

Translation for AJR article *Technical Adequacy of Fully Automated Artificial Intelligence Body Composition Tools: Assessment in a Heterogeneous Sample of External CT Examinations*, by lead author B. Dustin Pooler from the Department of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA.

Translation by Javier Cuetos from Hospital Universitario Donostia.

Exactitud técnica de herramientas de composición corporal de inteligencia artificial totalmente automatizadas: evaluación en una muestra heterogénea de exámenes TC externos.

RESUMEN:

Antecedentes: Las herramientas de inteligencia artificial (IA) de posible uso clínico en el análisis de los estudios de imagen deben ser sólidas frente a las variaciones esperadas en los parámetros del estudio.

Objetivo: Evaluar la idoneidad técnica de un conjunto de herramientas automatizadas de IA de composición corporal por TC abdominal en una muestra heterogénea de exámenes de TC externos realizados fuera del sistema hospitalario de los autores, así como explorar las posibles razones de fallo de la herramienta.

Métodos: Métodos: Este estudio retrospectivo incluyó a 8949 pacientes (edad media, 55,5 ± 15,9 años; 4256 hombres, 4693 mujeres) que se sometieron a 11699 exámenes de TC abdominal realizados en 777 instituciones externas diferentes utilizando 82 modelos de escáner diferentes de 6 fabricantes distintos, y que fueron posteriormente transferidos a los PACS locales con fines clínicos. Se implementaron tres herramientas de IA automatizadas independientes que evaluaron la composición corporal (atenuación ósea, cantidad y atenuación de músculo, cantidades de grasa visceral y subcutánea) en una serie axial en cada examen. La idoneidad técnica se definió como los valores de salida de la herramienta dentro de los rangos de referencia derivados empíricamente. Los errores (es decir, la salida de la herramienta fuera del rango de referencia) fueron revisados para identificar las posibles causas.

Resultados: Las tres herramientas fueron técnicamente adecuadas en 11431/11699 (97,7 %) de los exámenes, y al menos una herramienta falló en 268/11 699 (2,3 %) casos. Las tasas de adecuación individual fueron del 97,8 %, 99,1 % y 98,0 % para las herramientas de hueso, músculo y grasa respectivamente. Un solo tipo de error de procesamiento de imágenes (error de anisometría debido a información incorrecta de la dimensión del vóxel del encabezado DICOM) representó 81/92 (88 %) de los exámenes en los que fallaron las tres herramientas, y las tres fallaron cada vez que se produjo este error. El error de anisometría fue la causa específica más común de error para todas las herramientas (31,6 % para hueso, 81,0 % para músculo y 62,8 % para grasa). Se produjeron un total de 79/81 (97,5 %) errores de anisometría en escáneres de un solo fabricante; 80/81 (98,8%) ocurrieron en el mismo modelo de escáner. No se identificó ninguna causa específica de fallo en el 59,4 %, 16,0 % y 34,9 % de los errores para las herramientas de hueso, músculo y grasa respectivamente.

Conclusión: Las herramientas automatizadas de composición corporal de IA tuvieron altas tasas de adecuación técnica en una muestra heterogénea de exámenes de TC externos, lo que respalda la generalización y el potencial uso extendido de las herramientas.

Impacto clínico: Determinadas razones de error de la herramienta de IA relacionadas con factores técnicos pueden prevenirse en gran medida a través de protocolos de adquisición y reconstrucción adecuados.

Autor para correspondencia: B. Dustin Pooler, MD, Departamento de Radiología, Facultad de Medicina y Salud Pública de la Universidad de Wisconsin, Universidad de Wisconsin-Madison, E3/311 Clinical Science Center, 600 Highland Ave., Madison, WI 53792-3252; correo electrónico: bpooler@uwhealth.org ; teléfono: 608-628-5850; fax: 608-263-0140