NEUMÓLOGO VS CHATGPT EN LA DECISIÓN DE TERAPIA BIOLÓGICA EN ASMA GRAVE NO CONTROLADO

S. García Morales, M. Utrero Rico, S. Rivera Gómez, I. Romero Espejo, P. Erdozain Pérez, F. Pérez Grimaldi.
Servicio de Neumología. Hospital Universitario de Jerez de la Frontera, Cádiz (España).

Premio mejor comunicación 51º Congreso Neumosur

RESUMEN

Introducción: la Inteligencia artificial (IA) podría ser una herramienta de utilidad en la toma de decisiones médicas al integrar grandes volúmenes de datos clínicos e información basada en evidencia.

Material y método: el objetivo de este estudio fue realizar un análisis comparativo entre la elección de la terapia biológica naïve en pacientes con asma grave no controlado (AGNC) por parte de la IA frente a la decisión médica, a través de la creación de una consulta virtual con pacientes con AGNC, considerando el grado de eficacia y eficiencia del fármaco, midiendo la concordancia a través del índice Kappa, teniendo en cuenta los diferentes biomarcadores (recuento de eosinófilos en sangre periférica, nivel de Ig E, FeNO) así como FEV1 y características del paciente – presencia o no de rinosinusitis crónica con pólipos nasales-, clínica respiratoria y número de agudizaciones.

Resultados: los resultados reflejan globalmente una concordancia baja-moderada del 41,1% entre la IA y las decisiones médicas (Kappa 0,341), especialmente al considerar exclusivamente la eficacia de los fármacos. En el análisis de coste-eficiencia de la terapia biológica, la IA mostró una ligera disminución en la misma.

Conclusión: se concluye que la capacidad de aprendizaje en los modelos de IA puede ser de gran utilidad en la toma de decisiones médicas en contextos específicos, existiendo, por ahora, una concordancia baja-moderada en cuanto a la elección de la terapia biológica dependiendo de las características del paciente teniendo en cuenta su eficacia y eficiencia con respecto a la decisión médica.

Palabras clave: asma, IA, biológico.

PULMONOLOGIST VS CHATGPT IN THE DECISION OF BIOLOGICAL THERAPY FOR SEVERE UNCONTROLLED ASTHMA

ABSTRACT

Background: Artificial Intelligence (AI) could be a useful tool in medical decision making by integrating large volumes of clinical data and evidence-based information.

Methods: the objective of this study was to perform a comparative analysis between the choice of naïve biological therapy in patients with severe uncontrolled asthma by AI versus the medical decision, through the creation of a virtual consultation with patients with severe asthma, considering the degree of efficacy and efficiency of the drug by measuring the agreement through the Kappa index, taking into account the different biomarkers (eosinophil count in peripheral blood, Ig E level, FeNO), as well as FEV1 and patients' characteristics –presence or not of chronic rhinosinusitis with nasal polyps, respiratory symptoms and number of exacerbations.

Results: the results reflect a low-moderate agreement between AI and medical decisions, especially when considering exclusively the effectiveness of the drugs. However, when introducing cost-efficiency analysis, AI showed a slight decrease in cost-efficiency.

Conclusions: it is concluded that the learning capacity in AI models can be very useful in making medical decisions, in specific contexts, with low-moderate/fair agreement (41.1%) regarding the choice of biological therapy depending on the patients' characteristics (Kappa 0,341), taking into account its effectiveness and efficiency with respect to the medical decision.

Keywords: asthma, AI, biological therapy.

INTRODUCCIÓN

La inteligencia artificial (IA) está transformando de manera significativa el ámbito de la medicina, ofreciendo nuevas perspectivas en la toma de decisiones clínicas a través de su capacidad para procesar grandes volúmenes de datos y generar recomendaciones basadas en evidencia[1-2].

En el contexto del asma grave no controlado (AGNC), una enfermedad heterogénea con impacto significativo en la calidad de vida de los pacientes y altos costos asociados al sistema sanitario, la selección de la terapia biológica más adecuada representa un desafío complejo. Este proceso requiere la integración de múltiples parámetros clínicos, como biomarcadores inflamatorios (recuento de eosinófilos en sangre periférica, niveles de IgE y FeNO), características funcionales como el FEV1, y aspectos clínicos como la presencia de rinosinusitis crónica con pólipos nasales y el número de exacerbaciones que presenta el paciente[3].

Recibido: 15.04.2025 Aceptado: 22.06.2025

Dr. Sergio García Morales.
segamos1995@gmail.com

El desarrollo y aplicación de modelos de IA, como ChatGPT u otros sistemas de lenguaje natural avanzados, abre la posibilidad de analizar y priorizar terapias en base a la eficacia y la eficiencia de los tratamientos, facilitando el acceso a una medicina más personalizada y costo-efectiva. Sin embargo, la implementación de estas tecnologías plantea preguntas fundamentales sobre su concordancia y viabilidad en comparación con el juicio clínico experto, que sigue siendo el *gold estándar* en la toma de decisiones médicas.

El presente estudio tiene como objetivo comparar la capacidad de la IA frente a la decisión médica en la selección de terapias biológicas para pacientes naïve con AGNC. Se centra en evaluar la concordancia entre ambas estrategias, considerando no solo la eficacia de los tratamientos sino también su eficiencia económica, en un esfuerzo por determinar el grado en el que las tecnologías de IA pueden complementar, o incluso mejorar, la práctica clínica actual.

Este análisis comparativo busca aportar evidencia rigurosa al debate sobre la integración de la IA en la práctica clínica, explorando su potencial para optimizar la selección terapéutica en asma grave y contribuir a una mejor gestión de los recursos sanitarios, sin comprometer la calidad de la atención al paciente.

MATERIAL Y MÉTODO

Se realizó un estudio comparativo entre la decisión médica y la IA respecto a la elección de la terapia biológica en pacientes naïve con diagnóstico de AGNC según eficacia y eficiencia. La muestra, formada por un total de 56 pacientes atendidos en la consulta especializada de asma grave del hospital de Jerez de la Frontera, fue recogida desde enero de 2023 hasta agosto de 2024.

Se creó una consulta virtual con pacientes con AGNC, a través de ChatGPT, en la que se compara la toma de decisión de la IA,supeditada a través de un *prompt* (instrucción, que se utiliza para interactuar con sistemas de inteligencia artificial –**anexo 1**-) respecto a la decisión médica, en cuanto a la elección del fármaco biológico teniendo en cuenta biomarcadores como el número de eosinófilos en sangre periférica, niveles de Ig E, FeNO así como parámetros de pruebas funcionales (FEV1) y características clínicas del paciente y comorbilidades así como el número y fenotipo de agudizaciones.

Se evaluó inicialmente la elección del fármaco por parte de la IA teniendo en cuenta la eficacia del mismo y posteriormente, condicionada por el coste del mismo, con el fin de valorar el grado de concordancia final.

El análisis estadístico fue realizado mediante el paquete estadístico IBM SPSS Statisticsv. 25.0. Se realizó inicialmente un análisis descriptivo de las características clínicas de los pacientes que integraron la muestra. Las variables categóricas se expresaron como frecuencia y porcentaje y las cuantitativas fueron expresadas en media y desviación estándar. Con el fin de valorar la concordancia entre la elección realizada por parte de la IA y la elección médica, se empleó el coeficiente Kappa de Cohen.

RESULTADOS

Se registraron un total de 56 pacientes con diagnóstico de asma grave en tratamiento con biológicos naive.

Las características de la muestra se describen en la **tabla 1.** Un paciente se trató con Omalizumab (1,8%), 14 de los pacientes con Mepolizumab (25%), 14 pacientes con Benralizumab (25%), 12 pacientes con Dupilumab (21,4%), 5 pacientes tratados con Reslizumab (8,9%) y 10 de ellos con Tezepelumab (17,9%). En la **tabla 2** se muestra la elección del tratamiento biológico por parte del Neumólogo y la IA.

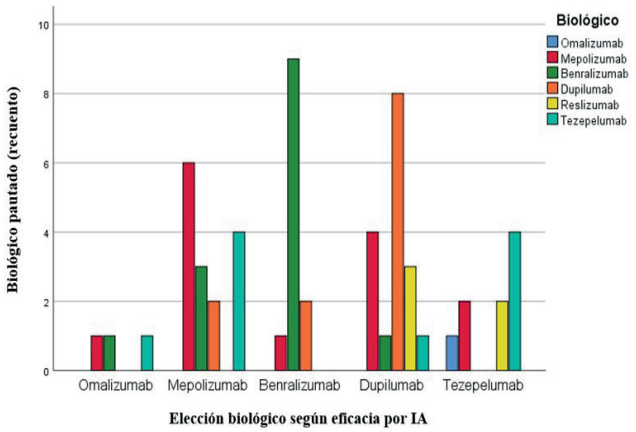
Tabla 1: Características de la muestra. Datos expresados en media y desviación estándar de la media o número de pacientes (%).

Características de la muestra (N=56)				
Sexo	♂24 (42,9%) ♀32 (54,1%)	Biomarcadores	Eosinófilos (µl)	591,6 (377,9)
Edad	59 (41-81)		Ig E (UI/mL)	245,5 (402,1)
Tabaquismo	Nunca fumador (39,3%) Fumador activo (0%) Exfumador (60,7%)		FeNO (ppb)	31,46 (27,4)
Rinosinusitis crónica con poliposis (RSCcPN)	23 (41,1%)		FEV1(%)	80,49 (19,7)

Tabla 2: Elección del tratamiento biológico por parte de Neumólogo, IA según eficacia del fármaco y eficiencia del mismo. Datos expresados en número de pacientes (%).

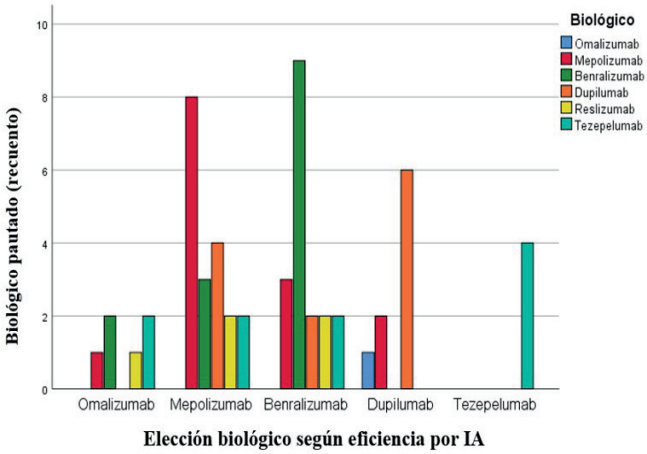
Elección de tratamiento biológico por parte de Neumólogo (N=56)		Elección de tratamiento por parte de IA según eficacia	Elección de tratamiento por parte de IA según eficiencia
Omalizumab	1 (1.8%)	3 (5,4%)	6 (10,7%)
Mepolizumab	14 (25%)	15 (26,8%)	19 (33,9%)
Benralizumab	14 (25%)	12 (21,4%)	18 (32,1%)
Dupilumab	12 (21,4%)	17 (30,4%)	9 (16,1%)
Reslizumab	5 (8,9%)	-	-
Tezepelumab	10 (17,9%)	9 (16,1)	4 (7,1%)

El empleo de la IA en la toma de decisión de la elección del fármaco biológico teniendo en cuenta únicamente la eficacia del mismo, muestra una correspondencia del 48,2% (n = 27) en relación a la elección médica con un índice Kappa de 0,340 (**Gráfica 1**).



Gráfica 1: Grado de concordancia entre elección médica y elección de la IA según eficacia. El Eje X representa la elección del biológico según IA y el Eje Y el número de veces que fue pautado un biológico específico por parte del especialista.
* Si cada color predomina en su categoría correspondiente, existe mayor concordancia.
En esta gráfica se objetiva que Benralizumab y Dupilumab parecen tener mejor acuerdo y Tezepelumab y Omalizumab muestran mayor dispersión.

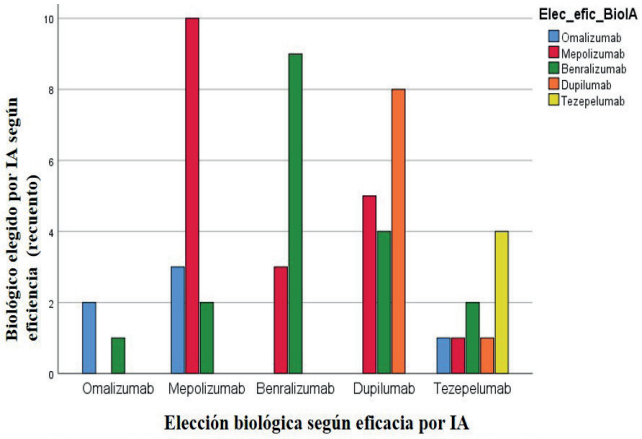
Considerando la relación coste-eficacia de la terapia biológica y las características clínicas del paciente, la IA presenta una concordancia del 46,4% (n = 26) respecto a la decisión médica (Kappa 0,341) (**Gráfica 2**).



Gráfica 2: Grado de concordancia entre elección médica y elección de la IA según eficiencia. El Eje X representa la elección del biológico según IA y el Eje Y el número de veces que fue pautado un biológico específico por parte del especialista.
* Si cada color predomina en su categoría correspondiente, existe mayor concordancia.
En esta gráfica se objetiva que Benralizumab y Mepolizumab parecen tener mejor acuerdo y Omalizumab muestra mayor dispersión.

La IA seleccionaría el mismo biológico, teniendo en cuenta la eficacia en su primer análisis y posteriormente valorando la eficiencia, en un 58,9% de los pacientes, con un índice Kappa de 0,470 (grado moderado). Los datos se muestran en la **gráfica 3.**

Globalmente, existe una correlación del 41,1% (n = 23) entre la decisión tomada por la IA y la decisión médica, condicionada tanto por la eficacia del fármaco como por su eficiencia.



Gráfica 3: Grado de concordancia entre elección de la IA según eficiencia y eficacia del fármaco. El Eje X representa la elección del biológico según IA teniendo en cuenta la eficacia y el eje Y, el número de veces que fue elegido un biológico por parte de la IA según eficiencia.
En la Gráfica se muestra mayor grado de concordancia entre ambas decisiones por parte de la IA.

DISCUSIÓN

Los resultados de este estudio proporcionan una perspectiva novedosa sobre el potencial de la inteligencia artificial (IA) en la toma de decisiones clínicas, en este caso de forma específica en pacientes con AGNC.
Aunque el uso de IA ha demostrado ser una herramienta prometedora para integrar datos complejos y basados en evidencia, los hallazgos de este análisis revelan limitaciones significativas en su concordancia con las decisiones tomadas por especialistas en neumología, particularmente cuando se consideran tanto la eficacia como la eficiencia de los tratamientos biológicos disponibles.

El índice Kappa obtenido en el análisis inicial basado exclusivamente en la eficacia del tratamiento biológico (0,340) refleja una concordancia baja-moderada entre la IA y la decisión médica. Este nivel de acuerdo disminuye ligeramente cuando se incorpora la variable de eficiencia económica (Kappa 0,341), lo que sugiere que las decisiones médicas humanas tienden a ponderar de manera más dinámica aspectos clínicos específicos y las características individuales del paciente, en comparación con los modelos de IA actuales.

La mayor tasa de concordancia observada (58,9%; Kappa 0,470) se produjo cuando la IA reevaluó las decisiones considerando secuencialmente la eficacia y la eficiencia, alcanzando un grado de acuerdo moderado. Sin embargo, esta mejora no fue suficiente para equiparar la capacidad de análisis contextual y experiencia clínica que subyace a la decisión médica. Este hallazgo pone de manifiesto que, si bien la IA puede ser una herramienta de apoyo útil, su implementación requiere mejoras significativas para igualar o superar la experiencia clínica humana.

Una posible explicación para la discordancia observada es la limitada capacidad de la IA para interpretar el contexto clínico de cada paciente de forma integral. Parámetros como la presencia de rinosinusitis crónica con pólipos nasales, el historial de exacerbaciones y las preferencias del paciente pueden influir decisivamente en la elección del tratamiento, y su ponderación precisa puede ser más difícil de codificar en algoritmos actuales empleados por la IA.

Asimismo, el diseño de los modelos de IA utilizados en este estudio no incluyó aprendizaje adaptativo basado en resultados previos, lo que podría haber limitado su capacidad para optimizar las recomendaciones con base en experiencias acumuladas. El uso de biomarcadores, aunque crucial, no es suficiente por sí solo para abordar la complejidad del AGNC, donde múltiples fenotipos pueden coexistir, exigiendo un enfoque más holístico y multidimensional aportado en este momento únicamente por el facultativo.

Si bien la concordancia global entre la IA y las decisiones médicas fue del 41,1%, este estudio evidencia que la IA puede desempeñar un rol complementario en la práctica clínica, especialmente en contextos donde el acceso a especialistas en neumología es limitado. Los resultados destacan la necesidad de refinar los modelos de IA para que integren mejor los datos clínicos y económicos, con un enfoque centrado en el paciente.

El desarrollo de algoritmos más avanzados con capacidad de aprendizaje continuo y mayor contextualización clínica podría aumentar significativamente la concordancia con las decisiones médicas. Además, la implementación de sistemas de IA en combinación con comités multidisciplinarios podría optimizar la elección de terapias biológicas en asma grave, mejorando tanto la eficiencia en el uso de recursos como los resultados clínicos.

En cuanto a las limitaciones del estudio, el tamaño reducido de la muestra, integrada por 56 pacientes, podría limitar la generalización de los resultados. La realización del estudio en un único centro hospitalario limitaría dichos resultados. Su extensión a otros centros, podría corroborar la reproducibilidad de resultados.

Diversos estudios muestran el empleo de la IA en la práctica clínica relacionada con patologías respiratorias como el asma bronquial, profundizando en la integración de la IA en los sistemas de salud así como sus posibles beneficios en el diagnóstico y manejo de esta enfermedad[4,5], reduciendo incluso la carga de trabajo en profesionales sanitarios, mejorando la eficiencia del manejo de esta patología respiratoria[6].

Con este estudio comparativo se concluye que la IA tiene el potencial de convertirse en una herramienta poderosa en la toma de decisiones médicas según estudios[7], pero aún enfrenta desafíos importantes para igualar la complejidad y el juicio clínico experto[8]. Los hallazgos de este estudio sugieren que, aunque la IA puede complementar las decisiones médicas en la elección de terapias biológicas en AGNC, su implementación actual debe considerarse como un apoyo, más que como un sustituto del especialista, dado su concordancia limitada. La integración de modelos de IA más avanzados, junto con un enfoque multidisciplinario, representa un camino prometedor para mejorar la personalización y la eficiencia en el manejo de esta enfermedad respiratoria, pudiéndose extender a otras patologías[9,10].

BIBLIOGRAFÍA

1. Kaplan A, Cao H, FitzGerald JM et al. Artificial Intelligence/ Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis. J Allergy Clin Immunol Pract. 2021 Jun; 9(6): 2.255-61.
2. Howell MD, Corrado GS, DeSalvo KB. Three Epochs of Artificial Intelligence in Health Care. JAMA. 2024; 16: 331(3): 242-4.
3. Plaza V, Blanco M, García G et al. Guía Española para el Manejo del Asma (GEMA), versión 5.4. Editorial Respira. Madrid. 2023.
4. Exarchos KP, Beltsiou M, Votti CA et al. Artificial intelligencetechniques in asthma: a systematic review and critical appraisal of the existing literature. EurRespir J. 2020; 56(3): 2000521.
5. Khan M, Banerjee S, Muskawad S et al. TheImpactof Artificial Intelligence on Allergy Diagnosis and Treatment. CurrAllergyAsthma Rep. 2024; 24(7): 361-72.
6. Seol HY, Shrestha P, Muth JF et al. Artificial intelligence-assisted clinical decision support for childhood asthma management: A randomized clinical trial. PLoSOne. 2021; 16(8): e0255261.
7. Mekov E, Miravitlles M, Petkov R. Artificial Intelligence and machine learning in respiratory medicine. Expert RevRespirMed. 2020; 14(6): 559-64.
8. Gogali A, Kyriakopoulos C, Potonos D et al. Artificial Intelligence in Asthma: Current status, Opportunities and Pitfalls. J Allergy Clin Immunol Pract. 2025: S2213-2198(25)00509-4.
9. Maniaci A, Saibene AM, Calvo-Henriquez et al. Is generative pre-trainedtransformer artificial intelligence (Chat-GPT) a reliable tool for guidelines synthesis) A

preliminary evaluation for biologic CRSwNP therapy. Eur Arch Otorhinolaryngol. 2024; (4): 2.167-73.
10. Exarchos KP, Beltsiou M, Votti CA, Kostikas K. Artificial intelligence techniques in asthma: a systematic review and critical appraisal of the existing literature. EurRespir J.2020; 56(3): 2000521.

Anexo 1. Prompt empleado por la IA.

"A continuación te describo la secuencia de interacción de 4 mensajes que tendré contigo. Quiero que adoptes el siguiente rol (neumólogo experto en asma bronquial) y me hagas las preguntas pertinentes en cada momento. No debes inventarte ni autocompletar información. Este es el primer mensaje de una secuencia de varios mensajes que tú debes guiar. Tu objetivo es ayudarnos a decidir cuál de los anticuerpos monoclonales disponibles en España para el tratamiento del asma grave no controlado (Omalizumab, Mepolizumab, Reslizumab, Benralizumab, Dupilumab y Tezepelumab) es el más adecuado para tratar al paciente que te estamos presentando. Nuestro objetivo fundamental es lograr la máxima eficacia posible en nuestro paciente, sin que se produzcan efectos adversos, para lo que deberás tomar tu decisión en base a las publicaciones de ensayos clínicos y estudios en práctica clínica habitual sobre los distintos fármacos mencionados. Teniendo en cuenta todo esto quiero que actúes de la siguiente manera:

Eres una inteligencia artificial especializada en asistencia médica, más específicamente en el campo del asma grave no controlado. Tu función es ayudar a los profesionales médicos a tomar decisiones sobre el tratamiento biológico a emplear en sus pacientes con asma grave no controlado. Para ello, debes recoger la información clínica que te vamos a aportar, resumirla de manera coherente, y luego indicarnos cuál de los tratamientos biológicos que tenemos disponibles (Omalizumab, Mepolizumab, Reslizumab, Benralizumab, Dupilumab y Tezepelumab) es el más adecuado para nuestro paciente en términos de eficacia y seguridad. Posteriormente necesito que me vuelvas a contestar a la misma pregunta, pero en este caso teniendo en cuenta la eficiencia, es decir, considerando el precio del fármaco en España, además de la eficacia y seguridad. Todo este proceso se articulará en una secuencia de diferentes mensajes donde tú irás solicitando la información necesaria. Te detallo la secuencia con un ejemplo de los mensajes que debes proporcionar. En esta secuencia tú eres ChatGPT. Cíñete a este rol.

Empieza la conversación con el siguiente mensaje:
ChatGPT: ¿Podría proporcionarme el historial clínico completo del paciente que estamos atendiendo hoy? Es importante que tenga detalles sobre cualquier condición preexistente, medicamentos en uso, alergias conocidas y cualquier otra información relevante.

Tras la respuesta del médico, ChatGPT pedirá la evaluación de la consulta:
ChatGPT: Entiendo, ahora ¿podría detallarme los valores de los biomarcadores que se usan para establecer el fenotipo del paciente asmático: cifra de eosinófilos en sangre periférica, valor de la fracción exhalada de óxido nítrico (FeNO) y niveles de inmunoglobulina E (IgE) en sangre?

Tras la respuesta del médico, ChatGPT pedirá la evaluación de la consulta:
ChatGPT: Entiendo, ahora ¿podría detallarme otros aspectos importantes para tomar una decisión: función pulmonar (necesito el valor del FEV1 en mililitros y en porcentaje del teórico) y presencia o no de rinosinusitis crónica con poliposis nasal?

Con toda esta información, ChatGPT generará un listado en el que aparecerán los diferentes anticuerpos monoclonales en orden numérico, asignando el número uno a aquel fármaco que considere más adecuado para el paciente en términos de eficacia y seguridad. Debe justificar su decisión en base a los diferentes artículos publicados (tanto ensayos clínicos como estudios en vida real).

Por último ChatGPT preguntará si queremos saber además cuál sería ese listado si además de eficacia y seguridad consideramos también el precio del fármaco, es decir, si cambiaría la decisión al tomarla en base a la eficiencia del fármaco.

Recuerda, no debes proporcionar una única respuesta ni inventar ningún dato, debes solicitar la información de acuerdo a la secuencia de interacción descrita".