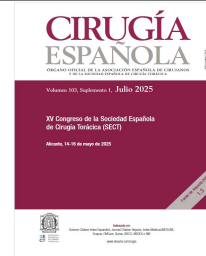




Cirugía Española



www.elsevier.es/cirugia

O-03 - ONCOLOGÍA: EVALUACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA TOMA DE DECISIONES DE UN COMITÉ MULTIDISCIPLINAR DE CÁNCER DE PULMÓN

Álvaro Fuentes Martín¹, Ángel Cilleruelo Ramos¹, Andrea Medina del Valle¹, Rafael López Castro¹, Blanca de Vega Sánchez¹, Juan Francisco Gutiérrez Cara¹, Itziar Gómez-Salvador¹, Manuel Carrasco-Moraleja¹ y Julio Mayol²

¹Hospital Clínico Universitario, Valladolid. ²Hospital Clínico San Carlos, Madrid.

Resumen

Objetivos: La inteligencia artificial (IA) se está posicionando como una herramienta emergente en la toma de decisiones médicas. Sin embargo, su fiabilidad en entornos multidisciplinares, como los comités de tumores (MTB), aún no ha sido suficientemente evaluada. Este estudio evalúa la concordancia entre un comité multidisciplinar y ChatGPT-4.0 en la selección del tratamiento óptimo para pacientes con cáncer de pulmón, comparando una versión sin entrenamiento con otra optimizada mediante guías clínicas.

Métodos: Diseño: estudio experimental basado en la simulación de decisiones clínicas mediante pacientes virtuales generados con la propia IA. Población: se generaron 102 pacientes virtuales tras el cálculo del tamaño muestral. Se excluyeron cinco pacientes por inconsistencias en la generación de datos, obteniendo una cohorte final de 97 casos con distribución representativa en términos epidemiológicos y de características clínicas, incluyendo estadio tumoral, biomarcadores y estado funcional. Evaluación de decisiones: Comité multidisciplinar de referencia; ChatGPT-4.0 sin entrenamiento; ChatGPT-4.0 entrenado con guías clínicas actualizadas (NCCN, ESMO, ESTS). Análisis de concordancia: selección de la decisión diagnóstica o terapéutica inicial (12 opciones, incluyendo la necesidad de pruebas diagnósticas adicionales). Selección del tratamiento definitivo tras completar el estudio diagnóstico (11 opciones). Análisis estadístico: se calculó el coeficiente kappa de Fleiss con intervalos de confianza del 95% para medir la concordancia interevaluador.

Resultados: Se evaluaron 97 pacientes virtuales, obteniéndose los siguientes valores de concordancia (kappa de Fleiss, IC95%): en la evaluación considerando 12 opciones posibles (incluyendo la opción de realizar pruebas diagnósticas adicionales): MTB vs. ChatGPT-4.0 sin entrenamiento: $k = 0,413$ (0,304-0,522). Concordancia moderada. MTB vs. ChatGPT-4.0 entrenado: $k = 0,886$ (0,809-0,963). Concordancia casi perfecta. En la evaluación considerando 11 opciones posibles (decisión definitiva): MTB vs. ChatGPT-4.0 sin entrenamiento: $k = 0,678$ (0,569-0,788). Concordancia sustancial. MTB vs. ChatGPT-4.0 entrenado: $k = 0,969$ (0,928-0,999). Concordancia casi perfecta.

Conclusiones: La IA sin entrenamiento muestra una concordancia limitada en la toma de decisiones médicas. Sin embargo, su optimización con guías clínicas mejora significativamente la precisión. Estos hallazgos sugieren que la IA, correctamente entrenada, podría integrarse como herramienta complementaria en la toma de decisiones.