assess the cost and non-monetary benefits of its implementation.	RBfracture (Radiobotics), a CE Class IIa AI solution with convolutional neural networks	Accuracy: 85.9%, Inter-rater reliability (Gwet's AC1): 0.837	Reduces workload and diagnostic delays, provides consistent and timely results, and minimizes patient callbacks.	Technical limitations with software and DICOM-tag processing, requiring updates and validation for broader deployment.
Detecting Pediatric Appendicular Fractures Using Artificial Intelligence	Kavak et al.	2024	Pediatric Radiographs (AP views)	To assess the diagnostic accuracy of a YOLOv8 AI model in detecting appendicular fractures in pediatric trauma cases and its impact on ED workflows.	YOLOv8 (Bounding Box Detection)	F1 score of 90%	Enhances diagnostic accuracy and efficiency; reduces diagnostic delays in pediatric fracture cases.	Limited ability to evaluate lateral radiographs; excluded cases of child abuse-specific fractures.
Diagnostic Performance of Artificial Intelligence in Chest Radiographs Referred from the Emergency Department	Lopez Alcolea et al.	2024	Chest X-ray (Digital)	To evaluate the diagnostic accuracy of AI in detecting fractures, pneumothorax, pulmonary opacities, pleural effusion, and pulmonary nodules on chest X-rays compared to human readers.	Arterys MICA v29.4.0 (CNN-based model)	AUC: 99.3% (fractures), 100% (Pneumothorax), 85.5% (opacity), 66.5% (nodules), 79.3% (Pleural effusion)	Enables efficient triage and prioritization in ED; potential to reduce workload for radiologists by screening non-pathological cases.	Low sensitivity for pulmonary nodules; high false-positive rates for subtle findings (e.g., effusions); limited to 5 predefined categories.
Evaluation of the impact of artificial intelligence-assisted image interpretation on the diagnostic performance of clinicians in identifying pneumothoraces on plain chest X-ray: a multi-case multi-reader study	Novak et al.	2024	Chest X-ray	To evaluate the impact of AI-assisted interpretation on diagnostic accuracy and confidence in detecting pneumothorax among clinicians of varying seniority.	GEHC CCS PTX algorithm (heatmap-based AI tool)	AUROC: 0.939	Improves diagnostic confidence and sensitivity, particularly for junior clinicians in acute care settings.	Single pathology focus; potential overreliance on AI findings without adequate training.

Multimodal deep learning models utilizing chest X-ray and electronic health record data for predictive screening of acute heart failure in emergency department	Lee et al.	2024	Chest X-ray (CXR)	To integrate multimodal data from CXR and EHR to predict NT-proBNP levels, improving the timeliness and accuracy of AHF diagnosis in emergency settings.	Multimodal deep learning using DenseNet121 for CXR and a numerical feature encoder for EHR data	AUROC: 0.89	Provides rapid and accurate AHF diagnosis, integrates easily into ED workflows, and reduces dependence on delayed lab results.	Requires validation in diverse populations; limited by demographic and positional data in current datasets.
The Potential Clinical Utility of an Artificial Intelligence Model for Identification of Vertebral Compression Fractures in Chest Radiographs	Ghatak et al.	2024	Chest Radiographs (CXR)	To assess the accuracy of the Annalise Enterprise CXR AI model for identifying vertebral compression fractures and its potential to improve care for undiagnosed osteoporosis patients.	Annalise Enterprise CXR AI model (version 2.2.0), consisting of CNNs for classification and segmentation	AUC: 0.955	Provides automated screening for vertebral compression fractures, enabling timely osteoporosis treatment and reducing healthcare costs.	Limited to binary output without localization; requires validation for broader deployment and subgroup-specific performance assessments.
A Deep Learning-Powered Diagnostic Model for Acute Pancreatitis	Zhang et al.	2024	Abdominal CT	To construct a DL-based diagnostic model for recognizing acute pancreatitis and evaluating its severity.	U-Net (segmentation), Baidu EasyDL platform	Classifier: AUC 0.993 (test set), AUC 0.850 (external); Segmentation: MIOU: 86.02 (pancreas segmentation), 61.81 (peripancreatic inflammatory exudation),57.73 (peripancreatic effusion), 66.36 (peripancreatic abscess necrosis)	Improves early diagnosis and severity evaluation of acute pancreatitis; supports deployment in primary hospitals.Supports rapid and accurate triage; potentially reduces misdiagnosis rates in emergency settings.	Limited dataset; suboptimal segmentation for effusion and necrosis; single-center study. limited segmentation accuracy for rare complications like abscess necrosis.
A Novel Deep Learning-based Artificial Intelligence System for Interpreting Urolithiasis in Computed Tomography.	Kim et al.	2024	Computed Tomography (CT)	To develop a deep learning-based AI system for detecting urolithiasis and calculating key stone parameters in real-time.	YOLOv4 architecture with advanced convolutional features and spatial pyramid pooling.	Internal validation: Detection accuracy 95%; External validation: Accuracy 94%.	Rapid and real-time urolithiasis diagnosis; provides reliable measurements aiding treatment decisions.	Overfitting issues and limited generalization for rare stone types or small UVJ stones.
Artificial Intelligence Application in Skull Bone Fracture with Segmentation Approach	Lu et al.	2024	Emergent CT scans (3-mm slices)	To develop an AI segmentation model for classifying and localizing skull fractures	2D U-Net with EfficientNet B6 backbone	F1 score 0.94	Enables precise localization and classification; reduces diagnostic duration significantly, aiding in emergency care.	High false positives in complex cases; lack of multi-center validation and reliance on 2D architecture for a 3D problem.
Automated Detection and Differentiation of Stanford Type A and Type B Aortic Dissections in CTA Scans Using Deep Learning	Liu et al.	2025	Aortic CTA	To develop an AI-based triage tool for automatic detection and differentiation of Stanford Type A and Type B aortic dissections.	RetinaNet (Object Detection), EfficientNet-B0 (Classification)	Accuracy 0.996, AUC 0.990	Accelerates diagnosis of life-threatening conditions, improving time-critical surgical interventions.	Limited generalizability due to single-institution data; excludes complex cases like thrombus and abdominal involvement.
Automated Hematoma Detection and Outcome Prediction in Patients With Traumatic Brain Injury	Xu et al.	2024	Head CT	To enhance the detection, segmentation, and outcome prediction of hematomas in TBI using a novel cascade deep learning framework.	DeepLabV3+, UNet (segmentation); ResNet-50, EfficientNet-B6 (classification)	AUC (mortality): 0.91; Accuracy (classification): 95.91%	Streamlines TBI management by improving hematoma detection and mortality prediction accuracy, aiding clinical decision-making.	Limited external validation, performance reliant on high-quality annotations and diverse datasets.

Diagnostic Performance of a Deep Learning-Powered Application for Aortic Dissection Triage Prioritization and Classification	Laletin et al.	2024	Chest and Thoraco-Abdominal Angiography (CTA)	To evaluate the diagnostic performance and clinical utility of a DL-powered tool for aortic dissection triage, prioritization, and type classification.	Hybrid 3D/2D U-Net-based segmentation and detection algorithm	Sensitivity: 94.2%, Specificity: 97.3%, AUC: 0.96; Type A accuracy: 99.5%, Type B accuracy: 97.5%; Time to notification: 27.9 s	Enables rapid triage and prioritization for emergent cases; robust across diverse scanner models and acquisition parameters.	Misses subtle cases with artifacts; limited validation for rare conditions like intramural hematoma and PAU.
Impact of a deep learning-based brain CT interpretation algorithm on clinical decision-making for intracranial hemorrhage in the emergency department	Choi et al.	2024	Non-enhanced Brain CT	To evaluate the impact of a DLHD algorithm on emergency clinicians' diagnostic accuracy and decision-making consistency.	2D U-Net with Inception module	Accuracy 83.32, AUROC: 0.788	Enhances sensitivity for inexperienced clinicians, supporting early hemorrhage detection.	Reduces specificity due to false positives; requires careful integration to avoid overreliance.
Improved Differentiation of Cavernous Malformation and Acute Intraparenchymal Hemorrhage on CT Using an AI Algorithm	Kim et al.	2024	Non-contrast Brain CT	To evaluate the utility of an AI algorithm in distinguishing CCM from AIH on brain CT, and its impact on diagnostic accuracy across different physician experience levels.	Medical Insight + Brain Hemorrhage (SK Inc. C&C)	Accuracy: 86.92% vs. 79.86% (Brain CT interpretation with/without AI assistance for radiology residents and emergency department physicians)	Improves diagnostic accuracy for less experienced clinicians; bridges gap between expert and non-expert practitioners.	Limited validation for small lesions; requires careful integration to avoid dependency among trainees.
SMART: Development and Application of a Multimodal Multi-organ Trauma Screening Model for Abdominal Injuries in Emergency Settings	Wang et al.	2024	Non-contrast Abdominal CT	To construct and evaluate a multimodal diagnostic model combining text and imaging data for abdominal trauma screening in emergency settings.	GPT-4 embeddings, nnU-Net, DenseNet121, Logistic Regression	SMART Sensitivity: 93.8%; SMART AUC: 0.88; Waiting time reduction: 64.24%	Facilitates rapid trauma diagnosis using multimodal integration; improves emergency department workflow.	Limited generalization due to single-center data; potential privacy concerns with GPT-4 embedding usage.
Towards clinical implementation of an AI-algorithm for detection of cervical spine fractures on computed tomography	Ruitenbeek et al.	2024	Non-contrast Cervical Spine CT	To evaluate the effect of an AI algorithm on diagnostic accuracy and time to diagnosis for cervical spine fractures in a clinical setting.	AIDoc Medical AI algorithm	Sensitivity: 89.8%, Specificity: 95.3%, Diagnostic Accuracy: 94.8%, DNT reduced by 16 minutes for fractured cases.	Enhances triage speed and diagnostic workflow efficiency; supports early stabilization decisions.	Missed clinically relevant fractures; limited generalization due to single-center data.
Using an artificial intelligence software improves emergency medicine physician intracranial haemorrhage detection to radiologist levels	Warman et al.	2024	Non-contrast Cranial CT (NCCT)	To evaluate whether the Caire ICH AI tool can improve emergency physicians' detection of intracranial hemorrhages to levels comparable to radiologists.	Caire ICH deep learning software using predetermined operating thresholds	Accuracy improved by 6.2% (from 85.82% to 92.02%, p=0.0092)	Accelerates emergency care by reducing diagnostic delays, enabling emergency physicians to make faster, accurate decisions.	Limited generalizability due to single-software evaluation; requires larger multicenter validation and training.
Assessment of Deep Learning-Based Triage Application for Acute Ischemic Stroke on Brain MRI in the ER	Kim et al.	2024	Brain MRI (DWI, FLAIR, optional T2WI)	To evaluate the performance of an AI-based application in detecting acute ischemic stroke on brain MRI in an emergency setting, and to examine the impact of optional T2WI sequences.	Deep learning-based 3D convolutional neural network (Neuro Triage Application, Siemens)	Sensitivity: 90%, Specificity: 89%, Accuracy: 89%, AUROC: 0.95	Reduces diagnostic time and enhances triage accuracy in ER settings; provides structured, consistent outputs for radiologists.	Limited to single-center data; optional T2WI sequences increase processing time without added diagnostic value.

Clinical Evaluation of a 2-Minute Ultrafast Brain MR Protocol for Evaluation of Acute Pathology in the Emergency and Inpatient Settings	Lang et al.	2024	Brain MRI (T1, T2, T2*, FLAIR, DWI)	To assess the diagnostic performance and feasibility of an ultrafast 2-minute brain MR protocol for acute pathology.	Machine-learning assisted reconstruction framework for noise reduction and improved image quality in ms-EPI.	98.5% agreement in diagnosis; motion artifacts reduced in ultrafast protocol ($P < .01$).	Improves accessibility to MR imaging in emergency settings, reduces scan time by 80%, and minimizes motion artifacts.	Higher image noise and geometric distortion; limited sensitivity for small findings; single-center study with a small sample size.
Deep learning model integrating radiologic and clinical data to predict mortality after ischemic stroke	Kim et al.	2024	Brain DWI and ADC	To validate a deep learning-based mortality prediction model integrating radiologic and clinical data for ischemic stroke patients.	DLP_DWI (DenseNet-169), Integrated Cox model (DLP_INTG)	TD AUC (internal): 0.859; TD AUC (external): 0.876; C-index (internal): 0.852	Enhances predictive accuracy for long-term outcomes in ischemic stroke, supporting treatment stratification.	Limited external validation, imbalanced dataset, and geographic variability in patient demographics.
AI-Enabled Assessment of Cardiac Function and Video Quality in Emergency Department Point-of-Care Echocardiograms	He et al.	2023	Point-of-care ultrasound (POCUS)	To develop EchoNet-POCUS, a deep learning model for assessing cardiac function and video quality in POCUS echocardiograms, and to reduce operator variability.	Convolutional Neural Network (CNN) with spatiotemporal convolutions	AUROC for cardiac function: 0.92, AUROC for video quality: 0.81	Real-time evaluation at bedside, low-cost implementation, reduces operator variability, and improves diagnostic accuracy.	Limited prospective study size; requires validation across multiple sites for generalizability.
Artificial Intelligence-Based Evaluation of Carotid Artery Compressibility via Point-of-Care Ultrasound in Determining the Return of Spontaneous Circulation During Cardiopulmonary Resuscitation	Park et al.	2024	Point-of-Care Ultrasound (Carotid Artery)	To validate RealCAC-Net for accurate ROSC determination using carotid artery compressibility quantification.. To develop RealCAC-Net for quantifying carotid artery compressibility and determining ROSC during CPR.	RealCAC-Net (TransUNet-based vascular segmentation, CACM algorithm for compression analysis)	Accuracy: 96%; F1-Score: 0.97	Improves real-time ROSC detection during CPR, reducing diagnostic delays and aiding clinical decision-making. Enhances real-time ROSC assessment during CPR; minimizes reliance on subjective manual pulse checks.	Needs high-quality imaging; limited generalizability due to single-center data and device/patient variability. Compression force impacts segmentation accuracy.
Exploring Deep Learning Applications using Ultrasound Single View Cines in Acute Gallbladder Pathologies	Ge et al.	2024	Ultrasound (Single View Cines)	To develop a deep learning model for distinguishing normal, non-urgent cholelithiasis, and acute cholecystitis using single-view ultrasound cines.	InceptionV3 CNN + GRU + MLP	Normal vs Abnormal: Accuracy 91%, Urgent vs Non-Urgent: Accuracy 82%	Promotes rapid triage for gallbladder pathologies; reduces reliance on expert sonographers.	Single-center study with limited external validation; reliance on specific ultrasound machines.
Using AI Segmentation Models to Improve Foreign Body Detection and Triage from Ultrasound Images	Holland et al.	2024	Ultrasound Imaging	To enhance foreign body detection and triage through AI-powered segmentation models for ultrasound images.	U-Net, YOLOv7	YOLOv7 Accuracy 99.1%; U-Net Accuracy: 99.1%	Improves objective triage decisions in remote or resource-limited settings; facilitates better injury detection with minimal human expertise.	Requires extensive labeling and computational resources; limited validation with in vivo data.
Application of a Deep Learning System to Detect Papilledema on Nonmydriatic Ocular Fundus Photographs	Biousse et al.	2024	Nonmydriatic ocular fundus photographs	To evaluate the BONSAI deep learning system for improving papilledema detection on NMFPs in ED settings.	BONSAI-DLS (U-Net for segmentation, EfficientNet-B5 for classification)	Normal vs abnormal AUC: 0.92; Papilledema AUC: 0.97	Reduces diagnostic errors in optic nerve assessments; facilitates rapid ED triage and improved patient outcomes.	Limited generalizability due to lower-quality NMFPs; suboptimal performance on mild papilledema cases.
Artificial Intelligence-Assisted Management of Retinal Detachment from Ultra-Widefield Fundus Images Based on a Weakly-Supervised Approach	Li et al.	2024	Ultra-Widefield Fundus Images	To develop a weakly supervised AI model for localizing and managing retinal detachment regions and providing postural guidance.	Weakly-supervised deep learning (ResNet-101 with AMM)	Accuracy: 0.8651 (AI model), 0.8639 (General ophthalmologist)	Supports efficient RD management and personalized postural guidance; addresses resource shortages in remote settings.	Limited generalization due to single-center data; issues with shallow detachment detection and artifacts in imaging.

Convolutional Neural Network to Classify Infrared Thermal Images of Fractured Wrists in Pediatrics	Shobayo et al.	2024	Infrared Thermal Imaging (IRT)	To develop and evaluate a CNN model for detecting wrist fractures using IRT images in pediatric patients.	Custom 34-layer CNN architecture	Accuracy: 76%, AUC: 0.82	Non-invasive and radiation-free diagnosis; potential to reduce unnecessary X-rays in pediatrics.	Limited sample size; requires further validation with larger datasets.
Prediagnosis recognition of acute ischemic stroke by artificial intelligence from facial images	Wang et al.	2024	Facial Images (2D Texture)	To develop a CNN model for rapid stroke prediagnosis using facial images.	EfficientNet, Xception, ResNet50, VGG19	AUC: 0.91 (cross-validation), 0.82 (independent test)	Provides rapid, noninvasive stroke risk screening; identifies novel auxiliary markers for stroke risk.	Single-center training dataset; requires multicenter validation and broader generalization.
Prediction of Anemia in Real-Time Using a Smartphone Camera Processing Conjunctival Images	Zhao et al.	2024	Smartphone-conjunctiva images (RAW format)	To evaluate a smartphone app's real-time performance in predicting anemia severity by processing conjunctival images.	32-bit high hue ratio algorithm (non-ML) computation of the tissue surface	Accuracy of HBc: 75.4%; AUC (Transfusion thresholds): 0.92 (HBc 7 g/dL), 0.90 (HBc 9 g/dL)	Non-invasive anemia detection; highly accessible for population-level screening in resource-limited settings.	Limited accuracy for mild anemia; potential variability due to lighting and operator performance.
RobOCTNet: Robotics and Deep Learning for Referable Posterior Segment Pathology Detection in an Emergency Department Population	Song et al.	2024	Robotically aligned optical coherence tomography (RAOCT)	To evaluate the diagnostic performance of a robotically aligned OCT system coupled with a deep learning model in detecting referable posterior segment pathology.	InceptionV3-based CNN	Internal AUC: 1.00, External AUC: 0.91; Sensitivity: 95%; Specificity: 76%	Improves emergency triage for ophthalmology referrals; reduces dependence on specialized personnel.	Single-center study; limitations in training data for volumetric images and lack of integration in real-world workflows.
Screening for RV Dysfunction Using Smartphone ECG Analysis App: Validation Study with Acute Pulmonary Embolism Patients	Choi et al.	2024	12-lead printed ECG images	To validate a smartphone app for detecting RV dysfunction in acute PE patients using ECG images.	QCG-RVDys AI-based digital biomarkers	AUC: 0.895 (Identifying RV dysfunction)	Non-invasive, rapid RV dysfunction detection; reduces reliance on echocardiography in resource-limited EDs.	Single-center study; suboptimal PPV; further validation required in diverse populations.
Smartphone pupillometry with machine learning differentiates ischemic from hemorrhagic stroke: A pilot study	Maxin et al.	2024	Smartphone Pupillometry (PLR Metrics)	To evaluate smartphone-based quantitative pupillometry combined with machine learning for differentiating ischemic and hemorrhagic strokes.	Random Forest (best-performing), SVM, KNN, Logistic Regression	Accuracy: 91.5%; ; AUC: 0.917	Offers rapid, non-invasive stroke differentiation, aiding prehospital triage and treatment decisions.	Small sample size, class imbalance, and lack of multi-center validation.
Smartwatch ECG and Artificial Intelligence in Detecting Acute Coronary Syndrome Compared to Traditional 12-lead ECG	Choi et al.	2025	Standard 12-lead ECG vs Smartwatch 9-lead ECG	To evaluate the feasibility and accuracy of a smartwatch-based ECG system combined with AI for diagnosing acute coronary syndrome.	QCG system (CNN-based ECG interpretation)	AUROC: qACS (0.987 smartwatch vs 0.991 standard)	Accessible ACS detection with smartwatches; facilitates early triage in resource-limited settings.	Limited leads on smartwatch ECG; dependency on stable internet connectivity for cloud processing.