

2026

The Leeuwenhoek paradox in dental artificial intelligence: a narrative review of artificial intelligence-detected subclinical periapical changes and human-labeled false positives

Etyene Schnurr

Department of Basic Science Institute of Health Federal Fluminense University, Nova Friburgo, Brazil,
etyene.schnurr@swiss-biohealth.com

Mario Parra

Biohealth Dental Spain, Benidorm, Spain

David M. Kim

Department of Oral Medicine, Infection and Immunity, Division of Periodontology, Harvard School of Dental Medicine, Boston, MA, USA

Follow this and additional works at: <https://jds.ads.org.tw/journal>

Recommended Citation

Schnurr, Etyene; Parra, Mario; and Kim, David M. (2026) "The Leeuwenhoek paradox in dental artificial intelligence: a narrative review of artificial intelligence-detected subclinical periapical changes and human-labeled false positives," *Journal of Dental Sciences*: Vol. 21: Iss. 3, Article 6.
Available at: <https://jds.ads.org.tw/journal/vol21/iss3/6>

This Review Article is brought to you for free and open access by Journal of Dental Sciences. It has been accepted for inclusion in Journal of Dental Sciences by an authorized editor of Journal of Dental Sciences. For more information, please contact cpchiang@ntu.edu.tw.

Available online at <https://jds.ads.org.tw/journal/>

Digital Commons

journal homepage: <https://jds.ads.org.tw/journal/>

Review Article

The Leeuwenhoek paradox in dental artificial intelligence: A narrative review of artificial intelligence-detected subclinical periapical changes and human-labeled false positives

Etyene Schnurr ^{a,b*}, Mario Parra ^c, David M. Kim ^d^a Department of Basic Science Institute of Health Federal Fluminense University, Nova Friburgo, Brazil^b Swiss Dental Solutions, Kreuzlingen, Switzerland^c Biohealth Dental Spain, Benidorm, Spain^d Department of Oral Medicine, Infection and Immunity, Division of Periodontology, Harvard School of Dental Medicine, Boston, MA, USA

Received 24 March 2026; accepted 15 April 2026

Available online 1 July 2026

KEYWORDS

Artificial intelligence;
Deep learning;
Diagnostic imaging;
Machine learning;
Periapical
periodontitis

Abstract This narrative review explores the “Leeuwenhoek Paradox,” a metaphorical and conceptual challenge that arises when artificial intelligence (AI) detects subtle periapical radiological changes that are not visible to human observers. Drawing on evidence from studies using two-dimensional and cone-beam computed tomography (CBCT) imaging, this review critically examined the limitations of human-centered reference standards and recurrent discrepancies in AI-based periapical detection. The literature shows that AI systems frequently detect more periapical alterations than human experts. These findings may reflect early bone changes, as indicated by CBCT-based analyses of bone density and longitudinal observations. The paradox arises because AI identifies subclinical radiographic changes (AI-SRCs) that fall below the threshold of human visual perception, leading to their classification as false positives by humans owing to the limitations of human vision. Framed within the Leeuwenhoek Paradox, this article highlighted the constraints of equating diagnostic truth exclusively with human visual perception. Finally, the review outlined conceptual pathways for validating AI-SRCs, including longitudinal clinical studies, standardized visualization tools such as heatmaps, and the definition of new diagnostic thresholds, with the aim of supporting the safe integration of AI into dental diagnostics while mitigating the risks of overdiagnosis.

* Corresponding author. Swiss Dental Solutions, Konstanzerstrasse 11, Kreuzlingen, 8280, Switzerland.
E-mail address: etyene.schnurr@swiss-biohealth.com (E. Schnurr).

Introduction

Apical periodontitis (AP) is a biologically driven inflammatory disease of periapical tissues, typically resulting from microbial infection of the root canal system, of which radiographic radiolucency represents one possible manifestation. AP prevalence in Europe is estimated between 34% and 61%.^{1,2} In root-canal-treated (RCT) teeth, radiographic evidence of chronic persistent AP has been observed in 30–65% of cases.¹ However, conventional periapical radiographs and panoramic images are inherently limited by their two-dimensional (2D) nature, which reduces sensitivity due to anatomical superimpositions and observer variability. Cone-beam computed tomography (CBCT) is the clinical reference standard for detecting periapical lesions, providing higher diagnostic accuracy and more reliable confirmation of apical pathologies.^{3–5}

Human interpretation of periapical radiography is constrained by limitations in both sensitivity and interobserver reliability. Bohay reported a mean sensitivity of 65% and specificity of 78% under ideal viewing conditions,⁶ underlining the diagnostic challenges associated with radiographic interpretation, even among trained observers.

Artificial intelligence (AI) refers to machines capable of performing tasks that typically require human cognition.^{7–11} Machine learning (ML), a subfield of AI, and deep learning (DL), based on multilayered neural networks, enable the detection of complex patterns in imaging data.^{8,9} Convolutional neural networks (CNNs), in particular, have demonstrated the ability to detect periapical lesions and other conditions with sensitivity comparable to or exceeding that of human experts.^{7–11} AI systems can also assist with tasks such as locating the apical foramen and improving working length determination on radiographs.⁸

A recurring limitation of DL is limited transparency. Explainable artificial intelligence (XAI) has emerged to address this issue by providing human-interpretable outputs (e.g., heatmaps and saliency maps).⁹

Radiology is currently the primary AI application area in dentistry.^{8–14} Across both 2D and three-dimensional (3D) imaging modalities, studies report promising diagnostic performance.^{2–5,15–22} Representative region of interest (ROI)- or tooth-level studies on periapical and panoramic images reported sensitivities ranging from 65% to 93% and specificities above 90%, while CBCT-based approaches frequently achieve or exceed 90% sensitivity for automated detection, segmentation, and classification of periapical lesions.^{2–4,15,17–29} However, these performance ranges reflect substantial methodological heterogeneity and should be interpreted descriptively rather than comparatively.

Although AI systems matched or exceeded clinician performance in a recent systematic review, most studies

were retrospective, single-center, relied on human visual annotations as references, and showed considerable variation in methodology, complicating direct comparison and limiting generalizability.² A common challenge is the high frequency of false positives. AI systems may identify radiographic anomalies that human experts do not recognize, leading to their classification as algorithmic errors.^{4,5,16–19} This discrepancy reflects differences in information processing: while conventional assessment relies on visual interpretation, AI analyzes voxel- or pixel-level intensity patterns,²⁰ enabling detection of subtle density fluctuations that fall below human perception thresholds.

This article proposes that a subset of these AI-positive/human-negative findings may represent AI-detected subclinical radiographic changes (AI-SRCs), defined as subtle structural or density variations detectable by AI but imperceptible through conventional radiographic interpretation at a given time point. This concept is introduced as a hypothesis-generating, heuristic framework rather than a validated clinical entity. This phenomenon is conceptualized as the Leeuwenhoek Paradox, drawing an analogy to early microscopic observations that preceded widespread scientific acceptance.³⁰ In this context, AI may be identifying AI-SRCs, potentially challenging prevailing diagnostic categories.

If AI consistently detects periapical alterations around RCT teeth deemed “successful” by conventional radiographic criteria, the definition of success in endodontic treatment may require reconsideration. However, such implications remain speculative. Concurrently, there is a risk of overdiagnosis or overtreatment if every AI detection is prematurely classified as pathological.^{2,16,18,31} Similar concerns regarding safety, bias, and real-world validation have been highlighted across medicine.^{25,32} Although AI can enhance periapical diagnosis by improving sensitivity and reducing observer variability, it remains a complementary tool rather than a replacement for clinical judgment. Larger, prospective, and standardized studies are necessary to confirm its robustness in real-world practice. Quantitative risk indices and ground thresholds in longitudinal outcomes are also required.^{1,7,10,18,33,34}

Given the emergence of the Leeuwenhoek Paradox, the need for a revised diagnostic framework in periapical imaging warrants careful consideration. Integrating explainable visualization approaches (XAI) and refined diagnostic thresholds could enable this framework to overcome the limits of the true-/false-positive dichotomy.^{6,32,35} Addressing this diagnostic paradox is important for the safe and effective integration of AI into clinical dental diagnostics.^{36–38} Accordingly, this study aimed to (i) review current evidence on AI detection of periapical lesions, (ii) critically analyze the limitations of human-based reference standards that label early AI findings as false-positives, and

(iii) propose a conceptual and methodological roadmap centered on AI-SRCs for future validation through longitudinal research, explainable visualization, and revised diagnostic thresholds.

Materials and methods

Study design

This study was conducted as a narrative review combined with a conceptual analysis to integrate current heterogeneous evidence on AI in dental radiology for the detection or segmentation of periapical pathology and to formulate the construct of AI-SRC within the Leeuwenhoek Paradox. The design was explicitly chosen for hypothesis generation and theoretical integration rather than exhaustive evidence synthesis or quantitative meta-analysis.

Data sources and search approach

PubMed/MEDLINE and Google Scholar were searched with no predefined start date, and searches were last updated in October 2025. Additional relevant studies were identified through backward and forward citation snowballing. These databases were selected to balance biomedical indexing (PubMed/MEDLINE) with broader interdisciplinary and methodological coverage (Google Scholar), particularly relevant for AI and imaging research that may not yet be fully indexed in traditional medical databases. The literature exploration drew on a combination of controlled vocabulary and free-text terms related to AI and dental imaging, such as ML, DL, panoramic radiography, CBCT, CNN, XAI, diagnostic accuracy, periapical lesion, and AP. Terminology and search expressions were refined iteratively as familiarity with the field increased. As this was not a systematic review, we did not preregister a protocol, claim exhaustive coverage, or produce a PRISMA flow diagram. Instead, search procedures are described to enhance transparency of the evidence base informing the conceptual work.

Eligibility and study selection

Eligibility criteria were applied flexibly to prioritize conceptual relevance rather than enforce rigid inclusion thresholds:

- i) Studies addressing AI-based detection or segmentation of periapical lesions in dental radiology (periapical, panoramic, or CBCT imaging)
- ii) Empirical studies, methodological papers, or reviews (including systematic reviews, meta-analyses, bibliometric analyses, diagnostic accuracy studies, randomized or prospective evaluations, and narrative conceptual reviews) relevant to the research questions
- iii) Peer-reviewed publications in English
- iv) Outcomes related to diagnostic detection, segmentation, clinical decision support, or endodontic treatment outcome prediction

Opinion pieces without substantive methodological content, non-imaging studies, non-peer-reviewed conference abstracts, duplicate reports, and studies without extractable diagnostic or segmentation outcomes were excluded. Study selection followed a pragmatic iterative process consistent with the narrative review methodology. Titles and abstracts were screened to remove irrelevant records, followed by a full-text assessment for conceptual relevance to AI-based periapical diagnosis and the emerging AI-SRC construct. Duplicate independent screening, formal inter-reviewer agreement analyses, and hierarchical evidence weighting were not performed, in line with the hypothesis-generating nature of the work.

Data extraction and qualitative appraisal

For each included article, we extracted: study type; imaging modality (2D periapical, 2D panoramic, CBCT); AI task (detection, segmentation, classification, outcome prediction); dataset characteristics (size, source, heterogeneity); reference standard; performance metrics (e.g., sensitivity, specificity, Dice similarity coefficient [DSC]); reported limitations (e.g., domain shift, class imbalance, generalizability); and use of explainability or visualization methods (e.g., heatmaps, saliency maps, 3D overlays). Given the narrative and conceptual objective, we did not perform formal tool-based risk-of-bias scoring or a quantitative meta-analysis. Instead, we conducted a qualitative methodological appraisal focusing on features known to affect AI validity in imaging (sampling strategies, annotation practices, reference standard construction, validation design, and explainability). Items from consensus guidance and checklists for dental AI research and imaging AI reports were used as sensitizing analytical lenses rather than formal scoring instruments.^{7–9,12,23} Particular attention was paid to guidance for diagnostic test accuracy and prediction models, including emerging frameworks such as Standards for Reporting of Diagnostic Accuracy Studies-AI, Transparent Reporting of a multi-variable prediction model for Individual Prognosis Or Diagnosis-AI, and PROBAST-AI.^{39–41} Formal evidence grades were not assigned.

Synthesis of findings

Findings were integrated via thematic narrative synthesis across imaging modality, AI task, and validation design. Rather than pooling effect sizes, we described recurrent patterns in the diagnostic behavior of AI systems, including:

- i) Tendency toward elevated sensitivity with lower specificity, with frequent classification of AI detections as false positives
- ii) Reference standards that position the human observer as the primary arbiter of “truth”
- iii) Emerging reports of AI-detected subclinical signals that appear before human-visible radiographic change

Where multiple studies presented comparable metrics, performance ranges were summarized for context only.

These summaries do not indicate equivalence or validate comparative results.

Conceptual framework development (AI-SRC)

The AI-SRC construct was developed through an abductive, concept-driven process. Patterns of early AI-positive/human-negative findings observed across imaging modalities and AI tasks were iteratively examined alongside the following:

- i) Established knowledge of the perceptual and cognitive limits of human readers
- ii) Methodological insights from AI-reporting guidance and checklists for imaging AI^{7–9,11,12}

Through iterative comparison and synthesis, we specified the defining operational characteristics of AI-SRC (AI-positive, human-negative at time *t*), candidate visualization tools (e.g., Grad-CAM heatmaps, 3D overlays),¹² and a proposed validation roadmap emphasizing prospective and longitudinal study designs that link early AI signals to subsequent clinical or radiographic outcomes. The framework is intended as a hypothesis-generating conceptual model, not as a validated biological or diagnostic entity.

Results

Study selection

After removal of duplicates, 63 unique articles on AI in dental imaging were screened at the title/abstract level for relevance to periapical pathology. Of these, 57 peer-reviewed publications met the eligibility guidance and were included in the main evidence synthesis as primary analytical studies reporting diagnostic detection, segmentation, classification, or outcome-prediction performance for periapical lesions. Six additional papers were identified as contextual or methodological references (e.g., framework or guideline papers, systematic or narrative reviews without extractable diagnostic or segmentation outcomes specific to periapical pathology). These informed contextual interpretations but were not included among the 57 primary analytical studies.

Diagnostic performance in periapical lesion detection

Across imaging modalities and AI tasks, reported diagnostic performance varied substantially depending on dataset characteristics, reference standards, and evaluation design. The following results are therefore presented descriptively, without implying equivalence across study types, and should not be interpreted as direct evidence of clinical superiority.

2D periapical and panoramic radiography

In 2D periapical/panoramic radiography at the tooth level (presence/absence of lesion per tooth) or ROI-level

(restricted to periapical ROI), CNN-based detectors often achieve sensitivities of ~ 0.72 – 0.93 , with specificities of ≥ 0.90 at the tooth/ROI-level.^{3,17,23–26,35,42–47} Pauwels et al. (ex vivo) reported an area under the curve (AUC) of ~ 0.86 – 0.93 with sensitivities of ~ 0.79 – 0.87 and specificities of 0.88 – 0.98 versus human observers (AUC ~ 0.75).⁴⁸

CBCT-based detection and segmentation

In CBCT, studies reported strong performance for automated lesion localization/segmentation and tooth/arch contextual modeling, approaching or matching expert-level readers in internal testing and sometimes surpassing less-experienced clinicians.^{15,27–29,36–38}

CBCT: periapical detection and segmentation

Screened CBCT studies suggested that pipelines integrating anatomical context (tooth and arch identification) with lesion detection and segmentation capabilities achieve high sensitivity values (often ≥ 0.90) and moderate-to-high specificity.^{15,27–29,36–38,49–51} Segmentation-focused studies using both CBCT and 2D radiographs reported DSCs within moderate ranges, reflecting morphological variability and labeling inconsistencies.^{3,17,23–26,35,42–47} Architectural approaches vary (U-Net-based models, spatial-configuration networks, and hybrid frameworks with transfer learning), with better performance when models incorporated multi-view context and calibrated thresholds.^{15,27–29,36–38,49–51}

Periapical/panoramic radiography: detection and reader comparisons

In 2D imaging, CNN-based detectors showed competitive accuracy but appeared more sensitive to exposure/protocol variability and case-mix shifts than CBCT.^{3,17,23–26,35,42–47} Comparative studies indicated that AI support can improve the sensitivity of junior clinicians while maintaining acceptable specificity.^{3,42} In some studies, findings initially classified as “false positives” on 2D radiographs were reclassified as true positives after CBCT evaluation, indicating that some features were not visible on baseline 2D imaging.^{3,5,17,35}

Observation of AI-positive/human-negative findings

Across studies, AI detections not recognized by clinicians at baseline were commonly labeled as false positives when human-eye-based reference standards were used.^{2,3,15,17,18,21,23–26,42} This pattern was particularly evident when single-timepoint visual adjudication was used rather than longitudinal or multimodal confirmation. Several authors highlighted this limitation.^{2,33,39}

Reference standards and adjudication

Human visual assessment (single or consensus) served as the reference standard in most studies. Few studies used CBCT as an adjudicator for 2D findings, and only a minority incorporated prospective/longitudinal confirmation.^{2,5,15,18,27,28}

Studies that benchmarked 2D AI findings against CBCT often demonstrated that some detections initially labeled as false positives were reclassified as true positives when evaluated with 3D imaging.^{5,18,35}

Explainability and visualization

XAI methods are increasingly incorporated in dental AI studies. Class-activation and saliency maps (such as Grad-CAM overlays) frequently highlighted apical trabecular regions or periapical rims, even in cases that were considered negative by human observers. This may indicate that AI models do not flag random image regions but rather anatomically plausible zones associated with apical lesions.^{3,24,25,29,36}

Limited literature concerning reader-blinded experiments exists, although preliminary evidence suggests that saliency overlays can enhance clinician confidence and facilitate triage by focusing attention on subtle ROIs.^{15,22,27} However, reporting of explainability methods is highly heterogeneous. Therefore, standardized visualization protocols and sanity checks remain necessary to minimize the risk of misleading saliency outputs.^{39,41}

External validation, generalizability, and domain shift

Performance typically attenuates during external validation due to scanner/protocol shifts, altered case mix, and differences in annotation policy.^{15,22,28,29,52} However, Fu et al. observed robustness across three external data sets with high performance.²⁷ Multicenter training, intensity harmonization of images, and domain-adaptation strategies may improve robustness, and external drops should be reported transparently with confidence intervals and calibration metrics.^{14,15,17,20,22,27–29,33,34,52–55}

Quantitative ranges (descriptive context only)

In papers reporting sensitivity for periapical lesions and other measurements, sensitivity was in the moderate-to-high range for 2D radiography (0.72–0.93) and was frequently ≥ 0.90 for 3D CBCT. Specificity was high (0.88–0.97) for both 2D and 3D CBCT (0.85–0.95), depending on the dataset and lesion definition.^{2–4,15,17,19,21,27,28,42} Szabó et al. reported a sensitivity of 0.86 and specificity of 0.92 on panoramic datasets, whereas Issa et al. observed a sensitivity of 0.83 and specificity of 0.95 in a 2D retrospective analysis.^{35,42} In 3D CBCT, for segmentation tasks, DSC values ranged from 0.65 to 0.78, measuring the spatial overlap between segmentation and ground-truth results.¹⁵ These moderate DSCs for CBCT lesion segmentation reflected annotation noise, class imbalance, and morphological heterogeneity.^{15,22,36–38} In contrast, accuracy across different teeth and segmentation of alveolar bone is consistently high.¹⁵ Cui et al. obtained an average DSC score of 92.54 for teeth and 93.8 for bone.¹⁵ Fu et al.²⁷ achieved a DSC of approximately 0.74 for periapical lesions in CBCT; Dot et al.²² reported a DSC of 0.70 across multicenter datasets. Systematic reviews and meta-analyses report competitive diagnostic accuracy of dental AI relative

to clinicians, although heterogeneity and risk of bias remain significant.^{2,10,18,20,53,54}

Observations consistent with AI-SRC (hypothesis-generating)

Three recurrent observations related to AI-SRC were identified across the reviewed studies.

- i) Consistent AI-positive/human-negative findings across modalities when adjudication is purely visual and cross-sectional.^{2,3,15,17,18,23–26,42,49}
- ii) Co-localization of the saliency maps with periapical trabecular regions in cases deemed negative by human observers.^{27,28,36,56}
- iii) Reduction in the number of false positives when CBCT is applied for 2D radiographs or when temporal follow-up is applied, consistent with early biological change below the threshold of human visibility.^{4,17,19}

These observations are reported to motivate conceptual discussion and do not constitute biological or clinical validation.

Risk of overdiagnosis or overtreatment and need for longitudinal evidence

Several studies emphasized that early AI detection may be clinically relevant but require adequately powered prospective/longitudinal studies to quantify progression from AI-SRC to clinically meaningful endpoints. Without calibrated thresholds and outcome-linked risk guidance, the risk of overdiagnosis and unnecessary treatment remains.^{2,15,18,27,28}

Discussion

This narrative and conceptual review examined current evidence on AI applications in dental radiology and introduced the concept of AI-SRCs within the framework of the Leeuwenhoek Paradox. Similar challenges where AI detects signals below human visual perception have been observed in broader medical and dental contexts.⁵⁷ Through the review of 63 publications, the main patterns were highly sensitive algorithms, variable specificity, dependence on reference standards based on human visual perception, and the recurrence of AI-positive/human-negative findings.

Labeling such AI findings as false positives assumes that human visual perception represents the diagnostic reference standard.^{2,33,39} However, this assumption may be methodologically limited rather than biologically incorrect, particularly in the periapical region where early density alterations can be subtle and affected by projection geometry, noise, and observer variability.^{3,5,15,17,21,23–28,42,49–51,58} At the same time, non-biological explanations must be considered, including annotation noise, class imbalance, threshold calibration favoring sensitivity, domain shift, and overfitting to local textural patterns. Interpreting some AI-positive/human-negative findings as AI-SRCs, therefore, represents one possible explanatory

hypothesis, rather than definitive evidence of early pathology.^{2,8–10,18,34,39,55,59}

If AI-SRCs are commonly detected around RCT teeth deemed radiographically successful, it suggests that defining success solely as the absence of visible periapical radiolucency may be insufficient.^{28,60} However, this implication remains speculative and should be interpreted as a long-term research consideration rather than an immediate clinical recommendation. A stability-oriented conceptual framework, focusing on the persistence, resolution, or progression of subclinical changes over time without clinical deterioration, may provide a useful direction for future research.^{33,39} This perspective emphasizes cautious monitoring while explicitly acknowledging the risks of overdiagnosis and overtreatment.⁶¹ Additional clinical context from endodontic literature underscores the importance of systemic and host factors when interpreting periapical findings over time.¹

Most existing studies rely on cross-sectional clinician adjudication (2D/3D CBCT) and rarely incorporate longitudinal confirmation, structured symptom tracking, or composite endpoints.^{2,33,34,39,55,62} Evidence derived from CBCT adjudication of 2D findings represents a stronger reference standard than single-timepoint visual assessment, yet it remains indirect without outcome-based validation. In studies using CBCT as an adjudicator, a subset of 2D false positives was reclassified as true positives, suggesting the presence of radiographic changes below the detection threshold of 2D imaging.^{5,17,18,63} Nevertheless, CBCT-based confirmation alone does not establish biological relevance or clinical impact. Future standards should integrate multimodal imaging, predefined follow-up intervals (e.g., 6–24 months), and patient-centered outcomes to establish the significance of AI detections.^{39,57,62}

Validation of the proposed AI-SRC concept should therefore be framed as a research roadmap, rather than a diagnostic directive, and may rest on four complementary pillars:

- i. Prospective longitudinal cohorts: enroll RCT-teeth with baseline AI analysis (CBCT); record AI-SRC index; follow-up at 6, 12, and 24 months to assess radiographic progression and symptoms (or interventions); and perform calibrations and predefine decision thresholds.^{2,18,33,34,55}
- ii. Standardization of AI outputs: pre-specify sanity checks and saliency methods such as heatmaps and overlays; quantify co-localization; perform reader-blinded studies to test whether visualization affects decisions, time, or confidence; and design standardized visualization protocols.^{3,15,22,24,25,27,29,36}
- iii. External validation and domain shift mitigation: evaluate performance across different scanners, protocols, and centers and report performance degradation (AUC, specificity, sensitivity) when moving from internal to external validation, as well as calibration metrics.^{15,27–29,33,34,55,57,62}
- iv. Clinical utility: beyond merely reporting metrics (AUC, sensitivity, DSC, or specificity), studies must demonstrate whether AI provides clinically meaningful benefits.^{39,41}

Studies should assess the net benefit using tools such as decision-curve analysis, which evaluates whether AI-guided surveillance or intervention at different risk thresholds truly improves decision-making compared with usual practice. Research must incorporate patient-centered outcomes such as symptom onset, exacerbation, need for retreatment or tooth extraction, quality of life, and consumption of clinical resources. Calibration and clinical reclassification analyses should be performed, when possible, to determine whether AI meaningfully improves risk stratification.^{28,59–61}

As a narrative review and conceptual analysis, this study does not include formal quality appraisal tools, structured risk-of-bias scoring, exhaustive search strategies, or a quantitative meta-analysis. Instead, it relies on purposive expert judgment and iterative synthesis, which, while suitable for hypothesis generation, introduces selection bias and limits generalizability. Accordingly, framing AI-SRC as a diagnostic construct extrapolates beyond limited empirical validation and integrates evidence thematically without rigid frameworks, potentially limiting generalizability, overstating evidence, and introducing inconsistencies. Future empirical studies are required to test, refine, or refute the proposed framework.

Reclassifying a subset of AI-positive/human-negative detections as potential AI-SRCs may offer a coherent conceptual explanation for recurrent saliency-guided findings reported across studies, without implying biological certainty or immediate clinical relevance. Framed within the Leeuwenhoek Paradox, this perspective highlights the temporal gap between detection and validation and underscores the need to decouple human visual perception from absolute diagnostic truth. Validation of this hypothesis requires longitudinal designs, standardized explainability, robust external validation, and outcome-based thresholds before any clinical implications can be drawn.^{2,18,33,34,39,55}

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Acknowledgments

The authors have no acknowledgments to declare.

References

1. Segura-Egea JJ, Martín-González J, Castellanos-Cosano L. Endodontic medicine: connections between apical periodontitis and systemic diseases. *Int Endod J* 2015;48:933–51.
2. Sadr S, Mohammad-Rahimi H, Motamedian SR, et al. Deep learning for detection of periapical radiolucent lesions: a systematic review and meta-analysis of diagnostic test accuracy. *J Endod* 2023;49:248–61.
3. Ekert T, Krois J, Meinhold L, et al. Deep learning for the radiographic detection of apical lesions. *J Endod* 2019;45: 917–22.
4. Allihaihi M, Koller G, Mannocci F. Diagnostic accuracy of an artificial intelligence-based platform in detecting periapical

- radiolucencies on cone-beam computed tomography scans of molars. *J Dent* 2025;160:105854.
5. Allilhaibi M, Koller G, Mannocci F. The detection of apical radiolucencies in periapical radiographs: a comparison between an artificial intelligence platform and expert endodontists with CBCT serving as the diagnostic benchmark. *Int Endod J* 2025; 58:1146–57.
 6. Bohay RN. The sensitivity, specificity, and reliability of radiographic periapical diagnosis of posterior teeth. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2000;89:639–42.
 7. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res* 2020;99:769–74.
 8. Ahmed N, Abbasi MS, Zuberi F, et al. Artificial intelligence techniques: analysis, application, and outcome in dentistry—a systematic review. *Biomed Res Int* 2021;2021:9751564.
 9. Thurzo A, Urbanová W, Novák B, et al. Where is the artificial intelligence applied in dentistry? Systematic review and literature analysis. *Healthcare (Basel)* 2022;10:1269.
 10. Xie B, Xu D, Zou XQ, Lu MJ, Peng XL, Wen XJ. Artificial intelligence in dentistry: a bibliometric analysis from 2000 to 2023. *J Dent Sci* 2024;19:1722–33.
 11. Katne T, Kanaparthi A, Gotoor S, Muppilala S, Devaraju S, Gantala R. Artificial intelligence: demystifying dentistry – the future and beyond. *Int J Contemporary Med Surg Radiol* 2019; 4:D6–9.5.
 12. Heo MS, Kim JE, Hwang JJ, et al. Artificial intelligence in oral and maxillofacial radiology: what is currently possible? *Dentomaxillofacial Radiol* 2021;50:20200375.
 13. Bonny T, Al Nassan W, Obaideen K, Al Mallahi MN, Mohammad Y, El-Damanhoury HM. Contemporary role and applications of artificial intelligence in dentistry. *F1000Res* 2023; 12:1179.
 14. Rahim A, Khatoon R, Khan TA, et al. Artificial intelligence-powered dentistry: probing the potential, challenges, and ethicality of artificial intelligence in dentistry. *Digit Health* 2024;10:20552076241291345.
 15. Cui Z, Fang Y, Mei L, et al. A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. *Nat Commun* 2022;13:2096.
 16. Pul U, Tichy A, Pitchika V, Schwendicke F. Impact of artificial intelligence assistance on diagnosing periapical radiolucencies: a randomized controlled trial. *J Dent* 2025;160:105868.
 17. Kazimierczak W, Wajer R, Wajer A, et al. Periapical lesions in panoramic radiography and CBCT imaging—assessment of AI's diagnostic accuracy. *J Clin Med* 2024;13:2709.
 18. Pul U, Schwendicke F. Artificial intelligence for detecting periapical radiolucencies: a systematic review and meta-analysis. *J Dent* 2024;147:105104.
 19. Allilhaibi M, Koller G, Mannocci F. Diagnostic accuracy of a commercial AI-based platform in evaluating endodontic treatment outcomes on periapical radiographs using CBCT as the reference standard. *J Endod* 2025;51:898–908.e8.
 20. Ghaffari M, Zhu Y, Shrestha A. A review of advancements of artificial intelligence in dentistry. *Dent Rev* 2024;4:100081.
 21. Güneç HG, Ürkmez ES, Danacı A, Dilmaç E, Onay HH, Aydın KC. Comparison of artificial intelligence vs. junior dentists' diagnostic performance based on caries and periapical infection detection on panoramic images. *Quant Imag Med Surg* 2023;13: 7494–503.
 22. Dot G, Chaurasia A, Dubois G, et al. DentalSegmentator: robust open-source deep learning-based CT and CBCT image segmentation. *J Dent* 2024;147:105130.
 23. Endres MG, Hillen F, Salloumis M, et al. Development of a deep learning algorithm for periapical disease detection in dental radiographs. *Diagnostics* 2020;10:430.
 24. Chen H, Li H, Zhao J, Wang Y. Dental disease detection on periapical radiographs based on deep convolutional neural networks. *Int J Comput Assist Radiol Surg* 2021;16:649–61.
 25. Li CW, Lin SY, Chou HS, et al. Detection of dental apical lesions using CNNs on periapical radiograph. *Sensors (Basel)* 2021;21: 7049.
 26. İçöz D, Terzioğlu H, Özel MA, Karakurt R. Evaluation of an artificial intelligence system for the diagnosis of apical periodontitis on digital panoramic images. *Niger J Clin Pract* 2023; 26:1085–90.
 27. Fu WT, Zhu QK, Li N, et al. Clinically oriented CBCT periapical lesion evaluation via 3D CNN algorithm. *J Dent Res* 2024;103: 5–12.
 28. Setzer FC, Li J, Khan AA. The use of artificial intelligence in endodontics. *J Dent Res* 2024;103:853–62.
 29. Ezhov M, Gusarev M, Golitsyna M, et al. Clinically applicable artificial intelligence system for dental diagnosis with CBCT. *Sci Rep* 2021;11:15006.
 30. Lane N. The unseen world: reflections on Leeuwenhoek (1677) 'concerning little animals. *Philos Trans R Soc Lond B Biol Sci* 2015;370:20140344.
 31. Parinitha MS, Doddawad VG, Kalgeri SH, Gowda SS, Patil S. Impact of artificial intelligence in endodontics: precision, predictions, and prospects. *J Med Signals Sens* 2024;14:25.
 32. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
 33. Putra RH, Doi C, Yoda N, Astuti ER, Sasaki K. Current applications and development of artificial intelligence for digital dental radiography. *Dentomaxillofacial Radiol* 2022;51: 20210197.
 34. Turosz N, Chęcińska K, Chęciński M, Brzozowska A, Nowak Z, Sikora M. Applications of artificial intelligence in the analysis of dental panoramic radiographs: an overview of systematic reviews. *Dentomaxillofacial Radiol* 2023;52:20230284.
 35. Issa J, Jaber M, Rifai I, Mozdziak P, Kempisty B, Dyszkiewicz-Konwińska M. Diagnostic test accuracy of artificial intelligence in detecting periapical periodontitis on two-dimensional radiographs: a retrospective study and literature review. *Medicina (Kaunas)* 2023;59:768.
 36. Zheng Z, Yan H, Setzer FC, Shi KJ, Mupparapu M, Li J. Anatomically constrained deep learning for automating dental CBCT segmentation and lesion detection. *IEEE Trans Autom Sci Eng* 2021;18:603–14.
 37. Ver Berne J, Saadi SB, Politis C, Jacobs R. A deep learning approach for radiological detection and classification of radicular cysts and periapical granulomas. *J Dent* 2023;135: 104581.
 38. Calazans MAA, Ferreira FABS, Alcoforado MLMG, Santos AA, Pontual ADA, Madeiro F. Automatic classification system for periapical lesions in cone-beam computed tomography. *Sensors (Basel)* 2022;22:6481.
 39. Sounderajah V, Ashrafian H, Golub RM, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open* 2021;11:e047709.
 40. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021; 11:e048008.
 41. Juarez-Orozco LE, Ruijsink B, Yeung MW, Benjamins JW, van der Harst P. Machine learning in practice—evaluation of clinical value, guidelines. In: Asselbergs FW, Denaxas S, Oberski DL, Moore JH, eds. *Clinical applications of artificial intelligence in real-world data*. Springer International Publishing, 2023:247–61.
 42. Szabó V, Orhan K, Dobó-Nagy C, et al. Deep learning-based periapical lesion detection on panoramic radiographs. *Diagnostics* 2025;15:510.
 43. Kwon O, Yong TH, Kang SR, et al. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a

- deep convolution neural network. *Dentomaxillofacial Radiol* 2020;49:20200185.
44. Arijji Y, Yanashita Y, Kutsuna S, et al. Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2019;128:424–30.
 45. Ba-Hattab R, Barhom N, Osman SA, et al. Detection of periapical lesions on panoramic radiographs using deep learning. *Appl Sci* 2023;13:1516.
 46. Bayrakdar IS, Orhan K, Çelik Ö, et al. A U-net approach to apical lesion segmentation on panoramic radiographs. *Biomed Res Int* 2022;2022:7035367.
 47. Ngoc VT, Viet DH, Anh LK, et al. Periapical lesion diagnosis support system based on x-ray images using machine learning technique. *World J Dent* 2021;12:189–93.
 48. Pauwels R, Brasil DM, Yamasaki MC, et al. Artificial intelligence for detection of periapical lesions on intraoral radiographs: comparison between convolutional neural networks and human observers. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2021;131:610–6.
 49. Kirnbauer B, Hadzic A, Jakse N, Bischof H, Stern D. Automatic detection of periapical osteolytic lesions on cone-beam computed tomography using deep convolutional neuronal networks. *J Endod* 2022;48:1434–40.
 50. Orhan K, Bayrakdar İŞ, Ezhov M, Kravtsov A, Özyürek T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *Int Endod J* 2020;53:680–9.
 51. Akalin F, Özkan Y. Deep learning-based hybrid scenario for classification of periapical lesions in cone-beam computed tomography. *Symmetry* 2025;17:1392.
 52. Krois J, Garcia Cantu A, Chaurasia A, et al. Generalizability of deep learning models for dental image analysis. *Sci Rep* 2021;11:6102.
 53. Karobari MI, Adil AH, Basheer SN, et al. Evaluation of the diagnostic and prognostic accuracy of artificial intelligence in endodontic dentistry: a comprehensive review of literature. *Comput Math Methods Med* 2023;2023:7049360.
 54. Alharbi SS, Alhasson HF. Exploring the applications of artificial intelligence in dental image detection: a systematic review. *Diagnostics* 2024;14:2442.
 55. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: a systematic review. *Dentomaxillofacial Radiol* 2020;49:20190107.
 56. Rajpurkar P, Irvin J, Zhu K, et al. Chexnet: radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv* 2017:05225. arXiv:1711.
 57. Vodanović M, Subašić M, Milošević D, Savić Pavićin I. Artificial intelligence in medicine and dentistry. *Acta Stomatol Croat* 2023;57:70–84.
 58. Boztuna M, Firincioglugulari M, Akkaya N, Orhan K. Segmentation of periapical lesions with automatic deep learning on panoramic radiographs: an artificial intelligence study. *BMC Oral Health* 2024;24:1332.
 59. Demir K, Sokmen O, Karabey Aksakalli I, Torenek-Agirman K. Comprehensive insights into artificial intelligence for dental lesion detection: a systematic review. *Diagnostics* 2024;14:2768.
 60. Marwaha J. Artificial intelligence in conservative dentistry and endodontics: a game-changer. *J Conserv Dent Endod* 2023;26:514–8.
 61. Sivari E, Senirkentli GB, Bostanci E, Guzel MS, Acici K, Asuroglu T. Deep learning in diagnosis of dental anomalies and diseases: a systematic review. *Diagnostics* 2023;13:2512.
 62. Iruvuri AG, Miryala G, Khan Y, et al. Revolutionizing dental imaging: a comprehensive study on the integration of artificial intelligence in dental and maxillofacial radiology. *Cureus* 2023;15:e50292.
 63. Chau KK, Zhu M, AlHadidi A, et al. A novel AI model for detecting periapical lesion on CBCT: CBCT-SAM. *J Dent* 2025;153:105526.