AMERICAN Journal of Pediatric Medicine and Health Sciences
Volume 4, Issue 08, 2026 ISSN (E): 2993-2149

Artificial Intelligence in the Diagnosis of Oral Diseases: Current Applications, Diagnostic Accuracy, and Future Perspectives

Absamatova Mohlaroyim Muzaffarovna
Second-Year Student at Samarkand State Medical University
absamatovamokhlaroyim@gmail.com

Abstract: Background: Artificial intelligence (AI), particularly machine learning and deep learning, is increasingly investigated as a decision-support technology for oral diagnosis. Dentistry is well suited to AI because diagnosis routinely integrates radiographs, photographs, three-dimensional imaging, pathology, and structured clinical information. Objective: This evidence-based literature review synthesizes current applications of AI in oral diagnosis, examines reported diagnostic performance, identifies methodological limitations, and outlines requirements for safe clinical translation. Methods: A structured narrative review was conducted using peer-reviewed literature identified through targeted searches of PubMed-indexed records and publisher databases, emphasizing systematic reviews, meta-analyses, and contemporary reporting and governance guidance. Reported performance measures were extracted without generating new patient-level estimates. Results: Evidence is most mature for dental caries, oral potentially malignant disorders and oral cancer, periapical lesions, periodontal disease, and automated tooth identification. Recent syntheses report promising performance, including accuracy ranges of 71–99% for caries models, 85–100% for deep-learning oral-cancer studies, 74–100% for oral mucosal lesion detection, pooled sensitivity/specificity of 0.94/0.96 for periapical radiolucencies, and pooled dental-diagnostic sensitivity/specificity of approximately 0.85/0.93 across low-bias systematic reviews. However, heterogeneity, limited external validation, single-center datasets, spectrum bias, inconsistent reference standards, and uncertain clinical benefit remain major barriers. Conclusion: AI can augment oral diagnosis, but current evidence supports clinician-supervised decision support rather than autonomous diagnosis. Multicenter prospective validation, transparent reporting, representative datasets, explainability, equity safeguards, and workflow-level outcome evaluation are essential before widespread implementation.

Keywords: Artificial intelligence; Deep learning; Oral diseases; Dental diagnosis; Diagnostic accuracy; Digital dentistry; Clinical decision support

Introduction

Artificial intelligence (AI) has moved from a largely experimental concept in dentistry toward a clinically relevant family of computational methods capable of detecting patterns in images, structured records, and multimodal health data. Machine learning algorithms learn statistical relationships from examples, while deep learning uses multilayer neural networks to learn increasingly complex representations directly from data. In oral healthcare, this distinction matters because many diagnostic tasks are image-intensive: intraoral photographs, bitewing and periapical radiographs, panoramic radiographs, cone-beam computed tomography (CBCT), cephalometric images, and digitized histopathology can all be converted into analyzable digital information. AI therefore has a natural role in computer-assisted detection, classification, segmentation, measurement, and decision support [1].

The clinical rationale is strong. Oral diseases are common, heterogeneous, and frequently diagnosed through combinations of visual inspection, tactile examination, radiographic interpretation, and pathology. Diagnostic performance can vary with clinician experience, image quality, lesion size, disease prevalence, fatigue, and interpretation thresholds. AI may reduce some sources of inter- and intra-observer variability by applying the same computational rules repeatedly, while also providing rapid localization of suspicious findings. Earlier reviews have documented applications across dental caries, periodontal disease, periapical pathology, maxillofacial lesions, salivary gland disease, osteoporosis-related findings, oral cancer, and orthodontic diagnosis. Nevertheless, the existence of a high-performing model in a retrospective dataset does not establish clinical effectiveness. A model may identify a lesion accurately but still fail to improve patient outcomes, increase inappropriate referrals, or perform poorly when transferred to another population or imaging device [2].

The evidence base has expanded substantially in recent years. Systematic reviews have examined AI for oral diseases as a broad category, while disease-specific reviews have separately evaluated caries, oral cancer, oral mucosal lesions, tooth identification, periapical radiolucencies, periodontal disease, and orthodontic diagnosis. A 2025 overview of systematic reviews synthesized 116 systematic reviews and reported pooled diagnostic performance estimates from low-bias reviews, while also demonstrating substantial heterogeneity between applications . These findings suggest that the central question is no longer whether AI can perform oral diagnostic tasks at all, but under what conditions its performance is reliable, clinically useful, equitable, and generalizable.

The methodological gap is particularly important. Many dental AI studies remain retrospective, use convenience datasets from one institution, rely on internal train-test splits, and report accuracy without sufficient attention to calibration, prevalence, external validation, or clinically meaningful endpoints. In some areas, the reference standard is also imperfect. Histopathology may provide a strong reference for oral malignancy, whereas radiographic labels may depend on expert interpretation and can vary across observers. Differences in image acquisition, annotation protocols, disease spectrum, and outcome definitions make direct comparison between reported accuracy values inappropriate [3].

The regulatory and ethical context adds another layer of complexity. AI systems may influence diagnosis, referral, treatment planning, or patient communication, creating responsibilities concerning safety, transparency, privacy, accountability, and equity. The World Health Organization has emphasized human autonomy, safety, transparency, accountability, inclusiveness, and sustainability as core principles for AI in health . For diagnostic AI research, the STARD-AI reporting guideline further emphasizes transparent reporting of datasets, index tests, model development, evaluation, algorithmic bias, and applicability .These principles are directly relevant to dentistry because a model that is technically accurate but poorly reported or insufficiently validated cannot be responsibly translated into routine care [4].

The present review therefore aims to synthesize current evidence on AI in the diagnosis of oral diseases, with four objectives: (1) to characterize the principal diagnostic applications; (2) to summarize reported diagnostic accuracy without creating unsupported empirical estimates; (3) to identify methodological, clinical, and ethical barriers to implementation; and (4) to define priorities for future research and responsible clinical adoption.

Materials and Methods

Study design and reporting approach. Because no original patient-level dataset, prospective clinical cohort, diagnostic test dataset, or experimental measurements were provided, the appropriate design for this manuscript is an evidence-based narrative literature review rather than an original empirical diagnostic study. The review was structured in an IMRAD-compatible format and informed by the principles of PRISMA 2020 for transparent evidence synthesis .It does not claim to be a de novo registered systematic review or meta-analysis, and no pooled estimate was calculated by the present authors. Reported numerical findings are reproduced only when they were available in verified systematic reviews or meta-analyses.

Information sources and search strategy. Literature was identified through targeted searches of PubMed-indexed records and publisher-hosted pages for peer-reviewed dental and medical literature. Search concepts combined terms for artificial intelligence, machine learning, deep learning, convolutional neural networks, oral diseases, dental diagnosis, diagnostic accuracy, sensitivity, specificity, oral cancer, oral mucosal lesions, caries, periodontitis, periapical lesions, dental radiography, and orthodontic diagnosis. Priority was given to systematic reviews, meta-analyses, comprehensive reviews, and methodological or governance guidance published between 2020 and August 2026, while earlier landmark evidence was retained when necessary to explain the development of dental AI. Reference verification was performed against PubMed or the official journal/publisher record whenever possible.

Eligibility and evidence selection. Studies were considered relevant when they evaluated AI or deep-learning methods for diagnosis, detection, classification, segmentation, identification, or diagnostic decision support involving oral or dental conditions in humans. Evidence was prioritized when it reported diagnostic performance or systematically synthesized diagnostic studies. Reviews focused exclusively on non-diagnostic applications were not central to the synthesis. Methodological guidelines were included when they directly addressed reporting, transparency, or ethical governance of AI-based diagnostic research. Because the objective was an evidence synthesis rather than a new systematic review, no numerical study-selection flow or fabricated count of retrieved records is reported.

Data extraction and synthesis. The following domains were extracted narratively from the selected evidence: disease or diagnostic target; input modality; principal AI task; model family when reported; diagnostic performance; comparison with clinicians when available; risk-of-bias or applicability concerns; and implications for clinical translation. Where reviews provided ranges, pooled sensitivity/specificity, or other summary measures, these values were retained with their original context. Results from different disease domains were not mathematically combined because their clinical targets, prevalence, reference standards, imaging modalities, and outcome definitions are not interchangeable.

Quality and bias assessment. Rather than assigning a new numerical quality score to every primary study, the review incorporated risk-of-bias findings reported by the included systematic reviews. Particular attention was given to patient-selection bias, spectrum restriction, reference-standard limitations, internal versus external validation, data leakage, applicability to routine clinical populations, and inconsistent reporting of performance metrics. These issues are especially important in AI because a model can appear highly accurate when tested on data that are closely related to its training distribution but perform substantially less well on unseen clinical environments. The recently published STARD-AI framework was used as a conceptual benchmark for transparent reporting of AI diagnostic accuracy studies .

Ethical considerations. No human participants were recruited, no identifiable patient data were accessed, and no intervention was performed by the authors for this review. Therefore, institutional ethics approval and informed consent were not applicable to the review itself. Ethical considerations concerning the deployment of AI in oral diagnosis were nevertheless incorporated into the interpretation, particularly privacy, human oversight, fairness, transparency, accountability, and patient safety, consistent with WHO guidance .

Results

Overview of the evidence. The literature demonstrates that AI applications in oral diagnosis have expanded from relatively narrow image-recognition experiments to a broader ecosystem of detection, segmentation, classification, automated identification, and clinical decision-support tasks. The strongest evidence is concentrated in image-based diagnosis, particularly dental radiography and clinical photographs. Reviews published from 2022 through 2026 consistently describe promising model performance while simultaneously identifying heterogeneity and limitations that prevent a single global estimate of AI diagnostic accuracy [5].

Dental caries. Dental caries is one of the most extensively investigated oral conditions for deep learning. Models have been developed for intraoral photographs, periapical and bitewing radiographs, panoramic images, near-infrared transillumination, and optical coherence tomography. A systematic review of 42 eligible deep-learning studies reported accuracy ranges of 71–96% for intraoral photographs, 82–99.2% for periapical radiographs, 87.6–95.4% for bitewing radiographs, 68–78% for near-infrared transillumination, 88.7–95.2% for optical coherence tomography, and 86.1–96.1% for panoramic radiographs . The same review found substantial methodological heterogeneity and concluded that reported accuracy was promising but study and reporting quality were frequently limited. This is clinically important because early enamel lesions are smaller and more visually ambiguous than established cavitated lesions; therefore, high performance in advanced lesions cannot automatically be extrapolated to early disease [6].

The evidence also indicates that AI can support more than binary caries detection. Object-detection and segmentation models can identify the location and extent of lesions, while automated tooth identification can provide the anatomical context needed to report which tooth is affected. This creates the possibility of an integrated radiographic workflow in which a system identifies teeth, detects suspicious areas, and presents candidate lesions to the clinician. However, such workflows should be evaluated as complete clinical systems rather than as isolated algorithms because errors at one stage can propagate into later stages [7].

Oral potentially malignant disorders and oral cancer. Oral cancer represents an important application because delayed recognition can increase treatment burden and worsen outcomes. Deep-learning systems have been investigated using clinical photographs, radiographic images, histopathological images, and other diagnostic data. A 2024 systematic review of deep learning in oral cancer included 54 studies, of which 51 were diagnostic and three addressed prognostic prediction. Reported performance included accuracy of 85.0–100%, F1 scores of 79.31–89.0%, Dice coefficients of 76.0–96.3%, and concordance indices of 0.78–0.95 across different tasks [8]. These results demonstrate that deep learning can recognize complex visual patterns relevant to malignancy, but the heterogeneity of input data and tasks means that the figures should not be interpreted as a universal accuracy range for clinical oral-cancer diagnosis.

A separate systematic review and meta-analysis focused on oral mucosal lesions photographed in human subjects. Across 36 eligible studies, reported AI accuracy for detecting oral mucosal lesions ranged from 74% to 100%, compared with 61% to 98% for clinicians without AI assistance. Only 7% of included studies were judged to have low risk of bias across all assessed domains. The pooled diagnostic odds ratio was 155 for potentially malignant lesions and 114 for cancerous lesions, although the authors concluded that the clinical benefit of accuracy gains remained uncertain . This distinction is critical: a screening tool may be valuable for triage or referral even when it is not suitable for definitive diagnosis, but its clinical role must be explicitly defined.

Radiographic and endodontic diagnosis. Periapical radiolucencies are another relatively mature target for AI. A 2024 systematic review identified 24 eligible studies, with four contributing to a meta-analysis. The pooled sensitivity was 0.94 and pooled specificity was 0.96, with corresponding 95% confidence intervals of 0.90–0.96 and 0.91–0.98 . Despite these encouraging figures, nearly all studies evaluated model-level accuracy rather than fully deployed clinical software, and the overall body of evidence had unclear-to-high risk of bias and applicability concerns . Consequently, high diagnostic performance should be interpreted as evidence of technical potential rather than proof of improved patient care.

Automated tooth identification and numbering. Accurate identification and numbering of teeth are foundational to dental charting and to downstream automated analysis. A 2024 systematic review and meta-analysis included 29 studies and reported accuracy ranging from 81.8% to 99%, sensitivity from 82.7% to 98%, specificity from 79.9% to 99%, and a pooled area under the curve of 0.96 [8]. Automated tooth identification is diagnostically relevant because it can provide spatial labels for caries, periodontal bone loss, restorations, periapical lesions, and treatment planning. Nevertheless, model performance can be affected by missing teeth, restorations, implants, overlapping structures, mixed dentition, and image quality, reinforcing the need for representative validation datasets [9].

Periodontal diagnosis. Periodontology has increasingly adopted AI for radiographic classification of periodontitis, assessment of alveolar bone loss, photographic screening, and prediction using structured clinical data. A 2026 systematic review found that binary periodontitis classification achieved reported accuracies of 81–99% on panoramic radiographs and 78% on CBCT. Staging or severity assessment generally showed lower performance, with reported accuracy of 64–91% on panoramic radiographs and 83% on intraoral radiographs; photograph-based screening reached an AUROC of 0.93 in the reviewed evidence . Importantly, applicability concerns were frequent, particularly because many datasets were single-center. This illustrates a recurring pattern in dental AI: identifying the presence of a disease may be easier than reliably staging its severity or determining individualized prognosis.

Orthodontic and treatment-planning applications. AI is also used for cephalometric landmark identification, skeletal classification, treatment-need assessment, extraction decisions, and treatment planning. A 2025 scoping review identified 71 relevant articles, with diagnostic applications, landmark identification, and treatment planning representing major domains The review concluded that AI can improve efficiency and reduce operator variability, but accuracy and reliability have not consistently surpassed expert clinicians and human supervision remains essential . These findings reinforce the distinction between automation of repetitive measurements and autonomous clinical decision-making.

Cross-specialty diagnostic performance. The broadest recent overview identified 116 systematic reviews of AI in dentistry. Among 12 low-bias reviews included in its meta-analysis, diagnostic accuracy across dental specialties ranged from 82% to 95%; pooled sensitivity was 0.85 and specificity was 0.93, while CNN-based models had a pooled accuracy of 93.1%. However, substantial heterogeneity was observed across sensitivity, specificity, area under the curve, and accuracy . These findings provide useful high-level context but should not be interpreted as a single performance benchmark for all oral diseases. The diversity of diagnostic targets, image modalities, reference standards, and clinical environments makes disease-specific interpretation necessary [10].

The principal evidence patterns are summarized in Table I. The table deliberately reports only values explicitly available from the verified evidence sources and identifies the principal limitation that constrains clinical interpretation.

Discussion

The present synthesis indicates that AI has reached a level of technical maturity at which clinically relevant diagnostic assistance is plausible across several oral-disease domains. The most consistent evidence concerns image-based tasks in which the diagnostic signal is visually represented and large numbers of labeled examples can be assembled. Caries, oral mucosal lesions, oral cancer, periapical radiolucencies, periodontal bone changes, and tooth identification all demonstrate promising performance in systematic reviews . However, the evidence does not support the conclusion that AI is ready to replace dentists. Instead, the literature supports a more conservative interpretation: AI is best positioned as an adjunct that highlights abnormalities, standardizes measurements, prioritizes cases, and provides a second reader.

Why performance can be high in research but uncertain in practice. Deep-learning systems can exploit subtle image features that may be difficult for human observers to quantify consistently. Convolutional architectures are particularly effective at extracting spatial hierarchies from radiographs and photographs. Yet the same feature-learning capability can make models sensitive to dataset-specific characteristics that are unrelated to disease, such as image acquisition parameters, cropping conventions, scanner signatures, institutional protocols, or demographic composition. If training and test images share hidden characteristics, internal performance can be inflated. The risk is especially relevant when datasets are small or derived from a single institution. Reviews of caries, oral mucosal lesions, periapical disease, and periodontology repeatedly identify heterogeneity, limited external validation, and applicability concerns .

Reference standards and disease spectrum. Diagnostic accuracy depends not only on the algorithm but also on what counts as the truth. For oral cancer, histopathological confirmation can provide a strong reference standard, but the spectrum of lesions in research may differ from that encountered in general dental practice. Studies enriched with obvious cancers and clear normal controls can overestimate discrimination compared with real-world populations containing inflammatory, traumatic, infectious, reactive, and potentially malignant lesions. For caries and periodontal disease, disease severity, lesion activity, imaging modality, and expert interpretation can influence labels. Consequently, an apparently impressive sensitivity or specificity cannot be interpreted without knowing the population, reference standard, disease prevalence, and decision threshold .

Clinical utility versus diagnostic accuracy. Diagnostic accuracy is necessary but not sufficient for adoption. A clinically useful system should improve a meaningful outcome such as time to diagnosis, referral appropriateness, detection of clinically important disease, treatment planning, patient understanding, or prevention of missed pathology. The periapical-lesion evidence is instructive: pooled sensitivity and specificity were high, but most studies assessed the accuracy of algorithms rather than the effect of complete software systems on clinical workflows [9]. Similarly, oral mucosal AI may be useful for screening and referral while remaining inappropriate for definitive diagnosis . Future studies should therefore evaluate AI-assisted clinicians, not merely isolated algorithms.

Human-AI interaction is another determinant of benefit. A model can improve clinician performance when it supplies complementary information, but it can also introduce automation bias if clinicians accept incorrect outputs without sufficient scrutiny. Conversely, poorly calibrated alerts can produce alarm fatigue and reduce attention to true abnormalities. The appropriate interface should therefore communicate uncertainty, show the location of the finding, provide interpretable evidence where feasible, and preserve the clinician's ability to inspect the original image. This aligns with the broader principle that AI should support, rather than displace, professional judgment [11].

Explainability and transparency. Explainability is particularly important in oral diagnosis because clinicians need to understand why an image was flagged and whether the model's reasoning is plausible. Heatmaps, saliency maps, lesion contours, confidence estimates, and example-based explanations may assist interpretation, but visual explanations should not be mistaken for proof of causal reasoning. Transparent reporting of model architecture, preprocessing, training data, validation strategy, reference standard, thresholds, and failure cases is equally important. STARD-AI specifically emphasizes reporting dataset practices, AI index tests, evaluation procedures, bias, and generalizability.

Generalizability and external validation. External validation should become a standard expectation for clinically oriented dental AI. A model trained on images from one dental hospital should be tested across different institutions, geographic regions, equipment manufacturers, image protocols, age groups, disease prevalence levels, and relevant demographic characteristics. This is particularly important for oral cancer and mucosal lesion screening, where disease prevalence and visual phenotypes may differ across populations. The 2026 periodontal review similarly identified limited data diversity and scarce external testing as central barriers . Without external validation, a reported accuracy estimate should be viewed as conditional on the development environment rather than a universal property of the algorithm.

Data quality, annotation, and leakage. AI performance is strongly dependent on the quality of labels. Multiple clinicians may disagree on lesion boundaries or diagnostic categories, and consensus procedures can reduce but not eliminate disagreement. In segmentation tasks, small differences in annotations can substantially affect overlap metrics. Data leakage is another major concern: if multiple images from the same patient appear in both training and test sets, apparent performance may be inflated because the model can learn patient-specific features. Patient-level separation and transparent reporting of data partitioning are therefore essential. The methodological limitations reported across recent dental reviews make these safeguards more than technical details; they directly affect the credibility of clinical claims .

Multimodal AI as a future direction. Oral diagnosis rarely depends on a single image. Dentists combine symptoms, medical history, medications, risk factors, clinical examination, periodontal measurements, radiographs, photographs, and pathology. Multimodal AI could integrate these sources to generate patient-specific decision support. However, multimodal systems also increase complexity and the potential for hidden bias. The system must distinguish correlation from clinically meaningful relationships and remain robust when one data source is missing or unreliable. Future development should therefore prioritize clinically grounded multimodal models rather than simply increasing model size.

Generative AI and conversational systems. Generative AI may eventually assist with clinical documentation, structured extraction from dental records, patient-facing explanations, literature retrieval, and decision-support interfaces. Nevertheless, generative outputs introduce risks of hallucination, inconsistent reasoning, and inappropriate confidence. For diagnostic purposes, generative models should be connected to validated clinical data and constrained by explicit decision pathways rather than used as unrestricted diagnostic authorities. The principles of human oversight, transparency, safety, and accountability described by WHO remain applicable to emerging AI modalities [12].

Equity, privacy, and governance. AI could reduce diagnostic inequities if it expands access to specialist-level screening support in underserved settings, but it could also worsen disparities if models are trained predominantly on data from well-resourced populations. Bias may arise from underrepresentation, different disease prevalence, image-quality differences, language and documentation differences, or unequal access to digital infrastructure. Privacy is equally important because dental images and clinical records are health data. Governance should therefore address consent, secure data handling, auditability, model updates, cybersecurity, accountability, and mechanisms for reporting harm. WHO guidance explicitly identifies equity, autonomy, transparency, accountability, and sustainability as central requirements for responsible AI in health

Implications for clinical practice. The evidence supports several near-term uses with relatively clear clinical logic: second-reader assistance for radiographs, automated tooth identification and charting, caries and periapical lesion highlighting, periodontal bone-loss assessment, and screening support for suspicious oral mucosal lesions. In these roles, AI can reduce repetitive work and help clinicians focus attention on ambiguous or high-risk findings. However, the final diagnosis should remain with a qualified clinician, especially when the output may trigger invasive procedures, cancer referral, or irreversible treatment. AI should be introduced with local validation, defined indications, user training, and continuous performance monitoring rather than treated as a one-time software purchase [13].

Implications for research. Future studies should move beyond retrospective algorithm development toward prospective, multicenter clinical evaluations. Diagnostic accuracy studies should use patient-level data separation, external validation, prespecified thresholds, clinically meaningful reference standards, calibration assessment, subgroup analysis, and transparent reporting. Comparative studies should test clinician performance with and without AI assistance, while randomized or pragmatic evaluations should examine workflow and patient outcomes where appropriate. STARD-AI provides a contemporary reporting framework for AI diagnostic accuracy research,while PRISMA 2020 supports transparent evidence synthesis [14].

Table 1. Evidence summary of major artificial-intelligence applications in oral diagnosis

Diagnostic domain	**Main input**	**AI task**	**Reported performance**	**Principal limitation**
Dental caries	Photographs; periapical, bitewing and panoramic radiographs; NILT; OCT	Classification, detection, segmentation	Reported accuracy varies by modality: approximately 68–99.2% across reviewed tasks	High heterogeneity; limited external validation; early lesions remain challenging
Oral cancer / OPMDs	Clinical photographs; histopathology; medical imaging	Classification, detection, segmentation, prognosis	Deep-learning studies reported accuracy 85–100%; oral-mucosal AI accuracy 74–100%	Dataset and lesion-spectrum bias; low proportion of studies at low risk of bias
Oral mucosal lesions	Clinical photographs	Classification, detection, segmentation	AI accuracy 74–100%; clinician accuracy without AI 61–98%	Clinical benefit over specialist assessment remains uncertain
Periapical radiolucencies	Periapical, panoramic and CBCT radiographs	Detection / classification	Pooled sensitivity 0.94; pooled specificity 0.96	Only a minority contributed to meta-analysis; most studies evaluated models rather than deployed software
Periodontitis	Panoramic, intraoral and CBCT imaging; photographs; clinical data	Classification, staging, screening, prediction	Panoramic classification accuracy 81–99%; staging 64–91%; photograph AUROC 0.93	Single-center datasets and generalizability concerns
Tooth identification	Dental and panoramic radiographs	Detection, segmentation, numbering	Accuracy 81.8–99%; pooled sensitivity 89%, specificity 99%, AUC 0.96	Performance can vary with missing teeth, restorations, mixed dentition and image quality

Orthodontic diagnosis	Cephalometric and other digital dental images	Landmark detection, diagnosis, treatment planning	Broad evidence of efficiency and potential accuracy gains; no single pooled accuracy estimate is appropriate	AI has not consistently surpassed experts; human supervision remains essential
Across dental specialties	Multimodal dental imaging and clinical data	Multiple diagnostic tasks	Low-bias systematic reviews: accuracy 82–95%; pooled sensitivity 0.85 and specificity 0.93	Marked heterogeneity limits interpretation of a single global performance estimate

Strengths and limitations of this review. A strength is the emphasis on recent systematic reviews and meta-analyses rather than isolated proof-of-concept studies. The review also avoids combining incompatible diagnostic tasks and does not invent patient-level data, sample sizes, or statistical estimates. A limitation is that this is a structured narrative review rather than a prospectively registered systematic review with a formal PRISMA flow diagram and de novo quantitative meta-analysis. Targeted searches can miss eligible studies, and the rapidly changing AI literature means that evidence continues to evolve. In addition, the small number of references required for this manuscript necessarily means that the review is selective rather than exhaustive. These limitations should be considered when interpreting the synthesis [15].

Conclusion

Artificial intelligence has demonstrated substantial potential to augment the diagnosis of oral diseases, particularly through image-based detection, classification, segmentation, and automated anatomical identification. The most mature evidence concerns dental caries, oral cancer and potentially malignant lesions, periapical radiolucencies, periodontal disease, and tooth identification, with multiple systematic reviews reporting promising diagnostic performance . Nevertheless, the evidence remains constrained by methodological heterogeneity, limited external validation, single-center datasets, variable reference standards, and uncertainty regarding real-world clinical benefit.

The appropriate near-term model is therefore clinician-supervised decision support rather than autonomous diagnosis. Responsible implementation requires representative multicenter data, patient-level separation, external validation, transparent reporting, explainability, calibration, subgroup evaluation, privacy protection, and continuous monitoring. Prospective studies should determine whether AI actually improves diagnostic decisions, referral quality, efficiency, and patient outcomes rather than merely increasing algorithmic accuracy. Ethical governance and human accountability should remain central to adoption. If these requirements are addressed, AI can become a valuable component of digital dentistry by strengthening diagnostic consistency, supporting earlier recognition of important disease, and enabling more efficient use of clinical expertise without replacing the clinician's responsibility for patient care.

References

[1] F. Schwendicke, W. Samek, and J. Krois, “Artificial Intelligence in Dentistry: Chances and Challenges,” J. Dent. Res., vol. 99, no. 7, pp. 769–774, 2020, doi: 10.1177/0022034520915714.

[2] S. B. Khanagar, A. Al-ehaideb, P. C. Maganur, S. Vishwanathaiah, S. B. Patil, H. A. Baeshen, S. C. Sarode, and S. Bhandi, "Developments, application, and performance of artificial intelligence in dentistry – A systematic review," J. Dent. Sci., vol. 16, no. 1, pp. 508–522, 2021, doi: 10.1016/j.jds.2020.06.019.
[3] S. Patil, S. Albogami, J. Hosmani, S. Mujoo, M. A. Kamil, M. A. Mansour, H. N. Abdul, S. Bhandi, and S. S. S. J. Ahmed, "Artificial Intelligence in the Diagnosis of Oral Diseases: Applications and Pitfalls," Diagnostics, vol. 12, no. 5, Art. no. 1029, 2022, doi: 10.3390/diagnostics12051029.
[4] H. Mohammad-Rahimi, S. R. Motamedian, M. H. Rohban, J. Krois, S. E. Uribe, E. Mahmoudinia, et al., "Deep learning for caries detection: A systematic review," J. Dent., vol. 122, Art. no. 104115, 2022, doi: 10.1016/j.jdent.2022.104115.
[5] S. H. Kolarkodi and K. Z. Alotaibi, "Artificial Intelligence in Diagnosis of Oral Diseases: A Systematic Review," J. Contemp. Dent. Pract., vol. 24, no. 1, pp. 61–68, 2023, doi: 10.5005/jp-journals-10024-3465.
[6] K. Warin and S. Suebnukarn, "Deep learning in oral cancer—A systematic review," BMC Oral Health, vol. 24, Art. no. 212, 2024, doi: 10.1186/s12903-024-03993-5.
[7] R. Rokhshad, H. Mohammad-Rahimi, J. B. Price, R. Shoorgashti, Z. Abbasiparashkouh, M. Esmaeili, B. Sarfaraz, A. Rokhshad, S. R. Motamedian, P. Soltani, and F. Schwendicke, "Artificial intelligence for classification and detection of oral mucosa lesions on photographs: a systematic review and meta-analysis," Clin. Oral Investig., vol. 28, no. 1, Art. no. 88, 2024, doi: 10.1007/s00784-023-05475-4.
[8] S. Sadr, R. Rokhshad, Y. Daghighi, M. Golkar, F. T. Kheybari, F. Gorjinejad, A. M. Kojori, P. Rahimirad, P. Shobeiri, M. Mahdian, and H. Mohammad-Rahimi, "Deep learning for tooth identification and numbering on dental radiography: a systematic review and meta-analysis," Dentomaxillofac. Radiol., vol. 53, no. 1, pp. 5–21, 2024, doi: 10.1093/dmfr/twad001.
[9] U. Pul and F. Schwendicke, "Artificial intelligence for detecting periapical radiolucencies: A systematic review and meta-analysis," J. Dent., vol. 147, Art. no. 105104, 2024, doi: 10.1016/j.jdent.2024.105104.
[10] A. Tichy, N. Werner, H. Dujic, C. Wetzel, V. Pitchika, C. V. Bumm, M. Folwaczny, and F. Schwendicke, "Artificial Intelligence in Periodontology: A Systematic Review," J. Periodontal Res., 2026, doi: 10.1111/jre.70107.
[11] R. S. Gracea, N. Winderickx, M. Vanheers, J. Hendrickx, F. Preda, S. Shujaat, M. Cadenas de Llano-Pérula, and R. Jacobs, "Artificial intelligence for orthodontic diagnosis and treatment planning: A scoping review," J. Dent., vol. 152, Art. no. 105442, 2025, doi: 10.1016/j.jdent.2024.105442.
[12] A. Saikia, T. Kvist, A. Fawzy, and R. Anthonappa, "Artificial Intelligence in dentistry: an overview of systematic reviews and meta-analysis," Evid.-Based Dent., vol. 26, Art. no. 180, 2025, doi: 10.1038/s41432-025-01190-z.
[13] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," BMJ, vol. 372, Art. no. n71, 2021, doi: 10.1136/bmj.n71.
[14] V. Sounderajah, A. Guni, X. Liu, G. S. Collins, A. Karthikesalingam, S. R. Markar, et al., "The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence," Nat. Med., vol. 31, pp. 3283–3289, 2025, doi: 10.1038/s41591-025-03953-8.
[15] World Health Organization, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva, Switzerland: World Health Organization, 2021, doi/ISBN: 978-92-4-002920-0.