Effectiveness of Artificial Intelligence in Detection of Apical Periodontitis: A Systematic Review

[1]Sidhant Goyal , [2]Geetu Singh, [3]Richa Wadhawan, [4]Lipika Jain, [5]Manish Anand, [6]Nidhi Arya

[1]Senior Resident, Department of Dentistry, Bhagat Phool Singh, Government Medical College for Women, Khanpur Kalan, Haryana, India sidhantgoyal23@gmail.com

[2]Senior Resident, Department of Dentistry, Bhagat Phool Singh, Government Medical College for Women, Khanpur Kalan, Haryana, India geetusingh9019@gmail.com

[3] Professor & Head, Department of Oral Medicine & Radiology, Faculty of Dental Sciences, PDM University, Bahadurgarh, Haryana, India wadhawanricha1@gmail.com

[4]Assistant Professor, Department of Conservative Dentistry & Endodontics, Dayananda Sagar College of Dental Sciences & Hospital, Bengaluru, Karnataka, India drlipikajain@gmail.com

[5]Resident, Department of Oral & Maxillofacial Surgery, Henry M Goldman School of Dental Sciences, Boston University, Boston, Massachusetts, United States of America manishanand028@gmail.com

[6]Reader, Department of Oral & Maxillofacial Pathology and Oral Microbiology, Seema Dental College & Hospital, Rishikesh, Uttarakhand, India drnidhiarya19@gmail.com

Corresponding author: Geetu Singh, geetusingh9019@gmail.com

HOW TO CITE:

Sidhant Goyal , Geetu Singh, Richa Wadhawan, Lipika Jain, Manish Anand, Nidhi Arya (2026). Effectiveness of Artificial Intelligence in Detection of Apical Periodontitis: A Systematic Review. International Journal of Special Education, 41(20s), 20 - 36

ABSTRACT:

Background:

Artificial intelligence (AI), particularly deep learning and convolutional neural networks (CNNs), is increasingly being investigated for the detection of periapical lesions and radiographic manifestations of apical periodontitis. However, reported diagnostic performance varies according to imaging modality, AI architecture, dataset characteristics, reference standards, and validation strategies.

Objective:

To synthesize the available evidence on the diagnostic effectiveness of AI-based systems for detecting periapical lesions and radiographic manifestations of apical periodontitis on periapical, panoramic, and cone-beam computed tomography (CBCT) images, and to summarize methodological limitations and potential sources of bias.

Methods:

This systematic review was structured and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 framework. Published systematic reviews and eligible primary diagnostic-accuracy studies were evaluated, with particular attention to sensitivity, specificity, accuracy, F1 score, area under the receiver operating characteristic curve (AUROC), imaging modality, AI architecture, reference standard, and validation strategy. Study-level methodological quality and risk of bias were considered using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) framework, with additional consideration of AI-specific issues including data leakage, train-test separation, and external validation.

Results:
A total of 12 primary studies were included in the present systematic review, encompassing AI-based detection, classification, localization, segmentation, and AI-assisted clinician assessment across periapical radiographs, panoramic radiographs, and CBCT. Existing systematic reviews consistently indicated promising diagnostic performance but substantial methodological heterogeneity. A 2024 systematic review and meta-analysis included 24 studies and reported pooled sensitivity of 0.94 and specificity of 0.96 from four studies contributing quantitative data. A 2026 systematic review of deep-learning detection of periapical radiolucent lesions on panoramic radiographs included 30 studies, of which six entered meta-analysis, reporting pooled sensitivity of 0.80 (95% CI: 0.49–0.94), specificity of 0.98 (95% CI: 0.87–1.00), diagnostic odds ratio (DOR) of 176.37, and AUROC of 0.93. Risk-of-bias concerns were identified in 16 of 30 studies and applicability concerns in 12. The primary studies included in the present review demonstrated promising diagnostic and detection performance; however, external validation, representative datasets, and clinically relevant reference standards were inconsistently reported.

Conclusion:
AI demonstrates considerable potential as an adjunctive tool for detecting radiographic manifestations of apical periodontitis and periapical lesions. Nevertheless, reported high diagnostic performance should be interpreted cautiously because of substantial methodological heterogeneity, potential spectrum and selection bias, limited external validation, variability in reference standards, and other AI-specific methodological limitations. Further well-designed studies using representative clinical datasets, independent external validation, standardized reference standards, and transparent reporting are required before widespread clinical implementation.

Keywords:
Apical periodontitis, Artificial intelligence, CBCT, Convolutional neural network, Deep learning, Diagnostic accuracy, Periapical lesion, QUADAS-2, Radiographic diagnosis.

Introduction

Apical periodontitis is an inflammatory disease affecting the periapical tissues and most commonly develops as a consequence of microbial infection of the root canal system. It represents one of the major complications of pulpal infection and is frequently encountered in routine dental practice. Accurate identification of periapical inflammatory changes is essential for establishing an appropriate diagnosis, determining the need for endodontic intervention, and evaluating treatment outcomes. However, early periapical lesions may be difficult to identify because their radiographic manifestations can be subtle, particularly when conventional two-dimensional imaging (2D) is used [1].

Periapical radiography remains one of the most widely used imaging modalities for the assessment of apical periodontitis because of its accessibility, relatively low radiation exposure, and cost-effectiveness. Nevertheless, conventional radiographs have inherent limitations, including image distortion, anatomical superimposition, and restricted visualization of three-dimensional (3D) periapical structures. These limitations may result in underdiagnosis or delayed recognition of early periapical bone changes [2]. Cone-beam computed tomography (CBCT) provides 3D visualization of the maxillofacial region and can demonstrate periapical lesions that may not be readily apparent on conventional radiographs. However, its routine use is limited by considerations related to radiation exposure, cost, and availability [3].

Recent advances in artificial intelligence (AI), particularly machine learning and deep learning, have created new opportunities for automated analysis of dental images. AI-based systems can process large numbers of radiographic images and identify complex image patterns that may be difficult to recognize consistently through conventional visual assessment. Convolutional neural network (CNN) and other deep-learning architectures have demonstrated promising performance in detecting and classifying various dental and maxillofacial conditions, including periapical pathology [4, 5].

The application of AI to the detection of apical periodontitis may provide several potential advantages. Automated image analysis could assist clinicians in identifying subtle periapical radiolucencies, improve diagnostic consistency, and reduce the possibility of lesions being overlooked during routine radiographic interpretation. Studies evaluating AI-assisted detection of periapical lesions have reported encouraging levels of diagnostic accuracy, sensitivity, and specificity, although performance varies according to the imaging modality, dataset characteristics, reference standard, and algorithm employed [6,7].

Despite these promising findings, several challenges remain before AI can be routinely incorporated into endodontic diagnosis. Differences in image acquisition protocols, image quality, lesion size, anatomical location, patient characteristics, and model architecture can influence diagnostic performance. Furthermore, variations in study methodology and the absence of standardized validation procedures make direct comparison between individual AI systems difficult. The clinical applicability of AI also requires consideration of model interpretability, external validation, generalizability to diverse patient populations, and integration into existing diagnostic workflows [8].

Therefore, a systematic evaluation of the available evidence is necessary to determine the effectiveness and reliability of AI in detecting apical periodontitis. This systematic review aims to critically assess the diagnostic performance of AI-based methods for detecting the radiographic manifestations of apical periodontitis and periapical lesions, with emphasis on accuracy, sensitivity, specificity, methodological quality, external validity, and risk of bias, as well as the potential role of AI as an adjunct to conventional radiographic diagnosis.

Aim & Objectives

Aim: To systematically evaluate the effectiveness and diagnostic performance of AI-based systems for detection of apical periodontitis / periapical lesions on dental radiographic images.

Objectives:

- To identify published studies evaluating AI, machine learning or deep learning for detection, classification or segmentation of apical/periapical lesions.
- To compare AI performance across periapical radiographs, panoramic radiographs and CBCT
- To summarize sensitivity, specificity, accuracy, F1 score, precision, recall, area under the receiver operating characteristic curve (AUROC) and other reported performance measures
- To evaluate methodological quality and risk of bias using Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2)
- To identify limitations affecting clinical translation, including spectrum bias, reference-standard limitations, data leakage, lack of external validation and heterogeneity of datasets

Methodology

Review Design and Reporting

A systematic review was conducted to evaluate the diagnostic performance of AI-based methods for detecting apical periodontitis and periapical lesions on dental imaging. This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines [9]. The review evaluated AI applications for the detection, classification, localization, and segmentation of apical and periapical pathology using periapical radiographs, panoramic radiographs, and CBCT.

Review Question and PICO Framework

The review question was:

How accurately do AI-based systems detect apical periodontitis or periapical lesions on dental imaging when compared with established diagnostic or reference standards? The review question was structured using a PICO-DTA-type framework. The Population comprised patients or clinically derived dental radiographic datasets containing teeth with or without radiographic evidence of apical periodontitis or periapical lesions. The Index Test included AI, machine learning, deep learning, convolutional neural networks (CNNs), object-detection algorithms, and image-segmentation models applied to dental radiographic images. The Reference Standard included expert clinician assessment or consensus, clinical and radiographic evaluation, CBCT, histopathological assessment where applicable, or other validated diagnostic reference standards. The Target Condition was apical periodontitis and/or radiographic manifestations of periapical disease, including periapical radiolucencies. The Primary Outcomes were sensitivity, specificity, diagnostic accuracy, and AUROC. Secondary outcomes included precision, recall, F1 score, positive predictive value (PPV), negative predictive value (NPV), diagnostic odds ratio (DOR), and measures of lesion localization and segmentation performance.

Search Strategy

A comprehensive literature search was conducted in PubMed/MEDLINE, Scopus, Web of Science, Embase, and the Cochrane Library. The reference lists of included studies and relevant systematic reviews were also manually screened to identify additional eligible publications. The search strategy combined controlled vocabulary and free-text terms relating to AI, machine learning, deep learning, CNNs, apical periodontitis, periapical lesions, apical radiolucencies, and dental imaging. Boolean operators AND and OR were used to combine the search concepts. A representative search strategy was :*"("artificial intelligence" OR "machine learning" OR "deep learning" OR "convolutional neural network" OR "neural network" OR "computer-aided diagnosis") AND ("apical periodontitis" OR "periapical lesion" OR "periapical radiolucency*" OR "apical lesion") AND ("dental radiograph" OR "periapical radiograph" OR panoramic OR CBCT OR "cone-beam computed tomography")". The strategy was adapted according to the indexing requirements of each database. Publications available up to the final search date were considered for inclusion.

Eligibility Criteria

Inclusion Criteria

Studies were included when they met all relevant eligibility requirements:

- Original research involving human participants or clinically derived dental images
- Evaluation of AI, machine learning, deep learning, CNN-based, or related computational methods
- Investigation of the detection, classification, localization, or segmentation of apical periodontitis, periapical lesions, or apical radiolucencies
- Use of periapical radiographs, panoramic radiographs, CBCT, or other clinically relevant dental imaging modalities
- Reporting of at least one diagnostic or model-performance outcome, such as sensitivity, specificity, accuracy, AUROC, PPV, NPV, detection, localization, or segmentation metrics
- Use of a defined reference standard, including expert assessment, clinical or radiographic diagnosis, CBCT, histopathology where applicable, or another established reference standard
- Full-text, peer-reviewed articles published in English

Exclusion Criteria

The following were excluded:

- Case reports, case series, editorials, letters, commentaries, conference abstracts, and protocols without sufficient primary diagnostic data
- Animal, phantom, or purely technical studies without clinically relevant lesion-detection data
- Studies evaluating AI applications unrelated to apical periodontitis or periapical pathology
- Studies focused exclusively on endodontic treatment planning, prognosis, instrumentation, root-canal treatment, or other applications without lesion detection or characterization
- Narrative reviews, systematic reviews, scoping reviews, and meta-analyses without original patient-level or image-level data
- Duplicate publications or secondary analyses of the same dataset that provided no additional relevant diagnostic information
- Articles for which the full text was unavailable.

Systematic and narrative reviews were considered only for background information and reference-list screening and were not included as primary studies in the evidence synthesis.

Study Selection

The study selection process was conducted in accordance with the predefined eligibility criteria and the PRISMA 2020 framework. A total of 350 records were identified through the electronic database search. After removal of 42 duplicate records, 308 records were screened based on their titles and abstracts, of which 250 were excluded as they did not meet the predefined inclusion criteria. The full texts of the remaining 58 articles were assessed for eligibility. Of these, 46 articles were excluded for the following reasons: wrong study design or population (n = 15), AI applications unrelated to apical periodontitis or periapical pathology (n = 12), absence of relevant diagnostic outcomes (n = 9), full text unavailable (n = 5), and duplicate or secondary use of datasets without additional diagnostic information (n = 5). Finally, 12 primary studies fulfilled all eligibility criteria and were included in the systematic review. These studies evaluated AI-based detection, classification, localization, segmentation, and AI-assisted clinician assessment of apical periodontitis and periapical lesions using periapical radiographs, panoramic radiographs, and CBCT imaging. As depicted in **Figure 1**.

Data Extraction

Data were extracted using a standardized data-extraction form. The following information was recorded where available:

- First author and year of publication
- Country and study design
- Number of patients, teeth, lesions, and/or images
- Patient and lesion characteristics
- Imaging modality and acquisition characteristics
- Definition and location of apical or periapical pathology
- AI technique and computational approach
- CNN or other deep-learning architecture
- Image preprocessing and annotation procedures
- Training, validation, and testing procedures
- Internal or external validation
- Reference standard and comparator
- Sensitivity, specificity, accuracy, and AUROC
- PPV and NPV
- Detection and localization metrics
- Segmentation metrics
- Principal study findings

When multiple AI models were assessed within the same study, the performance of individual models was extracted separately where sufficient information was available. Particular attention was given to whether the test dataset was independent of the training dataset, whether patient-level separation was maintained, and whether external validation was performed.

Risk of Bias and Applicability Assessment

Risk of bias and applicability concerns were assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. The assessment comprised four domains: patient selection, index test, reference standard, and flow and timing. Risk of bias was assessed across all four domains, while applicability concerns were evaluated for patient selection, index test, and reference standard. Two reviewers independently conducted the assessments, with disagreements resolved through consensus or consultation with a third reviewer. Because AI-based diagnostic studies have additional methodological considerations, the assessment also considered dataset construction, separation of training and testing images, potential data leakage, image augmentation, test-set selection, reference-standard procedures, threshold determination, and external validation.

Outcome Measures

The primary outcome was the diagnostic performance of AI systems in detecting apical periodontitis or periapical lesions.

Primary diagnostic outcomes included:

1. Sensitivity
2. Specificity
3. Diagnostic accuracy
4. AUROC

Secondary outcomes included PPV, NPV, false-positive and false-negative rates, precision, recall, F1 score, DOR, lesion localization, segmentation performance, and differences between AI-assisted and conventional or expert-based interpretation. Segmentation outcomes, including the Dice coefficient and related overlap measures, were interpreted separately from conventional diagnostic measures because they evaluate spatial agreement rather than diagnostic sensitivity and specificity.

Data Synthesis

The findings were synthesized according to the clinical and methodological characteristics of the included studies. Differences in imaging modality, AI architecture, dataset size, lesion definition, reference standard, and validation strategy were considered when interpreting diagnostic performance. Reported sensitivity, specificity, accuracy, and AUROC values were summarized across studies. Because substantial methodological and clinical heterogeneity was anticipated, particularly with respect to AI models, imaging modalities, study populations, reference standards, and outcome definitions, the primary synthesis was structured and narrative rather than based on direct comparison of individual numerical estimates. Studies using periapical radiographs, panoramic radiographs, and CBCT were considered separately where appropriate. Detection, classification, localization, and segmentation studies were also interpreted according to their respective outcome measures. Greater emphasis was placed on results derived from independent test datasets and external validation cohorts than on performance reported solely from training or internally validated datasets.

Assessment of Heterogeneity

Heterogeneity was assessed qualitatively by examining differences in study populations, sample sizes, imaging modalities, image-acquisition protocols, AI architectures, lesion definitions, reference standards, preprocessing methods, annotation procedures, and validation strategies. Potential sources of variation in reported diagnostic performance included differences in radiographic modality, lesion size and characteristics, dataset composition, diagnostic thresholds, expert reference standards, and model-validation procedures. These factors were considered when interpreting differences in performance among studies.

Publication Bias and Selective Reporting

The possibility of publication bias and selective reporting was considered when interpreting the evidence. Particular attention was given to studies with small or highly selected datasets and to the completeness of reporting of diagnostic outcomes. A formal statistical assessment of publication bias was not performed because the anticipated number and methodological diversity of studies were considered insufficient to support a reliable quantitative assessment.

Ethical Considerations

As this systematic review was based exclusively on previously published studies and did not involve the recruitment of human participants, collection of new patient data, or access to identifiable patient information, institutional ethical approval and informed consent were not required. The review was conducted using publicly available published data in accordance with established systematic-review methodology.

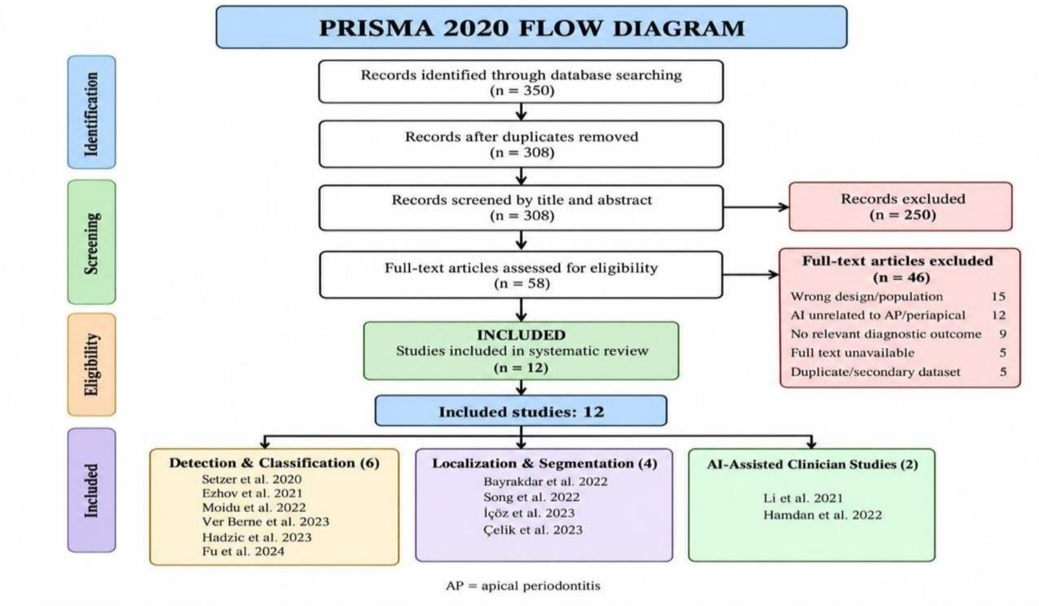

Figure 1: PRISMA flowchart depicting the methodology

Results

The findings of the present review are broadly in agreement with those reported in recent systematic reviews investigating the use of AI for the detection of periapical pathology. A systematic review published in 2026 evaluated deep-learning approaches for identifying periapical radiolucent lesions on panoramic radiographs. The authors searched six electronic databases and grey-literature sources up to March 18, 2026, and included 30 studies, of which six were eligible for quantitative synthesis. The pooled sensitivity was 0.80 (95% CI: 0.49–0.94), whereas the pooled specificity was 0.98 (95% CI: 0.87–1.00).

The pooled DOR was 176.37 (95% CI: 20.48–1515.70), and the pooled AUROC was 0.93 (95% CI: 0.82–0.98). Although these findings indicate strong diagnostic potential, considerable heterogeneity was observed among the included studies. Risk-of-bias concerns were identified in 16 studies, while 12 studies raised applicability concerns, resulting in an overall moderate certainty of evidence. These observations are also consistent with a 2024 systematic review and meta-analysis that investigated AI-assisted detection of periapical radiolucencies. Of 210 records initially identified, 24 studies met the eligibility criteria, although only four provided sufficient data for quantitative analysis. The pooled sensitivity was 0.94 (95% CI: 0.90–0.96), and the pooled specificity was 0.96 (95% CI: 0.91–0.98). Despite the high pooled diagnostic estimates, the authors emphasized important methodological limitations, including unclear or high risk of bias and concerns regarding the applicability of the study populations, index tests, and reference standards. Earlier evidence from a 2022 systematic review and meta-analysis similarly demonstrated promising performance of deep-learning systems for detecting periapical radiolucent lesions. The review included 18 studies, six of which were suitable for quantitative synthesis. The pooled sensitivity was 0.925 (95% CI: 0.862–0.960), while the pooled specificity was 0.852 (95% CI: 0.810–0.885). The reported DOR was 71.692. However, methodological weaknesses remained evident, with risk-of-bias concerns identified in 12 studies. **Table 1** summarizes the individual original research studies included in systematic review.

Table 1: Characteristics and Diagnostic Performance of the Included Primary Studies Evaluating AI for Periapical Disease

Study	**Country**	**Imaging**	**AI approach**	**Dataset**	**Reported performance**	**Key point**
A. Detection and Classification Studies						
Setzer et al., 2020 [10]	United States	CBCT	AI-based computer-aided detection	61 roots	Accuracy 93%; specificity 88%	Demonstrated the feasibility of AI-assisted detection of preriapical lesions on CBCT
Ezhov et al., 2021 [11]	United States	CBCT	AI-based dental diagnostic system	1,346 CBCT scans for system development; clinical evaluation included 30 scans	Overall sensitivity 85.4%; specificity 96.7% in the AI-assisted diagnostic evaluation	Demonstrated promising AI-assisted diagnostic performance on CBCT, although the reported metrics were not limited exclusively to apical periodontitis
Moidu et al., 2022 [12]	India	Periapical	Deep-learning/ You Only Look Once (YOLO)-based classification	540 periapical radiographs for testing	Sensitivity 92.1%; specificity 76%; accuracy 86.3%; F1 score approximately 0.89	Demonstrated strong detection/ classification performance for apical periodontitis using periapical radiographs
Ver Berne et al., 2023 [13]	Belgium	Panoramic	Deep learning CNN	80 cysts, 72 granulomas + controls	Cyst sensitivity 1.00, specificity 0.95; granuloma sensitivity 0.77, specificity 1.00	Demonstrated the potential of AI to differentiate periapical lesion types, particularly radicular cysts and granulomas
Hadzic et al., 2023 [14]	Austria	CBCT	CNN-based analysis	Clinically representative CBCT dataset	Sensitivity 86.7%; specificity 84.3%	Highlighted the importance of clinically representative datasets for evaluating generalizability

Fu et al., 2024 [15]	China	CBCT	3D CNN (PAL-Net)	4,951 teeth: 563 apical periodontitis and 4,388 non-apical periodontitis	Sensitivity 88–94%	Demonstrated robust performance using independent external datasets and highlighted the value of external validation
B. Localization and Segmentation Studies						
Bayrakdar et al., 2022 [16]	Türkiye	Panoramic	U-shaped Network Deep Convolutional Neural Network (U-Net/D-CNN)	470 panoramic radiographs	Sensitivity 0.92; precision 0.84; F1 0.88	Demonstrated promising automatic localization and segmentation of periapical lesions
Song et al., 2022 [17]	China	Panoramic	U-Net	1,000 panoramic radiographs	F1 0.828 at IoU 0.3	Supported automated localization and segmentation of periapical lesions on panoramic radiographs
İçöz et al., 2023 [18]	Türkiye	Panoramic	YOLO-based Computer-Aided Detection (CAD)	306 panoramic radiographs; 400 AP roots	Recall 0.98; precision 0.56; F1 0.71	Demonstrated high recall for AP detection, with better performance in mandibular than maxillary roots
Çelik et al., 2023 [19]	Türkiye	Panoramic	10 deep-learning detection frameworks	357 panoramic radiographs; 454 lesions/objects	mAP 0.832–0.953; accuracy 0.673–0.812; F1 0.80–0.895	Demonstrated variability among detection architectures, with Retina Net showing strong overall performance
C. AI-Assisted Clinician Studies						
Li et al., 2021	Taiwan	Periapic	CNN-bas	1,000		Demonstrated that a

		radiographs	detection of dental apical lesions	Periapical radiographs	Accuracy 0.86; sensitivity 0.88; specificity 0.84	CNN could automatically detect dental api deep-learning approaches for automated periapical lesion detection
Hamdan et al., 2022 [21]	United States	Periapical	Deep-learning diagnostic assistance tool	68 periapical radiographs with CBCT validation	Sensitivity 85% for 2–5 mm lesions; 92% for ≥5 mm lesions	Demonstrated that diagnostic performance varied according to lesion size and supported the potential role of AI as an adjunct to clinician interpretation

Risk-of-bias synthesis

The available systematic reviews indicate that methodological concerns are common. The 2024 review described the overall evidence as unclear-to-high risk of bias with applicability concerns. The 2026 panoramic-radiograph review found risk of bias in 16 of 30 studies (53.3%) and applicability concerns in 12 of 30 (40%). These findings support formal study-level QUADAS-2 assessment in the final review rather than interpreting reported accuracy values in isolation as depicted in Table 2.

Table 2: A Modified Assessment Framework

QUADAS-2 domain	Main concern to assess in AI studies	Likely consequence	Final-review requirement
Patient selection	Convenience/retrospective datasets; exclusion of difficult images; narrow lesion spectrum	Spectrum and selection bias; inflated accuracy	Assess consecutiveness, case-control design and representative sampling
Index test	Threshold selection after testing; lack of blinding; internal-only validation; data leakage	Optimistic performance estimates	Verify independent test data, pre-specified thresholds and leakage prevention
Reference standard	Expert consensus without independent confirmation; inconsistent reference standards	Misclassification and incorporation bias	Assess appropriateness and independence of reference standard
Flow and timing	Incomplete cases; exclusions after model testing; different	Attrition and verification bias	Track all images/teeth and timing between index

	datasets for reference and index test		test and reference

Discussion

The present systematic review demonstrates that AI, particularly deep-learning-based approaches, has considerable potential for detecting and characterizing apical and periapical lesions on dental radiographs. Across periapical, panoramic, and CBCT imaging, the included studies generally reported encouraging diagnostic or segmentation performance. These findings suggest that AI may serve as a useful supportive technology for identifying radiographic abnormalities and assisting clinicians during image interpretation.

The pooled findings reported in previous systematic reviews provide further context for the results of the present review. The 2024 systematic review and meta-analysis reported pooled sensitivity and specificity of 0.94 and 0.96, respectively, whereas Septina and Kiswanjaya (2026) [22] reported pooled sensitivity of 0.80 and specificity of 0.98 for studies evaluating deep-learning-based detection of periapical lesions. Differences in inclusion criteria, imaging modality, AI task, study populations, reference standards, and availability of complete 2×2 diagnostic data may explain some of this variation. High heterogeneity limits direct comparison of individual algorithms and indicates that pooled estimates should be interpreted within the methodological context of the contributing studies. The imaging modality appears to be an important determinant of AI performance. Periapical radiographs offer detailed visualization of individual teeth and the surrounding periapical region, while panoramic radiographs provide broader anatomical coverage but are more susceptible to magnification, distortion, superimposition, and positioning-related variation. CBCT provides 3D information and may improve visualization of lesions that are difficult to appreciate on 2D images. However, its higher radiation exposure, larger data volume, and greater technical complexity distinguish it from conventional dental radiography. Consequently, diagnostic performance should not be attributed to the AI algorithm alone, as it is also influenced by the characteristics of the imaging modality and the conditions under which the images are acquired.

The application of AI to panoramic radiographs has demonstrated potential beyond simple lesion detection. Orhan et al. (2023) [23] evaluated AI-assisted diagnosis and treatment-related assessment using panoramic radiographs, supporting the broader use of automated image interpretation in dental practice. Such systems may assist clinicians by drawing attention to radiographic abnormalities that could otherwise be overlooked. However, panoramic images remain susceptible to anatomical overlap and image-quality variation, factors that can influence both human interpretation and AI predictions.

CBCT-based approaches provide additional opportunities for lesion localization and segmentation. Huang et al. (2024) [24] investigated an uncertainty-based active-learning framework using Bayesian U-Net for multi-label segmentation of CBCT images. This approach illustrates the potential of advanced AI methods to move beyond binary classification toward more detailed identification and delineation of relevant structures. Similarly, Wang et al. (2021) [25] demonstrated the use of deep-learning-based segmentation for detecting oral lesions on CBCT images. Automated segmentation may ultimately assist clinicians in determining lesion boundaries, assessing spatial relationships, and supporting treatment planning. Nevertheless, segmentation performance is strongly dependent on the quality and consistency of the annotations used for model development and evaluation.

Lesion size represents another clinically important consideration. Small periapical lesions may be difficult to recognize because early mineralized tissue changes can be subtle on two-dimensional radiographs and may overlap with normal anatomical structures. Therefore, models developed predominantly using clearly visible or larger lesions may overestimate their ability to identify early disease. Future diagnostic studies should provide performance estimates according to lesion size and deliberately include radiographically subtle lesions to determine whether AI can reliably identify early periapical changes.

A further concern is the transferability of AI performance from research datasets to routine clinical practice. Many available investigations are retrospective and based on institutional or curated image collections. Such datasets may not adequately represent the variation encountered in everyday dental practice. Differences in patient characteristics, disease prevalence, imaging equipment, acquisition parameters, image quality, and anatomical presentation can affect model performance. Consequently, strong performance on an internal test set should not automatically be interpreted as evidence of clinical effectiveness in other settings.

The methodological characteristics of the available evidence also warrant consideration. Although high diagnostic estimates have been reported, variation in study design, reference standards, annotation procedures, and validation strategies makes the evidence difficult to compare directly. The relatively small number of prospective clinical investigations is particularly important because retrospective studies may not reproduce the workflow, patient spectrum, and diagnostic uncertainty encountered in clinical practice. More clinically representative research is therefore required to establish whether reported technical performance translates into meaningful improvements in diagnosis.

From a clinical perspective, the most appropriate current role of AI is likely to be as a supportive rather than autonomous technology. AI may assist with radiographic screening, lesion detection, localization, segmentation, and prioritization of images for detailed clinician assessment. Such applications could be particularly valuable in settings with large volumes of radiographic examinations or where subtle abnormalities may be overlooked. Nevertheless, radiographic evidence alone does not establish the complete diagnosis of apical periodontitis. Interpretation should remain integrated with clinical findings, symptoms, pulp sensibility testing, periodontal assessment, previous treatment history, and other relevant diagnostic information.

Overall, the evidence supports the feasibility and potential clinical usefulness of AI for the detection of apical and periapical disease. However, reported performance should be interpreted in relation to the characteristics of the datasets, imaging techniques, lesion definitions, and validation methods used in individual studies. The transition from promising experimental performance to dependable clinical application will require stronger evidence from representative clinical populations and independent testing environments.

Limitations

This systematic review has several limitations. Considerable methodological variation existed among the included studies with respect to AI architectures, imaging modalities, lesion definitions, datasets, preprocessing techniques, annotation methods, and outcome measures. This variation restricted direct comparison of diagnostic performance and contributed to heterogeneity across the evidence base.

Differences in reference standards and diagnostic criteria may also have influenced reported sensitivity, specificity, accuracy, and segmentation outcomes. In addition, many studies relied on retrospective or institution-specific datasets, which may introduce selection and spectrum bias. The frequent use of internal validation rather than independent external testing further limits confidence in the generalizability of the reported results. Quantitative synthesis was also constrained by incomplete reporting of diagnostic data. In particular, complete 2×2 contingency tables were not consistently available across studies, reducing the number of investigations that could contribute to pooled diagnostic estimates. Furthermore, metrics such as Dice coefficient, F1 score, and mean average precision measure different aspects of segmentation or detection performance and cannot be directly interpreted as equivalents of sensitivity or specificity.

Finally, AI technology is developing rapidly, meaning that newer architectures may differ substantially from models evaluated in earlier investigations. Consequently, the findings represent the current available evidence and may require reassessment as additional high-quality clinical studies become available.

Clinical Implications

The findings indicate that AI has the potential to support radiographic assessment of apical periodontitis by assisting clinicians in recognizing, localizing, and characterizing periapical abnormalities. Automated analysis may be particularly useful for screening large numbers of radiographs and highlighting areas that warrant closer examination.

CBCT-based AI systems may additionally provide information regarding lesion location and extent, potentially contributing to treatment planning in selected cases. Segmentation-based approaches could facilitate more objective assessment of lesion boundaries and longitudinal changes following endodontic treatment.

However, implementation in clinical practice should be cautious. AI predictions may be affected by image quality, acquisition parameters, lesion morphology, and differences between the population used for model development and the patients encountered in practice. False-positive and false-negative results therefore remain possible. AI outputs should be reviewed by appropriately trained clinicians and interpreted alongside clinical and radiographic findings rather than being used as an independent diagnostic decision.

Future Perspectives

Future research should prioritize large, multicenter datasets that represent the demographic, anatomical, and radiographic diversity encountered in routine dental practice. Studies should include lesions of different sizes and degrees of radiographic conspicuity to establish whether AI can reliably identify both early and well-established disease. Prospective clinical studies and independent external validation should become central components of future AI research. Models should be evaluated using datasets obtained from different institutions, imaging systems, and acquisition protocols to determine their robustness under real-world conditions. Standardized lesion definitions, reference standards, annotation procedures, and reporting of diagnostic outcomes would also facilitate meaningful comparison between AI systems. An important future direction is the development of multimodal AI systems capable of combining radiographic information with relevant clinical variables such as symptoms, pulp-testing results, periodontal findings, treatment history, and other patient-level information. Such approaches may provide a more comprehensive assessment than image analysis alone. Explainable AI may further improve clinical acceptance by indicating the image regions or features that contribute to a model's prediction. Future studies should also evaluate clinically meaningful outcomes, including changes in diagnostic accuracy, treatment planning, clinician workload, diagnostic efficiency, and patient outcomes, rather than relying exclusively on technical performance metrics. Before widespread clinical implementation, AI systems should additionally be assessed for calibration, robustness to changes in datasets, computational requirements, cost-effectiveness, interoperability with existing dental imaging systems, and ethical and data-governance considerations. These factors will determine whether technically successful AI models can provide reliable and sustainable benefits in everyday dental practice.

Conclusion

AI, particularly deep-learning-based methods, requires further evaluation to determine its appropriate integration into routine radiographic assessment and clinical diagnostic workflows. Current evidence suggests that these technologies may assist clinicians with lesion detection, localization, segmentation, and radiographic screening. Nevertheless, substantial methodological heterogeneity, retrospective study designs, variable reference standards, limited external validation, and incomplete diagnostic reporting restrict the certainty and generalizability of the available evidence. At present, AI should therefore be regarded as a supportive technology that complements rather than replaces clinical expertise. Well-designed multicenter prospective studies with standardized methodology and independent external validation are required to establish its true clinical effectiveness and determine its appropriate integration into routine diagnosis and treatment planning for apical periodontitis.

References

Gomes BPFA, Herrera DR. Etiologic role of root canal infection in apical periodontitis and its relationship with clinical symptomatology. Braz Oral Res. 2018 Oct 18;32(suppl 1):e69. doi: 10.1590/1807-3107bor-2018.vol32.0069. PMID: 30365610.

2.Stera G, Giusti M, Magnini A, Calistri L, Izzetti R, Nardi C. Diagnostic accuracy of periapical radiography and panoramic radiography in the detection of apical periodontitis: a systematic review and meta-analysis. Radiol Med. 2024 Nov;129(11):1682-1695. doi: 10.1007/s11547-024-01882-z. Epub 2024 Sep 3. PMID: 39225920; PMCID: PMC11554819.

Sadr S, Mohammad-Rahimi H, Motamedian SR, Zahedrozegar S, Motie P, Vinayahalingam S, Dianat O, Nosrat A. Deep Learning for Detection of Periapical Radiolucent Lesions: A Systematic Review and Meta-analysis of Diagnostic Test Accuracy. J Endod. 2023 Mar;49(3):248-261.e3. doi: 10.1016/j.joen.2022.12.007. Epub 2022 Dec 21. PMID: 36563779.

4.Pul U, Schwendicke F. Artificial intelligence for detecting periapical radiolucencies: A systematic review and meta-analysis. J Dent. 2024 Aug;147:105104. doi: 10.1016/j.jdent.2024.105104. Epub 2024 Jun 6. PMID: 38851523.

Ari T, Sağlam H, Öksüzoğlu H, Kazan O, Bayrakdar İŞ, Duman SB, Çelik Ö, Jagtap R, Futyma-Gąbka K, Różyło-Kalinowska I, Orhan K. Automatic Feature Segmentation in Dental Periapical Radiographs. Diagnostics (Basel). 2022 Dec 7;12(12):3081. doi: 10.3390/diagnostics12123081. PMID: 36553088; PMCID: PMC9777016.

Chen H, Li H, Zhao Y, Zhao J, Wang Y. Dental disease detection on periapical radiographs based on deep convolutional neural networks. Int J Comput Assist Radiol Surg. 2021 Apr;16(4):649-661. doi: 10.1007/s11548-021-02319-y. Epub 2021 Mar 2. PMID: 33651298.

Calazans, M.A.A.; Ferreira, F.A.B.S.; Alcoforado, M.d.L.M.G.; Santos, A.d.; Pontual, A.d.A.; Madeiro, F. Automatic Classification System for Periapical Lesions in Cone-Beam Computed Tomography. Sensors 2022, 22, 6481. https://doi.org/10.3390/s22176481

Ahmad, I. A., R. Ba-Hattab, S. A. Al-Maweri, and F. Tamimi. 2026. " Deep Learning Models for Detection of Periapical Radiolucent Lesions on Panoramic Radiographs: A Systematic Review and Meta-Analysis." International Endodontic Journal 1–39. https://doi.org/10.1111/iej.70207.

9.Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;372:n71. doi: 10.1136/bmj.n71. PMID: 33782057; PMCID: PMC8005924.

Setzer FC, Shi KJ, Zhang Z, Yan H, Yoon H, Mupparapu M, Li J. Artificial Intelligence for the Computer-aided Detection of Periapical Lesions in Cone-beam Computed Tomographic Images. J Endod. 2020 Jul;46(7):987-993. doi: 10.1016/j.joen.2020.03.025. Epub 2020 May 8. Erratum in: J Endod. 2026 Jan;52(1):156. doi: 10.1016/j.joen.2025.09.022. PMID: 32402466.

Ezhov, M., Gusarev, M., Golitsyna, M. et al. Clinically applicable artificial intelligence system for dental diagnosis with CBCT. Sci Rep 11, 15006 (2021). https://doi.org/10.1038/s41598-021-94093-923.

Moidu NP, Sharma S, Chawla A, Kumar V, Logani A. Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. Clin Oral Investig. 2022 Jan;26(1):651-658. doi: 10.1007/s00784-021-04043-y. Epub 2021 Jul 2. PMID: 34213664.

Ver Berne J, Saadi SB, Politis C, Jacobs R. A deep learning approach for radiological detection and classification of radicular cysts and periapical granulomas. J Dent. 2023 Aug;135:104581. doi: 10.1016/j.jdent.2023.104581. Epub 2023 Jun 7. PMID: 37295547.

Hadzic A, Urschler M, Press JA, Riedl R, Rugani P, Štern D, Kirnbauer B. Evaluating a Periapical Lesion Detection CNN on a Clinically Representative CBCT Dataset-A Validation Study. J Clin Med. 2023 Dec 29;13(1):197. doi: 10.3390/jcm13010197. PMID: 38202204; PMCID: PMC10779652.

Fu WT, Zhu QK, Li N, Wang YQ, Deng SL, Chen HP, Shen J, Meng LY, Bian Z. Clinically Oriented CBCT Periapical Lesion Evaluation via 3D CNN Algorithm. J Dent Res. 2024 Jan;103(1):5-12. doi: 10.1177/00220345231201793. Epub 2023 Nov 15. PMID: 37968798.

Bayrakdar IS, Orhan K, Çelik Ö, Bilgir E, Sağlam H, Kaplan FA, Görür SA, Odabaş A, Aslan AF, Różyło-Kalinowska I. A U-Net Approach to Apical Lesion Segmentation on Panoramic Radiographs. Biomed Res Int. 2022 Jan 15;2022:7035367. doi: 10.1155/2022/7035367. PMID: 35075428; PMCID: PMC8783705.

Song IS, Shin HK, Kang JH, Kim JE, Huh KH, Yi WJ, Lee SS, Heo MS. Deep learning-based apical lesion segmentation from panoramic radiographs. Imaging Sci Dent. 2022 Dec;52(4):351-357. doi: 10.5624/isd.20220078. Epub 2022 Jul 23. PMID: 36605863; PMCID: PMC9807797.

Icoz D, Terzioglu H, Ozel MA, Karakurt R. Evaluation of an artificial intelligence system for the diagnosis of apical periodontitis on digital panoramic images. Niger J Clin Pract. 2023 Aug;26(8):1085-1090. doi: 10.4103/njcp.njcp_624_22. PMID: 37635600.

Çelik B, Savaştaer EF, Kaya HI, Çelik ME. The role of deep learning for periapical lesion detection on panoramic radiographs. Dentomaxillofac Radiol. 2023 Nov;52(8):20230118. doi: 10.1259/dmfr.20230118. Epub 2023 Oct 18. PMID: 37641964; PMCID: PMC10968763.

Li CW, Lin SY, Chou HS, Chen TY, Chen YA, Liu SY, Liu YL, Chen CA, Huang YC, Chen SL, Mao YC, Abu PAR, Chiang WY, Lo WS. Detection of Dental Apical Lesions Using CNNs on Periapical Radiograph. Sensors (Basel). 2021 Oct 24;21(21):7049. doi: 10.3390/s21217049. PMID: 34770356; PMCID: PMC8588190.

Hamdan MH, Tuzova L, Mol A, Tawil PZ, Tuzoff D, Tyndall DA. The effect of a deep-learning tool on dentists' performances in detecting apical radiolucencies on periapical radiographs. Dentomaxillofac Radiol. 2022 Sep 1;51(7):20220122. doi: 10.1259/dmfr.20220122. Epub 2022 Sep 12. PMID: 35980437; PMCID: PMC9522978.

Septina F, Kiswanjaya B. Evaluating Deep Learning AI for Periapical Lesion Detection Across Panoramic, Periapical, and CBCT Radiographs: A Systematic Review., 2026; 20: e18742106391708. http://dx.doi.org/10.2174/0118742106391708260406104035

Orhan K, Aktuna Belgin C, Manulis D, Golitsyna M, Bayrak S, Aksoy S, Sanders A, Önder M, Ezhov M, Shamshiev M, Gusarev M, Shlenskii V. Determining the reliability of diagnosis and treatment using artificial intelligence software with panoramic radiographs. Imaging Sci Dent. 2023 Sep;53(3):199-208. doi: 10.5624/isd.20230109. Epub 2023 Aug 2. PMID: 37799743; PMCID: PMC10548159.

Huang J, Farpour N, Yang BJ, Mupparapu M, Lure F, Li J, Yan H, Setzer FC. Uncertainty-based Active Learning by Bayesian U-Net for Multi-label Cone-beam CT Segmentation. J Endod. 2024 Feb;50(2):220-228. doi: 10.1016/j.joen.2023.11.002. Epub 2023 Nov 17. PMID: 37979653; PMCID: PMC10842728.

Wang X, Meng X, Yan S. Deep Learning-Based Image Segmentation of Cone-Beam Computed Tomography Images for Oral Lesion Detection. J Healthc Eng. 2021 Sep 21;2021:4603475. doi: 10.1155/2021/4603475. PMID: 34594482; PMCID: PMC8478545.