



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS DE LA SALUD
CARRERA DE ODONTOLOGÍA**

TEMA:

**Correlación diagnóstica entre modelos de inteligencia artificial y
criterios clínicos radiográficos en patologías pulpares.**

AUTOR:

Herrero Quiroz, Aitana

**Trabajo de titulación previo a la obtención del título de
ODONTÓLOGA**

TUTOR:

García Guerrero, Yara Anna Paula

Guayaquil, Ecuador

10 de septiembre del 2026

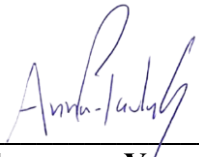


**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS DE LA SALUD
CARRERA DE ODONTOLOGÍA**

CERTIFICACIÓN

Certificamos que el presente trabajo de titulación fue realizado en su totalidad por **Herrero Quiroz, Aitana**, como requerimiento para la obtención del título de **Odontóloga**.

TUTOR (A)

f. 
García Guerrero, Yara Anna Paula

DIRECTOR DE LA CARRERA

f. _____
Bermúdez Velásquez, Andrea Cecilia

Guayaquil, a los 10 días del mes de septiembre del año 2026



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS DE LA SALUD
CARRERA DE ODONTOLOGÍA**

DECLARACIÓN DE RESPONSABILIDAD

Yo, Herrero Quiroz, Aitana

DECLARO QUE:

El Trabajo de Titulación, **Correlación diagnóstica entre modelos de inteligencia artificial y criterios clínicos radiográficos en patologías pulpares**, previo a la obtención del título de **Odontóloga** ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los 10 días del mes de septiembre del año 2026

LA AUTORA

f. 

Herrero Quiroz, Aitana



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS DE LA SALUD
CARRERA DE ODONTOLOGÍA

AUTORIZACIÓN

Yo, **Herrero Quiroz, Aitana**

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Correlación diagnóstica entre modelos de inteligencia artificial y criterios clínicos radiográficos en patologías pulpares**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 10 días del mes de septiembre del año 2026

LA AUTORA

f. 

Herrero Quiroz, Aitana



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS DE LA SALUD
CARRERA DE ODONTOLOGÍA

REPORTE COMPILATIO

	Informe de análisis Compilatio Magister+ UCSG-EC- Universidad Católica de Santiago de Guayaquil
PARA COMPILATIO AITANA ID : cf332170caa534fe582acefbfe4f137957fb6347	
0% Textos sospechosos	
Nombre del fichero : PARA COMPILATIO AITANA.txt Tamaño del archivo original : 312,23 kB Número de palabras : 5112	Depositante : Yara Anna-Paula García Guerrero Fecha de depósito : 30 de agosto de 2026 Tipo de carga : interface fecha de fin de análisis : 30 de agosto de 2026

TUTOR (A)

f. _____
García Guerrero, Yara Anna Paula

AGRADECIMIENTO

A mis papás, por darme la oportunidad de elegir mi profesión. Gracias por impulsarme a ser una mujer capaz de alcanzar todo lo que se proponga.

A mi familia y amigos de mis papás, gracias por sus buenos deseos durante este camino y por confiar en mí, incluso al convertirse en mis pacientes. Siempre llevaré en mi corazón su apoyo, pues fue una parte fundamental para mi trayectoria.

A mis queridas amistades, con las que inicié y las que conocí mejor este año, por compartir conmigo risas, largas horas de estudio y los desafíos de las clínicas. Gracias por estar hombro con hombro en todo momento. No puedo esperar para verlas/os cumplir sus sueños y alcanzar todos sus logros. Sé que lo mejor está por venir para cada una/o de ustedes.

A John Paul, por caminar conmigo el tramo final de nuestra carrera universitaria, siendo un pilar el uno para el otro y sé que nuestro constante esfuerzo nos llevará hacia los sueños que anhelamos. Gracias por prestarme aquel Endoblock que, sin saberlo, fue el inicio de una hermosa historia.

A mi tutora de tesis, la Dra. Anna Paula, y a cada doctor que dejó una huella en mi formación: les agradezco por exigirme, guiarme y ayudarme a crecer. Asimismo, a la Dra. Maru y su equipo de Healteeth, por abrirme las puertas a la práctica clínica cuando apenas conocía el sillón dental, y enseñarme el valor de transmitir confianza y seguridad a cada paciente.

Por último, pero no menos importante, a Dios, por darme fortaleza y guiar cada uno de mis pasos. A Él encomiendo este logro, sabiendo que es fruto de mi esfuerzo, pero también parte de Sus planes para mi vida.

DEDICATORIA

A mi mamá, Ceci, y a mi papá, Paco. Gracias por confiar siempre en mí y acompañarme en cada paso. Hoy, después de dar lo mejor de mí, me siento lista para nuevos caminos. Solo le pido a Dios que me permita llevarlos siempre conmigo, compartir cada logro y disfrutar juntos de todo lo que, con tanto amor y sacrificio, me han enseñado a construir.

A mis hermanos mayores, por apoyarme incondicionalmente y creer siempre en mis capacidades. Gracias por impulsarme a seguir adelante y por recordarme, con cariño, que tenía que graduarme rápido para convertirme en la dentista oficial de mis sobrinos/as. Este logro también lleva un poquito de ustedes.

Y a mi hermanito, el menor de la casa, pero quien estuvo muy cerca de mí durante todo este proceso. Gracias por apoyarme en los días difíciles, cuando llegaba cansada y sin energías de la clínica. Ahora que iniciaste tu propia etapa universitaria, espero poder hacer por ti lo mismo que tú hiciste por mí: estar a tu lado en las buenas y en las malas.

Los amo. Gracias de corazón, por todo y para siempre.



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS DE LA SALUD
CARRERA DE ODONTOLOGÍA**

TRIBUNAL DE SUSTENTACIÓN

f. _____

DRA. ANDREA CECILIA BERMÚDEZ VELÁSQUEZ
DECANO O DIRECTOR DE CARRERA

f. _____

DRA. ESTEFANÍA DEL ROCÍO OCAMPO POMA
COORDINADOR DEL ÁREA O DOCENTE DE LA CARRERA

f. _____

DRA. ANDREA CECILIA BERMÚDEZ VELÁSQUEZ
OPONENTE



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS DE LA SALUD
CARRERA DE ODONTOLOGÍA**

CALIFICACIÓN

TUTOR (A)

**f. _____
García Guerrero, Anna-Paula**

RESUMEN

Introducción: La inteligencia artificial ha adquirido relevancia en el área de la salud por su gran capacidad de procesamiento de datos clínicos en función de optimizar la atención. En endodoncia, los modelos de lenguaje de gran escala (LLMs) podrían ser una herramienta de apoyo para la identificación de patologías pulpares, cuyo rendimiento está sujeto a factores como la complejidad del caso clínico.

Objetivo: Determinar la correlación diagnóstica entre modelos de inteligencia artificial y criterios clínico-radiográficos en el diagnóstico de patologías pulpares y periapicales. **Materiales y métodos:**

Se realizó un estudio con 156 casos clínicos. El análisis se realizó en SPSS (IBM, versión 29), mediante coeficiente Kappa de Cohen, prueba de Friedman con ajuste de Bonferroni, y la prueba exacta de Fisher-Freeman-Halton. **Resultados:** Hubo concordancia moderada y estadísticamente significativa. No se

observaron diferencias significativas entre modelos, aunque la concordancia disminuyó conforme aumentó la complejidad clínica, particularmente en los casos Clase III. **Conclusiones:** ChatGPT,

Gemini y Claude mostraron utilidad potencial como herramientas complementarias de apoyo diagnóstico, aunque su desempeño depende de la complejidad del caso y no sustituye la valoración

clínica profesional.

Palabras clave: Inteligencia artificial, endodoncia, diagnóstico pulpar, diagnóstico periapical, complejidad diagnóstica, desempeño diagnóstico.

ABSTRACT

Introduction: Artificial intelligence has gained relevance in the field of health for its great capacity to process clinical data to optimize care. In endodontics, large language models (LLMs) could be a supporting tool for identifying pulpar pathologies, the performance of which is subject to factors such as clinical case complexity. **Objective:** To determine the diagnostic correlation between artificial intelligence models and clinical-radiographic criteria in the diagnosis of pulpar and periapical pathologies. **Materials and methods:** A study was conducted with 156 clinical cases. The analysis was performed using SPSS (IBM, version 29), applying Cohen's Kappa coefficient, the Friedman test with Bonferroni adjustment, and the Fisher-Freeman-Halton exact test. **Results:** There was a moderate and statistically significant agreement. No significant differences were observed between models, although concordance decreased as clinical complexity increased, particularly in Class III cases. **Conclusions:** ChatGPT, Gemini, and Claude showed potential usefulness as complementary diagnostic support tools, although their performance depends on the complexity of the case and does not replace professional clinical evaluation.

Keywords: Artificial intelligence, endodontics, pulpar diagnosis, periapical diagnosis, diagnostic complexity, diagnostic performance.

Introducción

La endodoncia es una de las ramas de la Odontología cuya misión se resume en manejar el dolor dental, en especial cuando se trata de urgencias. Aunque el estudio de los tejidos del interior del diente contribuye a identificar los agentes etiológicos que producen las lesiones pulpares, el examen clínico-radiográfico es la clave para un correcto diagnóstico endodóntico. Por ello, en el 2008, la Asociación Americana de Endodoncistas (AAE) estableció una guía en la que se pueda estandarizar los signos y síntomas de una enfermedad de la pulpa, en sus distintas etapas¹. El diagnóstico se basa en aspectos tanto subjetivos como objetivos, cuya integración requiere experiencia y criterio profesional².

La periodontitis apical constituye una de las principales consecuencias asociadas con la progresión de las patologías pulpares. Tibúrcio-Machado et al. (2021) reportaron una prevalencia global del 52% a nivel individual y del 5% a nivel dentario. Asimismo, se observó una prevalencia del 39% en dientes previamente tratados endodónticamente, frente al 3% en dientes sin tratamiento endodóntico previo³. La inflamación pulpar producida por diferentes agentes

etiológicos puede evolucionar hacia estados irreversibles y posteriormente hacia necrosis pulpar, por lo que una evaluación clínica precisa resulta fundamental para establecer el diagnóstico y manejo terapéutico correspondiente⁴. La extensión del proceso infeccioso hacia los tejidos perirradiculares puede, además, favorecer el desarrollo de lesiones periapicales⁵.

En Latinoamérica, Cuba-Cueto et al. (2022) reportaron que la pulpitis irreversible representó el 53,8% de las patologías pulpares evaluadas. En cuanto a las patologías periapicales, se observó una frecuencia del 44,5% de periodontitis apical y del 74,4% de absceso periapical agudo⁶. En Ecuador, específicamente en la ciudad de Cuenca, Cedillo-Orellana et al. (2019) encontraron que la pulpitis irreversible representó el 62% de las patologías pulpares estudiadas, mientras que la necrosis pulpar correspondió al 38%⁷. Estos datos evidencian la frecuencia de las patologías pulpares y periapicales y resaltan la importancia de establecer métodos diagnósticos precisos para orientar adecuadamente la conducta terapéutica.

A pesar de la existencia de criterios diagnósticos estandarizados, hay situaciones clínicas en las que un conjunto de signos y síntomas no puede encasillarse en un diagnóstico concreto. Las pruebas de sensibilidad pulpar, aunque son un componente importante del proceso diagnóstico, no siempre indican con absoluta certeza el verdadero estado biológico de la pulpa. Asimismo, el diagnóstico puede presentar variabilidad según la interpretación y experiencia del profesional. Azim et al. (2022) establecen que la terminología actual de la AAE puede resultar insuficiente para determinadas situaciones clínicas. Esto resulta en una falta de consenso entre especialistas de Endodoncia y odontólogos generales por sus distintas formas de aplicar los criterios diagnósticos, proponiendo incluso las categorías de condición pulpar inconclusa y condición perirradicular inconclusa para aquellos casos que no pueden clasificarse con seguridad dentro de la terminología diagnóstica existente⁸.

Ante estas limitaciones, las herramientas basadas en inteligencia artificial han comenzado a ser estudiadas como sistemas complementarios para procesar

información clínica y apoyar la toma de decisiones. Entre ellas se encuentran los modelos de lenguaje de gran escala o Large Language Models (LLMs), capaces de procesar grandes cantidades de información y generar respuestas a partir de las instrucciones o *prompts* proporcionados por el usuario. En Odontología, la aplicación de los LLMs ha avanzado más allá de una fase exclusivamente experimental, alcanzando principalmente un nivel de implementación práctica. Sin embargo, su incorporación requiere evaluación y monitoreo constante debido a los desafíos relacionados con su desempeño y confiabilidad⁹.

En este sentido, Mendonça de Moura et al. (2024), al comparar diferentes modelos de inteligencia artificial en el diagnóstico de patologías pulpares y perirradiculares, reportaron una exactitud diagnóstica del 86,4% para Bing y del 85,3% para ChatGPT 4.0, mientras que ChatGPT 3.5 alcanzó el 46,5% y Bard el 28,6%, con diferencias estadísticamente significativas entre los modelos ($p < 0,001$). En cuanto a las recomendaciones terapéuticas, ChatGPT 4.0 alcanzó una exactitud del 94,4%, seguido de Bing con 93,2%, ChatGPT 3.5 con 86,3% y Bard con 75%; además, se reportó una

consistencia global del 98,29% en las respuestas proporcionadas. Estos resultados demuestran que el potencial diagnóstico de los modelos de inteligencia artificial puede ser elevado, aunque no uniforme entre las diferentes plataformas¹⁰.

De forma complementaria, Özbay et al. (2025) evaluaron el desempeño de distintos modelos de lenguaje en la toma de decisiones clínicas en Endodoncia, encontrando diferencias significativas entre los sistemas analizados ($p = 0,008$) y un mejor desempeño de ChatGPT-4 frente a otros modelos evaluados. Los autores señalan que, con un desarrollo adecuado y el respaldo de literatura científica, estos sistemas podrían contribuir como herramientas complementarias en la identificación de patologías endodónticas y la planificación terapéutica basada en evidencia¹¹.

Sea estudiante de pregrado, odontólogo general o especialista, es fundamental reconocer los factores de dificultad antes del tratamiento, para evitar iatrogenias o un fracaso de este. Para ello, se han desarrollado herramientas de evaluación de complejidad como el Endodontic Complexity Assessment

Tool (E-CAT), presentado por Essam et al. (2021). Esta herramienta digital permite valorar sistemáticamente diferentes factores relacionados con la dificultad del caso endodóntico y clasificarlo en tres niveles de complejidad: Clase I, casos sin complicaciones; Clase II, casos de complejidad moderada; y Clase III, casos de alta complejidad. De esta manera, proporciona una estructura estandarizada para identificar factores que pueden modificar la dificultad del manejo endodóntico y orientar la planificación o derivación del caso¹².

La información proporcionada a los modelos de inteligencia artificial también puede condicionar su rendimiento. Atakır et al. (2025) demostraron que el desempeño diagnóstico de ChatGPT-4o varía según la información suministrada, incluyendo la disponibilidad de datos clínicos, imágenes y opciones de respuesta¹³. Este hallazgo adquiere relevancia en el contexto endodóntico, donde el diagnóstico depende de la integración de múltiples elementos clínicos y radiográficos y donde la complejidad del caso podría representar un factor adicional relacionado con el desempeño de estas herramientas.

En consecuencia, aunque los modelos de inteligencia artificial han demostrado resultados prometedores como herramientas complementarias para el diagnóstico, la variabilidad observada entre modelos y la influencia de la información clínica suministrada justifican continuar evaluando su desempeño frente a criterios diagnósticos establecidos y ante casos con diferentes grados de complejidad. Por ello, el objetivo de este estudio es determinar la correlación diagnóstica entre modelos de inteligencia artificial y criterios clínico-radiográficos en el diagnóstico de patologías pulpares y periapicales en pacientes atendidos en la clínica de Endodoncia II de la Universidad Católica de Santiago de Guayaquil durante el período B-2025 y A-2026.

Materiales y métodos

El estudio se desarrolla bajo un enfoque cuantitativo; de tipo retrospectivo, transversal y comparativo. Asimismo, emplea un diseño observacional y no experimental.

El universo de estudio abarca la totalidad de pacientes atendidos en la clínica de Endodoncia II de la Universidad Católica de Santiago de

Guayaquil durante los períodos académicos B-2025 y A-2026. Para determinar el tamaño de muestra, se aplicó la fórmula para proporciones en poblaciones finitas, estableciendo un nivel de confianza del 95% y un error máximo admisible del 5%. A partir de una población de 262 pacientes, se obtuvo una muestra mínima de 156 pacientes.

La unidad de análisis consistió en un único caso clínico, es decir, un diente por paciente. Si el paciente contaba con más dientes elegibles, se seleccionó uno de manera aleatoria para evitar la sobrerepresentación de datos y sesgos en el análisis estadístico.

Criterios de inclusión:

Se incluyeron casos clínicos con diagnósticos de pulpitis irreversible sintomática, pulpitis irreversible asintomática, necrosis pulpar y terapia previamente iniciada, que presentaran registros clínicos completos, considerando: historia clínica de Endodoncia de la UCSG con diagnóstico pulpar y periapical firmado y aprobado por el docente a cargo, consentimiento informado firmado y radiografía periapical inicial disponible.

Criterios de exclusión:

Se excluyeron casos clínicos con diagnósticos de pulpa normal, pulpitis reversible y pulpa previamente tratada, así como aquellos que presentaran registros clínicos incompletos, definidos por la ausencia de historia clínica de Endodoncia, diagnóstico pulpar y periapical firmado y aprobado por el docente a cargo, consentimiento informado y/o radiografía periapical inicial.

Procedimiento:

1°. Se seleccionaron plataformas de IA más utilizadas para automatización de tareas, razonamiento profundo y multimodalidad: ChatGPT5.6 Sol, Gemini 3.6 Flash y Claude Sonnet 5.

2°. Se diseñó un prompt de entrada, extrayendo los datos clínico-radiográficos de las historias clínicas para ser procesados de manera uniforme por las tres plataformas de IA (Anexo 1).

3°. Se revisó de manera sistematizada las carpetas con los registros clínicos de los pacientes desde el Semestre B-2025 hasta el A-2026 en su totalidad, tomando en cuenta los criterios de inclusión y exclusión.

4°. Por medio de una ficha digital creada especialmente para este estudio con las funciones de programación de ChatGPT Plus (versión 5.6 Sol), se registraron los datos referentes al examen clínico, sintomatología y análisis radiográfico, junto con el diagnóstico pulpar y periapical aprobado por los docentes del área (Anexo 2).

Se agregó funciones como completar automáticamente los datos en el prompt, adjuntando la imagen radiográfica en formato JPG; así como identificar si el nivel de completitud de los datos era alto (información clínica, hallazgos radiográficos y categorías diagnósticas disponibles), moderado (presenta algún dato incompleto) o bajo (insuficiente para emitir el diagnóstico).

Además, se incorporó un medidor del grado de complejidad basado en los criterios diagnósticos de la AAE, tomando como referencia la clasificación del software británico E-CAT, y adaptándola al contexto del estudio. A partir de ello, se establecieron las siguientes categorías:

- Clase I: Caso con información clínica suficiente y concordante, con radiografía

aceptable y acorde a los hallazgos.

- Clase II: Caso con una limitación concreta como calidad radiográfica u omisión de un signo, pero que aún permite una interpretación razonable.
- Clase III: Caso con dos o más limitaciones diagnósticas o dificultades, presentando alto riesgo de error diagnóstico.

5°. Se ingresó el prompt automatizado en las plataformas de IA de elección, y se exportaron todos los casos en un PDF para su análisis manual.

6°. Se realizó el registro de los datos en un documento de Microsoft Excel para su tabulación y posteriormente, iniciar su proceso estadístico.

Análisis y procesamiento de datos

Los datos se recolectaron en una matriz en el programa Microsoft Excel, posteriormente fueron importados al programa SPSS (IBM versión 29) para el procesamiento de los resultados. Se utilizó estadística descriptiva para las variables edad (media y desviación estándar) y sexo (frecuencia absoluta y porcentaje). Para establecer la concordancia diagnóstica entre los

modelos de inteligencia artificial y el estado de salud pulpar registrado según criterios clínicos se calculó el coeficiente Kappa de Cohen. La evaluación del desempeño diagnóstico de las respuestas emitidas por los modelos de inteligencia artificial frente al diagnóstico clínico-radiográfico de referencia se utilizó la prueba no paramétrica de Friedman para muestra dependientes, las comparaciones post hoc por pares se ajustaron mediante corrección de Bonferroni.

La influencia del nivel de complejidad del caso clínico en la concordancia diagnóstica entre los modelos de inteligencia artificial y el estado clínico-radiográfico de patologías pulpares y periapicales se realizó mediante análisis bivariado con la prueba exacta de Fisher-Freeman-Halton (casillas con recuentos inferiores a 5). Para todas las pruebas estadísticas se estableció un nivel de significación de $p < 0,05$.

Resultados

Se analizó el estado de salud pulpar y periapical clínico-radiográfico registrado según criterios clínicos en una muestra de 156 pacientes. Se observó predominio del sexo femenino, que representó el 67,3%

(n=105), mientras que el 32,7% (n=51) correspondió al sexo masculino. La edad estuvo disponible en 150 pacientes (96,2%), con una media de 44,36 años y una desviación estándar de 13,17 años; en seis casos (3,8%) no se contó con información sobre esta variable.

La Tabla 1 presenta la concordancia entre los diagnósticos pulpares y periapicales emitidos por los modelos de inteligencia artificial y el diagnóstico clínico-radiográfico de referencia.

1°. Salud pulpar: Los coeficientes Kappa mostraron una concordancia moderada para los tres modelos, con valores de entre $\kappa = 0,431$ y $\kappa = 0,513$, todos estadísticamente significativos ($p < 0,001$) (Tabla 1 y Figura 1). En la concordancia según el diagnóstico pulpar, la mayor concordancia se presentó para la necrosis pulpar, identificada correctamente en el 85,1% de los casos por ChatGPT y en el 86,2% tanto por Gemini como por Claude (Tabla 1).

Para la pulpitis irreversible sintomática, ChatGPT alcanzó una concordancia del 90,9%, seguido de Gemini con 86,4% y Claude con 68,2%. En los casos de terapia iniciada

previamente, la concordancia fue del 62,5% para ChatGPT y del 50,0% para Gemini y Claude. La pulpitis irreversible asintomática constituyó la categoría de mayor dificultad diagnóstica para los tres modelos, con concordancias de apenas 17,9% para ChatGPT, 25,6% para Gemini y 7,7% para Claude. En conjunto, estos resultados indican que los tres modelos presentaron una concordancia global moderada con el diagnóstico pulpar de referencia, aunque su capacidad de reproducir correctamente el diagnóstico varió considerablemente según la categoría clínica evaluada (Tabla 1).

2°. Salud periapical: Los coeficientes Kappa variaron entre $\kappa = 0,483$ y $\kappa = 0,525$ todos estadísticamente significativos ($p < 0,001$) (Tabla 1 y Figura 1). En la concordancia según el diagnóstico periapical, el absceso apical crónico presentó una concordancia elevada y homogénea entre los tres modelos, con 85,7% de identificaciones correctas. Para la periodontitis apical sintomática, ChatGPT y Claude alcanzaron una concordancia del 85,2%, mientras que Gemini presentó 74,1%. En los tejidos apicales normales, la concordancia fue menor, con valores de 68,2% para

ChatGPT, 65,2% para Gemini y 63,6% para Claude (Tabla 1).

La periodontitis apical asintomática mostró mayores dificultades diagnósticas, con una concordancia de 46,8% para ChatGPT y Gemini y de 44,7% para Claude. En cuanto al absceso apical agudo, ninguno de los modelos coincidió con el diagnóstico de referencia; sin embargo, esta

categoría estuvo representada por un solo caso. De manera similar, la concordancia del 100,0% observada para osteítis condensante corresponde también a un único caso, por lo que estos porcentajes no deben interpretarse como evidencia de un desempeño diagnóstico elevado o deficiente para dichas categorías (Tabla 1).

Tabla 1. Concordancia diagnóstica entre IA y el estado de salud pulpar y periapical registrado según criterios clínicos

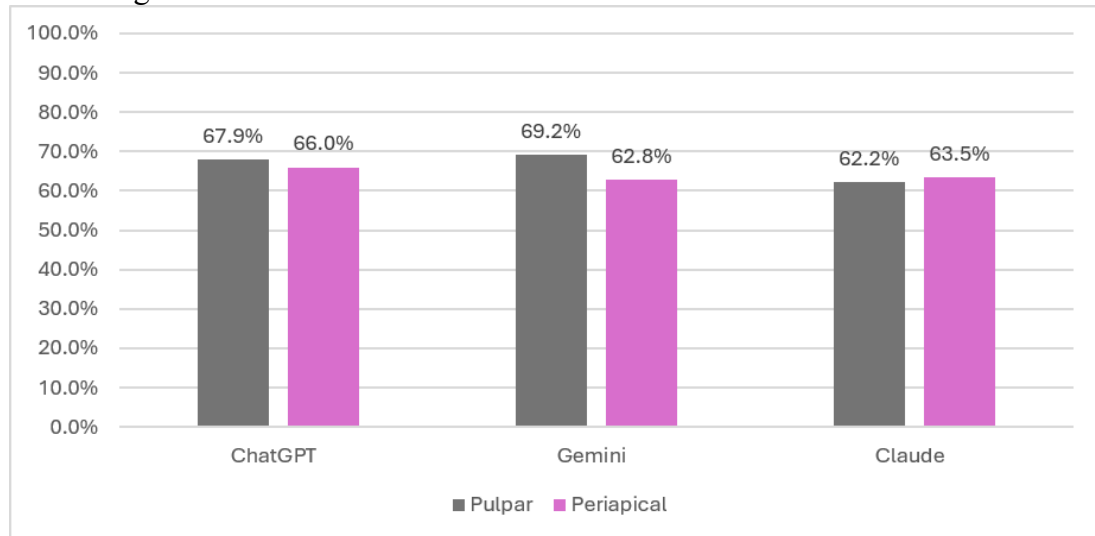
Diagnóstico pulpar de referencia	ChatGPT n (%)	Gemini n (%)	Claude n (%)
Pulpitis irreversible sintomática (n = 22)	20 (90,9)	19 (86,4)	15 (68,2)
Pulpitis irreversible asintomática (n = 39)	7 (17,9)	10 (25,6)	3 (7,7)
Necrosis pulpar (n = 87)	74 (85,1)	75 (86,2)	75 (86,2)
Terapia iniciada previamente (n = 8)	5 (62,5)	4 (50,0)	4 (50,0)
Acuerdo observado global (n = 156)	106 (67,9)	108 (69,2)	97 (62,2)
Kappa de Cohen (κ)	0,51	0,513	0,431
Valor-p	<0,001	<0,001	<0,001
Diagnóstico periapical de referencia	ChatGPT n (%)	Gemini n (%)	Claude n (%)
Tejidos apicales normales (n = 66)	45 (68,2)	43 (65,2)	42 (63,6)
Periodontitis apical sintomática (n = 27)	23 (85,2)	20 (74,1)	23 (85,2)
Periodontitis apical asintomática (n = 47)	22 (46,8)	22 (46,8)	21 (44,7)
Absceso apical agudo (n = 1)	0 (0,0)	0 (0,0)	0 (0,0)
Absceso apical crónico (n = 14)	12 (85,7)	12 (85,7)	12 (85,7)
Osteítis condensante (n = 1)	1 (100,0)	1 (100,0)	1 (100,0)
Acuerdo observado global (n = 156)	103 (66,0)	98 (62,8)	99 (63,5)
Kappa de Cohen (κ)	0,525	0,483	0,494
Valor-p	<0,001	<0,001	<0,001

Nota. Los valores por categoría representan el número y porcentaje de casos en los que el diagnóstico emitido por cada modelo coincidió exactamente con el diagnóstico pulpar de referencia. La concordancia global corresponde a la proporción total de coincidencias exactas. κ = coeficiente Kappa de Cohen. Los porcentajes por categoría se calcularon respecto del número de casos correspondiente a cada diagnóstico de referencia.

Para el diagnóstico de salud pulpar, Gemini presentó la mayor proporción de concordancia exacta con el diagnóstico de referencia, seguido de

ChatGPT y Claude. En el diagnóstico periapical, el mayor acuerdo global correspondió a ChatGPT, seguido de Claude y Gemini (Figura 1).

Figura 1. Concordancia diagnóstica exacta de los modelos de inteligencia artificial con el diagnóstico de referencia



Nota: Las barras representan el porcentaje de concordancia diagnóstica exacta de cada modelo de inteligencia artificial con el diagnóstico clínico-radiográfico de referencia, para los diagnósticos pulpar y periapical (n = 156).

Los resultados de la Figura 2 muestran el análisis del desempeño diagnóstico de las respuestas emitidas por los modelos de inteligencia artificial frente al diagnóstico clínico-radiográfico de referencia. Descriptivamente en el desempeño diagnóstico, ChatGPT presentó la mayor proporción de respuestas clasificadas como correctas, seguido de Gemini y Claude. Las respuestas parcialmente correctas fueron mayoría en Claude, seguido de Gemini y ChatGPT, mientras que las respuestas incorrectas fueron menos

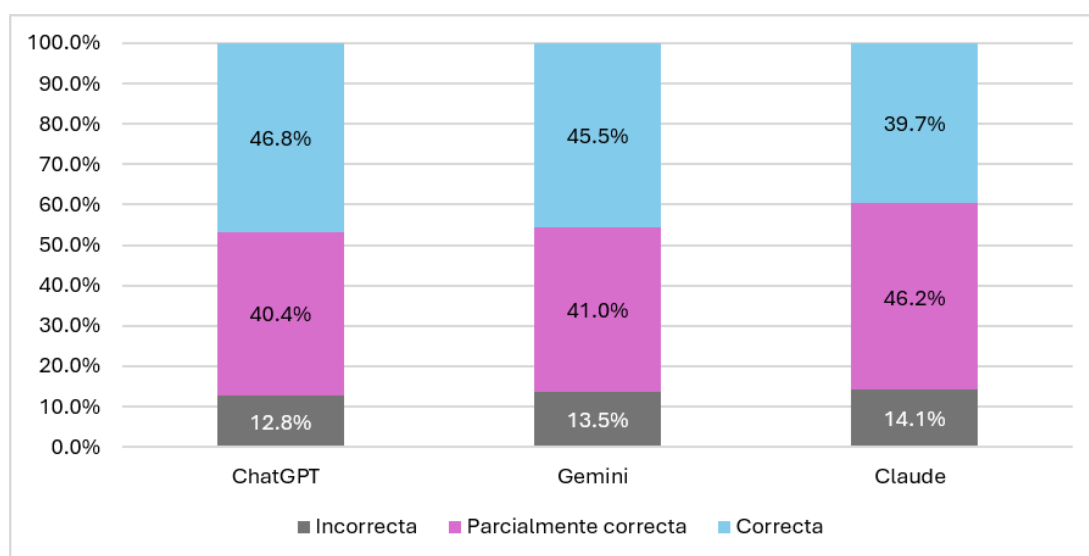
frecuentes en los tres modelos, con porcentajes entre 12,8% y 14,1%.

La prueba de Friedman evidenció diferencias estadísticamente significativas en la distribución global del desempeño entre los tres modelos ($\chi^2=9,624$; $gl=2$; $p=0,008$). Sin embargo, las comparaciones post hoc por pares con corrección de Bonferroni no mostraron diferencias estadísticamente significativas entre ChatGPT y Gemini (p ajustado=1,000), ChatGPT y Claude (p ajustado=0,846), ni Gemini y Claude (p

ajustado=1,000). Existe un patrón: ChatGPT y Gemini tienden a presentar mejores resultados que Claude, pero las diferencias son pequeñas. Friedman, al utilizar la información conjunta de los 156 casos relacionados, detecta

heterogeneidad global; sin embargo, cuando se fragmentó el análisis en pares y con la aplicación de Bonferroni, ninguna diferencia individual proporciona evidencia suficiente.

Figura 2. Distribución del desempeño diagnóstico según modelo de inteligencia artificial

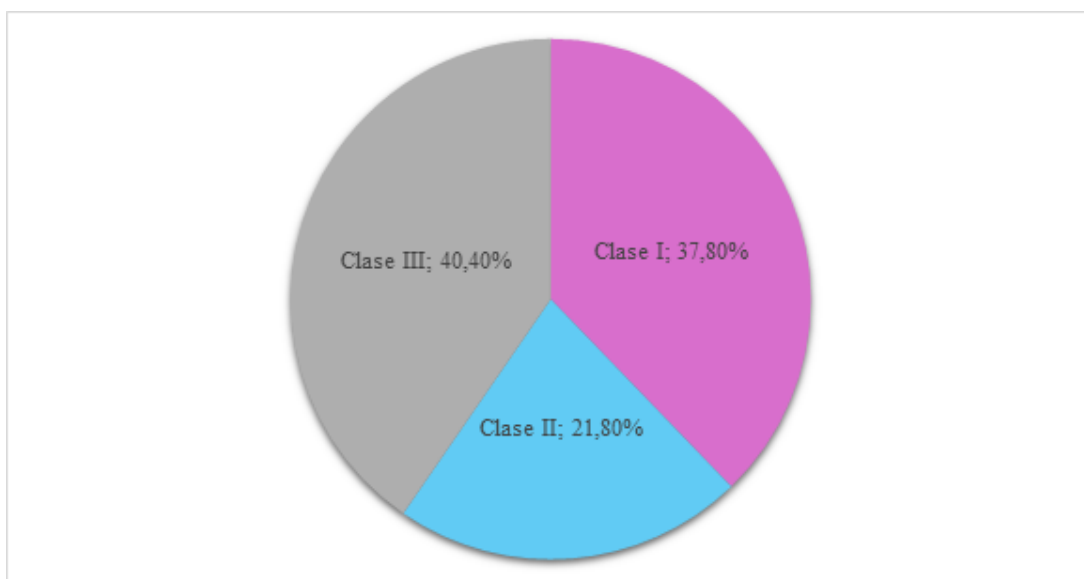


Nota: El desempeño diagnóstico fue clasificado como incorrecto (1 punto), parcialmente correcto (2 puntos) y correcto (3 puntos). La comparación global entre los tres modelos se realizó mediante la prueba no paramétrica de Friedman para muestras relacionadas. Las comparaciones post hoc por pares se ajustaron mediante corrección de Bonferroni.

La Figura 3 muestra la distribución de los casos clínicos según su nivel de complejidad. La Clase III, evidenció

predominio con mayor porcentaje de casos (n = 63), seguida de la Clase I (n = 59) y la Clase II (n=34).

Figura 3. Distribución del nivel de complejidad del caso clínico en la concordancia diagnóstica



Nota. El gráfico en pastel representa la distribución de los 156 casos clínicos según su nivel de complejidad diagnóstica: Clase I (n=59), Clase II (n=34) y Clase III (n=63).

En relación con la calidad radiográfica FGDP, fue del 80,77% para la categoría “A” y del 19,23% para la categoría “N”. En cuanto al nivel de información, el 85,26% fue alta y el 14,74% fue moderada.

Los resultados de la Tabla 2 evidencian una disminución progresiva del desempeño diagnóstico a medida que aumenta la complejidad del caso clínico. Además, se observó una asociación estadísticamente significativa entre el nivel de complejidad del caso clínico y el desempeño diagnóstico de los tres

modelos de inteligencia artificial (p-valor <0,001).

En los casos Clase I predominaron las respuestas correctas, alcanzando el 81,4% en ChatGPT y el 76,3% tanto en Gemini como en Claude. En los casos Clase II, la proporción de respuestas correctas disminuyó al 50,0%, 47,1% y 38,2%, respectivamente. Para los casos Clase III, esta reducción fue aún más marcada, donde el 12,7% de las respuestas de ChatGPT, el 15,9% de Gemini y el 6,3% de Claude fueron correctas (Tabla 2).

Tabla 2. Asociación del nivel de complejidad del caso clínico en la concordancia diagnóstica entre los modelos de IA y el estado clínico-radiográfico de patologías pulpares y periapicales

Modelo de IA / Complejidad	Incorrecta n (%)	Parcialmente correcta n (%)	Correcta n (%)	P-valor
ChatGPT				
Clase I	1 (1,7)	10 (16,9)	48 (81,4)	<0,001
Clase II	0 (0,0)	17 (50,0)	17 (50,0)	
Clase III	19 (30,2)	36 (57,1)	8 (12,7)	
Gemini				
Clase I	2 (3,4)	12 (20,3)	45 (76,3)	<0,001
Clase II	1 (2,9)	17 (50,0)	16 (47,1)	
Clase III	18 (28,6)	35 (55,6)	10 (15,9)	
Claude				
Clase I	1 (1,7)	13 (22,0)	45 (76,3)	<0,001
Clase II	1 (2,9)	20 (58,8)	13 (38,2)	
Clase III	20 (31,7)	39 (61,9)	4 (6,3)	

Nota: Los porcentajes se calcularon dentro de cada nivel de complejidad E-CAT. El desempeño diagnóstico se clasificó como incorrecto, parcialmente correcto o correcto considerando conjuntamente la concordancia de los diagnósticos pulpar y periapical con el diagnóstico clínico-radiográfico de referencia. La asociación entre el nivel de complejidad y el desempeño diagnóstico se evaluó mediante la prueba exacta de Fisher-Freeman-Halton.

Discusión

En el diagnóstico pulpar, los modelos de inteligencia artificial mostraron una concordancia moderada con el diagnóstico clínico-radiográfico de referencia, con coeficientes Kappa entre $\kappa=0,431$ y $\kappa=0,513$ ($p<0,001$). Sin embargo, el nivel de acierto cambió notablemente según la patología. La necrosis pulpar presentó una concordancia del 85,1% para ChatGPT

y del 86,2% para Gemini y Claude. En la pulpitis irreversible sintomática, ChatGPT alcanzó el 90,9%, Gemini el 86,4% y Claude el 68,2%. En cambio, la pulpitis irreversible asintomática fue la condición de mayor dificultad, con apenas 17,9%, 25,6% y 7,7%, respectivamente. Esta variación también fue observada por Mendonça de Moura et al. (2024), quienes reportaron exactitudes diagnósticas del

86,4% para Bing y 85,3% para ChatGPT-4.0, frente al 46,5% de ChatGPT-3.5 y 28,6% de Bard¹⁰. Aunque dicho estudio evaluó casos ficticios y consideró conjuntamente diagnósticos pulpaes y perirradiculares, ambos trabajos muestran que la capacidad diagnóstica de la IA puede variar de forma importante según el modelo y el tipo de caso evaluado.

Dentro de la distribución de los diagnósticos pulpaes, la necrosis representó 87 de los 156 casos estudiados (55,8%) y, al mismo tiempo, fue una de las condiciones con mayor concordancia diagnóstica, con porcentajes entre 85,1% y 86,2%. En pacientes atendidos en clínicas universitarias de Endodoncia en Portoviejo, Santos Zambrano et al. también identificaron una presencia importante de necrosis pulpar, aunque sus resultados fueron presentados principalmente según edad y sexo, lo que impide realizar una comparación porcentual global directa con el 55,8% encontrado en esta investigación¹⁴. Por ello, el predominio de la necrosis en la presente muestra debe interpretarse dentro del contexto de una población clínica específica y considerarse al

analizar los porcentajes globales de concordancia.

La diferencia entre patologías sintomáticas y asintomáticas fue especialmente marcada. Mientras la pulpitis irreversible sintomática alcanzó porcentajes de concordancia entre 68,2% y 90,9%, su presentación asintomática se situó únicamente entre 7,7% y 25,6%. Kaskova et al. (2024) señalan que determinar el estado pulpar puede resultar complejo cuando los signos y síntomas no permiten establecer una correspondencia clara con la condición de la pulpa¹⁵. Esto podría explicar, al menos parcialmente, la dificultad de los modelos ante la pulpitis irreversible asintomática, donde existen menos manifestaciones clínicas que orienten el diagnóstico y se requiere integrar con mayor cuidado los diferentes hallazgos del caso.

En el diagnóstico periapical, la concordancia también fue moderada, con valores Kappa entre $\kappa=0,483$ y $\kappa=0,525$ ($p<0,001$). El absceso apical crónico presentó una concordancia del 85,7% en los tres modelos, mientras que en la periodontitis apical sintomática ChatGPT y Claude alcanzaron el 85,2% y Gemini el 74,1%. Nuevamente, el desempeño

disminuyó frente a una condición asintomática: en la periodontitis apical asintomática, ChatGPT y Gemini alcanzaron un 46,8% y Claude un 44,7%. Akyüz et al. (2026), al evaluar modelos de inteligencia artificial mediante radiografías periapicales, también encontraron variaciones en el rendimiento según la condición endodóntica analizada¹⁶. Por su parte, Aloufi (2026) señala que las radiografías periapicales, al representar estructuras tridimensionales en una imagen bidimensional, pueden presentar superposiciones y distorsiones que dificultan la identificación de lesiones pequeñas o tempranas¹⁷. Estas limitaciones podrían influir especialmente cuando existen pocos signos clínicos que complementen la interpretación radiográfica.

La integración de información clínica y radiográfica adquiere especial importancia frente a estas dificultades. Huang et al. (2023) destacan el potencial de los modelos multimodales para procesar conjuntamente información textual e imágenes en el campo odontológico¹⁸. Esto resulta relevante en Endodoncia, donde el diagnóstico no depende de un hallazgo aislado, sino de la relación entre

antecedentes, signos, síntomas, pruebas clínicas y características radiográficas. Sin embargo, la variabilidad encontrada en este estudio muestra que la capacidad de procesar distintos tipos de información no necesariamente se traduce en un desempeño uniforme ante todas las patologías.

En cuanto al tipo de modelo de inteligencia artificial, Gemini presentó descriptivamente la mayor concordancia para el diagnóstico pulpar con 69,2%, seguido de ChatGPT con 67,9% y Claude con 62,2%. En el diagnóstico periapical ocurrió algo diferente: ChatGPT obtuvo el porcentaje más alto con 66,0%, seguido de Claude con 63,5% y Gemini con 62,8%. Por lo tanto, ningún modelo mantuvo el mayor porcentaje de concordancia en ambos tipos de diagnóstico. Madfa et al. (2026), al evaluar 117 casos odontológicos, también encontraron diferencias entre plataformas. En su segunda medición, Claude y Manus alcanzaron una exactitud del 92,3%, mientras que ChatGPT obtuvo un 76,9%; en la primera medición, los porcentajes fueron del 91,5% para Claude, 90,6% para Manus y 74,4% para ChatGPT¹⁹. Esto contrasta con el presente estudio,

donde Claude presentó el menor porcentaje de concordancia pulpar con 62,2%. Esta diferencia podría relacionarse con la metodología empleada, ya que Madfa et al. utilizaron preguntas de opción múltiple de distintas especialidades odontológicas, mientras que en esta investigación se evaluaron casos clínico-radiográficos específicamente endodónticos. Esto muestra que el comportamiento de una plataforma puede cambiar según el tipo de información y la forma en que se presenta el caso.

Al analizar el desempeño diagnóstico de los modelos de inteligencia artificial, ChatGPT presentó un 46,8% de respuestas completamente correctas, Gemini un 45,5% y Claude un 39,7%. Las respuestas parcialmente correctas correspondieron al 40,4%, 41,0% y 46,2%, respectivamente, mientras que las incorrectas oscilaron entre 12,8% y 14,1%. Aunque la prueba de Friedman mostró una diferencia global estadísticamente significativa entre los modelos ($\chi^2=9,624$; $gl=2$; $p=0,008$), las comparaciones por pares con corrección de Bonferroni no encontraron diferencias significativas entre ChatGPT y Gemini (p ajustado=1,000), ChatGPT y Claude (p

ajustado=0,846), ni Gemini y Claude (p ajustado=1,000). Por esta razón, las diferencias porcentuales observadas no permiten afirmar que un modelo haya sido estadísticamente superior a los demás. Este resultado guarda relación con lo reportado por Madfa et al. (2026), ya que a pesar de encontrar porcentajes de exactitud del 92,3% para Claude y Manus frente al 76,9% para ChatGPT, las diferencias entre modelos tampoco alcanzaron significación estadística. Además, reportaron coeficientes Kappa de consistencia de $\kappa=0,714$ para Claude, $\kappa=0,782$ para Manus y $\kappa=0,560$ para ChatGPT¹⁹. Ambos estudios muestran que una diferencia numérica entre plataformas no es suficiente para establecer que una de ellas tenga un rendimiento superior.

El nivel de complejidad del caso clínico fue la variable que mostró una asociación clara con el desempeño diagnóstico. En los casos Clase I, las respuestas correctas alcanzaron el 81,4% para ChatGPT y el 76,3% para Gemini y Claude. Al pasar a Clase II, estos porcentajes disminuyeron al 50,0%, 47,1% y 38,2%, respectivamente. La reducción fue aún mayor en Clase III, donde únicamente el 12,7% de las respuestas de ChatGPT,

el 15,9% de Gemini y el 6,3% de Claude fueron completamente correctas. Esta asociación fue estadísticamente significativa para los tres modelos mediante la prueba exacta de Fisher-Freeman-Halton ($p < 0,001$). Por tanto, a diferencia de las pequeñas variaciones encontradas entre plataformas, el aumento de la complejidad estuvo acompañado de una reducción clara y progresiva del desempeño. Atakır et al. (2025), aunque no estudiaron específicamente la complejidad endodóntica, demostraron cuantitativamente cuánto puede cambiar el rendimiento de un modelo según la información disponible. ChatGPT-4o alcanzó una exactitud de apenas 19,90% al recibir únicamente la imagen y de 20,67% cuando recibió solamente las opciones diagnósticas. Al combinar información clínica con opciones, la exactitud aumentó al 75,45%, y cuando se proporcionaron conjuntamente imagen, información clínica y opciones alcanzó el 80,88%. La concordancia también aumentó desde $\kappa = 0,023$ con opciones aisladas hasta $\kappa = 0,733$ con la combinación completa de información.¹³ Aunque estos resultados no equivalen directamente a las Clases I, II y III utilizadas en este estudio, muestran que la cantidad, calidad y

coherencia de la información disponible pueden modificar de forma importante la respuesta de un modelo.

Este aspecto también debe considerarse al interpretar la calidad radiográfica y la completitud de la información clínica. En la muestra estudiada, el 80,77% de las radiografías fue clasificado en categoría A según FGDP y el 19,23% en categoría N; a su vez, el nivel de información clínica fue alto en el 85,26% de los casos y moderado en el 14,74%. Estas variables no fueron analizadas como factores independientes, debido a que el análisis se centró en la asociación bivariada entre complejidad y desempeño. Además, formaron parte de los criterios considerados en la clasificación de complejidad empleada. Por esta razón, no es posible determinar cuánto del descenso observado en Clase III se debe específicamente a la calidad radiográfica, a la información clínica disponible o a otros factores presentes en los casos más complejos. Siddique et al. (2024) reconocen la importancia de contar con radiografías periapicales de calidad adecuada para favorecer una correcta interpretación diagnóstica²⁰. El análisis independiente de estas variables en futuras investigaciones permitiría determinar con mayor

precisión cuánto aporta cada una al desempeño de los modelos.

Conclusiones

Los resultados del presente estudio demostraron una concordancia diagnóstica significativa entre ChatGPT, Gemini y Claude con los criterios clínico-radiográficos de referencia para la evaluación del estado de salud pulpar y periapical. Aunque hubo variaciones en su desempeño, las comparaciones post hoc no demostraron una superioridad entre los modelos. Además, la concordancia disminuyó conforme aumentó la complejidad clínica, particularmente en los casos de mayor dificultad, lo que evidencia que el desempeño diagnóstico de estos modelos se asocia con las características del caso evaluado. Por tanto, ChatGPT, Gemini y Claude pueden considerarse como herramientas con potencial en la fase diagnóstica, si reciben un apropiado entrenamiento con la evidencia científica disponible y los criterios estándar en el área, pero su uso debe complementarse con la valoración clínica del profesional.

Limitaciones

Aunque se observó una asociación entre la complejidad diagnóstica sobre el desempeño de la IA, no se puede descartar la posible influencia residual de factores como la calidad radiográfica o información clínica disponible, aun cuando fueron previamente considerados para la designar el grado de dificultad diagnóstica.

Recomendaciones

Para futuras líneas de investigación, se recomienda que se evalúen otros factores clínicos y radiográficos que puedan influir en el desempeño de la IA, para determinar su efecto sobre el grado de dificultad del caso clínico. Asimismo, se podría perfeccionar el prompt para un mejor desempeño y aplicar el estudio en diferentes poblaciones con mayores tamaños muestrales, lo cual permita evidenciar los resultados obtenidos en un contexto más amplio.

Referencias

1. American Association of Endodontists.
<https://www.aae.org/specialty/newsletter/endodontic-diagnosis/endodonticdiagnosisfall2013/> [Internet]. 2017 [citado 9 de agosto de 2026]. p. 6. Endodontic diagnosis - American Association of

- Endodontists. Disponible en: <https://www.aae.org/specialty/newsletter/endodontic-diagnosis/endodonticdiagnosisfall2013/>
2. Donnermeyer D, Dammaschke T, Lipski M, Schäfer E. Effectiveness of diagnosing pulpitis: A systematic review. *Int Endod J.* 1 de octubre de 2023;56(S3):296–325. doi:10.1111/IEJ.13762. PubMed PMID: 35536159.
 3. Tibúrcio-Machado CS, Michelon C, Zanatta FB, Gomes MS, Marin JA, Bier CA. The global prevalence of apical periodontitis: a systematic review and meta-analysis. *Int Endod J.* 1 de mayo de 2021;54(5):712–35. doi:10.1111/IEJ.13467. PubMed PMID: 33378579.
 4. Sui H, Lv Y, Xiao M, Zhou L, Qiao F, Zheng J, et al. Relationship between the difference in electric pulp test values and the diagnostic type of pulpitis. *BMC Oral Health.* 1 de diciembre de 2021;21(1). doi:10.1186/S12903-021-01696-9. PubMed PMID: 34246278.
 5. Hilmi A, Patel S, Mirza K, Galicia JC. Efficacy of imaging techniques for the diagnosis of apical periodontitis: A systematic review. *Int Endod J.* 1 de octubre de 2023;56(S3):326–39. doi:10.1111/IEJ.13921. PubMed PMID: 37067066.
 6. Samanta Cuba-Cueto K, Caparó EV, Regina M, Apaza S, Alejandra C, Miranda M. Perfil epidemiológico de patologías pulpares y periapicales en Latinoamérica: Epidemiological profile of pulpal and periapical pathologies in Latin America. *Revista Científica Especialidades Odontológicas UG.* 5 de agosto de 2022;5(2). doi:10.53591/EOUG.V5I2.142 .
 7. Ivanna CO, Esteban TC, Boris OV, Evelyn CO, María José FS. Perfil epidemiológico de las patologías pulpares en pacientes que acuden a centros de atención odontológica en la ciudad de Cuenca-Ecuador. *Odontología Activa Revista Científica.* 10 de septiembre de 2019;4(3):15–20. doi:10.31984/OACTIVA.V4I3 .389.
 8. Azim AA, Merdad K, Peters OA. Diagnosis consensus among endodontic specialists and general practitioners: An international survey and a proposed modification to the current diagnostic terminology. *Int Endod J.* 1 de noviembre de 2022;55(11):1202. doi:10.1111/IEJ.13816. PubMed PMID: 35984730.
 9. Umer F, Batool I, Naved N. Innovation and application of

- Large Language Models (LLMs) in dentistry – a scoping review. *BDJ Open*. 1 de diciembre de 2024;10(1):90. doi:10.1038/s41405-024-00277-6.
10. Mendonça de Moura JD, Fontana CE, Reis da Silva Lima VH, de Souza Alves I, André de Melo Santos P, de Almeida Rodrigues P. Comparative accuracy of artificial intelligence chatbots in pulpal and periradicular diagnosis: A cross-sectional study. *Comput Biol Med*. 1 de diciembre de 2024;183:109332. doi:10.1016/J.COMPBIOMED.2024.109332. PubMed PMID: 39471663.
 11. Özbay Y, Erdoğan D, Dinçer GA. Evaluation of the performance of large language models in clinical decision-making in endodontics. *BMC Oral Health*. 1 de diciembre de 2025;25(1):648. doi:10.1186/S12903-025-06050-X. PubMed PMID: 40296000.
 12. Essam O, Boyle EL, Whitworth JM, Jarad FD. The Endodontic Complexity Assessment Tool (E-CAT): A digital form for assessing root canal treatment case difficulty. *Int Endod J*. 1 de julio de 2021;54(7):1189–99. doi:10.1111/IEJ.13506. PubMed PMID: 33682086.
 13. Atakır K, Işın K, Taş A, Önder H. Diagnostic accuracy and consistency of ChatGPT-4o in radiology: influence of image, clinical data, and answer options on performance. *Diagnostic and Interventional Radiology*. 22 de septiembre de 2025. doi:10.4274/DIR.2025.253460.
 14. Santos Zambrano TB, Parrales Espinoza LL, García Iturralde MG. Prevalencia de patologías pulpares en pacientes atendidos en las clínicas de Endodoncia de la Universidad San Gregorio de Portoviejo. *Revista Científica Especialidades Odontológicas UG*. 2022;5(1):1–6. doi:10.53591/eoug.v5i1.1402.
 15. Kaskova L, Yanko N, Vashchenko I, Khmil O. Problems of diagnostic of the state of pulp and pulpal diagnoses in children. *Wiad Lek*. 2024;77(9):1763–8. doi:10.36740/WLEK/191319. PubMed PMID: 39549004.
 16. Akyüz İE, Kıvrıkcık BE, Aslan T. Artificial Intelligence-assisted Periapical Radiographic Assessment: Lesion Detection, Endodontic Complication Analysis, and Review of Clinical Treatment Recommendations. *J Endod*. 1 de julio de 2026;52(7):1194–205. doi:10.1016/j.joen.2026.04.008. PubMed PMID: 42019676.
 17. Aloufi AS. Detection of Periapical Lesions Using

- Artificial Intelligence: A Narrative Review. *Diagnostics*. 17 de enero de 2026;16(2):301. doi:10.3390/DIAGNOSTICS16020301.
18. Huang H, Zheng O, Wang D, Yin J, Wang Z, Ding S, et al. ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. *International Journal of Oral Science*. 28 de julio de 2023;15(1):29. doi:10.1038/S41368-023-00239-Y. PubMed PMID: 37507396.
 19. Madfa AA, Alshammari AF, Anazi BA, Alenezi YE, Alkurdi KA. Accuracy and reliability of Manus, ChatGPT, and Claude in case-based dental diagnosis. *Frontiers in Oral Health*. 8 de enero de 2026;6:1686090. doi:10.3389/FROH.2025.1686090.
 20. Siddique SN, Anwar MA, Zaman H, Haider I, Ahmad A, Umair M, et al. Quality Assessment of Periapical Radiographs Taken by Dental Assistants Using the Recent Faculty of General Dental Practice (FGDP) Guidelines. *Cureus*. 3 de septiembre de 2024;16(9). doi:10.7759/CUREUS.68508.

ANEXOS

Anexo 1. PROMPT INGRESADO EN LOS MODELOS DE IA

07 PROMPT - EVALUACIÓN DIAGNÓSTICA

INSTRUCCIONES PARA LA IA (Copiar y pegar en ChatGPT, Gemini o Claude)

Actúa como especialista en Endodoncia. Analiza e integra los datos clínicos y radiográficos proporcionados y selecciona un diagnóstico pulpar y un diagnóstico periapical entre las categorías disponibles, según la clasificación de la Asociación Americana de Endodncistas.

La expresión "no consta" no debe interpretarse como ausencia del síntoma, y se utilizará únicamente en los datos del diente testigo, estímulos mecánicos y hallazgos radiográficos; y la expresión "no evaluable" se limita a la coincidencia entre la imagen radiográfica y los hallazgos registrados si no se puede interpretar de manera confiable.

Selecciona únicamente categorías incluidas en la guía diagnóstica de la AAE, y responde exactamente con este formato:

Diagnóstico pulpar: *(Categoría exacta)*
Diagnóstico periapical: *(Categoría exacta)*

DATOS DEL CASO

1. IDENTIFICACIÓN DE PACIENTE/CASO

- **Código paciente-caso:** *(PAC-XXX)*
- **Semestre:** *(B-2025 o A-2026)*
- **Edad:** *(Edad registrada)*
- **Género:** *(M o F)*
- **Diente evaluado:** *(# Pieza dental)*
- **Motivo de consulta/resumen clínico:** *(Información registrada en la historia clínica original)*

2. ESTADO DE SALUD PULPAR: CRITERIOS CLÍNICOS

- **Sintomatología:** *(Incluir únicamente los síntomas registrados)*
- **Examen clínico:** *(Incluir únicamente los hallazgos registrados)*
- **Pruebas de sensibilidad pulpar:** Frio/Calor *(Presente/positivo o Ausente/negativo)*
- **Diente testigo** *(# Pieza dental o No consta)*
 - Frio/Calor Dte. Testigo *(Presente/positivo, Ausente/negativo y/o No consta)*

3. ESTADO PERIAPICAL CLÍNICO-RADIOGRÁFICO

- **Estímulos mecánicos - Masticación, percusión, palpación:** *(Sí/No/No consta)*
- **Hallazgos radiográficos:**
 - Cuando el estado periapical sea "Tejidos apicales normales":
(Sin alteraciones periapicales observables)
 - Cuando el estado periapical tenga hallazgos presentes o expresamente descritos en observaciones en la historia clínica original: *(Con alteraciones periapicales observables)*
 - Cuando el estado radiográfico que respalda el diagnóstico periapical no esté marcado en la historia clínica original: *(No consta)*

4. RADIOGRAFÍA PERIAPICAL

Analiza la radiografía periapical adjunta junto con la información clínica y radiográfica descrita. Si no se puede visualizar correctamente, no inventes hallazgos y emplea como apoyo principal la información textual proporcionada.

CATEGORÍAS DIAGNÓSTICAS - AAE

- **Diagnóstico pulpar:** Pulpa Normal | Pulpitis Reversible | Pulpitis Irreversible Sintomática | Pulpitis Irreversible Asintomática | Necrosis Pulpar
- **Diagnóstico periapical:** Tejidos Apicales Normales | Periodontitis Apical Asintomática | Periodontitis Apical Sintomática | Absceso Apical Agudo | Absceso Apical Crónico



Anexo 2. FICHA DE RECOLECCIÓN DE DATOS

SEÑALADO

CÓDIGO PAC

Semestre B-2

PAC-001-01

X

Eliminar

01

Identificación del paciente/caso

SEÑALADO

CÓDIGO PACIENTE - CASO

SEÑALADO

SELECCIONAR

PAC-

36

N.º - HISTORIA CLÍNICA

ED - número

EDAD

32

GÉNERO

SELECCIONAR

MOTIVO DE CONSULTA / RESUMEN CLÍNICO

Resumen clínico breve

02

Criterios clínicos

Registra signos, síntomas y pruebas clínicas tal como constan en la historia clínica física.

SINTOMATOLOGÍA

☐ Ausencia de dolor
 ☐ Dolor provocado
 ☐ Dolor espontáneo

☐ Dolor localizado
 ☐ Dolor difuso
 ☐ Dolor constante

☐ Dolor intermitente

EXAMEN CLÍNICO

☐ Inflamación intraoral
 ☐ Inflamación extraoral
 ☐ Fístula

☐ Bolea periodontal
 ☐ Caries profunda
 ☐ Restauración defectuosa

Guardado: 09:48:09 a.m.

Guardar caso

SERIE TINE	CÓDIGO PAC		
Seriestro B-2	PAC-001-D1	X	

☐ Infección pericoronaria	☐ Inflamación extraoral	☐ Fístula
☐ Grieta o fisura	☐ Caries profunda	☐ Restauración defectuosa
☐ Fractura coronaria	☐ Cambio de color	☐ Movilidad

ESTÍMULOS MECÁNICOS PERIAPICALES

Aquí se registra la respuesta a masticación/mordida, percusión y palpación apical. No se repite en sintomatología para evitar doble registro.

MASTICACIÓN / MORDIDA	PERCUSIÓN	PALPACIÓN APICAL
No consta	No consta	No consta

PRUEBAS DE SENSIBILIDAD PULPAR

Registra la respuesta como presente/positiva o ausente/negativa, según corresponda.

FRIÓ	CALOR	DIENTE TESTIGO
No registrado	No registrado	Ej. 21,35 o 25
		☐ NO CONSTA

DIENTE TESTIGO - FRÍO	DIENTE TESTIGO - CALOR
No consta	No consta

Hallazgos radiográficos

ESTADO RADIOGRÁFICO PERIAPICAL

No consta	
-----------	--

☐ Radiolucidez apical	☐ Radiolucidez lateral	☐ Radiopacidad apical
☐ Reabsorción interna	☐ Reabsorción externa	☐ Raíz en formación

Guardado - 09:48:09 a.m.

Guardar caso

SERIE: PAC-001-01

☐ Radiolucidez lateral	☐ Radiopacidad apical
☐ Reabsorción externa	☐ Raíz en formación
☐ Tratamiento endodóntico previo	☐ Cámara calcificada
☐ Otros	

☐ Presencia de poste

☐ Conductos calcificados

DETALLE DE OTROS HALLAZGOS

Registra cualquier novedad, anomalía, limitación de la imagen u observación radiográfica adicional que no esté contemplada en los ítems anteriores.

04 Diagnóstico clínico-radiográfico de referencia

DIAGNÓSTICO PULPAR DE REFERENCIA	DIAGNÓSTICO PERIAPICAL DE REFERENCIA
Seleccionar	Seleccionar

05 Calidad de la imagen radiográfica – variable interviniente

Clasificación de dos niveles FGDP (Faculty of General Dental Practice): diagnósticamente aceptable (A) o diagnósticamente no aceptable (N). Registra también los errores técnicos observados.

CLASIFICACIÓN MANUAL FACULTY OF GENERAL DENTAL PRACTICE (FGDP)

CALIDAD DIAGNÓSTICA RADIOGRÁFICA

ERRORES TÉCNICOS OBSERVADOS

☐ Contraste/intidez incorrectos	☐ Posicionamiento incorrecto del sensor/película
☐ Angulación vertical incorrecta	☐ Angulación horizontal incorrecta

Guardado - 09:48:09 a.m.

Guardar caso

SERIE Y TEL.

CÓDIGO PAC

Semestre B-2

PAC-001-D1

almacenamiento y los

Seleccionar

▼

ERRORES TÉCNICOS OBSERVADOS

☐ Contraste/nitidez incorrectos

☐ Posicionamiento incorrecto del sensor/película

☐ Angulación vertical incorrecta

☐ Angulación horizontal incorrecta

☐ Procesamiento/calidad de radiografía periapical incorrecta

Calificación FGDP (Faculty of General Dental Practice): **PENDIENTE** 0 errores

Selecciona la clasificación FGDP (Faculty of General Dental Practice) y marca los errores técnicos observados.

Aquí aparecerá la radiografía.

Visor vertical 4:5 - máximo 1080 x 1350 px.

BRILLO 100%

CONTRASTE 100%

INVERTIR ESCALA NO

Restablecer visor

Eliminar radiografía

Guardado · 09:48:09 a. m.

Guardar caso

SERIE Y TEL.

CÓDIGO PAC

Semestre B-2

PAC-001-D1

miniar radiografía

66

Completitud de la información clínica proporcionada al modelo de IA – variable interviniente

Completitud del prompt: **BAJA** 1 ptos.

Alta (3): información clínica + hallazgos radiográficos + categorías diagnósticas. Moderada (2): dos componentes. Baja (1): información insuficiente o incompleta.

☐ Información clínica presente
Solo sección 02, síntomas, examen clínico y/o pruebas.

☐ Hallazgos radiográficos presentes
Solo sección 03, hallazgos mandibla o descripción.

☒ Categorías diagnósticas AAE disponibles
Incluidas en el prompt estandarizado.

Información clínica: **NO REGISTRADO**

No se registran datos en la sección 02.

Hallazgos radiográficos: **NO REGISTRADO**

No consta información sobre el estado radiográfico periapical.

Categorías diagnósticas AAE disponibles: **PRESENTE**

El prompt incluye categorías diagnósticas pulpares y periapicales.

67

Prompt estandarizado para modelos de IA

Uso metodológico: este prompt incluye datos clínicos, hallazgos radiográficos y las categorías diagnósticas AAE disponibles para el análisis. No incorpora diagnóstico de referencia, calidad de imagen, completitud, completitud ni resultados de modelos, para evitar sesgo en la respuesta de la IA. La radiografía se maneja como archivo JPG junto al prompt mediante la opción "Copiar prompt + radiografía".

Actualizar prompt

Copiar prompt + radiografía

PROMPT ESTANDARIZADO PARA CHATGPT, GEMINI Y CLAUDE

Actúa como especialista en Endodoncia. Analiza e integra los datos clínicos y radiográficos proporcionados y selecciona un diagnóstico pulpar y un diagnóstico periapical entre las categorías disponibles, según la clasificación de la Asociación Americana de Endododocistas.

La expresión "no consta" no debe interpretarse como ausencia del síntoma, y se utilizará únicamente en los datos del diente testigo, estímulos mecánicos, hallazgos radiográficos, y la expresión "no evaluable" se limita a la coincidencia entre la imagen radiográfica y los hallazgos registrados si no se puede

Guardado · 09:48:09 a. m.

Guardar caso

SERIE Y TEL.

CÓDIGO PAC

Semestre B-2

PAC-001-D1

información textual proporcionada.

68

Desempeño diagnóstico de los modelos de IA

Escala Likert de 3 puntos: Correcta (3 pts.) cuando coincide con el diagnóstico pulpar y periapical de referencia; Parcialmente correcta (2 pts.) cuando coincide de forma incompleta, acierta solo uno de los diagnósticos u omite parte del diagnóstico; Incorrecta (1 pts.) cuando no coincide o corresponde a una categoría diagnóstica equivocada.

A. DIAGNÓSTICOS EMITIDOS POR CADA MODELO

ChatGPT

DIAGNÓSTICO PULPAR EMITIDO

Seleccionar

DIAGNÓSTICO PERIAPICAL EMITIDO

Seleccionar

Escala Likert vs referencia:

PENDIENTE. Registra diagnóstico de referencia y diagnóstico emitido por el modelo.

Gemini

DIAGNÓSTICO PULPAR EMITIDO

Seleccionar

DIAGNÓSTICO PERIAPICAL EMITIDO

Seleccionar

Escala Likert vs referencia:

PENDIENTE. Registra diagnóstico de referencia y diagnóstico emitido por el modelo.

Claude

DIAGNÓSTICO PULPAR EMITIDO

Seleccionar

DIAGNÓSTICO PERIAPICAL EMITIDO

Seleccionar

Escala Likert vs referencia:

PENDIENTE. Registra diagnóstico de referencia y diagnóstico emitido por el modelo.

B. CONCORDANCIA ENTRE MODELOS DE IA

Concordancia entre modelos: **POR CALCULAR**

Esta clasificación compara únicamente los diagnósticos emitidos por ChatGPT, Gemini y Claude entre sí.

Concordancia total: los tres modelos coinciden en diagnóstico pulpar y periapical.

Concordancia parcial: coinciden solo en uno de los dos diagnósticos.

Sin concordancia: no coinciden de forma completa.

Guardado · 09:48:09 a. m.

Guardar caso

24

SEMESTRE

CÓDIGO PAC

Semestre B-2

PAC-001-01

X

Resultados emitidos por ChatGPT, Gemini y Claude entre sí.

Concordancia total: los tres modelos coincidieron en diagnóstico pulpar y periapical.
Concordancia parcial: coinciden solo en uno de los dos diagnósticos.
Sin concordancia: no coinciden de forma completa.

09 Nivel de complejidad del caso clínico – Endodontic Complexity Assessment Tool (E-CAT)

Clase I
Información clínica y radiográfica suficiente, clara y concordante. La imagen confirma los hallazgos y no existen factores importantes de dificultad.

Clase II
Existe una limitación concreta como calidad radiográfica u omisión de un signo, pero el caso todavía permite llegar a un diagnóstico razonablemente claro.

Clase III
Existen contradicciones importantes, varios factores de dificultad, información relevante insuficiente o un mayor riesgo de error diagnóstico.

¿LA IMAGEN CONFIRMA LOS HALLAZGOS RADIOGRÁFICOS MARCADOS?

☐ SÍ

☐ NO

☐ NO EVALUABLE

OBSERVACIÓN BREVE

Ej. No evaluable por angulación, encuadre o calidad radiográfica

CLASE E-CAT AUTOMÁTICA

Clase III / Caso altamente complicado

Resultado E-CAT: CLASE III / CASO ALTAMENTE COMPLICADO

Criterio aplicado:
Información clínico-radiográfica discordante o con alto riesgo de error

Motivo:

- Faltan ambos diagnósticos de referencia y la información es insuficiente para valorar razonablemente el caso.
- No se registró el hallazgo radiográfico, lo que limita parcialmente la valoración del caso, sin embargo, la información disponible permite una interpretación diagnóstica razonablemente clara.

Guardado - 09:48:09 p. m.

Guardar caso

Ficha digital exportada en PDF

FICHA DIGITAL - ENDODONCIA

Tema: Correlación diagnóstica entre modelos de inteligencia artificial y criterios clínico-radiográficos de patologías pulpares. Clínica de Endodoncia: Semestre B-2025 y A-2026.

Caso 1 de 195

Código: PAC-001

Semestre: Semestre A-2026

01. Identificación

Edad: 44
Género: Femenino
Diente evaluado: 12
Motivo de consulta: TENGO QUE REALIZARME UNA ENDODONCIA

02. Criterios clínicos

Sintomatología

- Ausencia de dolor

Examen clínico

- Restauración defectuosa

Estímulos mecánicosMasticación/mordida: No constaPercusión: No constaPalpación apical: No constaSensibilidad pulparFrí: Ausente / negativoCalor: Ausente / negativoDiente testigo: 11Diente testigo — frío: Presente / positivoDiente testigo — calor: Presente / positivo

03. Hallazgos radiográficos

Estado radiográfico periapical: Sin alteraciones periapicales observables

- Otros hallazgos: SE OBSERVA UNA RESTAURACION INVASIVA CON COMPROMISO PULPAR

04. Diagnóstico de referencia

Pulpar: Necrosis pulpar
Periapical: Tejidos apicales normales

05. Calidad radiográfica FGDP

Calidad: Diagnósticamente aceptable (A)
Sin errores técnicos seleccionados

06. Completitud del prompt

Nivel: Alta

08. Modelos de IA

ChatGPT: Pulpar: Necrosis pulpar · Periapical: Tejidos apicales normales · Desempeño: Correcta (3)
Gemini: Pulpar: Necrosis pulpar · Periapical: Tejidos apicales normales · Desempeño: Correcta (3)
Claude: Pulpar: Necrosis pulpar · Periapical: Tejidos apicales normales · Desempeño: Correcta (3)
Concordancia entre modelos: Concordancia total entre modelos

09. E-CAT

Imagen confirma hallazgos RX: Sí
Observación imagen-hallazgos: —
Clasificación: Clase I / Caso no complicado

- Caso claro, suficiente y concordante
- Sin una limitación importante

Documento generado desde la ficha digital.

25

DECLARACIÓN Y AUTORIZACIÓN

Yo, **Herrero Quiroz, Aitana**, con C.C: # **0930372925** autora del trabajo de titulación: **Correlación diagnóstica entre modelos de inteligencia artificial y criterios clínicos radiográficos en patologías pulpares**, previo a la obtención del título de **Odontóloga** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, 10 de septiembre de 2026

AUTORA

f. 

Nombre: **Herrero Quiroz, Aitana**

C.C: **0930372925**



REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TEMA Y SUBTEMA:	Correlación diagnóstica entre modelos de inteligencia artificial y criterios clínicos radiográficos en patologías pulpares.		
AUTOR(ES)	Herrero Quiroz, Aitana		
REVISOR(ES)/TUTOR(ES)	García Guerrero, Yara Anna Paula		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Ciencias de la Salud		
CARRERA:	Odontología		
TÍTULO OBTENIDO:	Odontóloga		
FECHA DE PUBLICACIÓN:	10 de septiembre de 2026	No. DE PÁGINAS:	24
ÁREAS TEMÁTICAS:	Inteligencia artificial, diagnóstico endodóntico, patologías pulpares, Endodoncia, Diagnóstico Asistido por Computador.		
PALABRAS CLAVES/KEYWORDS:	Inteligencia artificial, endodoncia, diagnóstico pulpar, diagnóstico periapical, complejidad diagnóstica, desempeño diagnóstico		
RESUMEN/ABSTRACT:	Introducción: La inteligencia artificial ha adquirido relevancia en el área de la salud por su gran capacidad de procesamiento de datos clínicos en función de optimizar la atención. En endodoncia, los modelos de lenguaje de gran escala (LLMs) podrían ser una herramienta de apoyo para la identificación de patologías pulpares, cuyo rendimiento está sujeto a factores como la complejidad del caso clínico. Objetivo: Determinar la correlación diagnóstica entre modelos de inteligencia artificial y criterios clínico-radiográficos en el diagnóstico de patologías pulpares y periapicales. Materiales y métodos: Se realizó un estudio con 156 casos clínicos. El análisis se realizó en SPSS (IBM, versión 29), mediante coeficiente Kappa de Cohen, prueba de Friedman con ajuste de Bonferroni, y la prueba exacta de Fisher-Freeman-Halton. Resultados: Hubo concordancia moderada y estadísticamente significativa. No se observaron diferencias significativas entre modelos, aunque la concordancia disminuyó conforme aumentó la complejidad clínica, particularmente en los casos Clase III. Conclusiones: ChatGPT, Gemini y Claude mostraron utilidad potencial como herramientas complementarias de apoyo diagnóstico, aunque su desempeño depende de la complejidad del caso y no sustituye la valoración clínica profesional.		
ADJUNTO PDF:	☒ SI	☐ NO	
CONTACTO CON AUTOR/ES:	Teléfono: +593979429423	E-mail: aitana.herrero@cu.ucsg.edu.ec	
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE)::	Nombre: Estefanía del Rocío Ocampo Poma		
	Teléfono: +593-0996757081		
	E-mail: estefania.ocampo@cu.ucsg.edu.ec		
SECCIÓN PARA USO DE BIBLIOTECA			
Nº. DE REGISTRO (en base a datos):			
Nº. DE CLASIFICACIÓN:			
DIRECCIÓN URL (tesis en la web):			