



Cirugía Española



www.elsevier.es/cirugia

P-293 - APLICABILIDAD DE LA INTELIGENCIA ARTIFICIAL EN LA VALORACIÓN DE PACIENTES DE UN SERVICIO DE URGENCIAS QUIRÚRGICAS. ESTUDIO EXPLORATORIO DE SU UTILIDAD EN EL TRIAJE INICIAL

Centeno Álvarez, Ana; López González, Ruth; Badia Closa, Jesús; Sobrerroca Porras, Laura; Castaño Ortega, Juan Carlos; Sala Carazo, Paula; Castellví I Valls, Jordi

Consorci Sanitari Integral-Hospital de l'Hospitalet, L'Hospitalet de Llobregat.

Resumen

Introducción: El alcance de grandes disrupciones tecnológicas como la inteligencia artificial pueden implicar cambios imprevisibles en la práctica médica tal y como la conocemos hoy en día. Pese a las críticas que han recibido otras herramientas interactivas con algoritmos de aprendizaje en la capacidad para emitir un diagnóstico, la potencial aplicabilidad del nuevo modelo de la empresa OpenAI en otros ámbitos lo convierte en una opción prometedora. El objetivo de este estudio es comparar la capacidad del nuevo sistema de chat para identificar patología abdominal y valorar su papel en el triaje inicial de un servicio de urgencias quirúrgico.

Métodos: Estudio unicéntrico retrospectivo transversal de una serie consecutiva de 100 pacientes valorados por el Servicio de Urgencias de Cirugía General para evaluar la capacidad de un sistema de inteligencia artificial para discernir entre presencia de patología intraabdominal aguda o patología banal tributaria de observación domiciliaria, comparándolo con el juicio clínico de los facultativos. Se realizaron las exploraciones complementarias pertinentes según práctica habitual. Los casos fueron presentados a un sistema de chat basado en el modelo de lenguaje por inteligencia artificial (ChatGPT, OpenAI GPT3) de forma secuencial. Se valoró el diagnóstico emitido por el sistema de chat como potencial herramienta diagnóstica, inicialmente con anamnesis y exploración física, y posteriormente con resultados de laboratorio sin pruebas de imagen. Se estudiaron los patrones de validez interna, los valores predictivos y razones de verosimilitud con Stata 15.1.

Resultados: Para la valoración inicial de los pacientes con anamnesis y exploración física, se obtuvo que el uso de ChatGPT como herramienta diagnóstica presentaba una Sensibilidad del 100% [90,4, 100] y una especificidad del 13,2% [5,75, 27,3], con una proporción de falsos positivos (FP) del 86,8% [72,7,94,2], un valor predictivo positivo (VPP) del 52,2% [40,6,63,5] y una eficiencia del 55,4% [44,1,66,2]. Al introducir los resultados de laboratorio, se conseguía mejorar su especificidad a 32% [17,2,51,6], los FP a 68% [48,4,82,2], el VPP a 61,4% [46,6,74,3] y su eficiencia a 67,3% [53,8,78,5].

Conclusiones: Los resultados parecen insuficientes en el momento actual para plantear la planificar una implementación de este tipo de algoritmo en un triaje a corto plazo. Sin embargo, el perfeccionamiento continuo del mismo y autoaprendizaje que presenta lo convierten en una posible

herramienta más a utilizar en un futuro a medio plazo. La aplicabilidad de un sistema de chat basado en lenguaje por inteligencia artificial es todavía limitada, sobre todo en el ámbito de un triaje inicial.