

Diagnostic Performance of Artificial Intelligence for Proximal Caries Detection on Bitewing Radiographs: A Systematic Review and Diagnostic Test Accuracy Meta-Analysis

Amun Jukisalapong¹



Amun Jukisalapong

¹ Dental Department, Bangkok Hospital Chiang Mai, Chiang Mai, Thailand.

* Address Correspondence to author:
Amun Jukisalapong
Bangkok Hospital Chiang Mai
88/8-9 Moo 6 Nong Pa Khrang,
Mueang Chiang Mai,
Chiang Mai 50000, Thailand
email: chopinconcerto@gmail.com

Received: March 13, 2026
Revision received: July 6, 2026
Accepted after revision: August 10, 2026
BKK Med J 2026;22(2): 117-125.
DOI: 10.31524/bkkmedj.2026.21.006
www.bangkokmedjournal.com

Abstract

OBJECTIVES: To systematically evaluate the diagnostic performance of artificial intelligence (AI) systems for detecting proximal caries lesions on bitewing radiographs.

MATERIALS AND METHODS: Electronic searches were conducted in PubMed/MEDLINE, Scopus, and Web of Science for studies published between January 2000 and March 2026. Studies evaluating AI-based systems for proximal caries detection on bitewing radiographs were included. True-positive, false-positive, true-negative, and false-negative values were extracted directly or reconstructed from reported diagnostic accuracy metrics when sufficient information was available. Pooled sensitivity and specificity were estimated separately using random-effects models on the logit scale.

RESULTS: Eleven studies were included in the qualitative synthesis, of which three studies provided sufficient 2×2 contingency table data for quantitative synthesis. The pooled sensitivity and specificity of AI systems for proximal caries detection were 0.73 (95% CI 0.70–0.75) and 0.92 (95% CI 0.69–0.99), respectively. Most included studies demonstrated promising diagnostic performance using deep learning architectures, particularly convolutional neural networks and YOLO-based models.

CONCLUSION: AI systems demonstrated promising diagnostic performance for detecting proximal caries lesions on bitewing radiographs. However, methodological heterogeneity and limited external validation remain important limitations. Further prospective multicenter studies are needed to confirm the clinical applicability of AI-assisted caries detection systems.

Keywords: artificial intelligence; deep learning; dental caries; bitewing radiography; diagnostic accuracy

Dental caries remains one of the most prevalent chronic diseases worldwide and continues to represent a major public health concern affecting individuals across all age groups.¹ Despite substantial improvements in preventive strategies, oral health education, and minimally invasive approaches, untreated carious lesions remain highly prevalent and may lead to pain, infection, tooth loss, and reduced quality of life. The development of dental caries is a multifactorial process involving complex interactions among oral microbiota, host susceptibility, dietary habits, and time.² Early detection of carious lesions is therefore essential to prevent disease progression and to support preventive and conservative treatment approaches.

Proximal caries lesions are particularly difficult to detect during routine clinical examination because the affected surfaces are not directly visible.³ Consequently, bitewing radiography is widely used as an adjunctive diagnostic method for identifying proximal lesions that may not be detectable through visual inspection alone.^{4,5} However, interpretation of bitewing radiographs is inherently subjective and may vary according to clinicians' experience, diagnostic thresholds, and image quality. Such variability may influence diagnostic consistency and treatment decision-making in daily clinical practice.

Recent advances in AI, particularly deep learning and convolutional neural network (CNN) algorithms, have enabled automated image analysis

and pattern recognition in medical and dental imaging.⁶⁻⁸ In dentistry, early investigations demonstrated that neural network models could assist in detecting proximal caries lesions on radiographic images.⁹ More recently, several studies have applied deep learning models to bitewing radiographs and reported promising diagnostic performance for proximal caries detection.¹⁰⁻²⁰ These AI systems have shown potential for improving diagnostic consistency, reducing observer variability, and supporting clinicians in radiographic interpretation.

Nevertheless, substantial heterogeneity remains among published studies. Variations in dataset characteristics, imaging protocols, annotation methods, reference standards, validation strategies, and AI architectures may significantly influence reported diagnostic performance and limit the reproducibility and generalizability of findings. Furthermore, many AI models have been developed using retrospective datasets with limited external validation, raising concerns regarding their applicability in real-world clinical settings. In addition, differences in lesion classification criteria and reporting of diagnostic outcomes complicate direct comparisons across studies.

Recent systematic reviews and meta-analyses have reported encouraging results regarding the diagnostic accuracy of AI systems for detecting proximal caries lesions on bitewing radiographs.²¹⁻²² Previous systematic review evidence has also evaluated the diagnostic accuracy of conventional methods for proximal caries detection, including visual examination and bitewing radiography.²³ However, rapid developments in deep learning methodologies and the increasing number of newly published validation studies necessitate an updated and comprehensive synthesis of the available evidence. Therefore, the aim of the present systematic review and meta-analysis was to evaluate the diagnostic accuracy of AI systems for detecting proximal caries lesions on bitewing radiographs and to summarize the current evidence regarding their potential clinical applicability.

Materials and Methods

This study was conducted as a systematic review and diagnostic test accuracy meta-analysis.²⁴ The review methodology followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement²⁵ and methodological recommendations from the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.²⁶ Reporting considerations specific to diagnostic accuracy studies were additionally guided by the PRISMA-DTA framework.²⁷

A predefined review protocol was developed prior to study selection and data extraction to improve methodological transparency and minimize potential bias. However, the protocol was not prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO).

Literature search strategy

Electronic database searches were conducted in PubMed/MEDLINE, Scopus, and Web of Science for studies published between January 2000 and March 2026. The search strategy

combined controlled vocabulary and free-text terms related to AI, proximal caries detection, and bitewing radiography.

The search strategy combined terms related to artificial intelligence, proximal caries, and bitewing radiography using the following Boolean structure:

("artificial intelligence" OR "deep learning" OR
"machine learning" OR "neural network")
AND
("dental caries" OR "proximal caries" OR
"interproximal caries")
AND
("bitewing radiograph" OR "bitewing radiography")

Reference lists of eligible studies and relevant review articles were also manually screened to identify additional potentially relevant studies.

Eligibility criteria

Original primary studies were eligible if they evaluated an AI-based system for detecting proximal caries in human bitewing radiographs and reported diagnostic performance outcomes sufficient for qualitative evaluation. Reviews, editorials, animal studies, studies using non-bitewing images, studies without a proximal-caries outcome, and studies without an AI or deep-learning model were excluded. Quantitative synthesis was restricted to studies with directly reported or reconstructable 2×2 contingency table data.

Study selection

Retrieved records were screened based on titles and abstracts according to predefined eligibility criteria. Potentially eligible studies subsequently underwent full-text assessment. The study selection process followed the PRISMA 2020 framework and is illustrated in Figure 1.

Data extraction

Data extraction was performed using a standardized extraction form. The following information was collected from each included study:

- First author and publication year
- Country
- Study design
- Dataset size
- AI architecture/model
- Validation method
- Reference standard
- Diagnostic performance outcomes

For quantitative synthesis, true-positive, false-positive, true-negative, and false-negative values were extracted directly from included studies or reconstructed from reported diagnostic accuracy metrics when sufficient information was available. Only studies providing sufficient 2×2 contingency table data were included in the diagnostic test accuracy meta-analysis.

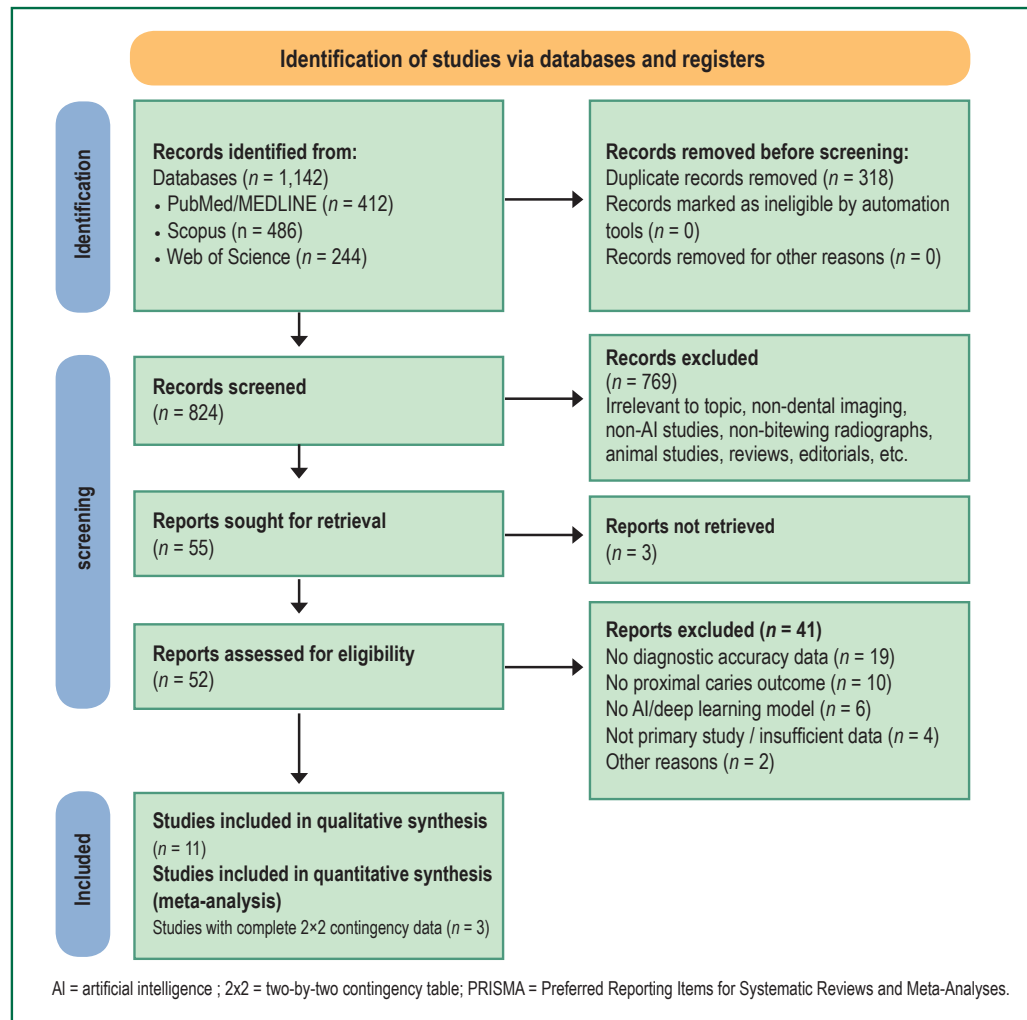


Figure 1: PRISMA flow diagram illustrating the study selection process.

Quality assessment

Methodological quality assessment was performed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.²⁸ The following domains were evaluated:

1. Patient selection
2. Index test
3. Reference standard
4. Flow and timing

Each domain was assessed for risk of bias and applicability concerns.

Statistical analysis

Pooled sensitivity and specificity estimates were calculated based on extracted 2×2 contingency table data. Pooled estimates and 95% confidence intervals were back-transformed to the probability scale. Although bivariate random-effects models are generally recommended for jointly synthesizing sensitivity and specificity²⁹, such a model was not used because only three studies provided complete 2×2 data, making reliable estimation of the between-study covariance difficult. Forest plots were generated to display study-level and pooled estimates. Statistical

analyses and graphical visualizations were performed using R statistical software with diagnostic meta-analysis packages. When contingency counts were reconstructed from reported sensitivity, specificity, prevalence, and sample size, values were rounded to the nearest whole number while preserving internal consistency.

Results

The literature search identified 1,142 records across PubMed/MEDLINE, Scopus, and Web of Science databases. After removal of 318 duplicate records, 824 records remained for title and abstract screening. Following initial screening, 769 records were excluded because they did not meet the eligibility criteria. Fifty-five reports were sought for retrieval, of which three could not be obtained. Consequently, 52 full-text articles were assessed for eligibility. After full-text evaluation, 41 studies were excluded for the following reasons: • No diagnostic accuracy data (n = 19) • No proximal caries outcome (n = 10) • No AI/deep learning model (n = 6) • Not a primary study or insufficient data (n = 4) • Other reasons (n = 2). Ultimately, 11 studies were included in the qualitative synthesis, and three studies with complete 2×2 contingency data were included in the quantitative synthesis.

The included studies were published between 2020 and 2025 and evaluated various AI systems for proximal caries detection on bitewing radiographs (Table 1).

Across the included studies, various deep learning architectures were evaluated, including convolutional neural networks (CNNs), U-Net models, Faster R-CNN frameworks, and YOLO-based detection systems. Dataset sizes varied substantially among studies, ranging from fewer than 500 radiographs to large multicenter datasets containing more than 13,000 bitewing images.

Most studies used retrospective datasets with internal validation approaches. Reference standards primarily involved expert consensus annotations by experienced dentists, oral radiologists, or calibrated evaluators. In some studies, diagnostic findings were additionally compared with clinical

examination results or established caries classification systems (e.g., ICCMS). None of the included studies reported external validation using a fully independent institutional dataset.

Risk of bias assessment

The methodological quality of the included studies was assessed using the QUADAS-2 tool. Overall, most studies demonstrated low risk of bias in the index test and reference standard domains. However, several studies showed unclear or high risk of bias in patient selection due to retrospective dataset selection and limited external validation. The summary of the risk-of-bias assessment is presented in Figure 2.

The diagnostic performance outcomes of the included AI systems are summarized in Table 2.

Table 1: Characteristics of included studies evaluating artificial intelligence systems for proximal caries detection on bitewing radiographs

Study	Year	Country	Study design	Dataset size and unit	AI model	Validation method	Reference standard	External validation	Main outcomes
Cantu et al. ¹⁰	2020	Germany	Retrospective diagnostic study	3,686 radiographs	CNN-based segmentation model	Training, validation, and test split	Pixel-wise annotations by four experienced dentists	No	Lesion-detection performance
Mertens et al. ¹¹	2021	Germany	Randomized diagnostic reader study	140 radiographs; 22 dentists	Fully convolutional neural network	Randomized crossover reader comparison	Consensus of four independent experts with adjudication by a fifth expert	No	Diagnostic performance with versus without AI assistance
Lee et al. ¹²	2021	South Korea	Retrospective diagnostic study	354 radiographs	U-Net-based CNN	Training and evaluation split	Consensus annotations by trained observers	No	Precision, sensitivity, and F1-score
Bayraktar and Ayan ¹³	2022	Turkey	Retrospective diagnostic study	1,000 radiographs: 800 training and 200 test radiographs	Modified YOLO-based CNN	Hold-out test set	Annotations by two experienced dentists	No	Accuracy, sensitivity, specificity, PPV, NPV, and AUC
Chen et al. ¹⁴	2022	China	Retrospective diagnostic study	978 radiographs; 10,899 proximal surfaces	Faster R-CNN	Hold-out test set	Consensus of two endodontists and one radiologist	No	Accuracy, sensitivity, specificity, PPV, NPV, F1-score, and AUC
García-Cañas et al. ¹⁵	2022	Spain	Diagnostic accuracy study	300 radiographs	CNN-based diagnostic software	Independent diagnostic evaluation	Clinical and radiographic reference assessment	No	Sensitivity, specificity, PPV, NPV, and accuracy
Kunt et al. ¹⁶	2023	Czech Republic	Retrospective model-development study	3,989 radiographs; 7,257 carious lesions	YOLOv5, Faster R-CNN, RetinaNet, and EfficientDet	Training, validation, and test split	Expert bounding-box annotations	No	Precision, recall, F1-score, and average precision
Pérez de Frutos et al. ¹⁷	2024	Norway	Retrospective diagnostic study	13,887 annotated radiographs; 197-image consensus test set	YOLOv5, RetinaNet, and EfficientDet	Five-fold cross-validation and consensus test set	Consensus of six dental clinicians	No	Object-detection performance
Tichý et al. ¹⁸	2024	Czech Republic	Experimental reader-comparison study	Test subset from the Kunt et al. dataset	Deep-learning detection models	AI-dentist comparison	Consensus caries annotations	No	Detection performance and interobserver agreement
Ayhan et al. ¹⁹	2025	Turkey	Comparative diagnostic study	2,150 radiographs; test set of 150 radiographs and 2,160 proximal surfaces	YOLO-based detection models	Hold-out test set and dentist comparison	Expert annotations	No	Precision, recall, specificity, F1-score, and Youden index
Alsolamy et al. ²⁰	2025	Saudi Arabia	Retrospective model-development study	1,354 radiographs	YOLOv11-based object-detection and instance-segmentation models	Training, validation, and test split	ICCMS-based expert consensus annotations ³⁰	No	Precision, recall, F1-score, mAP, lesion localization, and severity classification

Note: Dataset size and outcome measures are reported according to the unit of analysis used in each original study. Kunt et al.¹⁶ and Tichý et al.¹⁸ used related datasets and should not be interpreted as independent study populations.

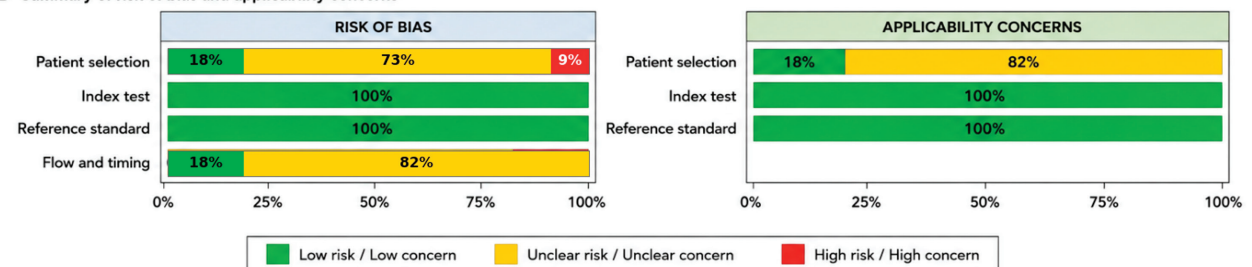
Abbreviations: AI, artificial intelligence; AUC, area under the receiver operating characteristic curve; CNN, convolutional neural network; ICCMS, International Caries Classification and Management System; mAP, mean average precision; NPV, negative predictive value; PPV, positive predictive value; R-CNN, region-based convolutional neural network; YOLO, You Only Look Once.

Diagnostic Performance of Artificial Intelligence for Proximal Caries Detection on Bitewing Radiographs: A Systematic Review and Diagnostic Test Accuracy Meta-Analysis

A Risk of bias and applicability concerns for each study

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
1. Cantu et al. (2020)	●	●	●	●	●	●	●
2. Mertens et al. (2021)	●	●	●	●	●	●	●
3. Lee et al. (2021)	●	●	●	●	●	●	●
4. Bayraktar & Ayan (2022)	●	●	●	●	●	●	●
5. Chen et al. (2022)	●	●	●	●	●	●	●
6. García-Cañas et al. (2022)	●	●	●	●	●	●	●
7. Kunt et al. (2023)	●	●	●	●	●	●	●
8. Pérez de Frutos et al. (2024)	●	●	●	●	●	●	●
9. Tichý et al. (2024)	●	●	●	●	●	●	●
10. Ayhan et al. (2025)	●	●	●	●	●	●	●
11. Alsolamy et al. (2025)	●	●	●	●	●	●	●

B Summary of risk of bias and applicability concerns



QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies-2.

Green indicates low risk (or low concern), yellow indicates unclear risk (or unclear concern), and red indicates high risk (or high concern).

Figure 2: QUADAS-2 assessment summarizing the risk of bias and applicability concerns of the included studies.

Table 2: Diagnostic performance data of studies included in the quantitative synthesis.

Study	True positive	False positive	True negative	False negative	Sensitivity	Specificity
Bayraktar and Ayan (2022) ¹³	201	31	1,682	77	0.72	0.98
Chen et al. (2022) ¹⁴	388	116	1,435	148	0.72	0.93
Ayhan et al. (2025) ¹⁹	648	347	922	243	0.73	0.73

Note: The unit of analysis was the proximal surface. The 2×2 contingency counts were reported directly or reconstructed from the original studies.

Quantitative synthesis

Three studies provided complete 2×2 contingency table data and were included in the diagnostic test accuracy meta-analysis. The contingency data used for quantitative synthesis are presented in Table 2.

Across the three studies included in the quantitative synthesis, the pooled sensitivity was 0.73 (95% CI 0.70–0.75), whereas the pooled specificity was 0.92 (95% CI 0.69–0.99). Study-level sensitivity estimates were consistent, ranging from 0.72 to 0.73. In contrast, specificity estimates ranged from 0.73 to 0.98, resulting in substantial between-study heterogeneity and a wide confidence interval for the pooled specificity estimate.

Forest plots of study-level and pooled sensitivity and specificity estimates are presented in Figures 3 and 4, respectively.

The pooled sensitivity and specificity estimates derived from the separate random-effects models are presented in Table 3.

Pooled sensitivity and specificity were estimated separately using random-effects models on the logit scale.

Because only three studies provided complete 2×2 contingency table data for quantitative synthesis, subgroup analyses and publication-bias assessments were not considered sufficiently robust for reliable interpretation. Diagnostic performance was therefore interpreted primarily from the pooled sensitivity and specificity estimates and their confidence intervals.

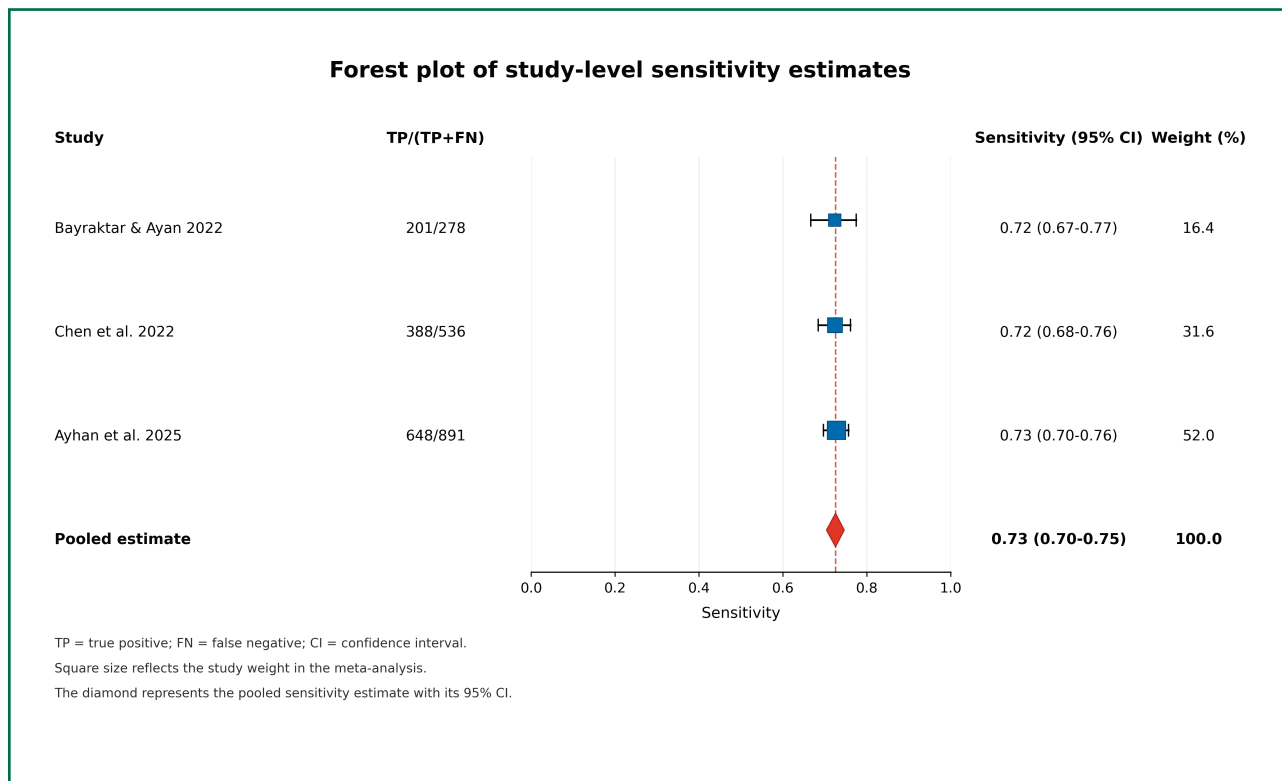


Figure 3: Forest plot of study-level sensitivity estimates for AI systems detecting proximal caries lesions on bitewing radiographs.

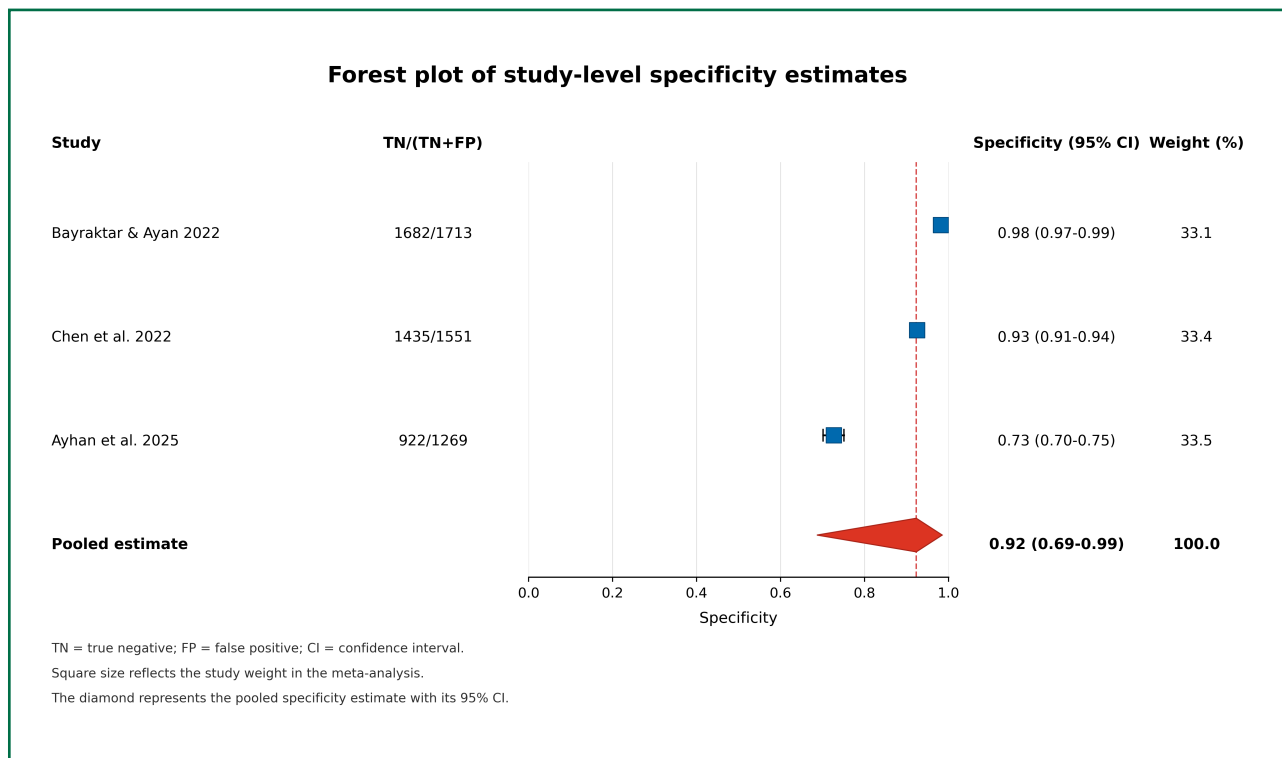


Figure 4: Forest plot of study-level specificity estimates for AI systems detecting proximal caries lesions on bitewing radiographs.

Table 3: Pooled diagnostic accuracy estimates of AI systems for detecting proximal caries lesions on bitewing radiographs

Outcome	Pooled estimate
Sensitivity	0.73 (0.70 – 0.75)
Specificity	0.92 (0.69 – 0.99)

Discussion

Our findings are broadly consistent with previous meta-analyses by Ammar et al.²¹ and Carvalho et al.²², which also reported favorable diagnostic performance of AI-assisted proximal caries detection systems on bitewing radiographs.

The present systematic review and diagnostic test accuracy meta-analysis evaluated the performance of AI systems for detecting proximal caries lesions on bitewing radiographs. The findings demonstrated promising diagnostic performance, with pooled sensitivity and specificity estimates of 0.73 and 0.92, respectively. These results suggest that AI algorithms may serve as useful adjunctive tools for supporting radiographic caries detection; however, the estimates should be interpreted cautiously because only three studies contributed to the quantitative synthesis.

Dental caries remains one of the most prevalent chronic oral diseases worldwide and continues to represent a significant public health burden. Early detection of proximal caries lesions is particularly important because non-cavitated lesions may be managed using preventive and minimally invasive approaches before extensive structural damage occurs. Bitewing radiography remains the standard imaging modality for detecting proximal lesions that are difficult to identify through visual examination alone. However, radiographic interpretation may vary considerably among clinicians depending on clinical experience, diagnostic thresholds, and image quality. Such variability has contributed to growing interest in computer-assisted diagnostic systems in dentistry.^{6,8}

Recent advances in AI, particularly deep learning techniques, have enabled automated image analysis with increasingly high diagnostic capability. Most studies included in the present review applied convolutional neural network (CNN)-based architectures, while more recent investigations evaluated object-detection frameworks such as YOLO-based systems. Overall, the included studies demonstrated favorable diagnostic performance across different AI architectures, suggesting that deep learning approaches may provide reliable support for proximal caries detection on bitewing radiographs. The pooled sensitivity estimate was moderate and consistent across the three studies. In contrast, specificity varied substantially across studies, ranging from 0.73 to 0.98, resulting in a wide confidence interval around the pooled specificity estimate. This heterogeneity may reflect differences in dataset composition, lesion prevalence, image acquisition protocols, annotation procedures, and AI training strategies. Although high specificity is clinically desirable because false-positive diagnoses may lead to unnecessary restorative intervention, the imprecision of the pooled specificity estimate warrants cautious interpretation.

Several methodological limitations should be considered when interpreting the findings of this review. First, most included studies used retrospective datasets and internal validation approaches, which may increase the risk of selection bias and limit external generalizability. Second, substantial heterogeneity was observed across studies regarding dataset size, AI architecture, validation methodology, and reference standards. Third, only a limited number of studies performed external validation using independent multicenter datasets. These limitations highlight the need for more standardized methodological frameworks in future AI diagnostic research.

Another important limitation relates to the availability of diagnostic contingency data. Several studies did not directly report true-positive, false-positive, true-negative, and false-negative values. Consequently, some contingency tables included in the quantitative synthesis were reconstructed from reported diagnostic accuracy metrics. Although this approach has been used in diagnostic meta-analysis research, reconstructed estimates may introduce additional uncertainty into pooled analyses.

Despite these limitations, the present study provides an updated synthesis of current evidence regarding AI-assisted proximal caries detection on bitewing radiographs. The findings support the growing potential of AI systems to enhance radiographic interpretation and improve diagnostic consistency in dental practice. Nevertheless, AI systems should be considered adjunctive decision-support tools rather than replacements for professional clinical judgment.

Future research should focus on prospective multicenter validation studies, standardized reporting of diagnostic contingency data, and external validation across diverse patient populations. Improved methodological transparency and reproducible reporting will be essential for determining the true clinical applicability of AI-assisted diagnostic systems in dentistry.

Overall, the findings of this systematic review and diagnostic test accuracy meta-analysis suggest that AI systems demonstrate promising diagnostic performance for detecting proximal caries lesions on bitewing radiographs. Continued methodological refinement and prospective clinical validation will be important for facilitating the integration of AI-assisted diagnostic tools into routine dental practice.

Strengths and Limitations

This study has several strengths. **First**, a comprehensive systematic literature search was conducted across multiple electronic databases following the PRISMA 2020 statement and PRISMA-DTA reporting considerations. **Second**, methodological quality was assessed using the QUADAS-2 tool, allowing structured evaluation of risk of bias and applicability concerns. **Third**, quantitative synthesis used separate random-effects meta-analyses of logit-transformed sensitivity and specificity based on 2×2 contingency data.

However, several limitations should be noted:

1. Limited Number of Quantitative Studies: Only three studies provided complete 2×2 contingency table data for meta-analysis, limiting the precision of the pooled estimates and precluding robust subgroup and publication bias analyses.
2. High Methodological Heterogeneity: Substantial variability was observed across studies regarding dataset characteristics, imaging protocols, AI architectures, and validation methodologies.
3. Retrospective Design: Most included studies relied on retrospective datasets and internal validation approaches, which may limit generalizability to prospective clinical workflows.
4. Data Reconstruction: In some cases, 2×2 contingency values were reconstructed from reported diagnostic performance metrics, which may have introduced rounding error and additional uncertainty.

Conclusion

AI systems demonstrated promising diagnostic performance for detecting proximal caries lesions on bitewing radiographs, with relatively high pooled specificity and moderate pooled sensitivity observed across the included studies. These findings suggest that AI-assisted systems may serve as supportive diagnostic tools for enhancing radiographic interpretation and improving diagnostic consistency in dental practice. However, methodological heterogeneity and limited external validation remain important limitations. Further prospective multicenter studies with standardized reporting and external validation are required before widespread clinical implementation of AI-based diagnostic systems can be recommended.

References

1. Petersen PE. The World Oral Health Report 2003: continuous improvement of oral health in the 21st century—the approach of the WHO Global Oral Health Programme. *Community Dent Oral Epidemiol.* 2003;31 (Suppl) 1:3-24. doi: 10.1046/j.2003.com122.x.
2. Kidd EAM, Fejerskov O. What constitutes dental caries? Histopathology of carious enamel and dentin related to the action of cariogenic biofilms. *J Dent Res.* 2004;83 Spec No C:C35-C38. doi: 10.1177/154405910408301s07.
3. Pitts NB. Clinical diagnosis of dental caries: a European perspective. *J Dent Educ.* 2001;65(10):972-8 doi 10.1002/j.0022-0337.2001.65.10.tb03472.x
4. Wenzel A. Bitewing and digital bitewing radiography for detection of caries lesions. *J Dent Res.* 2004;83 Spec No C:C72-C75. doi:10.1177/154405910408301s14
5. Schwendicke F, Tzschoppe M, Paris S. Radiographic caries detection: a systematic review and meta-analysis. *J Dent.* 2015;43(8):924-33. doi: 10.1016/j.jdent.2015.02.009.
6. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res.* 2020;99(7): 769-74. doi:10.1177/0022034520915714.
7. Hung M, Voss MW, Rosales MN, et al. Application of machine learning for diagnostic prediction of root caries. *Gerodontology.* 2019;36(4):395-404. doi: 10.1111/ger.12432.
8. Khanagar SB, Al-Ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry: a systematic review. *J Dent Sci.* 2021;16(1):508-22. doi: 10.1016/j.jds.2020.06.019.
9. Devito KL, de Souza Barbosa F, Felipe Filho WN. An artificial multilayer perceptron neural network for diagnosis of proximal dental caries. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2008;106(6):879-84. doi:10.1016/j.tripleo.2008.03.002.
10. Cantu AG, Gehrung S, Krois J, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent.* 2020;100:103425. doi:10.1016/j.jdent.2020.103425.
11. Mertens S, Krois J, Cantu AG, et al. Artificial intelligence for caries detection: randomized trial. *J Dent.* 2021;115:103849. doi:10.1016/j.jdent.2021.103849.
12. Lee S, Oh SI, Jo J, et al. Deep learning for early dental caries detection in bitewing radiographs. *Sci Rep.* 2021;11(1):16807. doi:10.1038/s41598-021-96368-7.

Acknowledgements

The author would like to thank Pimolnut Srisuwan for her support throughout this research.

Declarations

Author Contributions: Conceptualization, methodology, data collection, data analysis, and manuscript preparation were performed by Arnun Jukisalapong. The author has read and approved the final version of the manuscript.

Declaration of AI Use: AI was used solely for English-language editing and grammatical refinement. The author reviewed all AI-assisted revisions and takes full responsibility for the accuracy, integrity, and final content of the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflict of interest.

Ethical Considerations

Not applicable. This systematic review and meta-analysis used data from previously published studies and did not involve the collection of individual-level participant data.

**Diagnostic Performance of Artificial Intelligence for Proximal Caries Detection on Bitewing Radiographs:
A Systematic Review and Diagnostic Test Accuracy Meta-Analysis**

13. Bayraktar Y, Ayan E. Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. *Clin Oral Investig.* 2022;26(1):623-32. doi:10.1007/s00784-021-04040-1.
14. Chen X, Guo J, Ye J, et al. Detection of proximal caries lesions on bitewing radiographs using deep learning method. *Caries Res.* 2022;56(5-6):455-63. doi:10.1159/000527418.
15. García-Cañas Á, Bonfanti-Gris M, Paraíso-Medina S, et al. Diagnosis of interproximal caries lesions in bitewing radiographs using a deep convolutional neural network-based software. *Caries Res.* 2022;56(5-6):503-11. doi:10.1159/000527491.
16. Kunt L, Kybic J, Nagyová V, et al. Automatic caries detection in bitewing radiographs: Part I-deep learning. *Clin Oral Investig.* 2023;27(12):7463-71. doi:10.1007/s00784-023-05335-1.
17. Pérez de Frutos J, Helland RH, Desai S, et al. AI-Dentify: deep learning for proximal caries detection on bitewing x-ray-HUNT4 Oral Health Study. *BMC Oral Health.* 2024;24(1):355. doi:10.1186/s12903-024-04120-0.
18. Tichý A, Kunt L, Nagyová V, et al. Automatic caries detection in bitewing radiographs-part II: experimental comparison. *Clin Oral Investig.* 2024;28(2):133. doi:10.1007/s00784-024-05528-2.
19. Ayhan B, Ayan E, Karadağ G, et al. Evaluation of caries detection on bitewing radiographs: a comparative analysis of the improved deep learning model and dentist performance. *J Esthet Restor Dent.* 2025;37(7):1949-61. doi:10.1111/jerd.13470.
20. Alsolamy M, Nadeem F, Azhari AA, et al. Automated detection, localization, and severity assessment of proximal dental caries from bitewing radiographs using deep learning. *Diagnostics (Basel).* 2025;15(7):899. doi:10.3390/diagnostics15070899.
21. Ammar N, Kühnisch J. Diagnostic performance of artificial intelligence-aided caries detection on bitewing radiographs: a systematic review and meta-analysis. *Jpn Dent Sci Rev.* 2024;60:128-36. doi:10.1016/j.jdsr.2024.02.001.
22. Carvalho BKG, Nolden EL, Wenning AS, et al. Diagnostic accuracy of artificial intelligence for approximal caries on bitewing radiographs: a systematic review and meta-analysis. *J Dent.* 2024;151:105388. doi:10.1016/j.jdent.2024.105388.
23. Janjic Rankovic M, Kapor S, Khazaei Y, et al. Systematic review and meta-analysis of diagnostic studies of proximal surface caries. *Clin Oral Investig.* 2021;25(11):6069-79. doi:10.1007/s00784-021-04113-1.
24. Leeftang MMG, Deeks JJ, Gatsonis C, et al. Systematic reviews of diagnostic test accuracy. *Ann Intern Med.* 2008;149(12):889-97. doi: 10.7326/0003-4819-149-12-200812160-00008.
25. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. doi: 10.1136/bmj.n71.
26. Deeks JJ, Bossuyt PM, Leeftang MM, et al., editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy* [Internet]. Version 2.0 (updated July 2023). Cochrane; 2023 [cited 2026 Jan 14]. Available from: <https://training.cochrane.org/handbook-diagnostic-test-accuracy/current>
27. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA Statement. *JAMA.* 2018;319(4):388-96. doi:10.1001/jama.2017.19163.
28. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009.
29. Reitsma JB, Glas AS, Rutjes AWS, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005;58(10):982-90. doi:10.1016/j.jclinepi.2005.02.022.
30. Pitts NB, Ekstrand KR. International caries detection and assessment system (ICDAS) and its international caries classification and management system (ICCMS) methods for staging of the caries process and enabling dentists to manage caries. *Community Dent Oral Epidemiol.* 2013;41(1):e41-e52. doi: 10.1111/cdoe.12025.