



**Universidad
Europea** VALENCIA

Grado en ODONTOLOGÍA

Trabajo Fin de Grado

Curso 2024-2025

**REVISIÓN SISTEMÁTICA DEL DIAGNÓSTICO DE LA
HIPOMINERALIZACIÓN DE LOS MOLARES E INCISIVOS
(MIH) UTILIZANDO LA INTELIGENCIA ARTIFICIAL :
ESTADO ACTUAL Y PERSPECTIVAS FUTURAS.**

Presentado por: VILLETTE Valentin

Tutor: SELVA RIBERA, Inmaculada C

Campus de Valencia
Paseo de la Alameda, 7
46010 Valencia

universidadeuropea.com

Índice

1. RESUMEN	3
2. ABSTRACT	4
3. PALABRAS CLAVES Y ACRÓNIMOS	5
3.1 Listado de palabras claves	5
3.2 Siglas y acrónimos	5
4. INTRODUCCIÓN	6
4.1 Generalidades de la MIH	6
4.1.1. Impacto clínico y social de la MIH	7
4.1.2. Implicaciones en el manejo clínico	8
4.2 Complicaciones en el diagnóstico tradicional	8
4.2.1. Subjetividad en la evaluación clínica	8
4.2.2. Variabilidad en los criterios diagnósticos	8
4.2.3. Dificultad para detectar la MIH en etapas tempranas	9
4.2.4. Diagnóstico diferencial del MIH	9
4.3. Aplicación de la IA en el diagnóstico de la MIH	9
4.4. Ventajas específicas de la IA	10
4.4.1. Reducción del tiempo de diagnóstico y procesamiento de los datos	10
4.4.2. Mejora de la precisión diagnóstica	10
4.4.3. Aumento del acceso a diagnósticos de calidad	10
4.4.4. Identificación precoz antes del colapso post-eruptivo	11
4.4.5. Integración con tecnologías avanzadas	11
4.5. Desventajas en la implementación de la IA	11
4.5.1. Necesidad de datos de entrenamiento calidad y en cantidad	11
4.5.2 Aceptación por parte de profesionales	11
4.5.3 Aspectos éticos y legales	12
4.5.4. Desafíos técnicos	12
5. JUSTIFICACIÓN E HIPÓTESIS	13
5.1 Hipótesis	13
5.2 Justificación	13

6. OBJETIVOS	15
7. MATERIAL Y MÉTODO	16
7.1. Identificación de la pregunta PICO	16
7.2. Criterios de elegibilidad	16
7.3 Fuente de información y estrategia de la búsqueda de datos	18
7.3.1 Búsqueda cruzada	18
7.4. Proceso de selección de los estudios	19
7.5. Extracción de datos	19
7.6. Valoración de la calidad	20
7.7. Síntesis de datos	20
8. RESULTADOS	22
8.1. Selección de estudios. Flow chart	22
8.2. Análisis de las características de los estudios revisados	23
8.3. Evaluación de la calidad metodológica y riesgo de sesgo	25
8.4. Síntesis resultados	27
8.4.1. Resultados diagnósticos de los modelos de IA	27
8.4.2 Complicaciones relacionadas con el procedimiento de diagnóstico asistido por IA	28
9. DISCUSIÓN	30
9.1. Precisión diagnóstica y rendimiento general	30
9.2. Complicaciones técnicas y limitaciones metodológicas	30
9.4. Limitaciones de la presente revisión	31
10. CONCLUSIÓN	33
11. BIBLIOGRAFÍA	34
12. ANEXOS	37

1. RESUMEN

Introducción: La hipomineralización molar-incisiva (MIH) representa un desafío diagnóstico en odontología pediátrica debido a su variabilidad clínica y su frecuente solapamiento con otros defectos del esmalte. Los modelos de inteligencia artificial (IA) se están aplicando cada vez más para apoyar el diagnóstico precoz y mejorar la precisión en la clasificación. El objetivo de esta revisión sistemática fue evaluar el rendimiento diagnóstico de los sistemas de IA en la detección de MIH a partir de datos.

Material y Métodos: Se realizó una búsqueda electrónica en bases de datos como PubMed, Scopus y Web of Science hasta febrero de 2024. Se incluyeron nueve estudios que evaluaban modelos de IA como redes neuronales convolucionales (CNN) y arquitecturas basadas en transformadores entrenados con imágenes dentales clínicas o intraorales.

Resultados: En los estudios analizados, los modelos de IA alcanzaron precisiones diagnósticas que oscilaron entre el 83.98 % y el 97.8 %. Los valores de sensibilidad se situaron entre el 78.6 % y el 93.5 %, mientras que la especificidad fue superior al 90 % en la mayoría de los casos. Sin embargo, se identificaron varias limitaciones: ausencia de validación externa, reducción del rendimiento en imágenes de baja calidad y variabilidad según la zona anatómica.

Conclusiones: A pesar de sus limitaciones, la IA se presenta como una herramienta confiable y prometedora para la detección precoz de la MIH. Se requiere mayor validación clínica, estandarización de los protocolos de imagen y ensayos multicéntricos para consolidar su aplicación en contextos reales y su integración en los flujos de trabajo odontológicos rutinarios.

2. ABSTRACT

Introduction: Molar-incisor hypomineralization (MIH) poses a diagnostic challenge in pediatric dentistry due to its clinical variability and frequent overlap with other enamel defects. Artificial intelligence (AI) models are increasingly being applied to support early diagnosis and improve accuracy in classification. The aim of this systematic review was to evaluate the diagnostic performance of AI systems in detecting MIH based on clinical data.

Materials and Methods: An electronic search was conducted in databases such as PubMed, Scopus, and Web of Science up to February 2024. Nine studies were included, evaluating AI models such as convolutional neural networks (CNNs) and transformer-based architectures trained with clinical or intraoral dental images.

Results: In the analyzed studies, AI models achieved diagnostic accuracies ranging from 83.98% to 97.8%. Sensitivity values ranged between 78.6% and 93.5%, while specificity was above 90% in most cases. However, several limitations were identified: lack of external validation, reduced performance in low-quality images, and variability depending on the anatomical region.

Conclusions: Despite its limitations, AI emerges as a reliable and promising tool for the early detection of MIH. Further clinical validation, standardization of imaging protocols, and multicenter trials are needed to support its real-world application and integration into routine dental workflows.

3. PALABRAS CLAVES Y ACRÓNIMOS

3.1 Listado de palabras claves

MIH	Hipomineralización incisivo molar
IA	Inteligencia artificial
CNN	Convolutional Neural Network

3.2 Siglas y acrónimos

MIH	Hipomineralización Incisivo-Molar
IA	Inteligencia Artificial
CNN	Red Neuronal Convolucional (Convolutional Neural Network)
CAD-CAM	Diseño Asistido por Computadora - Manufactura Asistida por Computadora
EAPD	Academia Europea de Odontología Pediátrica (European Academy of Paediatric Dentistry)
PRISMA	Elementos de Reporte Preferidos para Revisiones Sistemáticas y Metaanálisis
PROSPERO	Registro Internacional Prospectivo de Revisiones Sistemáticas
VV	Villette Valentin
ISR	Inmaculada Selva Ribera
JBI	Instituto Joanna Briggs (herramienta de evaluación de calidad)
MOGA	Guía Metodológica para Series de Casos
RESNEXT-101	Arquitectura de red neuronal convolucional avanzada
VGG-19	Arquitectura de red neuronal convolucional desarrollada por Oxford
RESNET	Red Neuronal Residual
DENSENET	Red Neuronal Densa
OPENCV	Biblioteca de Visión por Computadora de Código Abierto
SAM	Modelo Segment Anything (modelo de segmentación de imágenes)
SEG FORMER-BS	Modelo tipo Transformer para segmentación de imágenes
AP	Precisión Promedio (Average Precision)
IOU	Intersección sobre Unión

4. INTRODUCCIÓN

La hipomineralización de los molares e incisivos (MIH) es un desafío clínico creciente en la odontología pediátrica, afecta hasta el 40 % de los niños en algunas regiones dependiendo de los métodos de diagnóstico utilizados (1,2). Esta condición no solo genera hipersensibilidad dental y un mayor riesgo de caries, sino que también tiene un impacto significativo en la calidad de vida de los niños afectados (3,4,5). La alta prevalencia y las consecuencias clínicas hacen que esta patología sea una prioridad en los programas de salud oral infantil (6,7).

La introducción de la inteligencia artificial (IA) en el diagnóstico de la MIH ofrece una oportunidad sin precedentes para optimizar la atención odontológica y ofrece soluciones prometedoras para abordar las limitaciones tradicionales en el diagnóstico de la MIH; facilita intervenciones preventivas oportunas y reduce la carga clínica en etapas avanzadas (8,9,10).

4.1 Generalidades de la MIH

La hipomineralización molar-incisiva (MIH) es un defecto cualitativo del esmalte causado por factores prenatales, perinatales y postnatales que afectan la amelogénesis. Entre ellos se incluyen enfermedades maternas, parto prematuro, bajo peso al nacer y enfermedades infantiles como fiebre alta y uso prolongado de antibióticos (4,5). Estos factores interfieren en la mineralización del esmalte, dando lugar a opacidades bien delimitadas de color blanco, amarillo o marrón, reduciendo su dureza y haciéndolo más propenso a fracturas (3). Histológicamente, el esmalte afectado muestra una menor concentración de calcio y fósforo, un aumento del contenido proteico y una mayor porosidad, lo que compromete su resistencia (7).

La prevalencia de la MIH varía significativamente entre regiones y poblaciones, con tasas reportadas de entre 2.8 % y 44 % a nivel mundial. Según metaanálisis recientes, la prevalencia global media es de aproximadamente 13.5 %, con diferencias marcadas entre continentes: América muestra la prevalencia más alta (15.3 %), mientras que Asia presenta la más baja (10.7 %) (1).

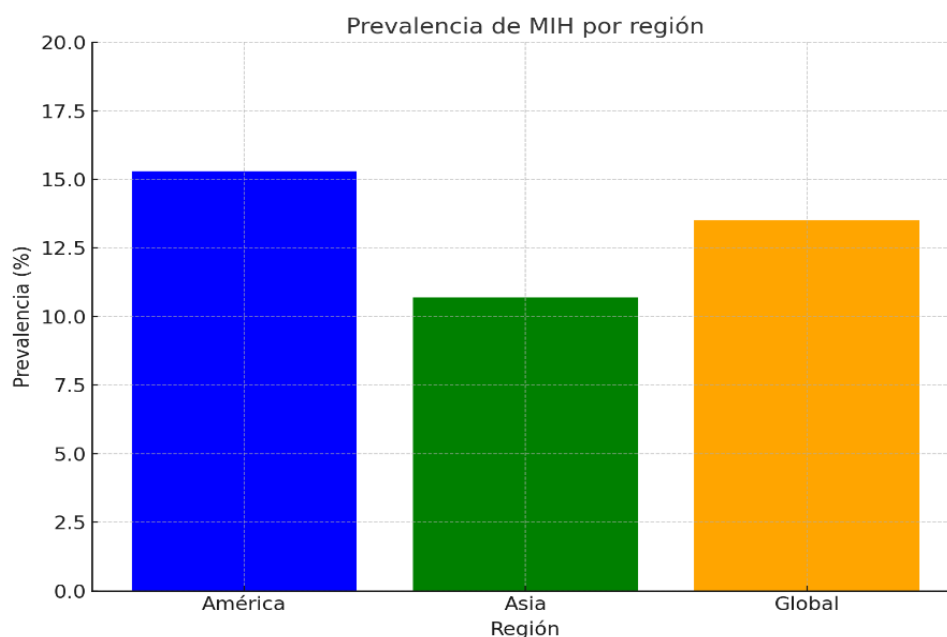


Figura 1: Prevalencia de la MIH por región.

4.1.1. Impacto clínico y social de la MIH

La MIH tiene consecuencias clínicas significativas como la hipersensibilidad dental que varía de leve a extrema debido a la porosidad del esmalte cuando existe una exposición a estímulos térmicos y mecánicos, lo que dificulta realizar actividades cotidianas como la masticación o el consumo de alimentos fríos y calientes (6). También existe un mayor riesgo de caries y una mayor progresión de la lesión, siendo los primeros molares permanentes los más afectados (2,6,9).

Otra consecuencia que se presenta a diario en la clínica odontológica, quizá la más importante, es la dificultad del éxito de los tratamientos restauradores, ya que la adhesión de los materiales al esmalte hipomineralizado es baja, reduciendo así su durabilidad (6,9). En los casos más severos, las complicaciones pueden incluir fracturas extensas del esmalte y la necesidad de intervenciones repetidas, lo que aumenta el impacto en la calidad de vida del paciente e incrementa el riesgo de infecciones severas, abscesos y la necesidad de tratamientos extensivos o extracciones (2,6,9).

La MIH representa un desafío tanto clínico como social. Los pacientes con MIH pueden experimentar una reducción en su calidad de vida debido a problemas estéticos y funcionales, como la hipersensibilidad dental y las opacidades visibles en los incisivos, afectando la autoestima de estos y pudiendo generar ansiedad y fobia en relación con los tratamientos dentales repetitivos, dificultando la ejecución de los

tratamientos necesarios (2,6,9). Esta preocupación estética es más prominente en la etapa escolar, donde la apariencia juega un rol importante (9).

4.1.2. Implicaciones en el manejo clínico

El manejo de la MIH requiere un enfoque integral que abarque tanto los aspectos funcionales como emocionales de la condición. Las estrategias incluyen:

- *Diagnóstico temprano*: Identificación de los dientes afectados antes de que se produzcan fracturas o caries severas. Las tecnologías avanzadas, como la inteligencia artificial, están mostrando promesas en este ámbito (9).
- *Protocolos de manejo del dolor*: Uso de técnicas desensibilizantes como pastas dentales específicas y la aplicación de barnices de flúor para reducir la sensibilidad (6,10).
- *Restauraciones personalizadas*: Utilización de materiales avanzados como cerámicas CAD/CAM o coronas preformadas para maximizar la eficacia de las restauraciones en dientes afectados (2,6).

El manejo exitoso de la MIH no solo mejora los resultados clínicos, sino que también puede impactar positivamente en la calidad de vida de los pacientes y sus familias (6).

4.2 Complicaciones en el diagnóstico tradicional

4.2.1. Subjetividad en la evaluación clínica

El diagnóstico de la MIH se basa principalmente en la inspección visual de opacidades demarcadas en el esmalte y fracturas posteriores a la erupción. Sin embargo, la experiencia del clínico juega un papel crucial, lo que genera una alta variabilidad interobservador e intraobservador (1,5). Además, las diferencias en la percepción de los colores y la gravedad de las opacidades pueden llevar a interpretaciones inconsistentes (7).

4.2.2. Variabilidad en los criterios diagnósticos

Aunque existen directrices establecidas por la Academia Europea de Odontología Pediátrica (EAPD), la falta de uniformidad en la aplicación de estos

criterios complica la comparación de estudios epidemiológicos y la práctica clínica (6,7). Los criterios diagnósticos incluyen la presencia de opacidades bien delimitadas y fracturas post-eruptivas, pero no siempre se implementan de manera consistente (9).

4.2.3. Dificultad para detectar la MIH en etapas tempranas

La detección temprana de la MIH es un desafío, ya que las alteraciones del esmalte pueden no ser evidentes antes de la erupción completa de los molares permanentes. Las fracturas post-eruptivas complican aún más el diagnóstico porque pueden alterar las características iniciales del esmalte afectado, lo que lleva a diagnósticos erróneos o tardíos (5).

4.2.4. Diagnóstico diferencial del MIH

La MIH comparte características clínicas con otras patologías del esmalte, como la fluorosis dental, las lesiones blancas iniciales de caries y la hipoplasia del esmalte. Estas similitudes dificultan la diferenciación clínica, especialmente sin el uso de herramientas diagnósticas avanzadas (5,7,9).

El MIH presenta características clínicas que nos permite diferenciarlo de las otras patologías.

- Presencia de opacidades bien delimitadas en molares permanentes e incisivos.
- Coloración blanco cremoso, amarillenta o marrón.
- Mayor susceptibilidad a fracturas post-erupción.
- Sensibilidad dental al frío o al cepillado.
- Afectación asimétrica, es decir, no todos los dientes presentan las mismas lesiones.

La tabla 1 diseñada para ayudar a diagnosticar correctamente el MIH de otras condiciones se adjunta en anexos.

4.3. Aplicación de la IA en el diagnóstico de la MIH

La inteligencia artificial (IA) es una tecnología que permite a los ordenadores aprender y tomar decisiones basadas en datos. Mediante herramientas como las redes

neuronales convolucionales (CNNs) y la visión por computadora, la IA analiza radiografías, fotos intraorales y escaneos digitales para identificar signos de MIH (8,10).

La inteligencia artificial (IA) ha transformado la odontología, especialmente en el ámbito del diagnóstico basado en imágenes. Modelos como las redes neuronales convolucionales (CNNs) y los transformadores se han utilizado con éxito para detectar patologías dentales como caries, periodontitis y anomalías del esmalte (8,10). En el caso de la MIH, estas tecnologías han demostrado un gran potencial para superar las limitaciones de los métodos tradicionales, proporcionando diagnósticos más precisos y estandarizados (8,9).

4.4. Ventajas específicas de la IA

4.4.1. Reducción del tiempo de diagnóstico y procesamiento de los datos

Los sistemas de IA pueden procesar grandes volúmenes de datos en segundos e imágenes en tiempo real, permitiendo a los odontólogos concentrarse en la planificación del tratamiento y proporcionando diagnósticos rápidos que mejoran significativamente la eficiencia. En entornos clínicos con alta carga de trabajo, esta capacidad mejora la atención odontológica (9,11).

4.4.2. Mejora de la precisión diagnóstica

Estudios recientes han mostrado que los algoritmos de IA alcanzan una sensibilidad y especificidad superiores al 90 % en la detección de la MIH frente a los métodos manuales tradicionales. Esta precisión permite un diagnóstico más confiable, especialmente en casos complejos (9,11). Los modelos basados en redes neuronales convolucionales (CNNs) y otros algoritmos de aprendizaje profundo son capaces de clasificar las lesiones de manera uniforme y precisa, mejorando la reproducibilidad diagnóstica entre diferentes clínicos y entornos. Esta estandarización es esencial para facilitar la comparación de estudios y garantizar la consistencia en la práctica clínica (9).

4.4.3. Aumento del acceso a diagnósticos de calidad

Las aplicaciones móviles y las plataformas basadas en IA democratizan el acceso a diagnósticos de alta calidad, incluso en regiones con recursos limitados. Esto permite a los odontólogos generales realizar evaluaciones preliminares precisas (11).

4.4.4. Identificación precoz antes del colapso post-eruptivo

La IA permite detectar signos iniciales de MIH antes de que las fracturas en el esmalte sean evidentes. Esto es posible gracias al análisis a nivel de píxeles que identifica incluso las alteraciones más pequeñas en las imágenes dentales (8,9).

4.4.5. Integración con tecnologías avanzadas

La IA se puede integrar con tecnologías como cámaras intraorales y escáneres digitales para facilitar el diagnóstico en tiempo real. Además, el uso de plataformas basadas en la nube permite la gestión y el análisis de grandes bases de datos de imágenes, mejorando la accesibilidad y la colaboración entre profesionales de diferentes ubicaciones. La integración con sistemas de gestión clínica también puede optimizar el flujo de trabajo en consultorios dentales (10,12).

4.5. Desventajas en la implementación de la IA

A pesar de sus ventajas, la implementación de la IA enfrenta varios desafíos, entre ellos:

4.5.1. Necesidad de datos de entrenamiento calidad y en cantidad

Los modelos de IA requieren grandes volúmenes de imágenes dentales bien etiquetadas y de alta calidad (como imágenes clínicas estandarizadas) para garantizar su correcto entrenamiento y precisión. Sin embargo, la recopilación de estas imágenes y datos puede ser costosa y llevar tiempo, suele ser un desafío ya que pueden variar en resolución, iluminación y ángulo de captura (8-9).

Estas inconsistencias limitan la precisión del modelo y su capacidad de generalización. Además, la falta de representatividad de datos provenientes de diversas regiones geográficas puede afectar la utilidad global de los modelos (9,10,12).

4.5.2 Aceptación por parte de profesionales

Algunos odontólogos son reacios a adoptar la IA debido a la falta de familiaridad con la tecnología o preocupaciones sobre su fiabilidad. La capacitación y sensibilización es esencial para superar estas barreras, ya que muchos odontólogos carecen de formación específica para integrar estas herramientas en la práctica clínica.

Esto destaca la necesidad de programas de capacitación que permitan a los profesionales utilizar la IA de manera efectiva y con confianza (8,11,12).

Los modelos de IA, especialmente aquellos basados en aprendizaje profundo, funcionan como "cajas negras", lo que significa que las decisiones que toman no siempre son comprensibles para los clínicos. Esta falta de interpretabilidad genera desconfianza entre los profesionales de la salud, quienes pueden dudar en basar decisiones críticas en estas herramientas (10,12).

4.5.3 Aspectos éticos y legales

La privacidad de los datos y la responsabilidad por errores en el diagnóstico son cuestiones críticas que deben abordarse antes de la adopción generalizada de estas tecnologías. Aunque se han desarrollado marcos regulatorios para otras áreas de la salud, la odontología aún carece de directrices específicas para implementar modelos de IA de manera segura y ética (9). Garantizar la transparencia y la protección de datos es fundamental para fomentar la confianza en estas tecnologías (12).

4.5.4. Desafíos técnicos

Los modelos de IA también enfrentan desafíos técnicos, como la necesidad de grandes recursos computacionales y el tiempo de procesamiento requerido para entrenar algoritmos complejos. Además, las tecnologías actuales tienen limitaciones al manejar datos heterogéneos o imágenes de baja calidad, lo que afecta la precisión y la eficiencia en entornos clínicos reales (8,12).

5. JUSTIFICACIÓN E HIPÓTESIS

5.1 Hipótesis

Este estudio analiza el impacto de la inteligencia artificial (IA) en el diagnóstico de la hipomineralización incisivo molar (MIH) y compara su rendimiento con los métodos tradicionales utilizados por los odontólogos. El objetivo es evaluar si la IA mejora la precisión, la eficiencia y la fiabilidad del diagnóstico.

- **Hipótesis principal (H1):** La IA permite un diagnóstico más preciso y eficiente de la MIH en comparación con los métodos tradicionales empleados por los clínicos. Esta hipótesis se basa en la capacidad de los algoritmos de aprendizaje profundo, en particular las redes neuronales convolucionales (CNN), para analizar imágenes dentales con alta sensibilidad y especificidad, reduciendo los errores humanos y facilitando la detección temprana.
- **Hipótesis nula (H0):** No existe una diferencia significativa entre el diagnóstico de la MIH realizado por un odontólogo mediante métodos convencionales y aquel obtenido a través de la IA. Esta hipótesis se confirmaría si los resultados muestran que el desempeño de los modelos de IA no supera al de los profesionales en términos de precisión, sensibilidad y reproducibilidad.

Con esta investigación, se pretende determinar si la IA puede convertirse en una herramienta complementaria o incluso superior a los métodos tradicionales para el diagnóstico de la MIH, contribuyendo así a una mejor atención y tratamiento de los pacientes.

5.2 Justificación

La hipomineralización de los molares e incisivos (MIH) afecta aproximadamente al 13.5 % de la población mundial, con rangos que varían entre el 2.4 % y el 40 %, dependiendo de la región y los métodos de diagnóstico utilizados (1). Esta condición no solo genera hipersensibilidad dental y un mayor riesgo de caries, sino que también tiene un impacto significativo en la calidad de vida de los niños afectados, especialmente en términos de autoestima y bienestar social (3,5). La alta prevalencia y las consecuencias clínicas hacen que esta patología sea una prioridad en los programas de salud oral infantil (6,7).

El diagnóstico de la MIH a menudo depende de la inspección visual, lo que introduce subjetividad y variabilidad interobservador. Además, la falta de criterios uniformes limita la comparabilidad entre estudios y dificulta la implementación de estrategias globales de manejo clínico (1). La detección temprana también es un desafío, ya que los defectos pueden pasar desapercibidos antes de que ocurra la fractura del esmalte (1).

La inteligencia artificial (IA) ofrece soluciones prometedoras para abordar las limitaciones tradicionales en el diagnóstico de la MIH. Los algoritmos de aprendizaje profundo, como las redes neuronales convolucionales, han demostrado una precisión diagnóstica superior al 90 %, incluso en casos complejos (1,6). Además, la IA permite la detección temprana de opacidades demarcadas y fracturas incipientes, lo que facilita intervenciones preventivas oportunas y reduce la carga clínica en etapas avanzadas (5,6).

A pesar de los avances en el desarrollo de herramientas diagnósticas, persisten importantes lagunas en la investigación, particularmente en la estandarización de criterios diagnósticos y la implementación de tecnologías de IA en entornos clínicos reales (2,10). Este estudio busca cerrar estas brechas evaluando la efectividad de modelos basados en IA en comparación con métodos convencionales, proporcionando una base para la adopción de estas tecnologías a nivel global (3,6).

6. OBJETIVOS

Objetivo principal:

1. Comparar la efectividad de los modelos de IA con los métodos convencionales utilizados por odontólogos en términos de sensibilidad, especificidad y tiempo de diagnóstico.

Objetivo secundario:

1. Analizar la precisión diagnóstica de los modelos de inteligencia artificial en la identificación de MIH mediante imágenes clínicas y radiográficas.
2. Identificar la efectividad del uso de inteligencia artificial en el diagnóstico de MIH, considerando aspectos como interpretabilidad, aplicabilidad clínica y aceptación por parte de los profesionales de la salud.

Este estudio está alineado con los Objetivos de Desarrollo Sostenible (ODS), en particular con:

ODS 3: (Salud y Bienestar): Mejora del acceso a diagnósticos precisos y tratamientos efectivos para los niños con MIH, reduciendo las desigualdades en salud (6).

- Meta 3.4: Mejorar la prevención, detección y tratamiento de enfermedades no transmisibles, como las afecciones dentales, mediante tecnologías innovadoras como la inteligencia artificial.
- Meta 3.8: Garantizar el acceso a servicios de salud esenciales y de calidad, promoviendo el uso de IA para diagnósticos más precisos y accesibles en odontología.

ODS 4: (Educación de Calidad): Promoción de la capacitación en tecnologías avanzadas, como la IA, para odontólogos en formación y en ejercicio (1,12).

- Meta 4.3: Fomentar la capacitación de profesionales de la salud en el uso de inteligencia artificial para mejorar la precisión diagnóstica en odontología.
- Meta 4.7: Integrar el aprendizaje de nuevas tecnologías en los programas de formación en odontología y salud bucal.

7. MATERIAL Y MÉTODO

La presente revisión sistemática se llevó a cabo siguiendo la declaración de la Guía PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses). La declaración de la guía PRISMA se presenta en la tabla 2 en el apartado de anexos.

7.1. Identificación de la pregunta PICO

Se utilizó el registro internacional de las revisiones sistemáticas (PROSPERO) para confirmar que no existía una revisión similar al estudio realizado.

Se utilizaron las bases de datos Medline-PubMed (United States National Library of Medicine), Web of Science y Scopus para realizar una búsqueda de los artículos indexados sobre pacientes con posible afectación de MIH diagnosticados con la ayuda de la inteligencia artificial, publicados hasta diciembre de 2024 para responder a la siguiente pregunta: **¿Cuál es la precisión y efectividad de los modelos de inteligencia artificial en el diagnóstico de la Hipomineralización de los Molares e Incisivos (MIH) en comparación con los métodos tradicionales utilizados por odontólogos?**

Esta pregunta de estudio se estableció de acuerdo con la pregunta estructurada PICO. El formato de la pregunta se estableció de la siguiente manera:

P (población): Pacientes con sospecha de hipomineralización de los molares e incisivos (MIH).

I (intervención): Diagnóstico mediante modelos de inteligencia artificial.

C (comparación): Diagnóstico realizado por odontólogos utilizando métodos tradicionales.

O (resultados):

O1: Efectividad en la detección de MIH.

O2: Precisión del diagnóstico.

Un resumen de la pregunta PICO se presenta en la tabla 3 en el apartado de anexos.

7.2. Criterios de elegibilidad

Los criterios de inclusión fueron:

- **Tipo de estudio,** según la pirámide metodológica se eligieron los siguientes estudios de inclusión: Estudios clínicos observacionales, incluyendo estudios transversales, de cohortes, casos y controles. Ensayos clínicos aleatorizados y controlados que evalúen el diagnóstico de MIH utilizando inteligencia artificial. Revisiones sistemáticas y metaanálisis enfocados en inteligencia artificial aplicada al diagnóstico dental. Estudios experimentales que involucren modelos de aprendizaje profundo o aprendizaje automático para análisis de imágenes dentales. Se incluyen publicaciones en inglés, español y francés.
- **Tipo de paciente:** Niños y adolescentes con diagnóstico confirmado o sospecha clínica de hipomineralización de molares e incisivos (MIH). Pacientes en edades comprendidas entre los 6 y 18 años, correspondientes al periodo de erupción de los primeros molares e incisivos permanentes.
- **Tipo de intervención:** Uso de herramientas de inteligencia artificial (IA) para el diagnóstico de MIH, incluyendo algoritmos de aprendizaje profundo, redes neuronales convolucionales (CNNs) y sistemas de aprendizaje automático. Evaluación mediante análisis de imágenes dentales clínicas (fotografías intraorales) y radiografías. Comparación entre diagnóstico realizado con IA y métodos tradicionales (evaluación clínica visual realizada por odontólogos).
- **Tipo de variables de resultados:** Estudios que proporcionan datos relacionados con la precisión diagnóstica de la inteligencia artificial (sensibilidad, especificidad y valores predictivos positivos/negativos) así como la eficiencia en el tiempo para realizar el diagnóstico. Y como variables secundarias: la tasa de concordancia entre el diagnóstico realizado por IA y los métodos tradicionales, el impacto clínico del diagnóstico temprano en la planificación y tratamiento.

Criterios de exclusión fueron:

- Pacientes con otras patologías dentales que puedan interferir con el diagnóstico, que no sean del tipo desmineralización y solo en primeros molares permanentes.
- Excluimos aquellos artículos de más de 10 años de antigüedad y aquellos que no utilizan métodos de diagnóstico con imágenes.

7.3 Fuente de información y estrategia de la búsqueda de datos

Se llevó a cabo una búsqueda automatizada en las tres bases de datos anteriormente citadas (PubMed, Scopus y Web of Science) con las siguientes palabras clave: “Molar incisive hypomineralisation”, “MIH”, “Artificial intelligence”, “AI”, “Diagnosis”, “Diagnosis imaging”, “Deep learning”, “CNN”, “Convolutional Neural Network”. Las palabras claves fueron combinadas con los operadores booleanos AND, OR y NOT, así como con los términos controlados (“MeSH” para Pubmed) en un intento de obtener los mejores y más amplios resultados de búsqueda.

La búsqueda en Pubmed fue la siguiente: (("molar hypomineralization"[MeSH Terms] OR "molar incisor hypomineralization"[Text Word] OR "MIH"[Text Word]) AND ("artificial intelligence"[MeSH Terms] OR "deep learning"[Text Word] OR "neural networks"[Text Word] OR ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields])))Filters: in the last 10 years, Full text, English, Spanish.

La búsqueda en SCOPUS fue la siguiente: (TITLE-ABS-KEY ("molar hypomineralization") OR TITLE-ABS-KEY ("molar incisor hypomineralization") OR TITLE-ABS-KEY ("MIH") AND TITLE-ABS-KEY ("artificial intelligence") OR TITLE-ABS-KEY ("AI artificial intelligence") OR TITLE-ABS-KEY ("machine learning") OR TITLE-ABS-KEY ("deep learning") OR TITLE-ABS-KEY (neural) AND TITLE-ABS-KEY ("diagnostic imaging") OR TITLE-ABS-KEY ("diagnosis") OR TITLE-ABS-KEY ("computer-assisted diagnosis")).

La búsqueda en Web of Science fue la siguiente: ((ALL= (("molar hypomineralization" OR "molar incisor hypomineralization" OR "MIH"))) AND ALL= (("artificial intelligence" OR "deep learning" OR "neural networks" OR "artificial intelligence"))).

Un resumen de las búsquedas de cada una de las bases consultadas se presenta en la tabla 3 en el apartado de anexos.

7.3.1 Búsqueda cruzada

Con el fin de identificar cualquier estudio elegible que la búsqueda inicial podría

haber perdido, se completó la búsqueda con una revisión de las referencias proporcionadas en la bibliografía de cada uno de los estudios.

7.4. Proceso de selección de los estudios

Se realizó un proceso de selección en tres etapas. La selección de los estudios fue llevada a cabo por dos revisores (VV, ISR). En la primera etapa se filtró por títulos con el objetivo de eliminar publicaciones irrelevantes aplicando combinaciones de palabras clave y operadores booleanos para optimizar la recuperación de artículos. En la segunda etapa se realizó el cribado por los resúmenes. Se excluyeron aquellos estudios que no abordaban directamente el diagnóstico de la MIH mediante inteligencia artificial o que no cumplían con los criterios de elegibilidad. En la tercera etapa se filtraron los artículos según texto completo y se procedió a la extracción de datos previamente elaborado para confirmar la elegibilidad de los estudios. Los desacuerdos entre los revisores, en cada una de las fases, se resolvieron mediante discusión y, si fuese necesario, la consultación de un tercer revisor.

7.5. Extracción de datos

- **Datos generales del estudio:** Autor(es) y año de publicación, diseño del estudio (observacional, ensayo clínico, revisión sistemática, etc.).
- **Características de la población:** Número de participantes, edad y distribución por género, criterios de inclusión y exclusión aplicados en cada estudio.
- **Métodos de diagnóstico utilizados:** Tipo de herramientas de inteligencia artificial utilizadas (redes neuronales convolucionales, aprendizaje profundo, etc.). Métodos tradicionales de diagnóstico utilizados como referencia.
- **Criterios diagnósticos aplicados:** Parámetros técnicos de las imágenes utilizadas (tipo de imagen, resolución, software empleado).
- **Resultados principales:** Sensibilidad y especificidad de los modelos de IA en comparación con los métodos tradicionales. Precisión diagnóstica (valores predictivos positivos y negativos). Tasa de concordancia entre IA y diagnóstico clínico convencional.
- **Evaluación del impacto clínico:** Beneficios clínicos identificados con el uso de IA. Limitaciones y desafíos reportados en los estudios.

Un resumen de la extracción de datos se presenta en la tabla 4 en el apartado de anexos.

7.6. Valoración de la calidad

La valoración del riesgo de sesgo fue evaluada por dos revisores (VV, ISR) con el objetivo de analizar la calidad metodológica de los artículos incluidos. Se examinó el riesgo de sesgo considerando factores como la metodología de selección de pacientes, el estándar de referencia utilizado y la validez de los métodos de inteligencia artificial empleados.

Se clasificaron los estudios en categorías de bajo, moderado o alto riesgo de sesgo, y se tomaron medidas para minimizar la influencia de sesgos en la síntesis de resultados.

Para la evaluación de la calidad de los estudios clínicos controlados aleatorizados se utilizó la guía Cochrane 5.1.0 (<http://handbook.cochrane.org>); las publicaciones fueron consideradas de «bajo riesgo de sesgo» cuando cumplían todos los criterios, «alto riesgo de sesgo» cuando no se cumplen uno o más criterios y «riesgo de sesgo incierto» por falta de información o incertidumbre sobre el potencial de sesgo.

La calidad de reportes de casos se evaluó mediante la escala de Joanna Briggs Institute (JBI) (<http://joannabriggs.org/research/critical-appraisal-tools.html>). Las publicaciones fueron consideradas «bajo riesgo de sesgo» cuando cumplían todos los criterios, «alto riesgo de sesgo» cuando no se cumplen uno o más criterios y «riesgo de sesgo incierto» por falta de información o incertidumbre sobre el potencial de sesgo.

Los estudios de series de casos se evaluaron mediante la escala MOGA (<https://cobe.paginas.ufsc.br/files/2014/10/MOGA.Case-series.pdf>). Un resultado de ≥ 6 se consideró un «alto riesgo de sesgo» y un resultado de < 6 se consideró un «bajo riesgo de sesgo».

7.7. Síntesis de datos

Los datos extraídos de los estudios seleccionados fueron sintetizados de manera cualitativa y cuantitativa. Se realizó una comparación de los resultados en términos de precisión diagnóstica, concordancia entre IA y métodos tradicionales. Se

utilizó un enfoque narrativo para resumir las principales tendencias observadas. Se compararon las medidas de sensibilidad y especificidad entre diferentes estudios. Se discutieron los hallazgos en relación con las limitaciones metodológicas de los estudios. Los hallazgos obtenidos proporcionan una visión integral de la efectividad de la inteligencia artificial en el diagnóstico de la MIH y sus implicaciones clínicas y metodológicas.

8. RESULTADOS

8.1. Selección de estudios. Flow chart

Del proceso de búsqueda inicial se obtuvo un total de 68 artículos: Medline - PubMed (n=8), SCOPUS (n=8) y Web of Science (n=51). Además, se obtuvieron 6 estudios adicionales mediante búsqueda manual (lista de referencias y fuentes primarias). De estas publicaciones, 13 se identificaron como artículos potencialmente elegibles mediante el cribado por títulos y resúmenes. Posteriormente, se obtuvieron los artículos a texto completo y se evaluaron minuciosamente. Como resultado, 8 artículos cumplieron los criterios de inclusión y se incluyeron en la presente revisión sistemática (Fig. 2). La información relativa a los artículos excluidos (y las razones de su exclusión) se presenta en la Tabla 5 en anexos.

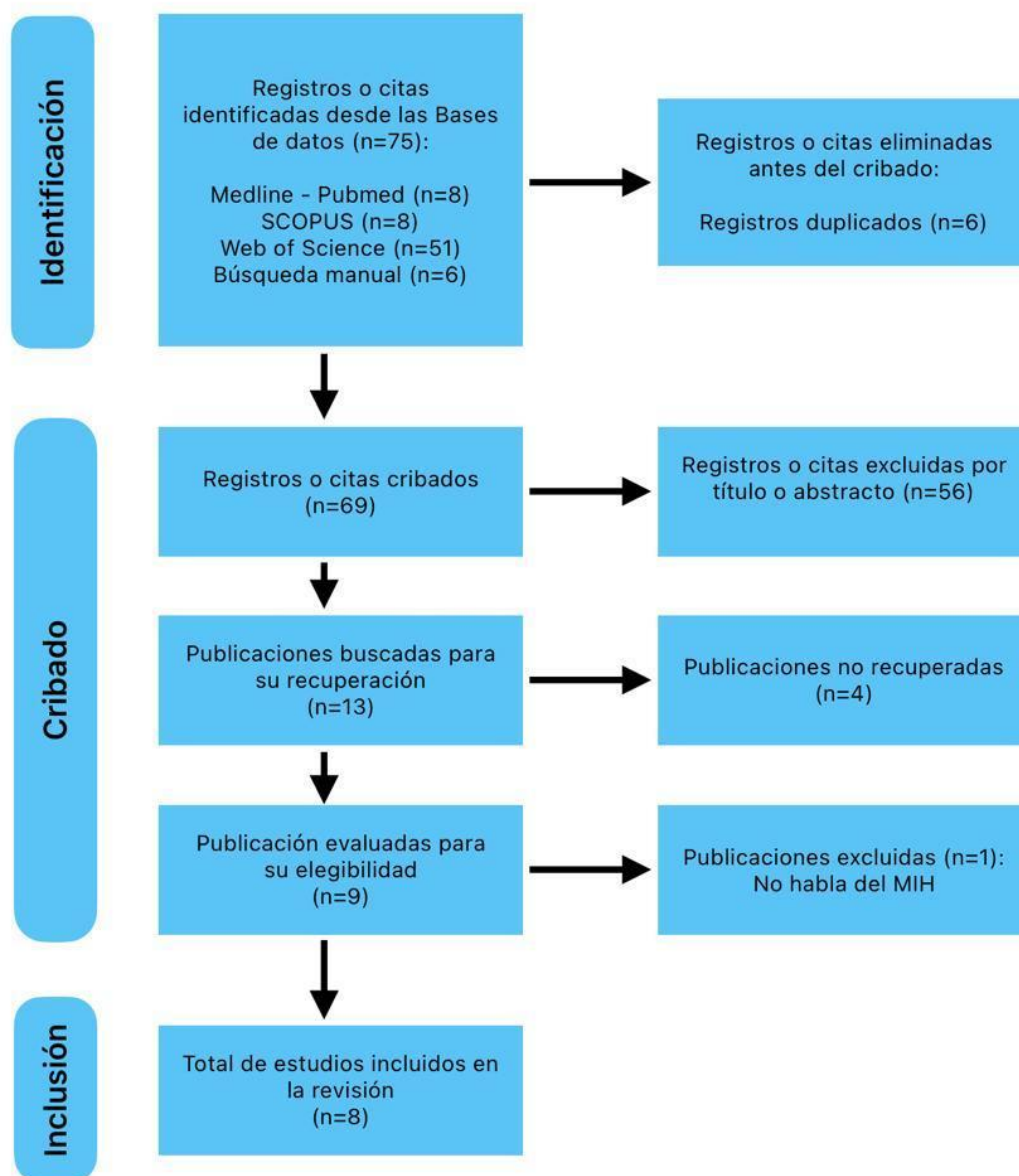


Fig. 2. Diagrama de flujo de búsqueda y proceso de selección de títulos durante la revisión sistemática.

8.2. Análisis de las características de los estudios revisados

En esta revisión sistemática se incluyeron un total de 8 estudios publicados entre 2022 y 2025, que abordaron el uso de inteligencia artificial (IA) en el diagnóstico de la hipomineralización molar-incisiva (MIH) a partir de imágenes clínicas y dentales.

En cuanto al diseño metodológico de los estudios, se observó una diversidad significativa. Cuatro estudios fueron observacionales no randomizados, centrados en el análisis clínico de imágenes dentales en contextos reales o simulados (8,9,12,13). Tres estudios fueron clasificados como in vitro, destacando por el uso de protocolos

experimentales controlados en laboratorio (10,14,15). Finalmente, se incluyó un estudio de carácter descriptivo en formato de carta al editor, que, si bien no proporciona métricas cuantitativas, ofrece una visión práctica sobre la aplicabilidad clínica de la IA (16).

Las muestras analizadas fueron muy variables entre los diferentes trabajos. Se reportaron desde bases reducidas con 50 imágenes estandarizadas de niños diagnosticados con MIH (12), hasta grandes bases de datos de más de 18.000 imágenes clínicas dentales (10). En total, la revisión reúne una evidencia significativa basada en el análisis de más de 25.000 imágenes, todas ellas procesadas digitalmente bajo distintos criterios de calidad y resolución.

En relación con los modelos de inteligencia artificial utilizados, se emplearon principalmente arquitecturas de redes neuronales convolucionales (CNN), como ResNeXt-101, VGG-19, ResNet y DenseNet (14,15), así como modelos híbridos con algoritmos de aprendizaje supervisado y herramientas de procesamiento de imagen como OpenCV y Segment Anything Model (SAM) (12,13). Algunos estudios aplicaron modelos tipo transformador como SegFormer-B5 (10), reflejando un avance técnico relevante en el análisis automatizado de imágenes.

La validación de los modelos también varió según el enfoque de cada estudio. Algunos realizaron validaciones internas con sus propios conjuntos de datos (8,10,15), mientras que otros compararon sus resultados con evaluaciones clínicas realizadas por expertos humanos, como fue el caso de los artículos de: Neumayr y cols., He y cols. y Veseli y cols. En el estudio de Jaiswal et al. la evaluación de la precisión diagnóstica se realizó mediante comparación con mediciones en Adobe Photoshop, considerada una herramienta de referencia para análisis digital.

Por último, el tipo de imagen empleado fue clínico en todos los estudios. Se utilizaron fotografías intraorales, imágenes oclusales de alta resolución, y capturas digitales en múltiples vistas de la cavidad oral, todas ellas adecuadamente etiquetadas por expertos para el entrenamiento o validación de los modelos. Esta consistencia en el tipo de dato analizado fortalece la aplicabilidad de los modelos propuestos a escenarios reales de diagnóstico odontológico.

Un resumen de las características de los estudios incluidos se presenta en la tabla 6 en el apartado de anexos.

8.3. Evaluación de la calidad metodológica y riesgo de sesgo

Para los estudios observacionales no randomizados, 1 de riesgo moderado y 2 artículos fueron considerados de bajo riesgo de sesgo según la escala Newcastle-Ottawa (Tabla 7). En cambio, un artículo fue clasificado con alto riesgo de sesgo, principalmente por la falta de grupo control y limitada comparabilidad.

Para los estudios in vitro, evaluados mediante la escala modificada ARRIVE/CONSORT (Tabla 8), todos presentaron buena calidad metodológica. Las principales debilidades fueron la falta de justificación muestral y la omisión de condiciones ambientales del laboratorio.

En el caso del estudio de Veseli y cols., al tratarse de una carta al editor de carácter descriptivo, no se puede aplicar una escala estandarizada para valorar el riesgo de sesgo. No se presentan datos cuantitativos ni se detalla una metodología experimental estructurada, lo que limita la evaluación objetiva de la validez interna. Por tanto, se considera que el riesgo de sesgo es alto debido a la falta de control metodológico, ausencia de grupo comparativo y dependencia del juicio clínico sin validación estadística.

	Definición de los casos	Representatividad	Selección de los controles	Definición de los controles	Comparabilidad (factor más importante)	Comparabilidad (cualquier otra variable)	Comparación de la exposición	Mismo método para ambos grupos	Tasa de abandonos	Total
Schönewolf y cols., 2022 (8)	★	★	-	-	★	-	★	★	★	6
Kühnisch y cols., 2024 (9)	★	★	★	★	★	★	★	★	★	9
Jaiswal y cols., 2024 (12)	★	-	-	-	-	-	★	-	-	2
He y cols., 2025 (13)	★	★	★	★	★		★	★	★	8

Tabla 7: Medición del riesgo de sesgo de los estudios observacionales no randomizados con la escala Newcastle-Ottawa – estudios observacionales con grupo control no randomizado.

	Criterio	Portella y cols., 2023 (15)	Alevizakos y cols., 2022 (14)	Felsch y cols., 2023 (10)
1	¿Se define claramente el objetivo del estudio desde el resumen o la introducción?	✓	✓	✓
2	¿Se describe adecuadamente el modelo in vitro utilizado (tipo de muestras, origen)?	✓	✓	✓
3	¿Se especifica el procedimiento experimental completo paso a paso?	✓	✓	✓
4	¿Se incluyen características del entorno de laboratorio (luz, temperatura, etc.)?	x	x	✓
5	¿Se justifica el tamaño muestral o se refiere a estudios previos?	x	x	✓
6	¿Se controlaron sesgos (aleatorización de muestras, cegamiento del evaluador, etc.)?	✓	✓	✓
7	¿Se define con claridad la variable principal de resultado?	✓	✓	✓
8	¿Las herramientas de medición están validadas o estandarizadas?	✓	✓	✓
9	¿Se repitió el experimento (reproducibilidad / replicabilidad) y se informa el nº de repeticiones?	✓	✓	✓
10	¿Se describen los análisis estadísticos usados y son adecuados?	✓	✓	✓
11	¿Los resultados están claramente expresados en tablas, gráficos o porcentajes claros?	✓	✓	✓
12	¿Se discuten claramente las limitaciones del estudio?	✓	✓	✓
13	¿Se declaran fuentes de financiación y conflictos de interés?	✓	✓	✓
	Puntaje	11	11	13

Tabla 8: Estudios in vitro – Escala ARRIVE/CONSORT Modificada

8.4. Síntesis resultados

8.4.1. Resultados diagnósticos de los modelos de IA

Los resultados obtenidos en los estudios incluidos demuestran que los modelos de inteligencia artificial aplicados al diagnóstico de la hipomineralización molar-incisiva (MIH) alcanzan en general altos niveles de precisión y rendimiento diagnóstico. Sin embargo, existen diferencias considerables en las métricas reportadas y en los métodos de validación empleados.

El estudio de Schönewolf y cols., que utilizó un modelo basado en ResNeXt-101, reportó una precisión general del 95.2%, con una sensibilidad del 78.6% y una especificidad del 97.3%, lo que indica un excelente rendimiento en la identificación de casos negativos, pero una menor capacidad para detectar ciertas categorías complejas de MIH. De forma similar, el modelo validado externamente por Kühnisch y cols. mostró una precisión del 94.3%, con una sensibilidad entre 81.7% y 92.8%, y una especificidad que osciló entre el 91.9% y el 98.7%, destacándose por su acuerdo clínico del 89.2% con el juicio de expertos.

Por su parte, el estudio de Felsch y cols. centrado en un modelo tipo transformer (SegFormer-B5), obtuvo una precisión global del 97.8% y valores de precisión promedio (AP) entre el 82.7% y el 90.2%, lo que refleja un rendimiento excelente en términos de segmentación y localización de lesiones. Aunque no se reportaron valores de sensibilidad ni especificidad desglosados, se destacó un Intersección sobre Unión (IoU) de 95.9%, indicador de una alta concordancia entre la predicción del modelo y las áreas reales de lesión.

En otros estudios, como el de Portella et d. se reportó una precisión del 87.9%, con sensibilidad del 82.7%, especificidad del 80% y un F1-score de 0.887, valores que superaron significativamente el rendimiento de los evaluadores humanos en el diagnóstico de lesiones cariosas incipientes. Igualmente He y cols. evidenció mejoras importantes en el rendimiento diagnóstico asistido: los dentistas junior incrementaron su F1-score desde 0.713 hasta 0.949 con el apoyo del modelo basado en Mask R-CNN + SAM, lo que refuerza el valor de la IA como herramienta complementaria en la práctica clínica.

Los autores Alevizakos y cols. y Veseli y cols., en sus estudios no presentan métricas detalladas de sensibilidad o especificidad, pero indican una alta efectividad estimada

del modelo. Se destacó que el modelo DenseNet ofreció los mejores resultados entre varias arquitecturas evaluadas, con una precisión que osciló entre 83.98% y 92.86%. Finalmente, el estudio Jaiswal y cols., no reportó directamente parámetros de rendimiento como sensibilidad o F1-score, ya que su objetivo principal fue cuantificar la extensión de las lesiones de MIH a través de imágenes estandarizadas. El modelo mostró un valor medio de extensión del 34.51%, comparable al obtenido mediante software Photoshop (35.32%), lo que evidencia una concordancia satisfactoria.

Autor (Año)	Precisión (%)	Sensibilidad (%)	Especificidad (%)	F1-Score	Valor Predictivo Positivo	Hallazgos Relevantes
Schönewol y cols., 2022	95.2	78.6	97.3	No especificado	No indicado	Alta precisión general; menor sensibilidad en casos complejos.
Kühnisch y cols., 2024	94.3	81.7–92.8	91.9–98.7	No indicado	No indicado	Acuerdo 89.2% con expertos; buen desempeño.
Felsch y cols., 2023	97.8	No específico	No específico	AP: 82.7–90.2	No indicado	Rendimiento excelente, IoU: 95.9%.
Jaiswal y cols., 2024	Media 34.51%	No aplica	No aplica	No aplica	No aplica	Extensión >60% se asocia con 80% preocupación estética.
He y cols., 2025	No directo	93.5	78.9	0.856	0.968	IA mejora F1 de 0.713 a 0.949 con ayuda.
Alevizakos y cols., 2022	83.98–92.86	No indicado	No indicado	No indicado	No indicado	Mejor modelo: DenseNet121.
Portella y cols., 2023	87.9	82.7	80.0	0.887	0.949	Mejora significativa en precisión con IA.
Veseli y cols., 2024	No indicado	No indicado	No indicado	No indicado	No indicado	Alto rendimiento estimado, sin métricas exactas.

Tabla 9: Resultados diagnósticos de los estudios

8.4.2 Complicaciones relacionadas con el procedimiento de diagnóstico asistido por IA

A diferencia de los procedimientos clínicos invasivos, las técnicas de diagnóstico asistido por inteligencia artificial (IA) no presentan complicaciones físicas directas para los pacientes. No obstante, los estudios analizados identifican diversas limitaciones técnicas y metodológicas que pueden considerarse como “complicaciones” en la implementación y aplicación de estos sistemas en entornos clínicos reales.

Una de las principales complicaciones fue reportada por Schönewolf y cols.,

quienes observaron una disminución significativa de la sensibilidad en subcategorías complejas, como las lesiones con ruptura de esmalte y presencia de sellantes, lo que puede llevar a falsos negativos clínicamente relevantes. De manera similar, Jaiswal y cols. destacó que los artefactos de reflejo especular en las fotografías reducen la precisión diagnóstica hasta en un 57 % cuando superan el 15 % de la superficie total de la imagen, lo cual compromete la fiabilidad del sistema.

Por otro lado, Felsch y cols. y Alevizakos y cols. señalaron la ausencia de validación externa como una limitación importante, lo que plantea el riesgo de sobreajuste de los modelos y sugiere la necesidad de nuevas pruebas en contextos clínicos más variados. Además, Veseli y cols. no proporciona métricas cuantitativas de rendimiento, lo que impide una evaluación objetiva y estandarizada del modelo propuesto.

Otra complicación destacada fue la variabilidad del rendimiento según la zona anatómica. He y cols. documentó una menor efectividad en la detección de lesiones en molares inferiores ($F1 = 0.727$), en comparación con premolares superiores ($F1 = 0.963$), lo cual puede limitar la aplicabilidad universal del sistema si no se ajusta a las diferencias morfológicas dentales.

Finalmente, algunos estudios, como el de Portella y cols., sugieren que la efectividad del diagnóstico asistido por IA puede depender en gran medida del nivel de experiencia del clínico que lo utilice. Si bien la IA mejora el rendimiento de estudiantes o profesionales menos experimentados, su valor añadido es menor en especialistas altamente entrenados.

Un resumen de las complicaciones o limitaciones encontradas en los estudios incluidos se presenta en la tabla 10 en el apartado de anexos.

9. DISCUSIÓN

9.1. Precisión diagnóstica y rendimiento general

La eficacia de los modelos de IA desarrollados para el diagnóstico de lesiones del esmalte como la hipomineralización molar-incisiva (MIH) ha sido un área de creciente exploración en los últimos años. Los resultados de esta revisión sistemática, basada en nueve estudios relevantes, mostraron que la precisión general de los modelos osciló entre el 83.98 % y el 97.8 %. Los modelos más precisos, como el propuesto por Felsch y cols., utilizaron arquitecturas modernas como transformadores visuales y bases de datos de imágenes clínicas muy amplias.

Schönewolf y cols., empleando una red neuronal convolucional (ResNeXt-101), alcanzó una precisión del 95.2 %, mientras que Kühnisch y cols., a través de una validación externa rigurosa, demostró la fiabilidad clínica del modelo al concordar en un 89.2 % con el juicio de expertos. En cuanto a la sensibilidad, los modelos lograron identificar correctamente lesiones entre el 78.6 % y el 93.5 % de los casos, mostrando así su utilidad en la detección precoz del MIH (8,15). La especificidad fue igualmente elevada, superando el 90 % en varios estudios (8,9).

Estas cifras reflejan un rendimiento comparable o incluso superior al de evaluadores humanos en entornos clínicos, y evidencian el papel potencial de la IA como herramienta diagnóstica complementaria, especialmente útil en contextos con limitaciones de tiempo o de recursos.

9.2. Complicaciones técnicas y limitaciones metodológicas

A pesar del alto rendimiento, la mayoría de los estudios revisados señalaron diferentes complicaciones metodológicas y limitaciones técnicas (8,10,12,13,14,15,16). Entre las más destacadas figura la dependencia de la calidad de la imagen. Jaiswal y cols. mostró que la presencia de reflejos especulares en las fotografías (>15%) podía reducir la precisión diagnóstica hasta en un 43 %, lo que pone de manifiesto la necesidad de condiciones estandarizadas de captura.

Asimismo, se identificaron diferencias de rendimiento según la localización anatómica del diente: se reportó una caída significativa del F1-score en molares inferiores frente a premolares superiores (13). Otros estudios carecieron de validación

externa, limitando la generalización de sus resultados (10,16). Además, la experiencia del clínico también fue un factor condicionante: se evidenció que la IA era más útil para estudiantes que para profesionales experimentados (15).

Estas complicaciones subrayan la necesidad de desarrollar algoritmos robustos, validados en entornos clínicos reales y adaptables a distintos perfiles de usuario.

9.3. Aplicaciones clínicas, formativas y perspectivas futuras

Uno de los aportes más relevantes de esta revisión es la demostración del potencial que la IA tiene para apoyar tanto el diagnóstico clínico como la formación de profesionales en odontología. Algunos modelos fueron diseñados no solo para detectar la presencia de MIH, sino también para cuantificar la extensión de la lesión (12) o diferenciar MIH de otras patologías similares como caries, fluorosis o amelogenénesis imperfecta (14).

Además, varios autores propusieron la implementación de estas tecnologías en herramientas accesibles como aplicaciones móviles, plataformas web interactivas o escáneres intraorales. Esta accesibilidad tecnológica podría democratizar el diagnóstico de calidad y reducir desigualdades en salud bucodental infantil. Asimismo, su uso en entornos educativos permitiría una curva de aprendizaje más rápida para los estudiantes de odontología (10,14,16).

Sin embargo, para consolidar este impacto, se requiere la validación multicéntrica, la incorporación de datasets heterogéneos y el desarrollo de modelos explicables y éticamente transparentes.

9.4. Limitaciones de la presente revisión

La presente revisión sistemática presenta varias limitaciones metodológicas que deben ser consideradas al interpretar los resultados obtenidos. En primer lugar, la heterogeneidad de los diseños de los estudios incluidos constituye una de las principales restricciones. Se incluyeron trabajos observacionales no aleatorizados, estudios in vitro, estudios de validación interna sin controles clínicos, y una carta al editor de naturaleza descriptiva. Esta variedad impide realizar comparaciones directas entre los resultados y limita la posibilidad de realizar un metaanálisis cuantitativo.

Asimismo, la ausencia de estudios clínicos aleatorizados controlados representa un factor importante que disminuye el nivel de evidencia global de esta revisión. La mayoría de los trabajos analizados se basaron en conjuntos de datos limitados o imágenes clínicas estandarizadas, sin validación externa multicéntrica, lo cual afecta su aplicabilidad real en entornos clínicos diversos.

Otra limitación destacada fue la falta de homogeneidad en las métricas reportadas. Algunos estudios informaron precisión general y sensibilidad, mientras que otros utilizaron métricas como el F1-score, el IoU o simplemente describieron resultados cualitativos sin valores cuantitativos. Esta variabilidad dificultó la comparación transversal entre modelos.

Además, varios artículos presentaron riesgos de sesgo moderados a altos, especialmente en los criterios relacionados con la representatividad de la muestra, la aleatorización de los casos o la falta de cegamiento de los evaluadores, como se evidenció en la evaluación mediante las escalas Newcastle-Ottawa, ARRIVE/CONSORT modificada y MOGA.

Finalmente, se debe señalar que la mayoría de los modelos de IA analizados fueron entrenados y validados con imágenes de alta calidad y en condiciones controladas de laboratorio, lo que puede no reflejar con precisión las condiciones clínicas reales. Esta brecha entre el entorno experimental y el clínico plantea un desafío importante para la generalización de los resultados y su implementación rutinaria en la práctica odontológica.

Por lo tanto, se recomienda realizar futuras investigaciones con diseños metodológicos más robustos, validaciones externas multicéntricas, y evaluaciones longitudinales que permitan estimar con mayor precisión el impacto de la IA en el diagnóstico del MIH en la práctica clínica habitual.

10. CONCLUSIÓN

Conclusiones principales

Los modelos de inteligencia artificial (IA) mostraron una mayor efectividad diagnóstica que los métodos convencionales, con valores más altos de sensibilidad, especificidad y rapidez diagnóstica, especialmente en la detección precoz de la hipomineralización molar-incisiva (MIH).

Conclusiones secundarias

La precisión diagnóstica de la IA en el análisis de imágenes clínicas fue alta, alcanzando valores superiores al 90 % en la mayoría de los estudios, lo que demuestra su fiabilidad en el reconocimiento de lesiones MIH.

El uso de IA demostró una alta aplicabilidad clínica, mejorando la reproducibilidad diagnóstica, especialmente en profesionales menos experimentados, y facilitando la formación, lo cual refuerza su valor como herramienta complementaria en odontología pediátrica.

Contribución a los Objetivos de Desarrollo Sostenible (ODS)

ODS 3: Este estudio respalda el acceso a diagnósticos más precisos y equitativos para niños afectados por MIH, promoviendo la prevención y detección precoz de enfermedades no transmisibles mediante tecnologías innovadoras como la inteligencia artificial (metas 3.4 y 3.8).

ODS 4: A través del análisis del impacto de la IA en la práctica clínica, se evidencia su valor pedagógico para odontólogos en formación y en ejercicio, promoviendo la incorporación de nuevas tecnologías en los programas educativos (metas 4.3 y 4.7).

Así, la implementación clínica de la IA no solo mejora la calidad del diagnóstico, sino que también reduce desigualdades en salud y fortalece la capacitación tecnológica en odontología, pilares fundamentales del desarrollo sostenible.

11. BIBLIOGRAFÍA

- 1) Lopes LB, Machado V, Mascarenhas P, Mendes JJ, Botelho J. The prevalence of molar-incisor hypomineralization: a systematic review and meta-analysis. *SciRep*. 2021 Nov 17;11(1):22405. doi: 10.1038/s41598-021-01541-7. PMID: 34789780; PMCID: PMC8599453.
- 2) Villani FA, Aiuto R, Dioguardi M, Paglia L, Caruso S, Gatto R, Re D, Garcovich D. Caries prevalence and molar incisor hypomineralisation (MIH) in children. Is there an association? A systematic review. *Eur J Paediatr Dent*. 2023 Dec 1;24(4):312-320. doi: 10.23804/ejpd.2023.1985. PMID: 38015112.
- 3) Mazur M, Corridore D, Ndokaj A, Ardan R, Voza I, Babajko S, Jedeon K. MIH and Dental Caries in Children: A Systematic Review and Meta-Analysis. *Healthcare (Basel)*. 2023 Jun 18;11(12):1795. doi: 10.3390/healthcare11121795. PMID: 37372913; PMCID: PMC10298042.
- 4) Rodd HD, Graham A, Tajmehr N, Timms L, Hasmun N. Molar Incisor Hypomineralisation: Current Knowledge and Practice. *Int Dent J*. 2021 Aug;71(4):285-291. doi: 10.1111/idj.12624. PMID: 34286697; PMCID: PMC9275314.
- 5) Alzahrani AY, Alamoudi NMH, El Meligy OAES. Contemporary Understanding of the Etiology and Management of Molar Incisor Hypomineralization: A Literature Review. *Dent J (Basel)*. 2023 Jun 23;11(7):157. doi: 10.3390/dj11070157. PMID: 37504223; PMCID: PMC10378579.
- 6) Lygidakis NA, Garot E, Somani C, Taylor GD, Rouas P, Wong FSL. Best clinical practice guidance for clinicians dealing with children presenting with molar-incisor-hypomineralisation (MIH): an updated European Academy of Paediatric Dentistry policy document. *Eur Arch Paediatr Dent*. 2022 Feb;23(1):3-21. doi: 10.1007/s40368-021-00668-5. PMID: 34669177; PMCID: PMC8926988.
- 7) Enax J, Amaechi BT, Farah R, Liu JA, Schulze zur Wiesche E, Meyer F. Remineralization Strategies for Teeth with Molar Incisor Hypomineralization (MIH): A Literature Review. *Dent J (Basel)*. 2023 Mar 13;11(3):80. doi: 10.3390/dj11030080. PMID: 36975577; PMCID: PMC10047667.
- 8) Schönewolf J, Meyer O, Engels P, Schlickerrieder A, Hickel R, Gruhn V, Hesenius M, Kühnisch J. Artificial intelligence-based diagnostics of molar-incisor-

hypomineralization (MIH) on intraoral photographs. Clin Oral Investig. 2022 Sep;26(9):5923-5930. doi: 10.1007/s00784-022-04552-4. PMID: 35608684;PMCID: PMC9474479.

9) Neumayr J, Frenkel E, Schwarzmaier J, Ammar N, Kessler A, Schwendicke F, Kühnisch J, Dujic H. External validation of an artificial intelligence-based method for the detection and classification of molar incisor hypomineralisation in dental photographs. J Dent. 2024 Sep;148:105228. doi: 10.1016/j.jdent.2024.105228. PMID: 38972447.

10) Felsch M, Meyer O, Schlickerrieder A, Engels P, Schönewolf J, Zöllner F, Heinrich-Weltzien R, Hesenius M, Hickel R, Gruhn V, Kühnisch J. Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model. NPJ Digit Med. 2023 Oct 25;6(1):198. doi: 10.1038/s41746-023-00944-2. PMID: 37880375; PMCID: PMC10600213.

11) Tchito Tchapgá C, Mih TA, Tchagna Kouanou A, Fozin Fonzi T, Kuetché Fogang P, Mezatio BA, Tchiotso D. Biomedical Image Classification in a Big Data Architecture Using Machine Learning Algorithms. J Healthc Eng. 2021 May 30;2021:9998819. doi: 10.1155/2021/9998819. PMID: 34122785; PMCID: PMC8191587.

12) Jaiswal M, Mukhtar U, Shakya KS, Laddi A, Singha LA. Computerised assessment-a novel approach for calculation of percentage of hypomineralized lesion on incisors and its correlation with aesthetic concern. J Oral Biol Craniofac Res. 2024 Sep-Oct;14(5):570-577. doi: 10.1016/j.jobcr.2024.07.004. Epub 2024 Jul 23. PMID: 39139516; PMCID: PMC11320481.

13) He K, Zhang R, Liang M, Tian K, Luo K, Chen R, Ren J, Wang J, Li J The application of deep learning in early enamel demineralization detection. PeerJ. 2025 Jan 2;13:e18593. doi: 10.7717/peerj.18593. PMID: 39763696; PMCID: PMC11700490.

14) Alevizakos V, Bekes K, Steffen R, von See C. Artificial intelligence system for training diagnosis and differentiation with molar incisor hypomineralization (MIH) and similar pathologies. Clin Oral Investig. 2022 Dec;26(12):6917-6923. doi: 10.1007/s00784-022-04646-z. Epub 2022 Sep 6. PMID: 36065023. 27

15) Portella PD, de Oliveira LF, Ferreira MFC, Dias BC, de Souza JF, Assunção LRDS. Improving accuracy of early dental carious lesions detection using deep learning based automated method. Clin Oral Investig. 2023 Dec;27(12):7663-7670. doi: 10.1007/s00784-023-05355-x. Epub 2023 Oct 31. PMID: 37906303.

16) Veseli E. Early detection of MIH in children by using artificial intelligence. Eur Arch Paediatr Dent. 2024 Dec;25(6):899-900. doi: 10.1007/s40368-024-00931- 5. Epub 2024 Aug 1. PMID: 39090506.

12. ANEXOS

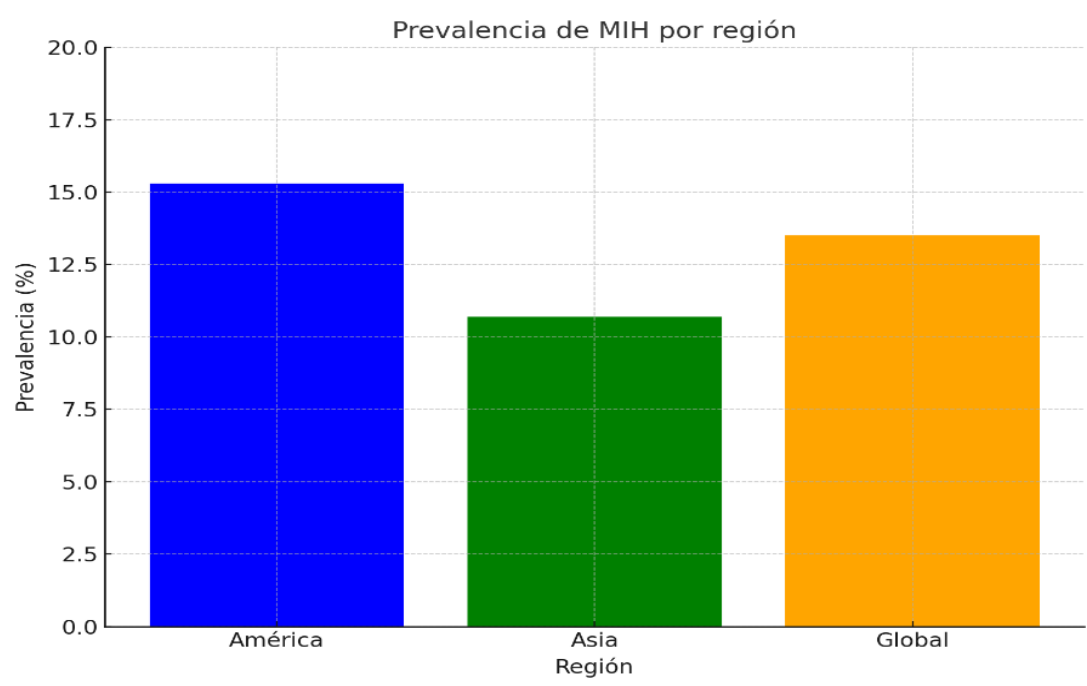


Figura 1: Prevalencia de la MIH por región.

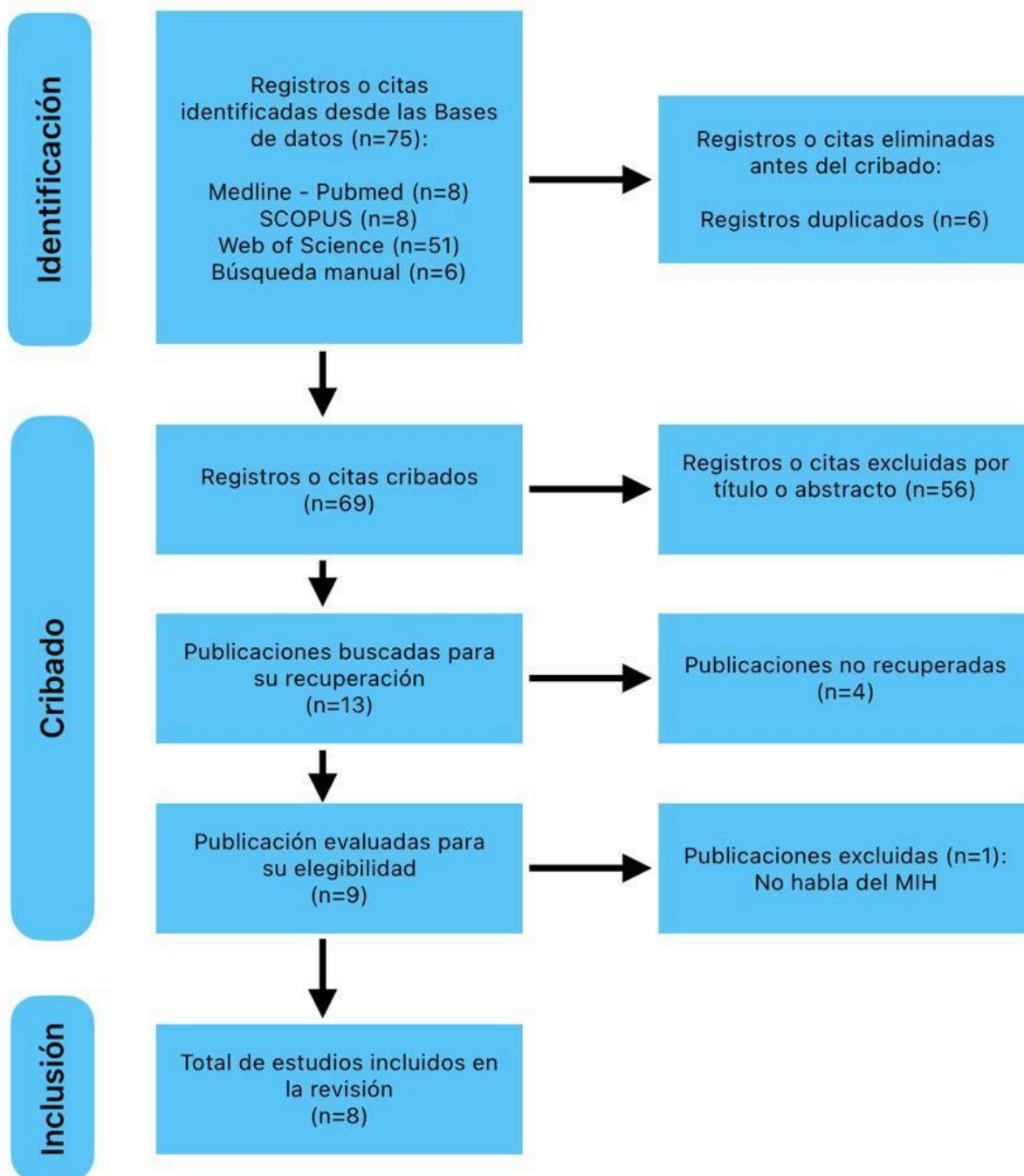


Fig. 2. Diagrama de flujo de búsqueda y proceso de selección de títulos durante la revisión sistemática.

Patología	Características diferenciales
Fluorosis dental	Opacidades difusas, sin bordes bien definidos, afectan varios dientes simétricamente. Coloración blanca tiza o marrón. Se asocia con consumo excesivo de flúor en la infancia.
Hipoplasia del esmalte	Reducción en el grosor del esmalte en vez de defectos en la mineralización. Puede haber surcos o fosas en el esmalte. Se debe a alteraciones en la formación del esmalte, no en su mineralización.
Amelogénesis imperfecta	Afecta todos los dientes primarios y permanentes de manera generalizada. Puede haber esmalte muy delgado, quebradizo o ausente. Es de origen genético.
Caries dental	Lesiones cavitadas con bordes irregulares, suele afectar zonas de retención de placa bacteriana. Evoluciona con pérdida de tejido dental y puede haber dolor.
Defecto traumático o hipoplasia postnatal	Se presentan en dientes individuales, con patrones de afectación localizados y relacionados con trauma o infecciones en la infancia.

Tabla 1: Diagnóstico diferencial para el MIH

P		I		C		O
Pacientes con sospecha de Hipomineralización de los Molares e Incisivos (MIH).		Diagnóstico mediante modelos de inteligencia artificial.		Diagnóstico realizado por odontólogos utilizando métodos tradicionales.		Efectividad en la detección de MIH. Precisión del diagnóstico.
"Molar incisor hypomineralization" OR "MIH" OR "molar hypomineralization"	AND	"Artificial intelligence" OR "AI" OR "deep learning" OR "machine learning" OR "neural networks"	AND	No se especifica directamente en la ecuación de búsqueda, ya que se espera que los estudios comparen IA con métodos tradicionales.	AND	"Diagnostic imaging" OR "diagnosis" OR "computer-assisted diagnosis"

Tabla 2: Pregunta PICO

Base de datos	Búsqueda	Numero de artículos	Fecha
PubMed	((("molar hypomineralization"[MeSH Terms] OR "molar incisor hypomineralization"[Text Word] OR "MIH"[Text Word]) AND ("artificial intelligence"[MeSH Terms] OR "deep learning"[Text Word] OR "neural networks"[Text Word] OR ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields])))Filters: in the last 10 years, Full text, English, Spanish.	8	6/02/25
Scopus	(TITLE-ABS-KEY ("molar hypomineralization") OR TITLE-ABS-KEY ("molar incisor hypomineralization") OR TITLE-ABS-KEY ("MIH") AND TITLE-ABS-KEY ("artificial intelligence") OR TITLE-ABS-KEY ("AI artificial intelligence") OR TITLE-ABS-KEY ("machine learning") OR TITLE-ABS-KEY ("deep learning") OR TITLE-ABS-KEY (neural) AND TITLE-ABS-KEY ("diagnostic imaging") OR TITLE-ABS-KEY ("diagnosis") OR TITLE-ABS-KEY ("computer-assisted diagnosis"))	8	6/02/25
Web of Science	((ALL= ((("molar hypomineralization" OR "molar incisor hypomineralization" OR "MIH")))) AND ALL=((("artificial intelligence" OR "deep learning" OR "neural networks" OR "artificial intelligence"))))	51	6/02/25

Tabla 3: Resumen de todas las búsquedas de cada una de las bases consultadas

Datos generales del estudio	Autor(es) y año de publicación, diseño del estudio (observacional, ensayo clínico, revisión sistemática, etc.).
Características de la población	Número de participantes, edad y distribución por género, criterios de inclusión y exclusión aplicados en cada estudio.
Métodos de diagnóstico utilizados	Tipo de herramientas de inteligencia artificial utilizadas (redes neuronales convolucionales, aprendizaje profundo, etc.). Métodos tradicionales de diagnóstico utilizados como referencia.
Criterios diagnósticos aplicados	Parámetros técnicos de las imágenes utilizadas (tipo de imagen, resolución, software empleado).
Resultados principales:	Sensibilidad y especificidad de los modelos de IA en comparación con los métodos tradicionales. Precisión diagnóstica (valores predictivos positivos y negativos). Tasa de concordancia entre IA y diagnóstico clínico convencional.
Evaluación del impacto clínico	Beneficios clínicos identificados con el uso de IA. Limitaciones y desafíos reportados en los estudios.

Tabla 4 : Resumen de la extracción de datos

Autor y Año	Publicación	Motivo de exclusión
Tchito Tchappa C, Mih TA, Tchagna Kouanou A, Fozin Fonzin T, Kuetché Fogang P, Mezatio BA, Tchiosop D. 2021	Biomedical Image Classification in a Big Data Architecture Using Machine Learning Algorithms	No hace referencia al MIH

Tabla 5: Artículos excluidos (y su razón de exclusión) de la presente revisión sistemática.

Autor (Año)	Tipo de Estudio	Muestra	Modelo de IA	Validación	Tipo de Imagen
Schönewolf y cols., 2022	Estudio diagnóstico observacional no randomizado	3,241 imágenes	ResNeXt-101-32x8d	Interna	Fotografías intraorales
Kühnisch y cols., 2024	Estudio observacional no randomizado	455 imágenes web	IA profunda (modelo abierto)	Externa	Imágenes web dentales
Felsch y cols., 2023	Estudio in vitro	18,179 imágenes	SegFormer-B5	Interna	Fotografías clínicas HD
Jaiswal y cols., 2024	Estudio observacional no randomizado	50 niños	SML+ OpenCV	Photoshop CS7	Fotos intraorales estandarizadas
He y cols., 2025	Estudio observacional no randomizado	208 pacientes (624 imágenes)	Mask R-CNN + SAM	Comparación con clínicos	Fotos intraorales múltiples vistas
Alevizakos y cols., 2022	Estudio in vitro	462 imágenes	VGG16, DenseNet, ResNet, AlexNet	Cruzada	Fotos dentales a color
Portella y cols., 2023	Estudio in vitro	2,481 dientes (140 imágenes test)	VGG-19	Clínica	Fotos oclusales alta resolución
Veseli y cols., 2024	Carta al editor / Descriptivo	No especificado	Modelo DL	Expertos clínicos	Fotografías clínicas

Tabla 6: Resumen – Características de los estudios incluidos

	Definición de los casos	Representatividad	Selección de los controles	Definición de los controles	Comparabilidad (factor más importante)	Comparabilidad (cualquier otra variable)	Comparación de la exposición	Mismo método para ambos grupos	Tasa de abandonos	Total
Schönewolf y cols., 2022	★	★	-	-	★	-	★	★	★	6
Kühnisch y cols., 2024	★	★	★	★	★	★	★	★	★	9
Jaiswal y cols., 2024	★	-	-	-	-	-	★	-	-	2
He y cols., 2025	★	★	★	★	★		★	★	★	8

Tabla 7: Medición del riesgo de sesgo de los estudios observacionales no randomizados con la escala

Newcastle-Ottawa – estudios observacionales con grupo control no randomizado.

	Criterio	Portella y cols., 2023	Alevizakos y cols., 2022	Felsch y cols., 2023
1	¿Se define claramente el objetivo del estudio desde el resumen o la introducción?	✓	✓	✓
2	¿Se describe adecuadamente el modelo in vitro utilizado (tipo de muestras, origen)?	✓	✓	✓
3	¿Se especifica el procedimiento experimental completo paso a paso?	✓	✓	✓
4	¿Se incluyen características del entorno de laboratorio (luz, temperatura, etc.)?	x	x	✓
5	¿Se justifica el tamaño muestral o se refiere a estudios previos?	x	x	✓
6	¿Se controlaron sesgos (aleatorización de muestras, cegamiento del evaluador, etc.)?	✓	✓	✓
7	¿Se define con claridad la variable principal de resultado?	✓	✓	✓
8	¿Las herramientas de medición están validadas o estandarizadas?	✓	✓	✓
9	¿Se repitió el experimento (reproducibilidad / replicabilidad) y se informa el nº de repeticiones?	✓	✓	✓
10	¿Se describen los análisis estadísticos usados y son adecuados?	✓	✓	✓
11	¿Los resultados están claramente expresados en tablas, gráficos o porcentajes claros?	✓	✓	✓
12	¿Se discuten claramente las limitaciones del estudio?	✓	✓	✓
13	¿Se declaran fuentes de financiación y conflictos de interés?	✓	✓	✓
	Puntaje	11	11	13

Tabla 8: Estudios in vitro – Escala ARRIVE/CONSORT Modificada

Autor (Año)	Precisión (%)	Sensibilidad (%)	Especificidad (%)	F1-Score	Valor Predictivo Positivo	Hallazgos Relevantes
Schönewol y cols., 2022	95.2	78.6	97.3	No especificado	No indicado	Alta precisión general; menor sensibilidad en casos complejos.
Kühnisch y cols, 2024	94.3	81.7–92.8	91.9–98.7	No indicado	No indicado	Acuerdo 89.2% con expertos; buen desempeño.
Felsch y cols., 2023	97.8	No específico	No específico	AP: 82.7–90.2	No indicado	Rendimiento excelente, IoU: 95.9%.
Jaiswal y cols., 2024	Media 34.51%	No aplica	No aplica	No aplica	No aplica	Extensión >60% se asocia con 80% preocupación estética.
He y cols., 2025	No directo	93.5	78.9	0.856	0.968	IA mejora F1 de 0.713 a 0.949 con ayuda.
Alevizakos y cols., 2022	83.98–92.86	No indicado	No indicado	No indicado	No indicado	Mejor modelo: DenseNet121.
Portella y cols., 2023	87.9	82.7	80.0	0.887	0.949	Mejora significativa en precisión con IA.
Veseli y cols., 2024	No indicado	No indicado	No indicado	No indicado	No indicado	Alto rendimiento estimado, sin métricas exactas.

Tabla 9: Resultados diagnósticos de los estudios

Autor (Año)	Fracaso reportado	Descripción detallada
Schönewolf y cols., 2022	Baja sensibilidad en subcategorías complejas	El modelo mostró baja sensibilidad (40%) en categorías como 'ruptura de esmalte con sellante', a pesar de una alta precisión global.
Kühnisch y cols., 2024	Ninguno reportado	No se documentaron fracasos ni fallos técnicos; validación externa satisfactoria con acuerdo clínico del 89.2%.
Felsch y cols., 2023	Falta de validación externa	El modelo no fue validado fuera del conjunto de entrenamiento; riesgo de sobreajuste no evaluado clínicamente.
Jaiswal y cols., 2024	Reducción crítica de precisión con artefactos	Presencia de reflejos >15% redujo la precisión del modelo a un 57%, afectando la fiabilidad diagnóstica.
He y cols., 2025	Rendimiento desigual por anatomía	Menor rendimiento (F1 = 0.727) en molares inferiores frente a valores >0.95 en premolares superiores.
Alevizakos y cols., 2022	Rendimiento variable según patología	Precisión inestable entre subgrupos; caries se clasificó con alta precisión, pero MIH mostró menor exactitud.
Portella y cols., 2023	Diferencias según experiencia del evaluador	El modelo fue menos útil en manos de clínicos experimentados; mejora significativa solo en estudiantes.
Veseli y cols., 2024	Falta de métricas objetivas	No se incluyen medidas cuantitativas de eficacia; se basa únicamente en juicio clínico no reproducible.

Tabla 10: Complicaciones o limitaciones encontradas en los estudios.

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	0
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	4
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	6
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	15
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	16
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	17
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	18
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	19
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	19
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	19
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	19
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	20
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	20
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	20
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	20
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	20
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	20
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	20

Section and Topic	Item #	Checklist item	Location where item is reported
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	19
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	20
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	22
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	23
Study characteristics	17	Cite each included study and present its characteristics.	24
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	25
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	27
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	28
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	28
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	27
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	27
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	29
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	30
	23b	Discuss any limitations of the evidence included in the review.	30
	23c	Discuss any limitations of the review processes used.	31
	23d	Discuss implications of the results for practice, policy, and future research.	31

Section and Topic	Item #	Checklist item	Location where item is reported
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	
Competing interests	26	Declare any competing interests of review authors.	
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71. This work is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Guia prisma

Declaración detallada de uso de IA:

En la elaboración del presente trabajo, se ha recurrido a herramientas de inteligencia artificial para guiar el procedimiento metodológico, concretamente Chat GPT 4o, Chat GPT 4.5, Grock 3.0, Grock 3.0, Mendeley, Prospero, Deep L.

-Herramientas: Chat GPT 4o, Chat GPT 4.5, Grock 3.0, Mendeley, Prospero, Deep L.

-Funciones:

- Apoyo en el desarrollo de un esquema para un artículo de revisión sistemática sobre el diagnóstico del MIH con la inteligencia artificial.
- Creación de tablas.
- Correcciones gramaticales, traducciones de artículos.
- Conversión de docx.pdf a docx.word
- Consejos para mejorar el tfg.
- Funcion de tutor

Prompts utilizados:

- Traduce este artículo en francés: (artículo x)
- Traduce este texto en francés: (texto)
- Traduce este texto en español: (texto)
- ¿Puedes explicarme cómo aplicar la guía PRISMA en mi revisión sistemática?
- ¿Cómo puedo hacer un diagrama de flujo según la guía prisma?
- Con estos datos: (datos) genera una tabla dividida en :(parte x/parte x).
- Cómo podría mejorar esta parte : (parte)
- Mejora el estilo académico de este párrafo
- Reformula este texto para que sea más claro y fluido
- Que nota pondrías a mi tfg, que me recomendarías añadir o quitar.

Title: Diagnostic accuracy of artificial intelligence for molar-incisor hypomineralization (MIH): A systematic review

Running Title: AI-based diagnosis of MIH

Authors: Valentin Villette, Inmaculada Selva Ribera

Contact Address:

Campus de Valencia, Paseo de la Alameda, 7, 46010 Valencia

Email: [villette.valentin@outlook.fr]

Abstract

Background: Molar-Incisor Hypomineralization (MIH) is a prevalent qualitative enamel defect in pediatric dentistry, posing diagnostic difficulties due to its clinical variability and similarity to other enamel anomalies. Artificial intelligence (AI) has emerged as a potential tool to enhance diagnostic precision and reduce human error.

Material and Methods: Following PRISMA guidelines, an electronic search was conducted in PubMed, Scopus, and Web of Science databases up to February 2024. Eight studies were included that applied AI models, such as convolutional neural networks (CNNs) and transformer-based architectures, for the diagnosis of MIH using clinical and intraoral images. Methodological quality and bias were assessed using Newcastle-Ottawa, ARRIVE/CONSORT and MOGA tools.

Results: AI models achieved diagnostic accuracy ranging from 83.98% to 97.8%, sensitivity between 78.6% and 93.5%, and specificity above 90% in most studies. CNN-based models like ResNeXt-101 and DenseNet showed high reliability, while transformer models reached peak segmentation performance (IoU > 95%). Limitations included dependency on image quality, reduced sensitivity in complex cases, absence of external validation, and ethical concerns regarding data use.

Conclusions: AI demonstrates high potential for early, accurate MIH detection. Its implementation may support clinical decision-making and dental education. However, further multicenter validation and regulatory frameworks are needed.

Key Words: *Molar-Incisor Hypomineralization, Artificial Intelligence, Deep Learning, Diagnostic Imaging, Pediatric Dentistry*

Introduction:

Molar-incisor hypomineralization (MIH) is a qualitative enamel defect of developmental origin, primarily affecting the first permanent molars and, often, the incisors. This anomaly represents a growing challenge in pediatric dentistry, affecting up to 40% of children in some regions, depending on the diagnostic criteria employed (1,2). In addition to its high prevalence, MIH is associated with dental hypersensitivity, increased caries risk, restorative complications and reduced quality of life in affected children (3,4,5).

Clinically, MIH manifests as well-demarcated opacities, white, yellow or brown in color, which can progress to post-eruptive fractures. These signs impair masticatory function and aesthetics, and are often accompanied by pain when brushing or eating hot/cold foods (6). The psychosocial consequences are significant, particularly in schoolchildren, where visible incisor anomalies can affect self-esteem and generate anxiety about dental care (7).

Despite the existence of diagnostic protocols established by institutions such as the European Academy of Paediatric Dentistry (EAPD), the diagnosis of MIH remains difficult, especially in its early or atypical forms. It still relies heavily on visual assessment, resulting in high inter- and intra-examiner variability, compounded by possible confusion with other enamel anomalies such as fluorosis, incipient caries or hypoplasia (4,5,7).

Faced with these challenges, artificial intelligence (AI) has recently attracted growing interest as a complementary diagnostic tool. Thanks to its image analysis capabilities, it can automatically identify MIH lesions from intraoral photographs or digital scans. Studies have shown that convolutional neural networks (CNNs), transformer-type architectures or computer vision-based segmentation tools achieve performance equal to or better than that of experienced clinicians, while improving reproducibility and speed of diagnosis (8,9,10). The integration of AI in dentistry is also part of a broader digital transformation, aimed at improving access to quality care, particularly in under-serviced regions, and supporting the training of professionals. However, the clinical adoption of these technologies still poses several challenges: the need for

diversified and standardized training data, acceptability by practitioners, regulatory requirements, and ethical concerns related to the protection of personal data (8,9,11). In this context, this systematic review aims to assess the diagnostic accuracy of artificial intelligence models in the detection of MIH, through a critical analysis of published studies on the subject. The aim is to identify the extent to which these tools can complement or improve on traditional methods, and to examine their current limitations as well as their prospects for clinical and educational integration.

Material and Methods:

This systematic review was conducted in accordance with the PRISMA guidelines and was registered in the international database PROSPERO. The aim was to evaluate the diagnostic performance of artificial intelligence (AI) systems in detecting molar-incisor hypomineralization (MIH) using image-based data.

The research question was structured using the PICO framework, as follows:

P (Population): Patients with suspected molar-incisor hypomineralization (MIH).

I (Intervention): Diagnosis using artificial intelligence models.

C (Comparison): Diagnosis performed by dentists using traditional methods.

O (Outcomes):

O1: Effectiveness in detecting MIH.

O2: Diagnostic accuracy.

A comprehensive literature search was conducted in three major electronic databases: PubMed (Medline), Scopus, and Web of Science, including publications up to February 2024. The search terms included: *“molar-incisor hypomineralization,” “MIH,” “artificial intelligence,” “deep learning,” “CNN,”* and *“image-based diagnosis,”* combined using appropriate Boolean operators.

Original studies that evaluated AI-based tools for the diagnosis of MIH using clinical or intraoral images were included. The accepted study designs comprised observational studies, in vitro experiments, randomized clinical trials (RCTs), systematic reviews, and meta-analyses. Publications had to be written in English, Spanish, or French, and published within the last ten years. Studies focusing on enamel defects other than

MIH, those that did not use image-based AI analysis, or those that lacked peer review (e.g., editorials or grey literature) were excluded.

Study selection was carried out independently by two reviewers (VV and ISR) through a multi-stage process, including the evaluation of titles, abstracts, and full texts. Any disagreement was resolved by consensus.

Data were extracted using a standardized form that collected the following variables: study characteristics, sample size, AI model architecture, performance metrics (accuracy, sensitivity, specificity, F1 score, IoU), and reported limitations.

To assess methodological quality and risk of bias, specific tools were applied depending on the study type: the Newcastle-Ottawa Scale for observational studies, the ARRIVE or CONSORT checklist for in vitro investigations, and the MOGA tool for case series.

Inter-rater reliability was calculated using Cohen's kappa coefficient.

Due to heterogeneity in study designs and outcome metrics, a narrative synthesis approach was chosen for the analysis and presentation of results.

Results:

In this systematic review, eight studies published between 2022 and 2025 were included, all focusing on the evaluation of artificial intelligence (AI) models applied to the diagnosis of molar-incisor hypomineralization (MIH) from clinical or dental images. Methodological designs were heterogeneous: four studies were non-randomized observational (8,9,12,13); three studies were experimental in vitro (10,14,15); and one was a descriptive letter to the editor (16).

Sample sizes varied considerably, from 50 to over 18,000 images, totalling over 25,000 images used for training or model validation. All studies used intraoral or manually annotated multiple-view photographs.

The most frequently used AI models were convolutional neural networks (CNNs), such as ResNeXt-101, VGG-19, DenseNet and ResNet (8,14,15). Other studies have introduced more advanced architectures such as the SegFormer-B5 transformer-type

model (10), or combinations with image analysis tools such as OpenCV and the Segment Anything Model (SAM) (12,13).

Reported diagnostic performance is generally high. Schönewolf et al. obtained an accuracy of 95.2%, a sensitivity of 78.6% and a specificity of 97.3%. Kühnisch et al., with external validation, report an accuracy of 94.3%, a sensitivity ranging from 81.7% to 92.8%, a specificity from 91.9% to 98.7% and an agreement of 89.2% with the experts. The transform model tested by Felsch et al., demonstrated an overall accuracy of 97.8%, with a mean accuracy (PA) between 82.7% and 90.2% and an IoU of 95.9%, attesting to excellent segmentation capability.

Portella et al., reported an accuracy of 87.9%, a sensitivity of 82.7%, a specificity of 80% and an F1-score of 0.887. These results exceeded the performance of human examiners for early lesions. He et al., using a Mask R-CNN model combined with SAM, showed that AI assistance enabled junior dentists to increase their F1-score from 0.713 to 0.949, with a sensitivity of 93.5% and a specificity of 78.9%.

Alevizakos et al., did not detail the classic metrics (sensitivity, specificity), but reported an accuracy between 83.98% and 92.86%, with DenseNet being the best-performing model among those tested. Jaiswal et al., for their part, developed a tool for quantifying the extent of MIH lesions, with a mean estimate of 34.51%, very close to the results obtained by Photoshop (35.32%), thus validating the usefulness of the model for extension measurement. Lastly, Veseli et al., did not provide quantitative data, but reported a good overall estimated yield.

With regard to complications, several technical limitations have been raised.

Schönewolf et al., observed a decrease in sensitivity in complex cases, particularly in the presence of sealants or fractures. Jaiswal et al., demonstrated that specular reflections on images could reduce diagnostic accuracy by 57% if more than 15% of the image was affected. Lack of external validation was mentioned as a major obstacle by Felsch et al. and Alevizakos et al., while Veseli et al., provided no objective data to support the efficacy of his model.

Another complication reported by He et al., concerns anatomical variability: performance was poorer on lower molars ($F1 = 0.727$) than on upper premolars ($F1 = 0.963$), underlining the need to adapt models to tooth morphology. Finally, Portella et al noted that the effectiveness of AI depended on the user's experience, with AI providing a more marked improvement in novice practitioners than in experts.

Discussion:

The results of this systematic review confirm the high level of diagnostic accuracy achieved by artificial intelligence (AI) models for the identification of molar-incisal hypomineralization (MIH). The performance of the various algorithms evaluated, with accuracy rates ranging from 83.98% to 97.8%, highlights their potential as a complementary tool in clinical diagnosis, particularly for non-specialist practitioners or in resource-limited settings. For example, the ResNeXt-101 model by Schönewolf et al., achieved an accuracy of 95.2%, although its sensitivity was relatively lower (78.6%), indicating difficulties in identifying certain complex cases. In contrast, the externally validated model by Kühnisch et al., showed a better balance between sensitivity (81.7-92.8%) and specificity (91.9-98.7%), with a high rate of agreement (89.2%) with experts, underlining its reliability in clinical practice.

Felsch et al., introduced a notable technological innovation with the use of a vision transform model, which achieved 97.8% accuracy and 95.9% spatial agreement (IoU), confirming the ability of recent architectures to accurately detect and locate MIH lesions in complex databases. This advanced performance echoes a growing trend towards the adoption of segmentation models in the dental field.

However, this accuracy is accompanied by several methodological and technical limitations noted in the included studies. Image quality has been identified as a critical factor: according to Jaiswal et al., the presence of light reflections (specular artifacts) over more than 15% of the image surface can reduce AI accuracy by up to 43%, highlighting the need for rigorous standardization of clinical imaging protocols.

Another limitation concerns the variability of performance depending on anatomical location. He et al., showed that results were significantly poorer for lower molars ($F1 =$

0.727) than for upper premolars ($F1 = 0.963$), probably due to morphological differences and visual access. This highlights the need to include varied data in model training sets to improve their robustness and general applicability.

From a technological point of view, several authors have noted the frequent absence of external validation, as in the work of Felsch et al. and Veseli et al., which compromises the generalizability of results obtained in controlled environments. In addition, models are often more useful for less experienced clinicians, as suggested by Portella et al., where performance improvement was significant in students or junior practitioners, but less so in experts.

Despite these limitations, the clinical and educational prospects are promising. Some models, such as that of Jaiswal et al., go beyond simple diagnosis and quantify the extent of MIH lesions, which could prove useful for longitudinal follow-up and therapeutic planning. Others, such as Alevizakos et al., aim to differentiate MIH from other similar pathologies such as fluorosis or amelogenesis imperfecta, reinforcing their usefulness in differential diagnosis.

Finally, several authors advocate integrating these tools into accessible media, such as mobile applications, web platforms or intelligent intraoral scanners. This technological accessibility, coupled with pedagogical interfaces, could facilitate their deployment in initial training, clinical contexts or public health programs (10,14,16).

To make this transition a reality, future work will need to be based on multicentric, diversified, ethically framed and technically explicable data, to ensure safe, equitable and effective adoption of AI in pediatric dentistry.

Conclusions:

This systematic review demonstrates that artificial intelligence, particularly deep learning models such as CNNs and transformer architectures, offers robust diagnostic capabilities for the detection of molar-incisor hypomineralization (MIH). Across various study designs and datasets, AI tools consistently achieved high levels of accuracy, sensitivity, and specificity, often equaling or surpassing traditional clinical diagnostic methods.

The use of AI presents a transformative opportunity in pediatric dentistry. Its ability to analyze clinical and intraoral images with speed and precision can significantly improve early detection, facilitate minimally invasive interventions, and reduce variability in clinical judgment especially in contexts where access to pediatric specialists is limited. Furthermore, AI holds potential as an educational tool, allowing dental students and general practitioners to benefit from diagnostic support and consistent feedback during clinical training.

Nevertheless, several limitations must be addressed before widespread implementation is feasible. The lack of multicentric external validation, dependency on high-quality image data, anatomical diagnostic inconsistencies, and ethical concerns regarding data privacy and algorithm transparency remain significant barriers. Future research should focus on developing explainable AI models, integrating diverse and representative datasets, and establishing clear regulatory frameworks to ensure ethical and effective deployment in clinical environments.

In conclusion, AI-based diagnostic systems for MIH are highly promising and represent a valuable adjunct to traditional diagnostic workflows. Their integration into everyday clinical and academic practice has the potential to enhance diagnostic accuracy, democratize access to specialized care, and contribute to improved oral health outcomes in children.

Conflict of Interest

The authors declare no conflicts of interest.

Ethics

Not applicable. Systematic review of previously published data.

Funding

No external funding was received.

Author Contributions

VV: Conceptualization, data collection, drafting. ISR: Supervision, methodology, revision.

References:

- 1) Lopes LB, Machado V, Mascarenhas P, Mendes JJ, Botelho J. The prevalence of molar-incisor hypomineralization: a systematic review and meta-analysis. *SciRep*. 2021 Nov 17;11(1):22405. doi: 10.1038/s41598-021-01541-7. PMID: 34789780; PMCID: PMC8599453.
- 2) Villani FA, Aiuto R, Dioguardi M, Paglia L, Caruso S, Gatto R et al. Caries prevalence and molar incisor hypomineralisation (MIH) in children. Is there an association? A systematic review. *Eur J Paediatr Dent*. 2023 Dec 1;24(4):312-320. doi: 10.23804/ejpd.2023.1985. PMID: 38015112.
- 3) Mazur M, Corridore D, Ndokaj A, Ardan R, Voza I, Babajko S et al. MIH and Dental Caries in Children: A Systematic Review and Meta-Analysis. *Healthcare (Basel)*. 2023 Jun 18;11(12):1795. doi: 10.3390/healthcare11121795. PMID: 37372913; PMCID: PMC10298042.
- 4) Rodd HD, Graham A, Tajmehr N, Timms L, Hasmun N. Molar Incisor Hypomineralisation: Current Knowledge and Practice. *Int Dent J*. 2021 Aug;71(4):285-291. doi: 10.1111/idj.12624. PMID: 34286697; PMCID: PMC9275314.
- 5) Alzahrani AY, Alamoudi NMH, El Meligy OAES. Contemporary Understanding of the Etiology and Management of Molar Incisor Hypomineralization: A Literature Review. *Dent J (Basel)*. 2023 Jun 23;11(7):157. doi: 10.3390/dj11070157. PMID: 37504223; PMCID: PMC10378579.
- 6) Lygidakis NA, Garot E, Somani C, Taylor GD, Rouas P, Wong FSL. Best clinical practice guidance for clinicians dealing with children presenting with molar-incisor-hypomineralisation (MIH): an updated European Academy of Paediatric Dentistry policy document. *Eur Arch Paediatr Dent*. 2022 Feb;23(1):3-21. doi: 10.1007/s40368-021-00668-5. PMID: 34669177; PMCID: PMC8926988.
- 7) Enax J, Amaechi BT, Farah R, Liu JA, Schulze zur Wiesche E, Meyer F. Remineralization Strategies for Teeth with Molar Incisor Hypomineralization (MIH): A Literature Review. *Dent J (Basel)*. 2023 Mar 13;11(3):80. doi: 10.3390/dj11030080. PMID: 36975577; PMCID: PMC10047667.
- 8) Schönewolf J, Meyer O, Engels P, Schlickerrieder A, Hickel R, Gruhn V. Artificial intelligence-based diagnostics of molar-incisor- 26 hypomineralization (MIH) on

- intraoral photographs. Clin Oral Investig. 2022 Sep;26(9):5923-5930. doi: 10.1007/s00784-022-04552-4. PMID: 35608684;PMCID: PMC9474479.
- 9) Neumayr J, Frenkel E, Schwarzmaier J, Ammar N, Kessler A, Schwendicke F et al. External validation of an artificial intelligence-based method for the detection and classification of molar incisor hypomineralisation in dental photographs. J Dent. 2024 Sep;148:105228. doi: 10.1016/j.jdent.2024.105228. PMID: 38972447.
- 10) Felsch M, Meyer O, Schlickerrieder A, Engels P, Schönewolf J, Zöllner F et al. Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model. NPJ Digit Med. 2023 Oct 25;6(1):198. doi: 10.1038/s41746-023-00944-2. PMID: 37880375; PMCID: PMC10600213.
- 11) Tchito Tchappa C, Mih TA, Tchagna Kouanou A, Fozin Fonzin T, Kuetché Fogang P, Mezatio BA et al. Biomedical Image Classification in a Big Data Architecture Using Machine Learning Algorithms. J Healthc Eng. 2021 May 30;2021:9998819. doi: 10.1155/2021/9998819. PMID: 34122785; PMCID: PMC8191587.
- 12) Jaiswal M, Mukhtar U, Shakya KS, Laddi A, Singha LA. Computerised assessment—a novel approach for calculation of percentage of hypomineralized lesion on incisors and its correlation with aesthetic concern. J Oral Biol Craniofac Res. 2024 Sep-Oct;14(5):570-577. doi: 10.1016/j.jobcr.2024.07.004. Epub 2024 Jul 23. PMID: 39139516; PMCID: PMC11320481.
- 13) He K, Zhang R, Liang M, Tian K, Luo K, Chen R et al. The application of deep learning in early enamel demineralization detection. PeerJ. 2025 Jan 2;13:e18593. doi: 10.7717/peerj.18593. PMID: 39763696; PMCID: PMC11700490.
- 14) Alevizakos V, Bekes K, Steffen R, von See C. Artificial intelligence system for training diagnosis and differentiation with molar incisor hypomineralization (MIH) and similar pathologies. Clin Oral Investig. 2022 Dec;26(12):6917-6923. doi: 10.1007/s00784-022-04646-z. Epub 2022 Sep 6. PMID: 36065023. 27
- 15) Portella PD, de Oliveira LF, Ferreira MFC, Dias BC, de Souza JF, Assunção LRDS. Improving accuracy of early dental carious lesions detection using deep learning based automated method. Clin Oral Investig. 2023 Dec;27(12):7663-7670. doi: 10.1007/s00784-023-05355-x. Epub 2023 Oct 31. PMID: 37906303.

16) Veseli E. Early detection of MIH in children by using artificial intelligence. Eur Arch Paediatr Dent. 2024 Dec;25(6):899-900. doi: 10.1007/s40368-024-00931- 5. Epub 2024 Aug 1. PMID: 39090506.

Anexos:

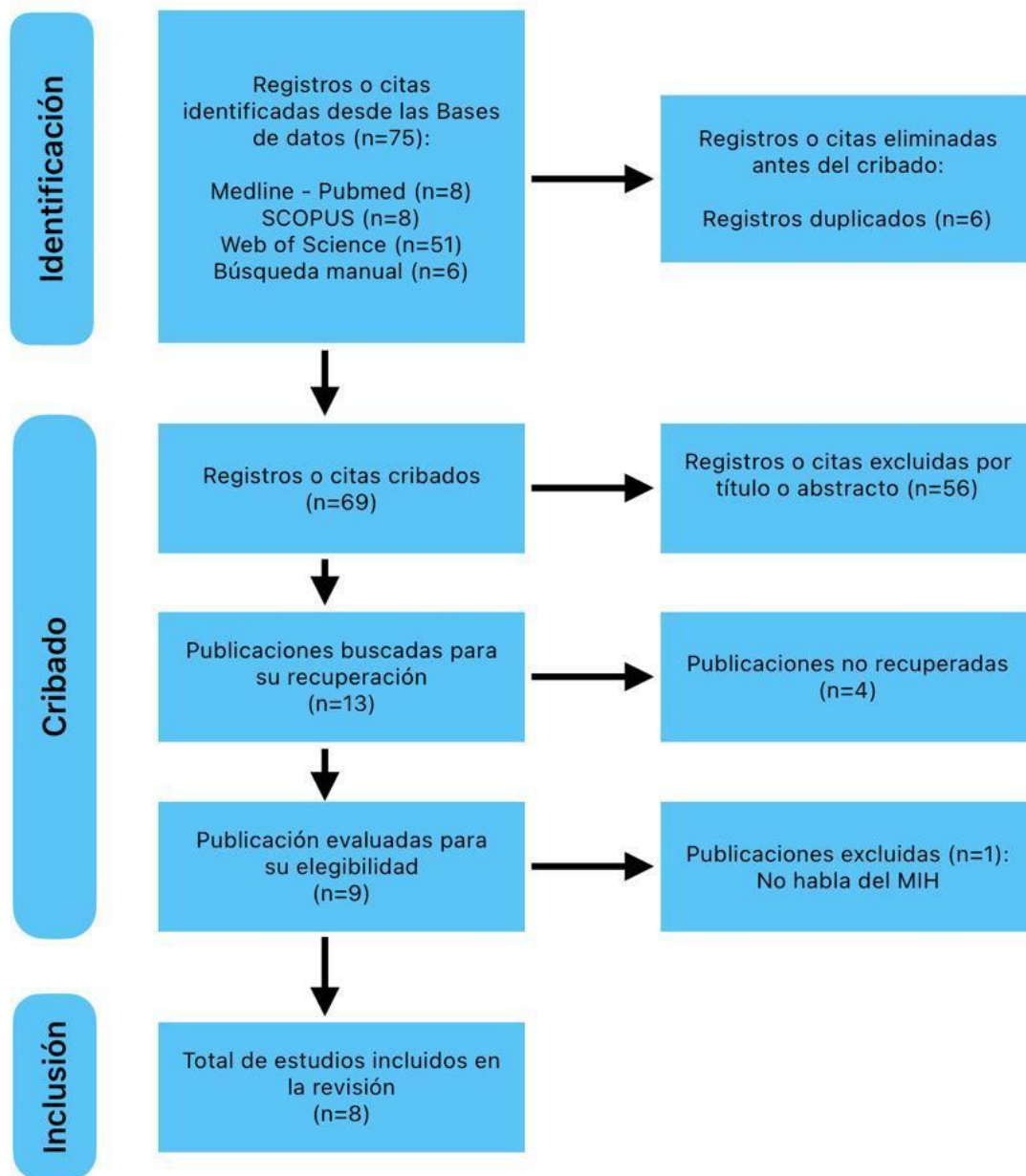


Fig. 1. Flowchart of the search and title selection process during the systematic review.

Autor (Año)	Tipo de Estudio	Muestra	Modelo de IA	Validación	Tipo de Imagen
Schönewolf y cols., 2022	Estudio diagnóstico observacional no randomizado	3,241 imágenes	ResNeXt-101-32×8d	Interna	Fotografías intraorales
Kühnisch y cols., 2024	Estudio observacional no randomizado	455 imágenes web	IA profunda (modelo abierto)	Externa	Imágenes web dentales
Felsch y cols., 2023	Estudio in vitro	18,179 imágenes	SegFormer-B5	Interna	Fotografías clínicas HD
Jaiswal y cols., 2024	Estudio observacional no randomizado	50 niños	SML+ OpenCV	Photoshop CS7	Fotos intraorales estandarizadas
He y cols., 2025	Estudio observacional no randomizado	208 pacientes (624 imágenes)	Mask R-CNN + SAM	Comparación con clínicos	Fotos intraorales múltiples vistas
Alevizakos y cols., 2022	Estudio in vitro	462 imágenes	VGG16, DenseNet, ResNet AlexNet	Cruzada	Fotos dentales a color
Portella y cols., 2023	Estudio in vitro	2,481 dientes (140 imágenes test)	VGG-19	Clínica	Fotos oclusales alta resolución
Veseli y cols., 2024	Carta al editor / Descriptivo	No especificado	Modelo DL	Expertos clínicos	Fotografías clínicas

Table 1: Summary – Characteristics of the Included Studies

Autor (Año)	Precisión (%)	Sensibilidad (%)	Especificidad (%)	F1-Score	Valor Predictivo Positivo	Hallazgos Relevantes
Schönewol y cols., 2022	95.2	78.6	97.3	No especificado	No indicado	Alta precisión general; menor sensibilidad en casos complejos.
Kühnisch y cols, 2024	94.3	81.7–92.8	91.9–98.7	No indicado	No indicado	Acuerdo 89.2% con expertos; buen desempeño.
Felsch y cols., 2023	97.8	No específico	No específico	AP: 82.7–90.2	No indicado	Rendimiento excelente, IoU: 95.9%.
Jaiswal y cols., 2024	Media 34.51%	No aplica	No aplica	No aplica	No aplica	Extensión >60% se asocia con 80% preocupación estética.
He y cols., 2025	No directo	93.5	78.9	0.856	0.968	IA mejora F1 de 0.713 a 0.949 con ayuda.
Alevizakos y cols., 2022	83.98–92.86	No indicado	No indicado	No indicado	No indicado	Mejor modelo: DenseNet121.
Portella y cols., 2023	87.9	82.7	80.0	0.887	0.949	Mejora significativa en precisión con IA.
Veseli y cols., 2024	No indicado	No indicado	No indicado	No indicado	No indicado	Alto rendimiento estimado, sin métricas exactas.

Table 2: Diagnostic Results of the Studie

Título: Precisión diagnóstica de la inteligencia artificial para la hipomineralización molar-incisiva (MIH): Una revisión sistemática

Running Title: Diagnóstico de la MIH basado en IA

Autores: Valentín Villette, Inmaculada Selva Ribera

Dirección de contacto:

Campus de Valencia, Paseo de la Alameda, 7, 46010 Valencia

Email: [villette.valentin@outlook.fr]

Resumen

Antecedentes: La Hipomineralización Incisivo-Molar (MIH) es un defecto cualitativo del esmalte prevalente en odontopediatría, que plantea dificultades diagnósticas debido a su variabilidad clínica y similitud con otras anomalías del esmalte. La inteligencia artificial (IA) ha surgido como una herramienta potencial para mejorar la precisión diagnóstica y reducir el error humano.

Material y métodos: Siguiendo las directrices PRISMA, se realizó una búsqueda electrónica en las bases de datos PubMed, Scopus y Web of Science hasta febrero de 2024. Se incluyeron ocho estudios que aplicaron modelos de IA, como redes neuronales convolucionales (CNN) y arquitecturas basadas en transformadores, para el diagnóstico de MIH utilizando imágenes clínicas e intraorales. La calidad metodológica y el sesgo se evaluaron mediante las herramientas Newcastle-Ottawa, ARRIVE/CONSORT y MOGA.

Resultados: Los modelos de IA alcanzaron una precisión diagnóstica que osciló entre el 83,98% y el 97,8%, una sensibilidad entre el 78,6% y el 93,5%, y una especificidad superior al 90% en la mayoría de los estudios. Los modelos basados en CNN como ResNeXt-101 y DenseNet mostraron una alta fiabilidad, mientras que los modelos de transformador alcanzaron un rendimiento de segmentación máximo (IoU > 95%). Las limitaciones incluyeron la dependencia de la calidad de la imagen, la sensibilidad reducida en casos complejos, la ausencia de validación externa y las preocupaciones éticas relativas al uso de los datos.

Conclusiones: La IA demuestra un alto potencial para la detección precoz y precisa del MIH. Su aplicación puede contribuir a la toma de decisiones clínicas y a la formación odontológica. Sin embargo, se necesitan más validaciones multicéntricas y marcos normativos.

Palabras clave: *Hipomineralización incisivo-molar, inteligencia artificial, aprendizaje profundo, diagnóstico por imagen, odontopediatría*

Introducción:

La hipomineralización molar-incisiva (MIH) es un defecto cualitativo del esmalte de origen evolutivo, que afecta principalmente a los primeros molares permanentes y, a menudo, a los incisivos. Esta anomalía representa un reto creciente en odontopediatría, afectando hasta el 40% de los niños en algunas regiones, dependiendo de los criterios diagnósticos empleados (1,2). Además de su elevada prevalencia, el MIH se asocia a hipersensibilidad dental, mayor riesgo de caries, complicaciones restaurativas y menor calidad de vida en los niños afectados (3,4,5).

Clínicamente, el MIH se manifiesta como opacidades bien delimitadas, de color blanco, amarillo o marrón, que pueden progresar a fracturas post-eruptivas. Estos signos deterioran la función masticatoria y la estética, y suelen acompañarse de dolor al cepillarse o al comer alimentos calientes o fríos (6). Las consecuencias psicosociales son importantes, sobre todo en los escolares, donde las anomalías visibles de los incisivos pueden afectar a la autoestima y generar ansiedad ante el cuidado dental (7).

A pesar de la existencia de protocolos diagnósticos establecidos por instituciones como la Academia Europea de Odontopediatría (EAPD), el diagnóstico del MIH sigue siendo difícil, especialmente en sus formas tempranas o atípicas. Todavía depende en gran medida de la evaluación visual, lo que da lugar a una gran variabilidad inter e intraexaminador, agravada por la posible confusión con otras anomalías del esmalte como la fluorosis, la caries incipiente o la hipoplasia (4,5,7).

Frente a estos retos, la inteligencia artificial (IA) ha despertado recientemente un creciente interés como herramienta diagnóstica complementaria. Gracias a sus capacidades de análisis de imágenes, puede identificar automáticamente lesiones MIH a partir de fotografías intraorales o escaneados digitales. Los estudios han demostrado que las redes neuronales convolucionales (CNN), las arquitecturas de tipo transformador o las herramientas de segmentación basadas en la visión por ordenador logran un rendimiento igual o mejor que el de los clínicos experimentados, al tiempo que mejoran la reproducibilidad y la velocidad del diagnóstico (8,9,10). La integración de la IA en la odontología también forma parte de una transformación digital más amplia, destinada a mejorar el acceso a una atención de calidad, en particular en las

regiones insuficientemente atendidas, y a apoyar la formación de los profesionales. Sin embargo, la adopción clínica de estas tecnologías sigue planteando varios retos: la necesidad de datos de formación diversificados y estandarizados, la aceptabilidad por parte de los profesionales, los requisitos normativos y las preocupaciones éticas relacionadas con la protección de datos personales (8,9,11). En este contexto, esta revisión sistemática pretende evaluar la exactitud diagnóstica de los modelos de inteligencia artificial en la detección del MIH, mediante un análisis crítico de los estudios publicados sobre el tema. El objetivo es identificar en qué medida estas herramientas pueden complementar o mejorar los métodos tradicionales, y examinar sus limitaciones actuales, así como sus perspectivas de integración clínica y educativa.

Materiales y Métodos:

Esta revisión sistemática se llevó a cabo siguiendo las directrices PRISMA y fue registrada en la base de datos internacional PROSPERO. El objetivo fue evaluar el rendimiento diagnóstico de los sistemas de inteligencia artificial (IA) en la detección de la hipomineralización incisivo-molar (MIH) utilizando datos basados en imágenes. Esta pregunta de estudio se estructuró siguiendo el modelo PICO, del siguiente modo:

P (población): Pacientes con sospecha de hipomineralización de los molares e incisivos (MIH).

I (intervención): Diagnóstico mediante modelos de inteligencia artificial.

C (comparación): Diagnóstico realizado por odontólogos utilizando métodos tradicionales.

O (resultados):

O1: Efectividad en la detección de MIH.

O2: Precisión del diagnóstico.

Se realizó una búsqueda exhaustiva de la literatura en tres bases de datos electrónicas principales: PubMed (Medline), Scopus y Web of Science, incluyendo publicaciones hasta febrero de 2024. Los términos de búsqueda incluyeron: *“hipomineralización incisivo-molar”*, *“MIH”*, *“inteligencia artificial”*, *“aprendizaje profundo”*, *“CNN”* y *“diagnóstico por imagen”*, combinados mediante operadores booleanos apropiados.

Se incluyeron estudios originales que evaluaran herramientas basadas en IA para el diagnóstico de MIH mediante imágenes clínicas o intraorales. Se aceptaron los siguientes diseños de estudio: estudios observacionales, experimentos in vitro, ensayos clínicos aleatorizados (ECA), revisiones sistemáticas y metaanálisis. Las publicaciones debían estar redactadas en inglés, español o francés, y haber sido publicadas en los últimos diez años. Se excluyeron los estudios centrados en defectos del esmalte distintos de la MIH, aquellos que no utilizaran análisis de imágenes basados en IA, o los que no contaran con revisión por pares.

La selección de los estudios fue realizada de forma independiente por dos revisores (VV e ISR) mediante un proceso en múltiples etapas, que incluyó la evaluación de títulos, resúmenes y textos completos. Cualquier desacuerdo fue resuelto por consenso.

Los datos se extrajeron utilizando un formulario estandarizado que recopilaba las siguientes variables: características del estudio, tamaño de la muestra, arquitectura del modelo de IA, métricas de rendimiento (precisión, sensibilidad, especificidad, puntuación F1, IoU) y limitaciones reportadas.

Para evaluar la calidad metodológica y el riesgo de sesgo, se aplicaron herramientas específicas según el tipo de estudio: la escala de Newcastle-Ottawa para estudios observacionales, la lista de verificación ARRIVE o CONSORT para investigaciones in vitro, y la herramienta MOGA para series de casos. La fiabilidad entre evaluadores se calculó mediante el coeficiente kappa de Cohen.

Dada la heterogeneidad en los diseños de estudio y las métricas de resultado, se optó por un enfoque de síntesis narrativa para el análisis y la presentación de los resultados.

Resultados:

En esta revisión sistemática se incluyeron ocho estudios publicados entre 2022 y 2025, todos centrados en la evaluación de modelos de inteligencia artificial (IA) aplicados al diagnóstico de la hipomineralización incisivo-molar (MIH) a partir de imágenes clínicas o dentales. Los diseños metodológicos fueron heterogéneos: cuatro estudios fueron observacionales no aleatorizados (8,9,12,13); tres estudios fueron experimentales in

vitro (10,14,15); y uno fue una carta descriptiva al editor (16).

El tamaño de las muestras varió considerablemente, desde 50 hasta más de 18,000 imágenes, totalizando más de 25,000 imágenes utilizadas para el entrenamiento o validación de los modelos. Todos los estudios utilizaron fotografías intraorales o anotadas manualmente con múltiples vistas.

Los modelos de IA más frecuentemente utilizados fueron las redes neuronales convolucionales (CNN), como ResNeXt-101, VGG-19, DenseNet y ResNet (8,14,15). Otros estudios introdujeron arquitecturas más avanzadas como el modelo tipo transformador SegFormer-B5 (10), o combinaciones con herramientas de análisis de imágenes como OpenCV y el modelo Segment Anything (SAM) (12,13).

El rendimiento diagnóstico reportado fue generalmente alto. Schönewolf y cols. obtuvieron una precisión del 95.2%, una sensibilidad del 78.6% y una especificidad del 97.3%. Kühnisch y cols., con validación externa, informaron una precisión del 94.3%, una sensibilidad entre el 81.7% y el 92.8%, una especificidad entre el 91.9% y el 98.7% y una concordancia del 89.2% con los expertos. El modelo transformador probado por Felsch y cols., demostró una precisión general del 97.8%, con una precisión media (PA) entre el 82.7% y el 90.2% y un IoU del 95.9%, lo que demuestra una excelente capacidad de segmentación.

Portella y cols., informaron una precisión del 87.9%, una sensibilidad del 82.7%, una especificidad del 80% y una puntuación F1 de 0.887. Estos resultados superaron el rendimiento de los examinadores humanos para lesiones tempranas. He y cols., utilizando un modelo Mask R-CNN combinado con SAM, mostraron que la asistencia de IA permitió a odontólogos principiantes aumentar su puntuación F1 de 0.713 a 0.949, con una sensibilidad del 93.5% y una especificidad del 78.9%.

Alevizakos y cols., no detallaron métricas clásicas (sensibilidad, especificidad), pero informaron una precisión entre el 83.98% y el 92.86%, siendo DenseNet el modelo con mejor rendimiento entre los probados. Jaiswal y cols., por su parte, desarrollaron una herramienta para cuantificar la extensión de las lesiones de MIH, con una estimación media del 34.51%, muy cercana a los resultados obtenidos con Photoshop (35.32%),

validando así la utilidad del modelo para la medición de la extensión. Por último, Veseli y cols, no proporcionaron datos cuantitativos, pero reportaron un buen rendimiento estimado general.

En cuanto a las complicaciones, se han señalado varias limitaciones técnicas. Schönewolf y cols, observaron una disminución de la sensibilidad en casos complejos, particularmente en presencia de selladores o fracturas. Jaiswal y cols., demostraron que los reflejos especulares en las imágenes podían reducir la precisión diagnóstica en un 57% si más del 15% de la imagen se veía afectada. La falta de validación externa fue mencionada como un obstáculo importante por Felsch y cols. y Alevizakos y cols., mientras que Veseli y cols., no proporcionaron datos objetivos que respaldaran la eficacia de su modelo.

Otra complicación señalada por He et al., se refiere a la variabilidad anatómica: el rendimiento fue inferior en molares inferiores ($F1 = 0.727$) que en premolares superiores ($F1 = 0.963$), lo que subraya la necesidad de adaptar los modelos a la morfología dental. Finalmente, Portella y cols., observaron que la eficacia de la IA dependía de la experiencia del usuario, siendo la mejora más marcada en profesionales noveles que en expertos.

Discusión:

Los resultados de esta revisión sistemática confirman el alto nivel de precisión diagnóstica alcanzado por los modelos de inteligencia artificial (IA) para la identificación de la hipomineralización incisivo-molar (MIH). El rendimiento de los distintos algoritmos evaluados, con tasas de precisión que oscilan entre el 83.98% y el 97.8%, destaca su potencial como herramienta complementaria en el diagnóstico clínico, especialmente para profesionales no especialistas o en contextos con recursos limitados. Por ejemplo, el modelo ResNeXt-101 de Schönewolf y cols., alcanzó una precisión del 95.2%, aunque su sensibilidad fue relativamente más baja (78.6%), lo que indica dificultades en la identificación de ciertos casos complejos. En contraste, el modelo validado externamente por Kühnisch y cols., mostró un mejor equilibrio entre sensibilidad (81.7–92.8%) y especificidad (91.9–98.7%), con una alta tasa de concordancia (89.2%) con los expertos, lo que subraya su fiabilidad en la práctica

clínica.

Felsch y cols., introdujeron una innovación tecnológica notable con el uso de un modelo de visión tipo transformador, que alcanzó una precisión del 97.8% y una concordancia espacial (IoU) del 95.9%, lo que confirma la capacidad de las arquitecturas recientes para detectar y localizar con precisión las lesiones de MIH en bases de datos complejas. Este rendimiento avanzado refleja una tendencia creciente hacia la adopción de modelos de segmentación en el campo odontológico.

Sin embargo, esta precisión se acompaña de varias limitaciones metodológicas y técnicas señaladas en los estudios incluidos. La calidad de la imagen ha sido identificada como un factor crítico: según Jaiswal y cols., la presencia de reflejos luminosos (artefactos especulares) en más del 15% de la superficie de la imagen puede reducir la precisión de la IA hasta en un 43%, lo que pone de relieve la necesidad de una estandarización rigurosa de los protocolos de imagen clínica.

Otra limitación se refiere a la variabilidad del rendimiento según la localización anatómica. He y cols., demostraron que los resultados fueron significativamente peores en los molares inferiores ($F1 = 0.727$) que en los premolares superiores ($F1 = 0.963$), probablemente debido a diferencias morfológicas y al acceso visual. Esto subraya la necesidad de incluir datos variados en los conjuntos de entrenamiento de los modelos para mejorar su robustez y aplicabilidad general.

Desde un punto de vista tecnológico, varios autores han señalado la frecuente ausencia de validación externa, como en los trabajos de Felsch y cols. y Veseli y cols, lo que compromete la generalización de los resultados obtenidos en entornos controlados. Además, los modelos suelen ser más útiles para clínicos con menor experiencia, como sugieren Portella y cols., donde la mejora del rendimiento fue significativa en estudiantes o profesionales con poca experiencia, pero menos evidente en expertos.

A pesar de estas limitaciones, las perspectivas clínicas y educativas son prometedoras. Algunos modelos, como el de Jaiswal y cols., van más allá del diagnóstico simple y cuantifican la extensión de las lesiones de MIH, lo que podría ser útil para el seguimiento longitudinal y la planificación terapéutica. Otros, como Alevizakos y cols,

tienen como objetivo diferenciar la MIH de otras patologías similares como la fluorosis o la amelogénesis imperfecta, reforzando su utilidad en el diagnóstico diferencial.

Finalmente, varios autores abogan por integrar estas herramientas en medios accesibles, como aplicaciones móviles, plataformas web o escáneres intraorales inteligentes. Esta accesibilidad tecnológica, combinada con interfaces pedagógicas, podría facilitar su implementación en la formación inicial, en contextos clínicos o en programas de salud pública (10,14,16).

Para hacer realidad esta transición, los trabajos futuros deberán basarse en datos multicéntricos, diversificados, enmarcados éticamente y técnicamente explicables, con el fin de garantizar una adopción segura, equitativa y eficaz de la IA en la odontología pediátrica.

Conclusión:

Esta revisión sistemática demuestra que la inteligencia artificial, en particular los modelos de aprendizaje profundo como las redes neuronales convolucionales (CNN) y las arquitecturas tipo transformador, ofrece capacidades diagnósticas sólidas para la detección de la hipomineralización incisivo-molar (MIH). En diversos diseños de estudio y conjuntos de datos, las herramientas de IA lograron de manera constante altos niveles de precisión, sensibilidad y especificidad, igualando o incluso superando en ocasiones los métodos diagnósticos clínicos tradicionales.

El uso de la IA representa una oportunidad transformadora en la odontología pediátrica. Su capacidad para analizar imágenes clínicas e intraorales con rapidez y precisión puede mejorar significativamente la detección precoz, facilitar intervenciones mínimamente invasivas y reducir la variabilidad en el juicio clínico, especialmente en contextos donde el acceso a especialistas pediátricos es limitado. Además, la IA tiene un gran potencial como herramienta educativa, permitiendo que estudiantes de odontología y profesionales generales se beneficien de apoyo diagnóstico y retroalimentación constante durante su formación clínica.

No obstante, deben abordarse varias limitaciones antes de que su implementación generalizada sea factible. La falta de validación externa multicéntrica, la dependencia

de imágenes de alta calidad, las inconsistencias diagnósticas según la anatomía y las preocupaciones éticas en torno a la privacidad de los datos y la transparencia de los algoritmos siguen siendo barreras importantes. Las investigaciones futuras deben centrarse en desarrollar modelos de IA explicables, integrar conjuntos de datos diversos y representativos, y establecer marcos regulatorios claros para garantizar un uso ético y eficaz en entornos clínicos.

En conclusión, los sistemas de diagnóstico basados en IA para la MIH son altamente prometedores y representan un valioso complemento a los flujos de trabajo diagnósticos tradicionales. Su integración en la práctica clínica y académica cotidiana tiene el potencial de mejorar la precisión diagnóstica, democratizar el acceso a la atención especializada y contribuir a mejores resultados de salud bucodental en la población infantil.

Conflicto de intereses

Los autores declaran no tener conflictos de interés.

Ética

No aplicable. Revisión sistemática de datos previamente publicados.

Financiación

No se recibió financiación externa.

Contribuciones de los autores

VV: Conceptualización, recopilación de datos, redacción. ISR: Supervisión, metodología, revisión.

Bibliografía:

- 1) Lopes LB, Machado V, Mascarenhas P, Mendes JJ, Botelho J. The prevalence of molar-incisor hypomineralization: a systematic review and meta-analysis. *SciRep*. 2021 Nov 17;11(1):22405. doi: 10.1038/s41598-021-01541-7. PMID: 34789780; PMCID: PMC8599453.
- 2) Villani FA, Aiuto R, Dioguardi M, Paglia L, Caruso S, Gatto R y cols. Caries prevalence and molar incisor hypomineralisation (MIH) in children. Is there an association? A systematic review. *Eur J Paediatr Dent*. 2023 Dec 1;24(4):312-320. doi: 10.23804/ejpd.2023.1985. PMID: 38015112.
- 3) Mazur M, Corridore D, Ndokaj A, Ardan R, Voza I, Babajko S y cols. MIH and Dental Caries in Children: A Systematic Review and Meta-Analysis. *Healthcare (Basel)*. 2023 Jun 18;11(12):1795. doi: 10.3390/healthcare11121795. PMID: 37372913; PMCID: PMC10298042.
- 4) Rodd HD, Graham A, Tajmehr N, Timms L, Hasmun N. Molar Incisor Hypomineralisation: Current Knowledge and Practice. *Int Dent J*. 2021 Aug;71(4):285-291. doi: 10.1111/idj.12624. PMID: 34286697; PMCID: PMC9275314.
- 5) Alzahrani AY, Alamoudi NMH, El Meligy OAES. Contemporary Understanding of the Etiology and Management of Molar Incisor Hypomineralization: A Literature Review. *Dent J (Basel)*. 2023 Jun 23;11(7):157. doi: 10.3390/dj11070157. PMID: 37504223; PMCID: PMC10378579.
- 6) Lygidakis NA, Garot E, Somani C, Taylor GD, Rouas P, Wong FSL. Best clinical practice guidance for clinicians dealing with children presenting with molar-incisor-hypomineralisation (MIH): an updated European Academy of Paediatric Dentistry policy document. *Eur Arch Paediatr Dent*. 2022 Feb;23(1):3-21. doi: 10.1007/s40368-021-00668-5. PMID: 34669177; PMCID: PMC8926988.
- 7) Enax J, Amaechi BT, Farah R, Liu JA, Schulze zur Wiesche E, Meyer F. Remineralization Strategies for Teeth with Molar Incisor Hypomineralization (MIH): A Literature Review. *Dent J (Basel)*. 2023 Mar 13;11(3):80. doi: 10.3390/dj11030080. PMID: 36975577; PMCID: PMC10047667.
- 8) Schönewolf J, Meyer O, Engels P, Schlickerrieder A, Hickel R, Gruhn V. Artificial intelligence-based diagnostics of molar-incisor- 26 hypomineralization (MIH) on

- intraoral photographs. Clin Oral Investig. 2022 Sep;26(9):5923-5930. doi: 10.1007/s00784-022-04552-4. PMID: 35608684;PMCID: PMC9474479.
- 9) Neumayr J, Frenkel E, Schwarzmaier J, Ammar N, Kessler A, Schwendicke F y cols. External validation of an artificial intelligence-based method for the detection and classification of molar incisor hypomineralisation in dental photographs. J Dent. 2024 Sep;148:105228. doi: 10.1016/j.jdent.2024.105228. PMID: 38972447.
- 10) Felsch M, Meyer O, Schlickerrieder A, Engels P, Schönewolf J, Zöllner F y cols. Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model. NPJ Digit Med. 2023 Oct 25;6(1):198. doi: 10.1038/s41746-023-00944-2. PMID: 37880375; PMCID: PMC10600213.
- 11) Tchito Tchappa C, Mih TA, Tchagna Kouanou A, Fozin Fonzin T, Kuetché Fogang P, Mezatio BA y cols. Biomedical Image Classification in a Big Data Architecture Using Machine Learning Algorithms. J Healthc Eng. 2021 May 30;2021:9998819. doi: 10.1155/2021/9998819. PMID: 34122785; PMCID: PMC8191587.
- 12) Jaiswal M, Mukhtar U, Shakya KS, Laddi A, Singha LA. Computerised assessment-a novel approach for calculation of percentage of hypomineralized lesion on incisors and its correlation with aesthetic concern. J Oral Biol Craniofac Res. 2024 Sep-Oct;14(5):570-577. doi: 10.1016/j.jobcr.2024.07.004. Epub 2024 Jul 23. PMID: 39139516; PMCID: PMC11320481.
- 13) He K, Zhang R, Liang M, Tian K, Luo K, Chen R y cols. The application of deep learning in early enamel demineralization detection. PeerJ. 2025 Jan 2;13:e18593. doi: 10.7717/peerj.18593. PMID: 39763696; PMCID: PMC11700490.
- 14) Alevizakos V, Bekes K, Steffen R, von See C. Artificial intelligence system for training diagnosis and differentiation with molar incisor hypomineralization (MIH) and similar pathologies. Clin Oral Investig. 2022 Dec;26(12):6917-6923. doi: 10.1007/s00784-022-04646-z. Epub 2022 Sep 6. PMID: 36065023. 27
- 15) Portella PD, de Oliveira LF, Ferreira MFC, Dias BC, de Souza JF, Assunção LRDS. Improving accuracy of early dental carious lesions detection using deep learning based automated method. Clin Oral Investig. 2023 Dec;27(12):7663-7670. doi: 10.1007/s00784-023-05355-x. Epub 2023 Oct 31. PMID: 37906303.

16) Veseli E. Early detection of MIH in children by using artificial intelligence. Eur Arch Paediatr Dent. 2024 Dec;25(6):899-900. doi: 10.1007/s40368-024-00931- 5. Epub 2024 Aug 1. PMID: 39090506.

Anexos:

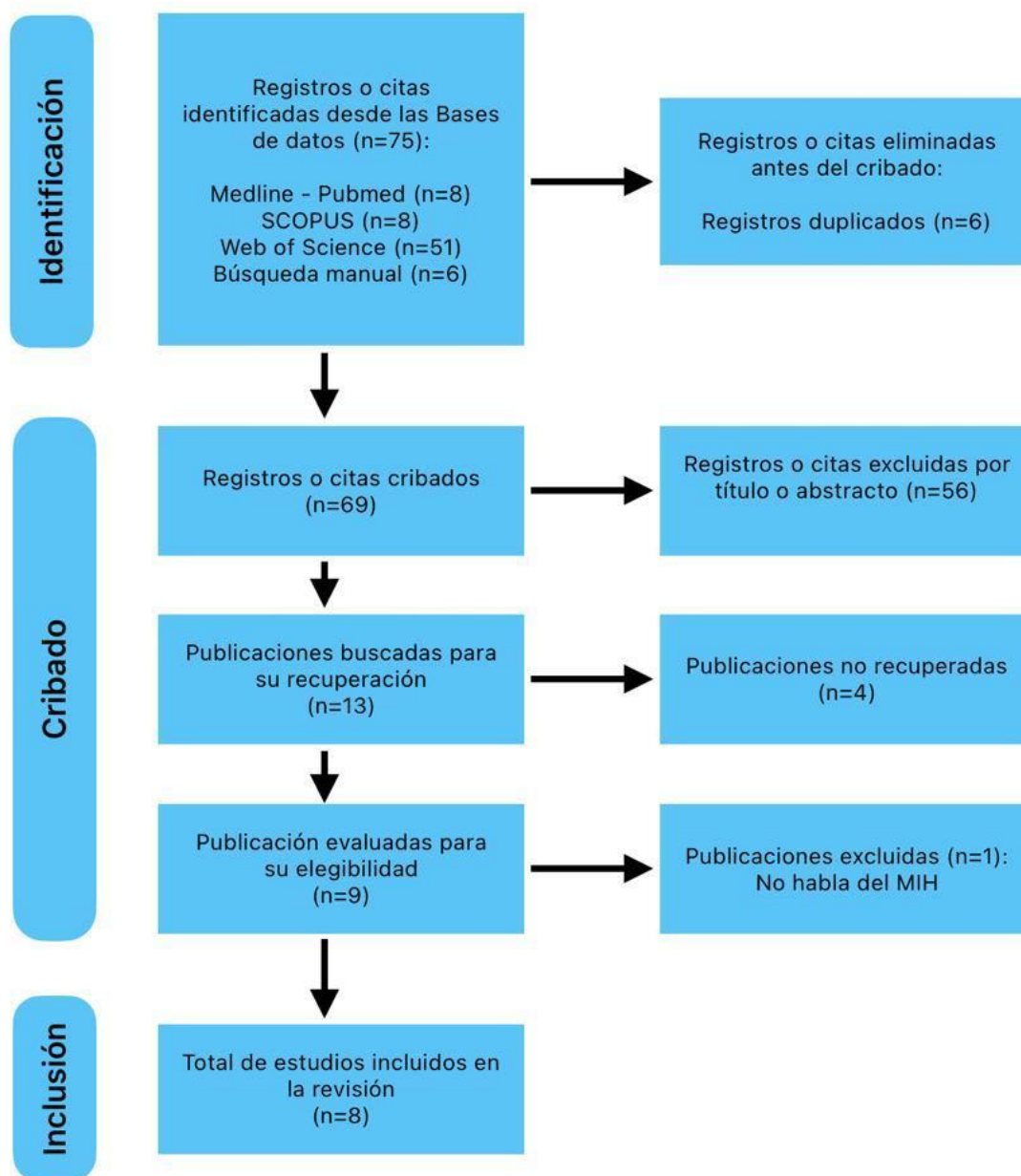


Fig. 1. Diagrama de flujo de búsqueda y proceso de selección de títulos durante la revisión sistemática.

Autor (Año)	Tipo de Estudio	Muestra	Modelo de IA	Validación	Tipo de Imagen
Schönewolf y cols., 2022	Estudio diagnóstico observacional no randomizado	3,241 imágenes	ResNeXt-101-32×8d	Interna	Fotografías intraorales
Kühnisch y cols., 2024	Estudio observacional no randomizado	455 imágenes web	IA profunda (modelo abierto)	Externa	Imágenes web dentales
Felsch y cols., 2023	Estudio in vitro	18,179 imágenes	SegFormer-B5	Interna	Fotografías clínicas HD
Jaiswal y cols., 2024	Estudio observacional no randomizado	50 niños	SML+ OpenCV	Photoshop CS7	Fotos intraorales estandarizadas
He y cols., 2025	Estudio observacional no randomizado	208 pacientes (624 imágenes)	Mask R-CNN + SAM	Comparación con clínicos	Fotos intraorales múltiples vistas
Alevizakos y cols., 2022	Estudio in vitro	462 imágenes	VGG16, DenseNet, ResNet AlexNet	Cruzada	Fotos dentales a color
Portella y cols., 2023	Estudio in vitro	2,481 dientes (140 imágenes test)	VGG-19	Clínica	Fotos oclusales alta resolución
Veseli y cols., 2024	Carta al editor / Descriptivo	No especificado	Modelo DL	Expertos clínicos	Fotografías clínicas

Tabla 1: Resumen – Características de los estudios incluidos

Autor (Año)	Precisión (%)	Sensibilidad (%)	Especificidad (%)	F1-Score	Valor Predictivo Positivo	Hallazgos Relevantes
Schönewol y cols., 2022	95.2	78.6	97.3	No especificado	No indicado	Alta precisión general; menor sensibilidad en casos complejos.
Kühnisch y cols, 2024	94.3	81.7–92.8	91.9–98.7	No indicado	No indicado	Acuerdo 89.2% con expertos; buen desempeño.
Felsch y cols., 2023	97.8	No específico	No específico	AP: 82.7–90.2	No indicado	Rendimiento excelente, IoU: 95.9%.
Jaiswal y cols., 2024	Media 34.51%	No aplica	No aplica	No aplica	No aplica	Extensión >60% se asocia con 80% preocupación estética.
He y cols., 2025	No directo	93.5	78.9	0.856	0.968	IA mejora F1 de 0.713 a 0.949 con ayuda.
Alevizakos y cols., 2022	83.98–92.86	No indicado	No indicado	No indicado	No indicado	Mejor modelo: DenseNet121.
Portella y cols., 2023	87.9	82.7	80.0	0.887	0.949	Mejora significativa en precisión con IA.
Veseli y cols., 2024	No indicado	No indicado	No indicado	No indicado	No indicado	Alto rendimiento estimado, sin métricas exactas.

Tabla 2: Resultados diagnósticos de los estudios