Original Article

Performance of Artificial Intelligence using Oral and Maxillofacial CBCT Images: A Systematic Review and Meta-Analysis

FF Badr, FM Jadu

Department of Oral Diagnostic Sciences, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

Background: Artificial intelligence (AI) has the potential to enhance health care efficiency and diagnostic accuracy. **Aim:** The present study aimed to determine the current performance of AI using cone-beam computed tomography (CBCT) images for detection and segmentation. **Materials and Methods:** A systematic search for scholarly articles written in English was conducted on June 24, 2021, in PubMed, Web of Science, and Google Scholar. Inclusion criteria were peer-reviewed articles that evaluated AI systems using CBCT images for detection and segmentation purposes and achieved reported outcomes in terms of precision and recall, accuracy, based on DICE index and Dice similarity coefficient (DSC). The Cochrane tool for assessing the risk of bias was used to evaluate the studies that were included in this meta-analysis. A random-effects model was used to calculate the pooled effect size. **Results:** Thirteen studies were included for review and analysis. The pooled performance that measures the included AI models is 0.85 (95%CI: 0.73,0.92) for DICE index/DSC, 0.88 (0.77,0.94) for precision, 0.93 (0.84, 0.97) for recall, and 0.83 (0.68, 0.91) for accuracy percentage. **Conclusion:** Some limitations are identified in our meta-analysis such as heterogenicity of studies, risk of bias and lack of ground truth. The application of AI for detection and segmentation using CBCT images is comparable to services offered by trained dentists and can potentially expedite and enhance the interpretive process. Implementing AI into clinical dentistry can analyze a large number of CBCT studies and flag the ones with significant findings, thus increasing efficiency. The study protocol was registered in PROSPERO, the international registry for systematic reviews (ID number CRD42021285095).

Keywords: *Artificial intelligence, cone-beam computed tomography, dentistry, machine learning*

Received: 09-Jun-2022; **Revision:** 22-Jul-2022; **Accepted:** 22-Aug-2022; **Published:** 18-Nov-2022

Recent advances in artificial intelligence (AI) and state of art neural networks[1] have been used for various applications that include speech, vision, robotics, natural language processing, and machine learning to name a few. Deep learning is a subset of machine learning commonly used in diagnostic imaging. Deep learning AI systems, known as deep neural networks, are capable of learning by extracting features from training data and interpreting test data, without explicit instructions.[1] Convolutional neural networks are a deep learning architecture used for large and complex images such as cone-beam computed tomography (CBCT) and magnetic resonance imaging.[1]

AI represents a significant paradigm shift in the field of diagnostic imaging. This is observed because AI systems are now capable of performing tasks such as disease detection, prediction, image segmentation, and classification at a level that equals and even exceeds human ability.[1] Computer-aided diagnosis represents a new era where machines are capable of rectifying human error during diagnosis.[2] In the field of Oral and

Access this article online

Quick Response Code:

Website: www.njcponline.com

DOI: 10.4103/njcp.njcp_394_22

Address for correspondence: Dr. FF Badr, Department of Oral Diagnostic Sciences, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia. E-mail: ffbadr@kau.edu.sa

How to cite this article: Badr FF, Jadu FM. Performance of artificial intelligence using oral and maxillofacial CBCT images: A systematic review and meta-analysis. Niger J Clin Pract 2022;25:1918-27.

Maxillofacial Radiology (OMR), periapical, bitewing, panoramic, and lateral cephalometric conventional radiographs are being used along with CBCT images to detect dental caries, periapical and periodontal disease, root fractures, osteoporosis, cyst and tumors of the jaws.[3-10] This study offers a promising contribution in demonstrating that AI systems offer high accuracy and excellent reliability.[3] In addition, integrating AI into the workflow significantly reduces manual labor and time wasted.[11]

Nevertheless, it is important to address some issues before AI can be efficiently used in clinical practice. These issues include the voluminous amount of data that is needed to train, validate, and test AI systems.[1,3] In addition, these data sets must be properly labeled which is a time-consuming task. The datasets must also be accurately interpreted which is an aspirational goal even for the most experienced radiologists.[3] Moreover, reliability of AI results may be difficult to comprehend, justify and accept especially for tasks that involve human judgement.[12,13] Privacy is another issue because the terabytes of data are being shared and used for the development of AI systems without a guarantee to protect the privacy of patient information.[13] Ethics or ethical consideration is of concern since no laws govern AI development and its application, thus far.[14] Finally, there is a significant risk of bias in AI studies that is difficult to quantify from the start of data selection to the final interpretation of the results.[13]

Several reviews have been published regarding the application of AI in the field of OMR, but no meta-analysis has been performed that qualitatively evaluate the performance of AI systems in OMR applications. Therefore, this meta-analysis was undertaken to review, quantify, and summarize the current performance of AI applications in the field of OMR.

Methods

The current study was conducted according to the preferred reporting items for systematic review and meta-analysis (PRISMA) guidelines. The study protocol was registered in PROSPERO, the international registry for systematic reviews[15] (ID number CRD42021285095).

Data sources and search strategy

A systematic search of the literature was conducted in the following databases: PubMed, Web of Science, and

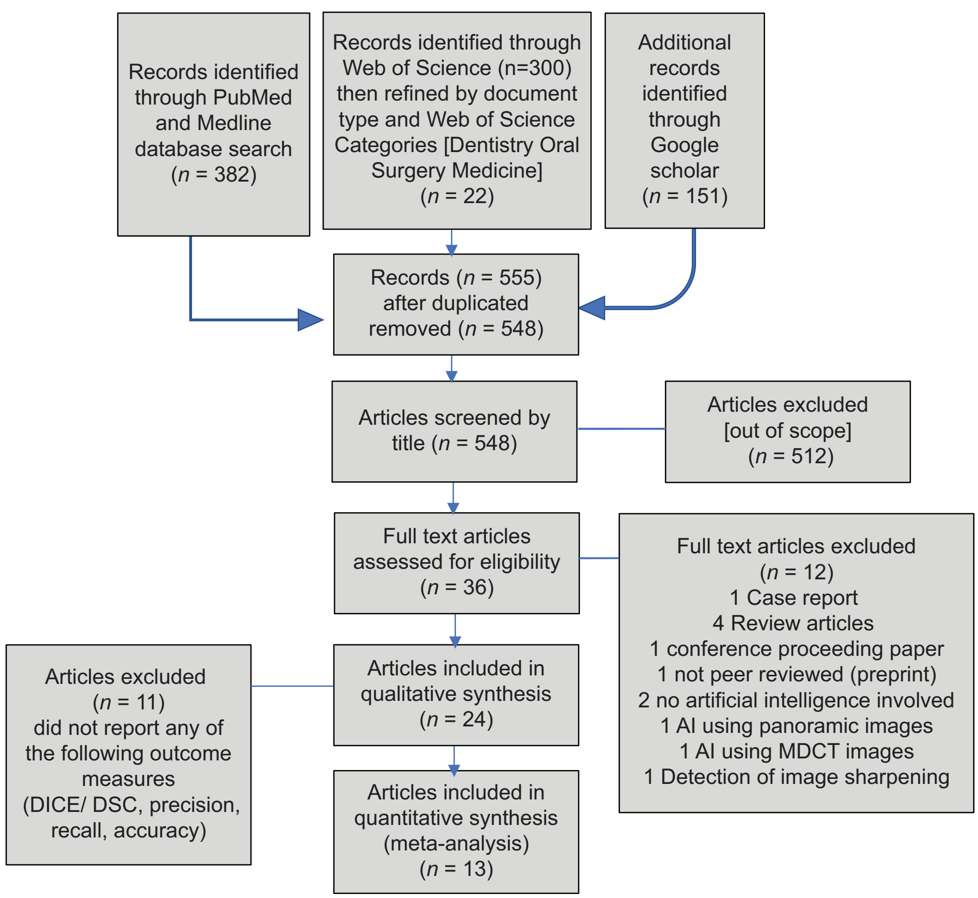

Figure 1: Study selection flowchart. MDCT, multidetector computed tomography; DICE, dice coefficient; DSC, dice similarity coefficient

Google Scholar. The aim was to identify studies that used CBCT images to develop any type of AI model and to perform any task. Electronic searches were augmented by searching references. The search strategy was designed by two OMR consultants.

PUBMED search strategy

((((“artificial intelligence”[Title/Abstract] OR “machine learning”[Title/Abstract] OR “deep learning”[Title/Abstract]) AND “cbct”[Title/Abstract]) OR “cone beam computed tomography”[Title/Abstract]) AND “dentistry”[Title/Abstract]) AND ((y_10[Filter]) AND (english[Filter]))

Web of Science search strategy

(TS = (artificial intelligence OR deep learning OR machine learning)) AND TS = (cbct OR cone beam computed tomography OR cone-beam computed tomography)). Refined by document type (excluding proceedings papers, meeting abstracts, review articles, early access, editorial materials, and data papers). Further refined by Web of Science categories to include only: Dentistry-Oral Surgery Medicine

Figure 2: Cochrane Collaboration’s tool for assessing the risk of bias (adapted from Higgins and Altman), omitting attrition bias due to the nature of AI studies

Author, year	Selection bias	Performance bias	Detection bias	Reporting bias	Other bias
Zheng Z 2021[17]	−	−	+	+	+
Zheng Q 2021[18]	−	−	+	+	+
Setzer F 2020[19]	−	−	+	+	+
Chen S 2020[20]	−	?	+	+	+
Shujaat 2021[21]	?	?	+	+	+
Wang H 2021[22]	−	−	+	+	+
Wang X 2021[23]	−	−	+	+	+
Jaskari J 2020[11]	?	−	+	+	+
Leonardi 2021[24]	−	−	+	+	+
Orhan K 2020[9]	−	?	+	+	+
Lee K 2020[25]	−	−	+	+	+
Shaheen E 2021[26]	?	?	+	+	+
Shoukri B 2019[27]	?	?	+	+	+

Key − High risk of bias. + low risk of bias. ? unclear risk of bias

Google Scholar search strategy

allintitle: artificial OR intelligence OR deep OR learning OR machine OR learning “CBCT “ OR “cone-beam computed tomography”.

Inclusion criteria

All studies were screened, and those that met the following inclusion criteria were selected: (1) peer-reviewed full-text articles published in the English language, (2) articles that evaluated AI systems using CBCT images of the head and neck of adult patients, (3) articles that explored automatic detection or segmentation of anatomical landmarks or pathological lesions, (4) and articles that reported the outcome based on DICE index, DICE ratio, DICE score or dice similarity coefficient (DSC) or precision and recall, or accuracy percentage. Studies excluded from this meta-analysis were those that assessed AI for nondiagnostic purposes such as prediction, image quality improvement, or dose adjustment.

Selection process

Figure 1 details the process of article review and selection. The Cochrane tool for assessing the risk of bias was used to evaluate the studies that were included in this meta-analysis [Figure 2][16] .

Data extraction

A data extraction tool was used to extract relevant information including total sample size, training sample, validation sample, testing sample, use of multivendor images, use of the external dataset, prior image manipulation/preparation, type of AI model, purpose, benchmarking to experts, commercial availability, and reported performance measure after the CLAIM (Checklist for Artificial Intelligence in Medical Imaging).[28] The authors of this study extracted the data independently. Disagreements were resolved by discussion.

Outcome

Three types of AI outcome measures were commonly reported: DICE/DSC, precision and recall, and accuracy percentage. Therefore, the studies were grouped into three based on the outcome measure. The first group reported AI performance in terms of DICE/DSC, which was the most frequently used index to validate segmentation performance.[29] The DICE/DSC index provides the degree of overlapping between automated and ground truth pixels ranging from 0 (no overlap) to 1 (complete overlap). The second group reported AI performance as precision and recall. Precision is defined as the volume of the correctly segmented region over the volume of the segmentation results. However, recall is defined as the

Table 1: Group 1 included nine articles that reported the outcome as DICE index, DICE ratio, DICE score, or DICE similarity coefficient (DSC)

First author, year	Total sample	Sample					AI model						Outcome
		Training set	Validation set	Testing set	External test set	Multivendor images	Preparation of images (annotation or cropping)	type	Purpose	Benchmark AI to radiology experts	Demonstrate how AI makes decision	Is it commercially available?	AI performance measure DICE/DSC
Zheng Z 2021[17]	100 images	15 pts	5 pts	-	-	-	Augmentation, flip, zoom	Anatomically constrained dense U-Net	Segmentation and lesion detection	1 radiology expert	-	no	DICE= 0.67-0.74
Zheng Q 2021[18]	180 CBCT	37	10	133	-	-	-	DL and LS (Deep Learning and Level Set)	Age estimation based on 3D pulp segmentation	Not clear who performed manual segmentation	-	no	DICE for segmentation =87.8%
Setzer F 2020[19]	20 pts	16	4	-	-	-	Normalization of parameters, labelling, presegmentation, initialization, evolution	U-Net DLS	Detection of periapical lesions	3 trained calibrated reviewers	-	no	DICE=0.67
Chen S 2020[20]	60 pts (30 SG+30 CG)	30	-	6	-	-	-	LINKS	Segmentation and landmark finder	-	-	-	DICE ratio=0.8
Shujaat 2021[21]	103 CT and CBCT	48	30	25	-	Yes CT+2 CBCT	-	CNN	Automatic segmentation of pharyngeal airway	2 trained calibrated experts	-	-	DICE score= 0.97+/-0.02
Wang H 2021[22]	30	21	unclear	7	No	No	Cropping	MS-D network	Segmentation performance of jaws and teeth	4 dentists	-	-	DSC jaw=0.934 DSC teeth=0.945
Wang X 2021[23]	60 pts	24 x 2	3 x 2	3 x 2	No	No	orientation	3D U-Net	To quantify 3D asymmetry of the maxilla in cleft pts	2 specialists	No	No	DSC=0.77
Jaskari J 2020[11]	594 pts (637 CBCT volumes)	457	52	128	Yes	No	Rescaling to 0.4 mm	Fully convolutional deep NN	Segmentation of mandibular canal	radiologists	No	No	DSC=0.575 for both canals
Leonardi 2021[24]	40 CBCT	20	-	20	No	No	Reorientation, magnification, fixed threshold	CNN	Segmentation of sinonasal cavity and pharyngeal airway	orthodontist	No	No	DSC=96.7

Table 2: Group two included five articles that reported the outcome as precision and recall

First author, year	Total sample	Sample						AI model					Outcome
		Training set	Validation set	Testing set	External test set	Multivendor images	Preparation of images (annotation or cropping)	type	Purpose	Benchmark AI to radiology experts	Demonstrate how AI makes decision	Is it commercially available?	AI performance measure Precision/recall
Zheng Z 2021[17]	100 images	15 pts	5 pts	-	No	No	Augmentation, flip, zoom	Anatomically constrained dense U-Net	Segmentation and lesion detection	1 radiology expert	-	no	Precision= 0.83-0.9 Recall=0.8-0.84
Orhan K 2020[9]	109 pts (153 PA lesions)	-	-	-	No	No	-	Deep CNN (U-Net)	Detection of periapical pathosis	radiologist	-	no	Precision=0.95 Recall=0.89
Lee K 2020[31]	314 pts	1757	1757	300	No	No	resizing	Single shot detection	Detection of TMJ osteoarthritis	2 orthodontists and 1 TMJ specialist	-	no	Precision=0.85 Recall=0.84
Shujaat 2021[21]	103 CT and CBCT	48	30	25	-	Yes CT+2 CBCT	-	CNN	Automatic segmentation of pharyngeal airway	2 trained calibrated experts	-	-	Precision= 0.97+/-0.01 Recall= 0.96+/-0.03
Shaheen E 2021[26]	186 CBCT	140	35	11	-	2 CBCT units	-	3D U-Net	Segmentation and classification of teeth	-	-	-	precision (0.98±0.02) Recall (0.83±0.05)

Table 3: Group three included three articles that reported the outcome as accuracy percentage

First author, year	Total sample	Sample					AI model					Outcome	
		Training set	Validation set	Testing set	External test set	Multivendor images	Preparation of images (annotation or cropping)	type	Purpose	Benchmark AI to radiology experts	Demonstrate how AI makes decision	Is it commercially available?	AI performance measure Accuracy %
Orhan K 2021[32]	65 pts (130 3rd molars)	-	-	-	No	No	yes	CNN (U-Net)	Detection of impacted 3rd molars	Radiologist	-	no	Accuracy=86.2%
Setzer F 2020[19]	20 pts	16	4	-	No	No	Normalization of parameters, labelling, presegmentation, initialization, evolution	U-Net DLS	Detection of periapical lesions	3 trained calibrated reviewers	-	no	Accuracy=93%
Shoukri B 2019[27]	293 condyles	34 condyles (internal sample)			Yes 259 condyles	No	unclear	unclear	Detection and staging of TMJ osteoarthritis	2 clinical experts Also, salivary and serum biomarkers	-	no	Accuracy=73.5% (Compared to human accuracy 91.2%)

size of the correctly segmented region over the ground truth. The third group of studies reported performance as accuracy percentage, defined as the degree to which the segmentation results agreed with the ground truth segmentation, in percentage.[30]

Each of the AI outcome groups mentioned above was further subdivided based on their purpose into either segmentation or detection. One study was excluded from the first group because it was testing detection while the rest of the studies tested segmentation in that group.

Quality assessment

A risk of bias assessment tool, specific to diagnostic and prediction models in AI research, does not exist. Assessing the risk of bias in studies that evaluate the performance of AI is somewhat ambiguous owing to the novelty of these studies. Nevertheless, we used the Cochrane tool to assess the risk of bias and evaluate the studies included in this meta-analysis which revealed moderate certainty of evidence [Figure 2].[16]

Statistical methodology

This study used Borenstein and Rothstein (1999) Comprehensive Meta-Analysis: A Computer Program for Research Synthesis, Version 1.0. 23 [Computer Software], Biostat, Englewood Cliffs. A random-effects model was used to calculate the pooled effect size. A Q test was used to determine heterogeneity. A funnel plot, classic fail-safe N, and Begg and Mazumdar Rank Correlation were used for publication bias.

Results

Thirteen studies were included in this meta-analysis. Nine were in the first group that reported the outcomes based on the DICE/DSC [Table 1]. The second group consisted of five studies that reported the outcome using precision and recall [Table 2]. The third and final group included five studies that measured the outcome in terms of accuracy percentage [Table 3].

The combined performance of the first group is described in [Table 4]. The pooled DICE/DSC for the group was 0.85. The funnel plot of the first group [Figure 3] demonstrated the combined effect size of more studies on the right side, suggesting publication bias. In the absence of publication bias, we expect the studies to be distributed symmetrically around the combined effect size. The fail-safe N for this group was calculated as 367, indicating that we must locate and include 367 'null' studies for the combined 2-tailed P value to exceed 0.050. Begg and Mazumdar Rank Correlation Test was performed, and Kendall's tau b (corrected for ties, if any) was 0.46, with a 1-tailed P value of 0.05 or a 2-tailed P value of 0.11 (based on continuity-corrected normal approximation).

In the second group of studies, the combined performance is described in [Table 5] for precision outcome and [Table 6] for recall outcome. The pooled precision was 0.92, and the pooled recall was 0.88. The funnel plot of the second group [Figures 4 and 5] demonstrates how smaller studies, (which appear toward the bottom) are more likely to be published if they have larger than average effects, which makes them more likely to meet the criterion for statistical significance.

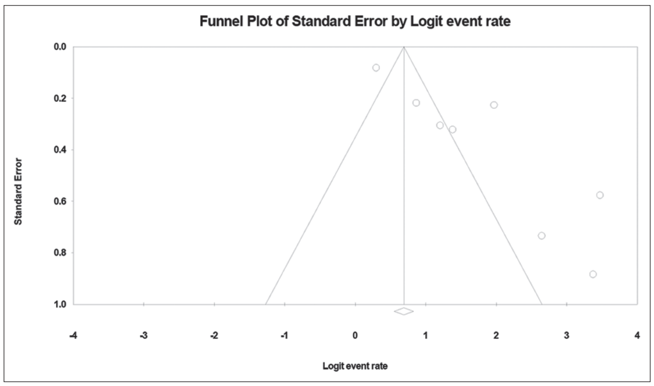

Figure 3: Funnel plot for AI performance as DICE/DSC (group 1)

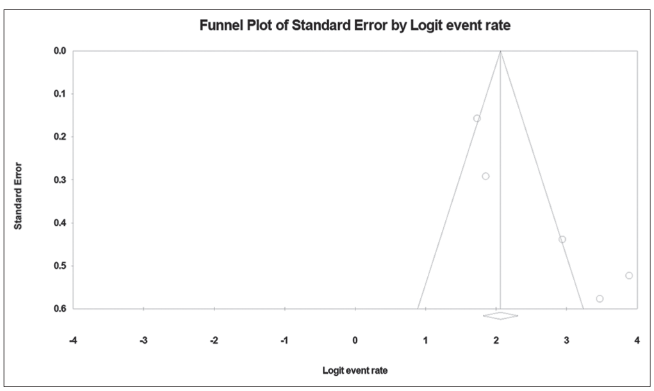

Figure 4: Funnel plot for AI performance as precision (group 2)

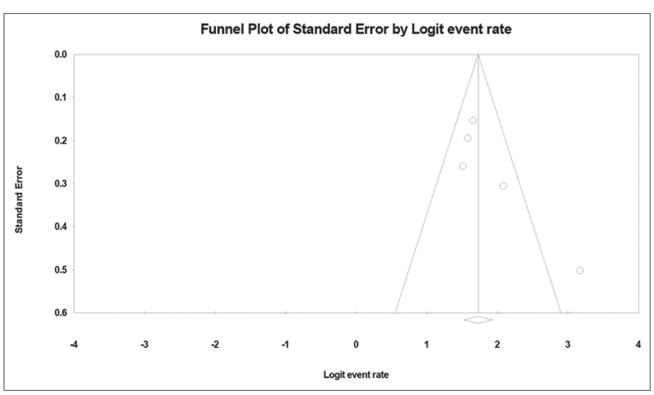

Figure 5: Funnel plot for AI performance as recall (group 2)

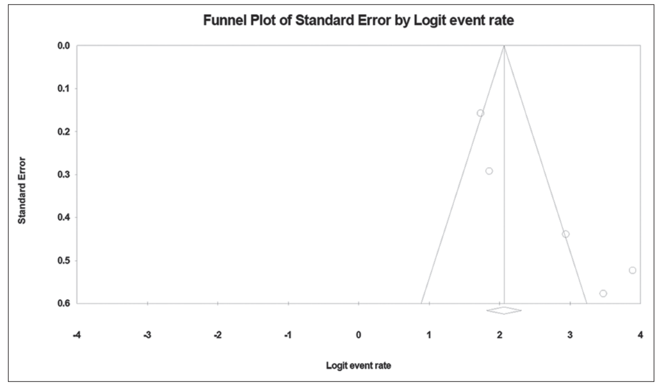

Figure 6: Funnel plot for AI performance as accuracy percentage (group 3)

The fail-safe N for this group was 361 (precision) and 369 (recall). Begg and Mazumdar Rank Correlation Test was performed and Kendall's tau b (corrected for ties, if any) was 0.70 (precision) 0.30 (recall). The combined performance of the third group of studies is described in [Table 7]. The pooled accuracy percentage was 83%. The funnel plot of the third group [Figure 6] is limited due to the small number of studies included in this group. Sensitivity analysis was not applicable because all variables were clear without missing values, and no assumptions were made, unlike systematic reviews of clinical trials.

Discussion

This study evaluates the performance of AI using CBCT images, which are three-dimensional (3D) images commonly used for diagnostic purposes of the head and neck. Interpreting CBCT images requires specialized knowledge and skills to manipulate the images and translate the findings into meaningful clinical data. This process is labor-intensive and time-consuming. Therefore, a pressing need to develop an automatic process is required to save time, improve clinician performance, and be seamlessly integrated into the workflow.

The performance of AI regarding the tasks of detection and segmentation of CBCT images is comparable to the works of trained dentists, in which a pooled performance measure is 0.85 (95%CI: 0.73,0.92), 0.88 (0.77,0.94), 0.93 (0.84, 0.97), 0.83 (0.68, 0.91) in studies using DICE/DSC, precision, recall, and accuracy percentage, respectively. The findings of this study agree with the results of numerous studies that examine the capabilities of AI for detection and segmentation. Hung *et al*.[3] investigated 50 studies that used AI for numerous clinical applications in dental and maxillofacial radiology. From their analysis of photographs, 2D, and 3D radiography, they concluded that the diagnostic performance of the AI models varies among different algorithms, although

Table 4: Performance of the AI model for studies in Group 1 that reported DICE/DSC. *Setzer F 2020 study was excluded because it was the only study that used DSC as an outcome measure for detection rather than segmentation and there was no other study that was used for detection

Model	Study name	Purpose	Statistics for each study					Event rate and 95% CI
			Event rate	Lower limit	Upper limit	*Z*	*P*	
	Zheng Z 2021	Segmentation	0.705	0.609	0.786	3.973	0.000	-1.00 -0.50 0.00 0.50 1.00
	Zheng Q 2021	Segmentation	0.878	0.822	0.918	8.666	0.000	
	Chen S 2020	Segmentation	0.800	0.680	0.883	4.295	0.000	
	Shujaat 2021	Segmentation	0.970	0.912	0.990	6.018	0.000	
	Wang H 2021	Segmentation	0.934	0.770	0.984	3.603	0.000	
	Wang X 2021	Segmentation	0.770	0.647	0.859	3.939	0.000	
	Jaskari J 2020	Segmentation	0.575	0.535	0.614	3.642	0.000	
	Leonardi 2021	Segmentation	0.967	0.838	0.994	3.816	0.000	
Random			0.848	0.731	0.920	4.691	0.000	

Table 5: Performance of the AI model for studies in group 2 that reported precision

Model	Group by Subgroup within	Study name	Statistics for each study					Event rate and 95% CI
			Event rate	Lower limit	Upper limit	*Z*	*P*	
	Detection	Orhan K 2020	0.950	0.889	0.978	6.700	0.000	-1.00 -0.50 0.00 0.50 1.00
	Detection	Lee K 2020	0.850	0.806	0.885	10.975	0.000	
Random	Detection		0.906	0.749	0.969	3.778	0.000	
	Segmentation	Zheng Z 2021	0.865	0.783	0.919	6.347	0.000	
	Segmentation	Shujaat 2021	0.970	0.912	0.990	6.018	0.000	
	Segmentation	Shaheen E 2021	0.980	0.946	0.993	7.431	0.000	
Random	Segmentation		0.953	0.834	0.988	4.211	0.000	
Random	Overall		0.929	0.842	0.970	5.600	0.000	

Table 6: Performance of the AI model for studies in group 3 that reported recalls

Model	Group by Subgroup within	Study name	Statistics for each study					Event rate and 95% CI
			Event rate	Lower limit	Upper limit	*Z*	*P*	
	Detection	Orhan K 2020	0.890	0.816	0.936	6.830	0.000	-1.00 -0.50 0.00 0.50 1.00
	Detection	Lee K 2020	0.840	0.795	0.877	10.772	0.000	
Random	Detection		0.857	0.802	0.899	8.944	0.000	
	Segmentation	Zheng Z 2021	0.820	0.732	0.884	5.826	0.000	
	Segmentation	Shujaat 2021	0.960	0.900	0.985	6.320	0.000	
	Segmentation	Shaheen E 2021	0.830	0.769	0.877	8.123	0.000	
Random	Segmentation		0.875	0.773	0.935	5.286	0.000	

the authors were unable to conduct a meta-analysis due to the heterogeneity of the studies. In the current study, we pooled the results because our research question was more focused, and we demonstrated that AI performance was excellent across different algorithms for detection and segmentation.

The detection tasks comprised of detection of periapical lesions,[9,19] temporomandibular joint (TMJ) osteoarthritis,[31] and impacted third molars.[32] Unfortunately, most of these detection studies failed to compare human intelligence with that of AI. Moreover, these studies failed to compare AI against an objective gold standard. The segmentation tasks included segmentation of pulp,[18] teeth,[26] jaws,[22] maxillae in cleft patients,[23] mandibular canal,[11] sinonasal cavity,[24] and pharyngeal airway. The comparative average DICE score for humans ranges between 0.97 and 0.98 for manual segmentation.[33] However, AI could match the performance through automation with less manual labor and in a shorter time.[22,34]

The internal validity of currently available studies is almost compromised owing to selection bias. Thus, assessing selection bias in these studies is vital because biased data can lead to algorithmic AI bias

Table 7: Performance of the AI model used for studies in Group 3 that reported accuracy percentage

Model	Study name	Statistics for each study					Event rate and 95% CI
		Event rate	Lower limit	Upper limit	*Z*	*P*	
	Orhan K 2021	0.862	0.755	0.927	5.094	0.000	-1.00 -0.50 0.00 0.50 1.00
	Setzer F 2020	0.930	0.705	0.987	2.952	0.000	
	Shoukri B 2019	0.735	0.682	0.782	7.707	0.000	
Random		0.826	0.680	0.914	3.800	0.000	

and compromise performance. Some authors randomly divided datasets into training, validation, and testing sets.[18] Nevertheless, this main sample selection is not blinded and not free from bias. Some studies used multivendor images[21,26] or external datasets[11,27] to reduce this bias but randomization of the selection process was not performed. Of the six core bias domains, "attrition bias" was excluded because it was not applicable in studies that use datasets, as it was only applicable for studies that used patients. In other words, dropping out of the study is not possible for datasets [Table 1].

In randomized controlled trials, performance bias is reduced by blinding participants and personnel. In AI studies, performance bias is unclear. Although computers are inherently unbiased, researchers can be selective by excluding datasets that have caries or restorations,[18] thereby significantly improving the performance of an AI model. Researchers can also manipulate the images to improve the performance of the AI algorithm. Image manipulation is conducted through normalization of parameters, pre-segmentation,[19] magnification, thresholding,[24] augmentation, flipping, zooming,[17] cropping,[22] reorientation,[23] rescaling,[11] which consequently introduce bias into the results. Detection bias was low since all included studies used computer-based detection software. In addition, all included studies reported the main outcome that was originally studied, therefore scored low on reporting bias.

Some limitations are identified in our meta-analysis. The studies included are not homogenous, assessing the performance of different tasks, however, regardless of the purpose for AI while using CBCT images, the reported performance was excellent across all tasks. Grey literature was not found and was excluded from the analysis. The results of all AI studies were positive, and no reported negative results were found, which could in itself be a form of bias. In addition to the risk of bias, lack of ground truth, relying on expert opinion in studies testing detection, and manual segmentation was the main limitation across all studies. However, this is currently the best noninvasive method to test AI. Future studies that use multivendor images and external datasets with minimal preparation of images are recommended to minimize algorithmic bias. Privacy about exporting CBCT DICOM images into the AI training model with embedded patient identifiers must be addressed by researchers through obtaining approval from an Institutional Review Board and strictly adhering to national standards that protect sensitive patient health information.

Conclusion

The application of AI for detection and segmentation using CBCT images is comparable to that of trained dentists with the potential to enhance and expedite the interpretive process. AI can analyze a large number of studies and flag ones with significant findings, increasing clinical efficiency. Future studies can focus on the ability of AI to recognize connections between imaging and clinical findings that may be oblivious to us humans thus improving patient care.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Heo MS, Kim JE, Hwang JJ, Han SS, Kim JS, Yi WJ, *et al.* Artificial intelligence in oral and maxillofacial radiology: What is currently possible? Dentomaxillofac Radiol 2021;50:20200375.
2. Hassan C, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, *et al.* Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: A systematic review and meta-analysis. Gastrointest Endosc 2021;93:77-85.e6.
3. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. Dentomaxillofac Radiol 2020;49:20190107.
4. Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schürmann F, *et al.* Caries Detection with Near-Infrared Transillumination Using Deep Learning. J Dent Res 2019;98:1227-33.
5. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent 2018;77:106-11.
6. Thanathornwong B, Suebnukarn S. Automatic detection of periodontal compromised teeth in digital panoramic radiographs

using faster regional convolutional neural networks. Imaging Sci Dent 2020;50:169-74.

7. Kim J, Lee HS, Song IS, Jung KH. DeNTNet: Deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs. Sci Rep 2019;9:17615.
8. Fukuda M, Inamoto K, Shibata N, Ariji Y, Yanashita Y, Kutsuna S, *et al.* Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. Oral Radiol 2020;36:337-43.
9. Orhan K, Bayrakdar IS, Ezhov M, Kravtsov A, Ozyurek T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. Int Endod J 2020;53:680-9.
10. Kwon O, Yong TH, Kang SR, Kim JE, Huh KH, Heo MS, *et al.* Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. Dentomaxillofac Radiol 2020;49:20200185.
11. Jaskari J, Sahlsten J, Jarnstedt J, Mehtonen H, Karhu K, Sundqvist O, *et al.* Deep Learning method for mandibular canal segmentation in dental cone beam computed tomography volumes. Sci Rep 2020;10:5842.
12. Schepman A, Rodway P. Initial validation of the general attitudes towards Artificial Intelligence Scale. Comput Hum Behav Rep 2020;1:100014.
13. Umer F, Khan M. A call to action: Concerns related to artificial intelligence. Oral Surg Oral Med Oral Pathol Oral Radiol 2021;132:255.
14. Keskinbora KH. Medical ethics considerations on artificial intelligence. J Clin Neurosci 2019;64:277-82.
15. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1.
16. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.
17. Zheng Z, Yan H, Setzer FC, Shi KJ, Mupparapu M, Li J. Anatomically constrained deep learning for automating dental CBCT segmentation and lesion detection. IEEE Transactions on Automation Science and Engineering 2021;18:603-14.
18. Zheng Q, Ge Z, Du H, Li G. Age estimation based on 3D pulp chamber segmentation of first molars from cone-beam-computed tomography by integrated deep learning and level set. Int J Legal Med 2021;135:365-73.
19. Setzer FC, Shi KJ, Zhang Z, Yan H, Yoon H, Mupparapu M, *et al.* Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. J Endod 2020;46:987-93.
20. Chen S, Wang L, Li G, Wu TH, Diachina S, Tejera B, *et al.* Machine learning in orthodontics: Introducing a 3D auto-segmentation and auto-landmark finder of CBCT images to assess maxillary constriction in unilateral impacted canine patients. Angle Orthod 2020;90:77-84.
21. Shujaat S, Jazil O, Willems H, Van Gerven A, Shaheen E, Politis C, *et al.* Automatic segmentation of the pharyngeal airway space with convolutional neural network. J Dent 2021;111:103705.
22. Wang H, Minnema J, Batenburg KJ, Forouzanfar T, Hu FJ, Wu G. Multiclass CBCT image segmentation for orthodontics with deep learning. J Dent Res 2021;100:943-9.
23. Wang X, Pastewait M, Wu TH, Lian C, Tejera B, Lee YT, *et al.* 3D morphometric quantification of maxillae and defects for patients with unilateral cleft palate via deep learning-based CBCT image auto-segmentation. Orthod Craniofac Res 2021;24(Suppl 2):108-16.
24. Leonardi R, Lo Giudice A, Farronato M, Ronsivalle V, Allegrini S, Musumeci G, *et al.* Fully automatic segmentation of sinonasal cavity and pharyngeal airway based on convolutional neural networks. Am J Orthod Dentofacial Orthop 2021;159:824-35.e1.
25. Lee JH, Kim DH, Jeong SN. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. Oral Dis 2020;26:152-8.
26. Shaheen E, Leite A, Alqahtani KA, Smolders A, Van Gerven A, Willems H, *et al.* A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. A validation study. J Dent 2021:103865.
27. Shoukri B, Prieto J, Ruellas A. Minimally Invasive Approach for Diagnosing TMJ Osteoarthritis. J Dent Res 2019;98:1103-11.
28. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A guide for authors and reviewers. Radiol Artif Intell 2020;2:e200029.
29. Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. J Med Imaging (Bellingham) 2018;5:015006.
30. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Med Imaging 2015;15:29.
31. Lee KS, Kwak HJ, Oh JM, Jha N, Kim YJ, Kim W, *et al.* Automated detection of TMJ osteoarthritis based on artificial intelligence. J Dent Res 2020;99:1363-7.
32. Orhan K, Bilgir E, Bayrakdar IS, Ezhov M, Gusarev M, Shumilov E. Evaluation of artificial intelligence for detecting impacted third molars on cone-beam computed tomography scans. J Stomatol Oral Maxillofac Surg 2021;122:333-7.
33. Weston AD, Philbrick KR, Conte GM, Boonrod A, Cai J, A Z, *et al.* How accurate is human segmentation? A comparison of 14 human tracers performing whole abdomen segmentation of 27 organs. Paper presented at: Conference on Machine Intelligence in Medical Imaging, Austin, TX, September 22-23, 2019.
34. Duan W, Chen Y, Zhang Q, Lin X, Yang X. Refined tooth and pulp segmentation using U-Net in CBCT image. Dentomaxillofac Radiol 2021;50:20200251.