Rev Colomb Cir. 2025;40:1080-7
https://doi.org/10.30944/20117582.3047

ARTÍCULO ORIGINAL

Concordancia de Inteligencia Artificial (ChatGPT, Bing) como predictor de apendicitis versus las escalas de Alvarado y AIR en pacientes adultos del Hospital Militar Central de Bogotá, Colombia

Artificial intelligence concordance (ChatGPT, Bing) as a predictor of appendicitis vs the Alvarado and AIR scales in adult patients at the Hospital Militar Central, Bogotá, Colombia

Álvaro Javier Bayona-Daza, MD[1], Erika Tatiana Mendizábal-Varona, MD[2], Jaime Andrés Marín, MD[2]

1 Programa de especialización en Cirugía General, Universidad Militar Nueva Granada, Bogotá, D.C., Colombia.
2 Servicio de Cirugía General Hospital Militar Central, Bogotá, D.C., Colombia.

Trabajo ganador del Tercer puesto en el Concurso Nacional de Investigación en Cirugía "José Félix Patiño Restrepo", Asociación Colombiana de Cirugía, 51° Congreso Semana Quirúrgica Nacional, agosto de 2025.

Resumen

Introducción. La apendicitis aguda es una patología quirúrgica que representa un desafío en el diagnóstico ya que sus síntomas pueden superponerse con varias afecciones abdominales no quirúrgicas. Las escalas de predicción clínica, como la escala de Alvarado y el AIR *score*, han sido herramientas utilizadas para evaluar la probabilidad de apendicitis en pacientes con dolor abdominal agudo.

Métodos. Se realizó un estudio analítico para determinar la concordancia entre la inteligencia artificial (ChatGPT 4.0 y Bing) y las escalas de predicción diagnóstica ya validadas en sospecha de apendicitis aguda. Se incluyeron los pacientes con dolor abdominal agudo con sospecha de apendicitis, desde octubre de 2023 hasta marzo de 2024.

Resultados. De los pacientes estudiados, mediante ChatGPT se corrieron las variables, y 129 (75 %) presentaron diagnóstico definitivo de apendicitis, que fue confirmado por patología. Mostraron significancia estadística las variables de dolor abdominal (75 %), migración del dolor (52,3 %), dolor a la palpación abdominal (72 %), signo de rebote positivo (43,5 %), leucocitosis, neutrofilia y PCR elevada. Respecto a la concordancia con el diagnóstico generado por la inteligencia artificial, el índice Kappa fue 0,005, lo cual representa una pobre concordancia. El desarrollador Bing no logró realizar el análisis de los datos recolectados.

Fecha de recibido: 20/08/2025 - Fecha de aceptación: 08/08/2025 - Publicación en línea: 14/10/2025
Correspondencia: Álvaro Javier Bayona-Daza, Carrera 16A # 48-40 Apt 102, Teléfono: +57 3186495807.
Dirección electrónica: javier.bayona3196@gmail.com
Citar como: Bayona-Daza AJ, Mendizábal-Varona ET, Marín JA. Concordancia de Inteligencia Artificial (ChatGPT, Bing) como predictor de apendicitis versus las escalas de Alvarado y AIR en pacientes adultos del Hospital Militar Central de Bogotá, Colombia. Rev Colomb Cir. 2025;40:1080-7. https://doi.org/10.30944/20117582.3047

Este es un artículo de acceso abierto bajo una Licencia Creative Commons - BY-NC-ND https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es

Conclusiones. En este estudio la Inteligencia Artificial demuestra muy pobre concordancia con las variables clínicas para diagnóstico de apendicitis aguda. Es probable que esta predicción mejore con el apoyo imagenológico, como lo demuestran otros estudios.

Palabras clave: apendicitis; dolor abdominal; inteligencia artificial; diagnóstico por computador; sensibilidad y especificidad; técnicas de apoyo para la decisión.

Abstract

Introduction. Acute appendicitis is a surgical condition that represents a diagnostic challenge as its symptoms can overlap with several nonsurgical abdominal conditions. Clinical prediction scales, such as the Alvarado scale and the AIR score, have been used to assess the likelihood of appendicitis in patients with acute abdominal pain.

Methods. An analytical study was conducted to determine the concordance between artificial intelligence (ChatGPT 4.0 and Bing) and previously validated diagnostic prediction scales for suspected acute appendicitis. Patients with acute abdominal pain and suspected appendicitis were included from October 2023 to March 2024.

Results. Of the patients studied, the variables were run using ChatGPT, and 129 (75%) presented a definitive diagnosis of appendicitis, which was confirmed by pathology. Statistical significance was shown in the variables of abdominal pain (75%), pain migration (52.3%), abdominal tenderness (72%), positive rebound sign (43.5%), leukocytosis, neutrophilia, and elevated CRP. Regarding the concordance with the diagnosis generated by artificial intelligence, the Kappa index was 0.005, which represents poor concordance. The developer Bing was unable to perform the analysis of the collected data.

Conclusions. In this study, AI demonstrated very poor concordance with clinical variables for diagnosing acute appendicitis. This prediction is likely to improve with imaging support, as demonstrated in other studies.

Keywords: appendicitis; abdominal pain; artificial intelligence; computer-assisted diagnosis; sensitivity and specificity; decision-support techniques.

Introducción

La apendicitis aguda es una afección médica que requiere una evaluación precisa y oportuna, debido a su potencial para causar complicaciones graves si no se trata adecuadamente [1]. Uno de los desafíos cruciales en el diagnóstico de la apendicitis aguda es diferenciarla de otras causas de dolor abdominal, ya que sus síntomas pueden confundirse con varias afecciones abdominales no quirúrgicas [2]. En este contexto, las escalas de predicción clínica, como la escala de Alvarado y el AIR *score* [1,3], han sido herramientas tradicionales utilizadas por los médicos para evaluar la probabilidad de apendicitis en pacientes con dolor abdominal agudo [4,5].

El desarrollo de la inteligencia artificial (IA) ha planteado la posibilidad de mejorar la precisión diagnóstica de múltiples patologías. La IA permite el análisis de grandes conjuntos de datos clínicos y radiológicos, lo que podría traducirse en una identificación más precisa de patologías, como la apendicitis aguda, en comparación con las escalas de predicción tradicionales [6,7].

La inteligencia artificial se empezó a desarrollar en el año 1956, basada en los fundamentos teóricos sentados por Alan Turing, Claude E. Shannon y Norbert Wiener. Así evolucionó una ciencia capaz de generar una variedad de técnicas y métodos que permiten realizar tareas que normalmente ejecuta la mente humana, basada en un subconjunto constituido por el aprendizaje profundo, el cual utiliza redes neuronales inspiradas en la estructura del cerebro humano para adquirir grandes cantidades de datos.

Este es un aprendizaje automático que se centra en el desarrollo de algoritmos y modelos estadísticos con técnicas supervisadas, no supervisadas y semi supervisadas, según sea el

algoritmo, generando modelos predictivos y la ciencia de datos en donde, con técnicas estadísticas y computacionales, obtiene conocimiento y perspectivas a partir de los datos. La IA se ha aplicado en diversos ámbitos a lo largo de la historia, más recientemente en la vida diaria, pasando por tres generaciones: inicialmente la inteligencia artificial estrecha, luego la inteligencia artificial general y más recientemente la superinteligencia artificial [7,8].

La aplicación de la IA en el ámbito de la medicina ha aportado al análisis de extensas bases de datos, encontrando cabida en todas las ramas de la salud. Los modelos de predicción para el diagnóstico de patologías médico-quirúrgicas son una herramienta para el clínico, en donde un algoritmo puede predecir la presencia de cuadros frecuentes, sin reemplazar la evaluación clínica. Aunque se han desarrollado revisiones en apendicitis aguda en donde se aplicó la inteligencia artificial en población pediátrica, para población adulta son escasos los estudios [9,10].

La concordancia entre la IA y las escalas tradicionales se convierte en un aspecto crítico de la evaluación de la efectividad y utilidad de la IA en la toma de decisiones clínicas. Es necesario comprender si los modelos de IA pueden complementar o incluso superar a las escalas de predicción tradicionales para mejorar la precisión diagnóstica, y por lo tanto, optimizar la atención de los pacientes. A medida que se ha avanzado en el desarrollo de modelos de IA para el diagnóstico de apendicitis aguda, surge una pregunta fundamental que pretendemos responder ¿en qué medida la IA puede concordar con las escalas de predicción tradicionales en el diagnóstico de apendicitis aguda en pacientes adultos con dolor abdominal?

Métodos

Diseño del estudio y selección de los pacientes

Se realizó un estudio de prueba diagnóstica para determinar la concordancia entre la IA y las escalas de predicción diagnóstica clásicas en abdomen agudo, en pacientes mayores de 18 años con cuadro clínico de dolor abdominal agudo con sospecha de apendicitis aguda, atendidos en la sala de urgencias del Hospital Militar Central de Bogotá, D.C., Colombia, desde octubre de 2023 hasta marzo de 2024. Los criterios de exclusión fueron antecedente de apendicectomía, estado de embarazo y pacientes sin datos completos para el cálculo de las escalas predictoras.

Análisis estadístico

Se creó una hoja de cálculo en Microsoft Excel® a partir del programa RedCap para el procesamiento de las variables. A todos los pacientes se les aplicaron las escalas de predicción diagnóstica de Alvarado y AIR score, además de someterlos a análisis con inteligencia artificial mediante ChatGPT 4.0 y Bing.

En la estadística descriptiva se obtuvieron medidas de tendencia central y dispersión para las variables cuantitativas según su distribución. En las variables categóricas, con frecuencias absolutas y relativas, se calculó el coeficiente de concordancia Kappa entre las escalas de predicción. Se obtuvo el mismo estadístico para determinar la fuerza de la concordancia de los resultados generados con cada uno de los desarrolladores ChatGPT y Bing. El análisis estadístico se realizó con el programa R Studio (R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/).

Resultados

Durante el periodo de estudio, basado en el registro de historias clínicas, utilizando la Clasificación Internacional de Enfermedades en su 10ª edición (CIE-10) se obtuvo una muestra de 230 pacientes, de los cuales 172 pacientes cumplieron los criterios de inclusión y se ingresaron para el análisis final del presente estudio. El 71 % (n=115) correspondieron al sexo masculino y la media de edad fue de 35,5 ± 17,3 años (rango intercuartílico [RIQ] 33,72 - 38,55) (Figura 1).

De los pacientes estudiados, 129 (75 %) presentaron diagnóstico definitivo de apendicitis confirmado por patología. Entre los hallazgos

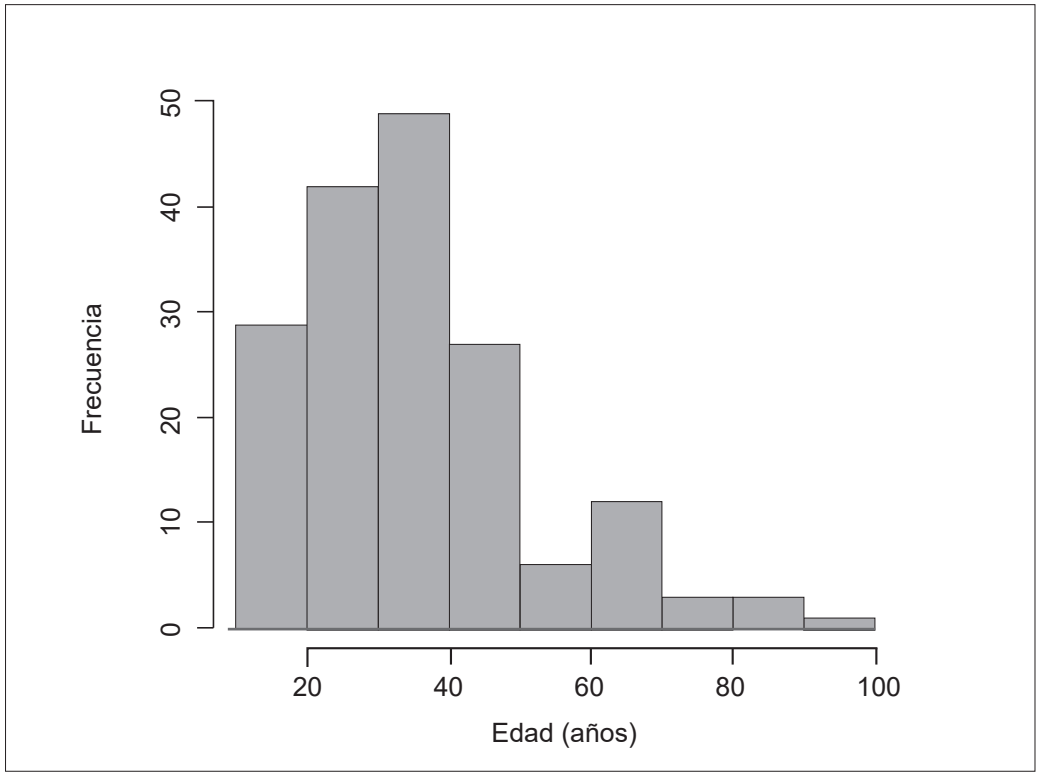

Figura 1. Distribución por grupos etarios de los pacientes con sospecha de apendicitis aguda incluidos en el estudio (n=172).

Fuente: Elaboración propia.

clínicos encontrados se destacaron el dolor abdominal (n=129, 75 %), la migración del dolor (n=90, 52,3 %), el dolor a la palpación abdominal 124 (72 %) y el signo de rebote positivo con p = 0.002 (Tabla 1). Los paraclínicos incluidos mostraron significancia estadística.

La herramienta ChatGPT generó un diagnóstico inicial, en donde en 17 (9,8 %) pacientes consideró un diagnóstico presuntivo de apendicitis y presentó otras opciones más probables como gastritis o úlcera y "cuadro no concluyente". Posterior a la impresión diagnóstica inicial de la inteligencia artificial, se midieron las escalas de apendicitis de Alvarado y AIR *score* (Figura 2), cuyos resultados en los casos de probabilidad intermedia y alta representaron significancia estadística para el diagnóstico de apendicitis con un valor de p = 0,000 (Tabla 2).

La concordancia entre el diagnóstico de apendicitis y la predicción generada por la inteligencia artificial fue la misma (índice Kappa de 0,005; $IC_{95\%}$ 0,060-0,048) al presentar las variables descritas y posteriormente al pedirle que tuviera en cuenta las escalas predictoras clásicas de Alvarado y AIR *score*, con una sensibilidad de 10,1 %, una especificidad de 90,7 %, un valor predictivo positivo (VPP) de 76,5 % y un valor predictivo negativo (VPN) de 25,2 %.

Discusión

El dolor abdominal agudo representa entre el 7 y el 10 % de todos los ingresos al departamento de emergencias, entre los cuales la prevalencia de apendicitis aguda oscila entre el 26,5 % y el 50 % [11,12]. En nuestro estudio encontramos que la frecuencia de apendicitis aguda en los pacientes que consultaron a nuestra institución y tuvieron un diagnóstico presuntivo de apendicitis aguda fue del 75 %. La relación entre hombres y mujeres reportada en las guías de Jerusalem de 2016

Tabla 1. Hallazgos clínicos y paraclínicos en los pacientes incluidos en el estudio (n=172)

Variable	Con apendicitis (n = 129)	Sin apendicitis (n = 42)	Total (n = 172)	p
Dolor abdominal	129 (100 %)	35 (83,3 %)	164 (95,3 %)	0
Dolor a la palpación	124 (96,1 %)	29 (69 %)	153 (88,9 %)	0,028
Rebote				0,002
Leve	10 (7,7 %)	2 (4,7 %)	12 (6,9 %)	
Severo	58 (44,9 %)	5 (11,9 %)	63 (36,6 %)	
Migración del dolor	90 (69,7 %)	17 (40,4 %)	107 (62,1 %)	0,024
Fiebre	25 (19,3 %)	6 (14,2 %)	31 (17,9 %)	0,934
Náuseas y/o vomito	98 (75,9 %)	26 (61,9 %)	124 (72 %)	1
Anorexia	64 (49,6)	14 (33,3)	78 (45,3)	0,391
Leucocitos (células por microlitro)				0,001
10.000 - 15.000	37 (28,6 %)	9 (21,4 %)	46 (26,7 %)	
> 15.000	58 (44,9 %)	6 (14,2 %)	64 (54,6 %)	
Neutrófilos				0,003
< 70 %	29 (22,4 %)	18 (12,8 %)	47 (27,3 %)	
70-84 %	68 (52,7 %)	12 (28,5 %)	80 (46,5 %)	
> 85 %	32 (24,8 %)	5 (11,9 %)	37 (21,5 %)	
Proteína C reactiva (mg/dL)				0,014
<10	32 (24,8 %)	15 (35,7 %)	47 (27,3 %)	
11-50	40 (31 %)	11 (26,1 %)	51 (29,6 %)	
>50	57 (44,1 %)	6 (14,2 %)	63 (36,6 %)	

Fuente: Elaboración propia.

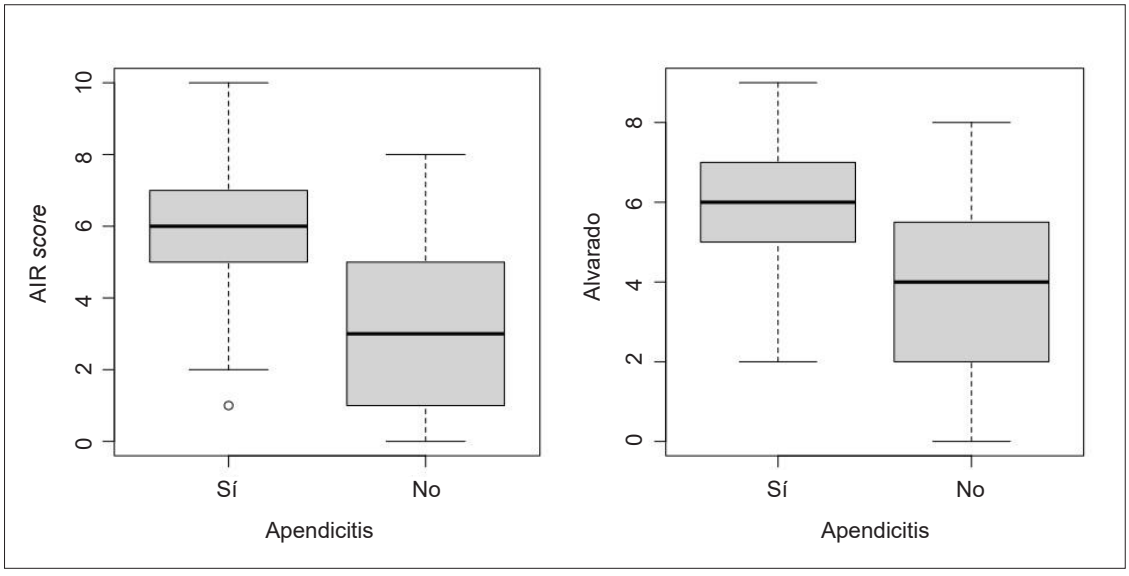

Figura 2. Probabilidad de apendicitis de acuerdo con el AIR *score* y la escala de Alvarado.
Fuente: Elaboración propia.

Tabla 2. Comparación de determinación del diagnóstico de apendicitis aguda con inteligencia artificial (IA) versus las escalas de riesgo.

Variable	Con apendicitis (n = 129)	Sin apendicitis (n = 42)	Total (n = 172)	p
Diagnóstico inicial de ChatGPT	No concluyente: 56 (32,5 %)	No concluyente: 25 (14,5 %)	No concluyente: 81 (47,0 %)	0,287
	Apendicitis: 13 (7,5 %)	Apendicitis: 4 (2,3 %)	Apendicitis: 17 (9,8 %)	
	Gastritis o úlcera: 56 (32,5 %)	Gastritis o úlcera: 14 (8,1 %)	Gastritis o úlcera: 70 (40,7 %)	
	Infección bacteriana: 4 (2,3 %)	Infección bacteriana: 0 (0 %)	Infección bacteriana: 4 (2,3 %)	
Alvarado	Poco probable (baja): 24 (13,9 %)	Poco probable (baja): 24 (13,9 %)	Poco probable (baja): 48 (27,9 %)	0,000
	Posible (intermedio): 66 (38,3 %)	Posible (intermedio): 12 (6,9 %)	Posible (intermedio): 78 (45,3 %)	
	Probable (alta): 39 (22,6 %)	Probable (alta): 7 (4,0 %)	Probable (alta): 46 (26,7 %)	
AIR *Score*	Poco probable (baja): 28 (16,2%)	Poco probable (baja): 28 (16.2 %)	Poco probable (baja): 56 (32,5 %)	0,000
	Posible (intermedio): 77 (44,7 %)	Posible (intermedio): 13 (7,5 %)	Posible (intermedio): 90 (52,3 %)	
	Probable (alta): 24 (13,9 %)	Probable (alta): 2 (1,1 %)	Probable (alta): 26 (15,1 %)	
ChatGPT con escalas de riesgo	No concluyente: 56 (32,5 %)	No concluyente: 25 (14,5 %)	No concluyente: 81 (47,0 %)	0,2
	Apendicitis: 6 (3,4 %)	Apendicitis: 4 (2,3 %)	Apendicitis: 10 (5,8 %)	
	Apendicitis posible: 4 (2,3 %)	Apendicitis posible: 0 (0 %)	Apendicitis posible: 4 (2,3 %)	

Fuente: Elaboración propia.

fue de 1,4:1 [13], mientras en nuestra población del Hospital Militar Central se identificó una relación de 2,4:1. En la literatura se reporta un pico de presentación entre los 10 y 30 años [12] y en este estudio el rango intercuartílico estuvo entre 33 y 38 años.

En la revisión sistemática de Lansdown AJ, et al. [14], reportaron que el dolor abdominal tuvo una sensibilidad de 90-95 % y una especificidad de 50-60 %; la migración del dolor, de 50-75 % y 70-90 %; el signo de rebote, 30-50 % y 80-90 %; y el dolor a la palpación de 70-90 % y 60-80 %, respectivamente.

Respecto a los resultados de laboratorio, en el metaanálisis de Andersson REB [15] se estimó que un valor elevado de proteína C reactiva presenta un área bajo la curva de 0,75 ($IC_{95\%}$ 0,71-0,78) y la leucocitosis un área bajo la curva de 0,72 ($IC_{95\%}$ 0,68-0,76). En concordancia con el análisis de nuestros datos, se identificó que la elevación de los leucocitos por encima de 10.000 células por microlitro, el porcentaje de neutrófilos mayor al 70 % y la PCR elevada mayor de 10 mg/dL tuvieron una diferencia estadísticamente significativa, con valores de $p < 0,05$.

Las variables anteriores corresponden a los ítems evaluados en las escalas de predicción de apendicitis aguda, donde el AIR *score* tiene una sensibilidad de 92 % y una especificidad de 63 % y la escala de Alvarado una sensibilidad de 82-99 % y una especificidad de 43-81 %, de acuerdo con lo reportado en la literatura, con una capacidad mayor del AIR *score* para discriminar el diagnóstico según las curvas ROC (p=0,0007) [15]. Esto concuerda con nuestro análisis, en donde ambas escalas tuvieron un mejor desempeño, con una diferencia estadísticamente significativa ($p<0,005$).

La inteligencia artificial como herramienta para el diagnóstico de patologías quirúrgicas está en auge en la medicina moderna, sin embargo, son limitados lo estudios que muestran beneficios al someter la IA a variables exclusivamente clínicas [16]. En este estudio se determinó la concordancia con las variables expuestas, que se correlacionaron con las que hacen parte de los criterios de las escalas de predicción clásica, y se encontró mediante el índice Kappa una pobre concordancia entre la información suministrada a la IA y el diagnóstico de apendicitis aguda. Así mismo, no se notó un cambio en la concordancia al indicarle a la IA que usara los valores de las escalas de predicción clásicas, alcanzando una sensibilidad muy baja (10,1 %), es decir, que no permite

confirmar la patología, y una especificidad elevada (90,7 %), que es mejor para identificar la ausencia de apendicitis aguda.

En este contexto, la inteligencia artificial ha demostrado en otros estudios su aporte en la interpretación de imágenes para el diagnóstico de patologías quirúrgicas, como la apendicitis aguda, sin embargo, para nuestro estudio, en donde se limitó a paramentos clínicos y paraclínicos, su aporte fue limitado para llegar al diagnóstico.

Conclusiones

El dolor abdominal es uno de los motivos más frecuentes de ingreso a urgencias y se relaciona con patología apendicular aguda en un alto porcentaje, por lo cual, se requiere de suspicacia clínica y herramientas que permitan realizar diagnósticos oportunos para guiar el manejo temprano. Es por ello que, el advenimiento de la inteligencia artificial aplicada a la medicina moderna se muestra como una ayuda adicional para esclarecer los diagnósticos; sin embargo, en este estudio donde se exploraron las variables clínicas y paraclínicas, la inteligencia artificial (específicamente ChatGPT) presentó una muy pobre concordancia con el diagnóstico, en comparación con estudios similares que identificaron sus beneficios y bondades cuando se expuso a ayudas imagenológicas. Con lo cual concluimos que no se recomienda el uso de estas herramientas de inteligencia artificial para hacer diagnóstico de apendicitis aguda.

Agradecimientos

Al Comité de Investigación del Hospital Militar Central de Bogotá, por ser nuestro apoyo durante la recolección de datos, en cabeza de Valentina Navas y Nathalia Aguirre; así como a Annuar de la Barrera, Coordinador del Laboratorio de Innovación INNLAB.

Cumplimiento normas éticas

Consentimiento informado: El desarrollo del presente estudio se ajustó a los principios señalados en la Declaración de Helsinki de la Asociación Médica Mundial, el Informe Belmont, las pautas CIOMS y la normativa colombiana según la Resolución 8430 de 1993, por la que se establecen las normas científicas, técnicas y administrativas para la investigación en salud y, para este caso en particular, la protección de datos clínicos derivados del manejo de la historia clínica, reglamentada por la Resolución 1995 de 1999 y la Ley Estatutaria de habeas data 1581 de 2012, por la cual se dictan las disposiciones generales para la protección de datos personales sancionada mediante la Ley 1581 de 2012 y reglamentada por el Decreto Nacional 1377 del 2013, que regula el manejo adecuado de datos sensibles. Además, fue aprobado por el Comité de Ética del Hospital Militar Central de Bogotá, D.C.

Conflictos de interés: Los autores declararon que no presentan ningún conflicto de interés.

Uso de Inteligencia Artificial: Los autores declararon que el uso de herramientas de Inteligencia Artificial constituye el eje central de este estudio.

Fuentes de financiación: El estudio no recibió recursos de ninguna fuente de financiación.

Contribución de los autores:

- Concepción y diseño del estudio: Álvaro Javier Bayona-Daza, Jaime Andrés Marín.
- Análisis de los datos: Álvaro Javier Bayona-Daza, Erika Tatiana Mendizábal-Varona.
- Redacción del manuscrito: Álvaro Javier Bayona-Daza, Erika Tatiana Mendizábal-Varona.
- Revisión crítica: Jaime Andrés Marín.

Referencias

1 Mán E, Zilágyi A, Simonka Z, Rárosi F, Pető Z, Lázár G. Validation of the modified Alvarado score on patients attending A&E units with suspected appendicitis. BMC Emerg Med. 2023;23:87. https://doi.org/10.1186/s12873-023-00846-2

2 Toney J, Rajesh PJ. Appendicitis inflammatory response score in comparison to Alvarado score in acute appendicitis. Surg J (NY). 2021;7:e127-e131. https://doi.org/10.1055/s-0041-1731446

3 Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986;15:557-64. https://doi.org/10.1016/S0196-0644(86)80993-3

4 Gupta S, Kolli VS, Da Costa K, Javed S, Ammar A, Rasheed A. A systematic review and meta-regression for validation of the Alvarado score as a tool for predicting acute appendicitis. Ann Med Surg (Lond). 2023;85:111-21. https://doi.org/10.1097/MS9.0000000000000238

5 Andersson M, Andersson RE. The appendicitis inflammatory response score: A tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. World J Surg. 2008;32:1843-9. https://doi.org/10.1007/s00268-008-9649-y

6 Ohle R, O'Reilly F, O'Brien KK, Fahey T, Dimitrov BD. The Alvarado score for predicting acute appendicitis: A systematic review. BMC Med. 2011;9:139. https://doi.org/10.1186/1741-7015-9-139

7 Bindra S, Jain R. Artificial intelligence in medical science: A review. Ir J Med Sci. 2024;193:1419-29. https://doi.org/10.1007/s11845-023-03570-9

8 Lam A, Squires E, Tan S, Swen NJ, Barilla A, Kovoor J, et al. Artificial intelligence for predicting acute appendicitis: A systematic review. ANZ J Surg. 2023;93:2070-8. https://doi.org/10.1111/ans.18610

9 Park JJ, Kim KA, Nam Y, Choi MH, Choi SY, Rhie J. Convolutional-neural-network-based diagnosis of appendicitis via CT scans in patients with acute abdominal pain presenting in the emergency department. Sci Rep. 2020;10:9556. https://doi.org/10.1038/s41598-020-66674-7

10 OpenAI. ChatGPT. Large language model. 2023. Disponible en: https://chat.openai.com/chat

11 Bejarano M, Gallego CX, Gómez JR. Frecuencia de abdomen agudo quirúrgico en pacientes que consultan al servicio de urgencias. Rev Colomb Cir. 2011;26:33-41.

12 Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. World J Emerg Surg. 2020;15:27. https://doi.org/10.1186/s13017-020-00306-3

13 Di Saverio S, Birindelli A, Kelly MD, Catena F, Weber DG, Sartelli M, et al. WSES Jerusalem guidelines for diagnosis and treatment of acute appendicitis. World J Emerg Surg. 2016;11:34. https://doi.org/10.1186/s13017-016-0090-5

14 Lamm R, Kumar SS, Collings AT, Haskins IN, Abou-Setta A, Narula N, et al. Diagnosis and treatment of appendicitis: Systematic review and meta-analysis. Surg Endosc. 2023;37:8933-90. https://doi.org/10.1007/s00464-023-10456-5

15 Andersson REB. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. Br J Surg. 2004;91:28-37. https://doi.org/10.1002/bjs.4464

16 Issaiy M, Zarei D, Saghazadeh A. Artificial intelligence and acute appendicitis: A systematic review of diagnostic and prognostic models. World J Emerg Surg. 2023;18:59. https://doi.org/10.1186/s13017-023-00527-2