



O-121 - EFICACIA DE CHATGPT-4 PARA LA VIGILANCIA EPIDEMIOLÓGICA DE INFECCIÓN DE LOCALIZACIÓN QUIRÚRGICA COLORRECTAL A PARTIR DE HISTORIAS CLÍNICAS ELECTRÓNICAS. ESTUDIO INFECT-IA3 PRECISIÓN DIAGNÓSTICA Y VALIDACIÓN

Vásquez Candela, Laura¹; Casanova-Portolés, Daniel¹; Castells, Julia¹; Sancho, Joan²; García-Forero, Carlos³; Pujol, Miquel⁴; Badia, Josep M¹

¹Fundació Hospital Asil de Granollers, Granollers; ²Hospital del Mar, Barcelona; ³Universitat Internacional de Catalunya, Barcelona; ⁴Hospital Universitari de Bellvitge, L'Hospitalet de Llobregat.

Resumen

Introducción: La vigilancia de infección de localización quirúrgica (ILQ) se basa en métodos manuales que requieren tiempo y son propensos a la subjetividad. La automatización mediante inteligencia artificial (IA), procesamiento de lenguaje natural y aprendizaje automático podría mejorar la eficiencia y ampliar sus posibilidades. Este estudio evalúa la precisión diagnóstica de ChatGPT para detectar ILQ a partir de la lectura de la historia clínica electrónica (HCE) tras cirugía colorrectal, comparándolo con los resultados de un programa de vigilancia de ILQ.

Métodos: Estudio retrospectivo de una cohorte de pacientes sometidos a cirugía colorrectal electiva entre 2010 y 2023. Se utilizaron las definiciones de ILQ de CDC-NHSN y ECDC: incisional superficial (ILQ-S), incisional profunda (ILQ-P) y de órgano-espacio (ILQ-O/E). Se confeccionó un *prompt* estandarizado para ChatGPT. Las notas clínicas de la HCE fueron extraídas de forma automática para ser analizadas por ChatGPT y comparadas con los resultados del programa de vigilancia manual. Se realizó un estudio univariado y bivariado. Se calcularon la sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN), AUC y curvas ROC.

Resultados: Se incluyeron 1.213 pacientes. La vigilancia manual identificó una tasa de ILQ del 11,2% (ILQ-S 3,1%, ILQ-P 1,2%, ILQ-O/E 6,8%). Para ILQ global, ChatGPT mostró sensibilidad y especificidad del 86%, con VPP 44%, VPN 99% y AUC 0,86 (80,6-89,9). Se observaron diferencias no significativas entre cirugía de colon y recto. Los resultados para ILQ-O/E fueron similares: sensibilidad 86%, especificidad 85%, VPP 99%, VPN 44% y AUC 0,86 (80,6-89,9), pero para ILQ-S e ILQ-P la capacidad diagnóstica fue muy inferior, con sensibilidad 21% y 30%, especificidad 98% y 99%, VPP 97% y 99%, VPN 44% y 44%, AUC 0,6 (49,4-70,3) y 0,6 (43,3-76,4), respectivamente.

Conclusiones: ChatGPT muestra una correcta sensibilidad y especificidad para la detección de ILQ global y ILQ-O/E. El VPN es alto, lo que permite descartar los casos sin infección y el VPP sugiere que se generan falsos positivos. ChatGPT parece útil como herramienta para vigilancia semiautomática y reducción de la carga de trabajo en la revisión manual. El VPP bajo indica la necesidad de ajustes adicionales. La integración del análisis por ChatGPT de HCE con datos

adicionales de tratamiento antibiótico y los resultados de pruebas de imagen o microbiología podría mejorar la vigilancia de la ILQ.