

## Ecografía de nódulos tiroideos en Niños y Adultos Jóvenes: Comparación del rendimiento diagnóstico de las Impresiones de los Radiólogos, el ACR TI-RADS y un algoritmo de Deep Learning

### RESUMEN:

**Contexto:** En la práctica clínica, la valoración de los nódulos tiroideos de los niños se basa generalmente en la impresión global de los radiólogos de las imágenes de ecografía.

**Objetivo:** Comparar el rendimiento diagnóstico de la impresión global de los radiólogos, el sistema de Datos e Informes de Imagen Tiroidea del Colegio Americano de Radiología (ACR TI-RADS) y un algoritmo de *deep learning* en la diferenciación por ecografía entre nódulos tiroideos benignos y malignos en niños y adultos jóvenes.

**Métodos:** Este estudio retrospectivo incluyó 139 pacientes (mediana de edad 17,5 años; 119 mujeres, 20 varones) evaluados desde el 1 de enero del 2004 hasta el 18 de septiembre de 2020 con una edad  $\leq 21$  y un nódulo tiroideo en ecografía con resultado de anatomía patológica definitivo de una punción aspiración con aguja fina (PAAF) y/o excisión quirúrgica como estándar de referencia. Se seleccionó un único nódulo por paciente, y se obtuvieron sólo una imagen transversal y otra longitudinal del nódulo para su posterior evaluación. Tres radiólogos caracterizaron el nódulo de forma independiente basándose en su impresión global (benigno vs maligno) y el ACR TI-RADS. Un algoritmo de *deep learning* desarrollado previamente determinó para cada nódulo la probabilidad de malignidad, y ésta fue utilizada para obtener un nivel de riesgo. Se calcularon la sensibilidad y la especificidad para malignidad. Se evaluó la concordancia utilizando los coeficientes kappa de Cohen.

**Resultados:** El rango de sensibilidad para la impresión global de los radiólogos fue de 32,1% a 75,0% [media, 58,3% (IC 95%: 49,2-67,3%)], y el de especificidad de 63,8% a 93,9% [media, 79,9% (IC 95%: 73,8-85,7%)]. El rango de sensibilidad del ACR TI-RADS fue de 82,1% a 87,5% [media, 85,1% (IC 95%: 77,3-92,1 %)] y el de especificidad de 47,0% a 54,2 % [media, 50,6% (IC 95%: 41,4-59,8%)]. El algoritmo de Deep learning tuvo una sensibilidad de 87,5% (IC 95%: 78,3-95,5%) y una especificidad de 36,1% (IC 95%: 25,6-46,8%). La concordancia interobservador pareada entre combinaciones de lectores, expresada como kappa, fue para la impresión global de 0,227 a 0,472 y para ACR TI-RADS de 0,597 a 0,643.

**Conclusión:** El ACR TI-RADS y el algoritmo de *deep learning* tuvieron mayor sensibilidad aunque menor especificidad en comparación con las impresiones globales. El algoritmo de *deep learning* tuvo una sensibilidad similar pero menor especificidad que el ACR TI-RADS. La concordancia interobservador fue mayor para el ACR TI-RADS que para las impresiones globales.

**Impacto clínico:** El ACR TI-RADS y el algoritmo de *deep learning* pueden servir como potenciales estrategias alternativas para guiar la decisión de realizar la punción aspiración de aguja fina de los nódulos tiroideos en niños.