


Review

# Evaluating the Checklist for Artificial Intelligence in Medical Imaging (CLAIM)-Based Quality of Reports Using Convolutional Neural Network for Odontogenic Cyst and Tumor Detection

Van Nhat Thang Le <sup>1,2,3,4</sup> , Jae-Gon Kim <sup>1,2,3</sup>, Yeon-Mi Yang <sup>1,2,3</sup> and Dae-Woo Lee <sup>1,2,3,\*</sup>

- <sup>1</sup> Department of Pediatric Dentistry, Institute of Oral Bioscience, School of Dentistry, Jeonbuk National University, Jeonju 54896, Korea; lvnthang@hueuni.edu.vn (V.N.T.L.); pedokjg@jbnu.ac.kr (J.-G.K.); pedo1997@jbnu.ac.kr (Y.-M.Y.)
- <sup>2</sup> Research Institute of Clinical Medicine, Jeonbuk National University, Jeonju 54907, Korea
- <sup>3</sup> Biomedical Research Institute, Jeonbuk National University Hospital, Jeonju 54907, Korea
- <sup>4</sup> Faculty of Odonto-Stomatology, Hue University of Medicine and Pharmacy, Hue University, Hue 52000, Vietnam
- \* Correspondence: oklee@jbnu.ac.kr; Tel.: +82-63-250-2826



**Citation:** Le, V.N.T.; Kim, J.-G.; Yang, Y.-M.; Lee, D.-W. Evaluating the Checklist for Artificial Intelligence in Medical Imaging (CLAIM)-Based Quality of Reports Using Convolutional Neural Network for Odontogenic Cyst and Tumor Detection. *Appl. Sci.* **2021**, *11*, 9688. <https://doi.org/10.3390/app11209688>

Academic Editors: Seongyong Moon, Yong-Dae Kwon and Kyoobin Lee

Received: 30 August 2021

Accepted: 14 October 2021

Published: 18 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** This review aimed to explore whether studies employing a convolutional neural network (CNN) for odontogenic cyst and tumor detection follow the methodological reporting recommendations, the checklist for artificial intelligence in medical imaging (CLAIM). We retrieved the CNN studies using panoramic and cone-beam-computed tomographic images from inception to April 2021 in PubMed, EMBASE, Scopus, and Web of Science. The included studies were assessed according to the CLAIM. Among the 55 studies yielded, 6 CNN studies for odontogenic cyst and tumor detection were included. Following the CLAIM items, abstract, methods, results, discussion across the included studies were insufficiently described. The problem areas included item 2 in the abstract; items 6–9, 11–18, 20, 21, 23, 24, 26–31 in the methods; items 33, 34, 36, 37 in the results; item 38 in the discussion; and items 40–41 in “other information.” The CNN reports for odontogenic cyst and tumor detection were evaluated as low quality. Inadequate reporting reduces the robustness, comparability, and generalizability of a CNN study for dental radiograph diagnostics. The CLAIM is accepted as a good guideline in the study design to improve the reporting quality on artificial intelligence studies in the dental field.

**Keywords:** odontogenic cyst; odontogenic tumor; convolutional neural network; medical imaging; methodological quality evaluation

## 1. Introduction

Advances in digital dentistry, along with rapid developments in diagnostic artificial intelligence (AI), have the potential to improve diagnostic accuracy. In addition, AI-based applications can assist dentists in making timely interventions and increase their working performance. Applications of AI in dentistry include detection, segmentation, and classification of anatomy (tooth, root morphology, and mandible) and pathology (caries, periodontal inflammation, and osteoporosis) [1,2].

In the last decade, deep-learning methods such as the convolutional neural network (CNN) have been demonstrated to achieve remarkable results on panoramic and cone-beam-computed tomographic (CBCT) images [3–5]. Consequently, an increasing number of studies are employing the CNN framework. Indeed, most studies for the automated detection of odontogenic cysts and tumors are based on this framework and achieved high performance [6–8]. However, AI has several challenges in terms of robustness, comparability, and generalizability in medical imaging.

In the medical field, several checklists are applied to report the evaluation of machine learning models [9–12]; these include the Standards for Reporting of Diagnostic Accuracy Studies (STARD) [13–16], Consolidated Standards of Reporting Trials (CONSORT)—AI extension [17], and the AI checklist in dental research [18]. Recently, the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) has been developed based on the consensus of radiological experts and is viewed as the best guideline for presenting research [19]. To the best of our knowledge, no previous systematic review has evaluated the methodological quality among studies on AI in dentistry. Therefore, this systematic review was methodologically performed on available studies using CNN for automated detection of odontogenic cysts and tumors to determine if the reports adequately adhered to the items of the CLAIM guideline.

## 2. Materials and Methods

### 2.1. Inclusion and Exclusion Criteria

All of the reports employing the CNN model to examine the performance of automated detection of odontogenic cysts and tumors on panoramic and CBCT images were eligible. We excluded methodological reviews, studies not employing CNN, studies unrelated to the topic, and studies not involving humans.

### 2.2. Information Sources and Search Strategy

#### 2.2.1. Electronic Search

A comprehensive literature search was conducted on electronic databases, including PubMed, EMBASE, Scopus, and Web of Science, from inception to 18 April 2021. The search strategy was a combination of MeSH (medical subject heading) terms and free text words, including “deep learning” (MeSH Terms) OR deep learning (Text Word) OR convolution neural network (Text Word) AND (“odontogenic tumors” (MeSH Terms) OR odontogenic tumor (Text Word) OR “odontogenic cysts” (MeSH Terms) OR odontogenic cysts (Text Word)). In this review, we provided the detailed search strategy in Supplementary Table S1. In addition, there was no language restriction in this review.

#### 2.2.2. Manual Searching

In addition to searching electronic databases, the list of bibliographic references of the included studies was screened to identify potentially relevant additional studies. Furthermore, we also searched opengrey.eu from inception to April 2021 for eligible studies in grey literature.

### 2.3. Study Selection

The title and abstract of each of the identified studies were independently screened by two reviewers (V.N.T.L. and D.-W.L.) to discard duplicates and studies that did not satisfy the inclusion criteria. After, the full-text articles were examined when information was provided insufficiently in the abstract. A third reviewer (Y.-M.Y.) resolved any disagreement during this process. Full-text articles that satisfied the inclusion criteria were independently assessed by two reviewers (V.N.T.L. and D.-W.L.) with clinical knowledge of odontogenic cysts and tumors and methodological knowledge of AI research.

### 2.4. Data Extraction

Two reviewers (V.N.T.L. and D.-W.L.) independently extracted the data from each included article into predesigned data collection forms on Microsoft Word: (1) General characteristics (primary author, country, date of publication, journal name); (2) Specific characteristics (studies objectives, dataset, CNN model, comparative analysis, outcome metrics, and performance). Discrepancies were resolved by discussion with a third reviewer (Y.-M.Y.).

## 2.5. Methods of Analysis

### 2.5.1. Reporting Epidemiological and Descriptive Characteristics

Among the included CNN studies, epidemiological and descriptive characteristics were assessed for journal category, location and job of corresponding author, guideline for reporting, and funding source.

### 2.5.2. Reporting of Methodological Elements of the Included CNN Studies

This systematic review was performed based on the CLAIM guideline [19], which includes 42 items. According to this guideline, we examined whether methodological elements were reported in the included CNN studies.

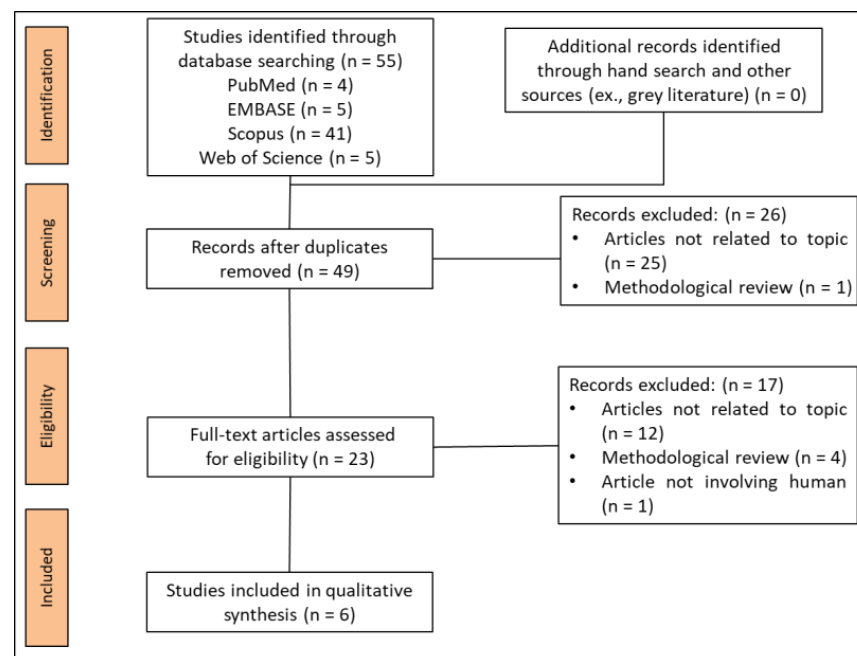
### 2.5.3. Statistical Analysis

Regarding categorical data, numbers (percentages) are used to summarize descriptive statistics. Among the included studies, absolute and relative frequencies are used to summarize the information extracted from the CLAIM items.

## 3. Results

### 3.1. Study Selection

The search strategy yielded a total of 55 studies from electronic databases and manual searching. After removing duplication, 49 studies were selected, of which 26 were removed after filtering the title and abstract. Finally, 23 studies were assessed for eligibility by full-text review. At this stage, the studies were excluded for some reasons, such as methodological review ( $n = 4$ ), articles unrelated to the topic ( $n = 12$ ), and articles not involving human participants ( $n = 1$ ). Finally, six reports were included in the systematic review [6–8,20–22] (Figure 1).



**Figure 1.** Flowchart summarizing the article-selection process.

### 3.2. Study Characteristics

#### 3.2.1. Epidemiological and Descriptive Characteristics

As presented in Table 1, six CNN studies were published in six journals, including those in the biomedical engineering field ( $n = 2$ , 33%) [20,22] and the dental or medical field ( $n = 4$ , 67%) [6–8,21]. In addition, all corresponding authors were located in Asia

(n = 6, 100%) [6–8,20–22]. The job of the corresponding author was that of a doctor, dentist (n = 6, 86%) [6–8,20–22], or engineer (n = 1, 14%) [20]. None of the CNN studies used the reporting guideline [6–8,20–22]. Regarding the source of funding, the majority of the studies were funded by public support (n = 4, 67%) [6,7,20,21]; the others had no funding (n = 2, 33%) [8,22].

**Table 1.** Descriptive characteristics of six reports using deep learning neural network for automated detection of cyst and tumor of the jaw.

Items and Subcategory	No. (%) of Reports
<i>Journal Category</i>	
Biomedical engineering field	2 (33%)
Dental or medical field	4 (67%)
<i>Location of corresponding author</i>	
Asia	6 (100%)
Europe	0 (0%)
USA	0 (0%)
<i>Job of corresponding author *</i>	
Doctor or dentist	6 (86%)
Engineer	1 (14%)
<i>Type of reporting guideline</i>	
STARD	0 (100%)
Other	0 (100%)
None	6 (100%)
<i>Funding source</i>	
Both private and public	0 (0%)
Private	0 (0%)
Public	4 (67%)
None	2 (33%)
Unclear	0 (0%)

\* The total does not equal 100% because multiple answers were possible in each study.

### 3.2.2. General Characteristics

The main characteristics of each included study are summarized in Table 2. The publication year of the included studies ranged from 2018 to 2021. In regard to study objectives, detection and classification were performed on five studies (83%) [6–8,21,22], and only classification was performed on one study (17%) [20]. All studies (n = 6, 100%) used private datasets for experiments [6–8,20–22]. For CNN architecture, all studies (n = 6, 100%) used transfer learning for training and testing [6–8,20–22]. In addition, the comparators were a radiologist (n = 1, 17%) [20], oral maxillofacial surgeons (n = 2, 33%) [21,22], a general practitioner (n = 1, 17%) [21], and not reported (n = 3, 50%) [6–8]. To assess the performance of the CNN models, the outcomes of the included studies were used, including sensitivity (n = 5, 83%) [6–8,20,22], specificity (n = 4, 67%) [6,7,20,22], accuracy (n = 4, 67%) [6,20–22], area under the curve (n = 3, 50%) [6,7,20], F1 score (n = 1, 17%) [21], precision (n = 1, 17%) [21], recall (n = 1, 17%) [21], false positive rate (n = 1, 17%) [8], and diagnostic time (n = 1, 17%) [22].

Table 2. Study characteristics of the included studies.

#	Study	Country (Year)	Journal	Study Objectives	Number of Images	Annotators	CNN Model	Comparative Analysis	Outcome Metrics	CNN Performance
1	Liu et al.	China (2021)	<i>International Journal of Computer Assisted Radiology and Surgery</i>	Classification	420 panoramic images: AM (209), OKC (211), Training (295), validation (42), and test (83)	Histopathologic diagnosis	VGG-19 and ResNet-50	Radiologists	Sensitivity, specificity, accuracy, and AUC	Sensitivity (92.88%), specificity (87.8%), accuracy (90.36%), and AUC (0.946)
2	Kwon et al.	South Korea (2020)	<i>Dentomaxillofacial Radiology</i>	Detection and classification	1282 maxillary and mandibular panoramic images: DC (350), periapical cyst (302), OKC (300), AM (230), no lesion (100) Training (946) and test (235)	Histopathologic diagnosis	A modified CNN from the YOLO v3	NR	Sensitivity, specificity, accuracy, and AUC	Sensitivity (88.9%), specificity (97.2%), accuracy (95.6%), and AUC (0.94)
3	Yang et al.	South Korea (2020)	<i>Journal of Clinical Medicine</i>	Detection and classification	1603 maxillary and mandibular panoramic images: DC (1094), OKC (316), AM (160), no lesion (33) Training (1422) and test (181)	Histopathologic diagnosis	YOLO v2	OMFS (3), general practitioner (2)	Precision, recall, accuracy, and F1 score	Precision (0.707), recall (0.68), accuracy (0.663), and F1 score (0.693)
4	Ariji et al.	Japan (2019)	<i>Oral Surgery Oral Medicine Oral Pathology Oral Radiology</i>	Detection and classification	285 mandibular panoramic images: AM (41), OKC (47), DC (90), radicular cyst (91), simple bone cyst (16) Training (21), test1 (50), test2 (25)	Histopathologic diagnosis	DIGITS using deep neural network Detect Net	NR	Sensitivity and false positive using IOU (threshold 0.6)	Detection of radiolucent lesions: sensitivity (0.88), false-positive rate per image for test1 (0.00) and test2 (0.04) Detection and classification sensitivity of each type of lesion using test1: AM (0.71 and 0.6), OKC (1 and 0.13), DC (0.88 and 0.82), and radicular cysts (0.81 and 0.82)
5	Lee et al.	South Korea (2019)	<i>Oral Diseases</i>	Detection and classification	1140 panoramic and 986 CBCT images: OKC (260 + 188), DC (463 + 396), periapical cyst (417 + 402)	Histopathologic diagnosis	Google Net inception v3	NR	AUC, sensitivity, and specificity	CBCT: AUC (0.914), sensitivity (96.1%), specificity (77.1%) Panoramic images: AUC (0.847), sensitivity (88.2%), specificity (77%)
6	Poedjiastoeti et al.	Thailand (2018)	<i>Health Informatics Research</i>	Detection and classification	500 panoramic images: AM (250), OKC (250) Training (400) and test (100)	Histopathologic diagnosis	16-layer CNN (VGG-16)	OMFS (5)	Sensitivity, specificity, accuracy, and diagnostic time	Sensitivity (81.8%), specificity (83.3%), accuracy (83%), and diagnostic time (38 s)

Abbreviations. AM: ameloblastoma; AUC: area under the curve; CBCT: cone-beam-computed tomography; CNN: convolutional neural network; DC: dentigerous cyst; DIGITs: deep learning GPU training system; NR: not reported; OKC: odontogenic keratocyst; OMFS: oral maxillofacial surgeon; YOLO: you only look once.

### 3.3. Synthesis of the Results

#### Reporting of CLAIM Items across the Included Studies

In all studies, 42 methodological items are reported (Supplementary Table S2). In the abstract section, two studies (33%) did not present a structured summary of the study design, methods, results, and conclusions (item 2) [7,20].

In the methods section, no methodological components were reported from the included CNN studies, including the study goal; item 6, model creation, exploratory study, feasibility study, noninferiority trial of study design (n = 6, 100%) [6–8,20–22]; item 7, data source (n = 1, 17%) [20]; item 8, eligibility criteria (n = 3, 50%) [6,20,22]; item 9, data-processing steps (n = 1, 17%) [21]; item 11, definitions of data elements, with reference to common data elements (n = 6, 100%) [6–8,20–22]; item 12, de-identification methods (n = 6, 100%) [6–8,20–22]; item 13, how missing data were handled (n = 6, 100%) [6–8,20–22]; item 14, definition of the ground-truth reference standard (n = 6, 100%) [6–8,20–22]; item 15, rationale for choosing the reference standard (n = 6, 100%) [6–8,20–22]; item 16, source of ground-truth annotations, qualifications and preparation of annotators (n = 3, 50%) [8,20,21]; item 17, annotation tools (n = 4, 67%) [7,20–22]; item 18, measurement of inter- and intra-rater variability (n = 6, 100%) [6–8,20–22]; item 20, how data were assigned to partitions, specify proportions (n = 1, 17%) [8]; item 21, level at which the partitions are disjointed (n = 6, 100%) [6–8,20–22]; item 23, software, libraries, frameworks, and packages (n = 2, 33%) [6,22]; item 24, initialization of model parameters (n = 3, 50%) [6,8,21]; item 26, method of selecting the final model (n = 5, 87%) [7,8,20–22]; item 27, ensembling techniques (n = 4, 67%) [6–8,21]; item 28, metrics of model performance (n = 1, 17%) [20]; item 29, statistical measure of significant and uncertainty (n = 3, 50%) [6,8,20]; item 30, robustness or sensitivity analysis (n = 2, 33%) [20,21]; and item 31, methods of explainability or interpretability (n = 5, 87%) [6,7,20–22].

In the results section, 5 (83%) studies did not report the flow of the participants or cases (item 33) [6–8,20,22]. In addition, 3 (50%) studies did not present the demographic and clinical characteristics of cases (item 34) [6,20,22]. Moreover, only 1 (17%) study did not report the estimates of diagnostic accuracy and their precision (item 36) [8] and failure analysis of incorrectly classified cases (item 37) [20]. In the discussion section, only 1 (17%) study did not report the study limitations (item 38) [20]. Regarding the other information, none of the studies registered the number and name of the registry (item 40). Moreover, none of the CNN studies presented where the full study protocol can be accessed (item 41) [6–8,20–22]. In addition, 2 (33%) studies did not show sources of funding and other support (item 42) [8,22].

## 4. Discussion

In our systematic review, the included CNN studies only improved the model performance for automated odontogenic cyst and tumor detection. This can help to reduce morbidity and mortality through long-term follow-up and early intervention. However, the application of AI must remain grounded in the fundamental tenets of science and scientific publication, which are evident in the design and reporting.

To examine the level of compliance with design and reporting standards, we evaluated six reports employing CNN for odontogenic cyst and tumor detection based on the CLAIM. Recently, CLAIM was used as evaluation guidance in the design and reporting of CNN studies for brain metastasis detection [23], knee imaging [24], and radiological cancer diagnosis [25] in the medical field. In our study, none of the CNN reports followed any previous reporting guidelines. After evaluation, the methodological reporting recommendations of the CLAIM guideline were missing in most CNN studies. Among the included CNN studies, we found a lack of adherence to the standards of CLAIM in the abstract, methods (study design, data, ground truth, data partitions, model, training, and evaluation), results, discussion, and supplementary sections. These findings indicate that the robustness, comparability, and generalizability of the CNN studies for the automated detection of odontogenic cysts and tumors are not guaranteed. Consequently, the reporting

quality related to the AI application for medical imaging must be improved for a clear, transparent, and reproducible CNN study.

Among the included studies, high heterogeneity of study design can influence the robustness, comparability, and generalizability of the CNN studies for automated odontogenic cyst and tumor detection. Regarding the sample characteristics, the location and category of lesions were inconsistent in the datasets. Furthermore, private datasets were different in size. Therefore, a benchmark dataset should be required to solve these issues. In addition to sample issues, comparators should be consistent across studies to reduce the bias of datasets. Especially, outcome measurement should be standardized to improve the comparability of the model performance. In general, these issues usually occur in novel studies; in particular, deep learning is an emerging approach, and the included studies have only been published in the last three years.

From previous studies, the quality of “AI for health” studies remains low, and reporting is often insufficient to fully comprehend and possibly replicate these studies [26–28]. In dental and oral sciences, the emergence of standards towards reporting is necessary given the increasing number of recent CNN studies [1]. CNN is one type of deep learning algorithm that is used in many branches of computer vision dealing with medical image analysis and represents a future computer-aided technology for medical and dental experts [29]. To improve the performance of future CNN studies, authors should examine assumptions in greater detail and report valid and adequate items following the CLAIM guidelines. Regarding medical imaging, the CLAIM is the best guideline for presenting research and is relatively new; this guideline should be applied widely to improve the reporting of AI research.

#### *Strengths and Limitations*

Our review is the first to investigate reporting quality of CNN studies for the automated detection of odontogenic cysts and tumors. However, our study has some limitations. From the reader’s position, we only evaluated the CNN reports for the automated detection of odontogenic cysts and tumors. Moreover, we realized that the AI researchers may have omitted or removed important details during publication despite using the proper methods. Further studies should be performed to compare our reported results with those of CONSORT-AI extensions and the STARD checklist to ascertain the reliability of the results. In addition, we did not investigate other dental issues because we only intended to evaluate the quality of CNN reports for odontogenic cyst and tumor detection. However, we recommend that the CLAIM should be considered as the best framework to help AI researchers for reporting any issue in dentistry.

## 5. Conclusions

This review revealed that the CLAIM-based quality of CNN reports for odontogenic cyst and tumor detection is considered low-level. Performing a CNN study with insufficient reporting raises the likelihood of producing invalid results. Therefore, the CLAIM is accepted as a good guideline in the study design to help authors writing AI manuscripts in dentistry.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/app11209688/s1>. Table S1: Detailed search strategies for each database. Mesh terms, search terms, and combinations of the two were used for each database search. Table S2: Evaluating the CLAIM-based quality of CNN reports for odontogenic cyst and tumor detection.

**Author Contributions:** Conceptualization, V.N.T.L. and D.-W.L.; methodology, V.N.T.L. and D.-W.L.; validation, Y.-M.Y.; investigation, Y.-M.Y.; resources, J.-G.K.; data curation, J.-G.K.; writing—original draft preparation, V.N.T.L., Y.-M.Y. and D.-W.L.; writing—review and editing, V.N.T.L., Y.-M.Y. and D.-W.L.; visualization, Y.-M.Y. and J.-G.K.; supervision, V.N.T.L. and D.-W.L.; project administration, V.N.T.L. and D.-W.L.; funding acquisition, V.N.T.L. and D.-W.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2020R1F1A1072484). This study was also supported by the Fund of Biomedical Research Institute, Jeonbuk National University Hospital.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Schwendicke, F.; Golla, T.; Dreher, M.; Krois, J. Convolutional neural networks for dental image diagnostics: A scoping review. *J. Dent.* **2019**, *91*, 103226. [[CrossRef](#)] [[PubMed](#)]
2. Shan, T.; Tay, F.R.; Gu, L. Application of artificial intelligence in dentistry. *J. Dent. Res.* **2021**, *100*, 232–244. [[CrossRef](#)] [[PubMed](#)]
3. Wang, H.; Minnema, J.; Batenburg, K.J.; Forouzanfar, T.; Hu, F.J.; Wu, G. Multiclass CBCT image segmentation for orthodontics with deep learning. *J. Dent. Res.* **2021**, *100*, 943–949. [[CrossRef](#)] [[PubMed](#)]
4. Jeon, S.J.; Yun, J.P.; Yeom, H.G.; Shin, W.S.; Lee, J.H.; Jeong, S.H.; Seo, M.S. Deep-learning for predicting C-shaped canals in mandibular second molars on panoramic radiographs. *Dentomaxillofac. Radiol.* **2021**, *50*, 20200513. [[CrossRef](#)]
5. Caliskan, S.; Tuloglu, N.; Celik, O.; Ozdemir, C.; Kizilaslan, S.; Bayrak, S. A pilot study of a deep learning approach to submerged primary tooth classification and detection. *Int. J. Comput. Dent.* **2021**, *24*, e1–e9.
6. Kwon, O.; Yong, T.H.; Kang, S.R.; Kim, J.E.; Huh, K.H.; Heo, M.S.; Lee, S.S.; Choi, S.C.; Yi, W.J. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. *Dentomaxillofac. Radiol.* **2020**, *49*, 20200185. [[CrossRef](#)] [[PubMed](#)]
7. Lee, J.H.; Kim, D.H.; Jeong, S.N. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral Dis.* **2020**, *26*, 152–158. [[CrossRef](#)] [[PubMed](#)]
8. Ariji, Y.; Yanashita, Y.; Kutsuna, S.; Muramatsu, C.; Fukuda, M.; Kise, Y.; Nozawa, M.; Kuwada, C.; Fujita, H.; Katsumata, A.; et al. Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2019**, *128*, 424–430. [[CrossRef](#)]
9. Luo, W.; Phung, D.; Tran, T.; Gupta, S.; Rana, S.; Karmakar, C.; Shilton, A.; Yearwood, J.; Dimitrova, N.; Ho, T.B.; et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J. Med. Internet Res.* **2016**, *18*, e323. [[CrossRef](#)]
10. Handelman, G.S.; Kok, H.K.; Chandra, R.V.; Razavi, A.H.; Huang, S.; Brooks, M.; Lee, M.J.; Asadi, H. Peering into the black box of artificial intelligence: Evaluation metrics of machine learning methods. *AJR Am. J. Roentgenol.* **2019**, *212*, 38–43. [[CrossRef](#)]
11. Park, S.H.; Han, K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology* **2018**, *286*, 800–809. [[CrossRef](#)]
12. Bluemke, D.A.; Moy, L.; Bredella, M.A.; Ertl-Wagner, B.B.; Fowler, K.J.; Goh, V.J.; Halpern, E.F.; Hess, C.P.; Schiebler, M.L.; Weiss, C.R. Assessing radiology research on artificial intelligence: A brief guide for authors, reviewers, and readers-from the radiology editorial board. *Radiology* **2020**, *294*, 487–489. [[CrossRef](#)]
13. Bossuyt, P.M.; Reitsma, J.B.; Bruns, D.E.; Gatsonis, C.A.; Glasziou, P.P.; Irwig, L.; Lijmer, J.G.; Moher, D.; Rennie, D.; de Vet, H.C.; et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Radiology* **2015**, *277*, 826–832. [[CrossRef](#)]
14. Bossuyt, P.M.; Reitsma, J.B.; Bruns, D.E.; Gatsonis, C.A.; Glasziou, P.P.; Irwig, L.M.; Lijmer, J.G.; Moher, D.; Rennie, D.; de Vet, H.C. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Radiology* **2003**, *226*, 24–28. [[CrossRef](#)]
15. Cohen, J.F.; Korevaar, D.A.; Altman, D.G.; Bruns, D.E.; Gatsonis, C.A.; Hooft, L.; Irwig, L.; Levine, D.; Reitsma, J.B.; de Vet, H.C.; et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open* **2016**, *6*, e012799. [[CrossRef](#)]
16. Bossuyt, P.M.; Reitsma, J.B. The STARD initiative. *Lancet* **2003**, *361*, 71. [[CrossRef](#)]
17. Schwendicke, F.; Krois, J. Better reporting of studies on artificial intelligence: CONSORT-AI and beyond. *J. Dent. Res.* **2021**, *100*, 677–680. [[CrossRef](#)] [[PubMed](#)]
18. Schwendicke, F.; Singh, T.; Lee, J.H.; Gaudin, R.; Chaurasia, A.; Wiegand, T.; Uribe, S.; Krois, J. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *J. Dent.* **2021**, *107*, 103610. [[CrossRef](#)] [[PubMed](#)]
19. Mongan, J.; Moy, L.; Kahn, C.E., Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A guide for authors and reviewers. *Radiol. Artif. Intell.* **2020**, *2*, e200029. [[CrossRef](#)]
20. Liu, Z.; Liu, J.; Zhou, Z.; Zhang, Q.; Wu, H.; Zhai, G.; Han, J. Differential diagnosis of ameloblastoma and odontogenic keratocyst by machine learning of panoramic radiographs. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 415–422. [[CrossRef](#)]
21. Yang, H.; Jo, E.; Kim, H.J.; Cha, I.H.; Jung, Y.S.; Nam, W.; Kim, J.Y.; Kim, J.K.; Kim, Y.H.; Oh, T.G.; et al. Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs. *J. Clin. Med.* **2020**, *9*, 1839. [[CrossRef](#)] [[PubMed](#)]



22. Poedjiastoeti, W.; Suebnukarn, S. Application of convolutional neural network in the diagnosis of jaw tumors. *Healthc. Inform. Res.* **2018**, *24*, 236–241. [[CrossRef](#)] [[PubMed](#)]
23. Cho, S.J.; Sunwoo, L.; Baik, S.H.; Bae, Y.J.; Choi, B.S.; Kim, J.H. Brain metastasis detection using machine learning: A systematic review and meta-analysis. *Neuro-Oncol.* **2021**, *23*, 214–225. [[CrossRef](#)]
24. Si, L.; Zhong, J.; Huo, J.; Xuan, K.; Zhuang, Z.; Hu, Y.; Wang, Q.; Zhang, H.; Yao, W. Deep learning in knee imaging: A systematic review utilizing a Checklist for Artificial Intelligence in Medical Imaging (CLAIM). *Eur. Radiol.* **2021**. [[CrossRef](#)] [[PubMed](#)]
25. O’Shea, R.J.; Sharkey, A.R.; Cook, G.J.R.; Goh, V. Systematic review of research design and reporting of imaging studies applying convolutional neural networks for radiological cancer diagnosis. *Eur. Radiol.* **2021**, *31*, 7969–7983. [[CrossRef](#)]
26. Liu, X.; Faes, L.; Kale, A.U.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdas, M.; Kern, C.; et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *Lancet Digit. Health* **2019**, *1*, e271–e297. [[CrossRef](#)]
27. Nagendran, M.; Chen, Y.; Lovejoy, C.A.; Gordon, A.C.; Komorowski, M.; Harvey, H.; Topol, E.J.; Ioannidis, J.P.A.; Collins, G.S.; Maruthappu, M. Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* **2020**, *368*, m689. [[CrossRef](#)]
28. Wynants, L.; Smits, L.J.M.; Van Calster, B. Demystifying AI in healthcare. *BMJ* **2020**, *370*, m3505. [[CrossRef](#)]
29. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)]