

Un plan reproducible para predecir el carcinoma folicular en ecografía tiroidea mediante análisis de datos topológicos y radiómica

David Llopis González¹, Artur Román Soler²

¹Institute of Diagnóstico per la Imatge. Duran i Reynals Hospital. Bellvitge

²Parc Taulí University Hospital

(Received: 28 March 2025. Accepted: 30 May 2025. Published online: 30 June 2025.)

Resumen

Antecedentes: El adenoma folicular y el carcinoma folicular presentan con frecuencia características ecográficas superpuestas, lo que limita el rendimiento discriminativo de los sistemas convencionales de estratificación de riesgo en ecografía tiroidea. Los métodos de análisis cuantitativo de imágenes, incluyendo la radiómica y el análisis de datos topológicos (TDA), ofrecen representaciones complementarias de la textura y la forma de la imagen que pueden mejorar la clasificación. **Métodos:** Este manuscrito reproduce el diseño metodológico reportado en el estudio de referencia [1]: selección retrospectiva de cohorte en un único centro (2010–2022), adenoma folicular versus carcinoma confirmados por anatomía patológica, una imagen ecográfica preoperatoria representativa por paciente, anotación manual de la región de interés, extracción de características radiómicas y de TDA, análisis de componentes principales (PCA) para compresión de alta dimensionalidad, y regresión logística regularizada con L2 evaluada mediante validación cruzada leave-one-out y *AUC*. **Resultados:** El modelo multimodal y un submodelo parsimonioso basado solo en TDA alcanzaron cada uno un *AUC* de 0.88 en validación cruzada leave-one-out, mientras que el mejor submodelo basado solo en radiómica alcanzó un *AUC* de 0.68. Se incluyen intervalos de confianza y composiciones de modelos en forma tabular para respaldar una implementación y validación transparentes. **Conclusión:** Este manuscrito sirve como un punto de partida completo para reproducir y ampliar la canalización de clasificación de neoplasias foliculares basada en radiómica+TDA en un conjunto de datos local. Preserva la lógica del estudio original, al tiempo que hace explícito, auditable y orientado a la implementación el flujo de trabajo.

Palabras clave: ecografía tiroidea; carcinoma folicular; adenoma folicular; radiómica; análisis de datos topológicos; homología persistente; regresión logística; reproducibilidad; aprendizaje automático

1. Introducción

Los nódulos tiroideos son frecuentes, y la ecografía (ultrasonido) sigue siendo la modalidad de imagen principal para la evaluación inicial [2,3]. Los marcos clínicos de estratificación del riesgo, como TI-RADS y los sistemas de patrones ecográficos de la ATA, han mejorado la estandarización [4,5]; sin embargo, las lesiones de patrón folicular siguen siendo diagnósticamente desafiantes porque los adenomas foliculares benignos y los carcinomas foliculares suelen compartir apariencias ecográficas similares [6,7]. La distinción final típicamente depende de la histopatología que demuestre invasión capsular y/o vascular después de la cirugía [6,7].

Esta limitación diagnóstica crea una oportunidad clínicamente importante para el análisis cuantitativo de imágenes. La radiómica proporciona descriptores de alta dimensionalidad derivados de textura e intensidad dentro de un nódulo anotado [8,9], mientras que el análisis de datos topológicos (TDA) captura explícitamente la estructura de forma multiescala (p. ej., componentes conexas y bucles) mediante homología persistente y resúmenes relacionados [10]. Debido a que estas modalidades enfatizan propiedades diferentes de la imagen, una estrategia de modelado combinada puede ser particularmente útil en conjuntos de datos pequeños y difíciles [8,10].

El presente manuscrito está escrito intencionalmente como una plantilla de artículo *orientada* que refleja el flujo de trabajo del estudio de prueba de concepto sobre neoplasias foliculares tiroideas. Está diseñado para ayudar a los investigadores a reproducir la misma metodología en sus propios datos, preservando la fidelidad metodológica, el reporte transparente y los detalles prácticos de implementación.

1.1 Objetivo y alcance del estudio

El objetivo de este trabajo no es afirmar un nuevo reanálisis a nivel de paciente sin acceso a las imágenes y anotaciones originales en bruto. Más bien, la meta es proporcionar un marco completo que:

1. replique la lógica de cohorte original y el flujo de trabajo de ingeniería de características,
2. reproduzca la taxonomía de modelos reportada y las tablas de resumen de desempeño,
3. formalice el procedimiento de evaluación estadística para un nuevo conjunto de datos, y
4. proporcione estructuras de figuras/tablas y lenguaje de reporte adecuados para adaptación directa.

2. Materiales y métodos

2.1 Diseño de la cohorte

El estudio de referencia [1] utilizó una cohorte retrospectiva de un único centro de adultos sometidos a cirugía tiroidea entre 2010 y 2022, con adenoma folicular o carcinoma folicular confirmados por anatomía patológica. Tras las exclusiones, la cohorte de imagen analizada comprendió 18 pacientes (7 adenomas foliculares y 11 carcinomas foliculares). Se seleccionó una imagen ecográfica tiroidea preoperatoria representativa por paciente tras revisión de imágenes, y el nódulo objetivo se anotó manualmente con una región de interés. Para replicar este diseño en un conjunto de datos local, la cohorte debe restringirse a casos quirúrgicos con adenoma folicular o carcinoma folicular confirmados por anatomía patológica, excluyendo categorías patológicas de confusión (como lesiones de células de Hürthle) si se desea una alineación estricta con el diseño de referencia. Las imágenes ecográficas tiroideas preoperatorias deben recuperarse de PACS o sistemas de archivo, debe seleccionarse una imagen representativa por paciente usando una regla predefinida y reproducible (o una selección revisada por expertos), el nódulo objetivo debe anotarse con una región de interés, y las imágenes deben reducirse de tamaño o remuestrearse a una resolución común para reducir la variabilidad en el número de píxeles entre casos.

2.2 Preprocesamiento de imágenes y manejo de la ROI

Para alinearse con la canalización, cada imagen ecográfica representativa en escala de grises debe procesarse de manera estandarizada para que la dimensionalidad y el manejo de intensidades sean consistentes antes de la extracción de características. En la práctica, esto significa convertir las imágenes a escala de grises si es necesario, normalizar el rango dinámico (por ejemplo, con escalado min–max a $[0, 1]$) y aplicar redimensionamiento o reducción de tamaño determinista a una resolución común en toda la cohorte. Cuando esté disponible, la metadata de espaciado de píxel debe conservarse para documentación y posibles análisis de armonización. La máscara de la región de interés debe aplicarse para la extracción de radiómica, mientras que tanto representaciones enmascaradas como basadas en contorno pueden conservarse para el cálculo de características de TDA, dependiendo de la estrategia de filtración. Todos los parámetros de preprocesamiento deben controlarse por versión y aplicarse de forma idéntica en los pliegues de entrenamiento y prueba para evitar fuga de datos y garantizar reproducibilidad.

2.3 Extracción de características de análisis de datos topológicos

El estudio de referencia [1] utilizó características de TDA basadas en homología persistente, con uso extensivo de filtraciones de altura sobre imágenes binarizadas generadas a múltiples escalas de suavizado, umbrales y direcciones. Específicamente, para cada imagen, se aplicó suavizado gaussiano en **3 anchos de banda**, seguido de binarización en **6 umbrales**. Luego se calcularon filtraciones de altura en **5 direcciones**, y se calculó homología persistente en dimensiones **0 y 1**. Este diseño produjo **180 diagramas de persistencia** por imagen, que se resumieron usando **entropía persistente**. Estos 180 valores formaron un bloque de características de TDA de alta dimensionalidad (denotado aquí como `TDA_height` antes de PCA). Además, se incluyeron dos características topológicas basadas en ALPS: `TDA_alps_boundary`, derivada de una filtración Vietoris–Rips sobre el contorno de la imagen como proxy de irregularidad del borde, y `TDA_alps_grey`, derivada de un resumen ALPS basado en filtración en escala de grises como proxy de estructura interna tipo cavidad o tendencia esponjiforme.

2.4 Extracción de características radiómicas

El estudio de referencia [1] extrajo **910 características radiómicas** por imagen utilizando un flujo de trabajo radiómico descrito previamente y posteriormente redujo este bloque de alta dimensionalidad mediante PCA. Además de la representación radiómica comprimida por PCA, se incluyeron directamente tres características radiómicas de menor dimensionalidad en la etapa de modelado: `Rad_mean` (brillo/luminosidad de la imagen), `Rad_variance` (un proxy relacionado con contraste) y `Rad_ratio` (relación de aspecto, que refleja comportamiento más ancho-que-alto frente a más alto-que-ancho). Para reproducibilidad, la implementación local debe documentar explícitamente el paquete de software y la versión (p. ej., PyRadiomics), los parámetros de discretización de imagen, la estrategia de normalización de intensidad, el espaciamiento de remuestreo, las clases de características habilitadas y el manejo de características fallidas, inestables o no computables.

2.5 Reducción de dimensionalidad

Para reducir el riesgo de sobreajuste en un entorno de muestra pequeña, el bloque de TDA de 180 dimensiones y el bloque radiómico de 910 dimensiones se proyectaron cada uno sobre sus dos primeros componentes principales. Esto produjo la representación bidimensional de TDA

$$\mathbf{z}_i^{(\text{tda})} = \text{PCA}_2(\mathbf{x}_i^{(\text{tda})}) = (\text{TDA_height_pca1}_i, \text{TDA_height_pca2}_i), \quad (1)$$

y la representación bidimensional de radiómica

$$\mathbf{z}_i^{(\text{rad})} = \text{PCA}_2(\mathbf{x}_i^{(\text{rad})}) = (\text{Rad_pca1}_i, \text{Rad_pca2}_i). \quad (2)$$

Nota importante: El ajuste de PCA debe realizarse *solo dentro de cada pliegue de entrenamiento*, y la transformación ajustada debe aplicarse después al caso retenido correspondiente para evitar fuga de datos.

2.6 Conjunto de predictores y familias de modelos

El estudio consideró un espacio de 9 predictores compuesto por `TDA_height_pca1`, `TDA_height_pca2`, `Rad_pca1`, `Rad_pca2`, `Rad_mean`, `Rad_variance`, `Rad_ratio`, `TDA_alps_grey` y `TDA_alps_boundary`. Se evaluaron seis familias de modelos con nombre: **Todos los predictores**, **Solo radiómica**, **Solo TDA**, **Mejor multimodal** (seleccionado mediante búsqueda exhaustiva de submodelos), **Mejor radiómica** (mejor submodelo solo radiómica) y **Mejor TDA** (mejor submodelo solo TDA). La selección de modelos se realizó mediante una búsqueda exhaustiva en todos los submodelos no vacíos, es decir, $2^9 - 1 = 511$ subconjuntos posibles, usando un parámetro de regularización por defecto de $C = 1$.

2.7 Clasificador y validación cruzada

El clasificador es un modelo de regresión logística regularizada con L2. Para el paciente i con vector de características \mathbf{x}_i , la probabilidad de carcinoma folicular se modela como

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \pi_i = \sigma(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}), \quad \sigma(t) = \frac{1}{1 + e^{-t}}, \quad (3)$$

con estimación de parámetros bajo una penalización L2 dada por

$$\hat{\boldsymbol{\beta}} = \arg \min_{\beta_0, \boldsymbol{\beta}} \left[- \sum_{i=1}^n \{y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)\} + \lambda \|\boldsymbol{\beta}\|_2^2 \right]. \quad (4)$$

El desempeño se estima mediante validación cruzada leave-one-out, en la que se deja fuera un paciente cada vez, los pasos de preprocesamiento (incluido PCA) y el modelo se ajustan con los $n - 1$ pacientes restantes, y se genera una probabilidad predicha para el paciente retenido. Este proceso se repite para todos los pacientes, tras lo cual se calculan curvas ROC y valores de *AUC* a partir de las probabilidades predichas fuera de pliegue. La sensibilidad y especificidad se reportan con un umbral de 0.5.

2.8 Cuantificación de la incertidumbre

El flujo de trabajo utilizó dos enfoques complementarios para la cuantificación de la incertidumbre. Primero, se utilizó bootstrap por submuestreo para obtener intervalos marginales generando 1,000 submuestras aleatorias de tamaño $b = 10$, con composición balanceada por clases (4 adenomas y 6 carcinomas), sin observaciones duplicadas dentro de cada submuestra, y recomputación de métricas de desempeño para estimación empírica de intervalos. Segundo, se calcularon intervalos de confianza asintóticos y comparaciones de modelos para AUC usando el método de DeLong.

2.9 Canalización reproducible paso a paso

Para replicar el artículo en un conjunto de datos local, el flujo de trabajo debe proceder como una canalización totalmente estandarizada que comienza con la curación de la cohorte, en la que se identifican casos con adenoma folicular y carcinoma folicular confirmados por anatomía patológica y se definen criterios de inclusión/exclusión para coincidir lo más estrechamente posible con el estudio de referencia [1]. Luego, las imágenes ecográficas tiroideas preoperatorias deben recuperarse del PACS, y debe seleccionarse una imagen representativa por paciente usando una regla predefinida (por ejemplo, la imagen que muestre la mayor sección transversal visible de la lesión). La lesión objetivo debe anotarse manualmente o bajo supervisión experta para crear una máscara binaria de región de interés, tras lo cual las imágenes deben estandarizarse mediante redimensionamiento o reducción de tamaño a una resolución común y normalización de intensidad. A continuación, deben extraerse características radiómicas de alta dimensionalidad (idealmente coincidiendo con las familias de características y el conteo aproximado utilizados en el estudio de referencia [1]), seguidas de extracción de características de TDA usando variantes binarias suavizadas con gaussiana y umbralizadas, homología persistente en dimensiones 0 y 1 en cinco direcciones, resúmenes de entropía persistente y características derivadas de ALPS. Dentro de cada pliegue de entrenamiento, debe aplicarse PCA por pliegue para reducir los bloques de radiómica y TDA a sus dos primeros componentes principales. Después, deben ajustarse modelos de regresión logística regularizada con L2 para las seis familias de modelos predefinidas, con búsqueda exhaustiva opcional de submodelos para identificar los mejores subconjuntos multimodales y específicos por modalidad. La validación debe realizarse usando validación cruzada leave-one-out para obtener probabilidades predichas fuera de pliegue, a partir de las cuales se calculan curvas ROC, AUC , sensibilidad y especificidad. Finalmente, la incertidumbre debe cuantificarse usando intervalos bootstrap e intervalos de AUC de DeLong, y las tablas y figuras resultantes deben completarse y compararse con los valores de referencia reportados en el estudio de referencia [1].

3. Resultados

Esta sección reporta los valores de referencia y las estructuras de modelos documentados en el estudio de prueba de concepto de referencia, y los presenta para comparación directa con futuros resultados en conjuntos de datos locales. Debido a que no se incluyen las imágenes crudas a nivel de paciente, las máscaras de región de interés ni las probabilidades predichas fuera de pliegue, los resultados numéricos que se muestran a continuación se tratan como objetivos de referencia para la alineación metodológica, y no como un reanálisis independiente. Por lo tanto, las tablas están destinadas a cumplir dos propósitos: primero, hacer explícito y reproducible el diseño de modelado original; y segundo, proporcionar un marco estructurado en el que puedan insertarse resultados generados localmente una vez que se haya ejecutado la canalización completa.

3.1 Resumen de la selección de la cohorte

La Tabla 1 reproduce la vía de depuración (attrition) reportada y el tamaño final de la cohorte de imagen analizada utilizada en el estudio de prueba de concepto de referencia. La cohorte se derivó de una población más amplia de cirugía tiroidea y se restringió a casos con adenoma folicular y carcinoma folicular confirmados por anatomía patológica, seguida de exclusiones basadas en patología e imagen. Una fuente importante de depuración fue la ausencia de imágenes ecográficas tiroideas preoperatorias disponibles, lo que redujo sustancialmente el tamaño de muestra analizable.

La cohorte final de imagen contenía 18 pacientes, incluidos 7 adenomas foliculares y 11 carcinomas foliculares. Esta distribución por clases refleja un desequilibrio modesto hacia carcinoma en la muestra analizada. En una cohorte tan pequeña, una o dos clasificaciones erróneas pueden alterar materialmente la sensibilidad, la especificidad y el área bajo la curva ROC (AUC), lo que subraya la importancia de reportar intervalos de incertidumbre y adoptar una estrategia de validación resistente a fuga de datos. En consecuencia, el flujo de trabajo de referencia utilizó validación cruzada leave-one-out (LOOCV) y múltiples formas de estimación de intervalos de confianza, como se resume más adelante en esta sección.

Table 1: Depuración de la cohorte y cohorte final de imagen analizada

Etapa de la cohorte	Conteo
Cohorte adulta de cirugía tiroidea (2010–2022)	908
Casos con adenoma/carcinoma folicular revisados por anatomía patológica	82
Excluidos: cáncer tiroideo previo	5
Excluidos: múltiples diagnósticos de cáncer en la anatomía patológica final	2
Excluidos: adenoma/carcinoma de células de Hürthle	7
Excluidos: sin imágenes ecográficas tiroideas preoperatorias disponibles	50
Cohorte final de imagen analizada	18
Adenoma folicular	7
Carcinoma folicular	11

Esta tabla resume la cohorte reportada utilizada en el estudio de referencia de prueba de concepto y debe reemplazarse por los conteos de la cohorte local al replicar en un nuevo conjunto de datos.

3.2 Inventario de características

La Tabla 2 resume los nueve predictores que finalmente se utilizaron para construir las familias de modelos de referencia tras la reducción de dimensionalidad y la ingeniería de características. El inventario de características es intencionalmente multimodal. Dos predictores se derivan de un bloque de características TDA de filtración de altura de alta dimensionalidad (comprimido mediante PCA), dos predictores se derivan de un bloque de características radiómicas de alta dimensionalidad (también comprimido mediante PCA), tres predictores son descriptores resumen radiómicos directamente interpretables y dos predictores representan resúmenes topológicos basados en ALPS.

Esta estructura es clínica y metodológicamente relevante. Las características radiómicas y de TDA derivadas de PCA proporcionan representaciones compactas de información compleja de imagen que, de otro modo, serían demasiado de alta dimensionalidad para un modelado estable en una cohorte muy pequeña. En paralelo, características directamente interpretables como `Rad_mean`, `Rad_variance` y `Rad_ratio` preservan descriptores clínicamente intuitivos (brillo, contraste y relación de aspecto relacionada con la morfología). Las características de TDA basadas en ALPS enriquecen aún más el espacio de características con proxies de topología del borde e interna que pueden capturar patrones morfológicos no bien representados por estadísticas convencionales basadas en píxeles.

3.3 Definiciones de modelos y submodelos seleccionados

La Tabla 3 reporta la composición de predictores de las seis familias de modelos de referencia. El modelo **Todos los predictores** incluye el espacio completo multimodal de 9 predictores, mientras que los modelos **Solo radiómica** y **Solo TDA** restringen el clasificador a subconjuntos específicos por modalidad. Los tres modelos restantes (**Mejor multimodal**, **Mejor radiómica** y **Mejor TDA**) fueron seleccionados a partir de una búsqueda exhaustiva sobre subconjuntos no vacíos de predictores bajo la estrategia de selección de referencia.

Varios patrones son dignos de mención. El modelo **Mejor multimodal** es relativamente parsimonioso e incluye tanto información topológica como derivada de radiómica (`TDA_height_pca1`, `TDA_height_pca2`, `Rad_mean` y `TDA_alps_boundary`), lo que sugiere que la solución multimodal de mejor desempeño no requirió el bloque radiómico completo. El modelo **Mejor TDA** es aún más compacto, al apoyarse solo en `TDA_height_pca2` y `TDA_alps_boundary`, lo que indica que un pequeño número de descriptores topológicos puede capturar una fracción

Table 2: Predictores del modelo y su derivación

Nombre del predictor	Derivado de	Características antes de PCA	Posible característica radiográfica correspondiente
TDA_height_pca1	TDA	180	Forma general del nódulo
TDA_height_pca2	TDA	180	Forma general del nódulo
Rad_pca1	Radiómica	910	Contraste complejo de imagen y correlación de píxeles
Rad_pca2	Radiómica	910	Contraste complejo de imagen y correlación de píxeles
Rad_mean	Radiómica	NA	Brillo/luminosidad de la imagen
Rad_variance	Radiómica	NA	Contraste de la imagen
Rad_ratio	Radiómica	NA	Más ancho que alto o más alto que ancho
TDA_alps_grey	TDA	NA	Naturaleza espongiforme (o ausencia de ella) del nódulo
TDA_alps_boundary	TDA	NA	Irregularidad de los bordes

TDA = análisis de datos topológicos; PCA = análisis de componentes principales.

sustancial de la señal discriminativa en este conjunto de datos. En contraste, el modelo **Mejor radiómica** retiene únicamente **Rad_ratio**, destacando la posibilidad de que gran parte de la información radiómica de alta dimensionalidad fuese redundante o inestable en el tamaño muestral considerado.

Table 3: Predictores incluidos en cada uno de los seis modelos principales (1 = incluido, 0 = no incluido)

Predictor	Todos	Solo Rad	Solo TDA	Mejor Multimodal	Mejor Radiómica	Mejor TDA
TDA_height_pca1	1	0	1	1	0	0
TDA_height_pca2	1	0	1	1	0	1
Rad_pca1	1	1	0	0	0	0
Rad_pca2	1	1	0	0	0	0
Rad_mean	1	1	0	1	0	0
Rad_variance	1	1	0	0	0	0
Rad_ratio	1	1	0	0	1	0
TDA_alps_grey	1	0	1	0	0	0
TDA_alps_boundary	1	0	1	1	0	1

3.4 Métricas de desempeño e intervalos de confianza

Las Tablas 4 y 5 resumen el desempeño de los modelos de referencia en las seis familias de modelos utilizando tanto intervalos por bootstrap de submuestreo como intervalos asintóticos de DeLong. El uso de dos métodos de estimación de intervalos es particularmente útil en un entorno de muestra pequeña porque cada enfoque hace supuestos diferentes y ofrece una visión complementaria de la incertidumbre del modelo.

A nivel de estimación puntual, la discriminación más fuerte se logra con los modelos **Mejor Multimodal** y **Mejor TDA**, cada uno con un AUC de 0.88. El modelo **Solo TDA** también presenta buen desempeño ($AUC = 0.82$), mientras que el modelo **Solo radiómica** está cerca del azar ($AUC = 0.49$). El submodelo **Mejor radiómica** mejora respecto al modelo completo solo-radiómica ($AUC = 0.68$), pero permanece por debajo de las alternativas topológicas y multimodales. Este patrón es consistente con la hipótesis de que los descriptores derivados de TDA codifican información discriminativa de forma/estructura que no es capturada eficazmente por el conjunto de características radiómicas por sí solo en la cohorte de referencia.

Los valores de sensibilidad son generalmente altos en la mayoría de los modelos (particularmente en los modelos que contienen TDA), mientras que la especificidad muestra mayor variación. Esto sugiere que los modelos de referencia fueron más eficaces para identificar casos de carcinoma que para clasificar correctamente como benignos los casos de adenoma al umbral fijo de 0.5. En la práctica, este comportamiento del punto operativo puede ser aceptable o incluso deseable en un entorno temprano de cribado de prueba de concepto, pero el ajuste de umbral y la calibración serían importantes en una futura traslación clínica. Los intervalos de confianza relativamente amplios para algunas métricas, especialmente la especificidad, enfatizan además la

incertidumbre asociada con una cohorte retrospectiva pequeña.

Table 4: Intervalos de confianza marginales por bootstrap de submuestreo al 95% para AUC , sensibilidad y especificidad

Métrica	Todos los predictores	Solo radiómica	Solo TDA	Mejor multimodal	Mejor radiómica	Mejor TDA
AUC	0.74 [0.66, 1.00]	0.49 [0.33, 0.71]	0.82 [0.74, 1.00]	0.88 [0.85, 1.00]	0.68 [0.61, 0.84]	0.88 [0.85, 1.00]
Sensibilidad	0.91 [0.86, 1.00]	0.73 [0.60, 0.84]	0.91 [0.89, 1.00]	0.91 [0.87, 1.00]	0.91 [0.88, 0.98]	0.91 [0.86, 1.00]
Especificidad	0.43 [0.16, 0.79]	0.43 [0.27, 0.64]	0.71 [0.54, 1.00]	0.57 [0.50, 0.70]	0.43 [0.22, 0.71]	0.71 [0.62, 0.87]

Umbral para sensibilidad/especificidad = 0.5 en el estudio de referencia [1].

Table 5: Intervalos de confianza asintóticos (DeLong) al 95% para AUC

Métrica	Todos los predictores	Solo radiómica	Solo TDA	Mejor multimodal	Mejor radiómica	Mejor TDA
AUC	0.74 [0.49, 1.00]	0.49 [0.12, 0.76]	0.82 [0.59, 1.00]	0.88 [0.72, 1.00]	0.68 [0.38, 0.97]	0.88 [0.71, 1.00]

3.5 Interpretación comparativa del comportamiento de los modelos

En conjunto, los resultados de referencia sugieren tres observaciones principales. Primero, los modelos que contienen TDA superan a las alternativas de solo radiómica en esta cohorte de prueba de concepto, tanto en la comparación completa restringida por modalidad (**Solo TDA** vs **Solo radiómica**) como en la comparación de submodelos seleccionados (**Mejor TDA** vs **Mejor radiómica**). Segundo, añadir una pequeña cantidad de información radiómica a un modelo predominantemente basado en TDA (**Mejor multimodal**) preserva una alta discriminación, pero no supera de forma evidente al mejor subconjunto parsimonioso de solo TDA en términos de estimación puntual de referencia. Tercero, incrementar la complejidad del modelo incluyendo los nueve predictores no produce el mayor desempeño, lo cual es consistente con lo esperado en una muestra muy pequeña, donde predictores irrelevantes o ruidosos pueden degradar la generalización pese a la regularización.

Estos hallazgos deben interpretarse como generadores de hipótesis y no como definitivos. La cohorte de referencia es pequeña, de un solo centro y ensamblada retrospectivamente, y no se dispone de un conjunto de validación externa en el diseño de referencia. No obstante, el patrón de desempeño observado proporciona una justificación sólida para reproducir el flujo de trabajo en cohortes más grandes y evaluar si la aparente señal topológica persiste entre instituciones, escáneres y estilos de adquisición ecográfica.

3.6 Cómo usar esta sección de resultados para su conjunto de datos

Cuando la canalización de replicación se ejecute en un conjunto de datos local, esta sección puede actualizarse de manera mínimamente disruptiva. Específicamente, la Tabla 1 debe reemplazarse con los conteos de la cohorte local y las razones de exclusión, la Tabla 3 debe reflejar los subconjuntos realmente seleccionados si se repite la búsqueda de submodelos, y las Tablas 4 y 5 deben completarse con las nuevas métricas puntuales e intervalos estimados. Los párrafos interpretativos deben revisarse entonces para reflejar si los resultados locales confirman, debilitan o contradicen el patrón de referencia de desempeño dominado por TDA.

La Figura 1 proporciona un resumen visual compacto del flujo de trabajo de referencia de extremo a extremo, desde la identificación retrospectiva de la cohorte hasta la extracción de características, reducción de dimensionalidad, ajuste del modelo, validación cruzada y estimación de incertidumbre. Este esquema es útil tanto para los lectores como para la documentación interna de reproducibilidad porque hace explícito el orden de operaciones y la estructura de ramificación multimodal (radiómica y TDA) antes de la fusión de características. En una implementación local, esta figura puede conservarse tal cual si el flujo de trabajo no cambia, o modificarse ligeramente para reflejar preprocesamiento específico del conjunto de datos, procedimientos de anotación o extensiones de validación.

La Figura 2 visualiza las estimaciones puntuales de referencia de AUC , sensibilidad y especificidad en las seis familias de modelos. El gráfico de barras proporciona una vista cualitativa inmediata de los compromisos (trade-offs) que, de otro modo, se distribuyen entre las Tablas 4 y 5. En particular, resalta el mejor desempeño discriminativo de los modelos que contienen TDA y la especificidad relativamente baja de varios modelos al

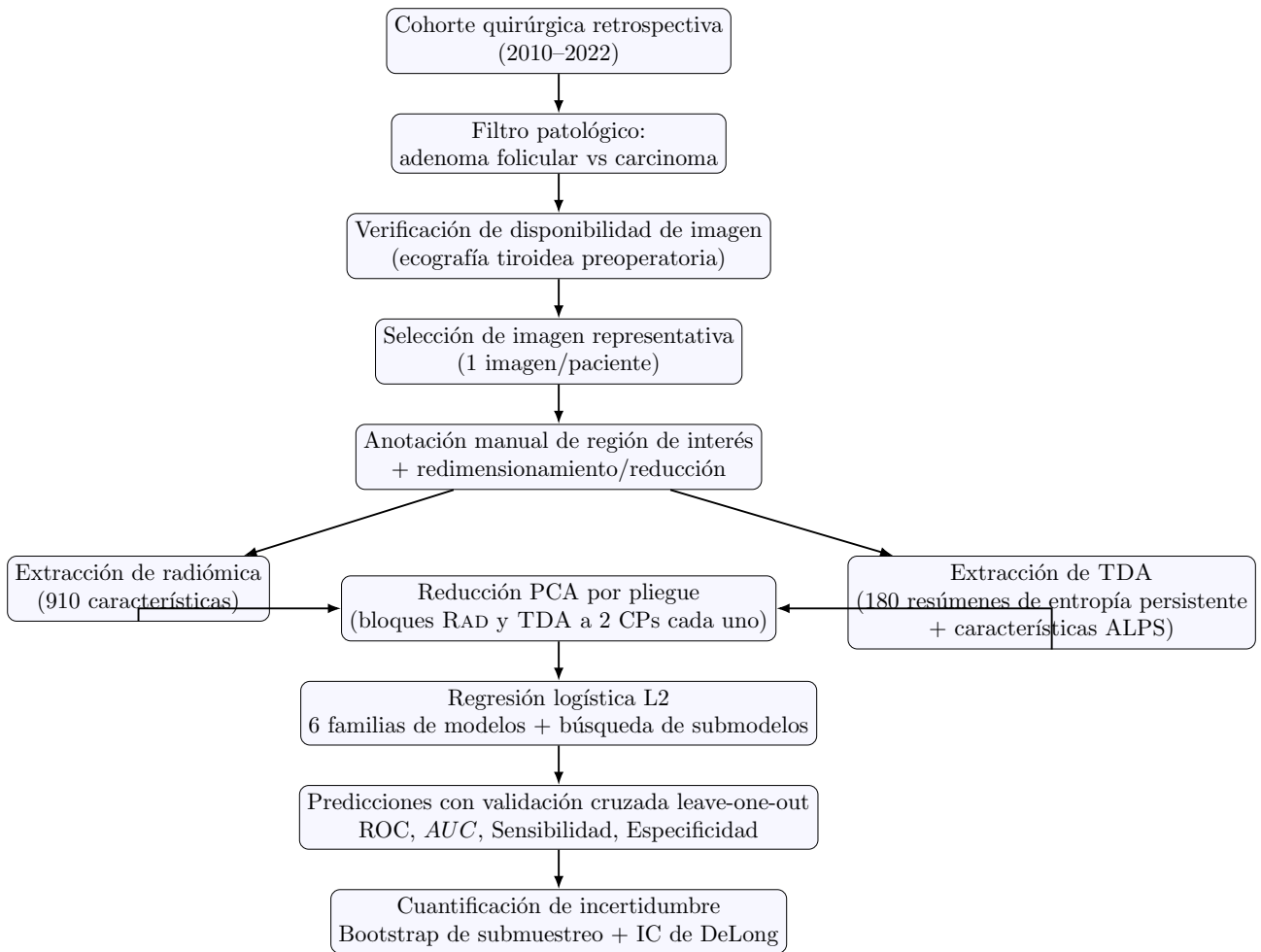


Figure 1: Canalización reproducible que refleja el flujo de trabajo de prueba de concepto para la predicción de carcinoma folicular a partir de ecografía tiroidea usando radiómica y análisis de datos topológicos.

umbral operativo fijo. Esta figura puede conservarse como panel de comparación de referencia y emparejarse con un segundo panel que muestre resultados locales una vez que esos valores estén disponibles.

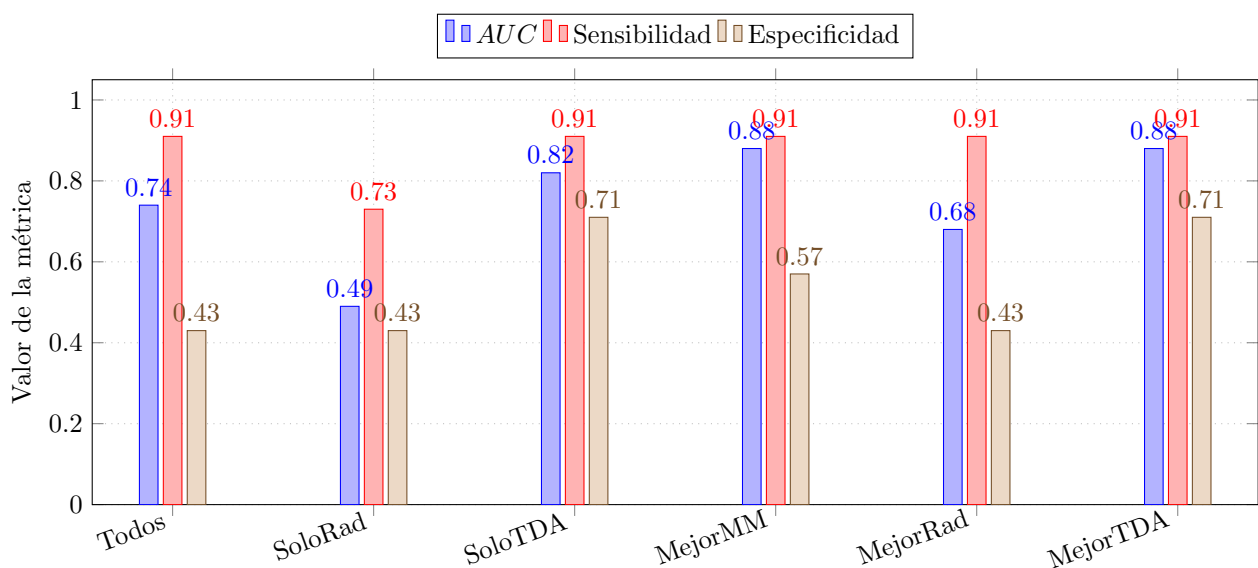


Figure 2: Estimaciones puntuales de referencia reportadas (*AUC*, sensibilidad, especificidad). Los intervalos de confianza se proporcionan en las Tablas 4 y 5.

El estudio contiene paneles para salidas dependientes del conjunto de datos que requieren imágenes crudas y

probabilidades predichas. Estos paneles se conservan intencionalmente en la plantilla porque corresponden a resúmenes visuales altamente informativos dentro del flujo de trabajo de referencia, incluyendo curvas ROC, barridos de regularización, ejemplos de casos representativos y visualizaciones de intervalos de confianza bootstrap.

Para maximizar la reproducibilidad, las figuras deben generarse a partir de artefactos exportados por código (p. ej., tablas CSV de predicciones fuera de pliegue, remuestreos bootstrap y trazas de ajuste) en lugar de valores curados manualmente. Esto ayuda a garantizar que los gráficos del manuscrito permanezcan sincronizados con las métricas finales reportadas y puedan regenerarse si cambian la definición de la cohorte o la canalización de preprocesamiento.

4. Discusión

Este manuscrito subraya varias fortalezas metodológicas del enfoque de prueba de concepto [1] y, al mismo tiempo, aclara las condiciones bajo las cuales el flujo de trabajo puede reproducirse válidamente en un nuevo conjunto de datos [12, 13]. Una fortaleza central del diseño de referencia es su énfasis deliberado en la interpretabilidad y la moderación metodológica [14]. En contraste con enfoques de aprendizaje profundo de alta capacidad que pueden ser difíciles de auditar en cohortes pequeñas [15], el uso de regresión logística regularizada con L2 proporciona un marco de modelado transparente en el que las contribuciones de las características, la composición del modelo y los compromisos de desempeño pueden examinarse directamente [16]. Esto es particularmente apropiado en el contexto actual, donde la cohorte de imagen disponible es pequeña y el objetivo científico principal es probar si los descriptores diseñados de radiómica y topología contienen señal discriminativa, más que maximizar el desempeño únicamente mediante complejidad del modelo [8, 17]. La comparación explícita entre variantes con todos los predictores, específicas por modalidad y seleccionadas por submodelos fortalece además el diseño al permitir una evaluación clara y clínicamente interpretable de si la radiómica, el TDA o su combinación contribuyen más a la discriminación [9, 10].

Una segunda contribución importante del flujo de trabajo de referencia es la inclusión de características de TDA basadas en homología persistente junto con radiómica convencional [11, 12]. Desde una perspectiva de imagen clínica, este diseño multimodal de características es convincente porque se alinea con los criterios visuales que los radiólogos suelen utilizar al evaluar nódulos tiroideos, incluyendo regularidad de la forma, características del borde, arquitectura interna y organización estructural global [13]. La homología persistente y resúmenes relacionados pueden codificar componentes conexas, estructuras tipo bucle y cambios de forma multiescala que son difíciles de representar únicamente mediante estadísticas estándar de intensidad o matrices de textura [14, 15]. El patrón de referencia observado, en el que los modelos que contienen TDA superan a los modelos de solo radiómica, no es por tanto solo un resultado empírico, sino también conceptualmente plausible [1, 16]. Sugiere que los descriptores topológicos pueden capturar información sensible a la morfología relevante para la discriminación de neoplasias foliculares, especialmente en dominios de imagen donde los descriptores ecográficos convencionales se superponen sustancialmente entre lesiones benignas y malignas [17].

Una tercera fortaleza importante es el cuidado puesto en la estimación del desempeño y la cuantificación de la incertidumbre a pesar de las restricciones de una cohorte retrospectiva muy pequeña [1, 18]. El uso de validación cruzada leave-one-out para generar probabilidades fuera de pliegue para estimación de ROC y *AUC* es apropiado en escenarios con datos limitados porque maximiza el uso de datos de entrenamiento en cada pliegue, al tiempo que preserva la separación a nivel de paciente durante la evaluación [19]. Del mismo modo, el uso de intervalos bootstrap por submuestreo (con muestreo balanceado por clases y sin duplicados dentro de las submuestras) proporciona una perspectiva adicional, motivada de manera práctica, sobre la incertidumbre que complementa los intervalos asintóticos de DeLong para *AUC* [20]. En conjunto, estas decisiones reducen el riesgo de reportes excesivamente optimistas y mejoran la credibilidad metodológica de un análisis de prueba de concepto [3, 18, 20]. Al mismo tiempo, el tamaño muestral sigue siendo la limitación dominante del estudio de referencia [1], y los intervalos de confianza reportados deben interpretarse como estimaciones amplias y potencialmente inestables, más que como límites de desempeño precisos [19, 20].

Al reproducir esta canalización en un conjunto de datos local, varias cuestiones de implementación requieren atención particular porque pueden afectar materialmente tanto la reproducibilidad como la validez [2–4]. La más crítica es la fuga de datos: operaciones de preprocesamiento como normalización y PCA deben ajustarse

exclusivamente dentro de cada pliegue de entrenamiento y luego aplicarse al caso retenido, en lugar de calcularse globalmente sobre el conjunto de datos completo [6, 18]. No imponer un preprocesamiento por pliegue puede inflar el desempeño e invalidar comparaciones [19]. La selección de imágenes es otra fuente importante de variabilidad potencial [13, 17]. Si se elige una imagen ecográfica “representativa” por paciente sin una regla determinista, puede introducirse subjetividad oculta en la canalización y debilitar la reproducibilidad [2]. Esto debe abordarse documentando explícitamente la regla de selección y, cuando sea factible, midiendo la concordancia interobservador e intraobservador [3, 13]. La variabilidad en la anotación de la región de interés es igualmente importante, dado que tanto la radiómica como las características topológicas sensibles al borde pueden ser sensibles a la colocación del contorno, particularmente en nódulos pequeños o heterogéneos [8, 11, 14]. En conjuntos de datos multicéntricos o multiescáner, la inestabilidad de características inducida por diferencias de adquisición y heterogeneidad del preprocesamiento puede reducir aún más la transportabilidad; en consecuencia, se recomiendan análisis de armonización y verificaciones de sensibilidad [7, 9, 17]. Finalmente, la búsqueda exhaustiva de submodelos, aunque informativa en un estudio de referencia, puede sobreajustar en muestras pequeñas, y las implementaciones futuras idealmente deberían incorporar validación anidada o validación externa para reducir el sesgo de selección cuando el tamaño muestral lo permita [5, 6, 19].

Los hallazgos de referencia también sugieren varias direcciones científicamente significativas para extensión en un estudio plenamente novedoso usando datos locales [1, 2]. La validación externa o multicéntrica sería el paso siguiente más importante porque prueba directamente si la señal aparentemente asociada a TDA se generaliza más allá del entorno original de adquisición de un solo centro [10, 16]. La evaluación comparativa frente a marcos establecidos de estratificación de riesgo por ecografía (como TI-RADS o los sistemas ecográficos de la ATA) mejoraría la relevancia clínica al situar el modelo en el contexto de flujos de decisión existentes [13, 17]. El análisis de calibración, incluyendo Brier score y curvas de calibración, complementaría las métricas de discriminación y aclararía si las probabilidades predichas son adecuadas para comunicación de riesgo o umbrales de triaje [18, 19]. La optimización de umbral para casos de uso clínico específicos (p. ej., cribado de alta sensibilidad frente a apoyo diagnóstico balanceado) fortalecería aún más el valor traslacional [20]. Extensiones metodológicas adicionales, como estudios de ablación sobre elecciones de filtración, resúmenes persistentes y variantes ALPS, podrían identificar qué decisiones de diseño topológico son realmente esenciales [11, 12, 14, 15]. También sería valioso comparar este marco de referencia interpretable con líneas base no lineales más potentes (p. ej., random forest, gradient boosting o redes neuronales superficiales cuidadosamente regularizadas), manteniendo el mismo protocolo de preprocesamiento por pliegue y evaluación para una comparación justa [4–6].

En conjunto, la prueba de concepto de referencia debe interpretarse como generadora de hipótesis, pero metodológicamente informativa [1, 3]. La contribución más fuerte del flujo de trabajo no son únicamente las estimaciones puntuales reportadas, sino la demostración de que una canalización de radiómica+TDA cuidadosamente construida puede implementarse en un entorno altamente restringido de cohorte pequeña con definiciones de modelo transparentes y reporte explícito de incertidumbre [7, 8, 11, 18]. Este manuscrito amplía ese valor al hacer explícita la lógica computacional y de reporte en una forma que puede reutilizarse, auditarse y mejorarse directamente para conjuntos de datos más grandes [2–4, 10, 16].

5. Conclusión

Presentamos una plantilla completa que refleja un flujo de trabajo de prueba de concepto para predecir carcinoma folicular en ecografía tiroidea usando radiómica y análisis de datos topológicos. El manuscrito formaliza la curación de la cohorte, el manejo de imágenes, la ingeniería de características, la compresión basada en PCA, la regresión logística regularizada con L2, la validación cruzada leave-one-out y la cuantificación de la incertidumbre, al tiempo que reproduce definiciones de modelos de referencia y tablas de resumen de desempeño en una estructura lista para replicación. Más allá de servir como un resumen narrativo del método, esta plantilla está diseñada como un andamiaje práctico de implementación que puede completarse con conteos de cohorte locales, salidas de características, resultados de modelos y figuras después de ejecutar la canalización en un nuevo conjunto de datos. De esta manera, respalda tanto la fidelidad metodológica al estudio de referencia como una extensión transparente hacia análisis más grandes, mejor validados y clínicamente más informativos.

Declaración de disponibilidad de datos y código

Para apoyar la reproducibilidad, el código utilizado para implementar y evaluar la canalización debe publicarse en un repositorio con control de versiones (p. ej., GitHub o GitLab) con especificación explícita del entorno (como `requirements.txt` o `environment.yml`), scripts de preprocesamiento, archivos de configuración de extracción de características, scripts de entrenamiento y validación del modelo, y cuadernos o scripts de generación de figuras. Los artefactos intermedios del análisis (por ejemplo, asignaciones de pliegues, probabilidades predichas fuera de pliegue, remuestreos bootstrap y resúmenes tabulados de métricas) también deben guardarse cuando sea posible para que las tablas y figuras del manuscrito puedan regenerarse sin transcripción manual. Los datos de imagen desidentificados y las anotaciones de región de interés deben compartirse de acuerdo con las políticas del comité de ética institucional, los acuerdos de uso de datos y las regulaciones de privacidad aplicables; cuando no sea posible compartir los datos directamente, debe proporcionarse una descripción clara de las restricciones de acceso y un ejemplo mínimo reproducible sintético o de metadatos para facilitar una implementación independiente.

Agradecimientos

Los autores agradecen a los equipos de radiología, cirugía endocrina, anatomía patológica y ciencia de datos que contribuyeron a la identificación de casos, recuperación de imágenes, anotación y revisión estadística.

Referencias

- [1] Thomas, Andrew M., et al. "A proof-of-concept investigation into predicting follicular carcinoma on ultrasound using topological data analysis and radiomics." *Imaging* 17.1 (2025): 39-48.
- [2] Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teehey SA, et al.: ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* 2017; 14(5): 587–95.
- [3] 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. *Thyroid* 2015; 26(1): 1–133.
- [4] Chng CL, Tan HC, Too CW, Lim WY, Chiam PPS, Zhu L, et al.: Diagnostic performance of ATA, BTA and TIRADS sonographic patterns in the prediction of malignancy in histologically proven thyroid nodules. *Singap Med J* 2018; 59(11): 578–83.
- [5] Kim J-S, Kim BG, Stybayeva G, Hwang SH: Diagnostic performance of various ultrasound risk stratification systems for benign and malignant thyroid nodules: a meta-analysis. *Cancers* 2023; 15(424): 1–13.
- [6] Huang B, Ebner SA, Jasnit S M, Bentley-Hibbert S, McConnell RJ, Lee JA, et al.: A multidisciplinary head-to-head comparison of American college of radiology thyroid imaging and reporting data system and American thyroid association ultrasound risk stratification systems. *The Oncologist* 2020; 25: 398–403.
- [7] Kuo T-C, Wu M-H, Chen K-Y, Hsieh M-S, Chen A, Chen C-N: Ultrasonographic features for differentiating follicular thyroid carcinoma and follicular adenoma. *Asian J Surg* 2020; 43(1): 339–46.
- [8] Castellana M, Piccardo A, Virili C, Scappaticcio L, Grani G, Durante C, et al.: Can ultrasound systems for risk stratification of thyroid nodules identify follicular carcinoma? *Cancer Cytopathol* 2020; 128(4): 250–9.
- [9] Lin J-D, Chao T-C: Follicular thyroid carcinoma: from diagnosis to treatment. *Endocr J* 2006; 53(4): 441–8.
- [10] Louis JDS, Leight GS, Tyler DS: Follicular neoplasms: the role for observation, fine needle aspiration biopsy, thyroid suppression, and surgery. *Semin Surg Oncol* 1999; 16(1): 5–11.
- [11] Sillery JC, Reading CC, Charboneau JW, Henrichsen TL, Hay ID, Mandrekar JN: Thyroid follicular carcinoma: sonographic features of 50 cases. *Neuroradiol/Head Neck Imaging* 2010; 194: 44–54.
- [12] Gillies RJ, Kinahan PE, Hricak H: Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278(2): 563–77.
- [13] Edelsbrunner H, Harer JL: *Computational topology: An introduction*: American Mathematical Society; 2010.
- [14] Bukkuri A, Andor N, Darcy IK: Applications of topological data analysis in oncology. *Front Artif Intell* 2021; 13(4): eCollection 2021.
- [15] Munch E: A user's guide to topological data analysis. *J Learn Anal* 2017; 4(2): 47–61.
- [16] Liang J, Huang X, Hu H, Liu Y, Zhou Q, Cao Q, et al.: Predicting malignancy in thyroid nodules: radiomics score versus 2017 American college of radiology thyroid imaging, reporting and data system. *Thyroid* 2018; 28(8): 1024–33.
- [17] Shin I, Kim YJ, Han K, Lee E, Kim HJ, Shin JH, et al.: Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. *Ultrasonography* 2020; 39: 257–65.
- [18] Vandaele R, Mukherjee P, Selby HM, Shah RP, Gevaert O: Topological data analysis of thoracic radiographic images shows improved radiomics-based lung tumor histology prediction. *Patterns* 2023; 4(1): 100657.
- [19] Lin AC, Liu Z, Lee J, Ranvier GF, Taye A, Owen R, et al.: Generating a multimodal artificial intelligence model to differentiate benign and malignant follicular neoplasms of the thyroid: a proof-of-concept study. *Surgery* 2024; 175(1): 121–7.
- [20] Chung YM, Hu CS, Lawson A, Smyth C, editors. *Topological approaches to skin disease image analysis*. 2018 IEEE international conference on big data (big data); 2018 10–13 Dec. 2018.
- [21] Maurya A, Stanley RJ, Lama N, Nambisan AK, Patel G, Saeed D, et al.: Hybrid topological data analysis and deep learning for basal cell carcinoma diagnosis. *J Imaging Inform Med* 2024; 37(1): 92–106.