

TRABAJO FIN DE GRADO

A stylized illustration of a human brain in shades of pink and light blue. The right side of the brain is overlaid with a blue circuit board pattern, featuring various lines, nodes, and circular components, symbolizing the intersection of medicine and technology.

**EVALUACIÓN DE UN SOFTWARE
DE APOYO AL DIAGNÓSTICO EN
LA EVALUACIÓN DEL NÓDULO
DE TIROIDES**

AUTORA: LAURA RUIZ FERNÁNDEZ

TUTOR: GONZALO DÍAZ SOTO

CURSO 2023-2024



Universidad de Valladolid

Facultad de Medicina

ÍNDICE

1. RESUMEN	2
2. ABSTRACT	3
3. INTRODUCCIÓN	4
4. MATERIAL Y MÉTODOS	6
5. RESULTADOS	8
6. DISCUSIÓN	14
7. BIBLIOGRAFÍA	16
8. PÓSTER	17

1. RESUMEN

Objetivos: Evaluar la utilidad clínica de un sistema de apoyo a la decisión (DSS) basado en inteligencia artificial (IA) en una cohorte de pacientes atendidos en una consulta de endocrinología general por endocrinólogos no especialistas. Así mismo, analizar el grado de concordancia en las características ecográficas definidas por los observadores, el DSS y su evaluación en una consulta específica de Alta Resolución de Nódulo Tiroideo (UART).

Material y métodos: Estudio prospectivo de todos aquellos pacientes a quienes se realizó una ecografía de cribado a criterio del endocrinólogo en una consulta no específica en el Hospital Clínico Universitario de Valladolid (HCUV) durante el año 2022. Se analizaron datos clínicos, bioquímicos, así como las imágenes ortogonales de los nódulos en formato DICOM, sus características ecográficas y criterio de derivación a la UART. Posteriormente, se analizaron las imágenes obtenidas mediante un sistema de IA DSS, así como los casos derivados a la UART.

Resultados: Se analizaron un total de 102 pacientes (94,7% mujeres, edad media $59,7 \pm 14,1$ años, 22.7% antecedentes familiares de bocio).

Los endocrinólogos no especialistas frente a la IA clasificaron los nódulos con un ACR TI-RADS ≤ 3 en el 77% frente al 44%, con una puntuación media de $2,4 \pm 2,0$ puntos frente a un $3,2 \pm 1,0$ ($p < 0,05$). Las categorías ACR TI-RADS mostraron diferencias en composición, ecogenicidad y focos ecogénicos entre en endocrinólogo no especialista y la UART. Finalmente, los endocrinólogos derivaron a la UART al 30,7% de los pacientes, frente a un 37,3% para la IA (ns). La UART realizó la PAAF en un 63,6% del total de pacientes derivados. Los nódulos derivados presentaron con más frecuencia características de agresividad (tamaño máximo, puntos TI-RADS, composición y ecogenicidad $p < 0,05$), con sobreestimación del riesgo al utilizar la clasificación ATA frente al ACR TI-RADS.

Además, se evaluó el grado de concordancia entre las características ecográficas definidas por los endocrinólogos no especialistas y la UART frente a la IA encontrándose que el acuerdo fue globalmente menor en el endocrinólogo no especialista $K_{media} = 0,113 \pm 0,086$ vs $K_{media} = 0,429 \pm 0,314$; $p < 0,05$. Por último, al evaluar la correlación entre la puntuación ACR TI-RADS del endocrinólogo general y la UART

frente a la IA, se encontró una correlación fuerte positiva mayor para la UART (Endocrino general-IA $r=0,337$, $p<0,001$ vs UART-IA $r=0,465$ y $p<0,05$).

Conclusiones: El programa de IA no redujo el número de derivaciones a la UART y mostró una escasa concordancia con la valoración realizada por los endocrinólogos no especialistas, en un contexto de patología nodular de bajo riesgo. Sin embargo, mostró un mejor rendimiento en aquellos nódulos de riesgo intermedio o alto derivados a la UART.

2. ABSTRACT

Objectives: To evaluate the clinical utility of a decision support system (DSS) based on artificial intelligence (AI) in a cohort of patients evaluated in a general endocrinology consult by non-specialist endocrinologists. Likewise, analyze the degree of agreement in the ultrasound characteristics defined by the observers, the DSS, and its evaluation in a specific High Resolution Thyroid Nodule unit (UART).

Methods: Prospective study of patients with thyroid nodules evaluated by ultrasound in the general endocrinology consult of the Hospital Clínico Universitario de Valladolid (HCUV) during the year 2022. Clinical and biochemical data were analyzed, as well as orthogonal images of the nodules in DICOM format, their ultrasound characteristics and criteria for referral to the UART. Subsequently, the images obtained using an AI DSS system were analyzed, as well as the cases referred to the UART.

Results: A total of 102 patients were evaluated (94.7% women, mean age 59.7 ± 14.1 years, 22.7% family history of goiter).

Non-specialist endocrinologists compared to AI classified nodules with an ACR TI-RADS <3 in 77% versus 44%, with a mean score of 2.4 ± 2.0 points versus 3.2 ± 1.0 ($p < 0.05$). The ACR TI-RADS categories showed differences in composition, echogenicity and echogenic foci between the non-specialist endocrinologist and the UART. Finally, endocrinologists referred 30.7% of patients to UART, compared to 37.3% for AI (ns). The UART performed FNA in 63.6% of all referred patients. The derived nodules more frequently presented aggressive characteristics (maximum size, TI-RADS points, composition and echogenicity $p < 0.05$), with overestimation of risk when using the ATA classification versus the ACR TI-RADS.

Furthermore, we evaluate the degree of agreement between the ultrasound characteristics defined by non-specialist endocrinologists and UART versus AI, finding that the agreement was overall lower in the non-specialist endocrinologist $K=0.113\pm 0.086$ vs $K=0.429\pm 0.314$; $p < 0.05$. Finally, when evaluating the correlation between the ACR TI-RADS score of the general endocrinologist and the UART with respect to the IA, a stronger positive correlation was found for the UART (General endocrine-IA $r=0.337$, $p < 0.001$ vs UART-IA $r=0.465$ and $p < 0.05$).

Conclusions: The AI program did not reduce the number of referrals to the UART and showed poor agreement with the assessment carried out by non-specialist endocrinologists, in a context of low-risk nodular pathology. However, it showed better performance in those intermediate or high risk nodules referred to the UART.

3. INTRODUCCIÓN

La patología nodular del tiroides es uno de los problemas clínicos más prevalentes en las consultas de endocrinología (1). La prevalencia de nódulos tiroideos palpables en zonas con aportes de yodo suficientes es del 5% en mujeres y del 1% en hombres (2). En cambio, la ecografía puede detectar nódulos en aproximadamente el 60% de la población, sobre todo en mujeres y personas mayores (2)(3).

La importancia clínica del nódulo tiroideo se debe a la necesidad de excluir un cáncer de tiroides. Sin embargo, sólo en torno al 5% de estos nódulos son finalmente malignos (4). El cáncer diferenciado de tiroides (CDT), donde se incluyen el cáncer papilar y el cáncer folicular comprende más del 90% de todos los cánceres de tiroides (2). La incidencia y prevalencia del diagnóstico de cáncer de tiroides ha aumentado en las últimas décadas; sin embargo, la mortalidad específica por cáncer se mantiene estable (5). Los factores de riesgo asociados con una mayor probabilidad de malignidad incluyen: antecedentes de irradiación del cuello, antecedentes familiares de carcinoma medular de tiroides o neoplasia endocrina múltiple (MEN2), edad < 20 años o > 60 años, sexo masculino, crecimiento rápido, firmeza y consistencia dura y presencia de ganglios linfáticos cervicales sospechosos (6).

En la evaluación de cualquier paciente con sospecha de nódulos tiroideos se incluye la anamnesis; antecedentes personales y familiares, exploración física, exploración funcional del tiroides y ecografía tiroidea. Las pruebas de laboratorio siempre deben incluir la determinación de hormona estimulante del tiroides (TSH). El uso de calcitonina

para la detección del carcinoma medular de tiroides proporciona un diagnóstico temprano y mejora potencialmente el pronóstico. Sin embargo, el valor de este tipo de detección aún está bajo debate. La determinación de calcitonina debe realizarse en caso de antecedentes personales o familiares de cáncer medular de tiroides o neoplasia endocrina múltiple tipo 2. Al contrario, la determinación de tiroglobulina sérica y de anticuerpos anti-Tg no tiene ningún papel en la evaluación inicial de la enfermedad nodular de la tiroides (3).

La ecografía es la principal prueba de imagen para evaluar los nódulos tiroideos. Permite confirmar la presencia, número, y dimensiones de los nódulos. Además, permite definir mediante la caracterización de la imagen por ultrasonidos aquellos que deben ser analizados por aspiración con aguja fina (PAAF) de aquellos que pueden ser seguidos por ecografía, según sus características (2)(3). Sin embargo, la evaluación de un nódulo tiroideo mediante ecografía tiene inconvenientes como la escasa reproducibilidad, debido a los diferentes equipos utilizados, la falta de un informe ecográfico estandarizado y la variabilidad inter e intraobservador. Además, las características de las imágenes de ultrasonido que sugieren malignidad como: hipoecogenicidad, una composición mayoritariamente sólida, una forma más alta que ancha, márgenes irregulares, la ausencia de un halo o la presencia de calcificaciones intranodulares no son lo suficientemente específicas como para diagnosticar definitivamente malignidad por sí solas (6).

Para resolver este problema, se han desarrollado escalas de estratificación del riesgo de malignidad para integrar la información ecográfica a la toma de decisiones clínicas (7)(8). Todas las escalas de estratificación de riesgo han demostrado niveles aceptables de sensibilidad; sin embargo, el American College of Radiology Thyroid Imaging, Reporting, and Data System (ACR TI-RADS) tiene el grado más alto de especificidad, reduciendo el número de PAAF innecesarias y manteniendo valores aceptables de sensibilidad (7)(8)(9). La escala ACR TI-RADS otorga puntos para las características ecográficas de un nódulo, y siendo aquellas altamente sospechosas las que reciben mayor puntuación. El total de puntos determina el nivel ACR TI-RADS del nódulo, que oscila entre TR1 (altamente benigno) y TR5 (alta sospecha de malignidad) (8).

Recientemente se han desarrollado sistemas de apoyo a la decisión (DSS) basados en inteligencia artificial (IA) para apoyar a los médicos en la interpretación de las imágenes ecográficas. Las imágenes médicas contienen información clave que refleja la fisiología subyacente de los tumores, esta información puede ser imperceptible para el ojo

humano (11). Estos sistemas ayudan a reducir el componente subjetivo, disminuyendo así la variabilidad inter- e intraobservador y mejorando del rendimiento diagnóstico de la ecografía de tiroides mediante la ayuda a la valoración de las características de malignidad del nódulo tiroideo (10)(11). Los DSS han demostrado su ayuda a la hora de automatizar procesos, mejorar la precisión y la coherencia del diagnóstico, personalizar los tratamientos, disminuir la carga de los médicos y acelerar la curva de aprendizaje de los profesionales menos experimentados, entre otros (12) (13).

El objetivo del presente trabajo fue valorar la utilidad clínica de la utilización de un DSS en una cohorte de nódulos evaluados en una consulta general de un servicio de endocrinología en práctica clínica real. Así mismo, se evaluó el grado de acuerdo con de las características ecográficas definidas por los observadores, el DSS y la modificación de las derivaciones a la consulta específica de Alta Resolución de Nódulo Tiroideo (UART) de haberse utilizado el DSS.

4. MATERIAL Y MÉTODOS

Se trata de un estudio prospectivo de aquellos pacientes con presencia de nódulos tiroideos evaluados mediante ecografía por parte de especialistas en endocrinología y nutrición en una consulta general de endocrinología. Así mismo se evaluó de manera retrospectiva las imágenes obtenidas por parte de los profesionales sanitarios mediante un sistema de IA DSS. Ambas evaluaciones fueron comparadas respectivamente.

Se incluyeron en el estudio todos los pacientes mayores de 18 años con nódulos tiroideos evaluados en la consulta general de endocrinología durante el año 2022.

Se recogieron datos clínicos y bioquímicos: año de nacimiento, sexo, antecedentes familiares y personales de patología tiroidea, edad al diagnóstico, método diagnóstico, niveles de hormona estimulante de la tiroides (TSH) y tiroxina libre (T4). Así mismo se recogieron las imágenes ortogonales de los nódulos en estudio (transversal/longitudinal) en el Formato DICOM, así como la evaluación de las características nodulares (tamaño, la composición del nódulo, su ecogenicidad, márgenes y focos ecogénicos) realizada por el endocrinólogo general y el análisis realizado por el programa de IA. Fueron excluidos aquellos pacientes que rechazaron entrar en el estudio, así como aquellas pacientes con imágenes ecográficas incompletas o de mala calidad (imagen borrosa o adquisición de imágenes no estándar). Por último, se analizaron aquellos pacientes que fueron derivados a la unidad de Alta Resolución de Nódulo Tiroideo (UART) y se recogió el número de PAAF realizadas, así como las recomendadas por el programa de IA.

Todas las imágenes se adquirieron durante la práctica clínica utilizando un sistema de ultrasonidos Toshiba Xario 100 mediante sonda lineal multifrecuencia. El protocolo incluía al menos dos imágenes ortogonales para cada nódulo, para capturar diferentes aspectos de la morfología y características del nódulo con la mejor resolución posible. Uno de los médicos eligió las imágenes más significativas (transversales y longitudinales) para revisarlas o excluirlas en función de la calidad de la imagen.

AI-Based DSS

El sistema de IA empleado en este estudio utiliza técnicas de visión por ordenador y aprendizaje automático para generar un motor capaz de analizar e interpretar la imagen ecográfica del nódulo tiroideo y su riesgo de malignidad. Las características de los nódulos tiroideos categorizadas por el DSS utilizado en este trabajo coinciden exactamente con las características de sospecha definidas en las guías ACR TI-RADS (con excepción de la extensión extratiroidea) (8).

El DSS evalúa la región de interés (ROI) definida por el usuario, de una imagen de ultrasonido correspondiente al nódulo tiroideo en estudio. En esta imagen, el DSS categoriza con una probabilidad cada uno de los componentes del ACR TI-RADS (composición, ecogenicidad, forma, márgenes y focos ecogénicos). También genera el AI Adapter, que es independiente de la clasificación y evaluación ACR TI-RADS. El adaptador permite la modificación opcional del riesgo total restando o sumando al total de la puntuación ACR TI-RADS -2, -1, 0, +1 o +2 puntos. Así, el AI Adapter permite la incorporación de una evaluación independiente del nódulo tiroideo basada en aprendizaje automático para mejorar la categorización ACR TI-RADS más allá de la evaluación de cada uno de los descriptores individuales del DSS. Finalmente, basándose en la puntuación total final del usuario (incluido el AI Adapter), el sistema realiza una recomendación de acción clínica (esencialmente si se debe realizar una PAAF) para un nódulo tiroideo específico, siguiendo los mismos umbrales de puntos y tamaño que las guías ACR TI-RADS (8).

Análisis estadístico

Los resultados se expresan como media y desviación estándar (DE) para las variables cuantitativas continuas. Las variables cuantitativas con distribución normal se analizaron con una prueba T de Student. Las variables no paramétricas se evaluaron mediante la prueba de Friedman y Wilcoxon. Las variables cualitativas se expresan como porcentaje (%) y se analizaron con la prueba de Chi-cuadrado, empleando el test exacto de Fisher en aquellas ocasiones necesarias.

Para evaluar la asociación de variables cuantitativas se utilizaron diagramas de dispersión y se calculó utilizando el coeficiente de correlación lineal de Pearson. Aquellos valores de $p < 0,05$ fueron considerados estadísticamente significativos.

Se utilizó el paquete estadístico IBM SPSS Statistics® versión 29.0.2.0 (SPSS Inc., Chicago, IL, EE. UU.).

5. RESULTADOS

Se evaluaron un total de 102 pacientes de los cuales fueron incluidos un 73,5% de la muestra por mala calidad de la imagen o imagen incompleta. Finalmente, de la muestra analizada un 94,7% fueron mujeres, y presentaban una edad media al diagnóstico de $59,7 \pm 14,1$ años. El 22,7% presentaba antecedentes familiares de bocio y el 18,7% de hiper o hipotiroidismo.

La forma de diagnóstico más frecuente fue la ecografía tiroidea (49,3%), seguida del nódulo palpable (22,7%). Los nódulos presentaron una lateralización derecha en un 57,3% de los casos. Finalmente, un 30,7% fueron derivados a la UART. (Tabla 1)

CARACTERÍSTICAS BASALES	MEDIA \pm DE/ PORCENTAJE
Número de pacientes	75
Porcentaje de mujeres	94,7
Edad	$59,7 \pm 14,1$
Antecedentes Personales (%)	
• RT Cabeza y Cuello	2,7
• Enfermedad oncológica	1,3
• Ingesta sal yodada	20
Antecedentes Familiares (%)	
• Cáncer tiroides	1,3
• Bocio	22,7
• Hiper/hipotiroidismo	18,7
TSH uUI/ml	$2,2 \pm 2,1$
T4L	$1,3 \pm 0,4$
T3L	$4,4 \pm 2,5$
AntiTPO	$99,7 \pm 138,4$
Dosis de Eutirox	75 ± 34
Forma de diagnóstico (%)	
• Nódulo palpable	22,7
• Incidental TC	5,3
• Ecografía tiroidea	49,3
• Síntomas compresivos	1,3
• Aumento tamaño glándula	2,7

• Alteración funcional	17,3
• PET	1,3
Lateralización derecha (%)	57,3
Ecogenicidad global (%)	
• Homogéneo	72
• Heterogéneo	16
• Tiroiditis	12
Diámetro AP	15,2± 9
Diámetro transversal	11,3±6,1
Diámetro longitudinal	18,5±11,4
Derivados a UART (%)	30,7

Tabla 1: Características descriptivas de la población evaluada.

En la tabla 2 se comparan las características de los nódulos que conforman la clasificación ACR TI-RADS para el endocrinólogo general, el endocrinólogo especialista en tiroides de la UART y la evaluación realizada por el programa de IA. Para el endocrinólogo general la composición más frecuente fue espongiiforme con un 38,7%, seguida de la composición sólida en un 36%. En cambio, para la IA el 80% de los nódulos fueron sólidos. La mayor parte de los nódulos se definieron como iso o hiperecogénicos tanto en las consultas de endocrinología general (57,3%) como en el programa de IA (46,7%). Los márgenes fueron bien definidos en más del 90% de los casos. La mayor parte de los nódulos no presentó focos ecogénicos, siendo la cola de cometa el más frecuente en las consultas (21,3%) y las microcalcificaciones en la IA (10,7%). El ACR TI-RADS calculado por el endocrino general estuvo por debajo de 3 en el 77% de los pacientes, con una puntuación ACR TI-RADS de 2,4±2.0 puntos. En cambio, la IA obtuvo un ACR TI-RADS ≤ 3 en el 44% con una puntuación de 3,2±1.0 puntos. Los endocrinólogos generales derivaron a la UART un 30,7% de los pacientes y la IA recomendó la PAAF en un 37,3%.

De los nódulos derivados a la UART (30.7% de la muestra), el 74% fueron sólidos y el 77% iso o hiperecogénicos. Los márgenes fueron bien definidos en el 95,2% y en su mayoría, no se encontraron focos ecogénicos. La clasificación ACR TI-RADS fue menor de 3 en el 65% de los nódulos con una puntuación de 3,2±1.0. Se recomendó la PAAF en un 24,6% del total. (Tabla 2).

Características (%)	Endocrino General	IA	UART
Más ancho que largo	93,3	77,3	8,7
Composición *			
• Sólido	36	80	73,9
• Solidoquístico	21,3	16	8,7
• Espongiforme	38,7	1,3	8,7
• Quístico	4	2,7	8,7
Ecogenicidad *			
• Muy hipoecoico	1,3	14,7	18,2
• Hipoecoico	14,7	37,3	77,3
• Iso/hiperecoico	57,3	46,7	4,5
• Anecoico	26,7	1,3	
Márgenes			
• Bien definidos	90,7	94,7	95,2
• Lobulados/irreg	4	5,3	4,8
• No visibles	5,3		
Focos ecogénicos *			
• Cola cometa	21,3	1,3	22,7
• Macrocalcificación	17,3	4	
• Calcificación anillo	2,7	1,3	4,5
• Microcalcificación	1,3	10,7	72,7
• Ninguno	57,3	76	
TIRADS ≤ 3 *	77,3	44	65
Puntos TIRADS $\bar{x} \pm DE$	2,4 \pm 2,0	4,4 \pm 4,6	3,2 \pm 1,0
ATA *			
• Benigno	22,7	1,6	
• Muy baja sospecha	32	4,8	
• Baja sospecha	29,3	46	
• Intermedia	12	7,9	
• Alta sospecha	4	39,7	
PAAF	30,7	37,3	24,6

Tabla 2: Características descriptivas de los nódulos en endocrinología general, en la IA y en la UART.

* $P < 0.05$ p valor para comparativa Endocrinólogo general vs IA.

Por otro lado, se evaluó el grado de acuerdo entre las diferentes características ecográficas que definen la escala de riesgo ACR TI-RADS entre el endocrinólogo general y el endocrinólogo especialista en tiroides de la UART con respecto a la IA. Se observó un grado de acuerdo globalmente inferior en el caso del endocrinólogo general y la IA frente al endocrinólogo especialista de la UART y la IA, siendo la ecogenicidad la variable con mayor grado de concordancia entre endocrinología general y el programa de IA ($k=0,511$, $p < 0.05$) así como la presencia de focos ecogénicos ($k=0,405$, $p < 0,05$). (Tabla 3).

CARACTERÍSTICAS	ENDOCRINO GENERAL vs IA		UART vs IA	
	<i>Kappa</i>	<i>p</i>	<i>Kappa</i>	<i>p</i>
Más ancho que largo	0,189	<0,05	0,181	<0,05
Composición	0,226	<0,05	0,352	<0,05
Ecogenicidad	0,107	P=0,08	0,511	<0,05
Márgenes	0,051	ns	1.000	ns
Focos ecogénicos	0,180	<0,05	0,405	<0,05
TIRADS	0,026	ns	0,123	<0,05
ATA	0,012	ns		

Tabla 3: Nivel de acuerdo en las características nodulares entre el endocrinólogo general y la UART con la IA.

Sin embargo, la comparación entre la evaluación del endocrinólogo general y la del endocrinólogo especialista de la UART mostró índices de acuerdo semejantes a los del endocrinólogo especialista de la UART frente al programa de IA.

Se observó un acuerdo $k=0,637$ en la ecogenicidad, $k=0,495$ en la composición y $k=0,472$ en los focos ecogénicos, estadísticamente significativos $p<0,05$. (Tabla 4).

CARACTERÍSTICAS	ENDOCRINO GENERAL vs UART	
	<i>Kappa</i>	<i>p</i>
Más ancho que largo	0,187	<0,05
Composición	0,495	<0,05
Ecogenicidad	0,637	<0,05
Márgenes	0,294	<0,05
Focos ecogénicos	0,472	<0,05
TIRADS	0,322	<0,05

Tabla 4: Nivel de acuerdo en las características nodulares entre el endocrinólogo general y la UART.

También se evaluó el grado de acuerdo mediante la correlación entre las puntuaciones obtenidas para el ACR TI-RADS del endocrinólogo general, el endocrinólogo especialista de la UART y el programa de IA.

En la gráfica se muestra la Correlación de Pearson para los Puntos de la escala ACR TI-RADS del endocrinólogo general respecto a la IA con $r=0,337$ y $p<0,001$. Así mismo, se obtuvieron correlaciones significativas en el resto de puntuaciones analizadas: UART-IA $r=0,465$ y $p<0,05$ y Endocrino general-UART $r=0,607$ y $p<0,05$. (Figura 1)

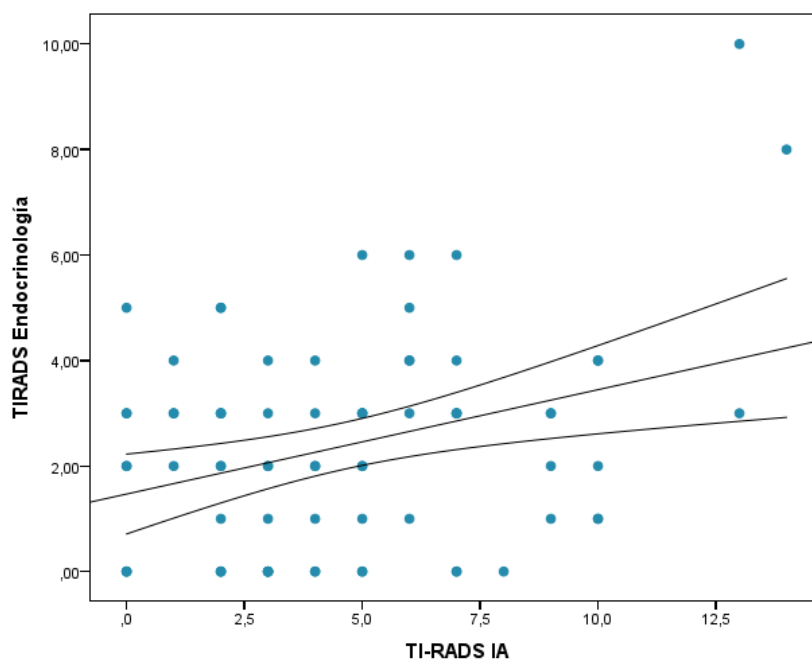


Figura 1: Correlación de Pearson para los Puntos ACR TI-RADS del Endocrinólogo general respecto a la IA.

Se compararon las características de los nódulos que fueron derivados a la UART respecto a los que no lo fueron. En la tabla 5 se recogen aquellas variables significativas. Se observa como los nódulos derivados fueron de mayor tamaño en las tres dimensiones analizadas ($p < 0.01$), así como la puntuación para la escala ACR TI-RADS fue mayor en los nódulos derivados a la UART, $3,8 \pm 2,2$ respecto a $1,8 \pm 1,6$ puntos, $p < 0,01$. Por otro lado, la ecogenicidad global de la glándula fue heterogénea en el 47,8% de los nódulos derivados frente al 19,2%, $p < 0,05$. El 60,9% de los nódulos derivados fueron sólidos, $p < 0,01$. Finalmente, respecto a las categorías ATA y ACR TI-RADS, el 39,1% de los nódulos derivados fueron de intermedia y alta sospecha y el 43,5% tuvieron un ACR TI-RADS 4 o 5. (Tabla 5).

CARACTERÍSTICAS	NO UART	SÍ UART	p
Diámetro anteroposterior	13±6,8	20±11,2	<0,01
Diámetro transversal	9,9±5,4	14,3±6,6	<0,01
Diámetro longitudinal	15,4±8,8	25,6±13,5	<0,01
Puntos TIRADS	1,8±1,6	3,8±2,2	<0,01
Más ancho que largo (%)	98,1%	82,6%	<0,05
Ecogenicidad global (%)			
• Tiroiditis/heterogéneo	19,2%	47,8%	<0,05
Composición (%)			
• Sólido	25%	60,9%	<0,01
Ecogenicidad (%)			
• Anecoico	36,5%	4,3%	<0,05

ATA (%) • Sospecha intermedia y alta	5,7%	39,1%	<0,01
TIRADS (%) • TIRADS 4-5	13,5%	43,5%	<0,01

Tabla 5: Comparación entre las características nodulares de los nódulos derivados a la UART con los que no lo fueron.

Por último, se analizaron las diferencias de clasificación entre las clasificaciones ATA2015 y ACR TI-RADS realizadas mediante el programa de IA. Se observó cómo los nódulos analizados mediante la clasificación ATA2015 como sospecha baja, intermedia y alta eran evaluados como ACR T-IRADS de menor riesgo en un 31.9% de los casos (Figura 2).

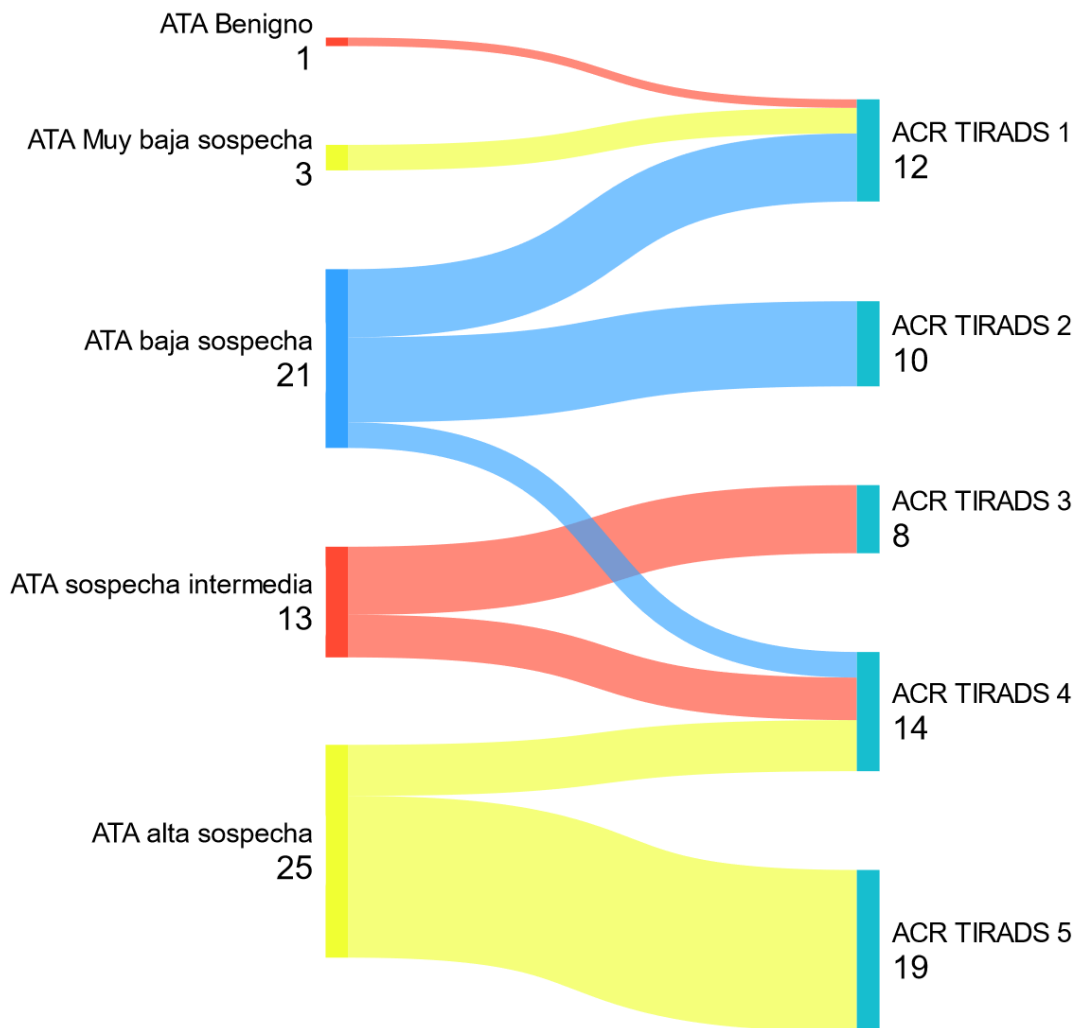


Figura 2: Clasificación de nódulos como ATA y ACR TI-RADS mediante el programa de IA.

6. DISCUSIÓN

La patología nodular del tiroides es uno de los problemas clínicos más prevalentes en las consultas de endocrinología (1). La importancia clínica del nódulo tiroideo se debe a la necesidad de excluir un cáncer de tiroides, aunque finalmente sólo en torno al 5% de estos nódulos son malignos (4). Por lo tanto, se trata de una patología altamente prevalente con un diagnóstico mayoritariamente benigno, lo que supone una sobrecarga asistencial muy importante (13). Estudios aún preliminares sugieren que los sistemas de apoyo a la decisión (DSS) basados en inteligencia artificial (IA) podrían ser una alternativa a la derivación de pacientes a las unidades de Alta resolución de tiroides (UART) para reducir el número de estas e incrementar la necesidad de realización de PAAF en aquellos pacientes derivados.

El presente estudio evalúa la utilidad clínica de la utilización de un DSS en una cohorte de nódulos evaluados en una consulta general de un servicio de endocrinología en práctica clínica real. Así como el grado de acuerdo en las características ecográficas definidas por los observadores de la consulta de endocrinología general, el DSS basado en IA y los endocrinólogos especialistas en tiroides de la UART. La principal conclusión del presente estudio demuestra como el sistema de IA no reduce significativamente el número de derivaciones a la UART, definidas como aquellos casos en los que aconseja la realización de PAAF. De hecho, el endocrinólogo general derivó el 30,7% de los pacientes a la UART para realizar una PAAF, mientras que el programa de IA recomendó la PAAF en un 37,3%.

Resulta significativo como las características de la población estudiada coinciden con aquellas que son más frecuentemente evaluadas en endocrinología por esta patología: mujeres en edades medias de la vida, lo cual se corresponde con la epidemiología de la patología tiroidea (2). De la misma manera, tanto las características ecográficas como el tamaño nodular e incluso las diferencias entre los distintos sistemas de clasificación nodular (Tabla 5 y Figura 2) corresponden a aquellas descritas en la literatura médica (13).

Sin embargo, llama la atención la poca concordancia entre el análisis de los componentes ACR TI-RADS del endocrinólogo general y de la IA, especialmente en composición, ecogenicidad y presencia de focos ecogénicos. Además, la mayoría de nódulos evaluados por el endocrinólogo general, el 77%, estuvo por debajo de un ACR TI-RADS 3 con una puntuación de $2,4 \pm 2,0$ puntos, frente al 44% y $4,4 \pm 4,6$ puntos para

el programa de IA ($p < 0.05$). Estas diferencias fueron mucho menores cuando se evaluó exclusivamente el número de pacientes derivados a la UART, de hecho, las características definidas por la IA no presentan diferencias entre el análisis del endocrinólogo especialista de la UART y la IA.

Son muchas las consideraciones que pueden realizarse a este respecto, en primer lugar, este tipo de sistemas están entrenados para asegurar el menor número de falsos negativos evitando punciones innecesarias pero priorizando el estudio de aquellos nódulos tiroideos que pudieran ser malignos aunque la sospecha sea baja (14). Además, el tipo de paciente evaluado en la consulta de endocrinología general correspondería a pacientes con patología de muy bajo riesgo, por lo que la población estudiada estaría evaluando nódulos cuya duda diagnóstica no hubiera exigido la utilización de la IA entrenada en el estudio del cáncer de tiroides. Sin embargo, aquellos nódulos derivados a la UART y por lo tanto con un ACR TI-RADS de riesgo alto, mostraron un grado de acuerdo más adecuado entre el endocrinólogo especialista de la UART y la IA.

De la misma manera, la evaluación de cada uno de los componentes del ACR TI-RADS mostró que el grado de acuerdo entre las características ecográficas definidas por el endocrinólogo general respecto a la IA fue pobre. El índice Kappa fue menor de 0,2 para todas las características excepto para la composición, $k=0,226$. En cambio, cuando comparamos la IA con el endocrino especialista de la UART observamos un nivel de acuerdo globalmente mayor. Características como la ecogenicidad y los focos ecogénicos obtuvieron un índice Kappa mayor de 0,4. Esta mejora de los resultados podría deberse a que el endocrino especialista está más formado en la evaluación del nódulo tiroideo mediante ecografía que el endocrino de la consulta general, sin embargo, no podemos descartar una peor evaluación del programa de IA cuando analiza imágenes de muy bajo riesgo. La buena correlación en el análisis entre el endocrinólogo general y el de la UART apoyarían la tesis de un peor rendimiento del programa de IA en consultas no específicas.

En cualquier caso, la correlación de Pearson realizada para las puntuaciones en la escala ACR TI-RADS fue una correlación significativa tanto en la comparación Endocrino general-IA, como UART-IA y Endocrino general-UART. La r más alta, una vez más, se obtuvo en la correlación Endocrino general-UART y fue $r=0,607$, apoyando mayores puntuaciones en aquellos nódulos más sospechosos independientemente de quién realizara el análisis.

Estudios previos han analizado sistemas de IA específicos individualmente, mostrando una alta especificidad especialmente si se basan en el algoritmo del sistema ACR TI-RADS. Algunos estudios muestran una mejora en la precisión diagnóstica, sobre todo, si hablamos de profesionales jóvenes con menos experiencia. También se ha visto que podrían evitar punciones innecesarias, aunque en el caso de nuestro estudio, no ha sido así (15). Sin embargo, a nuestro conocimiento es la primera vez que se evalúa en este tipo de entorno. Son necesarios ensayos multicéntricos prospectivos en una población amplia para mejorar la utilidad de la inteligencia artificial en la interpretación de los nódulos tiroideos (16).

Nuestro estudio presenta algunas limitaciones, en primer lugar, el número de pacientes evaluados es pequeño para un estudio en inteligencia artificial. Sin embargo, presenta importantes fortalezas, como son el conocimiento exhaustivo de las características clínicas de los pacientes evaluados y además la descripción prospectiva de endocrinólogos con mayor o menor grado de experiencia en ecografía. Por último, es el primer estudio en evaluar un programa de IA en derivaciones a una consulta específica de Alta Resolución de Nódulo Tiroideo (UART).

En conclusión, el programa de IA mantuvo el número de pacientes derivados a la UART y mostró una escasa concordancia con la valoración realizada por los endocrinólogos generales en un contexto de patología nodular de bajo riesgo. Sin embargo, mostró un mejor rendimiento en aquellos nódulos de riesgo intermedio o alto derivados a la UART.

7. BIBLIOGRAFÍA

1. Valentí PF, Farreras-Rozman. Medicina interna. 17.^a ed. Vol. 1. Barcelona: Elsevier; 2012. 3111 p.
2. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 1 de enero de 2016;26(1):1-133.
3. Durante C, Hegedüs L, Czarniecka A, Paschke R, Russ G, Schmitt F, et al. 2023 European Thyroid Association Clinical Practice Guidelines for thyroid nodule management. *Eur Thyroid J*. 1 de octubre de 2023;12(5):e230067.
4. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The Diagnosis and Management of Thyroid Nodules: A Review. *JAMA*. 6 de marzo de 2018;319(9):914-24.

5. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974-2013. *JAMA*. 4 de abril de 2017;317(13):1338-48.
6. Rago T, Vitti P. Risk Stratification of Thyroid Nodules: From Ultrasound Features to TIRADS. *Cancers*. 30 de enero de 2022;14(3):717.
7. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, et al. Comparison of Performance Characteristics of American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association Guidelines. *Am J Roentgenol*. mayo de 2018;210(5):1148-54.
8. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol*. 1 de mayo de 2017;14(5):587-95.
9. Hoang JK, Middleton WD, Langer JE, Schmidt K, Gillis LB, Nair SS, et al. Comparison of Thyroid Risk Categorization Systems and Fine-Needle Aspiration Recommendations in a Multi-Institutional Thyroid Ultrasound Registry. *J Am Coll Radiol*. 1 de diciembre de 2021;18(12):1605-13.
10. Yao W, Peng X, Guan Y, Du X, Xia C, Liu F. Thyroid Nodules: Emerging Trends in Detection and Visualization based on Citespace. *Endocr Metab Immune Disord Drug Targets*. 29 de diciembre de 2023;24(1):130-41.
11. Zhu YC, Jin PF, Bao J, Jiang Q, Wang X. Thyroid ultrasound image classification using a convolutional neural network. *Ann Transl Med*. octubre de 2021;9(20):1526.
12. Toro-Tobon D, Loo-Torres R, Duran M, Fan JW, Singh Ospina N, Wu Y, et al. Artificial Intelligence in Thyroidology: A Narrative Review of the Current Applications, Associated Challenges, and Future Directions. *Thyroid Off J Am Thyroid Assoc*. agosto de 2023;33(8):903-17.
13. Díaz-Soto G, Torres B, López Gómez JJ, Gómez Hoyos E, Villar A, Romero E, et al. Economic impact of and satisfaction with a high resolution thyroid nodule clinic at the endocrinology department. *Endocrinol Nutr Organo Soc Espanola Endocrinol Nutr*. octubre de 2016;63(8):414-20.
14. Fernandez-Velasco P, Pérez López P, Torres B, Delgado E, De Luis D, Diaz-Soto G. Clinical evaluation of an Artificial Intelligence–based decision support system for the diagnosis and ACR TI-RADS classification of thyroid nodules. *Thyroid®*. 18 de febrero de 2024;thy.2023.0603.
15. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health*. 1 de abril de 2021;3(4):e250-9.
16. Sorrenti S, Dolcetti V, Radzina M, Bellini MI, Frezza F, Munir K, et al. Artificial Intelligence for Thyroid Nodule Characterization: Where Are We Standing? *Cancers*. 10 de julio de 2022;14(14):3357.

8. PÓSTER

EVALUACIÓN DE UN SOFTWARE DE APOYO AL DIAGNÓSTICO EN LA EVALUACIÓN DEL NÓDULO DE TIROIDES



AUTORA: LAURA RUIZ FERNÁNDEZ

TUTOR: GONZALO DÍAZ SOTO

INTRODUCCIÓN

La patología nodular del tiroides es uno de los problemas clínicos más prevalentes en las consultas de endocrinología. Recientemente, se han desarrollado sistemas de apoyo a la decisión (DSS) basados en inteligencia artificial (IA) para apoyar a los médicos en la interpretación de las imágenes ecográficas. Estos sistemas ayudan a reducir el componente subjetivo, mejorando del rendimiento diagnóstico de la ecografía de tiroides.

OBJETIVOS

1. Evaluar la utilidad clínica de un sistema de apoyo a la decisión (DSS) basado en inteligencia artificial (IA) en una cohorte de pacientes atendidos en una consulta de endocrinología general.
2. Analizar el grado de concordancia en las características ecográficas definidas por los observadores, el DSS y su evaluación en una consulta específica de Alta Resolución de Nódulo Tiroideo (UART).

MATERIAL Y MÉTODOS

Estudio prospectivo de todos aquellos pacientes a quienes se realizó una ecografía de cribado a criterio del endocrinólogo en una consulta no específica en el Hospital Clínico Universitario de Valladolid (HCUV) durante el año 2022.

Se analizaron datos clínicos y bioquímicos, así como las imágenes ortogonales de los nódulos en formato DICOM, sus características ecográficas y el criterio de derivación a la UART. Posteriormente, se analizaron las imágenes obtenidas mediante un sistema de IA DSS, así como los casos derivados a la UART.

RESULTADOS

CARACTERÍSTICAS	ENDOCRINO GENERAL vs IA		UART vs IA	
	Kappa	p	Kappa	p
Más ancho que largo	0,189	<0,05	0,181	<0,05
Composición	0,226	<0,05	0,352	<0,05
Ecogenicidad	0,107	P=0,08	0,511	<0,05
Márgenes	0,051	ns	1,000	ns
Focos ecogénicos	0,180	<0,05	0,405	<0,05
TIRADS	0,026	ns	0,123	<0,05
ATA	0,012	ns		

Tabla 1: Nivel de acuerdo en las características nodulares entre el endocrinólogo general y la UART con la IA.

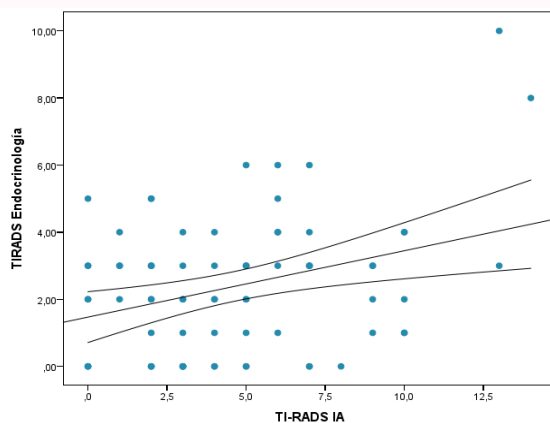


Figura 1: Correlación de Pearson para los Puntos ACR TI-RADS del Endocrinólogo general respecto a la IA.

CARACTERÍSTICAS	ENDOCRINO GENERAL vs UART	
	Kappa	p
Más ancho que largo	0,187	<0,05
Composición	0,495	<0,05
Ecogenicidad	0,637	<0,05
Márgenes	0,294	<0,05
Focos ecogénicos	0,472	<0,05
TIRADS	0,322	<0,05

Tabla 2: Nivel de acuerdo en las características nodulares entre el endocrinólogo general y la UART.

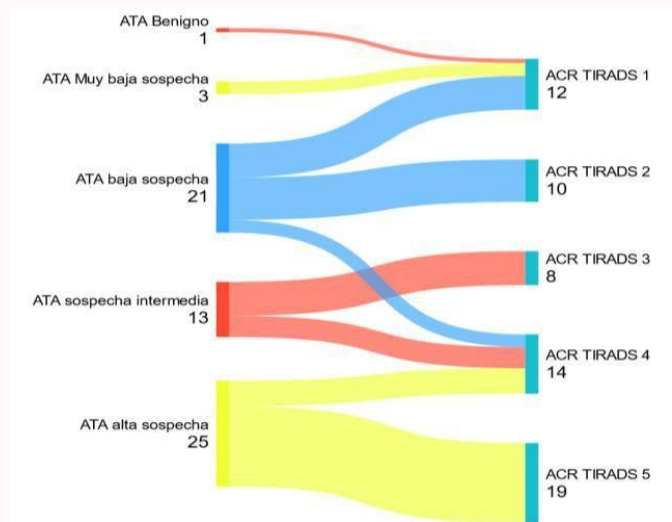


Figura 2: Clasificación de nódulos como ATA y ACR TI-RADS mediante el programa de IA.

CONCLUSIONES

- El programa de IA no redujo el número de derivaciones a la UART y mostró una escasa concordancia con la valoración realizada por los endocrinólogos no especialistas de la consulta de endocrinología general, en un contexto de patología nodular de bajo riesgo.
- Sin embargo, mostró un mejor rendimiento en aquellos nódulos de riesgo intermedio o alto derivados a la Unidad de Alta Resolución del Nódulo Tiroideo (UART).

