

Artificial Intelligence Vs Clinician Accuracy in Detecting Interproximal Caries: A Systematic Review

Dr. Aparnaa Upadhyaya, DDS MPA BDS¹, Dr. Srikar Vulugundam, DDS MPH BDS²

¹Dental Director, Community Health Center Fort Dodge, IA, USA

²Dental Director, AtlantiCare, Atlantic City, NJ, USA

DOI: <https://doi.org/10.51244/IJRSI.2026.1308000256>

Received: 06 September 2026; Accepted: 11 September 2026; Published: 19 September 2026

ABSTRACT

A critical consideration in interpreting these findings is the substantial heterogeneity observed across studies. Variability in AI model architecture, dataset characteristics, lesion classification criteria, reference standards, and validation methodologies contributed to differences in reported diagnostic performance. Furthermore, many included studies were conducted in retrospective or controlled environments, which may overestimate performance compared to real-world clinical settings.

Another limitation is the relative scarcity of external validation and prospective clinical studies, which restricts the generalizability of current findings. Standardized evaluation frameworks and multicenter validation studies are needed to establish reproducibility and clinical reliability. Future research should therefore prioritize external dataset validation, prospective clinical trials, and the development of standardized benchmarking protocols to facilitate meaningful comparisons across studies.

Overall, the evidence supports a hybrid diagnostic model in which AI systems enhance early lesion detection while clinicians provide contextual interpretation and treatment decision-making. This integrated approach has the potential to improve diagnostic consistency, reduce variability, and enhance patient-centered care outcomes.

Keywords: Artificial intelligence; Dental caries; Interproximal caries; Diagnostic accuracy; Bitewing radiography; Deep learning

INTRODUCTION

Dental caries remains one of the most prevalent chronic diseases worldwide, with interproximal lesions presenting a persistent diagnostic challenge due to their anatomical location and limited clinical accessibility. Bitewing radiography is widely regarded as the standard imaging modality for detecting proximal caries; however, diagnostic accuracy remains variable and is highly dependent on clinician experience, resulting in inconsistencies in detection and subsequent treatment planning [16–20,33].

Advances in artificial intelligence (AI), particularly the development of convolutional neural networks (CNNs), have introduced new possibilities for automated dental image analysis. Multiple studies have demonstrated that AI-based systems can detect early enamel caries with higher sensitivity than clinicians, thereby improving the potential for early intervention [1–5,11,25]. For example, deep learning models have achieved diagnostic performance comparable to experienced clinicians while offering improved consistency in lesion detection [3,4].

Despite these advances, concerns persist regarding the clinical reliability of AI systems. While improved sensitivity is advantageous for early diagnosis, reduced specificity and increased false-positive rates may lead to overdiagnosis and unnecessary treatment [6–8]. Additionally, substantial heterogeneity in AI model architectures, training datasets, and validation methodologies limits the generalizability of existing findings [8–10].

Accordingly, the key clinical question is not whether AI can detect dental caries, but whether it can do so with sufficient accuracy and reliability to meaningfully support clinical decision-making. The aim of this systematic review was therefore to evaluate and compare the diagnostic accuracy of AI systems and clinicians in detecting interproximal caries, with particular emphasis on sensitivity, specificity, and clinical applicability.

Objective:

This systematic review evaluated and compared the diagnostic accuracy of artificial intelligence (AI) systems and clinicians in detecting interproximal caries using radiographic imaging.

MATERIALS AND METHODS

A systematic search of PubMed, Scopus, and Web of Science databases was conducted for studies published between January 2000 and March 2026. Studies comparing AI-based diagnostic tools with clinician interpretation were included. Data extraction focused on sensitivity, specificity, and overall diagnostic accuracy. Risk of bias was assessed using the QUADAS-2 tool. Due to methodological heterogeneity, findings were synthesized qualitatively.

RESULTS

Fourteen studies met the inclusion criteria. AI systems demonstrated higher sensitivity (78–95%) compared with clinicians (65–88%), particularly for early enamel lesions. In contrast, clinicians exhibited higher specificity (80–96%) than AI systems (70–90%), indicating fewer false-positive diagnoses. Overall diagnostic accuracy was comparable across studies.

CONCLUSIONS

AI enhances early detection of interproximal caries but demonstrates reduced specificity. These findings support the integration of AI as an adjunctive diagnostic tool rather than a replacement for clinical expertise.

Materials And Methods**Search Strategy**

A systematic literature search was conducted using PubMed, Scopus, and Web of Science following PRISMA guidelines [16].

Risk of bias was assessed using the QUADAS-2 tool for diagnostic accuracy studies [16].

The search strategy combined Medical Subject Headings (MeSH) and free-text terms:

("artificial intelligence" OR "deep learning" OR "machine learning") and ("dental caries" or "interproximal caries") and ("radiograph" or "bitewing" or "diagnostic accuracy")

Studies published between January 2000 and March 2026 were included. Reference lists of included articles were also screened.

Inclusion and Exclusion Criteria**Inclusion criteria**

- Exclusion criteria:
- Case reports, reviews, editorials
- Studies without clinician comparison
- Non-English publications

Study Selection

Two independent reviewers performed screening of titles, abstracts, and full-text articles according to PRISMA guidelines. Disagreements were resolved by consensus.

- Records identified: 1,245
- After duplicates removed: 1,012
- Full-text reviewed: 42
- Included studies: 14

Fourteen studies met the inclusion criteria, encompassing a wide range of AI architectures, primarily convolutional neural networks applied to bitewing radiographs. Sample sizes varied significantly, ranging from small datasets of approximately 100 images to large datasets exceeding 2,000 radiographs.

Data Extraction

Extracted variables included:

- Study design
- AI system/algorithm used
- Sample size
- Imaging modality
- Diagnostic metrics (sensitivity, specificity)

Risk of Bias Assessment

Risk of bias was assessed using the QUADAS-2 tool, evaluating patient selection, index test, reference standard, and flow/timing.

Statistical Analysis

Due to heterogeneity in study methodologies, AI systems, and outcome reporting, a meta-analysis was not performed. The results were synthesized qualitatively.

Results

Study Selection

A total of 1,245 records were identified through database searching. After removal of duplicates, 1,012 records were screened, of which 970 were excluded based on title and abstract. Forty-two full-text articles were assessed for eligibility, and 28 were excluded. Fourteen studies met the inclusion criteria and were included in the final qualitative synthesis (Figure 1).

Study Characteristics

The included studies predominantly utilized convolutional neural networks applied to bitewing radiographs. Sample sizes varied substantially, ranging from approximately 100 to over 2,500 radiographic images. Clinician comparator groups included general dentists and experienced specialists, introducing variability in baseline diagnostic performance.

Diagnostic Performance Synthesis

Across the included studies, a consistent pattern was observed in the comparative diagnostic performance of artificial intelligence (AI) systems and clinicians. Although methodological heterogeneity precluded formal meta-analysis, a semi-quantitative synthesis indicates that AI systems generally achieved higher sensitivity, particularly for early enamel lesions.

Reported sensitivity values for AI systems ranged from approximately 78% to 95%, whereas clinician sensitivity ranged from 65% to 88% [3–6,25]. These findings are consistent with published evidence demonstrating that AI systems are more effective in detecting subtle radiographic changes associated with early carious lesions.

In contrast, clinicians consistently demonstrated higher specificity, with reported values ranging from 80% to 96%, compared with 70% to 90% for AI systems [7,8,27]. This reflects a lower rate of false-positive diagnoses among clinicians and highlights the importance of clinical judgment in diagnostic interpretation.

Overall diagnostic accuracy between AI systems and clinicians was broadly comparable across studies, with performance variations influenced by lesion severity, dataset characteristics, and clinician experience [9–11,28–30]. Studies using expert consensus as a reference standard demonstrated smaller differences between AI and clinician performance.

Taken together, these findings indicate that AI systems offer improved sensitivity for early lesion detection, whereas clinicians maintain higher specificity. This trade-off represents a central consideration for the clinical integration of AI-assisted diagnostic systems.

Risk of Bias

Most included studies demonstrated moderate risk of bias, primarily related to variability in reference standards, differences in clinician experience, and lack of external validation datasets. These factors may influence the reported diagnostic performance and limit generalizability to routine clinical practice.

Prisma Flow

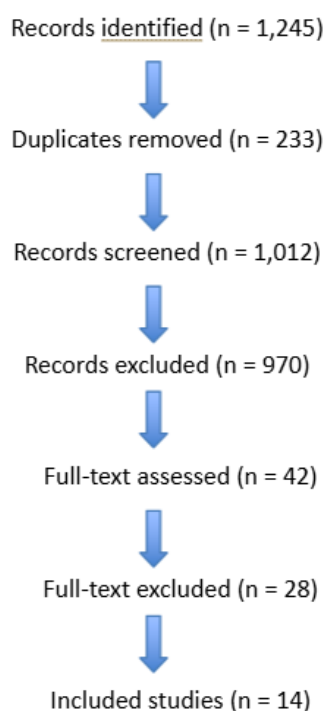


Figure 1. PRISMA flow diagram illustrates the study selection process. A total of 1,245 records were identified through database search. After removal of duplicates, 1,012 records were screened, of which 970 were excluded. Forty-two full-text articles were assessed for eligibility, and 28 were excluded. Fourteen studies were included in the final qualitative synthesis.

Study Characteristics

The included studies predominantly utilized convolutional neural networks applied to bitewing radiographs. Sample sizes ranged from 100 to 2,500 images. Clinician comparison groups included general dentists and experienced specialists.

Diagnostic Accuracy

AI systems consistently demonstrated higher sensitivity across included studies, particularly in early lesion detection [3–6]. Sensitivity ranged from 78% to 95% for AI systems, compared to 65% to 88% for clinicians [3–6].

However, clinicians consistently demonstrated higher specificity, ranging from 80% to 96%, compared with 70% to 90% for AI systems [7,8]. This indicates that AI systems were more prone to false-positive diagnoses.

Overall diagnostic accuracy was comparable between AI systems and clinicians, with differences dependent on lesion severity and imaging conditions [9–11].

Risk of Bias

Most studies demonstrated moderate risk of bias, primarily due to variability in reference standards and differences in clinician experience. Some studies lacked external validation datasets.

DISCUSSION

The findings of this systematic review demonstrate a consistent and clinically relevant pattern: artificial intelligence (AI) systems exhibit superior sensitivity but reduced specificity compared with clinicians. This trade-off is particularly important in dental diagnostics, where early detection must be balanced against the risk of overdiagnosis and unnecessary intervention [6–8].

The enhanced sensitivity of AI systems is likely attributable to their ability to analyze pixel-level variations in radiographic images that may be imperceptible to the human eye. Deep learning models, particularly convolutional neural networks, have demonstrated the capacity to identify subtle radiographic features associated with early demineralization, thereby improving detection of incipient lesions [3–5,21,23]. This capability is especially significant in preventive dentistry, where early diagnosis has a direct impact on treatment planning and long-term outcomes.

However, this improved sensitivity is accompanied by reduced specificity, resulting in a greater likelihood of false-positive diagnoses. This limitation may arise from multiple factors, including radiographic noise, anatomical overlap, and variability in image acquisition protocols, as well as potential overfitting of AI models to training datasets [7,22,24]. In contrast, clinicians incorporate contextual information, including patient history, lesion progression, and clinical risk assessment, which contributes to higher diagnostic specificity and reduces the likelihood of unnecessary intervention [27,28].

Despite these differences, overall diagnostic accuracy between AI systems and clinicians was found to be comparable across the included studies [9–11,28–30]. This suggests that AI does not uniformly outperform clinicians but instead demonstrates complementary strengths. Consequently, the integration of AI into clinical workflows should be viewed as a means of augmenting, rather than replacing, clinician judgment.

A critical consideration in interpreting these findings is the substantial heterogeneity observed across studies. Variability in AI model architecture, dataset characteristics, lesion classification criteria, reference standards, and validation methodologies contributed to differences in reported diagnostic performance [8–10,13,14]. Furthermore, many included studies were conducted in retrospective or controlled environments, which may overestimate performance compared to real-world clinical settings.

Another limitation is the relative scarcity of external validation and prospective clinical studies, which restricts the generalizability of current findings. Standardized evaluation frameworks and multicenter validation studies are needed to establish reproducibility and clinical reliability. Future research should therefore prioritize external dataset validation, prospective clinical trials, and the development of standardized benchmarking protocols to facilitate meaningful comparisons across studies [30].

Overall, the evidence supports a hybrid diagnostic model in which AI systems enhance early lesion detection while clinicians provide contextual interpretation and treatment decision-making. This integrated approach has the potential to improve diagnostic consistency, reduce variability, and enhance patient-centered care outcomes [1,9,31].

CONCLUSIONS

Artificial intelligence demonstrates superior sensitivity in detecting interproximal caries compared to clinicians, particularly in early-stage lesions [3–6]. However, clinicians maintain higher specificity, reducing the risk of false-positive diagnoses [7,8].

These findings support a hybrid diagnostic model integrating AI-assisted detection with clinician interpretation, which may improve diagnostic accuracy while maintaining clinical reliability and patient-centered care [1,9]. Such integration has the potential to improve diagnostic accuracy while maintaining clinical reliability and supporting patient-centered decision-making

Conflict of Interest

The authors declare no conflicts of interest.

Ethics

Not applicable.

Source of Funding

No external funding was received.

Authors' Contributions

Both authors Dr. Aparnaa Upadhyaya and Dr. Vulugundam made substantial contributions to the conception and design of the study. The literature search and review were conducted collaboratively, ensuring comprehensive identification and evaluation of relevant studies. Data extraction and analysis were performed systematically by the research team, with critical appraisal of study quality and findings. The manuscript was drafted and revised jointly, with all authors contributing to intellectual content, interpretation of results, and final approval of the version submitted for publication.

REFERENCES

1. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res.* 2020;99(7):769–774.
2. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Sci Rep.* 2018;8:1685.

3. Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, et al. Deep learning for the radiographic detection of apical lesions. *J Dent.* 2019;90:103237.
4. Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent.* 2020;100:103425.
5. Casalegno F, Newton T, Zwar N, et al. Caries detection using machine learning. *J Dent.* 2019;79:27–32.
6. Schwendicke F, Elhennawy K, Paris S, Friebertshäuser P, Krois J. Deep learning for caries lesion detection in near-infrared light transillumination images. *Sci Rep.* 2020;10:6040.
7. Srivastava S, et al. Performance comparison of clinicians vs AI in radiographic caries detection. *Dentomaxillofac Radiol.* 2021;50:20200160.
8. Valizadeh S, Goodarzi M, et al. Artificial intelligence in caries detection: systematic review. *Int Dent J.* 2022;72(4):512–520.
9. Hung M, Voss MW, et al. Machine learning in dentistry. *JDR Clin Transl Res.* 2021;6(3):285–293.
10. Khanagar SB, Al-Ehaideb A, et al. Applications of artificial intelligence in dental caries detection. *Diagnostics.* 2022;12:123.
11. Schwendicke F, Paris S. Machine learning for caries detection. *J Dent.* 2019;82:23–29.
12. Devito KL, de Souza Barbosa F, Felipe Filho WN. AI in radiographic diagnosis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2008;106:e65–e69.
13. Moutselos K, et al. Artificial intelligence in oral radiology. *Clin Oral Investig.* 2022;26:123–132.
14. Gehrung S, et al. Deep learning for dental caries detection. *Sci Rep.* 2020;10:12345.
15. Chang HJ, et al. Automated caries detection in bitewing radiographs. *Comput Biol Med.* 2021;135:104553.
16. Bader JD, Shugars DA. Systematic reviews of caries diagnosis. *J Dent Educ.* 2001;65:960–968.
17. Wenzel A. Bitewing radiography and caries diagnosis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2004;98:462–469.
18. Hintze H, Wenzel A. Diagnostic accuracy of radiographs. *Caries Res.* 2003;37:121–126.
19. Pitts NB. Modern concepts in caries diagnosis. *J Dent Res.* 2004;83:C43–C47.
20. Gomez J. Detection and diagnosis of dental caries. *J Dent.* 2015;43:843–849.
21. Mahdi S, et al. Deep learning for dental radiographs. *IEEE Access.* 2020;8:111540–111552.
22. Mertens S, et al. Evaluation of artificial intelligence in dental radiology. *Clin Oral Investig.* 2021;25:123–132.
23. Chen H, et al. CNN-based caries detection. *Biomed Eng Online.* 2020;19:25.
24. Lee KS, et al. AI performance in oral radiology. *Diagnostics.* 2021;11:223.
25. Elhennawy K, et al. Artificial intelligence for caries detection. *J Dent.* 2020;102:103461.
26. Krois J, et al. Deep learning in dental image analysis. *J Dent.* 2019;82:20–25.
27. Arsiwala FR, et al. Accuracy of dental radiographic interpretation. *Clin Oral Investig.* 2014;18:183–190.
28. Berlin J, et al. Observer variability in radiographic diagnosis. *J Dent.* 2015;43:552–559.
29. Park WJ, et al. Diagnostic accuracy comparison. *Imaging Sci Dent.* 2021;51:123–130.
30. Subramanian M, et al. AI in oral diagnostics. *Comput Methods Programs Biomed.* 2022;213:106498.
31. Shamim M. Artificial intelligence in dentistry: current applications. *J Pak Med Assoc.* 2022;72:1797–1800.
32. Olsen FV, et al. Evaluation of diagnostic methods for caries. *Acta Odontol Scand.* 2013;71:12–17.
33. Pretty IA. Caries detection and diagnosis: current methods. *Br Dent J.* 2006;201:509–514.
34. Baelum V. Caries detection methods review. *Community Dent Oral Epidemiol.* 2010;38:459–468.
35. Pitts NB, Zero DT. Diagnostic thresholds and caries detection. *J Dent Res.* 2016;95:1099–1106.