ACADEMIC JOURNAL OF HEALTH SCIENCES
MEDICINA BALEAR

Reseñas

Precisión diagnóstica de la inteligencia artificial en el diagnóstico de caries: revisión sistemática

Diagnostic accuracy of artificial intelligence in caries diagnosis: a systematic review

Selina Nunha[1], Sebastiana Arroyo-Bote*[1,2,3,4], Pere Riutord-Sbert[1,2], Catalina Bennasar-Verger[1]

1. Escuela Universitaria ADEMA, Palma de Mallorca, Islas Baleares, España
2. Instituto Universitario de Investigación en Ciencias de la Salud (IUNICS), Universitat de les Illes Balears (UIB), Palma de Mallorca, España
3. Universitat de Barcelona, Barcelona, España
4. Instituto de Investigación Biomédica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, España.

Recepción: 08-02-2026
Aceptación: 25-02-2026
Publicación: 27-03-2026

***Correspondencia:** Sebastiana Arroyo-Bote. s.arroyo@eua.edu.es

Resumen

Introducción: La caries dental es una de las enfermedades orales más prevalentes a nivel mundial. En los últimos años, la inteligencia artificial (IA), especialmente mediante redes neuronales convolucionales (CNN), se ha propuesto como herramienta para apoyar la detección y mejorar la precisión diagnóstica en odontología. El objetivo de este estudio fue evaluar la precisión diagnóstica de modelos de IA para la detección de caries a partir de imágenes diagnósticas.

Métodos: Se realizó una revisión sistemática con síntesis en PubMed, incluyendo artículos publicados entre 2020 y 2025. Se incluyeron estudios en humanos con texto completo disponible que aplicaran IA al diagnóstico de caries y reportaran métricas de precisión diagnóstica. La calidad metodológica se evaluó mediante AMSTAR-2 para revisiones sistemáticas y QUADAS-2 para estudios primarios de precisión diagnóstica.

Resultados: Se incluyeron 10 estudios. Las sensibilidades reportadas oscilaron entre 25% y 99,8% (predominando aproximadamente 44%–94%) y las especificidades entre 16,46% y 100% (mayoritariamente 77%–98%). Cinco estudios informaron valores de AUC entre 0,74 y 0,99. Las radiografías aleta de mordida (AMRX) fueron la modalidad más utilizada. Las arquitecturas más frecuentes incluyeron U-Net, ResNet y YOLO.

Conclusiones: La IA constituye una herramienta prometedora como complemento al diagnóstico de caries. Sin embargo, la evidencia disponible presenta heterogeneidad metodológica y limitada validación externa, por lo que se requieren protocolos estandarizados, validación multicéntrica y formación profesional para una implementación clínica segura y efectiva.

Palabras clave: IA. Caries dental. Diagnóstico. Radiografía. Aprendizaje profundo. Redes neuronales. Revisión sistemática.

Abstract

Background: Dental caries is among the most prevalent oral diseases worldwide. In recent years, artificial intelligence (AI), particularly convolutional neural networks (CNNs), has been proposed as a supportive tool to improve diagnostic accuracy for caries detection from diagnostic images.

Cite as: Nunha S, Arroyo-Bote S, Riutord-Sbert P, Bennasar-Verger C. Precisión diagnóstica de la inteligencia artificial en el diagnóstico de caries: revisión sistemática. Academic Journal of Health Sciences 2026; 41(2): 267-278. Available from: https://doi.org/10.65735/AJHS.2026.41.02.10

Academic Journal of Health Sciences 2026; 41(2): 267-278
https://doi.org/10.65735/AJHS.2026.41.02.10
www.ajhealthsciences.com

© 2026 The authors. This article is distributed under the terms of the **Creative Commons Attribution 4.0 International license (CC BY 4.0)**. Published with index on behalf of Academic journal of Health Sciences.

Methods: A systematic review with qualitative synthesis was conducted in PubMed, including studies published between 2020 and 2025. The last search was performed on May 30, 2025. Human studies with full-text access applying AI to caries diagnosis and reporting diagnostic accuracy metrics were included. Methodological quality was assessed using AMSTAR-2 for systematic reviews and QUADAS-2 for primary diagnostic accuracy studies.

Results: Ten studies were included. Reported sensitivities ranged from 25% to 99.8% (predominantly ~44%–94%) and specificities from 16.46% to 100% (mostly ~77%–98%). Five studies reported AUC values between 0.74 and 0.99. Bitewing radiographs were the most commonly used modality. Main architectures included U-Net, ResNet, and YOLO.

Conclusion: AI is a promising adjunct tool for caries diagnosis. Nevertheless, current evidence shows methodological heterogeneity and limited external validation; therefore, standardized protocols, multicenter validation, and professional training are required for safe clinical implementation.

Keywords: Artificial intelligence. Dental caries. Diagnosis. Radiography. Deep learning. Neural networks, Systematic review.

Introducción

La caries dental representa una de las enfermedades orales más prevalentes a nivel mundial y constituye un importante problema de salud pública debido a su alta frecuencia, impacto funcional y costes asociados. Se estima que aproximadamente 2 mil millones de adultos han presentado caries en dientes permanentes y alrededor de 520 millones de niños han padecido caries en dentición temporal[1].

En Europa, la situación epidemiológica es especialmente relevante. Según la Organización Mundial de la Salud (OMS), la Región Europea presenta la mayor prevalencia de caries en dientes permanentes a nivel mundial, afectando aproximadamente al 33,6 % de la población, lo que equivale a alrededor de 335 millones de casos en 2019[2]. Asimismo, la región registra una elevada proporción de pérdida dentaria, con una prevalencia estimada del 25,2 %, que afecta a unos 88 millones de personas mayores de 20 años[2]. En España, la Encuesta de Salud Oral 2020 muestra que la prevalencia de caries en adultos de 35–44 años se sitúa en torno al 92 %, con un índice CAOD medio cercano a 7,4, lo que evidencia la magnitud del problema y refuerza la necesidad de estrategias diagnósticas más precisas y accesibles[3].

La detección temprana, una adecuada higiene oral y la aplicación de intervenciones preventivas han demostrado ser eficaces para frenar la progresión de las lesiones cariosas, reduciendo la necesidad de tratamientos restauradores extensos y los costes asociados. Sin embargo, el diagnóstico de caries continúa siendo un reto clínico, ya que suele basarse en una combinación de inspección visual, sondaje táctil y, cuando es necesario, el apoyo de imágenes diagnósticas, principalmente radiografías intraorales y, en determinados contextos, fotografías clínicas[4,5].

En este escenario, la fotografía digital intraoral y la cámara intraoral han adquirido un papel complementario como herramientas de apoyo diagnóstico. Estos dispositivos permiten capturar imágenes de alta definición dentro de la cavidad oral y facilitan tanto la visualización como la comunicación de los hallazgos clínicos. Diversos estudios han señalado la viabilidad y utilidad de la fotografía digital como soporte al diagnóstico clínico, especialmente para la documentación y el seguimiento de lesiones[6].

Por otra parte, la IA (IA) se define como la capacidad de los sistemas computacionales para realizar tareas que tradicionalmente requieren inteligencia humana, incorporando técnicas como el aprendizaje automático y, más recientemente, el aprendizaje profundo (*deep learning*). En odontología, la IA se ha aplicado al diagnóstico, la planificación del tratamiento y la optimización de procesos clínicos. Se ha propuesto que los modelos basados en IA, al analizar grandes volúmenes de datos e imágenes, pueden reducir la variabilidad interobservador y la subjetividad inherente a la interpretación clínica, alcanzando niveles de rendimiento comparables a los de evaluadores expertos en determinados contextos[7,8].

En los últimos años se ha observado un crecimiento notable de la investigación sobre la integración de la IA en odontología, particularmente en el diagnóstico y la planificación terapéutica[9-12]. Este avance está estrechamente vinculado al desarrollo de las redes neuronales convolucionales (CNN), que han transformado el análisis automatizado de imágenes mediante técnicas de aprendizaje profundo[13]. Las CNN permiten aprender patrones complejos a partir de imágenes dentales y extraer automáticamente características relevantes, lo que puede mejorar la eficiencia diagnóstica y apoyar la toma de decisiones clínicas. En este contexto, la aplicación de la IA al diagnóstico de caries a partir de radiografías y fotografías intraorales se considera un campo de creciente interés y potencial impacto clínico[14].

Para que la aplicación clínica de la detección automática de caries mediante IA sea efectiva, es esencial identificar y numerar correctamente cada diente en

las imágenes diagnósticas, y posteriormente detectar y localizar las lesiones cariosas asociadas a cada uno. A partir de esta premisa, se formulan las siguientes hipótesis de investigación.

Hipótesis nula (H_0): La IA no mejora la precisión diagnóstica de la caries dental a partir de imágenes diagnósticas (radiografías intraorales y/o fotografías clínicas), en comparación con el diagnóstico clínico convencional realizado por odontólogos.

Hipótesis alternativa (H_1): La IA mejora la precisión diagnóstica de la caries dental a partir de imágenes diagnósticas (radiografías intraorales y/o fotografías clínicas), en comparación con el diagnóstico clínico convencional realizado por odontólogos.

Objetivo principal

Evaluar la precisión diagnóstica de los modelos de IA en la detección de caries dentales a partir de imágenes diagnósticas, mediante una revisión sistemática con síntesis cualitativa.

Objetivos secundarios

1. Analizar los valores de sensibilidad, especificidad y área bajo la curva (AUC) reportados en estudios que aplican modelos de IA para el diagnóstico de caries.
2. Identificar los tipos de imágenes diagnósticas más utilizados.
3. Describir las arquitecturas de redes neuronales más empleadas.
4. Comparar la precisión diagnóstica de los modelos de IA con la de evaluadores humanos (odontólogos o estudiantes, según el diseño del estudio).

Materiales y métodos

Tipo de estudio y pregunta PICO

Se realizó una revisión sistemática con síntesis cualitativa (sin metaanálisis propio) sobre la precisión diagnóstica de modelos de IA en la detección de caries dentales a partir de imágenes diagnósticas. La imposibilidad de realizar un metaanálisis cuantitativo se debió a una marcada heterogeneidad entre los estudios incluidos, que afectó a múltiples variables clave. Entre ellas se incluyen las diferencias en los estándares de referencia utilizados (consenso de expertos, criterios clínico-radiográficos o validación histológica), la categorización no uniforme de las lesiones cariosas (lesiones iniciales, moderadas o avanzadas), la diversidad de modalidades de imagen analizadas y la variabilidad en las métricas diagnósticas reportadas. Esta heterogeneidad metodológica limitó la comparabilidad directa entre los estudios y justificó la realización de una síntesis cualitativa.

Esta revisión sistemática no se registró prospectivamente en PROSPERO. Sin embargo, el protocolo de revisión se definió internamente antes de la selección de los estudios, incluyendo la pregunta PICO, los criterios de elegibilidad, la estrategia de búsqueda y los métodos de evaluación de la calidad. La falta de registro externo se reconoce como una limitación metodológica. Las futuras actualizaciones de este trabajo se beneficiarán del registro prospectivo del protocolo para mejorar la transparencia y la reproducibilidad.

La pregunta clínica se formuló siguiendo el modelo **PICO**:

- **P (Población):** pacientes con riesgo o presencia de caries dentales.
- **I (Intervención):** diagnóstico asistido por IA aplicado a imágenes diagnósticas.
- **C (Comparación):** diagnóstico convencional realizado por odontólogos (y/o evaluadores humanos según diseño del estudio).
- **O (Outcome/Resultado):** precisión diagnóstica medida mediante sensibilidad, especificidad y AUC.

Pregunta PICO

En pacientes con riesgo o presencia de caries dentales, ¿el uso de IA aplicada a imágenes diagnósticas mejora la precisión diagnóstica frente al diagnóstico convencional realizado por odontólogos?

Estrategia de búsqueda

La búsqueda bibliográfica se realizó en la base de datos PubMed entre abril y mayo de 2025, con fecha de última búsqueda el 30 de mayo de 2025. Se emplearon descriptores MeSH combinados con operadores booleanos mediante la siguiente estrategia:

[("Artificial Intelligence"[MeSH] OR "Diagnosis, Computer-Assisted"[MeSH]) AND "Dental Caries" [MeSH] AND "Diagnosis"[MeSH] AND ("Image Interpretation, Computer-Assisted"[MeSH] OR "Radiography, Dental"[MeSH])]

Se limitaron los resultados a artículos publicados entre 2020 y 2025, en inglés o español, con texto completo disponible.

La búsqueda bibliográfica se limitó a PubMed, que ofrece una cobertura completa de la investigación biomédica y odontológica, además de una indexación estandarizada mediante términos MeSH. Sin

embargo, restringir la búsqueda a una sola base de datos podría haber llevado a la omisión de estudios relevantes. Las futuras revisiones sistemáticas deberían considerar bases de datos adicionales como Scopus o Web of Science para aumentar la sensibilidad de la búsqueda y reducir el riesgo de sesgo de publicación.

Criterios de inclusión y exclusión

Criterios de inclusión:

- Artículos publicados entre 2020 y 2025.
- Estudios realizados en humanos.
- Texto completo disponible.
- Idioma: inglés o español.
- Aplicación explícita de modelos de IA al diagnóstico de caries.
- Reporte de métricas de precisión diagnóstica (sensibilidad, especificidad y/o AUC).
- Publicación en revistas indexadas (preferentemente JCR y/o SJR).

Criterios de exclusión:

- Estudios in vitro o en animales.
- Editoriales, cartas, revisiones narrativas u otros documentos sin metodología sistemática o sin datos pertinentes para el objetivo.
- Trabajos sin aplicación directa de IA al diagnóstico de caries.
- Estudios que no reportaran métricas de precisión diagnóstica (sensibilidad, especificidad o AUC).
- Duplicados o artículos con información incompleta o no pertinente.

La selección de estudios se resume en un diagrama de flujo PRISMA 2020.

Limitación por tipo de estudio

No se aplicó una restricción estricta por diseño metodológico. Se incluyeron revisiones sistemáticas y estudios primarios en humanos (observacionales, ensayos clínicos y estudios experimentales) siempre que cumplieran los criterios anteriores y aportaran datos útiles para responder a la pregunta PICO.

Proceso de selección de estudios

En la búsqueda inicial se identificaron 117 artículos. Tras aplicar filtros por idioma (inglés y español) y año de publicación (2020–2025), se obtuvieron 40 artículos. Posteriormente, se excluyeron aquellos sin texto completo disponible, quedando 29 artículos para lectura detallada.

La selección se realizó en dos fases:

1. Cribado de título y resumen: realizado por dos revisores de forma independiente (Selina Nunha y Dra. Sebastiana Arroyo Boté), aplicando los criterios de inclusión y exclusión.
2. Lectura a texto completo: los 29 artículos restantes se evaluaron íntegramente, y se excluyeron 19 por las siguientes razones principales:
    - 7 no aplicaban directamente modelos de IA al diagnóstico de caries dentales.
    - 5 correspondían a estudios in vitro o realizados en dientes extraídos.
    - 4 no reportaban métricas de precisión diagnóstica (sensibilidad, especificidad o AUC).
    - 3 eran revisiones narrativas sin metodología sistemática ni datos originales pertinentes.

Finalmente, 10 estudios cumplieron todos los criterios de inclusión y se incorporaron al análisis cualitativo. El flujo completo del proceso de selección se representa en la **Figura 1** (PRISMA).

Resolución de discrepancias:

Las discrepancias entre revisores se resolvieron por consenso;

Extracción y síntesis de datos

Los datos se extrajeron mediante un formulario prediseñado que incluyó: autor, año, revista y cuartil; diseño del estudio; tamaño muestral; tipo de imagen; tipo de modelo/arquitectura de IA; tipo de validación; métricas diagnósticas (sensibilidad, especificidad, AUC e intervalos de confianza cuando estuvieron disponibles) y comparación con evaluadores humanos.

Debido a la heterogeneidad en diseños, poblaciones, estándares de referencia y métricas reportadas, no fue posible realizar un metaanálisis cuantitativo. Por tanto, se realizó una síntesis narrativa cualitativa, presentando los resultados en tablas de evidencia para facilitar la comparación entre estudios.

Evaluación de la calidad metodológica

Para mejorar el rigor metodológico, la calidad de los estudios incluidos se evaluó con herramientas validadas[15]:

- Para las revisiones sistemáticas incluidas (n=5), se utilizó AMSTAR-2, interpretando los resultados mediante confianza global (sin puntuación numérica)[15,16].

- Para los estudios primarios de precisión diagnóstica (n=5), se empleó QUADAS-2, evaluando cuatro dominios: selección de pacientes, prueba índice (modelo de IA), estándar de referencia y flujo/temporalidad[17]. En el caso de estudios con componente educativo, la interpretación se realizó considerando principalmente su aplicabilidad, dado que no siempre corresponden a estudios diagnósticos clínicos en pacientes.

Los resultados de esta evaluación se presentan en tablas específicas en el apartado de Resultados (Tablas AMSTAR-2 y QUADAS-2).

Consideraciones éticas y registro

La presente revisión sistemática no requirió aprobación de comité de ética institucional al no implicar intervención clínica, ni uso de datos personales identificables. No se registró protocolo para la presente revisión.

La selección de estudios se resume en un diagrama de flujo PRISMA 2020.

Resultados

Estudios incluidos y características generales

Se incluyeron 10 estudios en la síntesis cualitativa, que abarcaron distintos diseños metodológicos: cinco revisiones sistemáticas, dos estudios clínicos u observacionales de precisión diagnóstica, un ensayo clínico aleatorizado, un estudio experimental diagnóstico y un estudio experimental de carácter educativo. En cuanto al cuartil de publicación, ocho trabajos se publicaron en revistas Q1, uno en Q2 y uno en Q3[5,18-26].

Los tamaños muestrales mostraron una amplia variabilidad. Szabó et al.[20] analizaron 323 dientes; Mertens et al.[23] incluyeron 22 dentistas y 140 AMRX (≈2.800 superficies); Zhang et al.[5] evaluaron 191 pacientes y 4.361 dientes mediante fotografías intraorales; Panyarak et al.[25] utilizaron 2.758 AMRX; y Ayan et al.[26] incluyeron 40 estudiantes y 1.200 aletas de mordida en un contexto educativo.

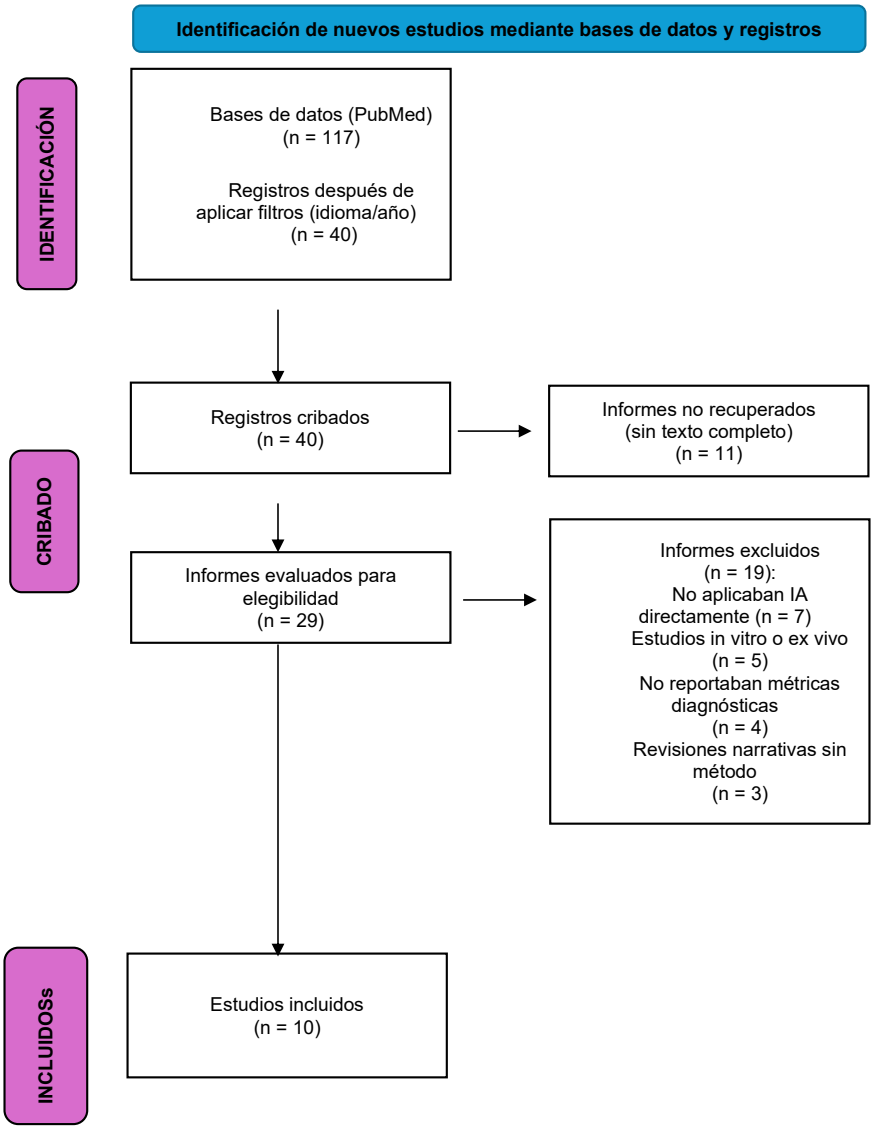

Figura 1. Diagrama de flujo PRISMA.

En relación con los procedimientos de validación, predominó la validación interna, mediante partición entrenamiento/prueba o validación cruzada. De forma general, no se reportó validación externa multicéntrica robusta, lo que limita la extrapolación de los resultados a otros centros, poblaciones y dispositivos de adquisición de imagen (**Tabla 1**).

Objetivo principal: precisión diagnóstica de la IA.

Todos los estudios incluidos apoyaron el objetivo principal, al reportar métricas de precisión diagnóstica —sensibilidad, especificidad y/o área bajo la curva (AUC)— que permitieron evaluar el desempeño de modelos de IA en la detección de caries a partir de imágenes diagnósticas. Las revisiones sistemáticas y metaanálisis (Abbott et al.[18], Albano et al.[19], Carvalho et al.[21], Mohammad-Rahimi et al.[22] y Prados-Privado et al.[24]) describieron un rendimiento globalmente favorable, aunque con variabilidad asociada al tipo de imagen analizada, al estándar de referencia empleado y a la heterogeneidad metodológica de los estudios primarios incluidos.

Objetivo secundario 1: sensibilidad, especificidad y AUC

Las sensibilidades reportadas oscilaron entre 25% y 99,8%, con un predominio aproximado de valores comprendidos entre 44% y 94% en gran parte de la evidencia (Albano et al.[19], Mohammad-Rahimi et al.[22], Prados-Privado et al.[24], Mertens et al.[23] y Szabó et al.[20]). En el estudio basado en el sistema ICCMS™, Panyarak et al.[25] reportaron sensibilidades entre 71,11% y 82,55%, lo que refleja la complejidad de la clasificación multicategoría.

La especificidad mostró un rango aún más amplio, desde 16,46% hasta 100 %. En la mayoría de los estudios, los valores se concentraron entre 77% y 98% (Abbott et al.[18], Albano et al.[19], Szabó et al.[20], Carvalho et al.[21], Zhang et al.[5], Mohammad-Rahimi et al.[22], Mertens et al.[23], Prados-Privado et al.[24] y Ayan et al.[26]). En contraste, Panyarak et al.[25] reportaron especificidades inferiores (16,46%–64,77%), compatibles con un mayor número de falsos positivos en la aplicación del sistema ICCMS™.

Cinco estudios informaron valores de AUC comprendidos entre 0,74 y 0,99 (Abbott et al.[18], Albano et al.[19], Mohammad-Rahimi et al.[22], Mertens et al.[23] y Prados-Privado et al.[24]). Los restantes estudios no reportaron esta métrica (Szabó et al.[20], Carvalho et al.[21], Zhang et al.[5], Panyarak et al.[25] y Ayan et al.[26]), lo que limita una comparación homogénea del rendimiento discriminativo entre todos los trabajos incluidos.

Objetivo secundario 2: tipo de imagen utilizada

Las radiografías aleta de mordida fueron la modalidad predominante, empleadas en siete estudios: Albano et al.[19], Szabó et al.[20], Carvalho et al.[21], Mertens et al.[23], Prados-Privado et al.[24], Panyarak et al.[25] y Ayan et al.[26], Zhang et al.[5] utilizaron exclusivamente fotografías intraorales.

Además, las revisiones sistemáticas incluyeron estudios que emplearon radiografías periapicales, panorámicas, imágenes de transiluminación por luz infrarroja cercana (NILT) y fotografías clínicas (Abbott et al.[18], Mohammad-Rahimi et al.[22] y Prados-Privado et al.[24]), lo que incrementó la heterogeneidad metodológica, pero evidenció la versatilidad del uso de la IA en distintos contextos diagnósticos.

Objetivo secundario 3: arquitecturas de IA

Las arquitecturas de IA empleadas se agruparon en familias de redes neuronales convolucionales para facilitar la comparación entre estudios. Predominaron modelos de segmentación y clasificación basados en U-Net y ResNet, así como arquitecturas de detección de objetos tipo YOLO o R-CNN, independientemente de sus variantes específicas (**Tabla 1**).

Todos los estudios utilizaron modelos basados en redes neuronales convolucionales (CNN). Las arquitecturas más frecuentes fueron U-Net, ResNet (en sus variantes 18, 50, 101 y 152) y YOLO, ¡junto con otros modelos como “Inception” y arquitecturas orientadas a tareas de detección o segmentación (**Tabla 2**).

En los estudios primarios, Szabó et al.[20] emplearon Cascade R-CNN combinado con ResNet-101; Zhang et al.[5] utilizaron MobileNet-v3 combinado con U-Net; y Mertens et al.[23] aplicaron una fully convolutional neural network (red totalmente convolucional) en un entorno de soporte clínico. Ayan et al.[26] emplearon YOLOv5xs, una CNN modificada, como herramienta con fines formativos.

Objetivo secundario 4: comparación con el diagnóstico clínico

Cinco estudios realizaron una comparación explícita entre el rendimiento de la IA y el de evaluadores humanos: Albano et al.[19], Szabó et al.[20], Carvalho et al.[21], Mertens et al.[23] y Ayan et al.[26]. En estos trabajos, el desempeño de la IA fue comparable o superior en determinadas métricas, especialmente cuando se utilizó como herramienta de apoyo diagnóstico.

En los estudios restantes, la comparación con el diagnóstico clínico fue parcial o indirecta, al basarse en consensos de expertos o en síntesis de la literatura (Abbott et al.[18], Zhang et al.[5], Mohammad-Rahimi et al.[22], Prados-Privado et al.[24] y Panyarak et al.[25]).

Tabla 1. Cuadro de los resultados. Las arquitecturas de inteligencia artificial se agruparon por familias de redes neuronales convolucionales para mejorar la legibilidad de la tabla, manteniendo los ejemplos representativos reportados en cada estudio.

Autor y año	Tipo de estudio	¿Apoya el objetivo principal?	Tamaño de la muestra	Precisión diagnóstica (Sens / Esp / AUC)	Tipo de imagen	Arquitectura de IA (agrupada)	Objetivo 4: Comparación con diagnóstico clínico	Q JCR
Abbott et al. 2025[18]	Revisión sistemática y metaanálisis	Sí	45 estudios	76 / 91 / 0,92	Clínicas y radiografías	CNN (U-Net, ResNet, YOLO, Inception, modelos tipo R-CNN)	Parcialmente	Q1
Albano et al. 2024[19]	Revisión sistemática	Sí	20 estudios	44–86 / 83–98 / 0,84–0,98	Aleta de mordida, periapicales, OPG	CNN (U-Net, ResNet, Inception, YOLO, R-CNN); ANN	Sí	Q1
Szabó et al. 2024[20]	Estudio diagnóstico retrospectivo	Sí	323 dientes	51–76 / 93 / —	Aleta de mordida y periapicales	CNN de detección (Cascade R-CNN con backbone ResNet)	Sí	Q1
Carvalho et al. 2024[21]	Revisión sistemática y metaanálisis	Sí	21 estudios	94 / 91 / —	Aleta de mordida	CNN (U-Net, ResNet, YOLO y modelos de clasificación/detección)	Sí	Q1
Zhang et al. 2024[5]	Estudio clínico prospectivo	Sí	191 pacientes / 4.361 dientes	81,31 / 95,65 / —	Fotografías intraorales	CNN de segmentación y clasificación (U-Net + MobileNet)	Parcialmente	Q1
Mohammad-Rahimi et al. 2022[22]	Revisión sistemática	Sí	42 estudios	25–99,8 / 17,2–100 / 0,99	Variadas	CNN (ResNet, U-Net, MobileNet, GoogleNet, R-CNN); autoencoders	Parcialmente	Q1
Mertens et al. 2021[23]	Ensayo clínico aleatorizado	Sí	22 dentistas / 140 aletas de mordida (~2.800 superficies)	81 / 97 / 0,89	Aleta de mordida	Red neuronal totalmente convolucional (FCNN)	Sí	Q1
Prados-Privado et al. 2020[24]	Revisión sistemática	Sí	13 estudios	59–97 / 83 / 0,74–0,971	Periapicales, aleta de mordida, OPG, NILT, fotos clínicas	CNN (ResNet, U-Net, Mask R-CNN, FCNN)	Parcialmente	Q1
Panyarak et al. 2023[25]	Estudio experimental diagnóstico	Sí	2.758 aleta de mordida	71,11–82,55 / 16,46–64,77 / —	Aleta de mordida	CNN de clasificación (ResNet y variantes)	Parcialmente	Q2
Ayan et al. 2024[26]	Estudio experimental (educativo)	Sí	40 estudiantes / 1.200 aletas de mordida	50–76 / 77–90 / —	Aleta de mordida	CNN de detección (YOLO, uso educativo)	Sí	Q3

Calidad metodológica de los estudios incluidos

Tabla 2. Calidad de las revisiones sistemáticas según AMSTAR-2.

Autor y año	¿Metaanálisis?	Protocolo registrado (PROSPERO)	Búsqueda exhaustiva	Selección/ extracción por duplicado	Evaluación de sesgo en estudios primarios	Síntesis (incl. heterogeneidad)	Ítems N/A	Calidad global (AMSTAR-2)
Abbott et al. 2025[18]	Sí	Sí	Sí	Sí	Sí	Metaanálisis bivariado de sensibilidad, especificidad, AUC e I^2	—	Alta
Albano et al. 2024[19]	No	Sí	Sí	Sí	Sí	Síntesis cualitativa estructurada	11, 12, 15	Alta
Carvalho et al. 2024[21]	Sí	Sí	Sí	Sí	Sí	Metaanálisis de sensibilidad y especificidad	—	Alta
Mohammad-Rahimi et al. 2022[22]	No	Sí	Sí	Parcial	Sí	Síntesis cualitativa con heterogeneidad elevada	11, 12, 15	Moderada
Prados-Privado et al. 2020[24]	No	No	Sí	No claramente duplicado	Sí	Síntesis narrativa (sin metaanálisis)	11, 12, 15	Moderada

Nota: En revisiones sin metaanálisis, los ítems 11, 12 y 15 de AMSTAR-2 se consignaron como N/A por ser específicos de la síntesis cuantitativa.

Tabla 3. Evaluación de estudios primarios mediante QUADAS-2.

Autor y año (Ref.)	Tipo de estudio	Selección de pacientes	Prueba índice (IA)	Estándar de referencia	Flujo y temporalidad	Riesgo de sesgo global	Aplicabilidad
Szabó et al. 2024[20]	Diagnóstico retrospectivo	Bajo	Bajo	Incierto	Incierto	Moderado	Bajas preocupaciones
Zhang et al. 2024[5]	Clínico prospectivo	Incierto	Bajo–incierto	Incierto	Incierto	Moderado	Moderadas (fotografías; 1 centro)
Mertens et al. 2021[23]	Ensayo clínico aleatorizado	Bajo	Bajo	Incierto	Bajo–incierto	Bajo–moderado	Bajas preocupaciones
Panyarak et al. 2023[25]	Experimental diagnóstico	Incierto	Incierto	Incierto	Incierto	Moderado	Moderadas (ICCMS™; baja especificidad)
Ayan et al. 2024[26]	Experimental (educativo)	N/A	Incierto	Incierto	N/A	N/A	Altas (contexto educativo; no pacientes)

Pie de tabla: N/A (no aplicable) en Ayan et al.[26] por tratarse de un estudio educativo sin población clínica de pacientes; los dominios de selección de pacientes y flujo/temporalidad de QUADAS-2 no son aplicables y no se calcula un riesgo global QUADAS-2. En el resto de los estudios, "Incierto" se utiliza cuando el artículo no reporta con suficiente detalle cegamiento, definición del estándar de referencia o el flujo completo/exclusiones.

Resumen de calidad metodológica

En AMSTAR-2, tres revisiones alcanzaron confianza global alta (Abbott et al.[18], Albano et al.[19] y Carvalho et al.[21]) y dos fueron moderadas (Mohammad-Rahimi et al.[22] y Prados-Privado et al.[24]), principalmente por limitaciones en el reporte de duplicación o ausencia de protocolo. En QUADAS-2, los estudios clínicos mostraron en conjunto riesgo de sesgo bajo a moderado; los dominios con mayor incertidumbre fueron el estándar de referencia y el flujo/temporalidad. El estudio educativo de Ayan et al.[26] se interpreta fundamentalmente en términos de aplicabilidad.

Discusión

Los resultados de la presente revisión sistemática deben interpretarse teniendo en cuenta una serie de limitaciones metodológicas y estructurales que afectan tanto al propio diseño del estudio como a la literatura disponible sobre IA aplicada al diagnóstico de caries. En primer lugar, la restricción a artículos con texto completo disponible pudo haber reducido el número de estudios potencialmente relevantes, introduciendo un posible sesgo de selección. Asimismo, la marcada heterogeneidad en los diseños metodológicos, los estándares de referencia y las métricas diagnósticas reportadas impidió la realización de un metaanálisis cuantitativo, limitando la obtención de estimaciones agrupadas de precisión diagnóstica. A ello se suma que la mayoría de los estudios incluidos se basan en validaciones internas y conjuntos de datos procedentes de uno o pocos centros, lo que condiciona de forma significativa la generalización de los resultados a otros entornos clínicos con diferente equipamiento, prevalencia de caries y características poblacionales. Además, es importante señalar que la definición de los estadios de caries no fue homogénea entre los estudios incluidos. Cuando esta información estuvo disponible, las caries iniciales se definieron generalmente como lesiones limitadas al esmalte, sin cavitación evidente o con cambios radiolúcidos incipientes, mientras que las caries avanzadas incluyeron lesiones con compromiso dentinario o cavitación radiográfica clara. La falta de estandarización en la clasificación de las lesiones constituye una limitación relevante y contribuye a la variabilidad observada en el rendimiento diagnóstico de los modelos de IA.

Se observó una considerable heterogeneidad clínica y metodológica entre los estudios incluidos. La heterogeneidad clínica se debió principalmente a diferencias en las modalidades de imagen (radiografías de aleta mordida, periapicales, panorámicas y fotografías intraorales), las superficies dentales evaluadas y los estadios de caries. La heterogeneidad metodológica se debió a la variabilidad en los estándares de referencia, los criterios de clasificación de lesiones, las arquitecturas de IA y las métricas diagnósticas reportadas. Esta heterogeneidad limitó la comparabilidad directa entre los estudios e impidió un metaanálisis cuantitativo.

Pese a estas limitaciones, el presente trabajo ofrece varias fortalezas relevantes. Se trata de una revisión sistemática desarrollada conforme a las directrices PRISMA 2020 y al marco PICO, con una estrategia de búsqueda transparente, criterios de inclusión y exclusión claramente definidos y un proceso de selección y evaluación de calidad realizado por dos revisores independientes. La inclusión de estudios recientes publicados entre 2020 y 2025, mayoritariamente en revistas indexadas de alto impacto, permite ofrecer una visión actualizada del estado del arte. Además, la evaluación de la calidad metodológica mediante herramientas validadas como AMSTAR-2 y QUADAS-2 aporta un análisis crítico de la evidencia disponible, evitando interpretaciones basadas únicamente en métricas de rendimiento.

Uno de los hallazgos más consistentes de esta revisión es el impacto de la variabilidad del conjunto de datos en el rendimiento de los modelos de IA. La mayoría de los sistemas analizados fueron entrenados con conjuntos de datos procedentes de centros académicos, con protocolos de adquisición homogéneos y equipos de alta calidad. Esta falta de diversidad limita la capacidad de generalización de los modelos cuando se aplican en la práctica clínica real, donde existe una considerable variabilidad en dispositivos radiográficos, calidad de imagen y perfiles de pacientes. En consecuencia, el rendimiento diagnóstico reportado en estudios controlados puede sobreestimar la precisión alcanzable en entornos clínicos habituales. Esta problemática ha sido destacada también por Schwendicke et al., quienes señalan la ausencia de validación externa multicéntrica como una de las principales barreras para la implementación clínica segura de la IA en odontología[27].

El análisis diferenciado según el estadio de la caries revela un patrón claro. Los modelos de IA muestran un rendimiento elevado en la detección de caries moderadas y avanzadas, caracterizadas por cavitación radiográfica evidente o compromiso dentinario, con sensibilidades que en algunos estudios superan el 90%[18,21,23]. Por el contrario, la detección de lesiones iniciales limitadas al esmalte presenta sensibilidades más bajas y una mayor variabilidad entre estudios, situándose aproximadamente entre el 44% y el 76%[18,22,25]. Esta diferencia evidencia una limitación fundamental de los sistemas actuales para identificar signos radiográficos sutiles, lo que resulta especialmente

relevante desde el punto de vista preventivo, dado que el diagnóstico temprano es el escenario donde la IA podría aportar un mayor beneficio clínico.

En relación con el tipo de imagen y la superficie dental evaluada, la evidencia disponible se concentra predominantemente en la detección de caries proximales mediante AMRX, utilizadas en la mayoría de los estudios incluidos. Si bien este enfoque refuerza la utilidad de la IA en un contexto clínico específico, introduce un sesgo estructural que limita la extrapolación de los resultados a superficies oclusales, donde el diagnóstico depende en mayor medida de la inspección visual y de métodos complementarios. La escasez de estudios centrados en caries oclusales representa una laguna relevante en la literatura actual y subraya la necesidad de ampliar la investigación a otros escenarios diagnósticos.

Respecto a las arquitecturas de IA, el predominio de redes neuronales convolucionales como U-Net, ResNet y YOLO indica un cierto grado de madurez tecnológica. En concordancia con los resultados resumidos en la **Tabla 1**, el rendimiento diagnóstico observado parece depender en mayor medida de la calidad y representatividad de los conjuntos de datos y del contexto clínico que de la arquitectura específica de la red neuronal empleada. En este contexto, Yang y Chen (2025) demostraron que la incorporación de estrategias de clasificación fina y mejoras de posprocesamiento puede optimizar la detección de caries iniciales, aunque los propios autores reconocen la necesidad de validación externa independiente antes de su aplicación clínica generalizada[28].

La comparación entre la IA y el diagnóstico clínico humano mostró que, en entornos experimentales controlados, la IA puede igualar o incluso superar la sensibilidad diagnóstica de los odontólogos, manteniendo especificidades comparables[18,20,21,23,26]. No obstante, estas comparaciones se realizaron bajo condiciones estandarizadas y con estándares de referencia definidos por consenso de expertos, lo que limita su extrapolación directa a la práctica clínica diaria, caracterizada por una mayor variabilidad interobservador. Por ello, la IA debe entenderse como una herramienta de apoyo destinada a complementar el juicio clínico, y no como un sustituto del profesional.

Finalmente, estudios recientes han explorado la generalización de estos modelos a otras modalidades diagnósticas. Turosz et al. analizaron en 2025 la aplicación de IA en radiografías panorámicas, destacando su potencial como ayuda clínica, pero enfatizando que el rendimiento del modelo depende en gran medida de la estandarización de la adquisición de imágenes y de la calidad del conjunto de datos. Estos hallazgos refuerzan la necesidad de validación externa multicéntrica antes de una implementación clínica generalizada, especialmente cuando se pretende extrapolar los modelos a modalidades distintas de la AMRX[29].

En conjunto, esta revisión pone de manifiesto que la IA constituye una herramienta prometedora para el apoyo diagnóstico en odontología, particularmente en la detección de caries proximales moderadas y avanzadas. Sin embargo, su utilidad clínica real sigue condicionada por limitaciones metodológicas, la variabilidad del conjunto de datos y la escasa validación externa. Abordar estos desafíos mediante estudios multicéntricos, protocolos estandarizados y modelos evaluados en condiciones clínicas reales será esencial para avanzar hacia una implementación ética, fiable y clínicamente efectiva de la IA en el diagnóstico de caries.

Conclusiones

La evidencia analizada en esta revisión sistemática indica que la IA aplicada al análisis de imágenes diagnósticas constituye una herramienta emergente de apoyo en el diagnóstico de caries dentales. Sin embargo, su utilidad clínica depende del tipo de lesión, de la modalidad de imagen empleada y del contexto metodológico de los estudios incluidos. En función de los objetivos planteados, se concluye que:

1. En relación con el objetivo principal, los modelos de IA muestran una capacidad diagnóstica relevante para la detección de caries dentales en entornos controlados, especialmente cuando se aplican a radiografías intraorales. No obstante, la evidencia disponible no permite generalizar estos resultados a todos los escenarios clínicos ni a todas las poblaciones.

2. Respecto al análisis de sensibilidad, especificidad y área bajo la curva (AUC), se observa que los valores más elevados se alcanzan en la detección de caries moderadas y avanzadas, particularmente aquellas con signos radiográficos evidentes. Por el contrario, la detección de caries iniciales limitadas al esmalte presenta una menor precisión diagnóstica y una mayor variabilidad entre estudios, lo que limita su aplicación en el diagnóstico precoz.

3. Las AMRX son el tipo de radiografía más utilizado y la que ofrece resultados más consistentes, especialmente en la detección de caries proximales. Sin embargo, la evidencia es escasa y heterogénea para otras modalidades como radiografías panorámicas, fotografías intraorales o sistemas multimodales, lo que restringe la extrapolación de los resultados a superficies oclusales y a otros contextos diagnósticos.

4. El uso de redes neuronales convolucionales como U-Net, ResNet y YOLO, son las arquitecturas de IA predominantes. Los resultados sugieren que el rendimiento diagnóstico depende en mayor medida de la calidad, diversidad y estandarización de los conjuntos de datos utilizados que de la arquitectura específica del modelo, lo que subraya la importancia de conjuntos de datos representativos y multicéntricos.

5. En los estudios que comparan la IA con el diagnóstico clínico realizado por profesionales, los modelos de IA pueden alcanzar un rendimiento comparable en términos de sensibilidad en determinados escenarios, especialmente para caries proximales y lesiones avanzadas. No obstante, no se observa una superioridad consistente frente al diagnóstico humano, y los resultados varían según el tipo de lesión, el estándar de referencia y el diseño del estudio.

   Desde el punto de vista clínico, la IA podría integrarse en el flujo asistencial actual como una herramienta de apoyo a la interpretación de radiografías intraorales, especialmente de AMRX. Su uso como sistema de segunda lectura podría contribuir a reducir la variabilidad interobservador y a mejorar la detección de lesiones evidentes, sin sustituir el juicio clínico del odontólogo.

En este sentido, la aplicación de la IA parece especialmente útil en escenarios de cribado radiográfico y en contextos con alta carga asistencial, donde puede facilitar la identificación temprana de lesiones moderadas y avanzadas y apoyar la toma de decisiones clínicas de forma más eficiente.

Sin embargo, la falta de una validación externa y multicéntrica sólida representa una de las principales barreras para la implementación clínica de sistemas de detección de caries basados en IA.

Investigaciones futuras

Con las limitaciones de este estudio, los autores concluyen respecto a futuras líneas de investigación que será fundamental abordar aspectos emergentes como la interpretabilidad de los modelos de IA, con el fin de aumentar la confianza clínica y facilitar su adopción. Así mismo, las investigaciones futuras deberían priorizar los estudios prospectivos multicéntricos que utilicen conjuntos de datos heterogéneos y protocolos de diagnóstico estandarizados.

También es importante considerar los marcos regulatorios y las implicaciones en términos de responsabilidad profesional derivadas de su uso en la práctica clínica. En último lugar, la realización de estudios de coste-efectividad será clave para concretar el impacto real de estas tecnologías y su viabilidad antes de una implementación generalizada en los sistemas de salud.

Abreviaturas utilizadas

CAOD: Caries, Ausentes, Obturados (índice de experiencia de caries)

OMS: Organización Mundial de la Salud (World Health Organization)

IA: Inteligencia Artificial (Artificial Intelligence)

CNN: Redes Neuronales Convolucionales (Convolutional Neural Networks)

RS: Revisión sistemática (Systematic Review)

RS+MA: Revisión sistemática con metaanálisis (Systematic Review with Meta-analysis)

ECA/RCT: Ensayo clínico aleatorizado (Randomized Clinical Trial)

AUC: Área bajo la curva ROC (Area Under the Curve)

PROSPERO: International Prospective Register of Systematic Reviews

JCR: Journal Citation Reports

Q1: Cuartil 1 (revistas de mayor impacto)

MeSH: Medical Subject Headings

PICO: Población, Intervención, Comparación, Resultado

OPG: Ortopantomografía / radiografía panorámica (Orthopantomogram)

NILT: Transiluminación con luz infrarroja cercana (Near-Infrared Light Transillumination)

FCNN: Fully Convolutional Neural Networks

YOLO: You Only Look Once

Financiación

No se recibió financiación para la investigación, autoría y/o publicación de este artículo.

Contribución de autoría

Selina Nunha: Adquisición de datos, Conceptualización, Redacción

Sebastiana Arroyo-Bote: Conceptualización, Análisis formal, Administración del proyecto

Pere Riutord-Sbert: Redacción, Supervisión

Catalina Bennasar-Verger: Administración del proyecto, Supervisión

Todos los autores revisaron críticamente el manuscrito, aprobaron la versión final y se responsabilizan de su contenido.

Conflictos de interés

Los autores declaran que no hay ningún conflicto de interés comercial o financiero para esta investigación.

Declaraciones éticas

Este trabajo (artículo original, ensayo clínico, investigación con seres humanos/animales, etc.) cumple con la normativa de investigación y cuenta con la aprobación de un Comité de Ética. Los autores disponen del documento oficial o código de aprobación correspondiente.

Uso de herramientas de inteligencia artificial

Declaro que no se ha utilizado ninguna herramienta de inteligencia artificial (como ChatGPT, Copilot, Gemini, u otras) en la redacción, análisis o revisión de este artículo.

Referencias

1. Mohammad-Rahimi M, et al. Artificial intelligence in dental caries detection and diagnosis: a systematic review. Oral Dis. 2024;30(4):1765-1783.
2. World Health Organization. WHO/Europe calls for urgent action on oral disease. Copenhagen: WHO Regional Office for Europe; 2023.
3. Bravo Pérez M, et al. Epidemiología de la caries dental en España. Rev Clin Esp Odontol. 2020;25(4):12-69.
4. Gomez J. Detection and diagnosis of dental caries using digital imaging. BMC Oral Health. 2015;15(Suppl 1):S3.
5. Zhang JW, et al. Deep learning for caries detection on intraoral photographs. BMC Oral Health. 2024;24:1095-1103.
6. Kargozar S, et al. Artificial intelligence applications in dental caries diagnosis: current evidence. BMC Oral Health. 2024;24:828-836.
7. Schwendicke F, et al. Artificial intelligence in dental diagnostics. J Dent Res. 2020;99(7):769-774.
8. Zeng Z, et al. Automated detection of dental caries using convolutional neural networks. Quintessence Int. 2025;56(1):82-91.
9. Park Y, et al. Performance of deep learning algorithms for caries detection. J Dent Res. 2022;101(11):1372-1379.
10. Yu H, et al. Convolutional neural networks for dental caries diagnosis. J Dent Res. 2020;99(3):249-256.
11. Hao J, et al. Deep learning-based detection of dental caries on radiographs. J Dent Res. 2022;101(3):304-311.
12. Yoon K, et al. Artificial intelligence-assisted dental diagnosis using radiographic images. Comput Methods Programs Biomed. 2023;233:107465.
13. Gu J, et al. Recent advances in convolutional neural networks. Pattern Recognit. 2018;77:354-377.
14. Hwang JJ, et al. Computer-aided diagnosis of dental caries. Imaging Sci Dent. 2019;49(1):1-7.
15. Page MJ, et al. The PRISMA 2020 statement: an updated guideline. BMJ. 2021;372:n71.
16. Shea BJ, et al. AMSTAR-2: a critical appraisal tool. BMJ. 2017;358:j4008.
17. Whiting PF, et al. QUADAS-2: a revised tool for diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-536.
18. Abbott LP, et al. Diagnostic accuracy of artificial intelligence for caries detection: systematic review and meta-analysis. J Evid Based Dent Pract. 2025;25(1):102077.
19. Albano D, et al. Artificial intelligence for caries detection: systematic review. BMC Oral Health. 2024;24:274-282.
20. Szabó V, et al. Deep learning-based detection of dental caries on radiographs. J Dent. 2024;147:105105.
21. Carvalho BKG, et al. Accuracy of artificial intelligence in caries detection: meta-analysis. J Dent. 2024;151:105388.
22. Mohammad-Rahimi H, et al. Artificial intelligence in dental caries diagnosis. J Dent. 2022;122:104115.
23. Mertens S, et al. Clinical evaluation of AI-assisted caries detection. J Dent. 2021;115:103849.
24. Prados-Privado M, et al. Artificial intelligence and dental imaging: systematic review. J Clin Med. 2020;9(11):3579.
25. Panyarak W, et al. Diagnostic performance of AI using ICCMS classification. Oral Surg Oral Med Oral Pathol Oral Radiol. 2023;135(2):272-281.
26. Ayan E, et al. Artificial intelligence as a teaching aid for caries detection. J Dent Educ. 2024;88(4):490-500.
27. Yang L, Chen GY. Evaluation of deep learning for caries detection with fine-grained classification and postprocessing improvements. Int Dent J. 2025;75(5):100898. doi:10.1016/j.identj.2025.100898.
28. Schwendicke F, Krois J, Gomez J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2024;103(1):12-20.
29. Turosz A, Nowak R, Kawala B, et al. Artificial intelligence in panoramic images—clinical aid to a dentist. Diagnostics (Basel). 2025;15(3):412.