

UCLA

UCLA Previously Published Works

Title

Evaluation of artificial intelligence—based clinical decision support systems for caries and periodontal bone loss

Permalink

<https://escholarship.org/uc/item/2pn574dc>

Journal

The Journal of the American Dental Association

ISSN

0002-8177

Authors

Farooqi, Owais A

Fru, Gilbert A

Gong, Yan Ming

et al.

Publication Date

2025-12-01

DOI

10.1016/j.adaj.2025.11.005

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nd/4.0/>

Peer reviewed

Investigation

Evaluation of artificial intelligence–based clinical decision support systems for caries and periodontal bone loss

An external validation study

Owais A. Farooqi, DDS, MDS, MPH^a; Gilbert A. Fru, DDS, MPH^b; Yan Ming Gong, DDS^b; Landon E. Oswald, DDS^c Donald J. DeNucci, DDS, MS^d

^aDental Service, Veterans Affairs Greater Los Angeles Health Care System, Los Angeles, CA; ^bDepartment of Periodontics, Veterans Affairs Greater Los Angeles Health Care System, Los Angeles, CA; ^cDepartment of Prosthodontics, Veterans Affairs Greater Los Angeles Health Care System, Los Angeles, CA; ^dVeterans Affairs Greater Los Angeles Health Care System, Los Angeles, CA.

ABSTRACT

Background. Artificial intelligence (AI) applications in dentistry are advancing rapidly, yet external validation studies are scarce. This pilot study evaluated 2 US Food and Drug Administration–cleared AI-based decision support systems for detecting caries and clinically relevant periodontal bone loss ($\geq 33\%$ or ≥ 5 mm of attachment loss) using longitudinal clinical data to establish ground truth.

Methods. A retrospective analysis of 90 patients (mean [SD] age, 61.4 [13.2] years) records was conducted at the Veterans Affairs Greater Los Angeles Health Care System in Los Angeles, California. Each case included full-mouth radiographs and standardized clinical documentation with 6 through 12 months of follow-up. Ground truth was determined through a multiphase validation protocol involving calibrated examiners and adjudication. Analyses emphasized specificity and negative predictive value (NPV) to assess diagnostic concordance.

Results. For caries detection, vendor A achieved 78.68% concordance, 80.17% specificity, and 96.92% NPV; vendor B achieved 82.43%, 83.76%, and 97.36%, respectively. For periodontal bone loss, vendor A achieved 78.47% concordance, 94.23% specificity, and 80.64% NPV; vendor B achieved 73.89%, 93.55%, and 75.20%, respectively.

Conclusions. The results of this external validation pilot study showed AI tools can reliably exclude disease and reduce false-positive results. The use of retrospective longitudinal ground truth with standardized validation protocols provides a practical template for future multisite studies.

Practical Implications. AI-based decision support tools are most effective as adjunctive screening aids in dental practice. Their high specificity and NPV allow clinicians to trust negative findings, avoid unnecessary interventions, and maintain clinical oversight of positive findings requiring confirmation.

Key Words. Artificial intelligence; dentistry; caries; periodontal bone loss; radiographic interpretation; clinical decision support; external validation; screening value; specificity; negative predictive value.

JADA 2025;■(■):■-■

<https://doi.org/10.1016/j.adaj.2025.11.005>

Artificial intelligence (AI) is rapidly transitioning from experimental promise to clinical application in dentistry. Investigators have reported its potential to support diagnosis, improve workflow efficiency, and reduce interoperator variability in clinical decision making.¹⁻⁵ Among the most studied areas are radiographic assessments for caries and periodontal bone loss, where machine learning and deep learning approaches have shown particular

Copyright © 2025
American Dental
Association. All rights are
reserved, including those
for text and data mining,
AI training, and similar
technologies.

promise.^{4,6,7} The earliest systems for proximal caries detection established feasibility more than a decade ago, laying the groundwork for the deep learning models that now dominate image analysis in dentistry.⁷⁻⁹

In AI research, the concept of ground truth refers to the benchmark or reference standard against which system outputs are judged. In dentistry, this is often limited to expert interpretation of radiographs at a single point in time. In our study, however, ground truth was defined by means of longitudinal clinical documentation and follow-up care, providing a more reliable reflection of patient outcomes and enhancing the clinical relevance of AI validation.

Despite the progress in AI-driven dentistry, a question remains: Where does AI provide dependable value in day-to-day practice? In general, in dental populations in the United States approximately 20% through 30% of adults have untreated caries, with most tooth surfaces remaining healthy.¹⁰ In such contexts, the ability to accurately rule out disease is often more valuable than identifying every potential lesion. High specificity and negative predictive value (NPV) reduce unnecessary interventions, support conservative treatment planning, and preserve patient trust.^{11,12}

In several systematic reviews of dental imaging, investigators emphasized the need to report clinically relevant operating points that align with real treatment decisions, such as whether to treat, monitor, or reassure.¹²⁻¹⁵ Although the number studies on AI systems for caries detection are increasing¹⁶ and studies on periodontal bone loss detection are encouraging,^{4,17} the need for rigorous external validation in diverse patient populations before widespread adoption is highlighted consistently in the literature.

Evidence from broader medical imaging underscores both the potential benefits and the challenges. Deep learning models can match or exceed expert performance in certain tasks, but their safe implementation depends on external validation, bias assessment, and integration into clinical workflows.^{18,19} Within dentistry, reviews have outlined opportunities while emphasizing limitations, particularly the lack of standardized evaluation methods and educational frameworks to ensure responsible adoption.^{20,21}

The clinical usefulness of AI must also be understood within the larger context of health-system implementation. High-performance medicine depends not only on accurate algorithms, but also on equity, generalizability, and transparent governance.^{4,22,23} Results from studies in other areas of medicine revealed that AI systems can unintentionally encode or amplify disparities, reinforcing the need for transparent evaluation frameworks and ongoing monitoring of deployed tools.^{24,25}

Against this backdrop, we designed our study as a pilot of external validation of 2 commercial AI systems for detecting caries and periodontal bone loss on full-mouth intraoral radiographs. By means of emphasizing pragmatic performance metrics (specificity, NPV, concordance with longitudinally established clinical outcomes) we addressed the question: When AI indicates no disease, how reliable is that judgment in everyday patient care?

METHODS

Study design and setting

Our retrospective external validation study was conducted at the Veterans Affairs (VA) Greater Los Angeles Health Care System in Los Angeles, California, in accordance with institutional policies for non-human research. The VA Greater Los Angeles Health Care System Institutional Review Board reviewed our study and determined it to be exempt from full review with a waiver of informed consent.

The data set comprised 90 consecutive patient records corresponding to comprehensive oral examinations (including complete periodontal charting and full-mouth radiographs) performed from August 1, 2022, through June 30, 2023. The mean (SD) patient age was 61.4 (13.2) years (range, 35-84 years), including 84 men (93%) and 6 women (7%).

Because our investigation served as a pilot external validation study intended to test feasibility and reproducibility, a formal power analysis was not performed. The sample size of 90 (45 patients per vendor) was consistent with previous external validation studies of comparable design, which generally range from 40 through 100 patients and provide sufficient precision to detect clinically meaningful differences in specificity and NPV ranging from 5 to 10 percentage points.

ABBREVIATION KEY

- AI:** Artificial intelligence.
- ALT-ID:** Alternate identifier.
- CEJ:** Cementoenamel junction.
- VA:** Veterans Affairs.

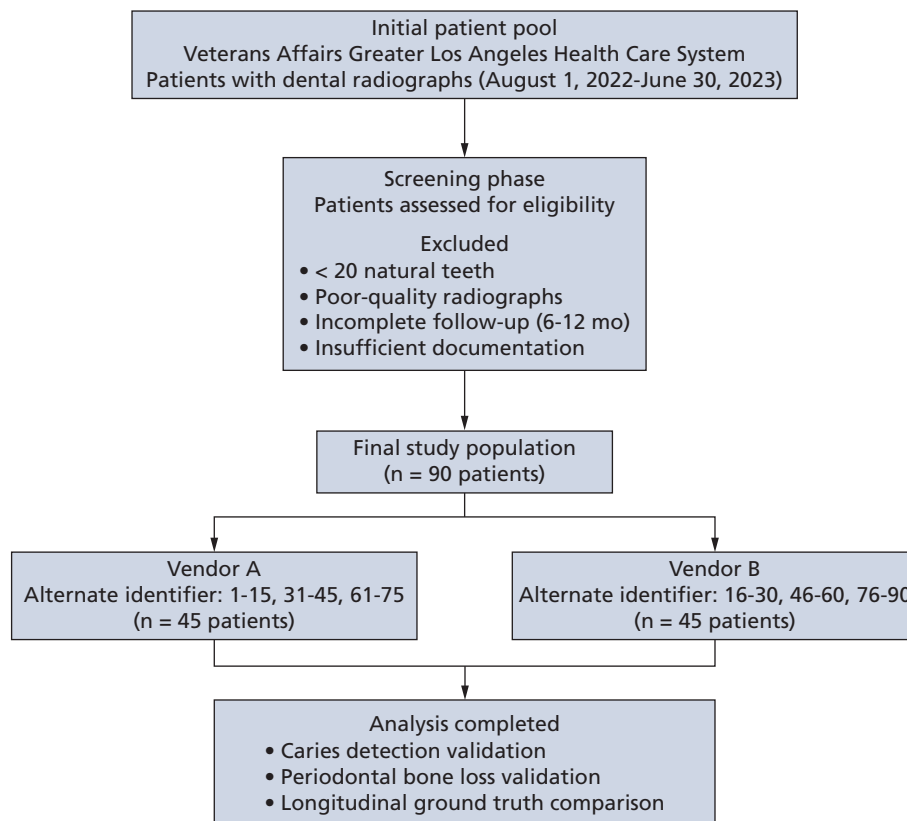


Figure 1. Study flow diagram of patient selection and vendor stratification.

Records were included if the corresponding patient's electronic health record met the following 3 inclusion criteria: (1) presence of at least 20 natural teeth, (2) availability of diagnostic-quality full-mouth intraoral radiographs, and (3) complete initial and follow-up documentation within 6 through 12 months of the index examination. Patient records were excluded if radiographs were of insufficient quality (eg, blur, overlapping contacts, or exposure errors), if follow-up visits were absent, or if documentation was incomplete.

To ensure reproducibility, data were restricted to cases from the periodontics postgraduate year 1 teaching clinics, where standardized diagnostic protocols and attending dentist cosignatures ensured clinical consistency. Each record included baseline radiographic interpretation, periodontal charting, treatment notes, and follow-up documentation confirming either continued health or treatment outcomes.

Internal or development data from either vendor were not available to investigators. All analyses were conducted independently using VA clinical data to ensure that findings reflected a true external validation rather than internal model testing (Figure 1).

Ground truth establishment

Ground truth, the benchmark for validating AI output, was derived from longitudinal VA dental records incorporating both baseline and follow-up documentation. Three trained residents (2 from the advanced prosthodontics program [L.E.O., Y. M.G.], 1 from the advanced periodontics program [G.A.F.], all graduates of US-accredited doctor of dental surgery and doctor of dental medicine programs) performed the structured data abstraction using a standardized written codebook.

Before data collection, all reviewers completed calibration sessions to standardize diagnostic criteria and interpretation thresholds. Discrepancies identified during abstraction were adjudicated by means of a board-certified periodontist (O.A.F.) with more than 16 years of clinical experience, who served as an independent reviewer. Postcalibration interrater reliability was assessed on a 10-record validation subset, yielding weighted κ of 0.83 (caries) and 0.87 (bone loss), reflecting strong agreement among evaluators.

This multiphase adjudication process established consensus on diagnostic outcomes and ensured uniform criteria for determining the presence or absence of caries and bone loss. This longitudinally anchored process ensured that validation reflected the true course of patient care.

Definitions and clinical criteria

Enamel-only caries lesions were excluded to reduce interpretive ambiguity in radiographic identification and to focus analysis on caries lesions with clear radiographic demarcation into dentin, while recognizing that enamel-level changes remain clinically relevant for preventive management.

Periodontal bone loss was quantified as the cementoenamel junction (CEJ) to alveolar crest distance divided by root length multiplied by 100%. Clinically relevant bone loss was defined as 33% or more per tooth.

To enhance clinical validity and align with the American Academy of Periodontology's 2017 classification of periodontal and peri-implant diseases,¹³ the following dual criterion was applied: cases with 5 mm or more of clinical attachment loss documented in the periodontal chart were considered corroborating evidence of advanced attachment loss consistent with radiographic findings. Although 5 mm or more clinical attachment loss does not always correspond directly to 5 mm or more or 33% or more bone loss, its inclusion ensured concordance between radiographic and clinical severity measures, as endorsed in the American Academy of Periodontology's classification framework as stage III. To confirm diagnostic accuracy, attending dentists cosigned all clinical findings previously.

A reduced periodontium was considered stable in the absence of active inflammation, distinguishing it from active periodontal disease.

AI systems and evaluation

Two US Food and Drug Administration–approved (510[k], class II²⁶) commercial AI systems were evaluated through collaboration with their respective vendors. Both systems processed full-mouth intraoral radiographs (bite-wing, periapical) and output binary detections indicating the presence or absence of dentinal caries and periodontal bone loss of 5 mm or more or 33% or more (extending to the middle one-third of the root or beyond) for each tooth.

Both systems used automated CEJ and alveolar crest landmarking for bone loss estimation; restoration type or margin detection was not analyzed separately.

Both systems used automated landmark detection algorithms incorporating CEJ identification, alveolar-crest mapping, and density-based differentiation between restored and unrestored tooth surfaces. Because system architectures remain proprietary, we focused on external performance validation rather than algorithmic design. Vendor participation was limited to data ingestion and output generation; all analyses and interpretations were performed independently by research team members.

Data management and coding

Data from all 90 patient records were consolidated into a single spreadsheet containing the following 4 workbooks: ground truth–caries, ground truth–bone loss, AI annotation–caries, and AI annotation–bone loss. For caries analysis, a blank cell denoted absence of dentinal caries and any marked entry indicated caries into dentin. For bone loss analysis, a blank cell denoted absence of bone loss 5 mm or more or 33% or more (extending to the middle one-third of the root or beyond) for each tooth, and any letter indicated presence.

Statistical analysis

Analyses prioritized performance metrics relevant to screening use, including specificity (correctly identifying healthy teeth), NPV (the likelihood that teeth classified as healthy by means of AI were confirmed healthy on follow-up), and concordance (overall agreement between AI detections and ground truth).

Sensitivity, precision (positive predictive value), and accuracy were also calculated for completeness, although the emphasis remained on metrics most relevant to rule-out use.

Comparisons between vendors were performed using χ^2 tests. Analyses were conducted at both the tooth and patient levels to assess robustness across units of analysis (Table).

Table. Performance of artificial intelligence systems for caries and periodontal bone loss detection.*

OUTCOME	VENDOR A, % (95% CI)	VENDOR B, % (95% CI)
Caries Detection		
Concordance	78.7 (76.54 to 80.69)	82.4 (80.38 to 84.31)
Specificity	80.2 (78.01 to 82.18)	83.8 (81.71 to 85.62)
NPV	96.9 (95.73 to 97.78)	97.4 (96.28 to 98.14)
Periodontal Bone Loss ($\geq 33\%$)		
Concordance	78.5 (76.28 to 80.52)	73.9 (71.56 to 76.09)
Specificity	94.2 (92.69 to 95.46)	93.6 (91.86 to 94.91)
NPV	80.6 (78.38 to 82.72)	75.2 (72.73 to 77.51)

* Performance metrics for 2 commercial artificial intelligence systems evaluated for caries detection and periodontal bone loss ($\geq 33\%$) assessment. Data include concordance with longitudinal ground truth, specificity, and negative predictive value (NPV) with 95% CIs calculated using Wilson score intervals.

Vendor assignment and comparison

Patients were evenly stratified between vendors using alternate identifiers (ALT-IDs). Vendor A included ALT-IDs from 1 through 15, from 31 through 45, and from 61 through 75 (45 patients), and vendor B included ALT-IDs from 16 through 30, from 46 through 60, and from 76 through 90 (45 patients). Identical analytic procedures were applied to each vendor's data set to enable balanced comparison of diagnostic performance.

RESULTS

Caries detection

Vendor A achieved 78.68% concordance (95% CI, 76.54% to 80.69%), 80.17% specificity (95% CI, 78.01% to 82.18%), and 96.92% NPV (95% CI, 95.73% to 97.78%). Vendor B achieved 82.43% concordance (95% CI, 80.38% to 84.31%), 83.76% specificity (95% CI, 81.71% to 85.62%), and 97.36% NPV (95% CI, 96.28% to 98.14%) (Figure 2).

Periodontal bone loss ($\geq 33\%$)

Vendor A achieved 78.47% concordance (95% CI, 76.28% to 80.52%), 94.23% specificity (95% CI, 92.69% to 95.46%), and 80.64% NPV (95% CI, 78.38% to 82.72%). Vendor B achieved 73.89% concordance (95% CI, 71.56% to 76.09%), 93.55% specificity (95% CI, 91.86% to 94.91%), and 75.20% NPV (95% CI, 72.73% to 77.51%) (Figure 3).

DISCUSSION

In our external validation study, we found that 2 US Food and Drug Administration–approved AI systems for detecting dentinal caries and clinically relevant periodontal bone loss achieved consistently high specificity and NPV. These findings confirmed that when such systems indicate the absence of disease, clinicians can have confidence in those results. Rather than serving as diagnostic replacements, these AI tools function most effectively as adjunctive screening aids that support conservative, evidence-based decision making in everyday practice.

Both systems performed strongly across independent VA data sets, which were entirely separate from their development data. This ensures that the findings represent a true external validation. Linking baseline radiographic findings with 6 through 12 months of follow-up documentation strengthened the ground truth reference and minimized interpretation bias that can occur when validation relies on single time-point annotations only. The use of standardized calibration, independent adjudication, and longitudinal chart review provided a clinically anchored benchmark for determining disease presence or stability.

Although both systems were evaluated within the same validation framework, this was not designed as a direct comparative analysis between the 2. Each AI system analyzed a distinct patient sample under identical inclusion criteria and analytic procedures. Our focus for this pilot study was the external validation of each system independently and an alternative approach to establishing

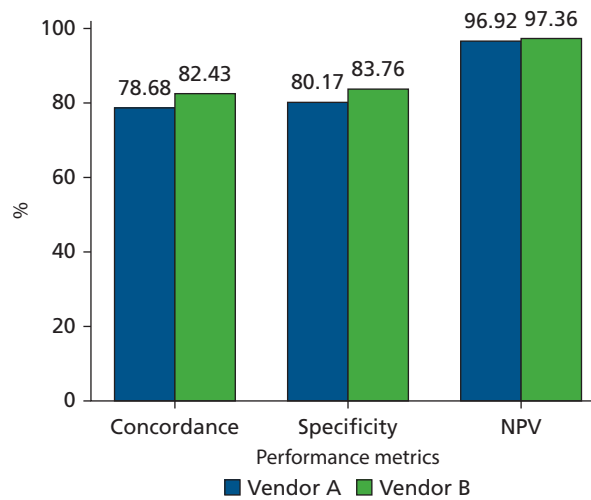


Figure 2. Caries detection performance. Bar chart comparing vendor A and vendor B across concordance, specificity, and negative predictive value (NPV). Both systems had high specificity and clinically meaningful NPV, supporting their role as adjunctive tools for screening and ruling out disease.

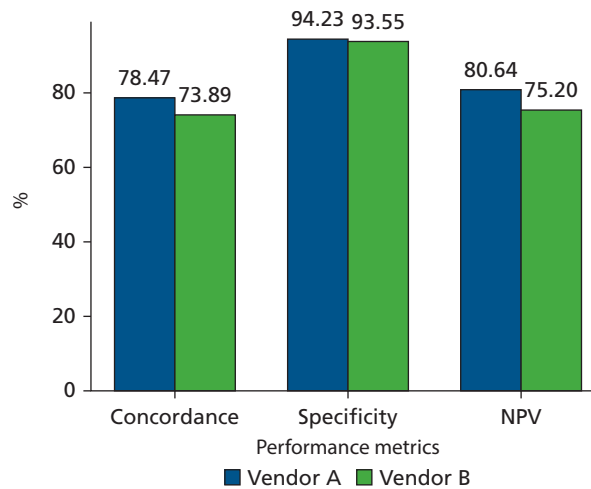


Figure 3. Periodontal bone loss detection performance. Bar chart comparing vendor A and vendor B across concordance, specificity, and negative predictive value (NPV). Both systems had high specificity and clinically meaningful NPV, supporting their role as adjunctive tools for screening and ruling out disease.

ground truth that extends beyond annotated radiographs. In this design, the reference standard was derived from chart-based data abstraction of comprehensive clinical examinations and prior radiographic assessments, each verified twice within the preceding 12 months by both the resident and the attending dentist.

The observed performance patterns align with AI's optimal role in dentistry; that is, to help clinicians reliably exclude disease. The high NPVs (97% for caries, $\approx$ 75%-81% for bone loss) indicate that teeth classified as healthy by means of AI were confirmed healthy at follow-up in nearly all cases. This rule-out reliability is especially valuable in dental practice, where most teeth are disease-free and the clinical priority often lies in avoiding unnecessary treatment. High specificity and NPV enable clinicians to reassure patients confidently and focus professional judgment on the few cases that require closer evaluation.

The nuanced distinction between active periodontal disease and stable reduced periodontium also helps explain the apparent performance differences between caries and bone loss detection. A tooth with historical bone loss but no signs of inflammation should not be categorized as diseased. AI systems that recognize such stable conditions as healthy are, in fact, performing correctly. This

interpretation reinforces the concept that AI validation must be context-specific and clinically informed, considering not only radiographic patterns but biological activity and disease trajectory as well.

By means of drawing on real-world VA data, we extended the evidence base beyond controlled development environments and toward the operational realities of clinical dentistry. The inclusion of adjudicated longitudinal ground truth rather than single time-point annotations directly addresses gaps identified in prior reviews in which researchers called for standardized, reproducible validation frameworks.^{1,12} Such methodology aligns with the broader movement toward high-performance medicine that balances computational precision with clinical responsibility, equity, and transparency.

These results are consistent with prior external validations in both dental and medical imaging domains, where specificity and NPV often exceed sensitivity at conservative thresholds. This pattern is desirable for screening applications because it minimizes false-positive results that can erode clinician trust or lead to overtreatment. In this context, AI's greatest contribution is to reinforce confidence in negative findings, enabling dentists to prioritize patients and allocate clinical attention more effectively.

Clinical and educational implications

Our findings emphasized that AI should be viewed as a collaborative partner in diagnostic reasoning. When properly calibrated, AI systems can enhance workflow efficiency and diagnostic consistency, particularly for screening radiographs in high-volume settings, such as Veterans Health Administration clinics. However, positive AI flags still require human verification, especially in cases of marginal or ambiguous findings. This balanced integration of using AI for triage and a clinician for confirmation optimizes both safety and efficiency.

For education and training, the growing availability of chairside AI tools offers an opportunity to improve dental students' and residents' interpretive consistency. Structured exposure to AI outputs and error modes (eg, misreads due to overlapping contacts or cervical burnout) can strengthen radiographic literacy and calibrate professional judgment.

Equity, governance, and generalizability

Although our study results confirmed AI reliability in an older, predominantly male veteran population, generalizability to broader dental settings should be approached cautiously. Algorithms can inadvertently amplify disparities if they perform unevenly across populations or imaging systems. Future work should include subgroup analyses, transparency in threshold reporting, and institutional oversight to ensure equitable application. Governance frameworks that assign responsibility for ongoing monitoring and error reporting will be essential as AI becomes embedded in clinical workflows.

Limitations

Our pilot study had several limitations. It was retrospective and single center in design, with a modest sample size that precluded detailed subgroup analyses. Ground truth was established from longitudinal documentation and adjudicated chart review rather than histologic or direct clinical reexamination. Although this approach mirrors real-world practice, it may not capture all aspects of disease activity. In addition, workflow integration and clinician-AI interaction were not evaluated prospectively.

Despite these limitations, the methodology achieved its primary objective: to externally validate 2 AI systems using independent data with rigorously standardized ground truth. The consistency of findings across vendors and conditions underscored the robustness of the validation framework.

Future directions

Building on this model, larger multisite prospective studies are warranted to confirm generalizability, assess temporal stability, and evaluate performance across diverse imaging systems and patient populations. Future research should also examine real-time workflow integration, clinician decision confidence, and patient outcomes. By means of incorporating automated monitoring tools and structured electronic health record data, continuous performance auditing could be achieved postdeployment to sustain transparency and fairness.

CONCLUSIONS

In an independent clinical cohort, 2 commercial AI systems had high specificity and NPV for caries and periodontal bone loss detection. These findings confirm that AI tools can safely assist clinicians by means of reliably excluding disease and minimizing false-positive results. When paired with professional oversight and standardized validation frameworks, such systems offer tangible value as screening adjuncts that enhance, rather than replace, clinical judgment. ■

DISCLOSURE

None of the authors reported any disclosures.

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used ChatGPT-5 (OpenAI) and Claude Sonnet 4 (Anthropic) for language refinement and data analysis. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Address correspondence to Dr. DeNucci, Veterans Affairs Greater Los Angeles Health Care System, 11301 Wilshire Blvd, Los Angeles, CA 90073, e-mail: donald.denucci@comcast.net

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US Government.

The authors thank the commercial partners who provided deidentified data outputs from their AI systems to support this independent validation project. Their collaboration made it possible to conduct the analyses presented. The study design, interpretation, and conclusions remain solely the responsibility of the authors.

ORCID Number. Donald J. DeNucci: <https://orcid.org/0009-0006-7496-8040>. For information regarding ORCID numbers, go to <http://orcid.org>.

- Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res*. 2020;99(7):769-774. doi:10.1177/0022034520915714
- Lee JH, Kim DH, Jeong SN, Choi SH. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontal Implant Sci*. 2018;48(2):114-123. doi:10.5051/jpis.2018.48.2.114
- Albano D, Galiano V, Basile M, et al. Artificial intelligence for radiographic imaging detection of caries lesions: a systematic review. *BMC Oral Health*. 2024;24:274. doi:10.1186/s12903-024-04046-7
- Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep*. 2019;9:8495. doi:10.1038/s41598-019-44839-3
- Park WJ, Park JB. History and application of artificial neural networks in dentistry. *Eur J Dent*. 2018;12(4):594-601. doi:10.4103/ejd.ejd_325_18
- Mohammad-Rahimi H, Motamedian SR, Pirayesh Z, et al. Deep learning in periodontology and oral implantology: a scoping review. *J Periodontol Res*. 2022;57(5):942-951. doi:10.1111/jre.13037
- Bayraktar Y, Ayan E. Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. *Clin Oral Investig*. 2022;26(1):623-632. doi:10.1007/s00784-021-04040-1
- Devito KL, de Souza Barbosa F, Felipe Filho WN. An artificial multilayer perceptron neural network for diagnosis of proximal dental caries. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2008;106(6):879-884. doi:10.1016/j.tripleo.2008.03.002
- Ekert T, Krois J, Meinhold L, et al. Deep learning for the radiographic detection of apical lesions. *J Endod*. 2019;45(7):917-922. doi:10.1016/j.joen.2019.03.016
- Centers for Disease Control and Prevention. *Oral Health Surveillance Report: Trends in Dental Caries and Sealants, Tooth Retention, and Edentulism, United States—1999–2004 to 2011–2016*. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2019. Accessed October 21, 2025. <https://www.cdc.gov/oral-health/media/pdfs/Oral-Health-Surveillance-Report-2019-h.pdf>
- Bader JD, Shugars DA, Bonito AJ. Systematic reviews of selected dental caries diagnostic and management methods. *J Dent Educ*. 2001;65(10):960-968.
- Arsiwala-Schepach LT, Chaurasia A, Müller A, Krois J, Schwendicke F. Machine learning in dentistry: a scoping review. *J Clin Med*. 2023;12(3):937. doi:10.3390/jcm12030937
- Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: framework and proposal of a new classification and case definition. *J Periodontol*. 2018;89(suppl 1):S159-S172. doi:10.1002/JPER.18-0006
- Casalegno F, Newton T, Daher R, et al. Caries detection with near-infrared transillumination using deep learning. *J Dent Res*. 2019;98(11):1227-1233. doi:10.1177/0022034519871884
- Shan T, Tay FR, Gu L. Application of artificial intelligence in dentistry. *J Dent Res*. 2021;100(3):232-244. doi:10.1177/0022034520969115
- Negi S, Mathur A, Tripathy S, et al. Artificial intelligence in dental caries diagnosis and detection: an umbrella review. *Clin Exp Dent Res*. 2024;10(4):e2532-e2548. doi:10.1002/cre2.70004
- Tariq A, Nakhi FB, Salah F, et al. Efficiency and accuracy of artificial intelligence in the radiographic detection of periodontal bone loss: a systematic review. *Imaging Sci Dent*. 2023;53(3):193-198. doi:10.5624/isd.20230092
- Haenssle HA, Fink C, Toberer F, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*. 2018;29(8):1836-1842. doi:10.1093/annonc/mdy166
- Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. doi:10.1016/j.media.2017.07.005
- Mörch CM, Atsu S, Cai W, et al. Artificial intelligence and ethics in dentistry: a scoping review. *J Dent Res*. 2021;100(13):1452-1460. doi:10.1177/00220345211013808
- Pitchika V, Büttner M, Schwendicke F. Artificial intelligence and personalized diagnostics in periodontology: a narrative review. *Periodontol*. 2020;2024;95(1):220-231. doi:10.1111/prd.12586
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
- Chen H, Zhang K, Lyu P, et al. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci Rep*. 2019;9:3840. doi:10.1038/s41598-019-40414-y
- Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:195. doi:10.1186/s12916-019-1426-2
- Overmeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi:10.1126/science.aax2342
- Class I and II device exemptions. US Food and Drug Administration. Accessed November 22, 2025. <https://www.fda.gov/medical-devices/classify-your-medical-device/class-i-and-class-ii-device-exemptions>