Radiómica e inteligencia artificial para la predicción preprocedimental de complicaciones en biopsias pulmonares guiadas por TC

María Cribillés-Pérez[1], Juan Luis Suárez-Díaz[1], Germán Maldonado-Vicente[2], José Luis Martín-Rodríguez[2,3] y Francisco Herrera[1]

[1] Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, España
mcribilles@ugr.es, jlsuarezdiaz@ugr.es, herrera@decsai.ugr.es
[2] Hospital Universitario Clínico San Cecilio de Granada, Granada, España
3agermanmaldonadovicente@gmail.com
[3] Instituto de Investigación Biosanitaria de Granada (ibs.GRANADA), Granada, España
joseluismartin.rx@hotmail.com

Resumen La biopsia pulmonar guiada por tomografía computarizada (TC) desempeña un papel fundamental en el diagnóstico y la caracterización de nódulos pulmonares. Sin embargo, este procedimiento conlleva un riesgo considerable de complicaciones. Aunque una parte importante de la investigación se ha centrado en diferenciar nódulos benignos y malignos, se ha prestado una atención limitada a la predicción de estos eventos adversos antes de realizar la biopsia. Este estudio aborda dicha limitación mediante la investigación de métodos de inteligencia artificial que combinan estudios de TC previos a la biopsia y variables clínicas para anticipar el riesgo de complicaciones. Se analizó un conjunto de datos de 125 pacientes del Hospital Universitario Clínico San Cecilio de Granada mediante enfoques clásicos basados en radiómica y modelos de aprendizaje profundo tridimensional. Aunque los modelos de aprendizaje profundo mostraron una capacidad de generalización limitada debido a la escasez de datos, el aprendizaje automático basado en características radiómicas ofreció un rendimiento más consistente e interpretable. Estos resultados constituyen un primer paso hacia el desarrollo de sistemas de inteligencia artificial explicables y fiables para la predicción preprocedimental del riesgo en radiología intervencionista.

Keywords: Biopsia pulmonar · Tomografía computarizada · Radiómica · Aprendizaje profundo · Inteligencia artificial

1. Introducción

El cáncer de pulmón continúa siendo una de las enfermedades más frecuentes y letales a escala mundial. En 2022, se registraron aproximadamente 2,5 millones de nuevos casos y 1,8 millones de muertes por esta enfermedad [4]. A pesar

de los avances en cribado y tratamiento, su tasa de mortalidad sigue siendo elevada, principalmente debido al diagnóstico tardío y al carácter invasivo de los procedimientos confirmatorios. El consumo de tabaco sigue siendo el principal factor de riesgo del cáncer de pulmón, mientras que las exposiciones ambientales y ocupacionales, como la contaminación atmosférica y la combustión de biomasa, también influyen en su incidencia mundial.

El cribado del cáncer de pulmón mediante tomografía computarizada de baja dosis permite detectar la enfermedad en fases tempranas y ha demostrado reducir la mortalidad en poblaciones de alto riesgo [1]. Aunque tradicionalmente este proceso se ha basado en la interpretación manual por parte de radiólogos, los avances en inteligencia artificial han impulsado sistemas capaces de detectar y caracterizar nódulos pulmonares con alta precisión, reduciendo falsos positivos y variabilidad entre observadores. En conjunto, estos resultados evidencian el potencial de la inteligencia artificial para mejorar la eficiencia, consistencia y escalabilidad del cribado en la práctica clínica [3].

La implementación de estos programas de cribado incrementará previsiblemente la detección de nódulos indeterminados y, con ello, el número de biopsias transtorácicas guiadas por TC. Este escenario refuerza la necesidad de herramientas que permitan identificar de forma anticipada a los pacientes con mayor riesgo de complicaciones.

La biopsia percutánea transtorácica con aguja, habitualmente guiada por TC, incluye la aspiración con aguja fina y la biopsia con aguja gruesa. Esta última se ha impuesto progresivamente por su mayor rendimiento diagnóstico y simplicidad técnica, aunque presenta más complicaciones. Pese a ser mínimamente invasiva, el procedimiento conlleva un riesgo relevante, con tasas globales de complicaciones del 39 % para la biopsia con aguja gruesa y del 24 % para la aspiración con aguja fina [8]. Las complicaciones más frecuentes son el neumotórax y la hemorragia pulmonar, lo que justifica la necesidad de una estimación precisa del riesgo preoperatorio.

Dado el éxito demostrado de los métodos de visión por computador e inteligencia artificial en tareas de imagen médica, como la detección de lesiones, la segmentación y el apoyo al diagnóstico, resulta razonable explorar su potencial para predecir complicaciones en procedimientos intervencionistas. Nuestra hipótesis es que la información extraída de estudios de TC previos a la biopsia, combinada con datos clínicos relevantes, podría utilizarse para identificar pacientes con mayor riesgo de sufrir eventos adversos. Un sistema automatizado de evaluación del riesgo de este tipo podría ayudar a los clínicos a optimizar la selección de pacientes, mejorar la seguridad del procedimiento y asignar los recursos sanitarios de forma más eficiente.

En este trabajo se presenta un estudio preliminar sobre la aplicación de técnicas de inteligencia artificial a la predicción de complicaciones en biopsias pulmonares guiadas por TC. El objetivo es evaluar si la información extraída de estudios de TC previos a la biopsia, junto con variables clínicas relevantes, puede utilizarse para construir un sistema automatizado capaz de identificar pacientes con mayor riesgo antes del procedimiento. Aunque algunos estudios han inves-

tigado modelos estadísticos o clínicos de riesgo para complicaciones asociadas a la biopsia [13,2], estos enfoques se basan principalmente en métodos convencionales de regresión o sistemas de puntuación, más que en inteligencia artificial. Nuestro trabajo extiende esta línea de investigación mediante la exploración de algoritmos clásicos y de aprendizaje profundo para evaluar su viabilidad en la predicción del riesgo de complicaciones en un entorno clínico real.

Las principales contribuciones de este trabajo son: (i) un estudio preliminar sobre una cohorte clínica real de 125 pacientes sometidos a biopsia pulmonar guiada por TC; (ii) una comparación experimental entre modelos basados en características radiómicas, variables clínicas y aprendizaje profundo tridimensional; y (iii) un análisis de explicabilidad mediante Grad-CAM y SHAP para interpretar el comportamiento de los modelos predictivos.

Se ha llevado a cabo un estudio experimental en colaboración con el Hospital Universitario Clínico San Cecilio de Granada (Granada, España). El conjunto de datos está compuesto por estudios de TC torácica anonimizados e información clínica asociada de pacientes sometidos a biopsia pulmonar guiada por TC. Las imágenes fueron preprocesadas, segmentadas y normalizadas para extraer características radiómicas, que posteriormente se combinaron con variables clínicas para entrenar y evaluar distintos modelos predictivos. Se evaluaron tanto algoritmos clásicos de aprendizaje automático como arquitecturas de aprendizaje profundo para estimar la probabilidad de complicaciones relacionadas con el procedimiento. Aunque el tamaño y la heterogeneidad del conjunto de datos impusieron ciertas limitaciones, los modelos fueron capaces de clasificar correctamente una proporción sustancial de los casos e identificar patrones relevantes asociados a eventos adversos. Estos resultados sugieren que, con conjuntos de datos más amplios y equilibrados, podría mejorarse la capacidad predictiva y la aplicabilidad clínica del enfoque propuesto.

El resto del artículo se organiza del modo siguiente. La Sección 2 presenta el contexto clínico y el trabajo relacionado. La Sección 3 describe el conjunto de datos. La Sección 4 detalla el preprocesamiento y los enfoques de modelado basados en radiómica y aprendizaje profundo. La Sección 5 analiza los resultados experimentales y la explicabilidad de los modelos. Finalmente, la Sección 6 recoge las conclusiones y las líneas de trabajo futuro.

2. Antecedentes

La biopsia pulmonar guiada por imagen permite confirmar y caracterizar lesiones pulmonares, proporcionando información clave para orientar estrategias terapéuticas personalizadas. Entre sus enfoques, la vía transtorácica guiada por TC, predominante en el centro donde se ha realizado este estudio, ofrece mayor rendimiento diagnóstico que la consolidada vía transbronquial guiada por ecografía endobronquial radial (83,45 % frente a 68,82 %), aunque con mayores tasas de neumotórax (21,43 % frente a 2,87 %) y hemorragia (9,0 % frente a 5,7 %) [9]. Estas constituyen las dos principales complicaciones asociadas al procedimiento: el neumotórax consiste en la entrada de aire en la cavidad pleural tras la in-

serción de la aguja, pudiendo producir un colapso pulmonar parcial o completo, mientras que la hemorragia pulmonar corresponde al sangrado en el trayecto de la aguja o en el interior de la lesión, generalmente autolimitado, aunque en algunos casos puede asociarse a hemoptisis o desaturación de oxígeno.

La TC es una modalidad de imagen esencial en medicina clínica, ya que proporciona imágenes transversales no invasivas sin superposición anatómica y permite evaluar cuantitativamente la densidad tisular mediante unidades Hounsfield (HU) [5]. Su capacidad para generar volúmenes tridimensionales y visualizar los planos axial, coronal y sagital la convierte en una herramienta clave en intervenciones guiadas por imagen, como las biopsias pulmonares.

La radiómica es un campo de la imagen médica centrado en la extracción de métricas cuantitativas, denominadas características radiómicas, para describir propiedades tisulares y lesionales como heterogeneidad, forma y textura [16]. Estas características pueden reflejar información biológica no perceptible y correlacionarse con parámetros genómicos, histológicos y clínicos. Suelen agruparse en características de primer orden, asociadas a distribuciones de intensidad, y de orden superior, que capturan relaciones espaciales mediante matrices de textura. Combinadas con aprendizaje automático, permiten predecir variables como respuesta terapéutica o supervivencia, contribuyendo a la medicina de precisión. No obstante, su sensibilidad a los parámetros de adquisición y reconstrucción exige estandarización y armonización para garantizar la reproducibilidad.

El aprendizaje profundo se ha consolidado como un enfoque central en el análisis del cáncer de pulmón, con resultados destacados en detección, clasificación, segmentación y pronóstico [12]. En particular, las redes neuronales convolucionales (*Convolutional Neural Networks*, CNN) permiten aprender representaciones jerárquicas directamente a partir de imágenes, evitando la ingeniería manual de características. En estudios de TC, estos modelos han alcanzado una elevada precisión para diferenciar nódulos benignos y malignos, superando con frecuencia a la radiómica tradicional. Además, han mostrado eficacia en segmentación tumoral, con delimitaciones comparables a las de radiólogos expertos. Recientemente, los enfoques híbridos y multimodales, que integran imagen con información clínica o genómica, han ampliado la capacidad predictiva y el potencial del diagnóstico personalizado.

El aprendizaje por transferencia y los modelos preentrenados, como ResNet, DenseNet y EfficientNet, han favorecido la convergencia en conjuntos de datos limitados, mientras que las CNN tridimensionales y los mecanismos de atención han mejorado el análisis volumétrico y contextual. Además, técnicas de inteligencia artificial explicable como Grad-CAM y SHAP han incrementado la interpretabilidad y aceptación clínica de estos modelos. No obstante, persisten retos relevantes, como la escasez de datos, el desbalance entre clases y la falta de estandarización en la anotación y el preprocesamiento. Superar estas limitaciones resulta esencial para garantizar su robustez, reproducibilidad e implementación clínica en el diagnóstico y la planificación terapéutica del cáncer de pulmón.

Aunque la inteligencia artificial ha logrado avances notables en imagen de cáncer de pulmón, su uso en procedimientos intervencionistas como la biopsia

pulmonar guiada por TC sigue siendo muy limitado. La investigación existente se ha centrado principalmente en el diagnóstico o el pronóstico, más que en la predicción de riesgos relacionados con el procedimiento. Uno de los pocos estudios que aborda esta brecha es el de Anzidei et al. [2], quienes desarrollaron un modelo clínico de predicción y una escala de riesgo para complicaciones en pacientes sometidos a biopsia pulmonar guiada por TC. Su trabajo se basó en análisis estadístico tradicional y factores clínicos, sin incorporar métodos basados en imagen ni aprendizaje automático. Por tanto, el potencial de la inteligencia artificial para anticipar complicaciones asociadas a la biopsia a partir de TC preprocedimentales y datos clínicos sigue estando ampliamente inexplorado, lo que proporciona la motivación del presente estudio.

3. Descripción de los datos

El conjunto de datos utilizado en este estudio se recopiló en colaboración con el Hospital Universitario Clínico San Cecilio de Granada y está formado por datos anonimizados de 125 pacientes que se habían sometido previamente a biopsia pulmonar guiada por TC. La recopilación de datos tuvo lugar progresivamente entre diciembre de 2024 y julio de 2025, aunque los casos clínicos subyacentes corresponden a procedimientos realizados desde 2018 en adelante. Cada registro de paciente incluye un volumen tridimensional de TC torácica y un conjunto asociado de variables clínicas. Todos los datos de imagen fueron completamente anonimizados de acuerdo con la normativa de privacidad aplicable.

Los volúmenes de TC muestran variabilidad en resolución espacial y número de cortes, ya que fueron adquiridos mediante distintos escáneres. Las dimensiones originales de las imágenes oscilan entre $512 \times 512 \times x$ y $768 \times 768 \times x$, donde x representa el número variable de cortes axiales por estudio. Para garantizar la consistencia, todos los volúmenes fueron remuestreados, normalizados en intensidad y redimensionados durante el preprocesamiento. Las imágenes originales se proporcionaron en formato `DICOM`. Los volúmenes preprocesados se almacenaron como arrays `NumPy` para optimizar la carga y la integración con flujos de aprendizaje profundo.

Además de las imágenes, se dispuso de un conjunto de datos clínicos con variables demográficas y prebiopsia, como sexo, edad, factores de riesgo y condiciones pulmonares. Para mantener el realismo clínico, solo se emplearon variables disponibles antes del procedimiento. El conjunto incluye varias etiquetas objetivo, como presencia de complicación, tipo de complicación y tipo tumoral; sin embargo, este último no se utilizó en el entrenamiento, al obtenerse mediante el análisis histopatológico posterior a la biopsia. De los 125 pacientes, 59 no presentaron complicaciones y 66 sí, incluyendo neumotórax, hemorragias, hemorragias leves y derrame pleural leve, con algunos casos de eventos múltiples. Dado el tamaño limitado de la muestra, el estudio se centró en la predicción binaria de complicaciones, reservando el análisis multiclase para trabajos futuros.



Figura 1. Cortes axial, coronal y sagital representativos de un volumen de TC torácica perteneciente al conjunto de datos utilizado en este estudio.

Estos datos clínicos y de imagen se integraron en el flujo de modelado para extraer características relevantes y entrenar algoritmos predictivos. En la Figura 1 se muestra un ejemplo de volumen de TC en tres planos anatómicos distintos.

4. Metodología

Esta sección describe las metodologías aplicadas para abordar el problema binario de predicción de complicaciones en procedimientos de biopsia pulmonar. Se exploraron varios enfoques, incluyendo aprendizaje automático clásico basado en radiómica y modelos de aprendizaje profundo aplicados a datos de TC. Se evaluaron múltiples configuraciones experimentales empleando características de imagen, variables clínicas o una combinación de ambas.

4.1. Preprocesamiento de los datos

Para los datos de imagen de TC se implementó un flujo modular de preprocesamiento utilizando el marco `MONAI` [6], complementado con transformaciones específicas del dominio. Todos los volúmenes DICOM se convirtieron a NIfTI y se remuestrearon mediante interpolación trilineal para obtener un espaciado de vóxel uniforme y dimensiones de entrada estandarizadas. Se probaron varias resoluciones espaciales, observándose que los volúmenes de menor tamaño (por ejemplo, $128 \times 128 \times 128$) mejoraban la generalización al evitar detalles redundantes. La normalización de intensidad se realizó aplicando ventanas HU para realzar el contraste del parénquima pulmonar. Se evaluaron dos configuraciones de ventana pulmonar: una centrada en -600 HU con una anchura de 1500 HU y otra centrada en -300 HU con una anchura de 1400 HU. Esta etapa permitió destacar eficazmente las estructuras pulmonares y suprimir tejidos irrelevantes y artefactos [10]. La segmentación pulmonar se realizó automáticamente con la herramienta `TotalSegmentator` [22], aislando los pulmones y enmascarando regiones irrelevantes para centrar los modelos en áreas clínicamente significativas.

Los datos clínicos tabulares se preprocesaron mediante `scikit-learn` [17]. Los valores categóricos y los campos multietiqueta, como factores de riesgo y

condiciones pulmonares, se limpiaron y codificaron mediante codificación one-hot. La edad se normalizó mediante min-max al intervalo $[0, 1]$. El conjunto resultante contenía características numéricas y binarias consistentes, preparadas para su integración con datos de imagen durante el entrenamiento de los modelos.

4.2. Extracción de características radiómicas y aprendizaje clásico

Las características radiómicas se extrajeron con la biblioteca `PyRadiomics` [21], que calcula un amplio conjunto de descriptores de primer orden, basados en forma y de textura a partir de regiones segmentadas. Tras la extracción de características, se eliminaron variables redundantes y de baja varianza, y se aplicaron técnicas de normalización para reducir la variabilidad entre estudios.

La matriz de características resultante se utilizó para entrenar varios modelos de aprendizaje automático [19], como regresión logística, árboles de decisión, máquinas de vectores soporte (*support vector machines*, SVM), k vecinos más cercanos (*k-nearest neighbors*, kNN) y métodos de ensamblado, como Random Forest para *bagging* y LightGBM, XGBoost y Gradient Boosting para *boosting*.

Dado el rendimiento competitivo de kNN en pruebas preliminares, se exploraron técnicas de aprendizaje de métricas de distancia (*distance metric learning*, DML) para mejorar la estructura del espacio de características e incrementar la separabilidad entre clases [20]. Se evaluaron dos algoritmos representativos: Neighborhood Components Analysis (NCA) y Large Margin Nearest Neighbors (LMNN). Ambos enfoques buscan aprender una métrica que capture mejor las relaciones entre patrones radiómicos asociados a complicaciones.

Finalmente, los experimentos basados en radiómica se realizaron tanto de forma independiente como en combinación con variables clínicas, añadiendo las características tabulares a los descriptores radiómicos. Esto permitió evaluar la información complementaria aportada por la imagen anatómica y los datos a nivel de paciente en la predicción del riesgo de complicaciones.

4.3. Modelos de aprendizaje profundo

Se llevaron a cabo experimentos preliminares con CNN bidimensionales, aunque no se exploraron de forma exhaustiva debido al contexto espacial limitado que proporcionan. Dada la naturaleza volumétrica de los datos de TC, un enfoque más adecuado consistió en procesar el volumen tridimensional completo mediante CNN 3D. Los principales modelos evaluados fueron DenseNet121 3D [11] y ResNet 3D [7], inicializados con pesos preentrenados y ajustados sobre el conjunto de datos de biopsia pulmonar. Se aplicó aumento de datos moderado, incluyendo transformaciones afines aleatorias, para mejorar la generalización. El entrenamiento utilizó el optimizador Adam con una pérdida de entropía cruzada binaria ponderada y *early stopping* para evitar el sobreajuste.

Para abordar el tamaño limitado del conjunto de datos, se emplearon estrategias de aprendizaje por transferencia. Las arquitecturas CNN 3D se preentrenaron primero sobre el conjunto de datos MedMNIST3D [23], que contiene estudios volumétricos de TC para tareas de clasificación anatómica. Esta etapa

de preentrenamiento permitió a las redes aprender representaciones morfológicas y texturales generalizables que podían transferirse a la tarea objetivo. Se evaluaron dos estrategias de ajuste fino: (1) actualizar únicamente la capa final de clasificación manteniendo congelado el extractor de características, y (2) reentrenar completamente la red con una tasa de aprendizaje menor.

Además, durante el preentrenamiento se probaron funciones de pérdida de aprendizaje profundo de métricas, como *contrastive loss* y *triplet loss* [14], junto con la pérdida estándar de entropía cruzada, con el objetivo de fomentar la compacidad intraclase y la separación interclase.

También se desarrolló un enfoque multimodal para integrar información de imagen y clínica. El volumen de TC segmentado se procesó mediante una DenseNet 3D, mientras que las variables clínicas se introdujeron en un perceptrón multicapa. Las salidas de ambas ramas se concatenaron y se pasaron a través de una capa completamente conectada para la clasificación binaria. Esta configuración híbrida permitió al modelo combinar indicios espaciales y clínicos, reflejando los procesos reales de toma de decisiones clínicas.

4.4. Protocolo de validación y control de fugas de información

El conjunto de datos se dividió mediante validación cruzada estratificada de cinco particiones. Para evitar sesgos de evaluación y posibles problemas de *data leakage*, todas las etapas dependientes de los datos se ajustaron de forma independiente dentro de cada partición. En particular, la normalización, la eliminación de variables redundantes o de baja varianza, el aprendizaje de métricas de distancia y el entrenamiento de los clasificadores se realizaron exclusivamente con los datos de entrenamiento de cada *fold*, aplicando después las transformaciones correspondientes al conjunto de prueba. Todas las transformaciones se implementaron mediante *pipelines* encapsulados dentro de cada *fold*.

5. Análisis de resultados

El rendimiento de los modelos se evaluó mediante *Accuracy*, F1-score, tasa de verdaderos positivos (*True Positive Rate*, TPR), tasa de verdaderos negativos (*True Negative Rate*, TNR) y media geométrica (*Geometric Mean*, G-Mean), calculada como G-Mean $= \sqrt{\text{TPR} \cdot \text{TNR}}$. Esta última métrica resulta especialmente informativa, ya que penaliza de forma sustancial valores muy bajos de TPR o TNR. La Tabla 1 resume los modelos con mejor rendimiento en todas las configuraciones experimentales evaluadas en este estudio.

Los modelos basados en radiómica alcanzaron el mejor rendimiento global, superando claramente tanto a los enfoques de aprendizaje profundo como a los modelos basados únicamente en variables clínicas. La combinación de características radiómicas con variables clínicas y aprendizaje de métricas de distancia proporcionó las predicciones más equilibradas y precisas, alcanzando una G-Mean de aproximadamente el 74,7 %. Este comportamiento se benefició de la información complementaria entre los datos anatómicos y los datos a nivel de

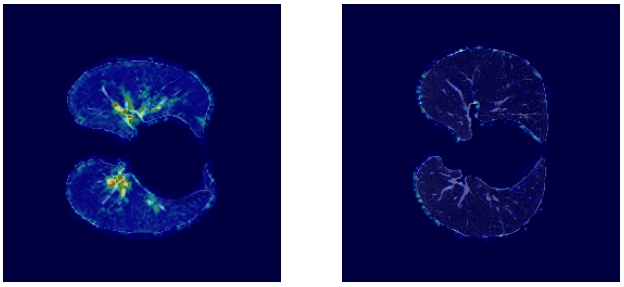
Tabla 1. Resumen de los modelos con mejor rendimiento en todas las configuraciones experimentales.

Categoría	Modelo	Acc.	F1	TPR	TNR	G-Mean
DL3D base	DenseNet121	0.696	0.680	0.679	0.709	0.691
DL3D multimodal	DenseNet121	0.512	0.459	0.456	0.559	0.469
DL3D preentrenado	ResNet3D	0.592	0.599	0.664	0.533	0.579
Solo variables clínicas	XGBoost	0.576	0.498	0.444	0.695	0.553
Radiómica	kNN	0.744	0.713	0.724	0.765	0.739
Radiómica + clínicas	LightGBM	**0.760**	0.725	0.700	**0.805**	0.746
Radiómica + DML	kNN + NCA	0.736	0.707	0.723	0.749	0.732
Radiómica + DML + clínicas	kNN + LMNN	0.747	**0.727**	**0.758**	0.738	**0.747**

paciente, así como de una representación mejorada de las características mediante aprendizaje de métricas. En cambio, los modelos de aprendizaje profundo 3D mostraron una capacidad de generalización limitada, con la DenseNet121 base alcanzando alrededor del 69,1 % de G-Mean, principalmente debido al reducido tamaño del conjunto de datos y a la variabilidad entre estudios de TC, que dificultaron el entrenamiento de arquitecturas de alta capacidad. Cabe destacar que ni el modelo multimodal (46,9 % de G-Mean) ni las configuraciones preentrenadas (57,9 % de G-Mean) mejoraron los resultados de la red 3D base, a pesar de ser teóricamente más expresivos. Esto sugiere que, bajo las restricciones actuales de datos, el aprendizaje por transferencia y la fusión de características no se aprovechan de forma efectiva. En conjunto, el aprendizaje automático clásico basado en radiómica demostró la mayor robustez y eficiencia de datos para predecir complicaciones de la biopsia.

Para interpretar el comportamiento de los modelos, se aplicaron dos técnicas de explicabilidad *post hoc*, Grad-CAM [18] y SHAP [15], sobre una partición entrenamiento-test procedente de la validación cruzada y utilizando varios de los modelos más relevantes del análisis experimental. Las visualizaciones Grad-CAM de los modelos de aprendizaje profundo resaltaron las regiones más relevantes para cada predicción y fueron revisadas en colaboración con médicos del Hospital Universitario Clínico San Cecilio de Granada para garantizar su coherencia clínica. Como se muestra en la Figura 2, los mapas confirmaron que las predicciones correctas se centraban en regiones pulmonares relevantes, mientras que las clasificaciones erróneas a menudo se concentraban en zonas periféricas o irrelevantes. Este análisis también orientó mejoras en el preprocesamiento y la segmentación.

SHAP se utilizó para ordenar las características en los modelos de aprendizaje profundo y de radiómica. Aunque su aplicación a volúmenes 3D presenta ciertas limitaciones, permitió analizar cómo las entradas influían en las predicciones. En la Figura 3, LightGBM se basó principalmente en métricas de textura transformadas mediante *wavelet* y logaritmo (GLDM, GLSZM, NGTDM), capturando heterogeneidad parenquimatosa multiescala coherente con fragilidad tisular y riesgo de neumotórax cuando la aguja atraviesa el pulmón. Los descriptores centrados en bordes e intensidad (GLRLM, LBP y características de primer orden) sugieren, en cambio, regiones más densas, vasculares o fibróticas, y pueden relacionarse con complicaciones hemorrágicas.



(a) Clasificación correcta. (b) Clasificación incorrecta.

Figura 2. Visualizaciones Grad-CAM para un caso clasificado correctamente y otro clasificado incorrectamente.

6. Conclusiones

Este estudio abordó el reto de predecir complicaciones en biopsias pulmonares guiadas por TC mediante inteligencia artificial, combinando información radiómica y clínica. Los experimentos mostraron que los enfoques basados exclusivamente en aprendizaje profundo, incluyendo modelos preentrenados y multimodales, no superaron a la arquitectura 3D base, probablemente debido al tamaño limitado y a la heterogeneidad del conjunto de datos. En cambio, el aprendizaje automático basado en radiómica resultó más robusto e interpretable, alcanzando el mejor rendimiento global al integrarse con variables clínicas y aprendizaje de métricas (G-Mean $\approx 74\,\%$). Aunque todavía existe margen de mejora, estos resultados son prometedores y sugieren que puede extraerse información relevante para la predicción de complicaciones a partir de estudios de TC preprocedimentales. Por tanto, los modelos clásicos pueden proporcionar resultados competitivos y más explicables incluso en conjuntos de datos médicos pequeños, ofreciendo una base prometedora para el apoyo a la decisión clínica y la futura automatización de la evaluación del riesgo de complicaciones.

Como trabajo futuro, se ampliará el conjunto de datos con cohortes externas y estrategias de aprendizaje federado. Se mejorará la segmentación pulmonar y se explorarán enfoques semisupervisados para aprovechar estudios no etiquetados. Estas extensiones permitirán avanzar hacia la predicción multietiqueta de complicaciones y reforzar la evaluación clínica de los modelos propuestos.

Agradecimientos

Esta investigación es parte del Proyecto TSI-100927-2023-1, financiado por Plan de Recuperación, Transformación y Resiliencia Europeo, de la Unión Europea-Next Generation EU a través del Ministerio para la Transformación Digital y de la Función Pública y está parcialmente financiada por el Ministerio de Ciencia y Tecnología de España en el marco del proyecto PID2023-150070NB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

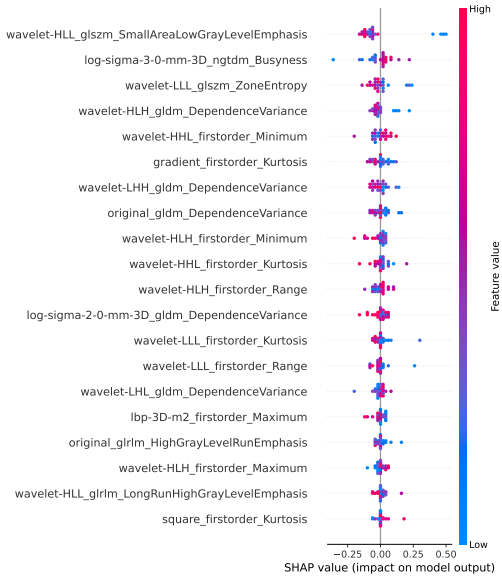


Figura 3. Gráfico resumen global de SHAP para el modelo de radiómica con mejor rendimiento (LightGBM) en el conjunto de los pacientes de prueba.

Referencias

1. Adams, S.J., Stone, E., Baldwin, D.R., Vliegenthart, R., Lee, P., Fintelmann, F.J.: Lung cancer screening. The Lancet **401**(10374), 390–408 (2023)
2. Anzidei, M., Sacconi, B., Fraioli, F., Saba, L., Lucatelli, P., Napoli, A., Longo, F., Vitolo, D., Venuta, F., Anile, M., et al.: Development of a prediction model and risk score for procedure-related complications in patients undergoing percutaneous computed tomography-guided lung biopsy. European Journal of Cardio-Thoracic Surgery **48**(1), e1–e6 (2015)
3. Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., et al.: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature medicine **25**(6), 954–961 (2019)
4. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **74**(3), 229–263 (2024)
5. Buzug, T.M.: Computed tomography. In: Springer handbook of medical technology, pp. 311–342. Springer (2011)
6. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
7. Hara, K., Kataoka, H., Satoh, Y.: Learning spatio-temporal features with 3d residual networks for action recognition. In: Proceedings of the IEEE international conference on computer vision workshops. pp. 3154–3160 (2017)

8. Heerink, W.J., de Bock, G.H., de Jonge, G.J., Groen, H.J., Vliegenthart, R., Oudkerk, M.: Complication rates of ct-guided transthoracic lung biopsy: meta-analysis. European radiology **27**(1), 138–148 (2017)
9. Ho, A.T.N., Gorthi, R., Lee, R., Chawla, M., Patolia, S.: Solitary lung nodule: Ct-guided transthoracic biopsy vs transbronchial biopsy with endobronchial ultrasound and flexible bronchoscope, a meta-analysis of randomized controlled trials. Lung **201**(1), 85–93 (2023)
10. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images. IEEE transactions on medical imaging **20**(6), 490–498 (2001)
11. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
12. Javed, R., Abbas, T., Khan, A.H., Daud, A., Bukhari, A., Alharbey, R.: Deep learning for lungs cancer detection: a review. Artificial Intelligence Review **57**(8), 197 (2024)
13. Lee, H., Murphy, C., Barbosa Jr, E.J.M.: Prediction of complication risk in computed tomography-guided thoracic biopsy: a prescription for improving procedure safety. Journal of Thoracic Imaging **38**(2), 88–96 (2023)
14. Li, X., Yang, X., Ma, Z., Xue, J.H.: Deep metric learning for few-shot image classification: A review of recent developments. Pattern Recognition **138**, 109381 (2023)
15. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. Advances in neural information processing systems **30** (2017)
16. Mayerhoefer, M.E., Materka, A., Langs, G., Häggström, I., Szczypiński, P., Gibbs, P., Cook, G.: Introduction to radiomics. Journal of Nuclear Medicine **61**(4), 488–495 (2020)
17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
18. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
19. Shalev-Shwartz, S., Ben-David, S.: Understanding machine learning: From theory to algorithms. Cambridge university press (2014)
20. Suárez, J.L., García, S., Herrera, F.: A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges. Neurocomputing **425**, 300–322 (2021)
21. Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. Cancer research **77**(21), e104–e107 (2017)
22. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
23. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data **10**(1), 41 (2023)