







# The efficiency of artificial intelligence methods for finding radiographic features in different endodontic treatments - a systematic review

Shaqayeq Ramezanzade<sup>a</sup> , Tudor Laurentiu<sup>b</sup> , Azam Bakhshandah<sup>a</sup> , Bulat Ibragimov<sup>b</sup> ,  
Thomas Kvist<sup>c</sup> , EndoReCo\* and Lars Bjørndal<sup>a</sup> 

<sup>a</sup>Department of Cariology and Endodontics, Section for Clinical Oral Microbiology, Faculty of Health and Medical Sciences, Department of Odontology, University of Copenhagen, Copenhagen, Denmark; <sup>b</sup>Department of Computer Science, University of Copenhagen, Copenhagen, Denmark; <sup>c</sup>Department of Endodontology, Institute of Odontology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

## ABSTRACT

**Objectives:** To assess the efficiency of AI methods in finding radiographic features in Endodontic treatment considerations.

**Material and methods:** This review was based on the PRISMA guidelines and QUADAS 2 tool. A systematic search was performed of the literature on cases with endodontic treatments, comparing AI algorithms (test) versus conventional image assessments (control) for finding radiographic features. The search was conducted in PubMed, Scopus, Google Scholar and the Cochrane library. Inclusion criteria were studies on the use of AI and machine learning in endodontic treatments using dental X-rays.

**Results:** The initial search retrieved 1131 papers, from which 24 were included. High heterogeneity of the materials left out a meta-analysis. The reported subcategories were periapical lesion, vertical root fractures, predicting root/canal morphology, locating minor apical foramen, tooth segmentation and endodontic retreatment prediction. Radiographic features assessed were mostly periapical lesions. The studies mostly considered the decision of 1–3 experts as the reference for training their models. Almost half of the included materials compared their trained neural network model with other methods. More than 58% of studies had some level of bias.

**Conclusions:** AI-based models have shown effectiveness in finding radiographic features in different endodontic treatments. While the reported accuracy measurements seem promising, the papers mostly were biased methodologically.

## ARTICLE HISTORY

Received 3 October 2022  
Revised 17 November 2022  
Accepted 11 December 2022

## KEYWORDS

Artificial intelligence; deep learning; endodontics; endodontic diagnosis; machine learning

## Introduction

Diagnosis and treatment of apical periodontitis are common and essential activities in dental practice. Since these entities of the disease frequently are asymptomatic and hidden from oral inspection and examination alone, detection is dependent on their appearance as radiolucent areas in dental radiographs [1]. General dental practitioners have routinely been relying on conventional intraoral radiographic techniques. Panoramic X-ray (PAN) has been used mainly in epidemiological surveys [2]. More recently, the Cone Beam-Computer Tomography (CBCT) technique has gained widespread use and many proponents in endodontics [3].


However, it is well known that dental radiography is encumbered with several sources of error. There are always serious concerns about both the validity as well as reliability when using a radiolucency as a surrogate of a lesion of apical periodontitis [4]. Any missed or late diagnosis of apical periodontitis can lead to unnecessary pain and/or suffering

for the affected patient, increase the complexity of the treatment, cause complications and jeopardize a favourable outcome [5]. On the other hand, predicting e.g. apical periodontitis when in fact it is not present may lead to unnecessary, costly and potentially harmful interventions.

In recent years, machine learning and image processing with Artificial Intelligence (AI) have been advocated not only to improve diagnostic accuracy in endodontics but also to save assessment time and allow a combined activity between the dentist and machines, while both human and machine are typically orchestrated by a centralized computer controller and denoted semi-automated documentation [6,7]. AI in diagnostics and treatment planning is among other things where an artificial network with many layers (the so-called deep learning) is trained with large amounts of data to recognize different types of patterns, such as faces or radiological signs of various diseases or degrees of a particular disease [8].

**CONTACT** Shaqayeq Ramezanzade  [Shaqayeq.ramezanzade@sund.ku.dk](mailto:Shaqayeq.ramezanzade@sund.ku.dk)  Department of Odontology, Cariology and Endodontics, Section for Clinical Oral Microbiology, University of Copenhagen, Nørre Allé 20, DK-2200 Copenhagen, Denmark

\*The researchers within the Endodontic Research Collaboration in Scandinavia contributed to this study. EndoReCo collaborators: L. Bjørndal, V.S. Dawson, H. Fransson, F. Frisk, P. Jonasson, T. Kvist, M. Markvart, M. Pigg, E. Wigsten.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/00016357.2022.2158929>

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of Acta Odontologica Scandinavica Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

AI is typically based on artificial neural networks (ANN) or convolutional neural networks (CNN). ANN represents a decision-making system inspired by the biological neural networks in the brain. The standard ANN consists of fully connected layers of neurons. The input is propagated forward through several hidden layers. The hidden layers are artificial neurons that take in a set of weighted inputs and are able to produce an output through an activation function [9]. CNN is a specialized type of Feedforward Neural Network that can recognize the spatial dependencies *e.g.* between a dental root and the surrounding bone [10]. The model learns filters that are able to detect lines, curves, edges and shapes [11].

The literature has reported increasing employment of CNNs for 2D and 3D dental image diagnostics in orofacial research [12]. A review of the literature up to October 2020 in this field has shown a limited number of AI algorithms being suggested for endodontic research [13]. The applications have been classification, detection or segmentation of periapical radiolucencies [14], root fractures [15], location of apical foramen and working length determination [16], as well as root and root canal system morphology determination [17]. Due to the rapid growth of AI research in endodontology, a systematic approach to reviewing the accuracy metrics and quality of data is warranted.

This systematic review aimed to update current endodontic knowledge on the efficiency of AI methods in finding radiologic features of relevance, diagnosis, treatment and retreatment planning.

## Materials and methods

### PRISMA registration

The PRISMA 2020 (updated version) guidelines were considered in conducting this systematic review [18]. This paper was aligned with the recommendations given by Nagendrababu et al. [19] on reporting abstracts of systematic reviews in endodontology. Also, the glossary used was aligned by Nagendrababu et al. [20]. The search protocol was defined and registered at PROSPERO (International prospective register of systematic reviews). The registration number is CRD42022320332.

### PICO question

The PICO question was as follows:

- (P) Problem:** examining features of relevance for endodontic diagnosis and (re-)treatment planning.
- (I) Intervention:** finding relevant endodontic radiographic features using AI models based on radiographic features.
- (C) Comparison:** performance of dental practitioners or no comparison.
- (O) Outcome:** the efficiency of AI methods in discovering radiographic features in endodontic treatment considerations using any kind of accuracy metric (*i.e.* accuracy, precision, sensitivity, specificity, negative predictive value (NPV),

Positive predictive value (PPV), area under the curve (AUC), Dice coefficient) on the pixel, voxel, tooth and image level.

### Inclusion and exclusion criteria

The inclusion criteria were *in-vivo*, *ex-vivo* and *in-vitro* studies on the use of machine learning models (supervised models of both regression and classification groups) in endodontic diagnoses or treatments using either periapical radiographs (PA), bitewings, PAN or CBCT. Case reports, review papers, book chapters, conference papers, letters, as well as commentaries were excluded. In addition, during study selection, duplicate reports and inaccessible reports (neither full text nor abstract) were excluded.

### Search strategy for the electronic database search

The literature was searched up to 20 March 2022 for related papers written in English without any time restriction. The following databases were searched: PubMed, Scopus, Google Scholar and the Cochrane library (Table 1). In addition, the reference list of included papers and similar systematic reviews was hand searched.

### Study selection

A duplicate study selection was performed, *i.e.* if the title and abstract of the paper suggested that the paper was eligible, the full text was retrieved for further assessment. Each title/abstract was assessed twice by one author. The study selection process was checked by another author as well.

### Data extraction

The following data were extracted from the full-text version of included papers:

Title, first author, year of publication, journal's name, sub-category of endodontic treatment, image type, sample size, AI algorithm and the evaluation metrics used for assessing the performance of AI models (including accuracy, precision, sensitivity, specificity, AUC, Dice coefficient) for different tasks, Scale to assess the size of radiographic features.

The Dice similarity coefficient, also known as the Sørensen–Dice index or simply the Dice coefficient is often used within AI analyses. It is a statistical tool that measures the similarity between two sets of data. The dice similarity coefficient is a spatial overlap index and a reproducibility validation metric. The value of the dice similarity coefficient ranges from 0, indicating no spatial overlap between two sets of binary segmentation results, to 1, indicating complete overlap [21]. It should be noted that when presenting a percentage for accuracy measures of an AI model or a comparator group, the ratio is expressed as a fraction of reference group performance (see also [Supplementary Material](#)).

The required data were retrieved from the included material by two authors based on a predefined checklist worksheet and two authors supervised data extraction.

**Table 1.** Search strategy for each database.

Search criteria	
PubMed (425 papers)	('Endodontics' [Title/Abstract] OR 'endodontic diagnosis' [Title/Abstract] OR 'endodontic prognosis' [Title/Abstract] OR 'dental caries' [Title/Abstract] OR 'deep dental caries' [Title/Abstract] OR 'pulp exposure'[Title/Abstract] OR 'periapical lesions'[Title/Abstract] OR 'periapical cysts'[Title/Abstract] OR 'periapical granuloma'[Title/Abstract] OR 'vertical root fractures'[Title/Abstract] OR 'Apical foramen'[Title/Abstract] OR 'working length'[Title/Abstract] OR 'root canal morphology'[Title/Abstract] OR 'root morphology'[Title/Abstract] OR 'retreatment'[Title/Abstract] OR 'dentistry'[Title/Abstract] OR 'dental field'[Title/Abstract] OR 'teeth'[Title/Abstract] OR 'tooth'[Title/Abstract]) AND ('artificial intelligence'[Title/Abstract] OR 'neural networks'[Title/Abstract] OR 'machine learning'[Title/Abstract] OR 'deep learning'[Title/Abstract] OR 'deep learning models'[Title/Abstract] OR 'convolutional neural network'[Title/Abstract] OR 'artificial neural network'[Title/Abstract])
Scopus (527 papers)	(TITLE-ABS-KEY (endodontics) OR TITLE-ABS-KEY ('periapical AND lesions') OR TITLE-ABS-KEY ('vertical AND root AND fractures') OR TITLE-ABS-KEY ('apical AND foramen') OR TITLE-ABS-KEY ('working AND length') OR TITLE-ABS-KEY ('root AND morphology') AND TITLE-ABS-KEY ('artificial AND intelligence') OR TITLE-ABS-KEY ('neural AND network'))
Google Scholar (138 papers)	Concept 1: allintitle: Endodontics OR 'periapical lesions' OR 'vertical root fracture' OR 'Apical foramen' OR 'working length' OR 'root' 'neural network' Concept 2: allintitle: Endodontics OR 'periapical lesions' OR 'vertical root fracture' OR 'Apical foramen' OR 'working length' OR root 'Artificial Intelligence'
Cochrane library (39 papers)	((Endodontics OR 'periapical lesions' OR 'vertical root fractures' OR 'Apical foramen' OR 'working length' OR 'root morphology')):ti,ab,kw AND (('artificial intelligence' OR 'neural networks' OR 'machine learning' OR 'deep learning')):ti,ab,kw (Word variations have been searched)

**Table 2.** The domains of QUADAS 2 tool and description of risk of bias assessment criteria for each domain.

Domain	Description of risk of bias assessment
Patient selection	The paper should have enrolled a consecutive or random sample of relevant patients/cases/variables in focus. Also, the paper should avoid inappropriate exclusions (for example studies should avoid excluding 'difficult-to-diagnose' patients). This part will address the 'Patient' section in the PICO question.
Index test	Index test is a diagnostic test being evaluated against a reference standard test. In this systematic review the index test is the dentist(s) which accuracy performance is compared to the model outcomes. An ideal paper should interpret the test results blinded to a reference standard results. This part will address the 'Comparison' section in the PICO question.
Reference standard	This domain assesses if the reference standard has classified the target condition correctly or not. The target condition is the processes that the neural network aims to improve. In this systematic review, the reference standard was mostly the dental expert group (radiologist, endodontists, etc.) that label the data for training the model. The target condition was the finding of the radiographic features (e.g. periapical radiolucency, vertical root fracture prediction, root/canal segmentation). This part will address the 'Intervention' section in the PICO question.
Flow and timing	It is ideal to collect the outcomes of the results of the index test and reference standard at the same time on the same cases since any interval might result in misclassification this is specifically important in active diseases and not when assessing radiographies. It is also important that all cases receive the same reference standard.

### Risk of bias assessment

The quality of the included studies was assessed in two steps. First, two examiners made a joint assessment followed by two other authors' assessments. Any disagreement was resolved by discussion to reach a consensus. Since all the included studies were diagnostic accuracy studies, we considered the QUADAS-2 tool for quality assessment [22]. QUADAS-2 is an update of the QUADAS tool that assesses the quality of diagnostic accuracy studies in a systematic review. Considering the type of the included material (diagnostic accuracy assessment of AI models based on a human reference) this tool was chosen. Accordingly, QUADAS-2 comprises the following four domains: Patient selection, Index test, Reference standard, Flow and timing. The description of the risk of bias assessment in each domain is shown in Table 2.

The domain of patient selection, index test and reference standard were also checked for concerns regarding applicability (not applicable for the flow and timing domain). The applicability refers to the extent to which the results in each domain are likely to reflect expected outcomes when the intervention is applied broadly across the routine practice. In this part, it was also judged how much the study matched the PICO question 'Finding features of relevance for endodontic diagnosis and (re-)treatment planning (P) using AI

methods (I) compared with the performance of dental practitioners (or no comparison) (C) using any kind of accuracy metric (*i.e.* accuracy, precision, sensitivity, specificity and AUC, Dice coefficient) on the pixel, voxel, tooth and image level (O)' and rated it as 'low', 'high' or 'unclear'.

### Data analysis and synthesis

Due to the heterogeneous nature of included material, a meta-analysis could not be conducted. In the included material, a wide range of accuracy measures have been implemented but sensitivity and specificity were the two most reported metrics. Therefore, the quantitative analysis in this systematic review was on studies reporting both the specificity and the sensitivity. The stratification was further employed in the type of studies *i.e.* periapical radiolucency segmentation, vertical root fracture prediction and root/canal segmentation. In addition, the qualitative data of included material were summarized in the results.

### Results

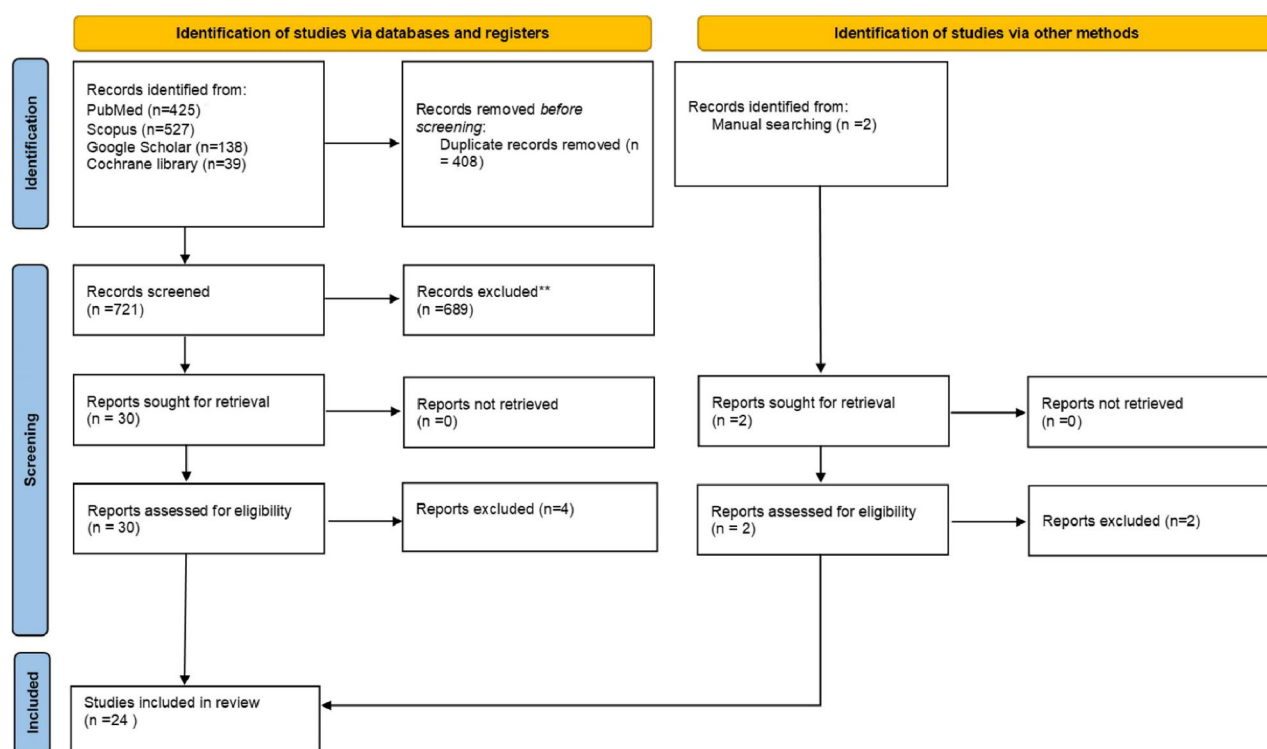
The initial search in datasets and manual search resulted in 1131 papers. After duplication removal, 723 papers remained.

Through the title/abstract assessment, 689 papers were removed. Of the 30 remaining papers, 6 were excluded for some reason (Table 3). A total of 24 papers were finally included. The PRISMA flow diagram of selection at each level is illustrated in Figure 1. Also, the study characteristics of the included material are shown in Table 4. The papers were published between 2012 and 2022. The selected studies report data from three types of radiographic sources of which a few reported PANs [7,14,23–25], some PAs [15,16,26–32] and CBCTs [5,17,25,26,33–38]. The AI models used were mostly CNN and ANN. One reported the use of a Bayesian neural network (BNN) for predicting the results of endodontic retreatment [32].

The reported subcategories were periapical lesion [5,7,14,23,27,28,30,31,36], vertical root fractures [24,26,39], predicting root morphology [25,40], predicting C-shaped canals [33,35], locating minor apical foramen [16,29], tooth segmentation [17,34,35,37,38] and predicting results of endodontic retreatment [32]. For the comparison between an AI model and human experts in classifying, detecting or segmenting periapical radiolucencies, some papers had 3 experts as reference [14,27,28,30], one study had more than three expert references [7], and the remaining studies had two or one human references [5,23,31,36]. The studies on pulp cavity or root morphology mostly had only one human reference [23,25,33,34,36,37,41] and a few with two

**Table 3.** Papers excluded with reason.

Title	Author, year, Journal	Excluded with reason
Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study	Herbst et al. 2022, Journal of Dentistry	No radiographies assessed
Neuro-fuzzy method for predicting the viability of stem cells treated at different time-concentration conditions	Blindal et al. 2017, Technology and Health Care	No radiographies assessed
Exploiting saliency in attention based convolutional neural network for classification of vertical root fractures	Xu et al. 2021, International Conference on Pattern Recognition	No clinical data
Clinical prediction of teeth periapical lesion based on machine learning techniques	Eid et al. 2019, International Conference on Electrical, Communication, and Computer Engineering	Insufficient/incorrect clinical data- vague process of detecting the various types of lesions and without information about the specific clinical examination leading to the specific lesion types.
Automatic lesion detection in periapical X-rays	Sajad et al. 2019, International Conference on Electrical, Communication, and Computer Engineering (ICECCE),	Insufficient/incorrect clinical data- vague process of detecting the various types of lesions and without information about the specific clinical examination leading to the specific lesion types.
Artificial intelligence system seems to be able to detect a high proportion of periapical lesions in cone-beam computed tomographic images	Brignardello-Petersen et al. 2020, JADA + Clinical Scans	No clinical data



**Figure 1.** PRISMA Flowchart of the study selection at different levels.

Table 4. Study characteristics of included materials.

Author (reference no)	Subcategory	Image type	Task	Sample size	AI model	Reference group	Metrics results	Comparison group
Li et al. [27]	Periapical lesion	PA	Detection	322	CNNs	Three professional dentists	Accuracy: 92.5%	-
Bayrakdar et al. [14]	Periapical lesion	PAN	Segmentation	470	U-Net	Three dental radiologists	The sensitivity, precision and F1-score for segmentation of periapical lesions at 70% IoU values were 0.92, 0.84 and 0.88, respectively.	-
Ekert et al. [7]	Periapical lesion	PAN	Detection	85	CNN	The majority vote of 6 independent and experienced dentists	The AUC of the CNN was 0.85 (0.04). Sensitivity and specificity were 0.65 (0.12) and 0.87 (0.04), respectively. The resulting PPV was 0.49 (0.10), and the NPV was 0.93 (0.03)	-
Orhan et al. [36]	Periapical lesion	CBCT	Detection	153	Deep CNN	An oral and maxillofacial radiologist	The reliability of correctly detecting a periapical lesion was 92.8%. The deep CNN volumetric measurements of the lesions were similar to those with manual segmentation ( $p > .05$ )	Volumes calculated by the manual segmentation
Setzer et al. [5]	Periapical lesion	CBCT	Detection	20	U-Net architecture	An oral maxil. radiol., an endodontist and a student	DLS lesion detection accuracy was 0.93 with specificity of 0.88, PPV of 0.87 and NPV of 0.93. The overall cumulative DICE indexes for the individual labels were lesion = 0.52	-
Endres et al. [23]	Periapical lesion	PAN	Detection	2902	Deep CNN	A single OMF-surgeon	An average precision of 0.60 ( $\pm 0.04$ ), and an F1 score of 0.58 ( $\pm 0.04$ ) corresponding to a PPV of 0.67 ( $\pm 0.05$ ) and TPR of 0.51 ( $\pm 0.05$ )	24 oral and maxillofacial (OMF) surgeons. These OMF surgeons represented a random sample comprising 13 residents, 11 attending physicians (6 female and 18 male)
Pauwels et al. [28]	Periapical lesion	PA	Detection	10	CNN	Three oral radiologists	When data were split up by socket, the mean sensitivity, specificity and ROC-AUC values were 0.79, 0.88 and 0.86, respectively; when split up by filter, they were 0.87, 0.98 and 0.93, respectively	For radiologists, the mean sensitivity, specificity and ROC-AUC values were 0.58, 0.83 and 0.75, respectively
Moidu et al. [30]	Periapical lesion	PA	Segmentation and classification	1950	CNN model YOLO version	Three endodontists	The model exhibited a 92.1% sensitivity/recall, 76% specificity, 86.4% PPV/precision and 86.1%NPV. The accuracy, F1 score and Matthews correlation coefficient were 86.3%, 0.89 and 0.71, respectively	-
Ngoc et al. [31]	Periapical lesion	Bitewings	Detection	130	Faster R-CNN	Two experienced endodontists	The sensitivity, specificity and accuracy of diagnoses of the software were 89.5, 97.9 and 95.6%, respectively	-
Johari et al. [26]	Vertical root fractures	CBCT and PA	Detection	240	ANN (probabilistic neural network)	<b>Examiner not mentioned in detail</b>	In the periapical radiographs, the maximum accuracy, sensitivity and specificity values in the three groups were 70.00, 97.78 and 67.7%, respectively	The maximum accuracy, sensitivity and specificity values in the CBCT images were 96.6, 93.3 and 100%, respectively, at the variance change range of 0.1–0.65. (PA and CBCT radiographs were compared)
Fukuda et al. [24]	Vertical root fractures	PAN	Detection	300	CNN	Two radiologists and one endodontist	Of the 330 VRFs, 267 were detected. 20 teeth without fractures were falsely detected. Recall was 0.75, precision 0.93 and F measure 0.83	-
Kositbownchai et al. [15]	Vertical root fractures	PA	Detection	200	ANN	<b>Examiner not mentioned in detail</b>	Sensitivity (98%), specificity (90.5%) and accuracy (95.7%)	-

(continued)

Table 4. Continued.

Author (reference no)	Subcategory	Image type	Task	Sample size	AI model	Reference group	Metrics results	Comparison group
Hatvani et al. [40]	Root morphology	CBCT	Detection	17	CNN (U-net)	<b>Examiner not mentioned in detail</b>	They reported Dice levels of 0.88 (original), 0.89 (SRR) and 0.91 (CNN). The results suggest the superiority of the proposed CNN-based approaches over reconstruction-based methods	Results are compared to a recent reconstruction-based SR method (SRR), implemented in MATLAB
Hiraiwa et al. [25]	Root morphology	CBCT PAN	Segmentation and classification	760	CNN	A radiologist	Extra roots were observed in 21.4% of distal roots on CBCT images. The deep learning system had diagnostic accuracy of 86.9% for the determination of whether distal roots were single or had extra roots	Compared with the performance of two radiologists
Jeon et al. [33]	Root canal morphology predicting C-shaped canals	PAN CBCT	Segmentation and classification	2040	CNN	All images were evaluated by a single oral and maxillofacial radiologist	The accuracy, sensitivity, specificity and precision of the CNN model were 95.1, 92.7, 97.0 and 95.9%, respectively	Two specialist, one radiologist (experience >6 years) and one endodontist (experience >6 years)
Sherwood et al. [35]	Root canal morphology predicting C-shaped canals	CBCT	Segmentation and classification	100	U-Net	Two experienced examiners	U-Net Sensitivity: $0.720 \pm 0.0495$ The mean positive predictive values were $77.6\% \pm 0.1998\%$	Two types of U-net models: Xception U-Net and residual U-Net the mean sensitivity values were $0.786 \pm 0.0378$ for Xception U-Net, $0.746 \pm 0.0391$ for residual U-Net//The mean PPV were $8.2\% \pm 0.01971\%$ for residual U-Net and $80.0\% \pm 0.1098\%$ for Xception U-Net
Saghiri et al. [29]	Locating minor apical foramen	PA	Classification	50	ANN	<b>Not mentioned in detail</b>	The correct assessment by the endodontists was accurate in 76% of the teeth. ANN determined the anatomic position correctly 96% of the time. The confidence interval for the correct result was 64.16–87.84 for endodontists and 90.57–101.43 for ANN	Two endodontists assessed the root apices under radiographs by using a Rinn XCP and a photostimulable phosphorus system
Saghiri et al. [16]	Locating minor apical foramen	PA	Classification	50	ANN	Two endodontists evaluated	Analysis of the images from radiographs (test samples) by ANN showed that in 93% of the samples, the location of the AF had been determined correctly by false rejection and acceptance error methods.	–
Shetty et al. [37]	Volume of the pulp space	CBCT	Segmentation	35	CNN (OsiriX MD / and 3D Slicer)	An experienced endodontist	The mean decreases in the pulp spaces measured using OsiriX MD were $25.06\% \pm 19.45\%$	The decreases in the pulp spaces using 3 D Slicer were $26.10\% \pm 18.90\%$ . The Slicer was found to be comparable with commercial software regarding vol. assessment of the post-REP pulp space
Lin et al. [41]	Tooth and pulp cavity segmentation	CBCT	Segmentation	30	U-Net network	An endodontic specialist	The segmentation accuracy of the experimental group (as measured by Dice similarity coefficient, precision rate, recall rate, average symmetric surface distance, Hausdorff distance) were better than the control group	Imaging Interaction Toolkit Workbench software (3D Slicer), which were used as the training samples for subsequent experiments
Li et al. [34]	Roots segmentation	CBCT	Segmentation	29	U-net and RNN	A dental technician	Applying to the testing dataset, the segmentation accuracy measured by the intersection over union (IOU), dice	–

(continued)

Table 4. Continued.

Author (reference no)	Subcategory	Image type	Task	Sample size	AI model	Reference group	Metrics results	Comparison group
Zheng et al. [38]	Pulp chamber segmentation	CBCT	Segmentation	180	DL(network not mention)	<b>Not mentioned in detail</b>	similarity coefficient (DICE), average precision rate (APR), average recall rate (ARR) and average symmetrical surface distance (ASSD) are 0.914, 0.955, 95.8%, 95.3%, 0.145 mm, respectively In the training and validation sets, the results showed high spatial overlaps between manual and automatic segmentation (dice = 87.8%). For the testing set, the estimated human ages were not significantly different with true human age ( $p = .57$ ), with a correlation coefficient $r = 0.74$ Accuracy: 0.99, AUC: 0.96, Specificity: 0.99, Sensitivity 0.9, Mean F 10.78	Manually labelled segmentation masks on validation data
Deleat-Besson et al. [17]	Automatic segmentation of dental root canal	CBCT	Segmentation	80	U-Net	<b>Not mentioned clearly</b>		Dental clinician (details not mentioned clearly)
Campo et al. [32]	Endodontic re treatment prediction	PA	Classification	205	BNNS	<b>Not mentioned in detail</b>	Prediction on final solution for treatment and retreatment in 84.4% of the cases, by applying the leave-one out techniques	-

references [16,31,35]. Within one vertical root fracture prediction study, three human experts were used as references [5,14,24,28,30]. The remaining root/canal morphology prediction or segmentation studies were unclear about the reference group and did not mention the expert information in detail [15,17,26,29,32,38,40].

Almost half of the included material made comparisons between their trained neural network model and other methods. Three papers on the periapical lesions compared the accuracy of their trained model with the performance of human experts [23,28,36]. Regarding the pulp cavity or root morphological studies, three papers compared their model with human experts [25,33,38]. Three compared their neural network model with non-AI technologies; a MATLAB-based method [40] Imaging Interaction Toolkit Workbench software [41] and a 3D Slicer software [37]. One paper made comparisons between different U-net (a subgroup of CNN) models [35]. A paper on vertical root fracture compared the accuracy of its model on different types of radiographies (PA and CBCT) [26]. The paper on endodontic retreatment compared its model with dental clinicians [17].

In the included studies, the evaluation metrics used for assessing the prediction performance of AI models were accuracy, precision, sensitivity, specificity and AUC while for segmentation tasks, the Dice coefficient was reported (see [Supplementary Material](#)).

### Interpretation of sensitivity–specificity plot

The reported sensitivity and specificity of the different types of studies by sample size are shown in [Figure 2](#). Papers on root/canal segmentation acted well both in sensitivity and specificity [17,33]. The highest sensitivities were related to VRF papers. This means that the two related models were able to detect true cases of VRFs greatly [26,39], while the papers on periapical radiolucencies reported mixed rates for both sensitivity and specificity [5,7,27,28,30,31].

### Results of risk of bias assessment

More than 58% of studies had some level of bias. The highest risk of bias was related to the patient selection domains both in the 'reference test' and 'the applicability' arms of the tool (37.5% and 29% respectively) ([Table 5](#)). Studies working on extracted teeth (*in-vitro*), as well as cadaver studies, were considered 'high-risk' or 'unclear risk' for bias in the patient selection domain of the applicability arm of the QUADAS 2 [15,16,26,29,40]. All studies had a low risk of bias and low concerns of applicability regarding the index test domain of the QUADAS 2, (the human comparators as index test were blinded to the reference test).

The details for quality assessment (QUADAS 2) summary of the risk of bias and applicability concerns are illustrated in [Table 5](#). In addition, the assessment of the individual risk of bias domains and concerns regarding applicability is shown ([Figure 3](#)).

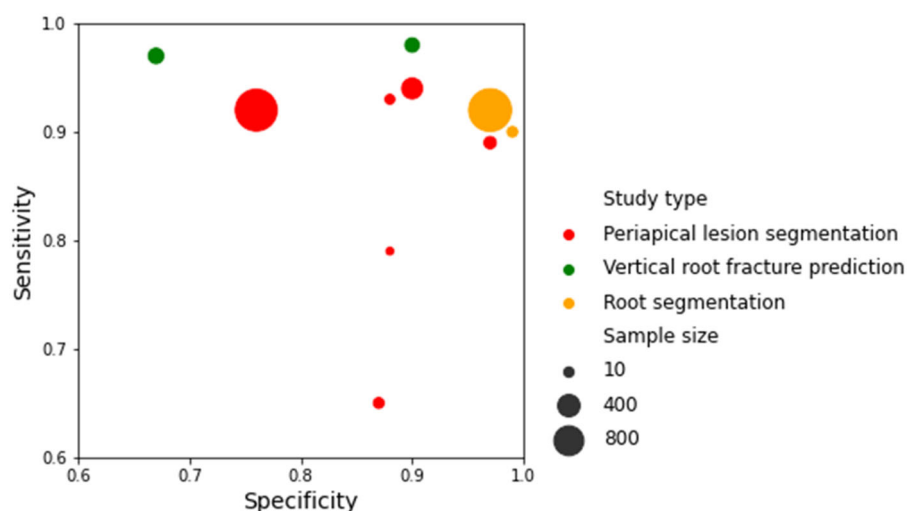


Figure 2. The reported sensitivity and specificity of different types of studies by sample size.

Table 5. Quality assessment (QUADAS 2) summary of risk of bias and applicability concerns.

First author/reference	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Li et al. [27]	Low	Low	Low	Low	Low	Low	Low
Bayrakdar et al. [14]	Low	Low	High	Unclear	Low	Low	Low
Ekert et al. [7]	High	High	Low	Low	High	High	Low
Orhan et al. [36]	Low	Unclear	Low	Low	Low	Low	High
Setzer et al. [5]	Low	Low	Low	Low	Low	Low	Low
Endres et al. [23]	Low	Low	Low	Low	Low	Low	Low
Pauwels et al. [28]	Low	Low	Low	Low	Low	Low	Low
Moidu et al. [30]	Low	Low	Low	Low	Low	Low	Low
Ngoc et al. [31]	Low	Low	Low	Unclear	Low	Low	Low
Johari et al. [26]	High	Low	Low	Low	High	Low	Low
Fukuda et al. [24]	Low	Low	Low	Low	Low	Low	Low
Kositbowornchai et al. [15]	Unclear	High	High	Low	High	Unclear	Low
Hatvani et al. [40]	Unclear	Unclear	High	Low	High	Low	High
Hiraiwa et al. [25]	Low	Low	Low	Low	Low	Low	Low
Jeon et al. [33]	Low	Low	Low	Low	Low	Low	Low
Sherwood et al. [35]	High	Low	Low	Low	Low	Low	Low
Saghiri et al. [29]	High	Unclear	Low	Low	High	Low	Low
Saghiri et al. [16]	High	Unclear	Low	Low	High	Low	Low
Shetty et al. [37]	High	Low	Low	Low	Unclear	Low	Low
Lin et al. [41]	Low	Low	Low	Low	Low	Low	Low
Li et al. [34]	Low	Low	High	Low	Low	Low	High
Zheng et al. [38]	Low	Low	Low	Low	Low	Low	Low
Deleat-Besson et al. [17]	Low	Unclear	Low	Low	Low	Low	Low
Campo et al. [32]	High	Unclear	Unclear	Low	Low	High	Unclear

### Neural network algorithms on periapical radiolucencies

Nine of 24 papers were on the use of neural networks in the automatic detection of periapical radiolucencies; of those 7 papers detected the periapical lesions and only two performed lesion segmentation [14,30]. In this, the studies reported excellent results, with accuracy values of 0.93 [27] and 0.96 [31] (Table 4). The methods used for diagnosis included CNNs [5,7,23,27,28,30,31,36], while some of the CNN architectures were the U-net [5,23,36]. The datasets consisted of PAs [27,28,30,31], CBCTs [5,36] and PANs [7,14,23]. The image count varied greatly across studies, with the smallest set including 10 radiographs [28] and the largest 2902 [23]. While six papers had no scale on the size of periapical radiolucencies, three presented a scale on the size of the radiolucency [7,30,36]. Moidu et al. [30] used the periapical index (PAI) scoring system

for training their CNN model. The best performance of the model was in the true detection of PAI score 1 (90.9%). In addition, PAI 3 and 4 were truly predicted in 60% and 71% of cases respectively [30]. When this multiple classification was turned into binary classification (scores 3, 4 and 5 were dichotomized as diseased) the accuracy increased and reached to 86%. Ekert et al. [7] considered an ordinal scale for training its neural network model and the scale was as follows: no periapical lesion (radiolucency), uncertain periapical lesion (widened periodontal ligament) and certain periapical lesion (clearly detectable lesion/radiolucency). Orhan et al. [36] trained its deep CNN regarding volumetric measurements of the periapical radiolucencies by manual segmentation and assessed the ability of the model; the results showed no significant difference between the manual and AI-based volume measurement methods ( $p > .05$ ).

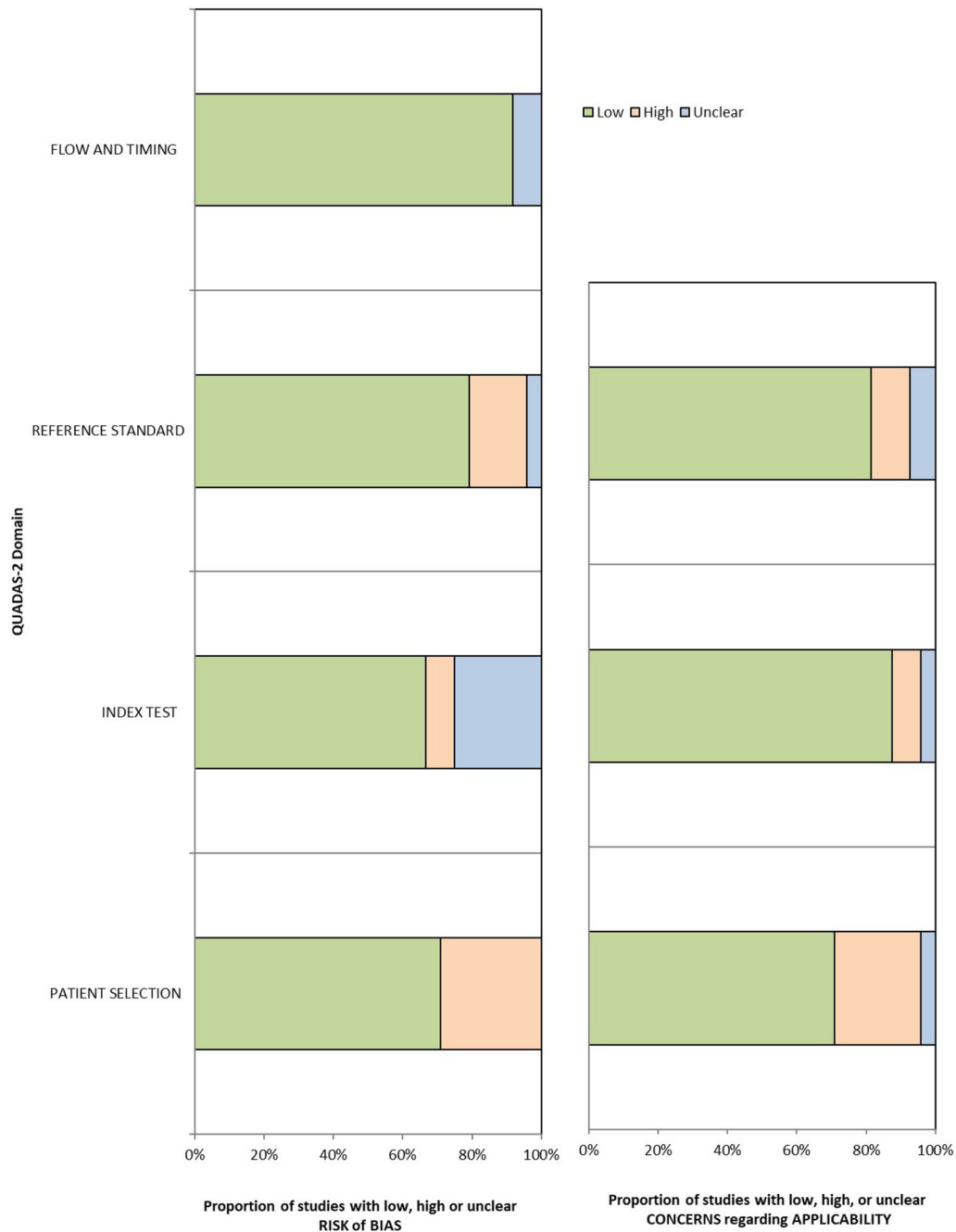


Figure 3. Assessment of individual risk of bias domains (left) and concerns regarding applicability (right).

### Vertical root fractures

Three papers reported on the use of neural networks in the automatic detection of vertical root fractures. The images assessed were CBCT [26], PAN [24] and PA [15,26]. The accuracy values reported were 0.70 [26] and 0.96 [15] (Table 4). The methods used for diagnosis included CNN [24] and ANN [15,26]. The image count varied from 200 [15] to 300 [24]. Only one *in-vivo* paper was found in this regard which performed detection and localization of VRFs in PANs using a CNN-based model, generating bounding boxes around the regions of interest. The reported F-score was 0.83 across all teeth [24].

Two out of three papers were *ex-vivo* studies working on extracted single-rooted premolars [15,26]. The study by Kositbowornchai et al. [15] created a fracture line in 150 extracted teeth, implemented a model that takes a horizontal line drawn across the root as input data (while 50 teeth remained without a created fracture line) and reported an accuracy level of 0.96. The other paper used both sound and endodontically treated extracted teeth with and without VRF and confirmed the presence or absence of fractures under a microscope ( $\times 20$ ). The performance of their probabilistic neural network model on CBCT images was more efficient than PA radiographs with maximum accuracy, sensitivity and

specificity values in the CBCT group of 96.6, 93.3 and 100% respectively [26].

### **Root morphology and predicting C-shaped canals**

Four papers discussed the 3D assessment of root/root canal morphology *via* CBCT Scans using CNN algorithms [25,33,35,40] while two papers specifically discussed the prediction of C-shaped canals [33,35]. The reported accuracy ranged from 89.9% to 95.1%. Sherwood et al. [35] reported a sensitivity rate of 72%. Hatvani et al. [40] compared two different CNN architectures: a subpixel network and a U-net model and reported a Dice coefficient of 0.91 and 0.89, respectively (full detail in Table 4). One paper combined both 2D (PAN) and 3D (CBCT) images in their assessment [25].

### **Tooth segmentation**

Five papers discussed tooth segmentation on CBCTs using CNN models [34,37,38,41] or a combination of CNN and recurrent neural network (RNN) [17]. The sample size of CBCTs ranged from 29 [34] to 180 [38]. Zheng et al. [38] reported a high spatial overlap between manual and AI-based segmentation (Dice = 87.8%). Lin et al. [41] assessed the Dice similarity coefficient for U-net based segmentation at tooth level (86.7%) and pulp level (96.2%). A study on a U-net model showed AUC, sensitivity, specificity and accuracy of 99% but an F1 score of 75% [17]. Another study compared the accuracy of different AI-based models named OsiriX MD and 3D Slicer and reported similar accuracy with regard to the volumetric assessment of pulp space for both models [37].

### **Locating minor apical foramen**

Only two cadaveric preliminary studies with an attempt to simulate the clinical situation of working length determination were found which both used straight single-rooted teeth and ANN as a decision-making system [16,29]. In the paper by Saghiri et al. [16], the location of the apical foramen had been determined correctly in 93% of the test dataset using false rejection and acceptance error methods. In another study, the performance of an ANN model was compared to a group of endodontists using the electronic apex locators while considering the stereomicroscope as the comparison group. The ANN model outperformed the endodontists with an accuracy of 96% vs 76% [29].

### **Prediction of endodontic retreatment**

Only one study by Campo et al. [32] presented a decision support system. They assessed 205 PAs using a BNN. By ordering the established and high-risk variables, they could predict that in 84.4% of the cases, retreatment should be performed because the chance of success seemed high.

## **Discussion**

The previous review on the use of AI in Endodontics (searching literature up to October 2020) has shown a limited number of models in endodontics [13]. Due to the rapid growth of AI research in endodontology, a systematic approach to update current knowledge on the efficiency of AI methods in finding radiologic features of endodontic treatments is warranted.

This systematic review revealed that most of the studies investigating the use of AI in endodontology were based on ANN or convolutional neural networks (CNN). Whenever there was a comparison between machine performance and human experts, the AI demonstrated good and often better accuracy than human performance alone, but whether it will benefit within a clinical setting is far too early to conclude. Ten of the included studies were on detecting periapical radiolucencies, three on detecting vertical root fracture, four on the classification of root morphology, four on pulp cavity segmentation, two on locating minor apical foramen and one on predicting outcomes involving radiographs in relation to endodontic retreatment. While the reported accuracy measurements seem promising, the quality of papers is considered low, as almost 60% of papers had some level of bias (Table 5). In particular, the reference standard domain was scored low in 37.5% of the papers.

### **Periapical lesions**

Almost half of the included papers have focussed on the potential of AI for the detection or segmentation of periapical radiolucencies [5,7,14,23,27,28,30,31,36]. The study researching the ability of oral and maxillofacial surgeons of varying levels of experience to detect radiolucencies reported a mean PPV of 0.69. The machine learning model outperformed 14 of the 24 annotators, reaching a PPV level of 0.67 [23].

Two studies trained their models based on the U-net architecture on 3D CBCT scans, reporting comparable detection accuracies of 0.93 [5,36]. Although when using a model for periapical lesion segmentation in CBCT images, there was no significant difference between human and machine performance indicating an inherent difficulty in CBCT segmentation [36].

Outlining periapical radiolucencies as signs of apical periodontitis is difficult because of the size, shape and low contrast with the surrounding structures. Of the two studies applying U-net architecture on PANs and CBCTs, respectively, there was a relatively high difference in Dice scores being 0.88 [14] and 0.52 [5], respectively. A causing factor of the suboptimal performance could be the fact that the model here was trained to predict multiple structures, *i.e.* lesions, teeth, bone, restorative materials and background. Using a specialized network only training for one particular feature may aid voxel classification and better prediction. In addition, the size of the dataset used differed. Within the low Dice score study, only 22 PANs were used [5], whereas the high Dice score included 470 PANs.

The literature has shown some potential for the AI algorithms to step deeper into the data to introduce scales for detecting the size of periapical radiolucency instead of just reporting the presence or absence of radiolucency. These multiple classifications might prevent class heterogeneity and oversimplification [42].

However, based on the sensitivity and specificity of papers on periapical lesions in Table 4 and the sensitivity-specificity plot (Figure 2), it could be noted that the study using radiolucency classification had a relatively poor performance in detecting certain from uncertain periapical lesions [5], that might also be attributed to the low prevalence of periapical lesions in the dataset. Similarly, the study using the PAI score had a relatively low specificity in truly detecting cases 'without periapical lesions', indicating a low specificity [30]. In fact, the false prediction of healthy teeth as diseased teeth appeared in 32% of total assessments. Although being widely used, this might show some deficiencies in the PAI scoring system. In short, studies using multiple classifications need further studies with huge datasets for verification.

#### **Vertical root fractures (VRFs)**

The radiographic feature of VRFs varies from a thin, radiolucent line [43], a diffuse widening of periodontal ligament to vertical bone loss and separation of root fragments [44]. Considering the reported metrics, the *in-vivo* study on VRFs showed good potential for detecting VRFs in PANs using AI models. They highlighted that the model performed best on mandibular premolars and molars, situated in regions that present less overlap of teeth on radiographic images [24]. The two other papers presenting a semi-automated machine learning approach for detecting VRFs on extracted teeth presented a good performance of their model while one created the fractures on extracted teeth [15] and the other confirmed presence or absence of fractures under a microscope. The performance of their probabilistic neural network model (PNN, a subgroup of ANN) on CBCT images was more efficient than on PA radiographs [26]. However, the efficacy of these *ex-vivo* models as a tool for clinical applications is not confirmed [15,26].

#### **Root morphology and predicting C-shaped canals**

It is challenging to obtain accurate radiographic visualization of dental roots due to their variety of size and shape. A study targeted detecting the presence of an extra root in the first mandibular molar in PANs [25]. The CNNs surpassed the performance of two experienced radiologists, obtaining an accuracy of 86.9% for the determination of whether the distal roots of mandibular molars were single or comprised an extra root canal. It must be noted that the prediction pipeline was not fully automated, requiring manual teeth segmentation. The need for manual initialization is one considerable limitation that could be addressed by training a deep learning model for computing an area of interest. Similar research on the ability of CNNs to detect C-shaped root structures in mandibular second molars (extracted from

PANs) showed the potential of their CNN model in outperforming human performance by reaching an accuracy score of 0.951 [33]. The authors performed an in-depth analysis of the decision-making system by visualizing regions with the highest activation incorrectly predicted radiographs using the Grad-Cam method [33,45].

#### **Tooth segmentation**

Despite the limited number of papers on AI-based models for tooth segmentation, the outcomes indicate the high potential of AI-based segmentation models to find their way into routine dental practice. A recent study developed an automatic segmentation method for pulp chamber segmentation using 180 CBCTs as training, validation and test datasets and the results illustrated a high spatial overlap between manual and AI-based segmentation (Dice = 87.8%) [38]. Another study on pulp cavity segmentation reported that the accuracy of segmentation as measured by Dice similarity coefficient for U-net based segmentation (experimental group) was better than 3D Slicer software (control group) both at tooth level (86.75%) and at pulp level (96.20%) [41]. Recent research on the classification and prediction of root canal and crown shape analysis workflow showed AUC, sensitivity, specificity and accuracy metrics close to 1 for a U-net model [17]. Their results indicated the success of this U net model, although, the F1 score reported was a bit low (75%). This might be rooted in the inconsistencies of human experts' segmentation of the thin and curved root canals [17].

#### **Locating minor apical foramen and working length determination**

Two preliminary cadaveric studies on straight single-rooted teeth presented a new approach for WL determination using ANN to enhance the accuracy of PA. Their results suggested the ANN model might be a good help to clinicians in routine practice [16,29] but based on the *in-vitro* nature of the studies, the data did not lead to any clinical conclusion.

#### **Prediction of endodontic retreatment**

One study presented a BNN model, which might reduce the number of unsuccessful retreatments and avoid unnecessary extractions as well. This system solves new problems by adapting learning solutions from experiences. They predicted the proper retreatment cases by ordering the established and high-risk variables [32]. Current literature on this application of AI is inconclusive.

#### **Current limitations**

Apart from the methodological biases seen at different levels of quality assessment (Figure 3), there were other shortcomings within the included papers. The included papers mostly have developed and trained their AI-models based on relatively small datasets. This might be attributed to the cost of time and labour of huge datasets. This limited number of labelled images for training models increases the risk of

'overfitting' the data with a loss of generalizability [46]. A potential solution is increasing the size of data used for training a model using data augmentation techniques as well as pre-trained CNN models. These might bring robust diagnostic performance in future neural networks [47]. In addition, the reliability of the human experts for model referencing was questionable in some papers as they did not mention the reference group in detail [15,17,26,29,32,38,40] and the number of experts was limited in some other papers [23,25,33,34,36,37,41]. Clarifying the details of the reference group (number of experts, their specialty and years of experience) can increase the robustness of the reference test and help to overcome this issue [7].

A common issue with the included papers was the restrictions in inclusion criteria, as they did not consider images with different qualities and only included high-quality images. One major task of neural networks is to direct the dental practitioner's diagnostic efforts towards difficult to assess images. Although removing images with too poor quality from the reference test database makes sense, the importance of some factors like image projection or quality contrast on the discrimination ability of models should be investigated in future works.

### *The future direction of AI in dentistry*

The recent decade has witnessed a noticeable development in the use of AI in the dental field including endodontology [48]. This review has shown one great challenge with the current literature on AI in dentistry; the presence of computer-based papers without proper understanding of relevant input data from a clinical point of view. Also, there are several models which are not trained based on clinical rigorous data [49,50] (Table 3). Therefore the outputs of such models are unreliable regardless of the accuracy metrics reported [48]. In this systematic review, papers without a robust source of data for model training were excluded. In future works, integrated collaboration of dental and computer researchers is warranted to enhance the practicality of the designed model along with accuracy measures.

The huge amounts of digital data, in form of image and report data, are archived as 'big data' that can act as a unique platform for the development of AI in endodontology. The current models on periapical lesion detection, VRF detection and root/canal segmentation are dealing with small datasets. But using these models with huge datasets (big data), is an obvious direction to follow. One objective for researchers and technology developers of AI for use in endodontics and dentistry as a whole, is to find models based on big data that quickly and accurately can support the practitioner with diagnosis of common radiological features for clinical decision making.

With new insights into the ways AI and big data correlate together, there will be some possibilities to link the medical and dental health data in order to get an optimal overview of potential prognostic variables in the complex interplay between oral cavity and systemic diseases; for instance the association between endodontic inflammatory disease and a first myocardial infarction (MI) [51]. However, such health

data connections have remained largely unexplored and their patient-based improvement potentials has yet to be harnessed.

One of the greatest challenges with AI in Endodontology is the clinical deployment of AI applications in dental imaging. Attention must be given to the ability of reproducibility of developed models. The well-understood pitfalls threatening the reproducibility of models are overfitting and under-specification. Most of the literature has proposed a machine learning model which is validated on single sites only. Future works should validate their developed models by cross-centre training and considering datasets with different populations to mitigate generalizability problems [52].

Recently, explainable AI (XAI) is emerging in 'black box' models of endodontology [33]. The 'black box' refers to the complex models to a level that is not straightforwardly interpretable to humans [53]. XAI is a new field in machine learning that can explain the decision-making process. With the transparency of XAI, neural network models are not complex black boxes anymore. In fact, XAI visualizes the pixels that are decisive for the prediction. This enables the clinicians to explore what was the most important feature screened by neural networks. XAI also enables clinicians to distinguish between safe meaningful prediction strategies and false predictors [33,54].

However, it is not certain whether AI-based models can add essential value to endodontics. Adding value to endodontic treatments except for more accurate diagnosis includes the establishment of more efficient treatment processes while being user-friendly for practitioners. AI-based models in endodontology are new advanced technologies and currently little is known about the learning curve for using such models in routine practice. Rather than replacing dental clinicians' diagnostic behaviour with machine learning models, the much more likely prospect would be the creative combination of human and AI models to solve challenging problems and oversee diagnostic procedures in routine dental practice.

## **Conclusions**

AI-based models have shown effectiveness in discovering radiographic features in different endodontic treatments. While the reported accuracy measurements seem promising, the papers mostly had different levels of methodological bias. Currently, there are not enough papers to come to a conclusion for the practical implementation of AI technologies in routine practice.

## **Disclosure statement**

No potential conflict of interest was reported by the author(s).

## **Funding**

This project has been funded by the University of Copenhagen

## ORCID

Shaquayeq Ramezanzade  <http://orcid.org/0000-0002-1012-0730>  
 Tudor Laurentiu  <http://orcid.org/0000-0002-4942-1389>  
 Azam Bakhshandah  <http://orcid.org/0000-0003-2974-1331>  
 Bulat Ibragimov  <http://orcid.org/0000-0001-7739-7788>  
 Thomas Kvist  <http://orcid.org/0000-0002-2685-3656>  
 Lars Bjørndal  <http://orcid.org/0000-0002-2183-6400>

## References

- [1] Ørstavik D. *Essential endodontology: prevention and treatment of apical periodontitis*. Oxford: John Wiley & Sons; 2020.
- [2] Choi J-W. Assessment of panoramic radiography as a national oral examination tool: review of the literature. *Imaging Sci Dent*. 2011;41(1):1–6.
- [3] Patel S, Brown J, Semper M, et al. European society of endodontology position statement: use of cone beam computed tomography in endodontics: European Society of Endodontology (ESE). *Int Endod J*. 2019 Dec;52(12):1675–1678.
- [4] Petersson A, Axelsson S, Davidson T, et al. Radiological diagnosis of periapical bone tissue lesions in endodontics: a systematic review. *Int Endod J*. 2012;45(9):783–801.
- [5] Setzer FC, Shi KJ, Zhang Z, et al. Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. *J Endod*. 2020;46(7):987–993.
- [6] Boreak N. Effectiveness of artificial intelligence applications designed for endodontic diagnosis, decision-making, and prediction of prognosis: a systematic review. *J Contemp Dent Pract*. 2020;21(8):926–934.
- [7] Ekert T, Krois J, Meinhold L, et al. Deep learning for the radiographic detection of apical lesions. *J Endod*. 2019;45(7):917.e5–922.e5.
- [8] Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol*. 2017;10(3):257–273.
- [9] Kavzoglu T. Determining optimum structure for artificial neural networks. Paper presented at: Proceedings of the 25th Annual Technical Conference and Exhibition of the Remote Sensing Society; 1999. Cardiff, UK.
- [10] Goodfellow I, Bengio Y, Courville A. *Deep learning*. China: MIT Press; 2016.
- [11] Wu J. Introduction to convolutional neural networks. *Natl Key Lab Novel Softw Technol Nanjing Univ China*. 2017;5(23):495.
- [12] Schwendicke F, Golla T, Dreher M, et al. Convolutional neural networks for dental image diagnostics: a scoping review. *J Dent*. 2019;91:103226.
- [13] Hagan MT, Demuth HB, Beale MH. *Neural network design*. Boston (MA): PWS Pub. Co; 1995.
- [14] Bayrakdar IS, Orhan K, Çelik Ö, et al. A U-Net approach to apical lesion segmentation on panoramic radiographs. *Biomed Res Int*. 2022;2022:7035367.
- [15] Kositbowornchai S, Plermkamon S, Tangkosol T. Performance of an artificial neural network for vertical root fracture detection: an ex vivo study. *Dent Traumatol*. 2013;29(2):151–155.
- [16] Saghiri MA, Asgar K, Boukani KK, et al. A new approach for locating the minor apical foramen using an artificial neural network. *Int Endod J*. 2012;45(3):257–265.
- [17] Deleat-Besson R, Le C, Al Turkestani N, et al. Automatic segmentation of dental root canal and merging with crown shape. *Annu Int Conf IEEE Eng Med Biol Soc*. 2021;2021:2948–2951.
- [18] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg*. 2021;10(1):89.
- [19] Nagendrababu V, Duncan H, Tsesis I, et al. Preferred reporting items for systematic reviews and meta-analyses for abstracts: best practice for reporting abstracts of systematic reviews in endodontology. *Int Endod J*. 2019;52(8):1096–1107.
- [20] Nagendrababu V, Dilokthornsakul P, Jinatongthai P, et al. Glossary for systematic reviews and meta-analyses. *Int Endod J*. 2020;53(2):232–249.
- [21] Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004;11(2):178–189.
- [22] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.
- [23] Endres MG, Hillen F, Salloumis M, et al. Development of a deep learning algorithm for periapical disease detection in dental radiographs. *Diagnostics*. 2020;10(6):430.
- [24] Fukuda M, Inamoto K, Shibata N, et al. Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. *Oral Radiol*. 2020;36(4):337–343.
- [25] Hiraiwa T, Arijji Y, Fukuda M, et al. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. *Dentomaxillofac Radiol*. 2019;48(3):20180218.
- [26] Johari M, Esmaeili F, Andalib A, et al. Detection of vertical root fractures in intact and endodontically treated premolar teeth by designing a probabilistic neural network: an ex vivo study. *Dentomaxillofac Radiol*. 2017;46(2):20160107.
- [27] Li CW, Lin SY, Chou HS, et al. Detection of dental apical lesions using CNNs on periapical radiograph. *Sensors*. 2021;21(21):7049.
- [28] Pauwels R, Brasil DM, Yamasaki MC, et al. Artificial intelligence for detection of periapical lesions on intraoral radiographs: comparison between convolutional neural networks and human observers. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2021;131(5):610–616.
- [29] Saghiri MA, Garcia-Godoy F, Gutmann JL, et al. The reliability of artificial neural network in locating minor apical foramen: a cadaver study. *J Endod*. 2012;38(8):1130–1134.
- [30] Moidu NP, Sharma S, Chawla A, et al. Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. *Clin Oral Investig*. 2022;26(1):651–658.
- [31] Ngoc VT, Viet DH, Anh LK, et al. Periapical lesion diagnosis support system based on X-ray images using machine learning technique. *World*. 2021;12(3):190.
- [32] Campo L, Aliaga IJ, De Paz JF, et al. Retreatment predictions in odontology by means of CBR systems. *Comput Intell Neurosci*. 2016;2016:7485250.
- [33] Jeon S-J, Yun J-P, Yeom H-G, et al. Deep-learning for predicting C-shaped canals in mandibular second molars on panoramic radiographs. *Dentomaxillofac Radiol*. 2021;49:20200513.
- [34] Li Q, Chen K, Han L, et al. Automatic tooth roots segmentation of cone beam computed tomography image sequences using U-net and RNN. *J Xray Sci Technol*. 2020;28(5):905–922.
- [35] Sherwood AA, Sherwood AI, Setzer FC, et al. A deep learning approach to segment and classify c-shaped anal morphologies in mandibular second molars using cone-beam computed tomography. *J Endod*. 2021;47(12):c907–c1916.
- [36] Orhan K, Bayrakdar IS, Ezhov M, et al. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *Int Endod J*. 2020;53(5):680–689.
- [37] Shetty H, Shetty S, Kakade A, et al. Three-dimensional semi-automated volumetric assessment of the pulp space of teeth following regenerative dental procedures. *Sci Rep*. 2021;11(1):21914.
- [38] Zheng Q, Ge Z, Du H, et al. 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(1):365–373.
- [39] Kositbowornchai S, Siriteptawee S, Plermkamon S, et al. An artificial neural network for detection of simulated dental caries. *Int J CARS*. 2006;1(2):91–96.
- [40] Hatvani J, Horváth A, Michetti J, et al. Deep learning-based super-resolution applied to dental computed tomography. *IEEE Trans Radiat Plasma Med Sci*. 2019;3(2):120–128.
- [41] Lin X, Fu Y, Ren G, et al. Micro-computed tomography-guided artificial intelligence for pulp cavity and tooth segmentation on

- cone-beam computed tomography. *J Endod.* 2021;47(12): 1933–1941.
- [42] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60–88.
- [43] Tsesis I, Rosen E, Tamse A, et al. Diagnosis of vertical root fractures in endodontically treated teeth based on clinical and radiographic indices: a systematic review. *J Endod.* 2010;36(9): 1455–1458.
- [44] Meister F Jr, Lommel TJ, Gerstein H. Diagnosis and possible causes of vertical root fractures. *Oral Surg Oral Med Oral Pathol.* 1980;49(3):243–253.
- [45] Selvaraju RR, Cogswell M, Das A, et al, editors. Grad-cam: visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis.* 2017:618–626.
- [46] Dash S, Acharya BR, Mittal M, et al. *Deep learning techniques for biomedical and health informatics.* Heidelberg: Springer; 2020.
- [47] Zheng L, Wang H, Mei L, et al. Artificial intelligence in digital cariology: a new tool for the diagnosis of deep caries and pulpitis using convolutional neural networks. *Ann Transl Med.* 2021;9(9): 763.
- [48] Thrall JH, Li X, Li Q, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol.* 2018;15(3 Pt B):504–508.
- [49] Sajad M, Shafi I, Ahmad J editors. *Automatic lesion detection in periapical X-rays.* Int Conf Electr Eng. 2019:1–6.
- [50] Eid Mahmoud Y, Safwat Labib S, MO Mokhtar H, editors. *Clinical prediction of teeth periapical lesion based on machine learning techniques.* London: ICDIP; 2015.
- [51] Sebring D, Buhlin K, Norhammar A, et al. Endodontic inflammatory disease: a risk indicator for a first myocardial infarction. *Int Endod J.* 2022;55(1):6–17.
- [52] Krois J, Garcia Cantu A, Chaurasia A, et al. Generalizability of deep learning models for dental image analysis. *Sci Rep.* 2021; 11(1):6102.
- [53] Stiglic G, Kocbek P, Fijacko N, et al. Interpretability of machine learning-based prediction models in healthcare. *WIREs Rev Data Min Knowl Discov.* 2020;10(5):e1379.
- [54] Schwendicke FA, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res.* 2020;99(7): 769–774.