

Impact of Artificial Intelligence-Assisted Radiographic Caries Detection on Clinical Decision-Making and Patient Outcomes: A Systematic Review

Fitri Yustika Sya'bani¹, Titiek Berniyanti^{2*}, Dina Izzatul Mila Azka¹

¹Undergraduate Student, Faculty of Dental Medicine,
Universitas Airlangga, Surabaya, Indonesia

²Department of Dental Public Health, Faculty of Dental Medicine,
Universitas Airlangga, Surabaya, Indonesia

E-mail: fitri.yustika.syabani-2023@fkg.unair.ac.id; titiek-b@fkg.unair.ac.id;
dina.izzatul.mila-2023@fkg.unair.ac.id

*Corresponding author details: Titiek Berniyanti; titiek-b@fkg.unair.ac.id

ABSTRACT

Dental caries remains a major global health problem, with approximately 2.24 billion people worldwide affected by untreated caries in permanent teeth in 2021. This persistent burden highlights the importance of accurate detection and appropriate management of carious lesions. Artificial intelligence (AI) is increasingly investigated for caries detection on dental radiographs, yet improved diagnostic performance does not necessarily translate into appropriate treatment decisions or better patient outcomes. This systematic review evaluated evidence linking AI-assisted radiographic caries assessment with treatment decisions, treatment intensity, clinical consequences, and economic outcomes. The method used is a Systematic Literature Review (SLR) approach guided by PRISMA 2020, with searches carried out in PubMed, Scopus, Web of Science, and SpringerLink. Eligible primary studies applied AI to dental radiographs and linked caries assessment to treatment decisions, treatment recommendations, treatment intensity or appropriateness, clinical consequences, or economic outcomes. Eleven studies were included from 998 records, comprising one cluster-randomized crossover trial, four economic/modeling evaluations, two diagnostic or ex vivo decision studies, and four treatment-planning or decision-support studies. AI assistance improved sensitivity and AUC but also altered treatment decisions, including increased invasive treatment for some lesions. Economic models generally favored AI when early lesions received non-restorative management, whereas trial-derived decisions showed no economic advantage. Treatment-recommendation systems demonstrated technical feasibility but lacked evidence of improved clinical appropriateness or patient outcomes. Overall, AI may influence treatment pathways, but clinical utility evidence remains less mature than diagnostic evidence. Future studies should evaluate AI using prospective designs, independent validation, lesion-specific management pathways, and patient-important outcomes.

Keywords: artificial intelligence; dental caries; radiography; clinical decision-making; treatment planning

INTRODUCTION

Dental caries remains one of the most prevalent oral diseases worldwide and represents a substantial public health burden because untreated lesions may progress to pain, tooth destruction, and tooth loss (Sun et al., 2024). In 2021, untreated dental caries in permanent teeth affected approximately 2.24 billion people worldwide, corresponding to a global age-standardized prevalence of 27.5% (Bernabe et al., 2025). This persistent burden highlights the importance of accurate detection and appropriate management of carious lesions (Alves-Costa et al., 2025). Contemporary caries management increasingly emphasizes early detection and preservation of tooth structure, making appropriate

differentiation between lesions that can be managed preventively or minimally invasively and those requiring operative treatment clinically important (Schwendicke et al., 2021; Schwendicke et al., 2022). Radiographic examination is an important adjunct to clinical caries assessment, particularly for lesions that may not be adequately identified by visual examination alone. However, radiographic interpretation remains subject to diagnostic limitations and variation among observers (Mertens et al., 2021; Ganss et al., 2026). These limitations are clinically relevant because radiographic findings may influence subsequent assessment of lesion severity and treatment selection (Mertens et al., 2021; Schwendicke et al., 2022).

Therefore, improvements in radiographic caries detection are clinically meaningful only if they contribute to appropriate clinical decision-making and patient care (Mertens et al., 2021; Ganss et al., 2026).

Artificial intelligence (AI), particularly machine-learning and deep-learning approaches, has emerged as a potential decision-support technology for dental radiographic interpretation and caries detection (Negi et al., 2024). Previous systematic and umbrella reviews have demonstrated considerable potential of AI for dental caries detection, with evidence predominantly assessed using diagnostic-performance measures such as sensitivity, specificity, and area under the curve (AUC) (Negi et al., 2024). Although these measures characterize diagnostic performance, they do not establish whether AI improves the clinical decisions made after a lesion is detected (Mertens et al., 2021; Ganss et al., 2026). This distinction is important because diagnostic accuracy and clinical utility represent related but different dimensions of AI performance.

Evidence from clinical research demonstrates that AI-assisted detection can influence treatment decisions. In a cluster-randomized crossover trial involving 22 dentists, Mertens et al. found that AI support increased the mean area under the receiver operating characteristic curve from 0.85 to 0.89 and sensitivity from 0.72 to 0.81, while also increasing non-invasive and invasive treatment decisions (Mertens et al., 2021). Thus, improved detection can influence management, but increased treatment decisions should not automatically be interpreted as improved treatment appropriateness or patient benefit (Mertens et al., 2021). This illustrates an important distinction between improved diagnostic performance and the clinical consequences of using the resulting information.

Economic evaluations further demonstrate that the value of AI depends on downstream clinical management. Schwendicke et al. reported that AI-supported proximal caries detection could be cost-effective in a modelling framework, but its economic benefit depended substantially on subsequent management, particularly the non-restorative management of early lesions (Schwendicke et al., 2021). Similarly, a subsequent randomized trial and economic evaluation incorporated treatment decisions into the assessment of AI-supported caries detection, demonstrating that economic consequences depend on the clinical pathway following detection rather than diagnostic performance alone (Schwendicke et al., 2022). Therefore, the value of AI is conditional on how clinicians act on the information generated by the system (Schwendicke et al., 2021; Schwendicke et al., 2022).

AI applications have also progressed from lesion detection toward treatment-oriented decision support. Orhan et al. evaluated an AI system using panoramic radiographs and included the appropriateness of treatment recommendations

among the outcomes, representing a shift from image interpretation toward treatment-related decision support (Orhan et al., 2023). Oruç et al. subsequently developed a machine-learning system designed to differentiate caries requiring filling from caries requiring root canal treatment and to provide treatment recommendations (Oruç et al., 2026). These studies demonstrate the technical feasibility of treatment-oriented AI; however, they do not establish that AI-generated recommendations improve patient outcomes in routine clinical practice (Orhan et al., 2023; Oruç et al., 2026).

Furthermore, the clinical value of AI may depend on how its outputs are integrated with clinician judgment rather than on AI performance alone. Ganss et al. compared dentist-only, AI-only, and combined human–AI decision strategies and demonstrated that the clinical implications of AI varied according to how its output was incorporated into the decision process (Ganss et al., 2026). Their findings indicate that structured human–AI decision rules can alter the balance between false-positive and false-negative consequences, supporting evaluation of AI as part of a clinical decision system rather than solely as an autonomous diagnostic instrument (Ganss et al., 2026). This is particularly relevant to caries management because the consequences of detecting an early lesion depend on the treatment threshold and subsequent management strategy (Schwendicke et al., 2021; Schwendicke et al., 2022; Ganss et al., 2026).

In this review, clinical decisions refer to clinical choices made following the interpretation of radiographic caries findings, including whether a lesion warrants intervention, the management strategy selected, and the invasiveness of treatment. Management may range from preventive and non-invasive approaches to micro-invasive procedures or conventional restorative treatment, depending on lesion characteristics and clinical judgment (Mertens et al., 2021; Schwendicke et al., 2022). These decisions are distinct from diagnostic classifications because identifying a lesion or determining its radiographic severity does not, by itself, establish the most appropriate management. Accordingly, this review examines whether AI-assisted assessment influences treatment selection and intensity while distinguishing changes in decision-making from evidence of treatment appropriateness.

Clinical outcomes refer to the implications that follow when AI-assisted assessment influences clinical decisions and subsequent management. In this review, these consequences encompass changes in treatment patterns and intensity, the potential implications of appropriate or inappropriate intervention, and reported clinical, patient-related, or economic outcomes where available (Mertens et al., 2021; Schwendicke et al., 2022; Ganss et al., 2026). These dimensions represent different levels of evidence and should not be treated as interchangeable. A change in treatment decisions does not establish improved patient outcomes, and increased treatment intensity alone does not demonstrate overtreatment.

Similarly, improved diagnostic accuracy does not establish clinical benefit unless the resulting information supports appropriate management and leads to demonstrated downstream benefits or harms (Mertens et al., 2021; Ganss et al., 2026).

Taken together, the evidence suggests that AI-assisted radiographic caries detection should be evaluated across a pathway linking diagnostic performance to treatment decisions, treatment appropriateness, and subsequent clinical, patient-related, and economic outcomes (Mertens et al., 2021; Schwendicke et al., 2021; Schwendicke et al., 2022; Ganss et al., 2026). However, evidence across this pathway remains heterogeneous, encompassing diagnostic performance, treatment decisions, treatment recommendations, economic modelling, and human–AI decision strategies. Existing systematic and umbrella reviews have predominantly focused on diagnostic performance, leaving the downstream relationship between AI-assisted detection, clinical management, and subsequent outcomes insufficiently synthesized (Negi et al., 2024). Consequently, it remains uncertain when AI-assisted detection supports more appropriate treatment and whether changes in clinical decisions translate into demonstrated patient or economic benefits.

Therefore, this systematic review aimed to evaluate evidence on AI-assisted radiographic caries detection beyond diagnostic performance, focusing on treatment decisions and intensity, treatment recommendations and appropriateness, and reported clinical, patient-related, and economic outcomes. By examining these interconnected but distinct dimensions, the review seeks to clarify the potential accuracy–decision gap between improved AI-assisted detection and the clinical consequences that follow.

MATERIAL AND METHODS

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement. The review question and eligibility criteria were defined before study selection. Because the included evidence was heterogeneous in design, population, imaging modality, and outcome definition, a narrative synthesis was used rather than meta-analysis.

Eligibility criteria

Eligibility criteria were predefined using the Population, Intervention, Comparison, Outcome, and Study design (PICOS) framework. The central eligibility principle was that diagnostic accuracy alone was insufficient; an eligible study had to connect AI-assisted radiographic caries assessment with a treatment decision, treatment recommendation, treatment intensity or appropriateness, or a downstream clinical, patient-relevant, efficiency, or economic consequence.

Population (P): Patients, teeth, tooth surfaces, extracted teeth/radiographs, or clinicians/trainees performing radiographic caries assessment or related treatment decisions.

Intervention (I): AI-assisted radiographic caries assessment, including machine learning, deep learning, convolutional neural networks, CAD, automated detection/classification, or decision-support systems.

Comparison (C): Unassisted/usual clinical assessment, alternative AI or decision strategy, clinician assessment, or an appropriate reference standard.

Outcomes (O): Primary: treatment decisions, treatment recommendations, treatment intensity, treatment appropriateness, or clinician decision impact. Secondary: clinical, patient-relevant, efficiency, or economic consequences.

Study design (S): Randomized or non-randomized comparative studies, clinician decision-impact studies, diagnostic/decision studies, economic evaluations, and AI systems linking radiographic findings to treatment planning or clinical decision support.

Studies were included if they met the following criteria: (1) original research evaluating AI-assisted radiographic caries detection; (2) involvement of dentists or other dental clinicians in interpreting radiographs or making treatment decisions; (3) assessment of treatment decisions with and without AI assistance, where applicable; (4) use of randomized, crossover, prospective, retrospective, or other appropriately designed observational methods; and (5) reporting of at least one clinically relevant outcome, including treatment recommendations, decision appropriateness, actual treatment, or patient-related outcomes.

Studies were excluded if they (1) reported only diagnostic performance metrics, such as sensitivity, specificity, or area under the receiver operating characteristic curve (AUC), without evaluating clinical decision-making or related outcomes; (2) investigated AI-based detection of conditions unrelated to dental caries; (3) were reviews, editorials, commentaries, or conference abstracts without sufficient original data; or (4) lacked a relevant comparison group when comparative evaluation of AI-assisted and non-AI-assisted decision-making was necessary to address the review question.

Information sources and search strategy

Four electronic databases, PubMed, Scopus, Web of Science, and SpringerLink, were searched on 17 September 2026. The search strategy used three core concepts, artificial intelligence or machine learning, dental caries, and dental radiography, connected using Boolean operators, structured as follows: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "deep neural network" OR "convolutional neural network" OR "neural network" OR "computer-aided diagnosis" OR "computer-assisted diagnosis" OR "AI-assisted" OR "AI-based") AND ("dental caries" OR caries OR "caries lesion" OR "carious lesion" OR "tooth decay" OR "dental decay") AND ("dental radiograph*" OR

"dental X-ray*" OR bitewing* OR "intraoral radiograph*" OR "periapical radiograph*" OR "panoramic radiograph*" OR orthopantomogram* OR "dental imaging" OR CBCT).

Study selection process

The search yielded 998 records: PubMed (n=274), Scopus (n=422), Web of Science (n=76), and SpringerLink (n=226). After removal of 297 duplicates, 701 records remained for title and abstract screening. Of these, 676 were excluded. Twenty-five reports were retrieved for full-text assessment, all of which were available. Fourteen reports were excluded because they involved AI model development or validation without a downstream clinical outcome (n=9), diagnostic accuracy only (n=2), insufficient clinical relevance to the review question (n=2), or non-radiographic AI caries detection (n=1). Eleven studies met the eligibility criteria and were included in the qualitative synthesis. The selection process is presented in Figure 1. All retrieved records were imported into Rayyan for screening and management. Following duplicate removal, title and abstract screening were conducted by both authors using the predefined eligibility criteria. Potentially relevant records were subsequently assessed through full-text review. Final inclusion and exclusion decisions were based on the screening results recorded in Rayyan.

Data extraction and synthesis

Data were extracted using a standardized framework covering author and year, study design and setting, population or dataset, radiographic modality, AI method, comparator or reference standard, clinical pathway role, diagnostic findings, treatment decisions or recommendations, treatment intensity, clinical or patient-relevant consequences, economic outcomes, and major study limitations. Economic outcomes were distinguished as modeled rather than observed outcomes, and related studies using overlapping datasets or assumptions were interpreted as a connected body of evidence rather than as fully independent replications.

The findings were synthesized narratively. Evidence was organized according to its position in the clinical pathway: clinician–AI decision impact, economic/ modeling consequences, ex vivo or diagnostic decision analysis, and AI-based treatment planning or decision support. Treatment decisions were interpreted according to management intensity, including monitoring, preventive/non-invasive care, micro-invasive care, and invasive/restorative treatment. Increased treatment intensity was not automatically classified as overtreatment unless inappropriateness was demonstrated or strongly supported by the study context.

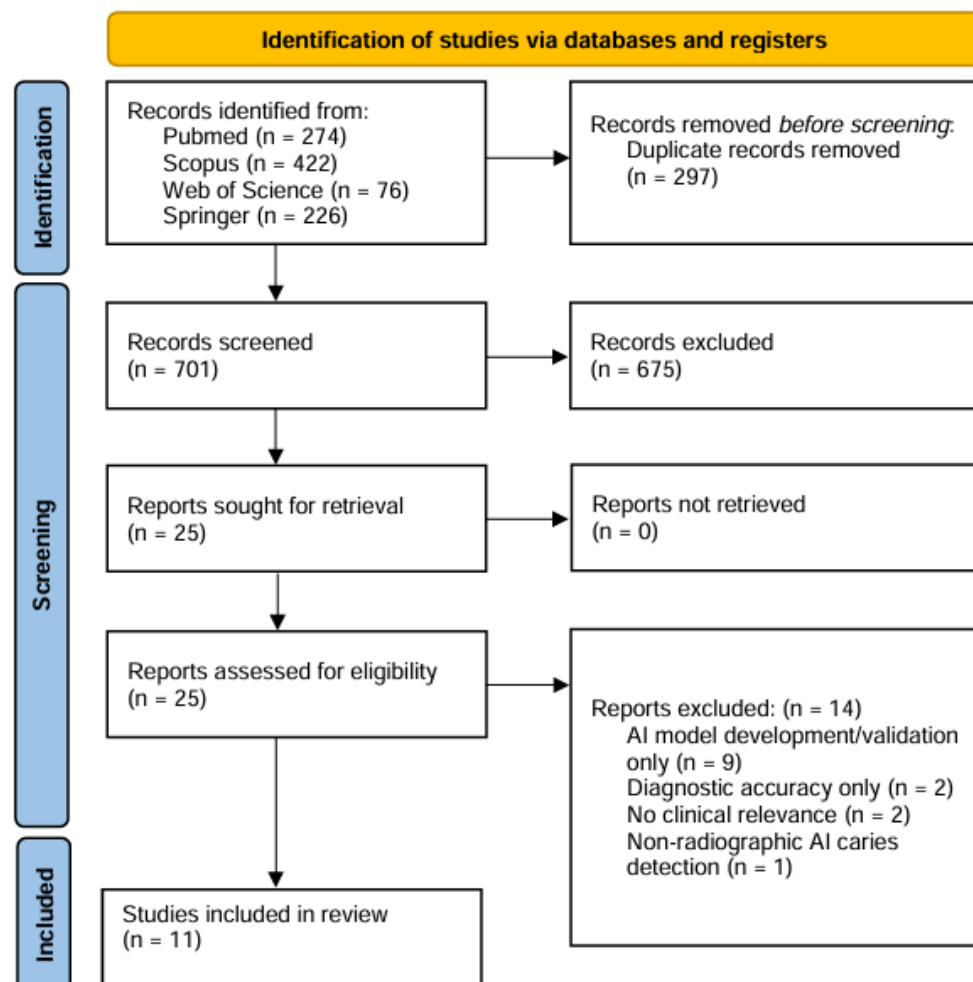


FIGURE 1: PRISMA 2020 flow diagram of study identification, screening, eligibility assessment, and inclusion.

RESULTS

Information sources and search strategy

Eleven studies published between 2021 and 2026 met the eligibility criteria. The evidence base comprised one cluster-randomized crossover trial [10], four economic/modeling evaluations [11–14], two *ex vivo* or diagnostic decision studies [15,16], and four studies evaluating AI-based treatment planning or broader decision support [17–20]. Imaging modalities included bitewing, panoramic, and periapical radiography, while one pediatric framework incorporated radiographic information together with clinical, genetic, and behavioral data [19]. Five economic or modeling reports were closely related in data source, modeling assumptions, or research group, limiting the extent

to which they should be interpreted as independent replications [11–14].

The most clinically informative evidence came from the randomized crossover trial by Mertens et al. [10], in which 22 dentists evaluated bitewing radiographs with and without AI assistance. The remaining evidence primarily evaluated modeled downstream outcomes, human–AI decision rules, retrospective software performance, or technical treatment-planning systems. These differences are important because technical model performance, modeled economic outcomes, and observed clinician behavior answer different questions. Characteristics of the included studies are summarized in Table 1.

TABLE 1: Summary of studies included in the systematic review.

Reference	Design and population/dataset	Radiograph and AI system	Comparator/reference	Main findings
Mertens et al., 2021 [10]	Cluster-randomized crossover trial; 22 dentists	Bitewing; dentalXrai Pro	Dentists without AI; expert reference	AUC 0.85→0.89; sensitivity 0.72→0.81; specificity similar. Non-invasive treatment 32%→36%; invasive treatment 17%→24%.
Schwendicke et al., 2022 [11]	Health-economic/value-of-information model	Bitewing; ResNet-18	Dentists with/without AI; expert reference	AI increased projected tooth retention and reduced detection costs; value of additional data varied with caries risk, training volume, and AI cost.
Schwendicke et al., 2021 [12]	Model-based cost-effectiveness; 12-year-old population; lifetime horizon	Bitewing; U-Net	Visual-tactile + bitewing assessment; expert consensus	AI projected tooth retention 64 vs 62 years and costs €298 vs €322; results depended on early-lesion management.
Schwendicke et al., 2022 [13]	Markov cost-effectiveness model using trial data	Bitewing; dentalXrai Pro	23-dentist trial dataset; expert reference	AI increased sensitivity but also invasive treatment; modeled retention and costs were similar, illustrating dependence on observed treatment behavior.
Gomez Rossi et al., 2022 [14]	Health-economic Markov evaluation	Bitewing; CNN decision-support	Standard screening without AI	Higher projected tooth retention (62.4 vs 60.9 years) and lower costs (€320 vs €342); cost-effectiveness depended strongly on post-diagnosis treatment.
Ganss et al., 2026 [15]	<i>Ex vivo</i> diagnostic/decision study	Periapical/bitewing digital radiography; Nostic AI	Dentists of varying experience; histological tooth sections	AI sensitivity exceeded dentists for enamel and dentine lesions; AND integration produced the lowest clinical loss when false positives were penalized.
Orhan et al., 2023 [16]	Retrospective diagnostic study; 100 panoramic radiographs/4,497 teeth	Panoramic; Diagnocat	Expert radiologists and verified clinical records	High agreement for several dental conditions, but caries sensitivity was low (36%), limiting reliability for caries treatment planning.
Ranjbar & Zamanifar [17]	Pilot treatment-planning study; 1,025 panoramic images	Panoramic; YOLOv7	Experienced dentist/clinical team labels	Generated plans across eight treatment classes; precision 100%, recall 96%, F1 76%, mAP@0.5 81.9%; technical feasibility only.
Oruç et al., 2025/2026 [18]	Retrospective comparative study; 1,253 augmented bitewings	Bitewing; Detectron-2, U-Net, YOLOv8	Consensus dentomaxillofacial radiologists	Segmentation accuracy 89.23–95.03%; treatment recommendation accuracy 61.90–88.09%, showing a gap between detection and recommendation.

Reference	Design and population/dataset	Radiograph and AI system	Comparator/reference	Main findings
Marie et al., 2025 [19]	Multicenter multimodal pediatric framework	Clinical images, panoramic/CBCT plus other data; neutrosophic-enzyme framework	Conventional clinical examination/IC DAS-II and other DL models	Reported accuracy 97.3%, sensitivity 94.7%, specificity 96.2%, AUC 0.973, time reduction 37.5%, and cost reduction 34.5%; multimodal design limits attribution to radiographic AI.
Jiang et al., 2026 [20]	Multicenter deep-learning development/validation; 1,170 panoramic radiographs	Panoramic; SMM-YOLOv8n	Expert-annotated ground truth	mAP@50 0.886, recall 0.960, F1 0.926; junior-dentist sensitivity improved 0.508→0.725 and interpretation time decreased by 8.79 min/60 images.

AI support and clinician treatment decisions

The strongest direct evidence that AI changes treatment decisions came from Mertens et al. [10]. AI assistance increased sensitivity from 0.72 to 0.81 and AUC from 0.85 to 0.89 while specificity remained similar. Importantly, the effect was not limited to detection: treatment choices also changed, with non-invasive treatment increasing from 32% to 36% and invasive treatment from 17% to 24%. The finding demonstrates that AI can alter the clinical pathway, but it does not establish that the resulting treatment distribution was more appropriate or beneficial to patients.

Ganss et al. [15] provide mechanistic evidence for this distinction. Human-AI integration rules generated different trade-offs between false-positive and false-negative consequences. When false positives were penalized, the AND rule minimized clinical loss for early dentine lesions, whereas an OR rule favored sensitivity. Thus, the optimal operating point is not determined by diagnostic accuracy alone; it depends on the clinical cost assigned to each type of error.

Orhan et al. [16] further illustrate that AI performance cannot be generalized across radiographic modalities. Although the panoramic system demonstrated high agreement for several dental conditions, sensitivity for caries was only 36%. A treatment-planning conclusion based on such performance would therefore be substantially less secure than a conclusion based on validated bitewing caries detection.

Projected clinical and economic consequences

Four economic evaluations [11–14] projected potential downstream benefits, but their results were conditional on the assumptions used to translate AI detection into management. The 2021 model projected tooth retention of 64 versus 62 years and lower costs (€298 versus €322) with AI [12]. Related analyses reported favorable economic outcomes when additional information from AI was assumed to support conservative management of early lesions [11,14].

In contrast, the model populated with decisions observed in the randomized trial did not show the same advantage: modeled tooth retention was similar, and costs were identical in the reported comparison [13]. This discrepancy is informative

rather than contradictory. It indicates that the economic value of AI is sensitive to what clinicians do after a lesion is detected. Accordingly, favorable cost-effectiveness estimates should not be interpreted as evidence that AI itself reduces costs; they represent the combined effect of AI performance and the downstream management pathway.

Factors influencing projected downstream outcomes

Four factors influenced the projected clinical and economic outcomes across the evaluations. First, the post-detection management pathway determines whether additional detections lead to preventive benefit or treatment escalation. Second, the source of the diagnostic input matters: models populated with favorable accuracy estimates may produce different outcomes from models populated with observed clinician decisions. Third, baseline caries risk modifies the value of additional diagnostic information [11]. Fourth, AI cost and training-data volume influence the economic value of deployment [11,14]. These determinants indicate that clinical and economic value is a property of the entire care pathway rather than of the algorithm in isolation.

AI systems generating treatment recommendations. Four studies evaluated systems that extended beyond detection toward treatment planning or broader clinical decision support [17–20]. Oruç et al. [18] demonstrated that high segmentation performance did not translate uniformly into treatment recommendation accuracy: recommendation accuracy ranged from 61.90% to 88.09% across the evaluated architectures. Ranjbar and Zamanifar [17] generated treatment plans from panoramic radiographs, while Jiang et al. [20] reported improved junior-dentist sensitivity and reduced interpretation time. These findings support technical feasibility and potential efficiency gains, but they do not demonstrate that AI recommendations are safer, more appropriate, or better for patients.

Marie et al. [19] reported very high performance and reductions in diagnostic time and cost, but the framework combined radiographic, clinical, genetic, and behavioral information. It therefore cannot be interpreted as evidence for radiographic caries AI alone. Across the recommendation studies, validation generally depended on retrospective data and expert-derived labels or plans.

Agreement with an expert-generated plan measures replication of the reference decision; it does not establish clinical appropriateness because treatment appropriateness depends on patient-specific factors and longitudinal outcomes.

DISCUSSION

This review demonstrates a clear difference between the maturity of evidence for AI diagnostic performance and the maturity of evidence for clinical utility. Eleven studies met the eligibility criteria, but only one was a randomized clinician-level trial, and none prospectively followed patients to determine whether AI-supported decisions improved patient-important outcomes. The evidence consistently suggests that AI can change detection and decision-making, while the direction and value of those changes depend on lesion characteristics, clinician behavior, management thresholds, modality, and economic assumptions.

The most important finding is therefore an accuracy–decision gap: improving a diagnostic metric does not guarantee improvement in the clinical decision generated from that metric. The randomized evidence showed higher sensitivity and AUC alongside more invasive treatment [10]. Economic models favored AI when early lesions were managed non-restoratively, whereas the trial-informed model did not show the same advantage [13]. Together, these findings indicate that downstream management is a critical mediator of AI's clinical and economic value.

The accuracy–decision gap is particularly consequential in caries because early lesions are not synonymous with an indication for restorative treatment. If an AI system detects more early lesions but the clinical response is predominantly operative, the algorithm may increase treatment intensity without improving disease control or tooth survival. Conversely, if increased detection is integrated into lesion-specific preventive or micro-invasive pathways, the same diagnostic improvement may produce greater benefit.

This has direct implications for algorithm development. Future systems should not report only sensitivity, specificity, or AUC; they should also characterize lesion depth or stage, uncertainty, and performance at clinically relevant thresholds. Where treatment recommendations are generated, the system should distinguish detection from management and provide recommendations that are explicitly linked to evidence-based lesion-specific care. The *ex vivo* findings of Ganss et al. [15] reinforce that decision thresholds should reflect the clinical consequences of false-positive and false-negative errors rather than maximizing a single diagnostic metric.

The impact of AI was not uniform across clinicians. The randomized study showed individual variation in performance, indicating that experience, interaction with the system, and user trust may modify the effect of AI assistance [10]. Consequently,

a pooled improvement in diagnostic performance should not be assumed to occur equally for every clinician or clinical setting.

Population and modality also influence expected benefit. Economic analyses indicated that baseline caries risk affects the value of additional information [11], while the panoramic study by Orhan et al. [16] showed poor caries sensitivity despite good performance for other dental conditions. Deployment decisions should therefore be risk- and modality-specific rather than based on the assumption that an AI system validated for one radiographic context will generalize to another.

Systems that directly generate treatment recommendations are attractive because they attempt to bridge the gap between detection and management. However, their current evidence primarily establishes technical feasibility. The distinction between segmentation accuracy, diagnostic classification, treatment recommendation accuracy, and actual treatment appropriateness is essential. A system can correctly reproduce an expert-labeled plan while still producing a clinically inappropriate recommendation for a patient whose lesion activity, cavitation, caries risk, restorative history, or preferences differ from the reference case.

The available studies also lack prospective comparisons of clinicians with versus without AI-generated treatment recommendations using patient-important outcomes. Reported reductions in reading time or increases in sensitivity therefore indicate efficiency or decision-support potential, not improved treatment quality. Independent external validation and prospective clinical evaluation are necessary before these systems can be considered autonomous or semi-autonomous treatment authorities.

For clinicians, AI-assisted radiographic detection should currently be regarded as an adjunct to clinical examination and structured treatment planning rather than an autonomous treatment authority. For developers, model reporting should include lesion stage/depth, clinically relevant thresholds, uncertainty, external validation, and the consequences of false-positive and false-negative decisions. For health systems and payers, economic evaluation should incorporate the actual treatment pathway and AI implementation cost rather than extrapolating from diagnostic accuracy alone.

Future trials should evaluate the complete pathway from AI-supported detection to treatment decision and patient outcome. Relevant outcomes include avoidance of unnecessary restoration, lesion progression or arrest, tooth survival, repeat operative treatment, treatment burden, patient-reported outcomes, and real-world cost. Pragmatic studies should include diverse clinicians and populations and should distinguish between AI assistance that changes a clinician's decision and AI that independently generates a treatment recommendation.

The principal strength of this review is its deliberate focus on the clinical pathway beyond diagnostic accuracy. It distinguishes direct clinician decision evidence from modeled economic projections and from technical treatment-planning studies. The review also identifies dependence among related economic analyses and avoids interpreting increased treatment intensity as overtreatment unless inappropriateness is demonstrated.

Several limitations should be considered. The evidence base was small and heterogeneous, preventing quantitative meta-analysis. Several economic studies originated from closely related datasets, models, and research groups, reducing evidence independence. The search was restricted to four databases and did not systematically include engineering-specific databases or grey literature, which may have omitted technical studies. Rapid changes in AI architectures and software also limit temporal generalizability. Finally, the included literature may be affected by publication and reporting bias, particularly because favorable technical AI results are more likely to be reported than unsuccessful or neutral deployments.

CONCLUSION

AI-assisted radiographic caries detection can improve diagnostic performance and can modify clinician treatment decisions, but improved detection does not automatically translate into more appropriate treatment or better patient outcomes. The strongest direct evidence indicates that AI can increase both non-invasive and invasive treatment decisions, demonstrating that diagnostic gains may be accompanied by treatment escalation. Economic benefits are conditional on downstream management assumptions, particularly conservative management of early lesions. AI systems that generate treatment recommendations demonstrate technical feasibility and potential efficiency gains, but current validation does not establish treatment appropriateness, safety, or long-term patient benefit. Future research should evaluate AI within lesion-specific management pathways using prospective, independently validated studies with longitudinal patient-important outcomes.

REFERENCES

- [1] Kühnisch, J., Aps, J.K., Splieth, C., Lussi, A., Jablonski-Momeni, A., Mendes, F.M., Schmalz, G., Fontana, M., Banerjee, A., Ricketts, D., Schwendicke, F., Douglas, G., Campus, G., van der Veen, M., Opdam, N., Doméjean, S., Martignon, S., Neuhaus, K.W., Horner, K. and Huysmans, M.-C.D. (2024). ORCA-EFCD consensus report on clinical recommendation for caries diagnosis. Paper I: Caries Lesion detection and depth assessment. *Clinical Oral Investigations*, 28(4). doi:10.1007/s00784-024-05597-3.
- [2] Mohammad-Rahimi, H., Motamedian, S.R., Rohban, M.H., Krois, J., Uribe, S.E., Mahmoudinia, E., Rokhshad, R., Nadimi, M. and Schwendicke, F. (2022). Deep learning for caries detection: A systematic review. *Journal of Dentistry*, [online] 122, p.104115. doi:10.1016/j.jdent.2022.104115.
- [3] Musri, N., Christie, B., Ichwan, S.J.A. and Cahyanto, A. (2021). Deep learning convolutional neural network algorithms for the early detection and diagnosis of dental caries on periapical radiographs: A systematic review. *Imaging Science in Dentistry*, 51(3), p.237. doi:10.5624/isd.20210074.
- [4] Carvalho, B.K.G., Nolden, E.-L., Wenning, A.S., Kiss-Dala, S., Gergely, A., Róth, I., Kerémi, B., Géczi, Z., Hegyi, P. and Kivovics, M. (2024). Diagnostic accuracy of artificial intelligence for approximal caries on bitewing radiographs: A systematic review and meta-analysis. *Journal of Dentistry*, [online] 151, p.105388. doi:10.1016/j.jdent.2024.105388.
- [5] Abbott, L.P., Saikia, A. and Anthonappa, R.P. (2025). Artificial intelligence platforms in dental caries detection: A systematic review and meta-analysis. *Journal of Evidence-Based Dental Practice*, 25(1), p.102077. doi:10.1016/j.jebdp.2024.102077.
- [6] Zanini, L.G.K., Rubira-Bullen, I.R.F. and Nunes, F. de L. Dos S. (2024). A Systematic Review on Caries Detection, Classification, and Segmentation from X-Ray Images: Methods, Datasets, Evaluation, and Open Opportunities. *Journal of Imaging Informatics in Medicine*, [online] 37(4), pp.1824–1845. doi:10.1007/s10278-024-01054-5.
- [7] Horvath, K.S.H., Gjerdet, N.R. and Shi, X.-Q. (2025). Advancements in Caries Diagnostics Using Bitewing Radiography: A Systematic Review of Deep Learning Approaches. *Caries Research*, pp.1–32. doi:10.1159/000546448.
- [8] Lai, J., Guo, S., Wang, K., Li, X., Yu, X. and Wang, X. (2026). Artificial Intelligence for Binary Dental Caries Diagnosis Using Intraoral Images and Dental Radiographs: A Systematic Review and Meta-analysis. *BMC Oral Health*, [online] 26(1), p.437. doi:10.1186/s12903-026-07770-4.
- [9] Lam, Q.T., Le, M.H.N., Fan, F.-Y., Le, N.Q.K. and Lee, I.-T. (2026). Deep Learning-Based Dental Caries Diagnosis: A Modality-Stratified Systematic Review and Meta-Analysis of Faster R-CNN and Mask R-CNN. *Diagnostics*, 16(5), p.731. doi:10.3390/diagnostics16050731.
- [10] Mertens, S., Krois, J., Cantu, A.G., Arsiwala, L.T. and Schwendicke, F. (2021). Artificial Intelligence for Caries Detection: Randomized Trial. *Journal of Dentistry*, [online] 115, p.103849. doi:10.1016/j.jdent.2021.103849.
- [11] Schwendicke, F., Oro, C.G. de J., Garcia, C.A., Meyer-Lueckel, H., Chaurasia, A. and Krois, J. (2022b). Artificial Intelligence for Caries Detection: Value of Data and Information. *Journal of Dental Research*, 101(11), pp.1350–1356. doi:10.1177/00220345221113756.

- [12] Schwendicke, F., Rossi, J.G., Göstemeyer, G., Elhennawy, K., Cantu, A.G., Gaudin, R., Chaurasia, A., Gehrung, S. and Krois, J. (2021). Cost-effectiveness of Artificial Intelligence for Proximal Caries Detection. *Journal of Dental Research*, 100(4), pp.369–376. doi:10.1177/0022034520972335.
- [13] Schwendicke, F., Mertens, S., Cantu, A.G., Chaurasia, A., Meyer-Lueckel, H. and Krois, J. (2022a). Cost-effectiveness of AI for Caries Detection: Randomized Trial. *Journal of Dentistry*, [online] 119, p.104080. doi:10.1016/j.jdent.2022.104080.
- [14] Rossi, G.J., Rojas-Perilla, N., Krois, J. and Schwendicke, F. (2022). Cost-effectiveness of artificial intelligence as a decision-support system applied to the detection and grading of melanoma, dental caries, and diabetic retinopathy. *JAMA Network Open*, [online] 5(3), p.e220269. doi:10.1001/jamanetworkopen.2022.0269.
- [15] Ganss, C., Jung, K., Schilling, L., Sonderegger, S. and Neuhaus, K.W. (2026). Human and Artificial Intelligence Performance in Radiographic Caries Detection: Ex Vivo Tooth Section-referenced Evaluation and Implications for Clinical Decision-making. *Journal of Dentistry*. *Journal of Dentistry*, [online] 168, p.106602. doi:10.1016/j.jdent.2026.106602.
- [16] Orhan, K., Belgin, C.A., Manulis, D., Golitsyna, M., Bayra, S., Aksoy, S., Sanders, A., Onder, M., Ezhov, M., Shamshiev, M., Gusarev, M. and Shlenskii, V. (2023). Determining the Reliability of Diagnosis and Treatment Using Artificial Intelligence Software with Panoramic Radiograph. *Imaging Science in Dentistry*, [online] 53(3), pp.199–199. doi:10.5624/isd.20230109.
- [17] Ranjbar, R.F. and Zamanifar, A. (2023). Autonomous Dental Treatment Planning on Panoramic X-ray Using Deep Learning Based Object Detection Algorithm. *Multimedia Tools and Applications*, 83(14), pp.42999–43033. doi:10.1007/s11042-023-17048-4.
- [18] Oruç, M.S., Yetik, İ.Ş., İncekürk, Ö.K., Çulhaoğlu, A.K., Kılıçarslan, M.A., Evli, C. and Kurt, M.H. (2025). Automated Differentiation of Caries Requiring Filling and Caries Necessitating Root Canal Treatment Using Machine Learning. *Oral Radiology*, 42(2), pp.421–431. doi:10.1007/s11282-025-00874-7.
- [19] Marie, H.S., Elbaz, M., Soliman, R.S., Hafez, M.E. and Elkhatib, A.A. (2025). Bio-inspired Neutrosophic-enzyme Intelligence Framework for Pediatric Dental Disease Detection Using Multi-modal Clinical Data. *Scientific Reports*, 15(1). doi:10.1038/s41598-025-21923-5.
- [20] Jiang, Y., Jin, J., Gao, Y., Tian, Y., Hu, L., Lu, Y. and Fu, Y. (2026). A Deep Learning Model for Second-Molar Lesions Related to Impacted Third Molars. *International Dental Journal*, [online] 76(3), p.109467. doi:10.1016/j.identj.2026.109467