

Evaluation of deep learning and convolutional neural network algorithms for mandibular fracture detection using radiographic images: A systematic review and meta-analysis

Mahmood Dashti^{1,*}, Sahar Ghaedsharaf², Shohreh Ghasemi³, Niusha Zare⁴,
Elena-Florentina Constantin⁵, Amir Fahimipour⁶, Neda Tajbakhsh⁷, Niloofar Ghadimi⁸

¹Dentofacial Deformities Research Center, Research Institute of Dental Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

²Department of Oral and Maxillofacial Radiology, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

³Department of Trauma and Craniofacial Reconstruction, Queen Mary College, London, England

⁴Department of Operative Dentistry, University of Southern California, Los Angeles, CA, USA

⁵Dental School, Carol Davila University of Medical Sciences, Bucharest, Romania

⁶Discipline of Oral Surgery, Medicine and Diagnostics, School of Dentistry, Faculty of Medicine and Health, Westmead Centre for Oral Health, The University of Sydney, Sydney, Australia

⁷School of Dentistry, Islamic Azad University Tehran, Dental Branch, Tehran, Iran

⁸Department of Oral and Maxillofacial Radiology, Dental School, Islamic Azad University of Medical Sciences, Tehran, Iran

ABSTRACT

Purpose: The use of artificial intelligence (AI) and deep learning algorithms in dentistry, especially for processing radiographic images, has markedly increased. However, detailed information remains limited regarding the accuracy of these algorithms in detecting mandibular fractures.

Materials and Methods: This meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Specific keywords were generated regarding the accuracy of AI algorithms in detecting mandibular fractures on radiographic images. Then, the PubMed/Medline, Scopus, Embase, and Web of Science databases were searched. The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool was employed to evaluate potential bias in the selected studies. A pooled analysis of the relevant parameters was conducted using STATA version 17 (StataCorp, College Station, TX, USA), utilizing the metandi command.

Results: Of the 49 studies reviewed, 5 met the inclusion criteria. All of the selected studies utilized convolutional neural network algorithms, albeit with varying backbone structures, and all evaluated panoramic radiography images. The pooled analysis yielded a sensitivity of 0.971 (95% confidence interval [CI]: 0.881-0.949), a specificity of 0.813 (95% CI: 0.797-0.824), and a diagnostic odds ratio of 7.109 (95% CI: 5.27-8.913).

Conclusion: This review suggests that deep learning algorithms show potential for detecting mandibular fractures on panoramic radiography images. However, their effectiveness is currently limited by the small size and narrow scope of available datasets. Further research with larger and more diverse datasets is crucial to verify the accuracy of these tools in practical dental settings. (*Imaging Sci Dent* 2024; 54: 232-9)

KEY WORDS: Artificial Intelligence; Image Processing, Computer-Assisted; Radiography, Panoramic; Wounds and Injuries; Deep Learning

Introduction

In recent years, the integration of artificial intelligence (AI) into dentistry has represented a major advancement, introducing a powerful tool for predicting, diagnosing, and developing treatment plans for dental conditions.¹ AI is commonly defined as the field focused on creating intelli-

Received March 1, 2024; Revised May 25, 2024; Accepted June 4, 2024

Published online August 12, 2024

*Correspondence to : Dr. Mahmood Dashti

Dentofacial Deformities Research Center, Research Institute of Dental Sciences, School of Dentistry, Shahid Beheshti University of Medical Sciences, Daneshju Blv., Velenjak St., Chamran Highway, Tehran 1983969411, Iran

Tel) 98-922-274-6706, E-mail) dashti.mahmood72@gmail.com

Copyright © 2024 by Korean Academy of Oral and Maxillofacial Radiology

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Imaging Science in Dentistry · pISSN 2233-7822 eISSN 2233-7830

gent machines capable of assisting humans in performing complex, repetitive tasks with increased efficiency and precision.² Machine learning (ML), a branch of AI, enables systems to autonomously learn and improve their performance by leveraging large datasets to refine their algorithms.³ Within ML, deep learning (DL) models are constructed using neural networks, which are computational architectures inspired by biological neural systems that allow computers to identify patterns in data.⁴ One type of deep feedforward network is the convolutional neural network (CNN), which is designed to process data in multiple array formats, such as images, and features a multi-stage architecture.⁵

Artificial neural networks are used within DL to simulate the operation of the human brain.⁶ These specialized branches of AI are employed to analyze patient dental records, assisting dentists in making predictive judgments and providing more precise diagnoses for treatment plans.⁷ Mandibular fractures, also known as jaw fractures, are a common type of facial injury.^{8,9} These are frequently caused by motor vehicle accidents, workplace accidents, certain medical conditions, falls, and sports-related injuries.¹⁰ Symptoms of a broken jaw can include bleeding, difficulty chewing, swelling, pain in the jaw or face, stiffness, bruising, and breathing difficulties.⁸ Mandibular fractures are classified according to 6 anatomical regions: the symphysis, body, angle, ramus, condyle, and coronoid process.¹¹

A suitable treatment plan for fracture depends on the severity and complexity of the injury, necessitating an accurate assessment.¹⁰ Radiographic imaging is essential for the evaluation and accurate diagnosis of jaw fracture.⁸ This modality is often used alongside other diagnostic tests, such as panoramic radiography, computed tomography, magnetic resonance imaging, and cone-beam computed tomography (CBCT).^{8,12} Panoramic radiography serves as the initial screening tool for patients with facial trauma, as it provides a view of the entire mouth, including both upper and lower jaws.¹² However, despite its diagnostic utility, panoramic radiography has several limitations; these include the absence of a 3-dimensional view, image homogeneity issues, distortions, and magnification errors.¹⁰ The ability of dental professionals to read panoramic radiographs with diagnostic precision varies, with an approximate success rate of 70% for detecting mandibular fractures. Consequently, diagnoses may be incorrect or missed.¹³

The current study aims to evaluate the accuracy of deep learning algorithms in detecting mandibular fractures on 2-dimensional (2D) radiographic images. The null hypothesis posits that deep learning algorithms will not accurately detect mandibular fractures on these images.

Materials and Methods

In the present meta-analysis, research articles were extracted, selected, and screened in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards.¹⁴ In line with these standards, the analysis involved the following steps. (1) Identification: A comprehensive database search was performed to identify potential studies for inclusion. The search terms used, databases searched, and timeframe of the search are detailed in Table 1. (2) Screening: Studies were initially screened based on their titles and abstracts, followed by full-text assessment to determine eligibility. Inclusion and exclusion criteria were strictly defined and followed. (3) Eligibility: The details of the selection process, including the numbers of studies screened, assessed for eligibility, and included in the review, are documented in a PRISMA flow diagram (Fig. 1). (4) Analysis of included studies: Detailed summaries were created for each study, including research characteristics, methodologies, and outcomes of interest (Table 2). (5) Risk of bias assessment: The risk of bias for each study was assessed using established tools and influenced the interpretation of the findings (Table 3). (6) Synthesis of results: The methods used for data extraction and synthesis were described, including the statistical methods employed for the meta-analysis (Figs. 2 and 3).

After the preliminary screening stage, the study protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) and assigned the code CRD42023433523.

The research question posed was, “Are 2D radiographic images suitable for the application of deep learning technologies to identify and predict mandibular fractures?”

Eligibility criteria

According to the inclusion criteria established for the meta-analysis, included studies were those that: 1) employed deep learning for prediction and diagnostic accuracy; 2) involved the assessment and evaluation of 2D radiographic images, such as identification and predictive accuracy regarding panoramic radiography scans; 3) reported claims of accuracy for the results; 4) were dated up to June 2023, to ensure the inclusion of the most recent deep learning-related data; and 5) were written in English. Studies were excluded if they: 1) were scoping reviews, systematic reviews, or meta-analyses; 2) were published in languages other than English; 3) utilized 3D radiographic modalities, such as CBCT; 4) or did not focus on the detection of mandibular fractures on radiographic images using deep learning algorithms.

Research strategy and screening

The systematic search for and evaluation of research articles was conducted across 5 databases: PubMed, Scopus, Scopus secondary, Embase, and Web of Science. The search was limited to articles published through June 2023. The PRISMA guidelines were followed to determine eligibility for inclusion in the meta-analysis. Table 1 presents the keywords used for each database, which were carefully selected to analyze articles from various disciplines. Two reviewers (M.D. and N.M.) independently assessed the titles and abstracts, with a third reviewer (J.L.) resolving any disagreements. All studies that met the eligibility criteria, and for which full texts were accessible, were included.

Table 1 presents the data extracted from the study articles. The information was collated based on study characteristics, including author, publication year, country of study, imaging modality, dataset size, model architecture, and conclusions (see Table 2). Studies that utilized multiple test datasets or model types were subjected to thorough extraction.

Quality evaluation

The methodological quality of the included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool.¹⁵ This instrument evaluates 4 key components: patient inclusion criteria, index tests,

reference standards, and the flow of participants through the study, including the precise sequence of tests and standards applied. Each component was independently assessed for potential risk of bias by 2 authors (M.D. and N.M.). The risk of bias was classified as low, high, or unclear, with arbitration by a third author when disagreements arose. The analysis focused on patient participation, methodology, and the adequacy of outcomes, while also exploring the heterogeneity of findings across the included studies.

Summary measures and data synthesis

Supporting the integrity and robustness of the findings, the Egger regression test was applied to assess potential publication bias. This test is a statistical method used to detect bias in meta-analyses by quantifying funnel plot asymmetry. A funnel plot is a scatter plot that maps the treatment effects estimated from individual studies against the precision of each study, typically using the inverse of the standard error as a proxy for precision.

In the analysis, effect sizes from studies employing CNN architecture with panoramic radiographs were plotted against their respective standard errors to generate a funnel plot. This plot could then be used to visually assess the presence or absence of publication bias, with asymmetry suggesting potential bias.

Table 1. Keywords for each database

Database	Keyword	Results
PubMed	("Machine Learning" [Mesh] OR "Deep Learning" [Mesh] OR "Supervised Machine Learning" [Mesh] OR "Unsupervised Machine Learning" [Mesh] OR "Neural Networks, Computer" [Mesh] OR "artificial intelligence") AND ("Mandibular Fractures" [Mesh] OR "Jaw Fractures" [Mesh])	6
Embase	('machine learning'/exp OR 'machine learning' OR 'deep learning'/exp OR 'deep learning' OR 'supervised machine learning'/exp OR 'supervised machine learning' OR 'unsupervised machine learning'/exp OR 'unsupervised machine learning' OR 'artificial neural network'/exp OR 'artificial neural network' OR 'artificial intelligence') AND ('mandible fracture'/exp OR 'mandible fracture' OR 'jaw fracture'/exp OR 'jaw fracture')	13
Scopus	(TITLE-ABS-KEY ("Machine Learning") OR TITLE-ABS-KEY ("Deep Learning") OR TITLE-ABS-KEY ("Supervised Machine Learning") OR TITLE-ABS-KEY ("Unsupervised Machine Learning") OR TITLE-ABS-KEY ("Neural Network") OR TITLE-ABS-KEY ("artificial intelligence")) AND (TITLE-ABS-KEY ("Mandibular Fracture") OR TITLE-ABS-KEY ("Jaw Fracture") OR TITLE-ABS-KEY ("Jaw Fractures") OR TITLE-ABS-KEY ("Mandibular Fractures"))	15
Scopus secondary	(TITLE-ABS-KEY ("Machine Learning") OR TITLE-ABS-KEY ("Deep Learning") OR TITLE-ABS-KEY ("Supervised Machine Learning") OR TITLE-ABS-KEY ("Unsupervised Machine Learning") OR TITLE-ABS-KEY ("Neural Network") OR TITLE-ABS-KEY ("artificial intelligence")) AND (TITLE-ABS-KEY ("Mandibular Fracture") OR TITLE-ABS-KEY ("Jaw Fracture") OR TITLE-ABS-KEY ("Jaw Fractures") OR TITLE-ABS-KEY ("Mandibular Fractures"))	1
WOS	(TS = ("Machine Learning" OR "Deep Learning" OR "Supervised Machine Learning" OR "Unsupervised Machine Learning" OR "Neural Networks" OR "artificial intelligence")) AND TS = ("Mandibular Fractures" OR "Jaw Fractures" OR "Mandibular Fracture" OR "Jaw Fracture")	4

WOS: Web of Science

Table 2. Extracted data

Author (year)	Country	Imaging modality	Dataset size	Architecture model	Accuracy	Sensitivity/recall	Specificity	Precision	F1-score/Dice coefficient	AUC	TP	TN
Nishiyama et al. ¹² (2021)	Japan	Panoramic radiograph	100	CNN	0.815	0.83	0.80			0.89	0.83	0.80
Shahnavazi et al. ¹³ (2023)	Iran	Panoramic radiograph	190	CNN (Faster R-CNN)	0.9167	100	0.8333			0.91	0.85	0.84
Son et al. ¹⁶ (2021)	Korea	Panoramic radiograph	60	CNN (YOLOv4) MLAT and SLAT Class 2		0.691		0.985	0.812			
Son et al. ¹¹ (2022)	Korea	Panoramic radiograph	60	CNN (YOLOv4) MLAT and SLAT Class 6		0.794		0.975	0.875			
Vinayahalingam et al. ¹⁰ (2022)	Netherlands/Germany	Panoramic radiograph	320	CNN (U-net and LAT YOLOv4)		0.866		0.955	0.908			
				CNN (Faster R-CNN)		0.960		0.935	0.947	0.977		

AUC: area under the curve, TP: true positives, TN: true negatives, CNN: convolutional neural network

Diagnostic accuracy was analyzed using hierarchical logistic regression. The meta-analysis was limited to the “sensitivity” and “precision” factors that were present in all studies. Specificity was defined as the number of true negatives divided by the sum of true negatives and false positives. Sensitivity was calculated as the number of true positives divided by the sum of true positives and false negatives. The positive likelihood ratio was determined by dividing sensitivity by (1 - specificity), while the negative likelihood ratio was calculated by dividing (1 - specificity) by specificity. The diagnostic odds ratio (DOR) was computed as the positive likelihood ratio divided by the negative likelihood ratio. These parameters were pooled using STATA version 17 (StataCorp, College Station, TX, USA) with the metandi command. A significance level of 0.05 was established.

Results

Identified studies

In the meta-analysis, 49 articles were found within the searched databases regarding the effectiveness of deep learning systems in identifying and predicting mandibular fractures. Of these, 19 studies were deemed relevant, reliable, and aligned with the goals of the study, according to the inclusion criteria. When the exclusion criteria were applied, 12 of the initial 19 articles were discarded, leaving 7 publications that met the requirements. Ultimately, 5 studies were included in the meta-analysis (Fig. 1). The final 2 studies were excluded due to not adequately describing the dataset (n = 1) and not utilizing 2D radiographic images (n = 1).

Descriptive analysis of identified studies

Of the 49 selected articles, 5 studies were included in the data extraction step. These studies, representing various global regions, consisted of model-based research conducted in South Korea, Japan, Iran, the Netherlands, and Germany. They employed CNNs to detect mandibular fractures and reported various outcomes, including accuracy, sensitivity/recall, specificity, precision, F1-score/Dice coefficient, and area under the receiver operating characteristic curve, as well as true positive and true negative rates (Table 2).

Risk of bias

None of the studies fully satisfied the quality criteria of the QUADAS-2 evaluation tool. However, the methodologies were consistent across studies. Regarding patient selection, a significant risk of bias was identified in 3 of the

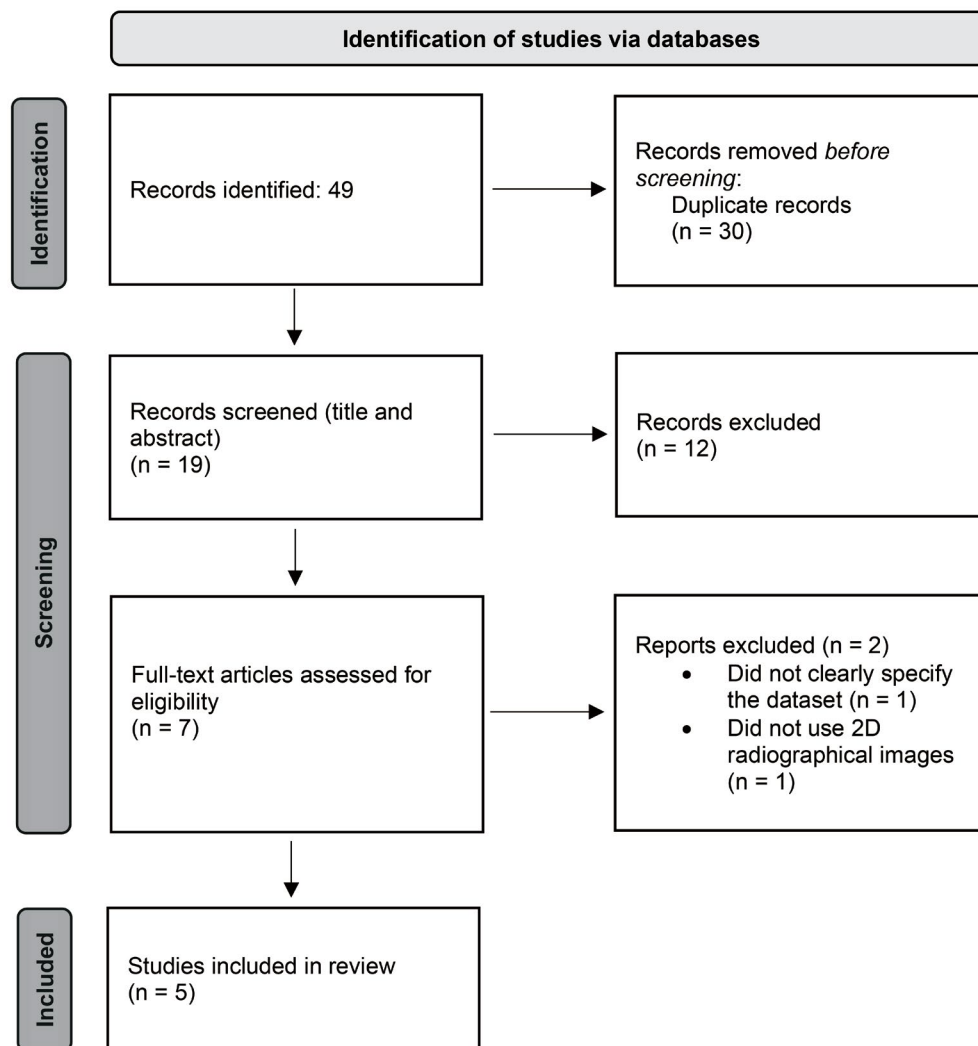


Fig. 1. PRISMA flowchart for article selection. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses, 2D: 2-dimensional.

5 articles. Due to the nature of the experimental designs, which did not involve direct patient intervention, none of the studies could address the question, “Was a consecutive or random sample of patients recruited?” Furthermore, the absence of control groups in the included studies meant that the second question (“Was a case-control design avoided?”) was answered in the affirmative. Two of the articles did not report interobserver or intraobserver agreement, which left the potential for bias in the reference standards unclear. These findings are summarized in Table 3. In terms of risk of bias, 1 study¹³ was deemed to have a high risk, while 2 studies^{10,16} had a low risk. No studies raised concerns regarding applicability.

Meta-analysis results

The meta-analysis incorporated 5 studies.^{10-13,16} Nishi-

yama et al.¹² and Son et al.¹⁶ each utilized 2 different CNN architecture modalities, which were treated as separate studies. Consequently, a total of 7 studies were included in the final meta-analysis. The pooled sensitivity, specificity, and DOR were 0.971 (95% confidence interval [CI]: 0.881-0.949), 0.813 (95% CI: 0.797-0.824), and 7.109 (95% CI: 5.271-8.913), respectively (Figs. 2 and 3). After the application of a logit transformation, the results demonstrated a positive association between sensitivity and specificity ($r=0.801$) (Fig. 2). Additionally, the beta variable demonstrated a significant result ($P=0.023$).

No subgroup analysis was conducted, as all studies used panoramic radiographs and CNN architecture. The results of the Egger test, as shown in the funnel plot in Figure 3, indicated that publication bias was not significant ($E=0.813$).

Table 3. QUADAS-2 quality assessment

Author (year)	Risk of bias				Applicability		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Nishiyama et al. ¹² (2021)	Unclear	Low	High	High	Low	Low	Low
Shahnavazi et al. ¹³ (2023)	Unclear	Low	Unclear	High	Low	Low	Unclear
Son et al. ¹⁶ (2021)	High	Low	Low	Low	Low	Low	Low
Son et al. ¹¹ (2022)	High	Low	Unclear	Low	Unclear	Low	Low
Vinayahalingam et al. ¹⁰ (2022)	Unclear	Low	Low	Unclear	Low	Low	Low

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies 2

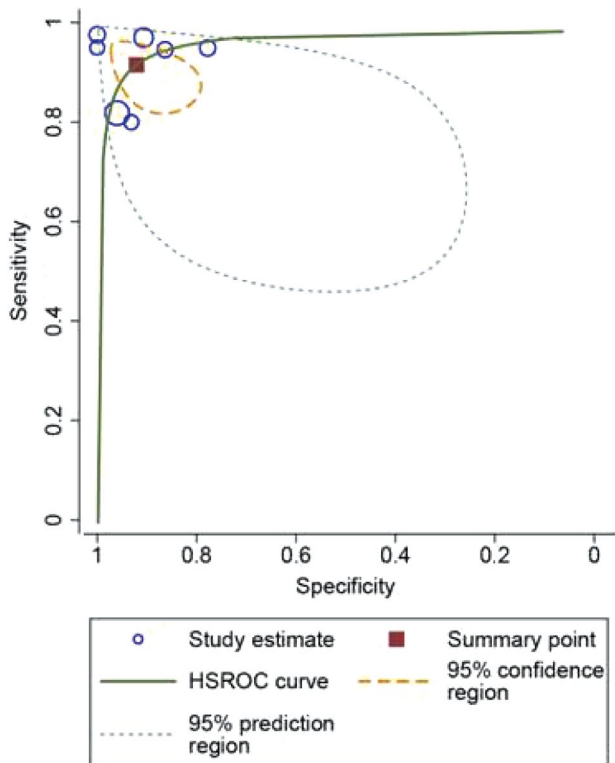


Fig. 2. Hierarchical summary receiver operating characteristic (HSROC) curve comparing sensitivity and specificity across studies.

Discussion

The aim of this systematic review and meta-analysis was to assess the accuracy, sensitivity, and specificity of DL algorithms in the detection of mandibular fractures using 2D radiographic images, specifically panoramic radiographs. The meta-analysis, which included 5 studies,^{10-13,16} revealed that the overall DOR of CNN algorithms for identifying mandibular fractures on panoramic radiography images was 7.109 (95% CI: 5.271-8.913). The sensitivity was 0.971 (95% CI: 0.881-0.949), and the specificity was 0.813 (95%

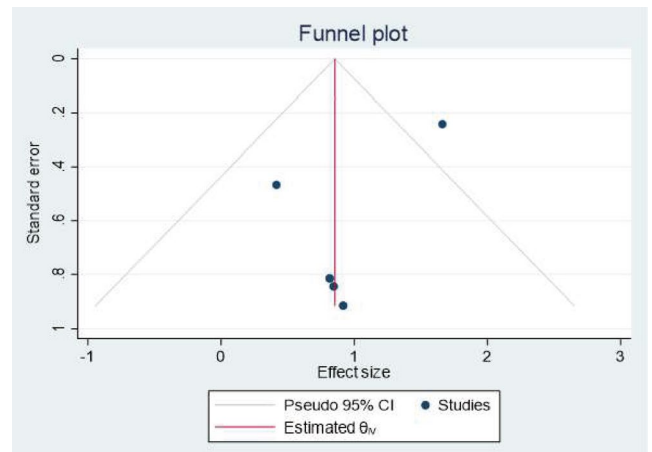


Fig. 3. Funnel plot of deep learning model in mandibular fracture detection. No significant publication bias was evident, as the distribution of studies was symmetrical (Egger test = 0.813). CI: confidence interval.

CI: 0.797-0.824).

The DOR is a generic indicator of diagnostic precision, used to estimate the discriminatory capacity of diagnostic tests and to compare the diagnostic accuracy between tests.¹⁷ It is calculated using the formula: $DOR = (\text{true positives}/\text{false negatives})/(\text{false positives}/\text{true negatives})$. The specificity and sensitivity of a test significantly impact the DOR. A test that demonstrates high specificity and sensitivity, with low false negative and false positive rates, will have a high DOR. As test specificity increases, the DOR also rises, provided that sensitivity remains constant.¹⁷ According to current standards, a DOR of 10.00 is considered excellent.¹⁷ The observed DOR of 7.109 (95% CI: 5.271-8.913) indicates a reliable and good result.

The studies included in this review^{10-13,16} utilized consistent imaging modalities and comparable model architectures, but with different backbones. Despite their similarities, all studies were limited by relatively small sample sizes.

To more thoroughly assess the efficacy of deep learning algorithms, further research involving larger databases is necessary.

The findings indicate that deep learning algorithms, particularly CNNs, can accurately detect mandibular fractures on panoramic radiography images. This aligns with previous studies demonstrating the effectiveness of CNNs in various dental applications, including caries detection, tooth segmentation, and landmark identification.¹⁸⁻²⁰ Furthermore, CNNs have been found to outperform other ML algorithms in dental contexts because they automatically extract features from images, making them highly effective in recognizing subtle patterns and irregularities.²¹

One potential application of deep learning algorithms in the detection of mandibular fractures is to improve the diagnostic accuracy and efficiency of dentists and surgeons. Accurate identification of these fractures on panoramic radiography images can provide crucial information to dentists early in the diagnostic process, as these images are more accessible than CBCT scans. Early detection helps ensure that patients receive timely and appropriate treatment plans. The observed high accuracy of deep learning algorithms in identifying mandibular fractures on panoramic radiography images suggests potential clinical applications in this area.

Nevertheless, the studies included in this review^{10-13,16} involved small datasets of panoramic radiography images. Further research is required to comprehensively evaluate the accuracy of deep learning algorithms in detecting mandibular fractures and to compare this accuracy across panoramic radiography images of varying quality.

The results of this review suggest that deep learning algorithms can assist dental professionals in detecting mandibular fractures on panoramic radiographs. Nevertheless, the effectiveness of these tools may be limited by the small size and limited scope of the datasets currently available. Thus, the conclusions of this study are preliminary and should be approached with caution. To validate the accuracy of deep learning algorithms in identifying mandibular fractures, future research with larger and more diverse datasets will be essential to establish their reliability in clinical dental practice.

Conflicts of Interest: None

Acknowledgments

During the preparation of this work, the authors used ChatGPT 4 (OpenAI, San Francisco, CA, USA) to improve the flow and grammar of the manuscript. After using this

tool, the authors reviewed and edited the content as needed. The authors take full responsibility for the content of the publication.

References

1. Agrawal P, Nikhade P. Artificial intelligence in dentistry: past, present, and future. *Cureus* 2022; 14: e27405.
2. Toh TS, Dondelinger F, Wang D. Looking beyond the hype: applied AI and machine learning in translational medicine. *EBioMedicine*. 2019; 47: 607-15.
3. Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. *Quintessence Int* 2020; 51: 248-57.
4. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015; 61: 85-117.
5. Yamashita R, Nishio M, Do RK, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018; 9: 611-29.
6. Zhang Z, Sejdić E. Radiological images and machine learning: trends, perspectives, and prospects. *Comput Biol Med* 2019; 108: 354-70.
7. Rodrigues JA, Krois J, Schwendicke F. Demystifying artificial intelligence and deep learning in dentistry. *Braz Oral Res* 2021; 35: e094.
8. Brown JS, Khan A, Wareing S, Schache AG. A new classification of mandibular fractures. *Int J Oral Maxillofac Surg* 2022; 51: 78-90.
9. Boffano P, Rocca F, Zavattoni E, Deddiol E, Uglešić V, Kovačič Ž, et al. European Maxillofacial Trauma (EURMAT) project: a multicentre and prospective study. *J Craniomaxillofac Surg* 2015; 43: 62-70.
10. Vinayahalingam S, van Nistelrooij N, van Ginneken B, Bresslem K, Tröltzsch D, Heiland M, et al. Detection of mandibular fractures on panoramic radiographs using deep learning. *Sci Rep* 2022; 12: 19596.
11. Son DM, Yoon YA, Kwon HJ, Lee SH. Combined deep learning techniques for mandibular fracture diagnosis assistance. *Life (Basel)* 2022; 12: 1711.
12. Nishiyama M, Ishibashi K, Ariji Y, Fukuda M, Nishiyama W, Umemura M, et al. Performance of deep learning models constructed using panoramic radiographs from two hospitals to diagnose fractures of the mandibular condyle. *Dentomaxillofac Radiol* 2021; 50: 20200611.
13. Shahnavaizi 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.
14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151: 264-9.
15. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529-36.
16. 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: 933.

17. Šimundić AM. Measures of diagnostic accuracy: basic definitions. *EJIFCC* 2009; 19: 203-11.
18. Khanagar SB, Alfouzan K, Awawdeh M, Alkadi L, Albalawi F, Alfadley A. Application and performance of artificial intelligence technology in detection, diagnosis and prediction of dental caries (DC) - a systematic review. *Diagnostics (Basel)* 2022; 12: 1083.
19. Londono J, Ghasemi S, Hussain Shah A, Fahimipour A, Ghadimi N, Hashemi S, et al. Evaluation of deep learning and convolutional neural network algorithms accuracy for detecting and predicting anatomical landmarks on 2D lateral cephalometric images: a systematic review and meta-analysis. *Saudi Dent J* 2023; 35: 487-97.
20. Dashti M, Londono J, Ghasemi S, Tabatabaei S, Hashemi S, Baghaei K, et al. Evaluation of accuracy of deep learning and conventional neural network algorithms in detection of dental implant type using intraoral radiographic images: a systematic review and meta-analysis. *J Prosthet Dent* (in press).
21. Chen L, Li S, Bai Q, Yang J, Jiang S, Miao Y. Review of image classification algorithms based on convolutional neural networks. *Remote Sens* 2021; 13: 4712.