

ARTICLE INFORMATION

Received: January, 22, 2025

Revised: March, 14, 2025

Available online: March, 16, 2025

at : <https://e-jurnal.iphorr.com/index.php/minh>

**Performance of artificial intelligence in detecting bone fractures in radiographic results:
A systematic literature review**

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

Abstract

Background: Artificial intelligence (AI), is currently very widely used in various lines of human life, including in the health sector. In radiology, the role of AI has begun to be involved in the interpretation of radiological imaging results with the aim of making radiological diagnosis more efficient.

Purpose: To assess the performance of artificial intelligence in detecting human bone fractures on radiological imaging results.

Method: A systematic literature review method using the PUBMED search application with a publication period of the last 5 years using search *queries* ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("bone fracture detection" OR "fracture detection" OR "bone injury detection") AND ("X-ray" OR "radiograph") AND ("sensitivity" OR "diagnosis speed" OR "cost efficiency").

Results: There are 27 articles showing that the use of AI in the field of radiology has been widely used in various countries, not only for X-ray radiology imaging, but also for CT Scan and MRI imaging applied to various fractures and its speed and effectiveness have also been compared for diagnosis using human radiology personnel.

Conclusion: Artificial intelligence methods have a fairly good ability to identify human fractures in radiological images, which can help doctors avoid misdiagnosis and speed up the time to establish a diagnosis.

Keywords: Artificial Intelligence (AI); CT Scan; MRI; Radiology; X-Ray.

INTRODUCTION

Technological developments in the field of radiology have undergone a significant transformation over the past few decades (van Leeuwen, Schalekamp, Rutten, van Ginneken, & de Rooij, 2021; Zhou, Greenspan, Davatzikos, Duncan, Van Ginneken, Madabhushi, Prince, Rueckert, & Summers, 2021). The digitization of radiology images has enabled clearer visualization and more detailed analysis, thereby improving diagnostic accuracy and operational efficiency in clinical practice (Naik, Rallapalli, Krishna, Vellara, Shetty, Patil, Zeeshan Hameed, Paul, Prabhu, Rai, Chlosta, & Somani, 2022; Martin-Leo, Frederick, Wojciech, Klaus-Robert, Emmanuel, Sonja, Sylvia, Ina, Jörg,

Liron, Hendrik, & Patrick, 2023). Radiology plays a crucial role in medical diagnosis, especially in detecting bone fractures which are one of the most common musculoskeletal injuries (Hussain, Mubeen, Ullah, Shah, Khan, Zahoor, Ullah, Khan, & Sultan, 2022). Precise and rapid identification of bone fractures through radiography is essential for effective treatment planning and optimal patient recovery (Qamar, Evans, Gibney, Redmond, Nasir, Wong, & Nicolaou, 2021; Joshi & Singh, 2020).

Along with these advancements, artificial intelligence (AI) has emerged as a promising revolutionary technology in various medical fields (Rossi, 2023; Kaur, Singla, Nkenyereye, Jha,

Prashar, Joshi, El-Sappagh, Islam, & Riazul Islam, 2020). AI offers great potential to improve diagnostic results through advanced algorithms capable of analyzing large amounts of data with high speed and accuracy (Mirbabaie, Stieglitz, & Frick, 2021; Ahmed, Mohamed, Zeeshan, & Dong, 2020; Acs, Rantalainen, & Hartman, 2020). The application of AI in the interpretation of medical images, particularly radiography, has been the main focus of many studies (Rezazade Mehrizi, van Ooijen, & Homan, 2021; Najjar, 2023; Barragán-Montero, Javaid, Valdés, Nguyen, Desbordes, Macq, Willems, Vandewinckele, Holmström, Löfman, Michiels, Souris, Sterpin, & Lee, 2021). AI's ability to provide consistent analysis and reduce variability caused by human factors makes it a valuable tool in supporting clinical decisions (Abass, Itua, Bature, & Eruaga, 2024; Harry, 2023).

However, despite the enormous potential of AI, the widespread adoption of this technology in clinical practice requires an in-depth evaluation of its performance compared to conventional methods (de Hond, Leeuwenberg, Hooft, Kant, Nijman, van Os, Aardoom, Debray, Schuit, van Smeden, Reitsma, Steyerberg, Chavannes, & Moons, 2022; Kishor & Chakraborty, 2022). This study aims to fill the gap by presenting a systematic review of the existing literature, evaluating the sensitivity and specificity of AI in detecting bone fractures, and analyzing its implications for speed of diagnosis and cost efficiency. As such, the study seeks not only to provide empirical evidence on the effectiveness of AI

diagnostics but also to discuss its impact on clinical practice and the health system as a whole.

RESEARCH METHOD

A systematic literature review (SLR) to evaluate the performance of artificial intelligence (AI) in detecting bone fractures in radiography results. The PICOS (Population, Intervention, Comparison, Outcome, Study Design) framework is used to compile research questions and determine relevant inclusion and exclusion criteria of studies.

The literature search was carried out using the PUBMED database with the following search queries: ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("bone fracture detection" OR "fracture detection" OR "bone injury detection") AND ("X-ray" OR "radiograph") AND ("sensitivity" OR "diagnosis speed" OR "cost efficiency").

Inclusion criteria were articles published in the last 5 years in Scopus indexed international journals, written in the UK, discussing the use of artificial intelligence in detecting human fractures in radiographic images, and reporting the results of the measurements performed. The flow of the study selection process will be visualized using a PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analysis*) diagram. This diagram will include the number of studies identified, screened, evaluated for eligibility, and finally included in the review, further the data is displayed in the form of a table.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan
Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

RESEARCH RESULTS

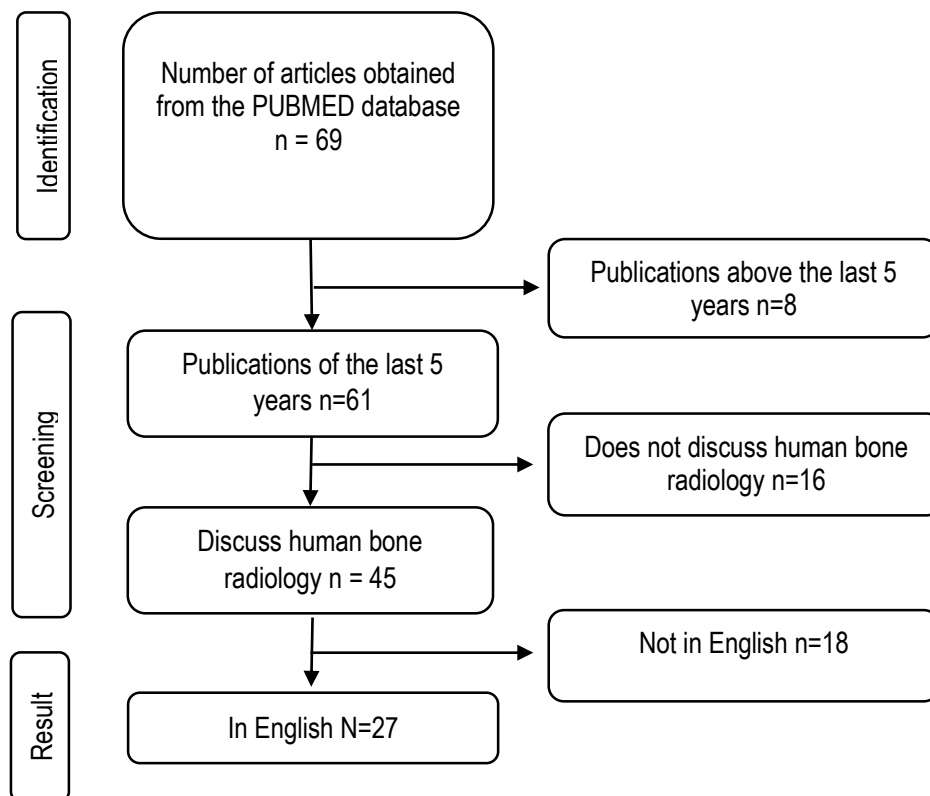


Figure Diagram PRISMA

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Table 1. Brief Description of the Reviewed Articles

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
(Yao et al., 2021) (China)	Developing an <i>AI deep learning</i> system for the diagnosis of rib fractures on chest CT scans	N = 1707 (1507 training set, 100 validation set, 100 testing set), Uni-Center Deep learning CNN F1-Score, negative predictive value (NPV), FROC, precision, recall Checklist Radiologist	A total of 50,170,399 blocks (91,574 fracture blocks; 50,078,825 normal blocks) are labelled for training. F1-score 0.890, <i>precision</i> 0.869, <i>recall</i> 0.913, dan NPV 0.969. The F1-score for junior radiologists increased from 0.796 to 0.925 and for experienced radiologists from 0.889 to 0.970. The recall score of junior radiologists increased from 0.693 to 0.920 and experienced radiologists from 0.853 to 0.972. The average diagnosis time is reduced by 65.3 seconds with the help of this system.	AI performance is on par with experienced radiologists and can automatically detect fractures on CT; reducing diagnosis time and radiologist workload in clinical practice.
(Choi et al., 2022) (South Korea)	To develop and evaluate a deep learning-based artificial intelligence (AI) model for detecting skull fractures in pediatric plain radiographs.	A retrospective multi-center study using data from three hospitals. The dataset included head radiographs and cranial CT scans of children with head trauma. The deep learning model used the YOLOv3 architecture to detect fractures in radiographic images. Validation tests were conducted by comparing the model's results with diagnoses from radiologists and emergency physicians. Evaluation was performed using the area under the receiver operating characteristic curve	The AI model showed an AUROC of 0.922 (95% CI, 0.842–0.969) in the internal test set and 0.870 (95% CI, 0.785–0.930) in the external test set. The sensitivity of the model is : 81.1% (95% CI, 64.8%–92.0%) and specificity of 91.3% (95% CI, 79.2%–97.6%) for internal test sets, 78.9% (95% CI, 54.4%–93.9%) and 88.2% (95% CI, 78.7%–94.4%) for external test sets. With the help of AI, there was a significant increase in AUROC in radiology residents (combined results) and	AI deep learning <i>models</i> improve the ability of inexperienced radiologists and emergency physicians to diagnose skull fractures on regular radiography in children.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
		(AUROC), sensitivity, and specificity. observer studies on an external test set with and without the help of AI.	emergency physicians (combined results) with a difference from AI-assisted readings of 0.094 (95% CI, 0.020–0.168; $P = 0.012$) and 0.069 (95% CI, 0.002–0.136; $P = 0.043$), but not in pediatric radiologists with a difference of 0.008 (95% CI, -0.074–0.090; $P = 0.850$).	
(Rahim et al., 2023) (Iran)	Investigating the Diagnostic Test Accuracy (DTA) of Machine Learning (ML) in the detection of osteoporosis through the image of the hip dual-energy X-ray absorptiometry (DXA).	Systematic Literature Review (SLR),	Combined sensitivity 0.844, with a specificity of 0.781. The diagnostic odds ratio was 18.91, with LR+ 3.7 and LR- 0.22. The combined ROC curve of the bivariate model has an AUC of 0.878.	Machine Learning (ML) can improve hip fracture prediction.
(Shen et al., 2023) (China)	Developing and validating deep learning for the diagnosis of vertebral fractures using the area loss ratio on conventional radiography as a fast and economical diagnostic tool.	N= 11,397 radiography from six Shanghai community centres for internal training and validation. External validation, N=1276 participants from a hospital clinic in Shanghai were recruited. Radiologists analyze all images and use the semi-quantitative tool Genant as	Sistem AI Osteoporotic vertebral fracture Shanghai Hospital (AI_OVF_SH) All fractures: 97.41% accuracy, 84.08% sensitivity, 97.25% specificity. Moderate fracture: sensitivity 88.55% and specificity 99.74%, Severe fractures: sensitivity of 92.30% and specificity of 99.92%.	The AI_OVF_SH system is efficient in helping radiologists and clinicians improve the diagnosis of vertebral fractures.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
		reference data for fracture diagnosis and assessment. Evaluation Accuracy, sensitivity, specificity, positive predictive values, and negative predictive values.	External validation <i>All fractures</i> , accuracy is 96.85%, sensitivity is 83.35 and 94.70%. <i>Moderate fracture</i> : sensitivity 85.61% Specificity 99.85% <i>Severe fractures</i> : sensitivity 93.46% specificity 99.92%.	
(Kaiume et al., 2021) (Japan)	Evaluation of the performance of rib fracture detection on CT images, software based on <i>deep convolutional neural network</i> (DCNN) and comparing it with the diagnosis of rib fractures performed by a doctor	N CT images of 39 patients with thoracic injuries who underwent CT scans. In the images, two radiologists detected 256 rib fractures and set the gold standard. The performance of rib fracture detection by software was compared with two young doctors using the McNemar test and the analysis of the jackknife alternative free-response receiver operating characteristic (JAFROC)	The sensitivity of the DCNN software was significantly higher compared to the two young doctors, A (0.645 VS 0.313; P<0.001), B (0.645 VS 0.258; P<0.001); Based on JAFROC's analysis, the <i>figure-of-merit</i> difference between the results obtained through DCNN software and the two young doctors was 0.057 and 0.071. With a <i>non-inferiority</i> limit of 0.10, the DCNN software is no worse than the rib fracture detection performed by the two young doctors.	Rib fracture detection using the DCNN device can be an alternative to interpretation performed by doctors who are less experienced in image interpretation.
(Anderson et al., 2023) (USA)	Evaluation of <i>FDA-approved deep learning</i> systems can improve the accuracy of fracture diagnosis on musculoskeletal	Medicare claims data were used to analyze trends in musculoskeletal radiograph interpretation by various clinician types. A multiple-reader, multiple-case study was conducted with 24 clinicians from different backgrounds (radiologists, orthopedic	Clinicians with AI and without AI: Accuracy (Without AI, AUC: 0.90 [95% CI 0.89 to 0.92]; With AI, AUC: 0.94 [95% CI 0.93 to 0.95]; Sensitivity (With AI: 90%[95%CI88%to92%]; Without AI: 82%	Clinicians are more accurate in diagnosing fractures with the help of deep learning systems, especially those with limited

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
	radiography by different types of clinicians, especially those with limited training in musculoskeletal imaging.	surgeons, physician assistants, primary care physicians, and emergency medicine physicians). Each clinician reviewed 175 musculoskeletal radiograph cases with and without the assistance of the deep learning system (totaling 4,200 case-clinician pairs per condition). Diagnoses were compared against a gold standard determined by three board-certified orthopedic surgeons or radiologists. Evaluation was performed using the area under the curve (AUC) of the receiver operating characteristic (ROC), sensitivity, and specificity	[95% CI 79% to 84%]]. Specificity (aided: 92% [95% CI 91% to 93%]; unaided: 89% [95% CI 88% to 90%]). Clinicians with limited training showed a high percentage in failing to diagnose fractures on radiographs without AI 20% [95% CI 17% to 24%]; <i>miss rate for radiologists</i> : 14% [95% CI 9% to 19%], and improved after using AI 9% [95% CI 7% to 12%]; <i>miss rate for radiologists</i> : 10% [95% CI 6% to 15%].	training. AI can reduce the number of missed fractures, improving patient care and mobility.
(Beyaz et al., 2020) (China)	Demonstrated <i>deep learning</i> to detect femur neck fractures on frontal pelvic radiography.	Retrospective studies. The study used 234 frontal pelvic X-ray images from 65 patients (32 men, 33 women; mean age 74.9 years; age range, 33 to 89 years), which was expanded to 2106 images. The CNN architecture consists of five blocks with layers such as convolution, batch normalization, and <i>dropout</i> . The training process stopped after 50 epochs with The Optimizer Adam, and the Genetic Algorithm (GA) approach was used to optimize the hyperparameters.	Studies have found that the best performance occurs when images are cropped to a size of 50x50 pixels. Sensitivity is 83% and Specificity reaches 73%. The use of the genetic algorithm (GA) increased sensitivity and specificity by 1.6%. The fracture detection rate reached 83%, and the kappa coefficient reached 55%, indicating an acceptable agreement rate.	Experimental studies use <i>deep learning</i> for the detection of bone fractures on radiography., although the data set is unbalanced, the results are promising. The use of smaller image sizes reduces computational costs and improves evaluation results.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
(Wang et al., 2023) (China)	To assess whether a diagnostic model using <i>deep learning</i> can be used to detect and assess how severe a fatigue fracture of the lower extremities (tibia-fibula) is in plain X-ray photographs.	Retrospective studies N= 1151 X-ray images of tibia-fibula fatigue fractures, N=2842 normal X-ray images. The diagnostic model was created from two clinical centres with <i>deep learning techniques</i> . (A ResNet-50, A triple branch network) Evaluation: Sensitivity, specificity, AUC The performance of the model is evaluated with standard metrics and compared to visual estimates by radiologists.	Detection model in tibiofibular Sensitivity 95.4%/85.5%, Specificity 80.1%/77.0%, AUC 0.965/0.877 on internal and external tests Detection model on the leg bone Sensitivity 96.4%/90.8%, a Specificity 76.0%/66.7%, AUC 0.947/ 0.911. Accuracy Overall diagnostic model 78.5%/62.9% Tibia fibula 74.7%/61.1%	Deep learning <i>models</i> can be used in radiological diagnosis through X-rays to help detect and classify fatigue fractures in the tibia fibula and leg bones.
(Li et al., 2021) (China)	Evaluate how well ResNet50 deep learning can distinguish between benign and malignant vertebral fractures in CT.	N = 433 patients (296 malignant fractures; 137 benign fractures) were selected retrospectively from a database of spinal CT images. Senior radiologists perform visual readings to evaluate six features of the image, while three junior radiologists provide diagnostic predictions. ROI placed on the most abnormal vertebrae The ResNet50 network is used for analysis. Diagnostic performance was evaluated using 10-fold cross-validation.	Radiological accuracy with 5.3 and 1 year experience (99%, 95.2%, and 92.8%). In the ResNet50 analysis, the sensitivity is 90%, the specificity is 79%, and the accuracy is 85%. Each slice of the diagnosis Slices combined for diagnosis per patient Sensitivity 95%, Specificity 80%, Accuracy 88%	In this study, ResNet50 achieved good accuracy in distinguishing between benign and malignant vertebral fractures on CT.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
(Jin et al., 2020) (China)	Developed an automated system to detect and map rib fractures from CT scans, to help radiologists work faster and more accurately in identifying fractures, as well as reduce the workload due to the large number of CT images and patients.	A dataset (RibFrac Dataset) containing 7,473 rib fractures from 900 patients from one service centre. The FracNet model, a <i>deep learning model</i> , was developed for fracture detection and segmentation. The dataset was divided into three parts: 720 patients for training, 60 patients for tuning, and 120 patients for testing. Performance evaluation uses <i>Free Response ROC (FROC)</i> for detection and <i>Intersection-over-Union (IoU)</i> and <i>Dice Coefficient (Dice)</i> for segmentation. Observer studies, including independent human studies and human-computer collaborations, were conducted to compare.	Detection sensitivity 92.9% False positives per scan 5.27 The segmentation Dice value was 71.5% on the test. Human experts showed lower false positives, but their detection sensitivity was lower than neural networks and took longer. Human-computer collaboration results in higher detection sensitivity than diagnosis by humans or computers separately.	FracNet improves the sensitivity of rib fracture detection with shorter clinical time,
(Inoue et al., 2022) (Japan)	Investigating whether automatic localization and classification using CNNs can be applied to pelvic, rib, and spine fractures. The study also examined whether this fracture detection algorithm can assist doctors in diagnosing	Dataset: 7,664 axial slices from whole-body CT scans of 200 trauma patients. Data Annotation: Two orthopedic specialists annotated the images to determine fractures as ground truth. AI Model: An object detection algorithm based on Faster R-CNN was used to localize and classify fractures in CT scans. Model Evaluation: The performance of the CNN model was assessed using sensitivity, precision, and F1-score.	CNN Model Performance: Sensitivity: 0.839 (pelvic), 0.713 (rib), 0.780 (spine). Precision: 0.645 (pelvic), 0.602 (rib), 0.683 (spine). Overall F1-score: 0.711. Improved Diagnosis by Orthopedic Surgeons: Sensitivity significantly increased with CNN assistance, especially for less experienced orthopedic surgeons. Precision remained stable, with no decline in accuracy when using the CNN model. Diagnosis time was significantly reduced with CNN assistance.	The CNN-based deep learning model demonstrated good performance in detecting pelvic, rib, and spine fractures from whole-body CT scans. Additionally, it helped orthopedic surgeons improve diagnostic sensitivity

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
	fractures.	Clinical Study: Three orthopedic surgeons (with 3 to 8 years of experience) reviewed CT scans with and without CNN assistance. Their performance was compared in terms of sensitivity, precision, and diagnosis time.		and reduce interpretation time without compromising precision. Implementing this technology in emergency departments could help reduce missed fractures and enhance the quality of care for trauma patients.
(Zhang et al., 2021) (China)	Investigating the impact of <i>deep learning</i> (DL) on the accuracy and efficiency of radiologists in detecting rib fractures on CT.	The study involved 198 chest trauma patients who underwent thin-cut CT scans. Two radiologists (R1 and R2) read CT images in three sessions: unassisted reading (S1), assisted by <i>deep learning</i> (DL) as a co-reader (S2), and DL as a second reader (S3). Detected fractures and total read times are documented. The reference standard for fractures is set by a panel of experts. Sensitivity and <i>false positives per scan</i> were compared between the three sessions.	Sensitivity increased when using <i>deep learning</i> (DL): from 82.8% (S1) to 88.9% (S2) and 88.7% (S3) for R1, as well as from 83.9% (S1) to 88.7% (S2) and 88.8% (S3) for R2. The sensitivity of S2 and S3 is significantly higher than that of S1 ($p < 0.05$), but there is no difference between S2 and S3. The number of false positives per scan was no different for R1 ($P = 0.24$), but lower for S2 and S3 than for S1 for R2 ($P < 0.05$). Reading time is reduced by 36% (R1) and 34% (R2) in S2 compared to S1.	The use of <i>deep learning</i> as a shared reader increases accuracy and efficiency in detecting rib fractures.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
(Yoon et al., 2021) (Taiwan)	Develop and validate <i>deep convolutional neural networks</i> (DCNN) in detecting scaphoid fractures, both visible and hidden from radiographic images.	The study used radiographic datasets of patients who may have had a <i>scaphoid fracture</i> between December 2019 and January 2001 from Chang Gung Memorial Hospital (Taipei, Taiwan) and Michigan Medicine (Ann Arbor). The data set is randomly divided into training, validation, and testing data sets. The images were processed through a detection model to cut around the scaphoid and train the DCNN model based on the EfficientNet B3 architecture. Data analysis was conducted from January to October 2021.	The results showed that of the 11,838 radiography analyzed, 41.5% had a scaphoid fracture and 58.5% did not. In the trial, the first DCNN had a sensitivity of 87.1% and specificity of 92.1% AUROC 0.955 to distinguish scaphoid fractures. The gradient-weighted class activation mapping matches the fracture location. The second DCNN has a sensitivity of 79.0% and a specificity of 71.6% with an AUROC of 0.810 in checking for negative cases from the first model. The two-stage examination identified 90.9% of the 22 cases of hidden fractures.	The DCNN model can identify scaphoid fractures, indicating that such a model can help detect hidden scaphoid fractures that are invisible to humans and can reliably detect fractures from other small bones.
(Niiya et al., 2022) (Japan)	Evaluating the clinical effectiveness of deep learning-based auto-detection algorithms for rib fractures in CT of high-energy trauma patients	The study aims to evaluate how well the artificial intelligence model detects vertebral fractures from lateral radiographs of the spine, compared to the results of human assessment based on the degree of Genant fractures.	The sensitivity of the algorithm is 89.8%, and the number of false positives per case is 2.5. After additional learning, the sensitivity increased to 93.5%, with a false positive count of 1.9 per case.	The sensitivity of the algorithm used in this study was sufficient to aid the rapid detection of rib fractures in the validated set of CT images evaluated.
(Li et al., 2021) (China)	AI evaluation on vertebral osteoporosis fracture detection	The study conducted a retrospective evaluation on 941 patients with an average age of 76 years who underwent a lateral radiography examination of the spine. Artificial intelligence models using such	The AI model with the ensemble method achieved high accuracy (93%), sensitivity (91%), and specificity (93%) in detecting lumbar vertebral fractures, with strong interobserver reliability. It performed well in grading fractures but showed lower specificity (91%) in	Artificial intelligence models can detect vertebral fractures on lateral radiography of the spine with high

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
		datasets are trained, validated, and tested. Human interpretation is compared to CT or MRI interpretation by orthopaedic surgeons and radiologists. Fractures are classified based on the Genant classification. Assessment: Sensitivity, specificity, accuracy.	distinguishing osteoporotic lumbar vertebrae. Sensitivity was higher for osteoporotic fractures (97%) than nonosteoporotic ones (83%). The model outperformed in detecting lumbar fractures over thoracic fractures, with external validation showing 89% accuracy, 83% sensitivity, and 95% specificity.	accuracy, sensitivity, and specificity, especially for osteoporosis fractures of the lumbar spine. The application of this model has the potential to improve the efficiency of vertebral fracture diagnosis.
(Lopez-Melia et al., 2024) (Switzerland)	Reviewing studies on <i>deep learning</i> (DL) models in detecting rib fractures on CT data,	Meta-analysis studies and research articles taken from PubMed, Embase, and Web of Science in April 2023. Studies were included if the DL model was used to detect rib fractures and trained with human CT data. Risk of Bias (ROB) assessment uses <i>the Quality Assessment of Diagnostic Accuracy Studies tool</i> . The performance of the model was meta-analyzed with <i>forest plots</i> .	ROB mostly comes from patient selection and improper data. Studies with low ROB showed better F1 scores in the detection of acute rib fractures using the DL model. Of the 27 studies selected, about 75% of them were at risk of bias because they did not report on patient selection criteria or used CT scans with a slice thickness of 5 mm. Studies with a low risk of bias show high sensitivity, precision, and F1 scores.	ROB in many studies is caused by patient selection and imprecise data. Studies with low ROB showed better performance in detecting acute rib fractures using the DL model. This review serves as a reference for the status of rib fracture detection with the DL model, providing benefits for future studies.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
(Weikert et al., 2020) (Switzerland)	Evaluating the performance of deep learning-based algorithms in detecting acute and chronic rib fractures on CT of body trauma.	Retrospective N=551 (positive 159; negative 352). Evaluation: sensitivity, specificity.	Sensitivity 87.4% Specificity 91.5%	Deep learning <i>algorithms</i> are effective in detecting rib fractures on traumatic CT with few <i>false positives</i> per case. It has potential as a screening tool to avoid <i>false-negative radiological reports</i> .
(Zhou et al., 2020) (China)	Evaluate the performance of a <i>convolutional neural network</i> (CNN) model that automatically detects and classifies rib fractures and generates a structured report from <i>computed tomography</i> (CT) images.	Involving 1079 patients (median age 55 years, male 718) from three hospitals in the period from January 2011 to January 2019. Patients are divided into training and validation sets with various image parameters. Fractures are classified and their locations are automatically identified in a structured report. The evaluation uses <i>precision</i> , <i>recall</i> , and F1-Score, as well as a comparison of diagnosis time and efficiency with experienced radiologists.	CNN Model Performance: F1-score for fresh fracture detection: 0.849, healing fractures: 0.856, old fractures: 0.770. The model demonstrated stable performance across five different validation sets (all F1-scores > 0.8, except for the low-resolution [512×512 pixels] set with an F1-score of 0.757). Improvement in Radiologists' Diagnosis: Precision increased from 80.3% to 91.1% with AI assistance. Sensitivity increased from 62.4% to 86.3% after using AI. Diagnosis time was reduced by an average of 73.9 seconds with AI assistance.	CNN models to automatically detect rib fractures can improve diagnostic efficiency, and reduce diagnosis time, and radiologist workload.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
(Achar et al., 2023) (Italy)	Developed a monitored <i>deep learning</i> tool to create CT-like images and fracture probability maps from <i>ultrashort echo time</i> (UTE) MRI to evaluate isthmic spondylolysis.	The research uses the preparation of the cadaver spine and the <i>deep learning</i> tools developed. Comparisons were made between UTE MRI modified for CT and CT-like appearance, as well as between CT and CT-like images using quantitative metrics such as contrast-to-noise ratio (CNR), mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM).	Quantitatively, the mean CNR of bones to deformity-filled tissue was 35, 97, and 146 for UTE MRI, CT-like, and CT images, respectively, significantly higher for CT-like than UTE MRI images. For image similarity metrics using CT images as a reference, CT-like images provide a much lower mean MSE (0.038 vs. 0.0528), a higher mean PSNR (28.6 vs. 16.5), and a higher SSIM (0.73 vs. 0.68) compared to UTE MRI images.	This concept study shows that the use of UTE MRI and deep learning tools can be very useful for the evaluation of isthmic spondylolysis.
(Li et al., 2023) (China)	Develop and evaluate radiomic-based CT with machine learning to detect <i>occult vertebral fractures</i> (OVFs).	The study involved 128 vertebrae, including 64 with MRI-confirmed OVFs, and 64 control vertebrae from 57 patients with chest/abdominal CT. Radiomic features were extracted and three machine learning algorithms were used to build the model. The ROC curve is used to verify the performance of the model.	For mid-axial CT, 6 radiomic parameters were used with a logistic regression (LR) algorithm, resulting in a training AUC of 0.682 and a validation of 0.775. For <i>mid-sagittal</i> CT, 5 parameters were used with logistic regression (LR), resulting in a training AUC of 0.832 and validation of 0.882. The sagittal LR model has an accuracy of 0.846, a sensitivity of 0.846, and a specificity of 0.846.	Machine learning with CT radiomic features can detect OVFs, especially sagittal LR models, suggesting combinations with <i>deep learning</i> for the automatic recognition of OVFs and reducing secondary injuries.
(Jiménez-Sánchez et al., 2020) (Spain)	Demonstrated whether a fully automated, deep learning-based computer-aided	The research methods included the collection of 1347 radiographic studies annotated by radiologists and trauma surgeons, as well as the distribution of datasets for model building and evaluation.	The results show that the proposed CAD can classify radiographs with high accuracy, achieving an F1-Score of 87% and an AUC of 0.95. In addition, the model successfully identifies fractures with a high degree of accuracy and provides satisfactory results in	The proposed CAD can classify proximal femur fractures with high accuracy, comparable to experts,

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
	diagnosis (CAD) tool can identify and classify proximal femur fractures on X-ray images according to the <i>Arbeitsgemeinschaft für Osteosynthesefragen</i> (AO) classification.	ResNet-50 and AlexNet deep learning models are used for fracture classification and localization, Evaluation using accuracy, precision, recall, and F1-score	localization.	and the best performance to date. The additional localization model is also accurate in predicting regions of interest in radiographs.
(Wu et al., 2023) (China)	Develop and verify the performance of artificial intelligence (AI) models in detecting rib fractures in chest radiography from various centres.	The research method used a <i>convolutional neural network</i> (CNN) with a dataset of 1080 rib fracture radiography which was divided into training and testing sets. CNN's object detection model, YOLOv3, was evaluated using <i>receiver operating characteristics</i> (ROC) and <i>free-response ROC</i> (FROC). Independent validation was carried out to compare the performance of the CNN model with the diagnostic efficiency of radiologists.	The results showed that the sensitivity and precision of the CNN model were 92.0% and 91.1%, respectively, for the training and testing sets. In the combined test, the accuracy, sensitivity, specificity, and area under the ROC curve were 85.1%, 93.2%, 79.4%, and 0.92. The CNN model also showed higher accuracy and sensitivity or close to the radiologist's readings in independent validation.	The CNN model based on YOLOv3 is sensitive in detecting rib fractures on chest radiography and has the potential to reduce missed diagnoses and ease the workload of radiologists.
(Zhang et al., 2023) (China)	Design and validate a multi-stage <i>deep learning</i> (DL) system for the automatic	This research method involved 1,217 CT <i>Scan images</i> of patients (January 2015 - December 2019). Fractures are marked and classified by two radiology residents,	The results showed the average age of patients was 61.87 years, with 760 women and 457 men. At the vertebral level, the sensitivity of fracture detection reached 95.23% in the test dataset, with an accuracy of	In conclusion, the multi-stage AI-DL system is effective in automatically detecting

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
	detection, localization, and classification of thoracolumbar vertebrae fractures according to The AOSpine thoracolumbar spine injury classification system (AO classification) using CT Scan.	with verification by a senior radiologist if there are any differences. Use U-net and U-GCN for localization and classification, then classification network to detect fractures, and double-branch output network to determine the AO type.	97.93% and a specificity of 98.35%. For the classification of vertebral body fractures, the accuracy of the balance reached 79.56%, with AUC of 0.904 for A1, 0.945 for A2, 0.878 for A3, and 0.942 for A4, respectively.	and classifying vertebral body fractures in the thoracolumbar spine on CT images with high accuracy according to AO classification.
(Ukai et al., 2021) (Japan)	Developing automated methods for detecting pelvic fractures on 3-dimensional CT-scan (3D-CT)	Using <i>deep convolutional neural networks</i> (DCNNs) for the detection of pelvic fractures on 3D-CT images. However, to reduce complexity, multiple real-time 2D object detection systems (YOLOv3 model) were evaluated in parallel with the reconstructed 2D slab images from 3D-CT. 2D fracture candidates of various orientations are then detected simultaneously to obtain a 3D fracture region.	Validation of the method on 93 subjects with bone fractures resulted in an <i>Area under the curve</i> (AUC) of 0.824, with a recall of 0.805 and a precision of 0.907. This method was then applied to 112 subjects without bone fractures, successfully detecting the absence of fractures in all but 4 non-fractured subjects (96.4%).	The proposed method is able to automatically detect hip fractures with a high degree of accuracy, providing hope in reducing the risk of post-fracture mortality in the elderly.
(Polzer et al., 2024) (Germany)	Develop and test neural networks for automatic detection	Study methods: Involving 257 CT patients with Institutional Review Board (IRB) approval, including 463 vertebral body	The average age of patients was 68 ± 14 years, with 140 female patients. This tissue can distinguish between fractured and non-fractured vertebral bodies	The developed artificial nerve network can automatically

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
	and stability analysis of vertebral body fractures on computed tomography (CT).	fractures and 1883 without, with 190 unstable fractures. Two readers identify the fracture and assess its stability. A combination of <i>Hierarchical Convolutional Neural Network</i> (hNet) and <i>fracture Classification Network</i> (fNet) is used to build a neural network for automatic detection and stability analysis of vertebral fractures on CT. Two final test settings were chosen: one with C1/2 levels included and one in which they were excluded.	with a sensitivity of 75.8% and a specificity of 80.3%. In addition, the tissue determines the stability of the vertebral body with a sensitivity of 88.4% and a specificity of 80.3%. The AUC is 87% and 91% for fracture detection and stability analysis. Deep tissue sensitivity indicating the presence of at least one fracture/unstable fracture in the entire spine reached values of 78.7% and 97.2% when C1/2 was excluded.	detect vertebral body fractures and evaluate their stability simultaneously with high diagnostic performance.
(Monchka et al., 2021) (Canada)	Evaluate <i>convolutional neural networks</i> (CNNs) to compare <i>vertebral fracture assessment</i> (VFA), <i>General Electric</i> (GE) <i>dual-energy</i> (DE) and GE <i>single-energy</i> (SE) scanning modes.	In this study, 12,742 VFA GE images from the Manitoba Bone Mineral Density Program between 2010 and 2017 were exported in DE and SE modes. VFAs are classified by imaging specialists using a modified algorithm-based qualitative method (mABQ) as cracks are present or not. The data was randomly divided into training sets (60%), validation (10%), and tests (30%). The three CNN models were built by training separately on DE only, and SE only, and the combined dataset consisted of both SE and DE VFAs. All separately trained CNN models were evaluated against both the SE and DE test	The DE scan used for training and evaluation (DE/DE) achieved a sensitivity of 87.9%, a specificity of 87.4%, and an area under the receiver operating characteristic curve (AUC) of 0.94. The SE scan used for training and evaluation (SE/SE) achieved a sensitivity of 78.6%, specificity of 90.6%, AUC = 0.92. In contrast, CNNs performed poorly when evaluated on different scan modes of their training set (AUC = 0.58). However, composite CNNs that train simulants simultaneously on SE and DE VFA provide comparable performance to DE/DE (sensitivity 82.4%, specificity 94.3%, AUC = 0.95); and provide improved performance over SE/SE (sensitivity 82.2%, specificity 92.3%, AUC = 0.94).	CNNs are highly sensitive to the scanning mode in the identification of vertebral fractures. Training of CNNs on the combined dataset of GE DE and GE SE VFAs allows both scan modes, facilitating the development of the manufacturer's independent machine-learning model for vertebral fracture

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

(Author, Year) (Country)	Purpose	Method	Result	Conclusion
		datasets.	Positive predictive values were higher with composite CNNs compared to models trained only on DE (74.5% vs 58.7%) or SE VFA (68.6% vs 62.9%).	detection.
(Ruitenbeek et al., 2024) (Netherlands)	Evaluate the effects of artificial intelligence (AI) applications in detecting posterior cervical fractures on CT scans and reduce time to diagnosis by prioritizing suspected cases.	The research methods included the evaluation of regulatory-approved AI software to prioritize posterior cervical fractures on CT, using historical controls and prospective evaluations. A total of 2973 patients were analyzed using AI, including 2036 retrospective cases and 938 prospective cases.	2973 (mean age, 55.4 ± 19.7 [standard deviation]; 1857 males) patients were analyzed by AI, including 2036 retrospective and 938 prospective cases. The overall prevalence of cervical vertebral fractures is 7.6%. Detection and notification time (DNT) was 18% (5 minutes) shorter in the prospective group. In a positive scan for cervical vertebral fracture according to the reference standard, the DNT was 46% (16 minutes) shorter in the prospective group. The overall sensitivity of the AI application is 89.8% (95% CI: 84.2-94.0%), the specificity is 95.3% (95% CI: 94.2-96.2%), and the diagnostic accuracy is 94.8% (95% CI: 93.8-95.8%). The negative predictive value was 99.1% (95% CI: 98.5-99.4%) and the positive predictive value was 63.0% (95% CI: 58.0-67.8%). 22 fractures were missed by the AI, 5 of which required stabilization therapy.	Although these AI applications show high diagnostic accuracy, there are clinically important cases that are missed.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

DISCUSSION

Various studies on the use of artificial intelligence in detecting fractures have been carried out a lot, from the results of mapping the results of literature searches, it can be seen that the research stage is not only at the trial stage, but there is research that has led to the evaluation and development stage of existing AI systems, as well as evaluating and comparing with the conventional radiological diagnosis of fractures using human radiologists. In terms of radiology imaging sources, the use of AI in the field of radiology can be divided into 3 clusters, namely the use of AI in interpreting simple radiological images (X-ray), the use of AI in identifying images from CT scans, and the use of AI in identifying fractures using magnetic resonance imaging (MRI). The identified bone fractures are rib fractures, vertebral bones, pelvic bones, femur bones, bones in the shoulder joint and scaphoid bones.

From the results of the literature search, it was found that the research method used consisted of retrospective research used to evaluate AI performance, while research that was developmental on AI models used prospective methods. In addition, systematic literature review (SLR) studies and meta-analyses were also found which were also used to evaluate the use of AI in detecting fractures radiologically. As a standard of truth in many studies, it was found that the results of fracture radiology readings carried out by several experienced radiology specialists or orthopedic specialists were used as a reference standard for the truth of fracture diagnosis by radiology specialists to be compared with the results of AI readings, readings by new radiology specialists, emergency room doctors, and internal medicine specialists.

The results of this literature study show that the artificial intelligence models currently used in identifying human bone fractures in radiological images are convolutional neural network (CNN) and deep convolutional neural network (DCNN) models. CNN is a type of deep machine learning algorithm that mimics human neural networks, this algorithm is inspired by the structure of the human visual cortex in recognizing an object or image (Kattenborn, Leitloff, Schiefer, & Hinz 2021; Li, Liu, Yang, Peng, & Zhou, 2022). CNN works using a series of layers in the form of convolutional layers, pooling layers, and

fully connected layers (Basha, Dubey, Pulabaigari, & Mukherjee, 2020; Zafar, Aamir, Mohd-Nawi, Arshad, Riaz, Alruban, Dutta, & Almotairi, 2022; Taye, 2023). Convolutional layer to extract image features where each convolutional layer applies a filter to the input image and the filter output is a feature map (Liu, Pu, & Sun, 2021; Tian, 2020). The feature map is then passed through a series of connected layers, which reduce their size and dimensions (Gholamalinezhad & Khosravi, 2020; Li, Liu, Yang, Peng, & Zhou, 2022). Finally, the output of the unification layer is fed into the fully connected layer, which results in the final output of the network (Yang, Zhang, Li, & Xie, 2021; Liu et al., 2021). The convolutional layer is a key layer of CNN, where filters are applied to the input image to extract features such as edges, textures and shapes (Wang, Turko, Shaikh, Park, Das, Hohman, Kahng, & Polo Chau, 2021; Kanwal, Ahmad, Khan, Abbasi, & Li, 2020). The output of the convolutional layer is then passed through the collection layer, which is used to sample the feature map, reducing spatial dimensions while retaining the most important information. The output of the unification layer is then passed through one or more fully connected layers, which are used to make predictions or classify images.

CNN deep learning methods are trained using large datasets of labeled images, where the network learns to recognize patterns, and features associated with a particular object or class. The training process for CNN involves steps such as data preparation, where the training images are preprocessed to ensure they are all in the same format and size, The next step is the loss function, which is used to measure how well the CNN performs on the training data, where the loss function is calculated by taking the difference between the predicted labels and the actual labels of the training images. The next process is optimization which is used to update the CNN weights to minimize the loss function (optimizer), and backpropagation is a technique used to calculate the gradient of the loss function regarding the CNN weights, this gradient is then used to update the CNN weights using optimization (backpropagation). The AI models developed are: A ResNet-50, a three-branch network (TBSNet), FracNet, EfficientNet B3. YOLOv3, Hierarchical

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan
Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Convolutional Neural Network (hNet) and Fracture Classification Network (fNet).

Based on the data above, it can be seen that CNN efficiency can be evaluated in popular metrics such as sensitivity, specificity, accuracy, calculation of area under the curve (AUC), area under the receiver operating characteristic (AUROC), F1-Score.

In this study, a combined sensitivity value of 88.89%, a combined specificity value of 85.48%, a combined accuracy value of 86.94%, and an average AUC of 88%. Based on previous research, it shows that the AI model in interpreting fractures on radiological images shows an area under receiver operating characteristics (AUROC) value of 92.2% on the internal test set (sensitivity 81.1% and specificity 91.3%) and AUROC 87% on the external test (sensitivity 78.9%, specificity 88.2%). A significant increase in the AUROC value occurred in radiology residents and emergency room doctors with a combined AUROC increase of 0.094. (Choi, Cho, Ha, Lee, Koh, Seo, Choi, Cheon, Phi, Kim, Yang, & Kim, 2022).

Another study that saw DCNN's ability to detect scaphoid fractures had an AUROC score of 0.955 (sensitivity of 87.1% and specificity of 92.1%) on DCNN 1 and an AUROC score of 0.810 (sensitivity of 79.0% and specificity of 71.6%) on DCNN2 (Yoon, Lee, Kane, Kuo, Lin, & Chung, 2021).

In previous studies, the F1-Score value of using AI to detect fractures in radiological images showed a figure of 0.89, with a precision value of 86.9% and a recall value of 93.1%. Meanwhile, the interpretation of fractures in radiological images by junior radiologists assisted by AI experienced an increase in F1-Score from 78.9% to 92.5%, while the use of AI by experienced radiologists showed an increase in F1-Score from 85.3% to 97.2% with an average reduction in diagnosis time of 65.3 seconds (Yao, Guan, Song, Tan, Wang, Jin, Chen, Wang, & Zhang, 2021).

In line with other studies, it shows that the use of the AI CNN model in detecting fractures in radiological images shows an F1-Score value of 78.6% for detecting pelvic fractures, 64.8% for detecting spinal fractures and 71.1% for detecting rib fractures, research conducted by Inoue shows an F1-Score value of 84.9% (precision 85.6%; recall 77%) (Inoue, Maki, Furuya, Mikami, Mizutani,

Takada, Okimatsu, Yunde, Miura, Shiratani, Nagashima, Maruyama, Shiga, Inage, Orita, Eguchi, & Ohtori; 2022; Jiménez-Sánchez, Kazi, Albarqouni, Kirchhoff, Biberthaler, Navab, Kirchhoff, & Mateus, 2020).

When viewed from the overall score obtained, it can be seen that AI can assist clinicians in establishing the diagnosis of bone fractures both conventionally by using ordinary X-ray photos and by using more specific imaging such as CT Scan and MRI.

CONCLUSION

Artificial intelligence methods have a fairly good ability to identify human bone fractures on radiological images, which can help clinicians avoid misdiagnosis and speed up the time for diagnosis.

REFERENCES

- Achar, S., Hwang, D., Finkenstaedt, T., Malis, V., & Bae, W. C. (2023). Deep-Learning-Aided Evaluation of Spondylolysis Imaged with Ultrashort Echo Time Magnetic Resonance Imaging. *Sensors*, 23(18).
- Acs, B., Rantalainen, M., & Hartman, J. (2020). Artificial intelligence as the next step towards precision pathology. *Journal of Internal Medicine*, 288(1), 62–81.
- Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. Q. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, 2020, 1–35.
- Anderson, P. G., Baum, G. L., Keathley, N., Sicular, S., Venkatesh, S., Sharma, A., Daluiski, A., Potter, H., Hotchkiss, R., Lindsey, R. V., & Jones, R. M. (2023). Deep Learning Assistance Closes the Accuracy Gap in Fracture Detection Across Clinician Types. *Clinical Orthopaedics and Related Research*, 481(3), 580–588.
- Barragán-Montero, A., Javaid, U., Valdés, G., Nguyen, D., Desbordes, P., Macq, B., Willems, S., Vandewinckele, L., Holmström, M., Löfman, F., Michiels, S., Souris, K., Sterpin, E., & Lee, J. A. (2021). Artificial intelligence and machine learning for medical imaging: A technology review. *Physica Medica*, 83, 242–256.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan
Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

- Basha, S. H. S., Dubey, S. R., Pulabaigari, V., & Mukherjee, S. (2020). Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, 378, 112–119.
- Beyaz, S., Açici, K., & Sümer, E. (2020). Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches. *Joint Diseases and Related Surgery*, 31(2), 175–183.
- Choi, J. W., Cho, Y. J., Ha, J. Y., Lee, Y. Y., Koh, S. Y., Seo, J. Y., Choi, Y. H., Cheon, J. E., Phi, J. H., Kim, I., Yang, J., & Kim, W. S. (2022). Deep Learning-Assisted Diagnosis of Pediatric Skull Fractures on Plain Radiographs. *Korean Journal of Radiology*, 23(3), 343–354.
- de Hond, A. A. H., Leeuwenberg, A. M., Hooft, L., Kant, I. M. J., Nijman, S. W. J., van Os, H. J. A., Aardoom, J. J., Debray, T. P. A., Schuit, E., van Smeden, M., Reitsma, J. B., Steyerberg, E. W., Chavannes, N. H., & Moons, K. G. M. (2022). Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *Npj Digital Medicine*, 5(1).
- Gholamalinezhad, H., & Khosravi, H. (2020). Pooling methods in deep neural networks, a review. *arXiv preprint arXiv:2009.07485*.
- Harry, A. (2023). Role of AI in Education. *Interdisciplinary Journal and Humanity (INJURY)*, 2(3), 260–268.
- Hussain, S., Mubeen, I., Ullah, N., Shah, S. S. U. D., Khan, B. A., Zahoor, M., Ullah, R., Khan, F. A., & Sultan, M. A. (2022). Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review. *BioMed Research International*. 2022(1)
- Inoue, T., Maki, S., Furuya, T., Mikami, Y., Mizutani, M., Takada, I., Okimatsu, S., Yunde, A., Miura, M., Shiratani, Y., Nagashima, Y., Maruyama, J., Shiga, Y., Inage, K., Orita, S., Eguchi, Y., & Ohtori, S. (2022). Automated fracture screening using an object detection algorithm on whole-body trauma computed tomography. *In Scientific Reports* (Vol. 12, Issue 1).
- Jiménez-Sánchez, A., Kazi, A., Albarqouni, S., Kirchhoff, C., Biberthaler, P., Navab, N., Kirchhoff, S., & Mateus, D. (2020). Precise proximal femur fracture classification for interactive training and surgical planning. *International Journal of Computer Assisted Radiology and Surgery*, 15(5), 847–857.
- Jin, L., Yang, J., Kuang, K., Ni, B., Gao, Y., Sun, Y., Gao, P., Ma, W., Tan, M., Kang, H., Chen, J., & Li, M. (2020). Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet. *In EBioMedicine* (Vol. 62).
- Joshi, D., & Singh, T. P. (2020). A survey of fracture detection techniques in bone X-ray images. *Artificial Intelligence Review*, 53(6), 4475–4517.
- Kaiume, M., Suzuki, S., Yasaka, K., Sugawara, H., Shen, Y., Katada, Y., Ishikawa, T., Fukui, R., Abe, O., & Lawal, I. (2021). Rib fracture detection in computed tomography images using deep convolutional neural networks. *Medicine (United States)*, 100(20), E26024.
- Kanwal, K., Ahmad, K. T., Khan, R., Abbasi, A. T., & Li, J. (2020). Deep learning using symmetry, FAST scores, shape-based filtering and spatial mapping integrated with CNN for large scale image retrieval. *Symmetry*, 12(4), 612.
- Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 24–49.
- Kaur, S., Singla, J., Nkenyereye, L., Jha, S., Prashar, D., Joshi, G. P., El-Sappagh, S., Islam, M. S., & Riazul Islam, S. M. (2020). Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives. *IEEE Access*, 8, 228049–228069.
- Kishor, A., & Chakraborty, C. (2022). Artificial Intelligence and Internet of Things Based Healthcare 4.0 Monitoring System. *Wireless Personal Communications*, 127(2), 1615–1631.
- Li, W. G., Zeng, R., Lu, Y., Li, W. X., Wang, T. T., Lin, H., Peng, Y., & Gong, L. G. (2023). The value of radiomics-based CT combined with machine learning in the diagnosis of occult vertebral fractures. *BMC Musculoskeletal Disorders*, 24(1), 1–9.
- Li, Y. C., Chen, H. H., Horng-Shing Lu, H., Hondar Wu, H. T., Chang, M. C., & Chou, P. H. (2021). Can a Deep-learning Model for the Automated Detection of Vertebral Fractures Approach the Performance Level of Human Subspecialists?. *Clinical Orthopaedics and Related Research*, 479(7), 1598–1612.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan

Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

- Li, Y., Zhang, Y., Zhang, E., Chen, Y., Wang, Q., Liu, K., Yu, H. J., Yuan, H., Lang, N., & Su, M. Y. (2021). Differential diagnosis of benign and malignant vertebral fracture on CT using deep learning. *European Radiology*, 31(12), 9612–9619.
- Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 6999–7019.
- Liu, Y., Pu, H., & Sun, D. W. (2021). Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. *Trends in Food Science and Technology*, 113, 193–204.
- Lopez-Melia, M., Magnin, V., Marchand-Maillet, S., & Grabherr, S. (2024). Deep learning for acute rib fracture detection in CT data: a systematic review and meta-analysis. *The British Journal of Radiology*, 97(1155), 535–543.
- Martin-Leo, H., Frederick, K., Wojciech, S., Klaus-Robert, M., Emmanuel, D., Sonja, S., Sylvia, H., Ina, K., Jörg, A., Liron, P., Hendrik, S., & Patrick, W. (2023). Imaging bridges pathology and radiology. *Journal of Pathology Informatics*, 14, 1–4.
- Mirbabaie, M., Stieglitz, S., & Frick, N. R. J. (2021). Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction. In *Health and Technology* (Vol. 11, Issue 4).
- Monchka, B. A., Kimelman, D., Lix, L. M., & Leslie, W. D. (2021). Feasibility of a generalized convolutional neural network for automated identification of vertebral compression fractures: The Manitoba Bone Mineral Density Registry. *Bone*, 150, 116017.
- Naik, N., Rallapalli, Y., Krishna, M., Vellara, A. S., Shetty, D. K., Patil, V., Zeeshan Hameed, B. M., Paul, R., Prabhu, N., Rai, B. P., Chłosta, P., & Somani, B. K. (2022). Demystifying the Advancements of Big Data Analytics in Medical Diagnosis: An Overview. *Engineered Science*, 19, 42–58.
- Najjar, R. (2023). Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics*, 13(17).
- Niia, A., Murakami, K., Kobayashi, R., Sekimoto, A., Saeki, M., Toyofuku, K., Kato, M., Shinjo, H., Ito, Y., Takei, M., Murata, C., & Ohgiya, Y. (2022). Development of an artificial intelligence-assisted computed tomography diagnosis technology for rib fracture and evaluation of its clinical usefulness. *Scientific Reports*, 12(1), 1–7.
- Polzer, C., Yilmaz, E., Meyer, C., Jang, H., Jansen, O., Lorenz, C., Bürger, C., Glüer, C. C., & Sedaghat, S. (2024). AI-based automated detection and stability analysis of traumatic vertebral body fractures on computed tomography. *European Journal of Radiology*, 173, 111364.
- Qamar, S. R., Evans, D., Gibney, B., Redmond, C. E., Nasir, M. U., Wong, K., & Nicolaou, S. (2021). Emergent Comprehensive Imaging of the Major Trauma Patient: A New Paradigm for Improved Clinical Decision-Making. *Canadian Association of Radiologists Journal*, 72(2), 293–310.
- Rahim, F., Zaki Zadeh, A., Javanmardi, P., Emmanuel Komolafe, T., Khalafi, M., Arjomandi, A., Ghofrani, H. A., & Shirbandi, K. (2023). Machine learning algorithms for diagnosis of hip bone osteoporosis: a systematic review and meta-analysis study. *BioMedical Engineering Online*, 22(1), 1–15.
- Rezazade Mehrizi, M. H., van Ooijen, P., & Homan, M. (2021). Applications of artificial intelligence (AI) in diagnostic radiology: a technography study. *European Radiology*, 31(4), 1805–1811.
- Rossi, N. (2023). Applications of Artificial Intelligence in Healthcare. *Chimica Oggi/Chemistry Today*, 41(2), 49–51.
- Ruitenbeek, H. C., Oei, E. H. G., Schmahl, B. L., Bos, E. M., Verdonschot, R. J. C. G., & Visser, J. J. (2024). Towards clinical implementation of an AI-algorithm for detection of cervical spine fractures on computed tomography. *European Journal of Radiology*, 173, 111375.
- Shen, L., Gao, C., Hu, S., Kang, D., Zhang, Z., Xia, D., Xu, Y., Xiang, S., Zhu, Q., Xu, G. W., Tang, F., Yue, H., Yu, W., & Zhang, Z. (2023). Using Artificial Intelligence to Diagnose Osteoporotic Vertebral Fractures on Plain Radiographs. *Journal of Bone and Mineral Research*, 38(9), 1278–1287.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan
Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>

Performance of artificial intelligence in detecting bone fractures in radiographic results: A systematic literature review

- Taye, M. M. (2023). Theoretical Understanding of Convolutional Neural Network. *Mdpi Ag*, 11(3), 52.
- Abass, T., Itua, E. O., Bature, T., & Eruaga, M. A. (2024). Concept paper: Innovative approaches to food quality control: AI and machine learning for predictive analysis. *World Journal of Advanced Research and Reviews*, 21(3), 823-828.
- Tian, Y. (2020). Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. *IEEE Access*, 8, 125731–125744.
- Ukai, K., Rahman, R., Yagi, N., Hayashi, K., Maruo, A., Muratsu, H., & Kobashi, S. (2021). Detecting pelvic fracture on 3D-CT using deep convolutional neural networks with multi-orientated slab images. *Scientific Reports*, 11(1)
- van Leeuwen, K. G., Schalekamp, S., Rutten, M. J. C. M., van Ginneken, B., & de Rooij, M. (2021). Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *European Radiology*, 31(6), 3797–3804.
- Wang, Y., Li, Y., Lin, G., Zhang, Q., Zhong, J., Zhang, Y., Ma, K., Zheng, Y., Lu, G., & Zhang, Z. (2023). Lower-extremity fatigue fracture detection and grading based on deep learning models of radiographs. *European Radiology*, 33(1), 555–565.
- Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., Kahng, M., & Polo Chau, D. H. (2021). CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1396–1406.
- Weikert, T., Noordtzi, L. A., Bremerich, J., Stieltjes, B., Parmar, V., Cyriac, J., Sommer, G., & Sauter, A. W. (2020). Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *In Korean Journal of Radiology* (Vol. 21, Issue 7, pp. 891–899).
- Wu, J., Liu, N., Li, X., Fan, Q., Li, Z., Shang, J., Wang, F., Chen, B., Shen, Y., Cao, P., Liu, Z., Li, M., Qian, J., Yang, J., & Sun, Q. (2023). Convolutional neural network for detecting rib fractures on chest radiographs: a feasibility study. *BMC Medical Imaging*, 23(1), 1–12.
- Yang, L., Zhang, R. Y., Li, L., & Xie, X. (2021). SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *Proceedings of Machine Learning Research*, 139, 11863–11874.
- Yao, L., Guan, X., Song, X., Tan, Y., Wang, C., Jin, C., Chen, M., Wang, H., & Zhang, M. (2021). Rib fracture detection system based on deep learning. *In Scientific Reports* (Vol. 11, Issue 1).
- Yoon, A. P., Lee, Y. L., Kane, R. L., Kuo, C. F., Lin, C., & Chung, K. C. (2021). Development and Validation of a Deep Learning Model Using Convolutional Neural Networks to Identify Scaphoid Fractures in Radiographs. *JAMA Network Open*, 4(5), 1–11.
- Zafar, A., Aamir, M., Mohd Nawi, N., Arshad, A., Riaz, S., Alruban, A., Dutta, A. K., & Almotairi, S. (2022). A Comparison of Pooling Methods for Convolutional Neural Networks. *Applied Sciences (Switzerland)*, 12(17), 1–21.
- Zhang, B., Jia, C., Wu, R., Lv, B., Li, B., Li, F., Du, G., Sun, Z., & Li, X. (2021). Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: A clinical evaluation. *British Journal of Radiology*, 94(1118).
- Zhang, J., Liu, F., Xu, J., Zhao, Q., Huang, C., Yu, Y., & Yuan, H. (2023). Automated detection and classification of acute vertebral body fractures using a convolutional neural network on computed tomography. *Frontiers in Endocrinology*, (14)1–10.
- Zhou, Q. Q., Wang, J., Tang, W., Hu, Z. C., Xia, Z. Y., Li, X. S., Zhang, R., Yin, X., Zhang, B., & Zhang, H. (2020). Automatic detection and classification of rib fractures on thoracic ct using convolutional neural network: Accuracy and feasibility. *In Korean Journal of Radiology* (Vol. 21, Issue 7, pp. 869–879).
- Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., Prince, J. L., Rueckert, D., & Summers, R. M. (2021). A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises. *Proceedings of the IEEE*, 109(5), 820–838.

Stevany Nurifin

Rumah Sakit Bhayangkara Tk II Medan
Corresponding author: E-mail: stevanystepi@gmail.com

DOI: <https://doi.org/10.33024/minh.v8i1.666>