

LOCALIZACIÓN Y CLASIFICACIÓN DE REGIONES ASOCIADAS CON  
MUTACIONES GENÓMICAS DE NÓDULOS PULMONARES EN IMÁGENES PET Y  
CT

ISABELLA CALLEJAS MANDON  
JOAN SEBASTIÁN PATIÑO JAIMES  
CRISTIAN STIVENS VILLARREAL PARRA

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2025

LOCALIZACIÓN Y CLASIFICACIÓN DE REGIONES ASOCIADAS CON  
MUTACIONES GENÓMICAS DE NÓDULOS PULMONARES EN IMÁGENES PET Y  
CT

ISABELLA CALLEJAS MANDON  
JOAN SEBASTIÁN PATIÑO JAIMES  
CRISTIAN STIVENS VILLARREAL PARRA

Tesis presentada en cumplimiento de los requisitos para optar por el título de:  
Ingenieros de Sistemas

Director:

Fabio Martínez Carrillo  
Doctor en Ingeniería de Sistemas y Computación

Codirectora:

Alejandra Moreno Tarazona  
Magíster en Ingeniería de Sistemas e Informática

Codirector:

Luis Carlos Guayacán Chaparro  
Magíster en Matemática Aplicada

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2025

## AGRADECIMIENTOS

### **Joan Sebastian Patiño Jaimes**

Quiero expresar mi más sincero agradecimiento a todas las personas que fueron parte importante en la construcción de esta tesis. En primer lugar, a mi familia, por su amor incondicional, paciencia y apoyo constante. Gracias por ser mi motor, por creer en mí incluso cuando yo dudaba, y por acompañarme en cada paso de este camino. Al grupo de investigación BIVL<sup>2</sup>ab, gracias por abrirme las puertas a un espacio donde no solo crecí como investigador, sino también como persona. Agradezco profundamente su guía, los recursos compartidos y, sobre todo, la calidez con la que siempre me hicieron sentir parte de esta familia. A mi director, Fabio Martínez Carrillo, agradezco su dedicación, visión y acompañamiento invaluable a lo largo de este proceso. Su orientación fue clave para que esta investigación tomara forma. Igualmente, gracias a mi codirector, Luis Carlos Guayacán, por su valioso aporte académico, paciencia, consejos y constante disponibilidad para resolver mis inquietudes durante todo este proceso. A mis compañeros de tesis, gracias por compartir este viaje. Juntos enfrentamos desafíos, celebramos logros y aprendimos más de lo que imaginábamos. Su colaboración y buena energía hicieron todo mucho más llevadero. Y, por supuesto, a mis amigos, gracias por ser ese refugio emocional en los momentos difíciles. Sus palabras de aliento, compañía y apoyo me ayudaron a mantener el rumbo y no rendirme. A todos ustedes, gracias de corazón. Esta tesis también es suya.

### **Isabella Callejas Mandón**

Quiero agradecer, en primer lugar y ante todo, a Dios por todas las oportunidades que me ha brindado y por nunca soltar mi mano. A mi mamá, a mi papá y a mis hermanos: Néstor, Alejandra y Karen, por ser mi refugio ante la adversidad y por su constante amor incondicional. Agradezco también al grupo de investigación BIVL<sup>2</sup>ab, no solo por los recursos, guía e infinita paciencia, si no también por ofrecerme un segundo hogar, en el que, en conjunto con expe-

riencias inolvidables, me inculcaron valores que llevaré conmigo a todas partes. En especial, quiero resaltar a mi director, el profesor Fabio Martínez, por su acompañamiento constante, las oportunidades ofrecidas y paciencia, no solo durante la realización de este proyecto, sino también a lo largo de toda mi trayectoria en el grupo. A mi codirectora Alejandra Moreno, a quien admiro enormemente y considero que, más que una mentora, es la mejor hermana científica del mundo. Y a mi codirector Luis Guayacán, por la motivación, sus consejos y su humor tan particular, que sin duda hicieron la experiencia de este proyecto más ameno. A mis compañeros de proyecto Joan y Cristian, con los que en la marcha construimos una amistad, compartiendo aciertos, errores, aprendizajes y experiencias que forjaron esta unión tanto personal como académica. Y finalmente, le agradezco a mis amigos de la universidad, que me han acompañado en todo este camino, que han llorado, trasnochado y reído incontables veces conmigo. Mil millones de gracias.

### **Cristian Stivens Villarreal Parra**

Quiero dar mi agradecimiento a todas las personas que hicieron posible la realización de este proyecto de grado. A mis padres y abuela por dar su apoyo constante y sincero, además de ser parte integral del porque me encuentro en este punto de mi vida. Agradecer al grupo de investigación BIVL<sup>2</sup>ab porque me permitió hacer parte de el y formarme como persona. A mi director Fabio Martínez Carrillo, por su contante guía hacia en cada paso de este proyecto. Agradecer a mi codirectora Alejandra Moreno Tarazona, por su dedicación, paciencia y generosidad al compartir sus conocimientos a lo largo de este camino. Igualmente a mi segundo codirector Luis Carlos Guayacán, ya que me apoyo en cada paso del proyecto. A mis amigos y compañeros que han estado presentes en cada paso de mi vida, que conocen el esfuerzo, dedicación y sacrificio vivido, muchas gracias.

## CONTENIDO

	pág.
<b>INTRODUCCIÓN</b> . . . . .	<b>12</b>
<b>1. FUNDAMENTOS Y TRABAJOS PREVIOS</b> . . . . .	<b>15</b>
1.1. CÁNCER DE PULMÓN Y MARCADORES DIAGNÓSTICOS . . . . .	15
1.2. ESTRATEGIAS COMPUTACIONALES PARA LA DETECCIÓN Y LA CLASIFICACIÓN . . . . .	23
1.2.1 Modelos de detección. . . . .	23
1.2.2 Modelos de clasificación . . . . .	26
1.3. DETECCIÓN Y CLASIFICACIÓN DE BIOMARCADORES DE CÁNCER DE PULMÓN . . . . .	33
1.3.1 Métodos de detección . . . . .	33
1.3.2 Métodos de clasificación . . . . .	36
<b>2. PROBLEMA DE INVESTIGACIÓN</b> . . . . .	<b>41</b>
<b>3. OBJETIVOS</b> . . . . .	<b>43</b>
3.1. OBJETIVO GENERAL . . . . .	43
3.2. OBJETIVOS ESPECÍFICOS . . . . .	43
<b>4. MÉTODO PROPUESTO</b> . . . . .	<b>44</b>
4.1. MODELO MULTIMODAL DE CT Y PET PARA LA LOCALIZACIÓN DE NÓDULOS PULMONARES . . . . .	45
4.2. MODELO SUPERVISADO MULTIMODAL BASADO EN TRANSFORMERS PARA LA CLASIFICACIÓN DE MUTACIONES GENÓMICAS EN NP CON IMÁGENES CT, PET E INFORMACIÓN TABULAR . . . . .	49

4.3. MODELO AUTOSUPERVISADO BASADO EN SÍNTESIS DE IMÁGENES PARA LA CLASIFICACIÓN DE MUTACIONES GENÓMICAS EN NP CON IMÁGENES CT Y PET . . . . .	55
<b>5. DISEÑO EXPERIMENTAL . . . . .</b>	<b>59</b>
5.1. DATOS . . . . .	59
5.2. CONFIGURACIÓN DE LAS ARQUITECTURAS . . . . .	61
5.3. VALIDACIÓN . . . . .	63
<b>6. EVALUACIÓN Y RESULTADOS . . . . .</b>	<b>64</b>
6.1. EVALUACIÓN DE MODELOS DE DETECCIÓN . . . . .	64
6.2. EVALUACIÓN DE MODELOS DE CLASIFICACIÓN SUPERVISADA Y AUTOSUPERVISADA . . . . .	68
<b>7. CONCLUSIONES Y TRABAJO FUTURO . . . . .</b>	<b>77</b>
<b>BIBLIOGRAFÍA . . . . .</b>	<b>79</b>

## LISTA DE FIGURAS

	<b>pág.</b>
Figura 1. Imágenes CT y PET para cada mutación del dataset NSCLC Radiogenomics.	19
Figura 2. Vías de transducción de señales relacionadas con EGFR y KRAS . . . . .	21
Figura 3. Fenotipos de EGFR y KRAS observados en CT y PET . . . . .	23
Figura 4. Esquema de la arquitectura YOLOv1 . . . . .	24
Figura 5. Esquema de la arquitectura YOLOv5 . . . . .	25
Figura 6. Estrategias de fusión multimodal . . . . .	28
Figura 7. Esquema de la Red Generativa Adversaria (GAN) . . . . .	31
Figura 8. Esquema de la arquitectura Pix2Pix . . . . .	32
Figura 9. Esquema de la red generativa antagónica CycleGAN . . . . .	34
Figura 10. Esquema del flujo de trabajo propuesto de localización y clasificación de NP	45
Figura 11. Esquema del método propuesto para detección multimodal de NP . . . . .	46
Figura 12. Modelo multimodal de clasificación supervisada basado en Transformers . .	50
Figura 13. Arquitectura supervisada individual para CT . . . . .	51
Figura 14. Arquitectura supervisada individual para PET . . . . .	52
Figura 15. Arquitectura supervisada individual para datos tabulares . . . . .	55
Figura 16. Esquema del método autosupervisado propuesto . . . . .	56
Figura 17. Distribución de pacientes del dataset NSCLC-Radiogenomics . . . . .	59
Figura 18. Evaluación del método de enmascaramiento en Tab Transformer . . . . .	70
Figura 19. Evaluación ViT con extractores de características en CT y PET . . . . .	71
Figura 20. Evaluación de modelos multimodales según su fusión y enmascaramiento . .	72
Figura 21. Evaluación de rendimiento de las probabilidades del modelo . . . . .	73
Figura 22. Cada modelo unimodal vs método propuesto . . . . .	74

Figura 23.Evaluación de AUC de los embebidos con diferentes clasificadores . . . . . 76

## LISTA DE TABLAS

	pág.
Tabla 1. Imágenes del <i>dataset NSCLC-Radiogenomics</i> en cada etapa de procesamiento para CT y PET . . . . .	61
Tabla 3. Resultados para la modalidad PET original (validación automática YOLO, 5 folds). . . . .	65
Tabla 4. Resultados para la modalidad PET-CT fusionada con $\alpha = 0.5$ (validación automática YOLO, 5 folds). . . . .	65
Tabla 5. Resultados unificados para CT segmentada (validación automática YOLO, 5 folds). . . . .	66
Tabla 6. Resultados para la modalidad PET segmentada (validación automática YOLO, 5 folds). . . . .	67
Tabla 7. Resultados <b>pre</b> -reductor en detección (5 <i>folds</i> ). . . . .	67
Tabla 8. Resultados <b>post</b> -reductor en detección (5 <i>folds</i> ). . . . .	68
Tabla 9. Estudio de diversos modelos ante diferentes métodos de codificación de datos tabulares . . . . .	69
Tabla 10. Métricas de clasificación del método propuesto en el problema multiclase . . .	74
Tabla 11. Resultados de clasificación con CycleGAN (5° bloque) . . . . .	76

## RESUMEN

**TÍTULO:** Localización y clasificación de regiones asociadas con mutaciones genómicas de nódulos pulmonares en imágenes PET y CT. \*

**AUTOR:** Isabella Callejas Mandon, Joan Sebastián Patiño Jaimés y Cristian Stivens Villarreal Parra. \*\*

**PALABRAS CLAVE:** Cáncer de pulmón, Nódulos pulmonares, Mutación genómica, CT, PET, Localización, Síntesis.

**DESCRIPCIÓN:** El cáncer de pulmón (CP) es el tipo de cáncer más frecuente y letal a nivel mundial, reportando 2.4 millones de incidencias y 1.8 millones de muertes en 2022 respectivamente. Los nódulos pulmonares (NP) son los principales biomarcadores de esta enfermedad, típicamente observados y caracterizados morfológicamente en tomografías computarizadas (CT). De forma complementaria y especializada, las tomografías por emisión de positrones (PET) brindan información funcional con alta correlación respecto a su malignidad. Además de su identificación y caracterización radiológica, la asociación con mutaciones genómicas específicas son claves para la definición de un tratamiento con pronóstico favorable. Sin embargo, este flujo de trabajo utilizando información morfológica (CT), funcional (PET) y genómica es poco común en la rutina clínica, debido a los altos costos, la exposición a la radiación en PET, el uso de biopsias invasivas y la poca disponibilidad de dispositivos para su captura. Además, persisten desafíos en la identificación y caracterización de los NP, debido a la alta subjetividad, variabilidad morfológica y la similitud con otras estructuras del parénquima. En este trabajo de investigación, se abordó un flujo de trabajo para la localización y clasificación de NP, considerando estudios imagenológicos de CT, PET, así como su mutación genómica. Para la localización de NP se empleó una arquitectura YOLO sobre CT, utilizando la información de PET como reducción de falsos positivos. Para la clasificación de mutaciones genómicas se implementó un modelo multimodal basado en Transformers, además de un enfoque autosupervisado que extrajo representaciones latentes mediante traducción CT-PET. El flujo completo se entrenó y validó en NSCLC-Radiogenomics (104 estudios). En localización, el modelo alcanzó un mAP50 de  $73.30\% \pm 4.30$ . Para clasificación, el enfoque multimodal obtuvo un AUC de  $94.14\% \pm 8.00$ , y el método autosupervisado logró un AUC de  $84.60\% \pm 9.00$ .

---

\* Trabajo de investigación

\*\* Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez, PhD. Codirectores: Alejandra Moreno Tarazona, Msc., Luis Carlos Guayacán Chaparro, Msc.

## ABSTRACT

**TITLE:** Localization and classification of regions associated with genomic mutations of pulmonary nodules in PET and CT images. \*

**AUTHOR:** Isabella Callejas Mandon, Joan Sebastián Patiño Jaimes y Cristian Stivens Villarreal Parra.  
\*\*

**KEYWORDS:** Lung cancer, Lung nodules, Genomic mutation, CT, PET, Localization, Synthesis.

### DESCRIPTION:

Lung cancer (LC) is the most common and lethal cancer worldwide, reporting 2.4 million cases and 1.8 million deaths in 2022, respectively. Pulmonary nodules (PNs) are the main biomarkers for this disease, typically observed and morphologically characterized using computed tomography (CT) scans. Complementarily, and in a specialized manner, positron emission tomography (PET) scans provide functional information with a high correlation to malignancy. In addition to their identification and radiological characterization, the association with specific genomic mutations is key to defining a treatment with a favorable prognosis. However, this workflow using morphological (CT), functional (PET), and genomic information is uncommon in routine clinical practice due to high costs, radiation exposure during PET scans, the use of invasive biopsies, and the limited availability of data acquisition devices. Furthermore, challenges persist in the identification and characterization of NPs due to high subjectivity, morphological variability, and their similarity to other parenchymal structures. In this research, a workflow for the localization and classification of NPs was developed, considering CT and PET imaging studies, as well as their genomic mutations. A YOLO architecture on CT was used for NP localization, utilizing PET information to reduce false positives. A multimodal Transformers-based model was implemented for genomic mutation classification, along with a self-supervised approach that extracted latent representations through CT-PET translation. The complete workflow was trained and validated in NSCLC-Radiogenomics (104 studies). For localization, the model achieved an mAP50 of  $73.30 \pm 4.30$ . For classification, the multimodal approach achieved an AUC of  $94.14\% \pm 8.00$ , and the self-supervised method achieved an AUC of  $84.60\% \pm 9.00$ .

---

\* Research work

\*\* Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Fabio Martínez Carrillo, PhD. Co-advisors: Alejandra Moreno Tarazona, Msc., Luis Carlos Guayacán Chaparro, Msc.

## INTRODUCCIÓN

El cáncer de pulmón (CP) es el tipo de cáncer más frecuente y letal a nivel mundial, estimándose en 2.4 millones de casos incidentes y 1.8 millones de pacientes fallecidos durante el 2022<sup>1</sup>. Su alta mortalidad está asociada con un diagnóstico tardío, al ser confundido con otras enfermedades de cuidado primario y la generalización en el tratamiento sin tener en cuenta el origen de mutación asociado en cada paciente<sup>2</sup>. Lo anterior, en consecuencia, conlleva a diagnósticos generalizados que limitan la personalización en la planificación y tratamiento de la enfermedad.

En la ruta integral de atención, diagnóstico y caracterización temprana del cáncer de pulmón, se abordan diferentes etapas que van desde la detección y estratificación de los nódulos pulmonares (NP), hasta, en estudios especializados, su categorización genómica. En cuanto a la detección de los NP, los cuales son los principales biomarcadores del CP, se realiza principalmente mediante tomografías computarizadas (CT, por sus siglas en inglés). Por otra parte, las tomografías por emisión de positrones (PET, por sus siglas en inglés) proporcionan información funcional del metabolismo tumoral, ofreciendo ventajas en la localización de nódulos incluso antes de que sean visibles en las imágenes anatómicas<sup>3</sup>. En estudios más sofisticados, se incluyen pruebas genómicas para identificar mutaciones en genes específicos como EGFR y KRAS, permitiendo definir opciones de tratamiento, como la terapia dirigida o personalizada,

---

<sup>1</sup> Ferlay J et AL. *Global Cancer Observatory: Cancer Today (version 1.1)*. <https://gco.iarc.who.int/today>. Lyon, France: International Agency for Research on Cancer. Accessed: [26 08 2024]. 2024.

<sup>2</sup> Neal Richard D BRADLEY STEPHEN H Kennedy Martyn PT. “Recognising lung cancer in primary care”. In: *Advances in therapy* 36.1 (2019), pp. 19–30.

<sup>3</sup> Liu HILLNER BRUCE E Siegel Barry A. “Impact of Positron Emission Tomography/Computed Tomography and Positron Emission Tomography (PET) Alone on Expected Management of Patients With Cancer: Initial Results From the National Oncologic PET Registry”. In: *Journal of Clinical Oncology* (2008). DOI: [10.1200/JCO.2007.14.5631](https://doi.org/10.1200/JCO.2007.14.5631).

y proporcionan información crucial sobre el pronóstico de la enfermedad y la probabilidad de respuesta a diferentes tipos de tratamientos<sup>4</sup>. Estos pasos en la ruta de atención son en general aislada y no existen mecanismos que permitan aprovechar información correlativa, desde las diferentes etapas, para hacer un seguimiento más riguroso de los pacientes. Por lo tanto, es aún inminente la necesidad de desarrollar análisis complementarios con las diferentes etapas diagnósticas, en la ruta de atención, para correlacionar diferentes hallazgos, que en últimas, permitan robustecer los procedimientos diagnósticos y de tratamiento del CP.

En la literatura se han propuesto diferentes herramientas computacionales para brindar soporte a las diferentes etapas de atención y caracterización de NP. En cuanto a la detección y clasificación de NP, se destacan principalmente enfoques unimodales basados en las imágenes CT, usando modelos como *YOLOv3*, *YOLOv5* y *YOLOv7*, así como variantes optimizadas mediante *Transformers* o análisis multiescala<sup>5,6</sup>. En cuanto a la estratificación de los nódulos, típicamente se emplean descriptores netamente texturales del CT<sup>7</sup>. Trabajos recientes han mostrado que la integración de diversas modalidades, como por ejemplo, información anatómica y funcional (CT-PET) brindan ventajas en la localización y clasificación de NP<sup>8</sup>. También, se han abordado trabajos que integran hallazgos visuales desde los estudios imagenológicos con datos clínicos para la respectiva caracterización de los NP. Incluso, estos

---

<sup>4</sup> Chen Jian-Hua YUAN MIN Huang Li-Li. “The emerging treatment landscape of targeted therapy in non-small-cell lung cancer”. In: *Signal transduction and targeted therapy* 4.1 (2019), p. 61.

<sup>5</sup> Jin LIU et al. “Detection of pulmonary nodules with YOLOv3 in chest CT”. in: *Quantitative Imaging in Medicine and Surgery* 10.10 (2020), pp. 1917–1929.

<sup>6</sup> X. WU et al. “YOLOv7 with multi-scale receptive field for small pulmonary nodule detection”. In: *Pattern Recognition* 145 (2024), p. 109876.

<sup>7</sup> Kawaji Hideya TAKAMOCHI KAZUYA Mogushi Kaoru. “Correlation of EGFR or KRAS mutation status with 18F-FDG uptake on PET-CT scan in lung adenocarcinoma”. In: *PLoS One* 12.4 (2017), e0175622.

<sup>8</sup> Nazari Mostafa SHIRI ISAAC Amini Mehdi. “Impact of feature harmonization on radiogenomics analysis: Prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images”. In: *Computers in biology and medicine* 142 (2022), p. 105230.

enfoques han permitido aproximar características de la mutación de un NP desde las características textuales y/o de imágenes CT, PET <sup>9</sup>. A pesar de los avances reportados en el análisis multimodal, aún la caracterización de NP se considera un problema abierto que requiere estudios adicionales y propuestas computacionales para contribuir tanto en la localización de NP, como en su aproximación genómica desde datos capturados en la ruta de atención diagnóstica.

En este trabajo de investigación, se abordó un flujo de trabajo para atender diferentes etapas de la ruta de atención diagnóstica, específicamente relacionadas con la localización y clasificación de NP, considerando estudios imagenológicos de CT, PET, datos clínicos y su correspondiente mutación genómica (EGFR o KRAS). En localización, se implementó una arquitectura *YOLOv9*, ajustado sobre secuencias CT, responsable de la detección inicial de potenciales regiones nodulares. Como esquema de reducción de falsos positivos, se integraron observaciones funcionales de la respuesta metabólica de los estudios PET para discriminar entre verdaderos nódulos y otras estructuras pulmonares. En cuanto a la clasificación de las observaciones nodulares, en este trabajo se implementó una arquitectura multimodal basada en *Transformers* con tolerancia a datos faltantes para la clasificación de mutaciones genómicas, en el cual, cada modalidad tiene su propia arquitectura y se unen mediante una fusión tardía. Además, se exploró una estrategia autosupervisada basada en la síntesis de imágenes CT y PET mediante una arquitectura *CycleGAN*. Durante la fase de evaluación, se extrajeron representaciones latentes en capas intermedias del generador empleando únicamente imágenes CT, estas representaciones se usaron como descriptores para la discriminación de mutaciones genómicas con regresión logística. El flujo de trabajo fue entrenado y validado sobre el conjunto de datos público *NSCLC-Radiogenomics* con 104 estudios.

---

<sup>9</sup> Augustin Alisha Isabelle TRIPATHI SATVIK Moyer Ethan Jacob. “Radgenets: Deep learning-based radiogenomics model for gene mutation prediction in lung cancer”. In: *Informatics in medicine unlocked* 33 (2022), p. 101062.

## 1. FUNDAMENTOS Y TRABAJOS PREVIOS

### 1.1. CÁNCER DE PULMÓN Y MARCADORES DIAGNÓSTICOS

El cáncer de pulmón (CP) es la principal causa de muerte por cáncer a nivel mundial. En el 2022, se diagnosticaron cerca de 2.4 millones de nuevos casos y se registraron aproximadamente 1.8 millones de muertes, lo que representa el 12.4% de todos los diagnósticos de cáncer y el 18.7% de las muertes por esta enfermedad <sup>1</sup>. En Colombia, las cifras también son alarmantes, dado que se reportan cerca de 7,196 nuevos casos y 6,124 muertes anuales <sup>10</sup>. Esta elevada mortalidad se debe, en gran medida, a que la mayoría de los casos se detectan en etapas avanzadas, cuando las opciones de tratamiento son limitadas y la supervivencia es menor.

El principal indicador radiológico del CP es la presencia de nódulos pulmonares (NP), pequeñas masas de tejido anormal que pueden formarse en los pulmones debido a diversas causas, incluyendo infecciones, inflamación, procesos neoplásicos, la exposición a sustancias nocivas como el humo de tabaco y la inhalación de asbesto, entre otras <sup>11</sup>. Aunque muchos NP son benignos, alrededor del 33% pueden llegar a ser cancerígenos especialmente en pacientes con antecedentes de tabaquismo. Asimismo, este riesgo de malignidad también puede variar con la edad, ya que el 65% de los NP detectados en pacientes mayores de 50 años resultan malignos, mientras que en los pacientes menores de 35 años esta probabilidad desciende

---

<sup>10</sup> Mejía Ricardo CARDONA DIANA and Serna Fernando VIOLA ANTONIO Rojas Pedro. “Lung cancer in Colombia”. In: *Journal of Thoracic Oncology* 17.6 (2022), pp. 719–727. DOI: [10.1016/j.jtho.2022.02.015](https://doi.org/10.1016/j.jtho.2022.02.015).

<sup>11</sup> AMERICAN CANCER SOCIETY. “Lung Cancer Risk Factors”. In: *American Cancer Society* (2025). <https://www.cancer.org/cancer/lung-cancer/causes-risks-prevention/risk-factors.html> [Accedido: 13-mar-2025].

al 3%, lo que hace crucial su correcta evaluación y seguimiento <sup>12,13</sup>.

La caracterización de los NP no solo requiere del análisis de su estructura, sino también el de su comportamiento metabólico y perfil molecular, lo que proporciona información esencial sobre su naturaleza y agresividad. Para esto, se emplean diversas estrategias de diagnóstico que combinan información anatómica, funcional y molecular. Las imágenes de tomografía computarizada (CT, por sus siglas en inglés) permiten identificar la presencia de nódulos y analizar su morfología, proporcionando detalles sobre características como cavitación, calcificación y localización intrabronquial <sup>14</sup>. Por otro lado, las tomografías por emisión de positrones (PET, por sus siglas en inglés) ofrecen una visión del metabolismo tumoral, facilitando la diferenciación entre lesiones benignas y malignas <sup>15</sup>. La PET puede detectar manifestaciones tempranas de enfermedades al identificar cambios a nivel celular antes que otros exámenes imagenológicos. Finalmente, el análisis de biomarcadores genómicos permite identificar mutaciones específicas asociadas al cáncer de pulmón, lo que no solo contribuye a su caracterización, sino que también ha impulsado el desarrollo de terapias dirigidas para mejorar el tratamiento de la enfermedad <sup>16</sup>.

A continuación, se detallan estas herramientas imagenológicas y su relevancia en la detección y manejo del cáncer de pulmón.

---

<sup>12</sup> Roswell Park Comprehensive Cancer CENTER. *When to Worry About Lung Nodules*. Accedido: 2025-03-19. 2024.

<sup>13</sup> Archivo de BRONCONEUMOLOGÍA. “El diagnóstico y manejo de los nódulos pulmonares en adultos”. In: *Archivos de Bronconeumología* 56.7 (2020). Accedido: 2025-03-19, pp. 452–461.

<sup>14</sup> J. et al. ROBERTS. “Advances in Lung Cancer Imaging: The Role of CT”. in: *Journal of Oncology Imaging* 45 (2023), pp. 120–135. DOI: [10.1016/j.joi.2023.04.003](https://doi.org/10.1016/j.joi.2023.04.003).

<sup>15</sup> A. et al. SMITH. “PET Imaging in Lung Cancer: A Review of Recent Studies”. In: *Lung Cancer Research* 38 (2021), pp. 88–102. DOI: [10.1016/j.lcr.2021.06.007](https://doi.org/10.1016/j.lcr.2021.06.007).

<sup>16</sup> B. et al. JOHNSON. “Genomic Biomarkers in Lung Cancer: Clinical Applications”. In: *Cancer Genomics Journal* 54 (2022), pp. 200–215. DOI: [10.1002/cgj.22045](https://doi.org/10.1002/cgj.22045).

**CT (Tomografía Computarizada).** El análisis de imágenes CT es una herramienta esencial para la detección y diagnóstico temprano del cáncer de pulmón. Permite la identificación de NP y una estimación inicial de la malignidad por medio del estudio de su morfología en aspectos como el tamaño, forma y densidad <sup>17</sup>. Sin embargo, la sensibilidad en la detección de NP con un tamaño reducido (menos de 8 a 10 mm) aún representa un desafío, como se muestra en la Figura 1-a <sup>17</sup>. Adicionalmente, una de las mayores dificultades en el análisis de CT es la diferenciación entre nódulos benignos y malignos, especialmente cuando la morfología del nódulo no es concluyente <sup>18</sup>. Si bien ciertos rasgos, como los bordes espiculados o la presencia de crecimiento acelerado, pueden sugerir malignidad, muchos NP presentan características intermedias que dificultan su clasificación. Algunos nódulos benignos, como los granulomas, pueden imitar la apariencia de un tumor maligno, mientras que ciertos cánceres de crecimiento lento pueden exhibir bordes bien definidos similares a los de nódulos benignos <sup>19</sup>. Esta ambigüedad en la interpretación puede llevar a incertidumbre diagnóstica, lo que hace necesario complementar el análisis estructural con información adicional.

**PET (Tomografía por Emisión de Positrones).** Estas imágenes permiten la evaluación funcional y metabólica de los tejidos, lo que resulta fundamental en la detección, caracterización y seguimiento del cáncer de pulmón. La técnica implica el uso de un análogo de la glucosa marcado radiactivamente, *18F-fluorodesoxiglucosa* (18F-FDG), que es absorbido

---

<sup>17</sup> Geoffrey D. RUBIN. “Lung Nodule and Cancer Detection in Computed Tomography Screening”. In: *Journal of Thoracic Imaging* 30.2 (2015), pp. 130–138. DOI: [10.1097/RTI.000000000000140](https://doi.org/10.1097/RTI.000000000000140).

<sup>18</sup> Martinez-Jaramillo Stella Isabel FERNÁNDEZ-ARRIETA ALEXANDER. “Características clinicopatológicas de nódulos pulmonares: Experiencia en Clínica Reina Sofía, Bogotá, Colombia”. In: *Revista Colombiana de Cirugía* 37 (2022), pp. 49–59. DOI: [10.30944/20117582.903](https://doi.org/10.30944/20117582.903).

<sup>19</sup> Liang Jianye HUANG CUIQING. “Diagnostic Performance of Perfusion Computed Tomography for Differentiating Lung Cancer from Benign Lesions: A Meta-Analysis”. In: *Med Sci Monit* 25 (2019), pp. 3485–3494. DOI: [10.12659/MSM.914206](https://doi.org/10.12659/MSM.914206).

por células metabólicamente activas, como las células cancerosas <sup>20,21</sup>. Los escáneres PET detectan los rayos gamma emitidos por la descomposición de este biomarcador, lo que permite la visualización de áreas con un alto metabolismo de glucosa <sup>20</sup>, como se observa en la Figura 1-b. De hecho, ofrece beneficios como diferenciar eficazmente entre nódulos y masas pulmonares solitarios benignos y malignos, mostrando una precisión superior en comparación con la CT en la estadificación mediastínica del cáncer de pulmón de células no pequeñas <sup>20,21</sup>. Sin embargo, presenta limitaciones. La PET excede la exposición a la radiación que una CT estándar, lo que aumenta el riesgo de cáncer, especialmente en poblaciones vulnerables como mujeres embarazadas, niños, adultos mayores y pacientes con exploraciones repetidas. Debido a esto, el uso de PET con *18F-FDG* no se considera apropiada para la detección del cáncer de pulmón a los niveles de dosis actuales. Además, los escáneres PET son costosos, complejos de adquirir, operar y mantener, lo que limita su uso a centros clínicos especializados <sup>22,23</sup>.

---

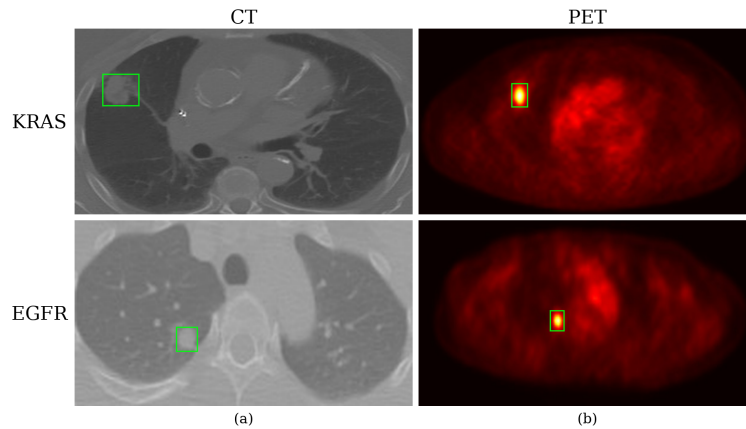
<sup>20</sup> Zhang Honghe WU YIHUA Li Peiwei. “Diagnostic value of fluorine 18 fluorodeoxyglucose positron emission tomography/computed tomography for the detection of metastases in non-small-cell lung cancer patients”. In: *International Journal of Cancer* (2013). DOI: <https://doi.org/10.1002/ijc.27779>.

<sup>21</sup> Vanderveen UNG YEE C. Maziak Donna E. “18 Fluorodeoxyglucose Positron Emission Tomography in the Diagnosis and Staging of Lung Cancer: A Systematic Review”. In: *JNCI: Journal of the National Cancer Institute* (Dec. 2007). DOI: [10.1093/jnci/djm232](https://doi.org/10.1093/jnci/djm232). eprint: <https://academic.oup.com/jnci/article-pdf/99/23/1753/7689425/djm232.pdf>.

<sup>22</sup> Karpinets Tatiana V. SALEHJAHROMI MORTEZA. “Synthetic PET from CT improves diagnosis and prognosis for lung cancer: Proof of concept”. In: *Cell Reports Medicine* (2024). DOI: [10.1016/j.xcrm.2024.101463](https://doi.org/10.1016/j.xcrm.2024.101463).

<sup>23</sup> Avi BEN-COHEN et al. “Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection”. In: *Engineering Applications of Artificial Intelligence* 78 (2019), pp. 186–194.

**Figura 1.** (a) Cortes axiales de la tomografía computarizada (CT) y (b) tomografía por emisión de positrones con  $^{18}F$ -FDG (PET) correspondientes a dos pacientes con cáncer de pulmón. La fila superior muestra un tumor con mutación KRAS, mientras que la fila inferior corresponde a una lesión con mutación EGFR. Se observa la delimitación del nódulo mediante un recuadro verde. Imágenes tomadas del conjunto de datos: *NSCLC Radiogenomics* <sup>24</sup>.



**Mutaciones genómicas.** Las células cancerígenas reportan mutaciones en los genes Receptor del Factor de Crecimiento Epidérmico (EGFR, por sus siglas en inglés) y *Kirsten rat sarcoma viral oncogene homolog* (KRAS, por sus siglas en inglés), los cuales son biomarcadores clave en el desarrollo y progresión del cáncer de pulmón. A continuación, se da una breve descripción de estas mutaciones:

- EGFR es una proteína de transducción de señales que regula el crecimiento y la división de las células. Cuando la secuencia de ADN dentro del gen está mutada (normalmente, entre los exones 18-21), las vías de transducción de señales (como las rutas JAK/STAT, MAPK o PAM) pueden funcionar de manera anormal, dando como resultado una pro-

liferación acelerada de células cancerosas y la formación de tumores<sup>25,26,27</sup>. Además, su incidencia es aproximadamente tres veces mayor en no fumadores que en fumadores, y en mujeres en comparación con hombres<sup>28</sup>.

- KRAS codifica una proteína GTPasa que actúa dentro de la célula como un mediador para algunas vías de señalización, transmitiendo señales desde EGFR hasta el núcleo. En condiciones normales, KRAS se activa cuando se une a GTP y se desactiva cuando lo hidroliza a GDP. Sin embargo, las mutaciones en KRAS (especialmente en los codones 12, 13 y 61) bloquean su capacidad de apagarse, lo que mantiene activas permanentemente las vías de señalización como MAPK y PAM, favoreciendo el crecimiento descontrolado de las células cancerosas, como se puede ver en la Figura 2<sup>25,4,29</sup>. Las mutaciones en KRAS están estrechamente relacionadas con el tabaquismo (presente en aproximadamente el 30% de los casos)<sup>30</sup>.

---

<sup>25</sup> Seonhwa et al. KIM. “Deep learning–radiomics integrated noninvasive detection of epidermal growth factor receptor mutations in non-small cell lung cancer patients”. In: *Scientific Reports* 14.1 (2024), p. 922.

<sup>26</sup> Sobti Ranbir Chander SHARMA PARUL Sharma Siddharth. “Molecular Pathogenesis of Lung Cancer: Evolving Landscapes Based on Past Experience”. In: *Molecular Biomarkers for Cancer Diagnosis and Therapy*. Springer, 2024, pp. 347–380.

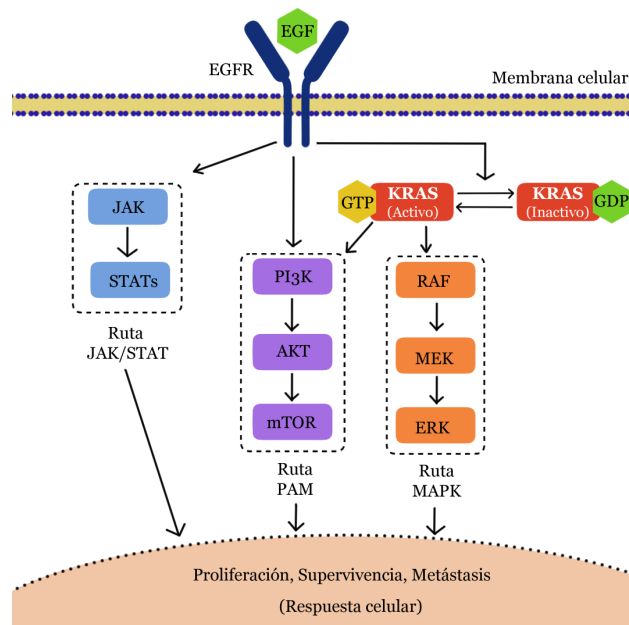
<sup>27</sup> Galdadas Ioannis IYER R SUMANTH Needham Sarah R. “Drug-resistant EGFR mutations promote lung cancer by stabilizing interfaces in ligand-free kinase-active EGFR oligomers”. In: *Nature Communications* 15.1 (2024), p. 2130.

<sup>28</sup> Del Re Marzia LEVANTINI ELENA Maroni Giorgia. “EGFR signaling pathway as therapeutic target in human cancers”. In: *Seminars in cancer biology*. Vol. 85. Elsevier. 2022, pp. 253–275.

<sup>29</sup> Jiao LIU; Rui KANG, and Daolin TANG. “The KRAS-G12C inhibitor: activity and resistance”. In: *Cancer gene therapy* 29.7 (2022), pp. 875–878.

<sup>30</sup> M et al. RECK. “Targeting KRAS in non-small-cell lung cancer: recent progress and new approaches”. In: *Annals of Oncology* 32.9 (2021), pp. 1101–1110.

**Figura 2.** Algunas vías de transducción relacionadas con EGFR y KRAS, un error dentro de estos componentes podrían causar el crecimiento de tumores cancerígenos.



No obstante, la identificación de las mutaciones mencionadas requieren biopsias tumorales. Por ello, existen limitaciones presentes dadas las complicaciones médicas del paciente o por el contrario, la ubicación del tumor <sup>31,32</sup>. Es por esto, que han emergido tecnologías tales como la radiogenómica, permitiendo relacionar mutaciones genómicas de acuerdo a la caracterización radiológica, logrando el desarrollo de métodos menos invasivos para guiar el tratamiento <sup>33</sup>. Por otro lado, la caracterización de mutaciones genómicas como EGFR y KRAS mediante

- 
- <sup>31</sup> Qinghua Zhou et AL. “The Society for Translational Medicine: indications and methods of percutaneous transthoracic needle biopsy for diagnosis of lung cancer”. In: *Journal of Thoracic Disease* 10.9 (2018).
- <sup>32</sup> Tian Yijun SUN WEI Yuan Xun. “Non-invasive approaches to monitor EGFR-TKI treatment in non-small-cell lung cancer”. In: *Journal of hematology & oncology* 8 (2015), pp. 1–9.
- <sup>33</sup> Buscarino Valentina RIZZO STEFANIA Petrella Francesco. “CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer”. In: *European radiology* 26 (2016), pp. 32–42.

imágenes radiológicas ha sido abordada en diversos estudios <sup>33,34,7,35</sup>. En particular, se han observado patrones distintivos en imágenes CT y PET, los cuales podrían ser útiles para la clasificación no invasiva de estos genotipos.

**Fenotipos de mutaciones genómicas.** Dentro de los fenotipos que pueden ser observados en imágenes CT, en el caso de mutaciones EGFR, se encuentran características como la ausencia de fibrosis y enfisema, el cual es un indicador de pulmones sin cicatrices anormales o daños en los alvéolos. Asimismo, también se puede identificar la presencia de broncograma aéreo, en donde se evidencian estructuras tubulares o ramificadas de aire dentro de una lesión pulmonar, así como la retracción pleural, que se caracteriza por el desplazamiento de la pleura visceral hacia un nódulo. Por otra parte, las mutaciones KRAS, se encuentra en nódulos cuya morfología es mayormente circular o en aquellos con presencia de enfisema, evidenciado por el daño en los sacos de los pulmones o alvéolos, como se pueden observar en la Figura 3.

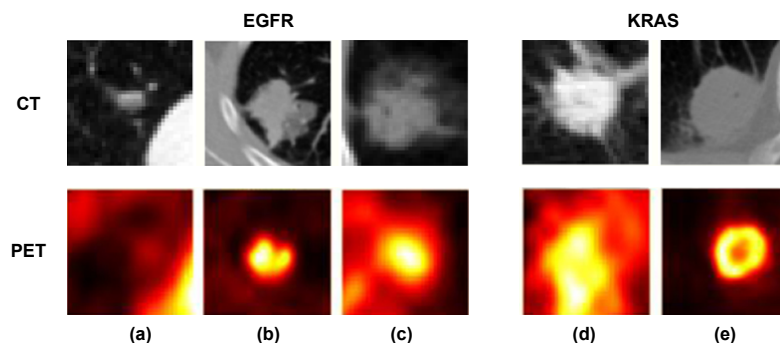
En el caso de las imágenes PET, el SUVmax (*Standardized Uptake Value maximum*) es un parámetro utilizado para cuantificar la actividad metabólica de tejidos o lesiones. Los tumores con mutaciones EGFR, suelen presentar un SUVmax más bajo que los pacientes que no lo presentan, debido a una menor actividad glucolítica. Sin embargo, en KRAS, los estudios muestran resultados variables: Algunos no encuentran correlación, mientras que otros reportan que tienen valores más altos de SUVmax en comparación a los tumores EGFR mutados, debido a su alta actividad metabólica y agresividad, como se muestra en la Figura 3 <sup>7,35</sup>.

---

<sup>34</sup> de Jong Evelyn EC RIZZO STEFANIA Raimondi Sara. “Genomics of non-small cell lung cancer (NSCLC): Association between CT-based imaging features and EGFR and K-RAS mutations in 122 patients—An external validation”. In: *European journal of radiology* 110 (2019), pp. 148–155.

<sup>35</sup> Lozano Maria Dolores CAICEDO CARLOS Garcia-Velloso Maria Jose. “Role of [18 F] FDG PET in prediction of KRAS and EGFR mutation status in patients with advanced non-small-cell lung cancer”. In: *European journal of nuclear medicine and molecular imaging* 41 (2014), pp. 2058–2065.

**Figura 3.** Fenotipos observados según la mutación en la modalidad radiológica. En CT, los nódulos con EGFR puede presentar ausencia de fibrosis y enfisema (a), broncograma aéreo (b), retracción pleural (c). Mientras que los presentados en KRAS son redondos (d) y pueden presentar enfisema (e). Imágenes tomadas del conjunto de datos: *NSCLC Radiogenomics* <sup>36</sup>.



## 1.2. ESTRATEGIAS COMPUTACIONALES PARA LA DETECCIÓN Y LA CLASIFICACIÓN

**1.2.1. Modelos de detección.** Los modelos de detección de objetos basados en redes neuronales convolucionales (CNN) han logrado detecciones rápidas y precisas en imágenes. Uno de los enfoques más utilizados es *You Only Look Once* (YOLO, por sus siglas en inglés) como se observa en la Figura 4, el cual permite la predicción de múltiples objetos en una imagen mediante un único paso a través de la red neuronal, a diferencia de los métodos tradicionales que dependen de múltiples etapas para proponer regiones y luego clasificarlas <sup>37</sup>.

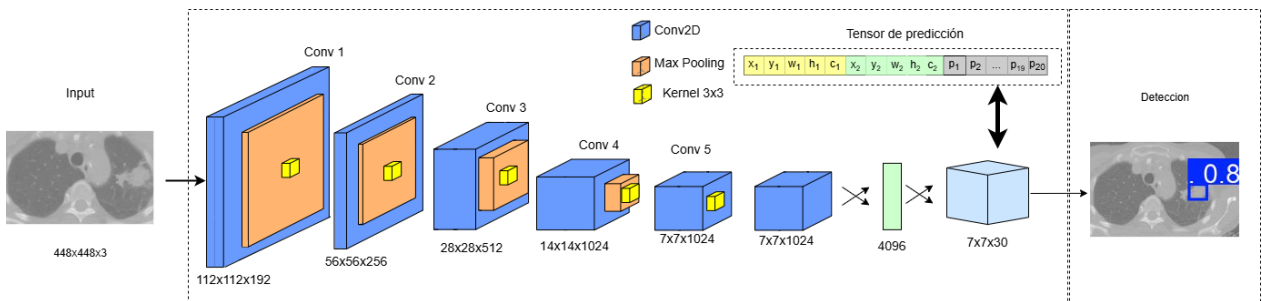
La arquitectura *YOLO* estándar divide la imagen en una cuadrícula de tamaño  $S \times S$ , asignando a cada celda la tarea de predecir un número determinado de cajas delimitadoras (*bounding boxes*, en inglés) junto con su respectiva puntuación de confianza. Logrando un aumento significativo en la velocidad de detección en comparación con modelos anteriores

---

<sup>37</sup> Girshick Ross REDMON JOSEPH Divvala Santosh. “You Only Look Once: Unified, Real-Time Object Detection”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 779–788.

basados en propuesta de regiones. Sin embargo, estos métodos presentan algunas limitaciones, especialmente en la detección de objetos pequeños o en condiciones de alta variabilidad (calidad de la imagen, ruido, entre otros) <sup>38</sup>.

**Figura 4.** Esquema de la arquitectura *YOLOv1*, donde la imagen de entrada se procesa mediante una serie de capas convolucionales hasta generar un tensor de predicción que contiene las coordenadas de las cajas delimitadoras, la puntuación de confianza y las clases detectadas, todo en un único paso por la red.

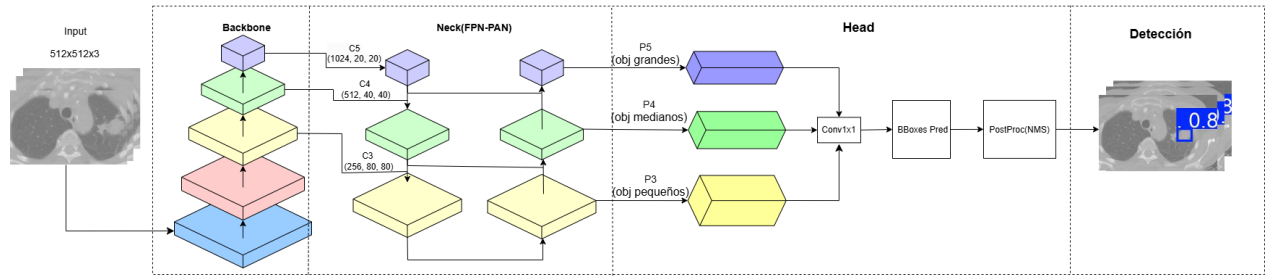


A medida que la *YOLO* ha evolucionado, se han agregado mejoras para la extracción de características, la representación de objetos en diferentes escalas y la precisión en la localización. En particular, la implementación de *CSPDarknet53* como módulo extractor de características (*backbone*) logra aumentar la eficiencia computacional sin comprometer la capacidad de detección, gracias a que divide y fusiona el flujo de características, lo que reduce la redundancia y el costo de cómputo sin perder riqueza representacional. Estas características son procesadas posteriormente por el cuello (*neck*), compuesto usualmente por una *Red piramidal de características* (FPN, por sus siglas en inglés) o una *Red de Agregación de Rutas* (PAN, por sus siglas en inglés). Esta etapa combina información de diferentes niveles de la red, permitiendo mejorar la detección de objetos de distintos tamaños al combinar capas con alta

<sup>38</sup> Joseph REDMON and Ali FARHADI. “YOLO9000: Better, Faster, Stronger”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 7263–7271.

información semántica y espacial <sup>39</sup>, como se muestra en la Figura 5.

**Figura 5.** Representación simplificada de *YOLOv5*, en la que se destacan los bloques funcionales principales: Backbone (CSPDarknet53), Neck (PAN) y Head. Esta arquitectura optimiza la extracción de características y la fusión de información multiescala, mejorando la precisión en la detección de objetos de distintos tamaños.



Finalmente, la cabeza de detección (*head*) se encarga de predecir las coordenadas de las *bounding boxes*, los cuales son representados mediante los parámetros  $[x, y, w, h]$ , donde  $[x, y]$  determina la ubicación del centro de la caja, y  $[w, h]$  representan su ancho y alto, normalizados con respecto a la imagen. Además, cada *bounding box* asigna una puntuación de confianza que indica la probabilidad de que la detección sea correcta <sup>40</sup>. Durante el entrenamiento, se minimiza una función de pérdida compuesta que penaliza los errores en la regresión de las coordenadas y en la predicción de la confianza, como se indica en las ecuaciones 1 y 2, respectivamente. El término de confianza regula la probabilidad asignada a cada detección,

<sup>39</sup> Alexey BOCHKOVSKIY and Chien-Yao WANG. “YOLOv4: Optimal Speed and Accuracy of Object Detection”. In: *arXiv preprint arXiv:2004.10934* (2020).

<sup>40</sup> Glenn JOCHER; Alex STOKEN, and CHAURASIA. *YOLOv5: Implementation of YOLO for object detection*. GitHub repository. 2021.

maximizándola cuando hay un objeto presente y minimizándola en su ausencia.

$$\begin{aligned} \mathcal{L}_{\text{coord}} = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{\text{obj}} \left[ (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \end{aligned} \quad (1)$$

$$\mathcal{L}_{\text{conf}} = \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \quad (2)$$

Para mejorar la precisión de detección, se han desarrollado métricas y funciones de pérdida como *Intersection over Union (IoU)* y su variante *Complete IoU (CIoU)*, que no solo evalúan el solapamiento entre la caja predicha y la caja real, sino que también optimizan la orientación y la relación de aspecto de las detecciones <sup>41</sup>.

Una vez generadas las predicciones, se aplican técnicas de posprocesamiento para mejorar los resultados. Una de las más utilizadas es *Non-Maximum Suppression (NMS)*, la cual elimina las detecciones redundantes manteniendo solo la predicción con la puntuación de confianza más alta en cada región. Además, se establecen umbrales de confianza para filtrar predicciones irrelevantes y reducir falsos positivos <sup>42</sup>.

### 1.2.2. Modelos de clasificación

**Modelo de clasificación supervisado** Un modelo de clasificación supervisado se utiliza para predecir la categoría a la que pertenece una entrada, basándose en ejemplos previos con etiquetas conocidas. El modelo se entrena con un conjunto de datos etiquetado para aprender

---

<sup>41</sup> Zhaohui ZHENG et al. “Distance-IoU loss: Faster and better learning for bounding box regression”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. 07. 2020, pp. 12993–13000.

<sup>42</sup> Axel NEUBECK and Luc VAN GOOL. “Efficient Non-Maximum Suppression”. In: *Proceedings of the International Conference on Pattern Recognition (ICPR)*. IEEE, 2006, pp. 850–855.

una función  $f : X \rightarrow Y$  que minimice el error de clasificación en datos no vistos, donde  $X^{CT}$  es el conjunto de imágenes CT y  $Y = \{EGFR, KRAS\}$  las etiquetas de la clase <sup>43,44</sup>. Normalmente este tipo de arquitecturas solo reciben un solo tipo de entrada para realizar una predicción, sin embargo, en muchos casos, la información que proviene de una sola fuente no es suficiente para lograr una clasificación precisa. En el contexto médico, la integración de diversos exámenes y/o características de un paciente pueden ayudar a discernir con mayor detalle el diagnóstico de este, por ejemplo, la unión de imágenes CT con datos clínicos han demostrado mejorar significativamente el rendimiento de los modelos de clasificación en el estado del arte. Este tipo de enfoque que toma diversas fuentes de información para predecir una etiqueta es conocido como aprendizaje multimodal supervisado.

**Modelos de clasificación supervisado multimodal** Dado que los dos tipos de imágenes radiológicas se complementan y se relacionan a través de fenotipos para la identificación de mutaciones genómicas, es necesario implementar estrategias de fusión multimodales que permitan explotar esa complementariedad que presentan estos datos radiológicos para mejorar la precisión en la clasificación de las mutaciones genómicas, facilitando la identificación de patrones radiogenómicos. Sean  $X_i^{CT} = \{x_1, \dots, x_i, \dots, x_N\}$  el conjunto de imágenes de CT y  $X_j^{PET} = \{x_1, \dots, x_j, \dots, x_N\}$  el conjunto de imágenes PET, ambos con una cantidad  $N$  de datos, busca proyectar ambas modalidades a un espacio común mediante vectores de características  $n$ -dimensionales, siendo estos vectores  $Z_i^{CT}$  y  $Z_j^{PET}$  para CT y PET respectivamente. Posteriormente, se realiza la concatenación de dichos vectores de características ( $Z_i^{CT} \oplus Z_j^{PET}$ ), asegurando que estén asociados a una misma etiqueta de clasificación

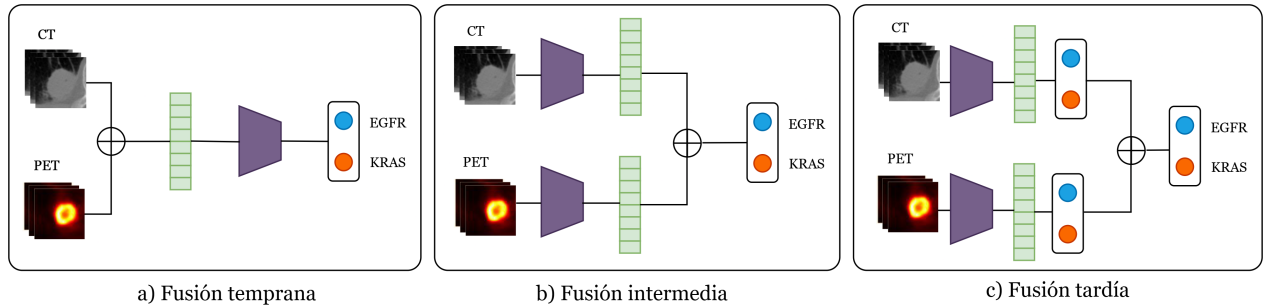
---

<sup>43</sup> Dolores Romero Morales EMILIO CARRIZOSA. “Supervised classification and mathematical optimization”. In: *Computers & Operations Research* 40 (2013), pp. 150–165. DOI: [10.1016/j.cor.2012.05.015](https://doi.org/10.1016/j.cor.2012.05.015).

<sup>44</sup> Mitadru Ghosh PRATAP CHANDRA SEN Mahimarnab Hajra. “Supervised Classification Algorithms in Machine Learning: A Survey and Review”. In: *Emerging Technology in Modelling and Graphics*. Ed. by J. K. MANDAL and D. BHATTACHARYA. Vol. 937. Advances in Intelligent Systems and Computing. Springer Nature Singapore Pte Ltd., 2020, pp. 99–109. DOI: [10.1007/978-981-13-7403-6\\_11](https://doi.org/10.1007/978-981-13-7403-6_11).

$Y = \{y_1, \dots, y_N\}$ , la cual define a qué tipo de mutación pertenece. Existen principalmente tres tipos de estrategias para la fusión multimodal: La fusión temprana, intermedia y tardía, estas definiciones varían según el punto de la arquitectura en el que se realiza la integración de los datos, como se ilustra en la Figura 6 <sup>45</sup>.

**Figura 6.** Diferentes estrategias para realizar clasificación multimodal mediante fusión de características.



- **Fusión temprana:** Esta fusión consiste en combinar las características de las diferentes modalidades CT y PET antes de entrenar el modelo. Para esto, se concatenan los vectores de características extraídos de ambas modalidades en un único vector de entrada, como se observa en la Figura 6a. De esta forma se puede capturar patrones compartidos en una única representación si existe una fuerte correlación entre las características.
- **Fusión intermedia:** En esta estrategia se extraen las representaciones embebidas de cada modalidad después de haber pasado los datos por una arquitectura de aprendizaje profundo, estas representaciones latentes se integran antes de ser clasificadas, como se puede ver en la Figura 6b. De esta manera, los modelos aprenden características relevantes para cada modalidad de forma independiente antes de la fusión, lo que permite

<sup>45</sup> Said Yacine BOULAHIA et al. “Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition”. In: *Machine Vision and Applications* 32.6 (2021), p. 121.

capturar relaciones no lineales más complejas y preservar la independencia de cada modalidad hasta la integración final antes de la clasificación.

- **Fusión tardía:** En esta fusión se entrena una red de aprendizaje profundo para cada tipo de modalidad y se integran las predicciones realizadas por estos modelos para ser posteriormente procesados en una capa final de clasificación, como se puede apreciar en la Figura 6c. Esta estrategia permite que cada modelo se enfoque en los patrones específicos de cada modalidad, similar a la intermedia. Sin embargo, al concatenar las predicciones, resulta útil cuando las modalidades son más complementarias que correlacionadas.

**Modelo de clasificación autosupervisado** Un modelo de clasificación autosupervisado es un enfoque de aprendizaje profundo que se entrena utilizando datos no etiquetados para generar representaciones latentes significativas, buscando crear sus propias etiquetas a partir de propiedades inherentes de los datos a diferencia de otros modelos supervisados. Estas representaciones son aprendidas en una tarea auxiliar obligando al modelo a capturar patrones latentes de los datos y posteriormente emplearlos en tareas de clasificación <sup>46,47</sup>.

**Modelos de síntesis** Los modelos de síntesis desempeñan un papel crucial en diversas áreas. Para el ámbito médico y radiológico sirven como marcos fundamentales que facilitan la creación y optimización de datos, imágenes basados en parámetros cualitativos y cuantitativos. En esencia, los modelos de síntesis se centran en cómo se puede integrar información de manera coherente mejorando las capacidades diagnósticas mediante una mejor calidad y precisión de la imagen. Existen dos estrategias destacadas para la traducción entre imágenes

---

<sup>46</sup> Saeed SHURRAB and Rehab DUWAIRI. “Self-supervised learning methods and applications in medical imaging analysis: a survey”. In: *PeerJ Computer Science* (2022), e1045. DOI: [10.7717/peerj-cs.1045](https://doi.org/10.7717/peerj-cs.1045).

<sup>47</sup> Et al. HUANG Shih-Cheng. “Self-supervised learning for medical image classification: a systematic review and implementation guidelines”. In: *NPJ Digital Medicine* (2023). DOI: [10.1038/s41746-023-00811-0](https://doi.org/10.1038/s41746-023-00811-0).

desde conjuntos de datos con diferentes configuraciones.

## Modelos pareados

- **Redes adversariales generativas**

Las redes adversariales generativas (GANs, por sus siglas en inglés) es una estrategia de aprendizaje profundo que emplea un enfoque de aprendizaje no supervisado basado en la competencia entre dos redes neuronales, que se entrenan en conjunto: Un generador ( $G$ ) crea muestras transformando un vector aleatorio ( $Z$ ) en datos sintéticos que imitan las distribuciones de datos reales. Por su parte, el discriminador ( $D$ ) evalúa estas muestras comparándolas con datos reales, distinguiendo entre resultados reales y sintéticos <sup>48</sup>. El objetivo es encontrar un equilibrio en el que la red generadora produzca datos tan realistas que la red discriminadora no pueda diferenciarlos de los datos reales, como se puede apreciar en la Figura 7.

La dinámica de entrenamiento se basa en el generador y el discriminador, cada una regida por sus respectivas funciones de pérdida. La pérdida del discriminador generalmente está diseñada para medir la diferencia entre las distribuciones de datos reales y generados, manteniendo la disponibilidad del gradiente para el generador <sup>49</sup>. Por otro lado, la pérdida del generador está intrínsecamente ligada al rendimiento del discriminador. Este utiliza la retroalimentación del discriminador para modificar sus ponderaciones y mejorar la calidad de las muestras generadas <sup>50</sup>.

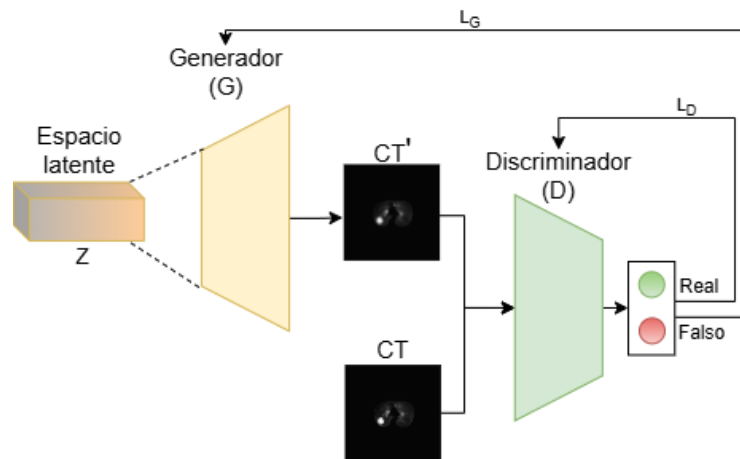
---

<sup>48</sup> Ian et al. GOODFELLOW. “Generative adversarial networks”. In: *Communications of the ACM* 63.11 (2020), pp. 139–144.

<sup>49</sup> Felix JUEFEI-XU; Vishnu Naresh BODDETI, and Marios SAVVIDES. “Gang of GANs: Generative Adversarial Networks with Maximum Margin Ranking”. In: *arXiv preprint arXiv:1704.04865* (2017). Accessed: 2025-05-01.

<sup>50</sup> Ward Tomás E WANG ZHENGWEI She Qi. “Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy”. In: *ACM Computing Surveys (CSUR)* 54.2 (2020), pp. 1–38. DOI: [10.1145/3439723](https://doi.org/10.1145/3439723). arXiv: [1906.01529](https://arxiv.org/abs/1906.01529) [cs.CV].

**Figura 7.** Esquema de una red generativa adversaria (GAN) diseñada para generar nódulos pulmonares sintéticos a partir de ruido aleatorio y discriminar entre los nódulos reales y los creados por la red.<sup>51</sup>



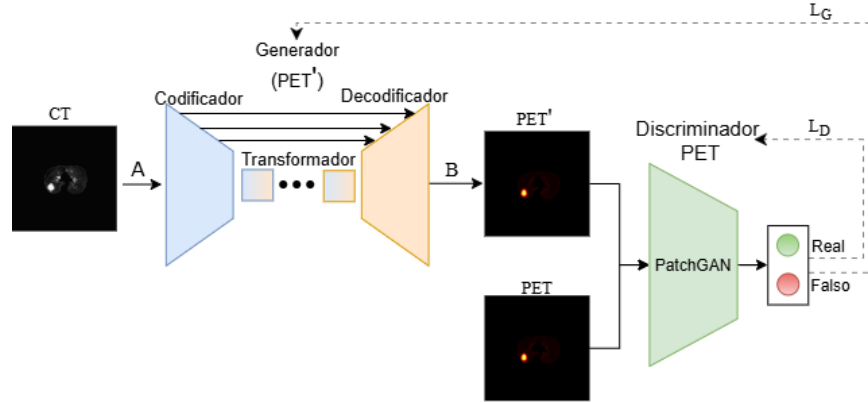
- **Pix2Pix**

Pix2Pix es un modelo de traducción de imagen a imagen basado en GANs condicionales mediante un aprendizaje supervisado de pares de imágenes. Esta aprende a convertir una imagen de entrada en otra de salida, aplicando una función de mapeo que traduce una representación visual en otra, lo que es especialmente relevante en la medicina, donde se requiere fusionar información de diferentes modalidades para obtener diagnósticos más precisos.

Este modelo tiene dos componentes principales. El generador  $G$ , se basa en una arquitectura  $U$ -Net, compuesto por un codificador con capas convolucionales que extraen características a la vez que reducen la dimensionalidad y un decodificador que implementa convoluciones transpuestas que incrementan la resolución espacial y reduce la dimensionalidad, a su vez, incorpora conexiones de saltos entre el codificador y el decodificador, siendo apreciable en la Figura 8, preservando detalles esenciales de la imagen. El discriminador  $D$ , emplea una arquitectura GAN que penaliza la estructura a escala de parches en imágenes (*PatchGAN*, por sus siglas en inglés), que evalúa pequeños parches de imagen tomando estructuras más locales. No obstante, para poder realizar

la tarea de traducción entre dominios, es necesario que los datos estén emparejados <sup>52</sup>.

**Figura 8.** *Pix2Pix* aplicado a la transferencia de imágenes pulmonares. El generador, recibe una imagen de entrada y genera una imagen similar a la imagen objetivo. Un discriminador evalúa la imagen generada con la real para y determina si es auténtica o sintética, por ultimo la perdida es calculada para entrenar el modelo <sup>53</sup>.



El entrenamiento del modelo se basa principalmente en dos funciones de pérdida: La pérdida adversarial condicional,  $\mathcal{L}_{cGAN}$ , esta función incentiva al generador a producir imágenes que el discriminador no pueda distinguir de las reales (3) y la pérdida de reconstrucción,  $\mathcal{L}_{L1}$  mide la diferencia promedio absoluta entre la imagen generada y la real, promoviendo que el generador produzca imágenes similares a las reales (4). Definidas matemáticamente como <sup>54,52</sup>:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \quad (3)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[||y - G(x, z)||_1] \quad (4)$$

<sup>52</sup> Han-Seop SONG; Jophous MUGABI, and Jae-Ho JEONG. “Pix2Pix and Deep Neural Network-Based Deep Learning Technology for Predicting Vortical Flow Fields and Aerodynamic Performance of Airfoils”. In: *Applied Sciences* 13.2 (2023), p. 1019. DOI: [10.3390/app13021019](https://doi.org/10.3390/app13021019).

<sup>54</sup> Alharbe Nawaf ALJOHANI ABEER. “Generating Synthetic Images for Healthcare with Novel Deep Pix2Pix GAN”. in: *Electronics* 11.21 (2022), p. 3470. DOI: [10.3390/electronics11213470](https://doi.org/10.3390/electronics11213470).

## Modelos no pareados

- **CycleGAN** *CycleGAN* es una arquitectura de aprendizaje profundo para la traducción entre dominios con datos no emparejados para su entrenamiento. Ideal para casos en los que obtener pares entre dominios resulta complicado. La arquitectura *CycleGAN*, a diferencia de la GAN tradicional y como se observa en la Figura 9, está compuesta por dos generadores y dos discriminadores. Los cuales por medio de un generador realizan el proceso de transformar una imagen de un dominio  $X$  a un dominio  $Y$ ,  $G : X \rightarrow Y$ , mientras que el generador  $F : Y \rightarrow X$  realiza el proceso contrario, pasando la imagen del dominio  $Y$  al  $X$ . A su vez el discriminador  $D_X$  distingue la imagen del dominio  $X$  original de las generadas por el generador  $F$ . De igual manera  $D_Y$  realiza el mismo proceso pero para las imágenes del dominio  $Y$  el generador  $G : X \rightarrow Y$ <sup>55,56</sup>.

### 1.3. DETECCIÓN Y CLASIFICACIÓN DE BIOMARCADORES DE CÁNCER DE PULMÓN

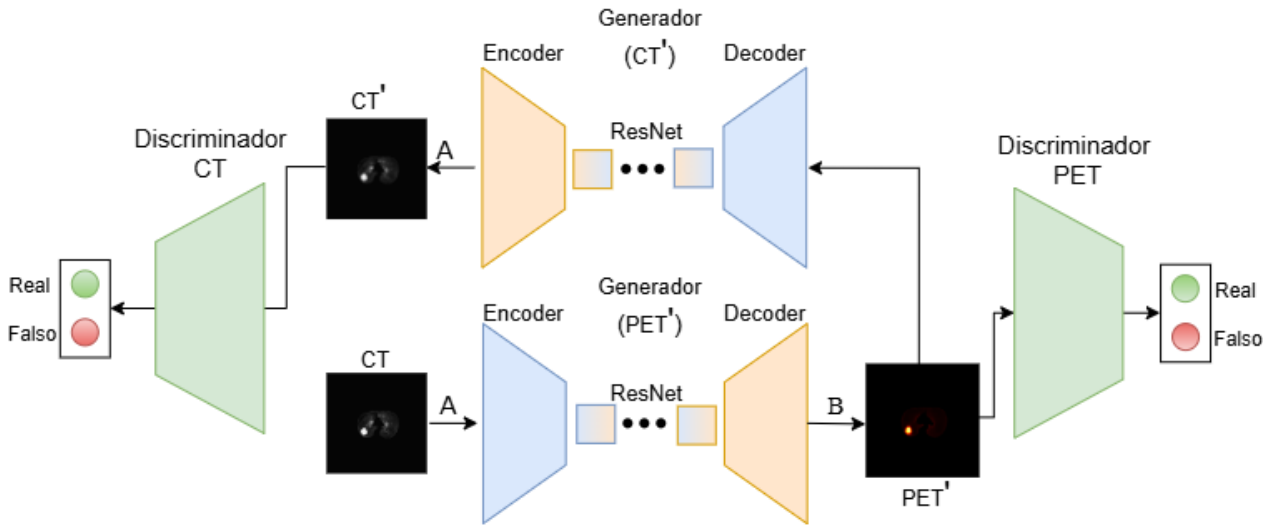
**1.3.1. Métodos de detección** En la literatura se han propuesto diversos enfoques que utilizan exclusivamente imágenes de tomografía computarizada (CT) debido a su amplia disponibilidad y detalle anatómico. Un ejemplo es la adaptación del modelo *YOLOv3*, que permite realizar detecciones rápidas, pero con una tasa alta de falsos positivos, particularmente en nódulos de menor tamaño (diámetro menor a 5 mm)<sup>5</sup>. Para reducir este problema, posteriormente se propuso una versión mejorada de *YOLOv5* mediante la integración de módulos *Transformers* y técnicas avanzadas de fusión cruzada de características, incrementando

---

<sup>55</sup> Tinghui Zhou PHILLIP ISOLA Jun-Yan Zhu. “Image-to-Image Translation with Conditional Adversarial Networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 1125–1134. DOI: [10.1109/CVPR.2017.632](https://doi.org/10.1109/CVPR.2017.632).

<sup>56</sup> Keisuke et al. USUI. “A cycle generative adversarial network for improving the quality of four-dimensional cone-beam computed tomography images”. In: *Radiation Oncology* 17.1 (2022), p. 69.

**Figura 9.** Marco de trabajo de la red generativa antagónica *CycleGAN* y estructura de red del generador y el discriminador. El modelo de entrenamiento consta de dos generadores y dos discriminadores. Para entrenar *CycleGAN*, se mejora el rendimiento general de la red a través de redes que actúan de forma bidireccional entre sí.



la precisión en la identificación de los nódulos <sup>57</sup>. También, Mammeri et al. combinaron el modelo *YOLOv7* con una red convolucional VGG16, permitiendo la detección y clasificación de los nódulos en categorías benignas, sospechosas o malignas <sup>58</sup>. De manera similar, Wu et al. desarrollaron otra variante de *YOLOv7*, enfocada especialmente en la detección de nódulos pequeños mediante un módulo multiescala específico <sup>6</sup>. En esta misma línea, Sun et al. propusieron un detector de dos etapas basado en pirámides de características e imágenes en 3D, complementado con una estrategia de *curriculum learning* que acelera significativamente el entrenamiento y mejora la sensibilidad en la detección de nódulos pequeños en conjuntos

<sup>57</sup> Y. ZHOU et al. “Improved YOLOv5 with Transformers for pulmonary nodule detection”. In: *Scientific Reports* 14 (2024), p. 1234.

<sup>58</sup> A. MAMMERI et al. “YOLOv7 and VGG16-based classification of pulmonary nodules”. In: *Scientific Reports* 14 (2024), p. 5678.

como LUNA16 y LIDC-IDRI <sup>59</sup>.

Sin embargo, la mayoría de estos métodos se limitan al uso de CT, desaprovechando la información funcional que aporta la modalidad PET, fundamental para caracterizar la actividad metabólica de los nódulos. Adicionalmente, incluso los enfoques más recientes tienden a generar una alta tasa de falsos positivos, lo que ha motivado el desarrollo de módulos especializados de reducción de falsos positivos. Ejemplos representativos incluyen la propuesta de Hao et al., basada en una red convolucional con atención central en escenarios de datos desbalanceados <sup>60</sup>, y el trabajo clásico de Setio et al., que introdujo redes convolucionales multivista para este mismo fin <sup>61</sup>.

En paralelo, algunos estudios abordan la combinación de modalidades de imagen PET y CT, aprovechando la información anatómica y metabólica para mejorar la precisión diagnóstica. Por ejemplo, Sibille et al. desarrollaron una red profunda multimodal que integra ambas modalidades para detectar lesiones pulmonares <sup>62</sup>. Asimismo, Chen et al. propusieron un modelo basado en la arquitectura CenterNet con mecanismos de atención multimodal, logrando una mejora significativa en la sensibilidad de detección, aunque con un coste computacional elevado <sup>63</sup>. También, Zia UrRehman et al. presentaron una CNN profunda con

---

<sup>59</sup> Benyuan SUN et al. “Integrating Feature and Image Pyramid: A Lung Nodule Detector Learned in Curriculum Fashion”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer. 2018, pp. 683–691.

<sup>60</sup> K. HAO et al. “Lung nodule false positive reduction using a central attention convolutional neural network on imbalanced data”. In: *Proceedings of SPIE Medical Imaging*. Vol. 12466. 2023, p. 124661X. DOI: [10.1117/12.2654216](https://doi.org/10.1117/12.2654216).

<sup>61</sup> A. A. A. SETIO et al. “Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks”. In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1160–1169. DOI: [10.1109/TMI.2016.2536809](https://doi.org/10.1109/TMI.2016.2536809).

<sup>62</sup> L. SIBILLE et al. “Multimodal deep learning for lung lesion detection in PET/CT”. in: *Journal of Nuclear Medicine* 61.4 (2020), pp. 521–527.

<sup>63</sup> H. CHEN et al. “CenterNet-based multimodal attention for lung nodule detection”. In: *European Radiology* 32 (2022), pp. 1234–1243.

mecanismos duales de atención espacial y de canal, permitiendo capturar características morfológicas detalladas de los nódulos pulmonares en imágenes de CT <sup>64</sup>.

**1.3.2. Métodos de clasificación** En la literatura existen diversos enfoques para los modelos de clasificación supervisados, estos se centran principalmente en predecir el estado de la mutación en EGFR empleando imágenes CT. Empezando con las estrategias unimodales, algunos autores han optado por esta opción para identificar qué tipos de datos son más relevantes en la clasificación, evaluando características clínicas del paciente, anotaciones semánticas del nódulo e imágenes del nódulo, como realizó de Shuo Wang *et al.*, que compara los diferentes tipos de datos en modelos separados utilizando arquitecturas densas y convolucionales respectivamente <sup>65</sup>. Bajo esta misma línea de trabajo, Francisco Silva *et al.* propusieron un enfoque basado en autoencoders convolucionales para extraer características en regiones específicas del parénquima pulmonar, enfocándose en diferentes perspectivas, del pulmón que solo contiene al nódulo o solo a nivel de parche con su tejido alrededor, y ambos pulmones dentro de la imagen CT <sup>66</sup>, mientras que Wei Zhao *et al.* introdujeron una red basada en transformers con aprendizaje débilmente supervisado para la clasificación de EGFR en imágenes CT <sup>67</sup>.

Otros trabajos han combinado machine learning y deep learning como estrategia híbrida. Jiangran Wu *et al.* integraron ambas metodologías para detectar pacientes con mutación

---

<sup>64</sup> Zia URREHMAN *et al.* “Dual attention CNN for pulmonary nodule detection in CT”. in: *European Radiology* 33 (2023), pp. 4567–4575.

<sup>65</sup> Ye Zhaoxiang WANG SHUO Shi Jingyun. “Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning”. In: *European Respiratory Journal* 53.3 (2019).

<sup>66</sup> Morgado Joana SILVA FRANCISCO Pereira Tania. “EGFR assessment in lung cancer CT images: analysis of local and holistic regions of interest using deep unsupervised transfer learning”. In: *IEEE Access* 9 (2021), pp. 58667–58676.

<sup>67</sup> Li Ge ZHAO WEI Chen Weidao. “GMILT: a novel transformer network that can noninvasively predict EGFR mutation status”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).

EGFR que solamente se encuentran en etapas tempranas del cáncer de pulmón <sup>68</sup>, y de manera similar, Seonhwa et al. implementa estas estrategias adicionándole los datos clínicos de los pacientes, fortaleciendo el modelo y sus resultados <sup>25</sup>.

Por otro lado, se han desarrollado enfoques multimodales que combinan diferentes modalidades como imágenes CT y PET. En el caso exclusivo de EGFR, Jay Kumar et al. emplearon las características radiómicas extraídas de la segmentación de nódulos de imágenes PET y CT para clasificar los subtipos de esa mutación <sup>69</sup>. Por otra parte, para abordar tanto EGFR como KRAS, Isaac *et al.* fusionaron información de ambas modalidades CT y PET en un modelo de machine learning, enfatizando en la armonización de características radiómicas para facilitar la generalización entre distintos centros médicos, sin embargo, dependen de la segmentación de los nódulos, lo cual implica la pérdida de información de fenotipos potenciales como la presencia o ausencia de enfisema y fibrosis <sup>8</sup>. Así mismo, este enfoque multimodal no solo se centra entre imágenes médicas, la integración de datos clínicos y genómicos ha sido explorada para enriquecer la clasificación de mutaciones. Mu Zhou *et al.* propusieron un mapa que asocia imágenes CT con datos genómicos para mejorar la identificación de patrones fenotípicos <sup>70</sup>. Satvik Tripathi *et al.*, aunque requirieron secuencias de RNA de biopsias, demostraron que la inclusión de estos datos, con los datos clínicos en conjunto con imágenes CT mejora significativamente los resultados. No obstante, el uso de datos tabulares con registros faltantes pueden inducir sesgos si se realiza algún tipo de imputación, asimismo, el uso del

---

<sup>68</sup> Zhou Lin WU JINGRAN Meng Hao. “Habitat radiomics and deep learning fusion nomogram to predict EGFR mutation status in stage I non-small cell lung cancer: a multicenter study”. In: *Scientific Reports* 14.1 (2024), p. 15877.

<sup>69</sup> McDougall Connor C NAIR JAY KUMAR RAGHAVAN Saeed Umar Abid. “Radiogenomic models using machine learning techniques to predict EGFR mutations in non-small cell lung cancer”. In: *Canadian Association of Radiologists Journal* 72.1 (2021), pp. 109–119.

<sup>70</sup> Echegaray ZHOU MU Leung Ann. “Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications”. In: *Radiology* 286.1 (2018), pp. 307–315.

slice completo puede ocultar detalles importantes del nódulo y aún más si este es reducido<sup>9</sup>. De manera similar, Gil Pinheiro *et al.* y Olivier *et al.* incorporaron anotaciones de los nódulos en sus modelos de machine learning, logrando una mejor clasificación de EGFR y KRAS, aunque este último trabajo destacó la dificultad de diferenciar KRAS, resaltando una poca significancia estadística con respecto a su método<sup>7172</sup>.

También, en la literatura se han reportado modelos con esquemas autosupervisados, que evitan etiquetas de la mutación, permitiendo tener un mayor carácter de generalización. Por ejemplo, se han usado autoencoders variacionales (VAE), que permiten la reducción de dimensionalidad y la generación de representaciones latentes de mutaciones<sup>73</sup>. Otros enfoques han empleado arquitecturas generativas para aumentar la diversidad de perfiles genéticos con formas sintéticas. Por ejemplo, Lazebnik y Simon-Keren (2023) diseñaron el modelo CGMM, que combina algoritmos genéticos con autoencoders profundos y modelos de predicción para generar secuencias genómicas. Esto, aunque logra que las secuencias sintéticas sean casi diferenciables en tareas de clustering no supervisado, aún enfrenta desafíos relacionados con la captación fiel de relaciones mutacionales complejas<sup>74</sup>. Pese a esto, los estudios abordados sobre la síntesis de imágenes no suelen explorar las características latentes aprendidas por los modelos, centrándose en la fidelidad visual y la propia aplicabilidad de la imagen resultante<sup>22</sup>.

---

<sup>71</sup> Gil et al. PINHEIRO. “Identifying relationships between imaging phenotypes and lung cancer-related mutation status: EGFR and KRAS”. in: *Scientific reports* 10.1 (2020), p. 3625.

<sup>72</sup> Khuong Amanda GEVAERT OLIVIER Echegaray Sebastian. “Predictive radiogenomics modeling of EGFR mutation status in lung cancer”. In: *Scientific reports* 7.1 (2017), p. 41674.

<sup>73</sup> Yabo Fu TONGHE WANG Yang Lei. “A review on medical imaging synthesis using deep learning and its clinical applications”. In: *Journal of Applied Clinical Medical Physics* 22.1 (2020), pp. 11–36. DOI: [10.1002/acm2.13121](https://doi.org/10.1002/acm2.13121).

<sup>74</sup> Teddy LAZEBNIK and Liron SIMON-KEREN. “Cancer-inspired genomics mapper model for the generation of synthetic DNA sequences with desired genomics signatures”. In: *Computers in Biology and Medicine* (2023). DOI: <https://doi.org/10.1016/j.compbiomed.2023.107221>.

Algunos enfoques han utilizado GANs para la traducción entre ct y pet, para poder enriquecer representaciones latentes sin necesitar etiquetas de mutaciones. En Salehjahreni et al. (2024) demostro que la PET sintética generada desde CT aporta aproximaciones funcionales utiles en tareas clínicas, lo que sugiere que la síntesis puede servir para estabilizar el aprendizaje y aporta actividad metabolica <sup>22</sup>. A su vez, trabajos como el de Jin et al. (2021), proponen un nuevo modelo generativo GAN (Richer Generative Adversarial Network) para la síntesis libre de tumores 3D en imágenes CT, introduciendo un generador de diálogo con compuertas (RicherDG)<sup>75</sup>. También, recientemente se ha demostrado que modelos de deep learning pueden sintetizar imágenes PET a partir de CT (cross-modality), empleando técnicas de image-to-image translation <sup>76</sup>. Kuang et al. (2020) proponen una MDGAN autosupervisada para detección de anomalías, entrenando con la reconstrucción de imágenes, esto solo con nódulos benignos. Posterior, para la clasificación usan el error de la reconstrucción y la diferencia de características aprendidas en el discriminador entre la imagen y su reconstrucción, sin embargo el método asume que lo benigno es normal maligno son anomalías, lo que requiere fijar un umbral para separar clases, lo que lo hace sensible al conjunto y al dominio <sup>77</sup>. También, se ha demostrado la viabilidad de la CT para proporcionar características que permiten extraer información funcional <sup>78</sup>, demostrando una buena correlación a nivel de voxel entre imágenes de perfusión sintetizadas a partir de CT y los parámetros funcionales

---

<sup>75</sup> Qiangguo JIN et al. “Free-form tumor synthesis in computed tomography images via richer generative adversarial network”. In: *Knowledge-Based Systems* 218 (2021), p. 106753.

<sup>76</sup> Yang Xiaofeng WANG TONGHE. “Take CT, get PET free: AI-powered breakthrough in lung cancer diagnosis and prognosis”. In: *Cell Reports Medicine* 5 (2024). <https://doi.org/10.1016/j.xcrm.2024.101486>, accessed [11 09 2024], p. 2. DOI: [10.1016/j.xcrm.2024.101486](https://doi.org/10.1016/j.xcrm.2024.101486).

<sup>77</sup> Wang Zhi KUANG YANG Liu Yuxin. “Unsupervised Multi-Discriminator Generative Adversarial Network for Lung Nodule Malignancy Classification”. In: *Medical Image Analysis* 64 (2020), p. 101713. DOI: [10.1016/j.media.2020.101713](https://doi.org/10.1016/j.media.2020.101713).

<sup>78</sup> Ge REN et al. “A Transfer Learning Framework for Deep Learning-Based CT-to-Perfusion Mapping on Lung Cancer Patients”. In: *Frontiers in Oncology* 12 (2022). DOI: [10.3389/fonc.2022.883516](https://doi.org/10.3389/fonc.2022.883516).

empleados en la evaluación y planificación de un tratamiento.

Más recientemente, algunos estudios han explorado el uso de técnicas de ensembles de modelos no supervisados para mejorar la robustez en la clasificación de mutaciones. Zhang et al. (2023) investigaron métodos que combinan múltiples algoritmos de clustering para incrementar la sensibilidad en la identificación de patrones mutacionales <sup>79</sup>. Sumado a esto, la integración de datos complementarios a imágenes CT han demostrado mejoría en modelos unimodales de clasificación <sup>71, 9</sup>. Sin embargo, estos enfoques pueden generar redundancias en los datos. Siguiendo la misma línea, trabajos como los de Wang et al., 2019, que desarrollo un modelo de CNN para predecir estados de mutaciones del EGFR, permitiendo una predicción directa sin segmentaciones precisas <sup>65</sup>.

---

<sup>79</sup> R. ZHANG et al. “Ensemble of 3D CNNs for enhanced pulmonary nodule detection”. In: *European Radiology* 33 (2023), pp. 3456–3465.

## 2. PROBLEMA DE INVESTIGACIÓN

El CP es la principal causa de muerte por cáncer a nivel mundial, debido a su elevada tasa de diagnósticos tardíos, reportando más de 1.8 millones de muertes durante el 2022 <sup>1</sup>. Los NP son masas anormales que aparecen en el pulmón, su localización y caracterización de malignidad son claves para un pronóstico favorable en el paciente. El tratamiento personalizado y preciso del CP debería demandar un análisis complementario y multimodal de los NP, integrando un flujo de trabajo que incluya información morfológica (CT), funcional (PET) y genómica. Sin embargo, su análisis multimodal es exclusivo en estudios especializados, lo que limita la transferencia en la personalización de los tratamientos.

Hoy en día, en la rutina clínica, la localización y cuantificación morfológica se realiza principalmente mediante tomografías computarizadas de bajo contraste. No obstante, los estudios de CT no proporcionan información detallada de la malignidad de los nódulos pulmonares y sus estimaciones se ven afectadas por la alta variabilidad morfológica, provocando una tasa alta de falsos positivos y falsos negativos <sup>80</sup>. Por otra parte, las PET proporcionan información funcional del metabolismo tumoral <sup>76,3</sup> aunque con una alta exposición a la radiación y un uso exclusivo para centros especializados <sup>81,20</sup>. Asimismo, las pruebas genómicas permiten identificar mutaciones en genes específicos como EGFR y KRAS, sin embargo, su obtención requiere procedimientos invasivos y dependientes de una apropiada localización del tumor <sup>31,32</sup>. Estas modalidades de análisis incluyen aún limitaciones propias de la caracterización de los NP y adicionalmente, el uso conjunto en un marco general sigue siendo un problema abierto.

---

<sup>80</sup> Contegiacomo P DEL CIELLO Franchi A. “Missed lung cancer: when, where, and why?” In: *Diagnostic and Interventional Radiology (Ankara, Turkey)* (2017), pp. 118–126.

<sup>81</sup> Zhiyong Yang GUOBIN ZHANG Shan Jiang. “Automatic nodule detection for lung cancer in CT images: A review”. In: *Computers in Biology and Medicine* 103 (2018). DOI: <https://doi.org/10.1016/j.combiomed.2018.10.033>.

Es por ello, que herramientas computacionales han emergido para soportar cada modalidad, combinando información anatómica, funcional y molecular, que pueda proporcionar información esencial sobre su localización, naturaleza y agresividad, para plantear un tratamiento adecuado. Sin embargo, estas herramientas se han enfocado en la detección y clasificación desde una perspectiva unimodal a partir del uso de imágenes CT. Por otro lado, enfoques multimodales que integran CT y PET dependen de la integración de los datos radiómicos los cuales pueden variar según el protocolo de adquisición de la imagen, ignorando potenciales fenotipos como la presencia de fibrosis. A su vez, la limitada disponibilidad de imágenes PET presentan dificultades en la fase de inferencia, ya que estos modelos requieren de ambas modalidades CT y PET, para poder hacer una detección precisa del nódulo. De hecho, se requiere realizar modelos que puedan aprender mutaciones genómicas desde hallazgos en PET y CT, para así poder hacer inferencias en escenarios con ausencia de alguna de estas modalidades, brindando un soporte en el flujo diagnóstico.

**Pregunta de investigación** ¿Cómo implementar representaciones de aprendizaje profundo para localizar y clasificar nódulos pulmonares según su mutación genómica, a partir de imágenes CT y PET?

## **3. OBJETIVOS**

### **3.1. OBJETIVO GENERAL**

Implementar representaciones de aprendizaje profundo para localizar y clasificar mutaciones genómicas observadas en imágenes diagnósticas PET y CT.

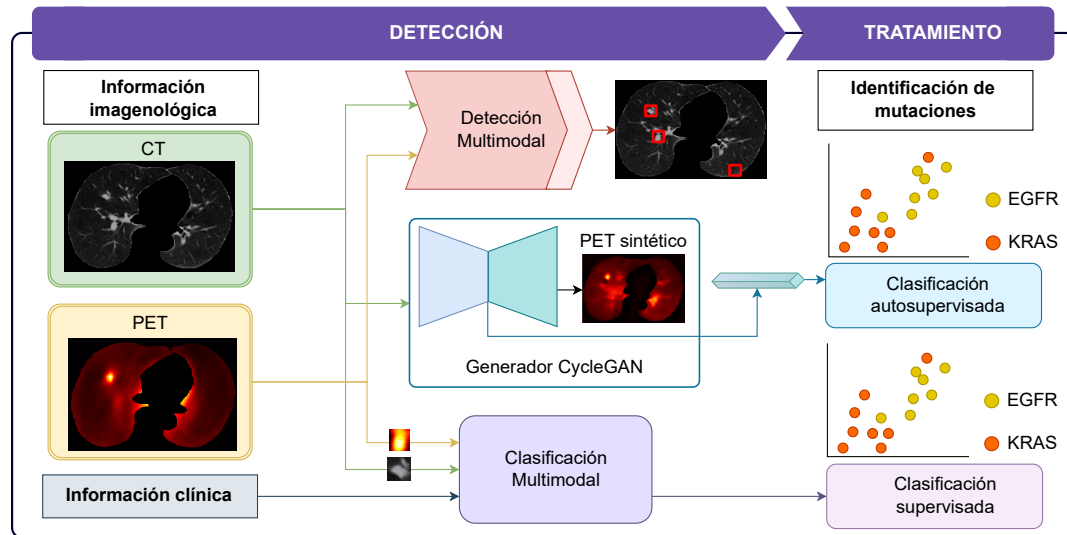
### **3.2. OBJETIVOS ESPECÍFICOS**

- Seleccionar un conjunto de datos con imágenes CT y PET de cáncer de pulmón que contengan anotaciones de localización, historia clínica y clasificación de mutaciones genómicas en nódulos pulmonares.
- Implementar una metodología basada en aprendizaje profundo para la localización de nódulos pulmonares en las imágenes CT y PET.
- Desarrollar un modelo de aprendizaje profundo supervisado multimodal para clasificar el tipo de mutación genómica presente en los nódulos pulmonares usando imágenes PET y CT.
- Desarrollar un modelo de aprendizaje profundo auto supervisado para codificar descriptores discriminatorios con respecto a mutaciones genómicas.
- Validar los modelos desarrollados en cuanto a su capacidad para localizar nódulos pulmonares y estimar la presencia de mutaciones genómicas.

## 4. MÉTODO PROPUESTO

En este proyecto se implementó un flujo de trabajo orientado a brindar soporte en la localización y clasificación de nódulos pulmonares (NP), enmarcadas en la ruta diagnóstica de caracterización del cáncer de pulmón. La caracterización se realizó con respecto a la mutación genómica, integrando información imagenológica y datos clínicos del paciente. Tal como se ilustra en la Figura 10, este proceso inicia con la implementación de una arquitectura para localización de NP en imágenes CT. Este esquema de localización es reforzado con un Reductor de Falsos Positivos (RFP), complementando la caracterización del NP, según su actividad metabólica. Una vez localizados los nódulos, se realizaron dos aproximaciones para apoyar la clasificación de mutaciones genómicas (EGFR o KRAS) de las observaciones nodulares. En esta fase, primero se realizó una arquitectura multimodal en conjunto con los datos clínicos del paciente para clasificar y determinar el estado mutacional del nódulo de forma supervisada. Adicionalmente, siguiendo un esquema autosupervisado (CT-to-PET), se obtuvieron vectores embebidos para clasificar el tipo de mutación genómica, como alternativa a la caracterización. A continuación, se detalla cada una de las arquitecturas involucradas en este proyecto.

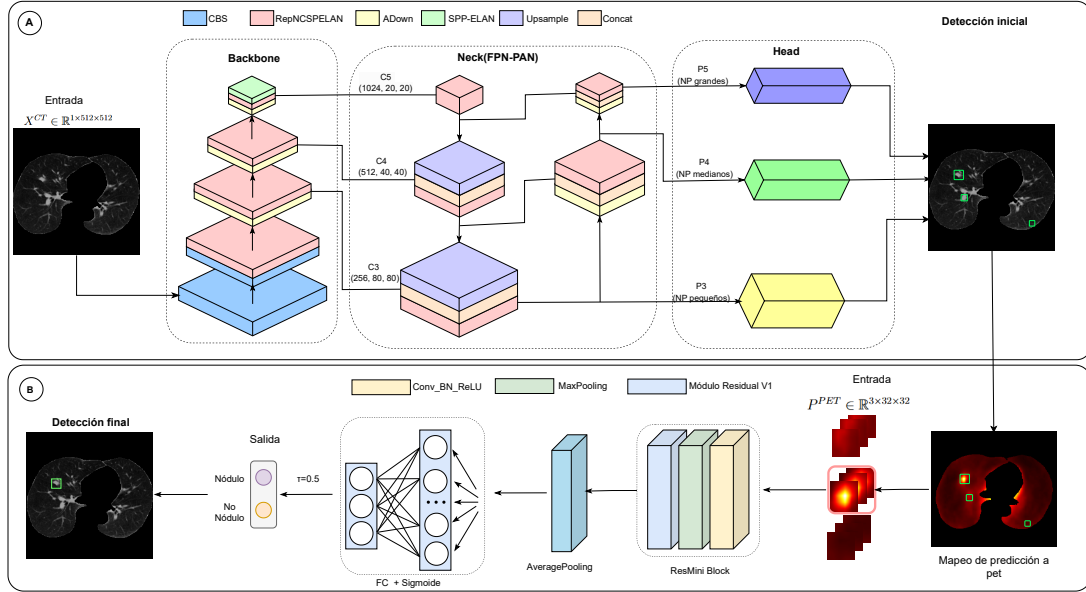
**Figura 10.** Esquema de la arquitectura propuesta para la localización y clasificación de NP, considerando estudios imagenológicos de CT, PET, datos clínicos y su correspondiente mutación genómica (EGFR o KRAS).



#### 4.1. MODELO MULTIMODAL DE CT Y PET PARA LA LOCALIZACIÓN DE NÓDULOS PULMONARES

En este trabajo se realizó una aproximación para la localización de nódulos desde la morfología observada en tomografía computarizada (CT). Una vez estimadas las regiones candidatas a ser NP, luego son filtradas según la actividad metabólica tumoral de glucosa en el tejido, sirviendo así como reductor de falsos positivos (RFP). La Figura 11 presenta el esquema completo de este proceso, en el que se articulan las etapas de detección inicial en CT y reducción de falsos positivos en PET.

**Figura 11.** Esquema de la arquitectura multimodal propuesta: (A) detección inicial en CT mediante *YOLOv9* y (B) reducción de falsos positivos mediante un clasificador *ResNet18* que procesa los parches proyectados en PET.



Para la etapa de detección inicial se emplearon cortes axiales de CT, representados como  $X^{CT} \in \mathbb{R}^{C \times H \times W}$ , donde  $C$  corresponde al canal y  $(H, W)$  pertenecen a las dimensiones espaciales de cada corte. El objetivo de este bloque es localizar regiones candidatas a NP mediante el modelo *YOLOv9*, el cual obtiene un conjunto de predicciones  $\mathcal{Y}_{YOLO} = \{(b_i, s_i)\}_{i=1}^N$ , donde cada  $b_i = (x_i, y_i, w_i, h_i)$  define las coordenadas normalizadas de una caja delimitadora alrededor del NP y  $s_i \in [0, 1]$  es la confianza asociada a la detección.

Particularmente, la arquitectura de *YOLOv9* se organiza en tres componentes principales que trabajan de forma secuencial. En primer lugar, el *backbone* se basa en la Red Generalizada de Agregación Eficiente de Capas (GELAN, por sus siglas en inglés) que agrupa cuatro bloques con funciones complementarias. El bloque Conv-BN-SiLU (CBS) compuesto por un bloque convolucional, normalización por lote y la activación SiLU. Este bloque transforma la imagen en mapas de características que resaltan patrones básicos de intensidad y textura. Luego, se incluye módulo *RepNCSC-ELAN*, el cual incluye conexiones normalizadas y repetidas con

una red de atención de *kernel*. Este bloque contiene ramas paralelas cortas con cuellos de botella residuales (*RepNbottleneck*) y una fusión por concatenación en la dimensión de canales seguida de una proyección  $1 \times 1$ . El *RepNCSC-ELAN* permite integrar detalles locales con representaciones de mayor alcance y conserva rutas de paso cortas que mantienen un flujo de gradientes estable hacia las primeras capas. El bloque de submuestreo asimétrico (*ADown*, por sus siglas en inglés) contiene dos rutas con *Max-pooling* y *Average-pooling* acompañadas de convoluciones, que luego se concatenan a lo largo de los canales, permitiendo reducir a la mitad ( $H, W$ ) y ampliando el campo receptivo, lo que contribuye, en nuestro caso a preservar características relevantes del tejido. Por otra parte, el bloque *SPP-ELAN* está compuesto por ramas paralelas de *Max-pooling* de tamaño  $5 \times 5$ ,  $9 \times 9$  y  $13 \times 13$ , seguidas de concatenación en canales y una proyección  $1 \times 1$ , lo que aporta contexto multiescala en las imágenes CT. Estos bloques en conjunto permiten que el *backbone* produzca salidas jerárquicas denominadas como  $C3, C4, C5$ , aproximadamente a  $1/8, 1/16$  y  $1/32$  de la resolución de entrada. Formalmente,  $F^{\text{backbone}} = f_{\text{backbone}}(X^{CT}) = \{C3, C4, C5\}$ . En general, en este bloque *backbone* se estructura la información primaria de la imagen, manteniendo una descripción amplia por las observaciones de la imagen, para luego ser utilizada en las tareas de localización

En segundo lugar, el *neck* emplea una estructura de agregación multi-escala denominada *Red de Agregación de Rutas-Red Piramidal de Características* (*PAN-FPN*, por sus siglas en inglés). El *FPN* conduce información semántica de alto nivel desde capas profundas hacia mapas de mayor resolución (flujo descendente) de manera que los mapas en alta resolución incorporan ese contexto y refuerzan las representaciones finas. De forma complementaria, *PAN* traslada detalles espaciales desde capas iniciales hacia representaciones más abstractas (flujo ascendente), incorporando contexto útil en la detección de NP. La interacción bidireccional de ambos flujos producen las representaciones fusionadas  $P3, P4$  y  $P5$  alineadas con  $C3, C4$  y  $C5$ , donde  $P3$  favorece NP pequeños,  $P4$  NP medianos y  $P5$  NP grandes; de forma compacta, el proceso se expresa como  $F^{\text{neck}} = f_{\text{neck}}(F^{\text{backbone}}) = \{P3, P4, P5\}$ .

En una última etapa, el *head* recibe  $F^{\text{neck}}$  y genera las predicciones mediante un diseño de-

sacoplado, en el que una rama estima la regresión de cajas  $(x, y, w, h)$  y la puntuación de objeto, mientras que otra rama produce la clasificación por categoría; las salidas se calculan por celda en  $P3, P4$  y  $P5$ , dando lugar a candidatos  $\{(b_i, s_i)\}_{i=1}^N$  donde  $b_i = (x_i, y_i, w_i, h_i)$  delimita un NP y  $s_i \in [0, 1]$  es la confianza asociada; este mapeo se resume como  $\mathcal{Y}_{YOLO} = f_{\text{head}}(F^{\text{neck}})$ , donde  $\mathcal{Y}_{YOLO}$  agrupa las cajas propuestas y sus puntajes para NP pequeños, medianos y grandes. Así se utilizó y ajustó la arquitectura *YOLOv9* para la tarea de localización, produciendo múltiples candidatos de regiones asociadas a nódulos pulmonares.

Sin embargo, es común que estas redes generen múltiples regiones con la asociación de nódulos, entre otras por las estructuras similares del pulmón y su reducido tamaño. Esto conduce a generar un amplio número de falsos positivos, lo cual puede ser contraproducente en la tarea de apoyo diagnóstico y análisis de estudios completos del parénquima pulmonar. Para mitigar esta alta tasa de falsos positivos característica de los detectores de NP, en este trabajo se incorporó un módulo reductor multimodal de falsos positivos, diseñado para explotar la complementariedad entre información anatómica (CT) y funcional (PET). Cabe resaltar que esta integración entre los estudios funcionales y anatómicos son una contribución para el estado-del-arte, logrando así implementar de forma natural las observaciones, de acuerdo a las ventajas de confirmación, brindadas por los estudios PET.

Particularmente, una vez obtenidas las detecciones iniciales en CT, estas regiones candidatas se proyectan sobre PET y se extraen en forma de parches  $P^{PET} \in \mathbb{R}^{C \times W' \times H'}$  con el fin de discriminar, mediante la actividad funcional, qué zonas corresponden realmente a un NP. Para ello, se usa la arquitectura *ResNet18*, la cual está compuesta por bloques residuales que mitigan el problema del desvanecimiento del gradiente al ofrecer un camino de retropropagación alternativo a través de conexiones residuales, lo que garantiza un flujo de gradiente más estable y previene su degradación. Cada bloque aplica la operación  $\sigma(BN_2(W_2 * \sigma(BN_1(W_1 * x))) + \delta(x))$ , donde  $x = P^{PET}$  es la entrada del bloque,  $W_1$  y  $W_2$  son los filtros convolucionales,  $BN$  corresponde a la normalización por lote,  $\sigma$  es la activación ReLU y  $\delta(x)$  es la conexión residual. En el caso de los NP en PET, estas representaciones permiten diferenciar capta-

ciones focalizadas y consistentes con actividad tumoral de aquellas difusas o irregulares, que suelen corresponder a estructuras normales o falsos positivos. De esta forma, la red genera un vector global  $F^{PET}$ , que se proyecta en una capa sigmoide para estimar la probabilidad  $\hat{y}$  de que el parche corresponda a un nódulo verdadero.

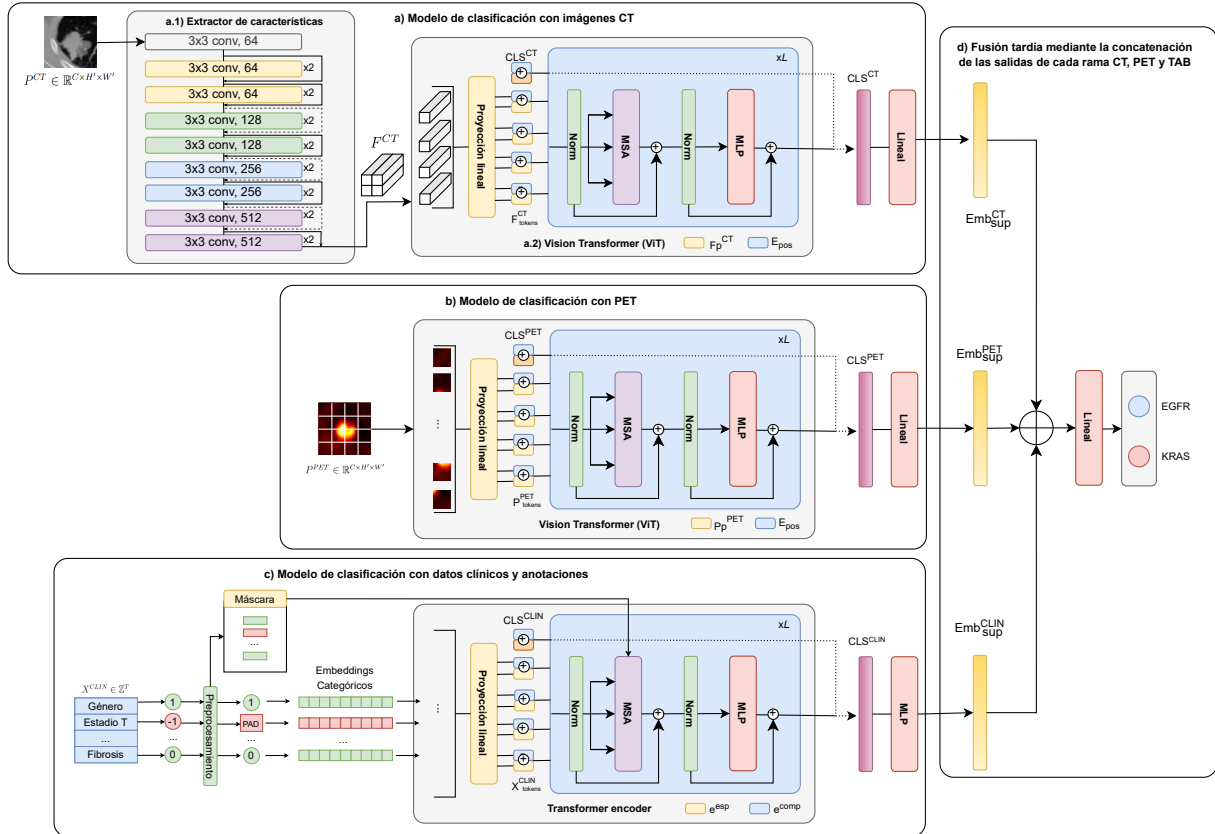
## 4.2. MODELO SUPERVISADO MULTIMODAL BASADO EN TRANSFORMERS PARA LA CLASIFICACIÓN DE MUTACIONES GENÓMICAS EN NP CON IMÁGENES CT, PET E INFORMACIÓN TABULAR

En una segunda etapa de soporte en la caracterización de los nódulos, se busco explotar la información visual de las observaciones nodulares (tanto en CT, como en PET) e integrarla con información tabular sobre caracterización clínica de estas masas. Particularmente en este trabajo, se usaron parches que delimitan al NP en CT y PET, representados como  $P^{CT} \in \mathbb{R}^{C \times H' \times W'}$  y  $P^{PET} \in \mathbb{R}^{C \times H' \times W'}$  respectivamente para extraer fenotipos asociados. Donde  $C$  son los canales,  $H'$  la altura y  $W'$  el ancho del parche. Asimismo, se integraron los datos clínicos de los pacientes y las anotaciones del nódulo en una matriz tabular,  $X^{CLIN} \in \mathbb{Z}^T$ , donde  $T$  es el número de categorías. Esta unión se realiza para capturar de manera integrada la relación entre las características del paciente y su NP.

Cada una de las modalidades consideradas en este trabajo fue mapeada y procesada por una rama de representación que emplea múltiples mecanismos de atención (i.e., una rama *Transformer*), pretendiendo no solo capturar información local, sino aprovechar el contexto no local en cada una de las observaciones. Estas ramas de tipo *Transformer* son luego fusionadas, logrando capturar y aprender simultáneamente patrones morfológicos, funcionales y clínicos, con el fin de integrar diferentes perspectivas del diagnóstico. En el esquema propuesto en este trabajo, para la arquitectura multimodal considerada, las salidas lineales de cada rama supervisadas, denotadas como  $Emb_{sup}^{CT}$ ,  $Emb_{sup}^{PET}$ , y  $Emb_{sup}^{CLIN}$  se concatenaron para integrar información complementaria de cada modalidad en un espacio común, con el objetivo de determinar la mutación genómica EGFR o KRAS de los NP, como se muestra en la Figura

12.

**Figura 12.** Diagrama completo de la arquitectura multimodal del método propuesto para clasificación supervisada. Cada modalidad tiene su propia arquitectura: a) CT pasa por un extractor de características basado en *ResNet18* obteniendo representaciones más enriquecidas que serán procesadas por un *ViT*, b) PET cuenta con un *Vision Transformer* sin caracterización previa y c) los datos clínicos pasan por un preprocesamiento previo en el que se estudian los datos faltantes para pasar al *Tab Transformer* en forma de vectores embebidos.



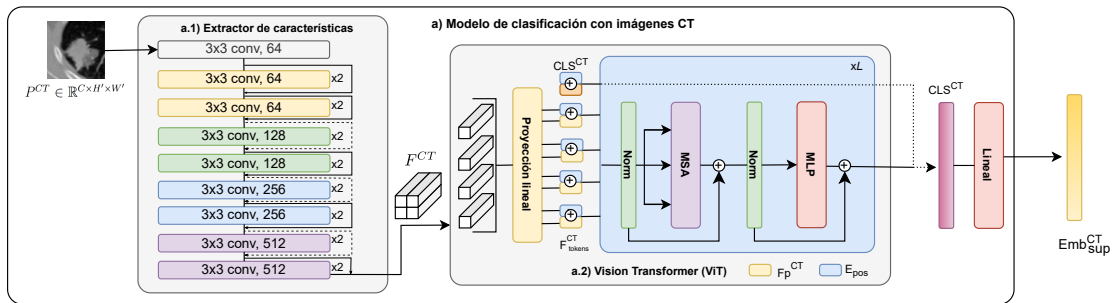
A continuación, se describirá a detalle como funciona cada rama de procesamiento *Transformer*, de acuerdo a la modalidad codificada:

### Modelo de Vision Transformer para imágenes CT y PET

Para la implementación de las ramas con estudios imagenológicos, se aprovecha la información que aporta cada modalidad individualmente: Mientras que CT describe la morfología del nódulo, PET ofrece in-

formación funcional basada en la captación metabólica. En la rama CT, los parches de los nódulos  $P^{CT}$  se procesan inicialmente a través de una arquitectura convolucional basada en *ResNet18* que actuará como extractor de características para obtener información más representativa del nódulo, capturando patrones locales como bordes o texturas, obteniendo un mapa de características  $F^{CT} \in \mathbb{R}^{C \times H \times W}$ . Este mapa será utilizado como entrada a un *Vision Transformer* (*ViT*, por sus siglas en inglés), dividiendo el mapa en una secuencia de parches o *tokens*  $F_p^{CT} \in \mathbb{R}^{N \times d}$  mediante una proyección lineal, en donde cada una de las  $N$  posiciones corresponde a un *token* del mapa de características y  $d$  es la dimensión de su respectivo vector embebido. A esta secuencia se añade un *token* de clasificación  $CLS^{CT}$ , cuyo propósito es condensar la información global del NP y, asimismo, cada *token* recibe un vector embebido posicional  $E_{pos}$ , para conservar el orden de los *tokens*, los cuales serán procesados de manera paralela, por lo que, la secuencia que servirá de entrada al *Vision Transformer* es  $F_{tokens}^{CT} = [CLS^{CT}; F_p^{CT}] + E_{pos}$ , en el que  $F_{tokens}^{CT} \in \mathbb{R}^{(N+1) \times d}$  (ver Figura 13-a.1).

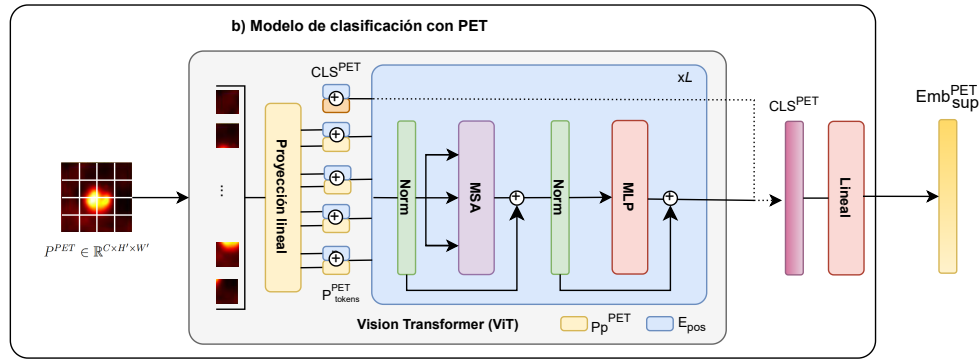
**Figura 13.** Rama de CT para la arquitectura multimodal supervisada basada en *Transformers*. a.1) Tiene un extractor de características basado en *ResNet18* que extrae información morfológica representativa del NP mediante la creación de un mapa de características a.2) Este mapa de características es usado como entrada a un *ViT* para modelar relaciones globales entre distintas partes del NP y generar un *token* de clasificación.



La rama PET sigue un procedimiento similar al descrito por la rama dedicada a explotar información morfológica desde la secuencia CT. Sin embargo, la naturaleza de PET tiene una menor cuantización, es decir, sus valores indican en menor o mayor medida la activación o aceleración funcional (comportamiento metabólico) de un tejido, por lo que los parches no

muestran una complejidad textural para tener que ser representados en un banco de características. Es por ello que los parches  $P^{PET}$  ingresan directamente a un *Vision Transformer* sin un extractor convolucional. Por ello, se aplican los mecanismos de atención de manera directa para capturar relaciones en zonas de captación, obteniendo una secuencia de *tokens* a partir de la división del parche nodular en PET,  $P_{tokens}^{PET} = [CLS^{PET}; P_p^{PET}] + E_{pos}$ , como se observa en la Figura 14.

**Figura 14.** Rama de PET para la arquitectura multimodal supervisada basada en *transformers*.



En cada rama, el *ViT* procesa su respectivo conjunto de *tokens* sea  $F_{tokens}^{CT}$  o  $P_{tokens}^{PET}$  de manera independiente. Cada bloque  $L$  de *Transformer* está compuesto por dos operaciones: *Auto-Atención Multicabeza* (MSA, por sus siglas en inglés), que permite que cada región del nódulo se relacione con otras para capturar dependencias globales, y una red *feed-forward* (*Multi-Layer Perceptron*, *MLP*) que refina dichas interacciones, ambas están relacionadas en conexiones residuales y antes de ellas una capa de normalización. En el mecanismo de atención, cada *token* de la secuencia se proyecta en tres espacios distintos mediante transformaciones lineales conocidas como *Query* (Q), *Keys* (K) y *Values* (V), representadas de la siguiente manera  $Q = XW^Q$ ,  $K = XW^K$ ,  $V = XW^V$ , en el que  $Q, K, V \in \mathbb{R}^{N \times d_k}$ , donde  $d_k = d/nh$  que es la dimensión por cabeza,  $d$  representa el tamaño del vector embebido y  $nh$  el número de cabezas. Para cada cabeza la atención se define de la siguiente manera:  $Attn(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V$ , lo que permite asignar mayor peso a interacciones relevantes entre regiones del nódulo. Cada una de estas  $nh$  cabezas se concatenan y proyectan,

obteniendo  $MSA(X) = [Attn_1, \dots, Attn_h]W^o$ . El resultado del  $MSA$  se agrega residualmente  $X' = X + MSA(LN(X))$ , de esta manera, los *token* se enriquecen no solo con información propia, sino también con dependencias globales respecto a otras regiones del NP. Posteriormente, cada *token* procesado  $X'$  pasa por un  $MLP$ ,  $MLP(x) = W_3 \cdot \sigma_1(W_4x + b_1) + b_2$  compuesto por dos capas densas con una tasa de expansión que incrementa temporalmente la dimensión del vector embebido, otorgando mayor capacidad al  $MLP$  de refinar las interacciones detectadas por la atención y una activación  $GELU$  ( $\sigma_1$ ), que suaviza el tratamiento de valores negativos, evitando perder información sutil que pueda ser relevante en la caracterización del NP. Nuevamente, se aplica una suma residual, por lo que la salida del bloque quedaría definida como  $X^l = X' + MLP(LN(X'))$ . Después de los  $L$  bloques *Transformer*, se obtiene una secuencia  $X^{(L)} \in \mathbb{R}^{(N+1) \times d}$ , en la que el primer *token* de la secuencia,  $CLS$ , resume la información global de la modalidad, ya sea CT con respecto a la información morfológica o PET en relación con su comportamiento funcional. Estos *tokens*  $CLS^{CT}$  y  $CLS^{PET}$  serán empleados para la fusión multimodal junto con el vector embebido clínico, ver Figura 13-a.2 y Figura 14-b.

### Modelo clínico para datos tabulares

De manera complementaria a la información visual, la representación multimodal de este trabajo también aprovechó la información clínica resumida en información tabular, que acompaña el dataset. Para este apartado del modelo se adoptó una red basada en *Tab transformer*<sup>82</sup> para integrar de manera conjunta las variables clínicas de cada paciente y las anotaciones categóricas asociadas al NP. En este caso, la matriz de entrada es definida como:  $X^{CLIN} \in \mathbb{Z}^T$ , donde  $T$  es el número de variables categóricas.

Teniendo en cuenta que, para cada columna  $t$  presente en  $X^{CLIN}$  existe un número de categorías válidas  $J_t \geq 0$ , el objetivo es proyectar cada uno de estos valores en forma de vectores

---

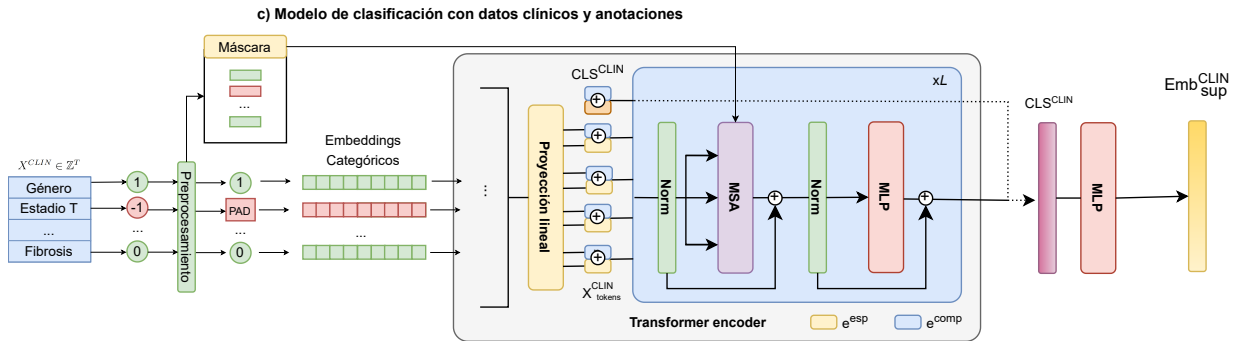
<sup>82</sup> Xin Huang et AL. “TabTransformer: Tabular Data Modeling Using Contextual Embeddings”. In: *CoRR* abs/2012.06678 (2020). arXiv: [2012.06678](https://arxiv.org/abs/2012.06678).

embebidos específicos en una tabla global de vectores embebidos  $E^{esp} \in \mathbb{R}^{(U+s+1) \times d_{esp}}$ , donde  $U = \sum_{t=1}^T J_t$  que corresponde al número total de categorías únicas,  $s$  el número de *tokens* especiales (como *CLS*) y  $+1$  corresponde al espacio reservado para el *padding*, cuya labor es convertir los datos faltantes en categorías válidas con un índice común. De esta tabla, cada categoría  $t$  observada se proyecta en un vector embebido específico,  $e_t^{esp} = E_{esp}[x_t + \text{offset}(t)] \in \mathbb{R}^{d_{esp}}$ , donde la función de *offset* se asegura que cada columna tenga un rango exclusivo de índices en la tabla global para evitar colisiones con otras categorías. Así evitando, por ejemplo, que la categoría "1" de la variable Género y la categoría "1" de la variable Fibrosis no apunten al mismo vector embebido, si no posiciones diferentes dentro de  $E_{esp}$ . Adicionalmente, a los vectores embebidos específicos se les asocian los vectores embebidos compartidos,  $e_t^{comp} \in \mathbb{R}^{d_{comp}}$ , los cuales permiten preservar la identidad de la variable clínica o anotación a la que pertenecen. Teniendo esto en cuenta, el vector embebido final de cada variable es la concatenación de ambos:  $e_t = [e_t^{esp}; e_t^{comp}] \in \mathbb{R}^{d'}$ , en el que  $d' = d_{esp} + d_{comp}$ .

Por otra parte, para evitar sesgos derivados de imputaciones, se define una máscara binaria  $M \in \{0, 1\}^T$ , donde  $M_t = 1$  si la entrada es faltante. Esta máscara se aplica al *MSA*, anulando las posiciones correspondientes en la matriz de similitud antes de aplicar *softmax*, convirtiendo las probabilidades a cero. Así, un valor ausente en una variable clínica o anotación no puede ser atendido como *Key* o *Value* por ningún otro *token*.

El conjunto de vectores embebidos enriquecidos  $\{e_1, \dots, e_N\}$  se combinan con un *token* especial de clasificación  $CLS^{CLIN}$ , de modo que la secuencia se define como  $X_{tokens}^{CLIN} = [CLS^{CLIN}; e_1, \dots, e_N]$ . Esta secuencia pasa por los  $L$  bloques de *Transformers* aplicando *MSA* y *MLP*, la secuencia resultante  $X^{(L)} \in \mathbb{R}^{(N+1) \times d}$ , donde se toma la representación del *token*  $CLS^{CLIN}$  que sintetiza la información clínica y de anotaciones para un paciente. Esta representación  $Emb_{sup}^{CLIN}$ , será utilizada en la capa de fusión multimodal en conjunto con las salidas de las ramas de CT y PET, como se puede apreciar en la Figura 15.

**Figura 15.** Rama de datos tabulares con método de enmascaramiento de datos que se desarrolla en el módulo MSA.



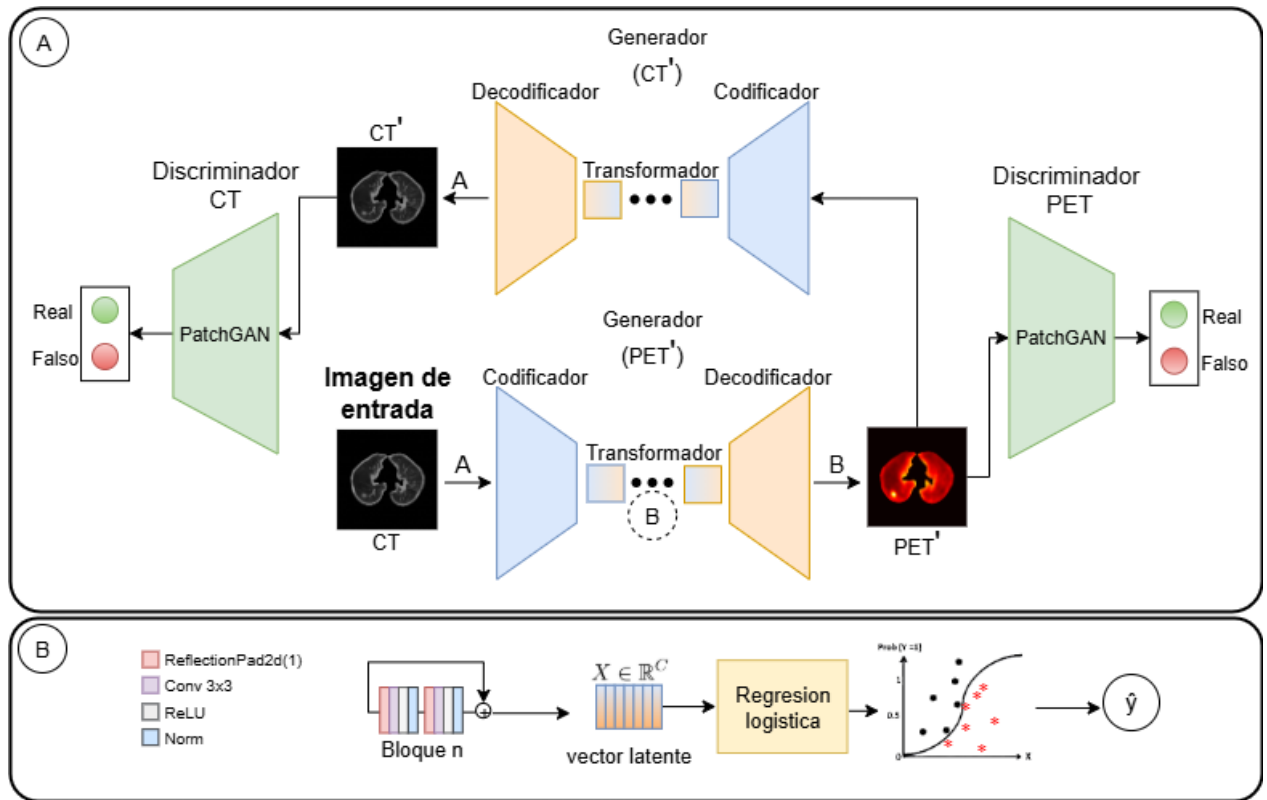
Una vez obtenida la representación de cada fuente de información, la fusión se realiza mediante la proyección de vectores embebidos de baja dimensionalidad, que contienen toda la semántica observable en los nódulos. Es así que la fusión tardía se logra mediante la concatenación de las salidas de cada rama:  $Emb_{sup}^{CT}$ ,  $Emb_{sup}^{PET}$ , y  $Emb_{sup}^{CLIN}$ , el cual es proyectado usando un clasificador lineal (ver Figura 12).

### 4.3. MODELO AUTOSUPERVISADO BASADO EN SÍNTESIS DE IMÁGENES PARA LA CLASIFICACIÓN DE MUTACIONES GENÓMICAS EN NP CON IMÁGENES CT Y PET

Como estrategia computacional alternativa para caracterizar las mutaciones, en este trabajo se exploró una alternativa no supervisada (sin requerimientos de etiquetas de mutaciones). Con esto, se pretendía explorar como descriptores desde las imágenes, pueden por si solos, tener una correlación con respecto a la mutación. Además, podría ser una alternativa para brindar un soporte en centros, donde solo se cuente con la información imagenológica. Particularmente, en este trabajo se siguió una tarea de pretexto para generar imágenes PET desde secuencias CT. Luego su descriptor embebido fue utilizado para discriminar entre mutaciones. Específicamente, se exploraron las representaciones latentes al extraer un vector embebido y utilizarlo como un espacio discriminativo para la clasificación genómica entre mutaciones

EGFR y KRAS. La Figura 16 presenta el esquema completo de este proceso, en el cual se articulan las etapas de síntesis de imágenes entre las modalidades CT a PET y clasificación genómica con las representaciones latentes aprendidas. A continuación, se detalla la obtención de este vector.

**Figura 16.** Diagrama general de la arquitectura para el método propuesto de clasificación autosupervisada. (A) Traducción entre dominios CT y PET, mediante la arquitectura de *CycleGAN* (B) Clasificación de mutaciones genómicas (EGFR/KRAS) a partir de la extracción de representaciones latentes del generador usando Regresión Logística.



**Modelo CycleGAN para la síntesis de imágenes CT a PET** Para la fase de síntesis se emplearon cortes axiales de CT, representados como  $X^{CT}$ . A partir de un generador  $G_{CT \rightarrow PET}$  basado en una arquitectura *ResNet* con  $n$  bloques residuales, el cual busca generar imágenes sintéticas en el dominio PET, denotadas como  $\hat{X}^{PET} = G_{CT \rightarrow PET}(X^{CT})$ . Adicionalmente, se entrena un generador inverso  $G_{PET \rightarrow CT}$  para asegurar la consistencia cíclica, es de-

cir, que al traducir una imagen de un dominio a otro y luego, retornarla a su dominio original, la reconstrucción se asemeja a la imagen de entrada  $G_{PET \rightarrow CT}(G_{CT \rightarrow PET}(X^{CT})) \approx X^{CT}$ . Asimismo, se introducen dos discriminadores adversariales  $D_{PET}$  y  $D_{CT}$ , a través de un modelo *PatchGAN*, que evalúan las imágenes generadas, con el fin de discernir correctamente entre una imagen sintética y una real.

El modelo *CycleGAN* se organiza en dos componentes principales: los generadores y los discriminadores. El generador  $G$  sigue una estructura *codificador–transformador–decodificador*. En la codificación se abstraen características convolucionales desde  $X^{CT}$ , obteniendo un mapa de características  $F^{enc} = f_{enc}(X^{CT})$ . El transformador, basado en un modelo *ResNet* preserva la resolución espacial  $(W, H)$  y refina representaciones latentes. El *decodificador* se encarga de reconstruir la imagen en el dominio de destino a partir de  $F^{trans}$ , representado como  $\hat{X}^{PET} = f_{dec}(F^{trans})$ ,  $f_{dec}(X) = \tanh(\sigma(BN(W^\top * F^{trans})))$ , donde  $W_i^\top$  son los filtros de convolución transpuesta y la función  $\tanh$  aplicada asegura que la salida permanezca en el rango  $[-1, 1]$ . A su vez, este proceso se replica de manera cíclica, en el cual la imagen de entrada sera una imagen  $\hat{X}^{PET}$  que realizará el mismo proceso de reconstrucción en el dominio original.

Esta arquitectura se minimiza integrando dos pérdidas adversariales  $\mathcal{L}_{ADV}(G_{CT \rightarrow PET}, D_{PET})$ ,  $\mathcal{L}_{GAN}(G_{PET \rightarrow CT}, D_{CT})$  que incentivan a los generadores  $G_{CT \rightarrow PET}, G_{PET \rightarrow CT}$  a que las imágenes generadas sean indistinguibles de las reales por los discriminadores. Complementando estas perdidas, el ajuste incluye dos perdidas de consistencia cíclica  $(\mathcal{L}_{cyc}^{CT}, \mathcal{L}_{cyc}^{PET})$ , que ayudan a que una imagen traducida a otro dominio y luego devuelta conserve la estructura de la original. Por ultimo, la perdida de identidad,  $\mathcal{L}_{idt} = \mathcal{L}_{idt}^{PET} + \mathcal{L}_{idt}^{CT}$  evalúa cuando una imagen ya pertenece al dominio objetivo. En general, estos términos dan pie a la función que engloba las perdidas de entrenamiento y validación de las imágenes generadas:  $\mathcal{L}(G_{CT \rightarrow PET}, G_{PET \rightarrow CT}, D_{CT}, D_{PET}) = \mathcal{L}_{GAN} + \lambda_{cyc} \mathcal{L}_{cyc} + \lambda_{idt} \mathcal{L}_{idt}$  donde  $\lambda_{cyc}$  y  $\lambda_{idt}$  son hiperparámetros que balancean la fidelidad estructural frente a la calidad visual de las imágenes generadas.

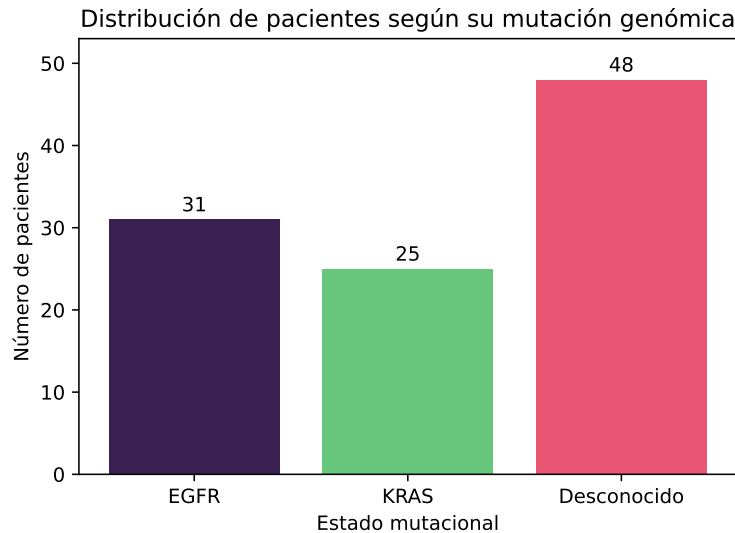
**Clasificación de mutaciones genómicas mediante Regresión Logística** Tras entrenar el generador  $G_{CT \rightarrow PET}$ , dado  $X^{CT}$ , pasa por el *encoder* y en el transformador se obtiene un vector latente a partir de un bloque residual  $Z^{(n)} = f_{\text{trans}}^{(n)}(F^{\text{enc}})$ ,  $Z^{(n)} \in \mathbb{R}^{C_n \times H_n \times W_n}$ . el cual es la salida posterior a la conexión residual en ese bloque  $n$ , de este, se obtiene un vector embebido global, el cual pasa por medio de una agrupación de promedios globales (GAP, por sus siglas en inglés) dado  $v = \text{GAP}(Z^{(n)})$  facilitando la clasificación ya que transforma un mapa complejo en un vector de características más representativo y compacto, que conserva la información relevante de cada canal y la entrega en un formato directamente utilizable por el clasificador, posteriormente, se realiza una normalización  $\ell_2$ , asegurando una magnitud unitaria entre los valores, estabilizando el clasificador y facilitando la interpretación de patrones relevantes. Para el proceso de clasificación binaria, buscando predecir la mutación genómica (EGFR o KRAS) se decidió utilizar una *Regresión Logística* (LR) sobre el vector embebido del proceso de síntesis, debido a que, a la vez que se basa en una estructura lineal, transforma las características en una probabilidad asociada a una clase específica, facilitando establecer los umbrales de decisión entre mutaciones.

## 5. DISEÑO EXPERIMENTAL

### 5.1. DATOS

Para entrenar y validar los diferentes métodos propuestos se empleó el conjunto de datos público NSCLC-Radiogenomics. Este conjunto de datos contiene imágenes CT, PET, información clínica y descripciones de las características morfológicas del nódulo en CT. La cantidad original de 211 pacientes se redujo bajo los siguientes criterios: todos los pacientes deben contar con todas las modalidades, los pacientes deben mostrar adenocarcinoma en sus registros histológicos y no deben presentar de manera simultánea mutaciones en EGFR y KRAS. Bajo estas condiciones, se obtuvo un total 104 pacientes cuyo estado mutacional se dividió en EGFR, KRAS y desconocido (ver Figura 17).

**Figura 17.** Distribución de pacientes del dataset NSCLC Radiogenomics, la mayoría de pacientes no poseen una mutación conocida.



La información tabular clínica fue modificada al retirar columnas no relevantes en la clasificación como la etnia, el hospital o la fecha de la toma de las tomografías, resultando en un conjunto de 21 variables categóricas, las cuales fueron ordenadas y reemplazadas por números

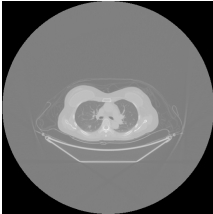
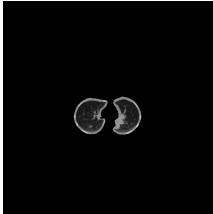
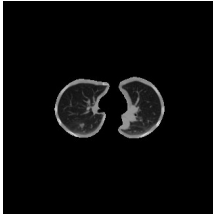
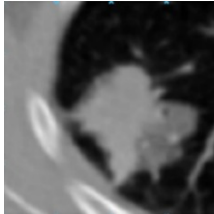
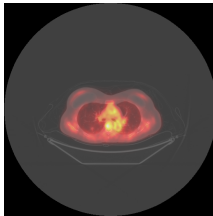
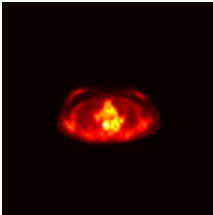
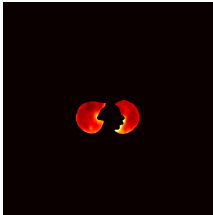
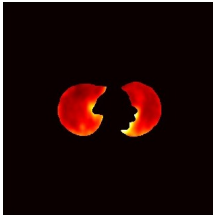
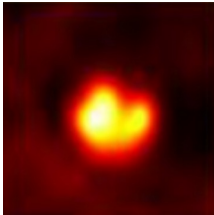
enteros positivos secuenciales.

Algunos datos faltantes (como estadio TNM, grado histológico, porcentaje de vidrio esmerilado, entre otros), fueron reemplazados con un marcador (-1). Por otra parte, en las 28 variables de anotaciones sobre los nódulos, los valores *NaN* fueron reemplazados por una categoría entera válida, debido a que estas características faltantes se relacionan a la presencia o ausencia de otras categorías como la fibrosis o enfisema.

El conjunto de imágenes radiológicas del *dataset NSCLC-Radiogenomics* incluye estudios de CT y PET, ambos originalmente en formato *DICOM* con una resolución de  $512 \times 512$  y  $192 \times 192$  respectivamente. Para este trabajo, se seleccionaron 104 pacientes que cumplieran con los criterios definidos anteriormente. Inicialmente, las modalidades CT y PET presentaban diferencias en resolución y orientación espacial. Para corregirlas, se empleó el software *3D Slicer*, tomando la tomografía CT como volumen de referencia. De esta manera, los volúmenes PET fueron reescalados y alineados al espacio de coordenadas de CT, garantizando la correspondencia espacial con los NP. Posteriormente, se realizaron correcciones de inversión de color en algunos estudios PET y se aplicó una normalización min-max para homogeneizar las intensidades por corte. Finalmente, todas las imágenes fueron almacenadas en matrices `.npy` y, para detección, estas imágenes fueron exportadas a formato `.png`.

Con el fin de reducir el sesgo causado por estructuras externas al parénquima pulmonar y el ruido de fondo de las imágenes originales. Adicionalmente, se consideraron 3 variantes del conjunto de datos, como se puede ver en la tabla 1. Para la segmentación se usó la librería de *LungMask* realizando ajustes para refinar la segmentación, se aplicaron operaciones morfológicas de cierre (*closing*), las cuales permitieron rellenar huecos en la máscara del parénquima pulmonar y suavizar los bordes. Para la máscara recortada, se utilizó la segmentación definiendo un margen variable según el tamaño del parénquima para que todas las imágenes cuenten con el mismo tamaño  $315 \times 315$  píxeles, tomando como referencia la segmentación binaria de mayor extensión que correspondía a ese tamaño. Finalmente, se generaron imágenes PET-CT sintéticas sobreponiéndolas bajo un parámetro de 0.5 en opacidad.

**Tabla 1.** Imágenes del *dataset NSCLC-Radiogenomics* en cada etapa de preprocesamiento para CT y PET

	Original	Segmentada	Segmentada recortada	Parches	Fusión
CT					
PET					

A partir de observaciones directas en *3D slicer* se tomó el corte axial medio con el fin de asegurar una región más representativa a nivel morfológico y una mayor concentración de captación metabólica, definiendo un rango  $z$  que abarca la anatomía y la actividad funcional del NP. Posteriormente, estas coordenadas fueron transformadas al formato *YOLO* para definir las *bounding boxes* (`clase`,  $x$ ,  $y$ ,  $w$ ,  $h$ ) y normalizadas respecto al tamaño original de cada imagen ( $512 \times 512$ ).

## 5.2. CONFIGURACIÓN DE LAS ARQUITECTURAS

**Localización** En la fase de detección se empleó el modelo *YOLOv9* sobre imágenes de tomografía (CT) previamente segmentadas y normalizadas por *slice*, con un tamaño de  $512 \times 512$  píxeles. El entrenamiento usó 300 épocas, con un tamaño de batch de 8 y una tasa de aprendizaje inicial de 0.00125. Además, antes de entrenar con el conjunto principal, se realizó un preentrenamiento con el dataset *LIDC-IDRI*, para después entrenar la arquitectura con el dataset *NSCLC-Radiogenomics*.

En la fase de reducción de falsos positivos se implementó una *ResNet18* adaptada a un enfoque multimodal, ya que recibe como entrada información tanto de CT como de PET. Esta red procesa parches de  $32 \times 32$  píxeles recortados con la ayuda de las coordenadas detectadas por *YOLO* en la fase anterior. De esta forma, la *ResNet18* actúa como un filtro adicional que clasifica cada parche en las categorías *nódulo* o *no-nódulo*. Además, el entrenamiento utilizó el optimizador Adam junto con la función de pérdida *Cross Entropy*.

**Clasificación supervisada multimodal** Para clasificación usando el modelo multimodal transformer, se usó un optimizador Adam, con una tasa de aprendizaje de 0.001, la función de *Cross Entropy* como pérdida, se entrenó con 140 épocas y un *batch size* de 32. Para la rama de CT, el *Vision Transformer* tiene un *patch size* de 1, dimensión del embebido de 512, 8 cabezas y 2 bloques de *Transformers* como profundidad con un *MLP ratio* de 4. Para la rama de PET el *Vision Transformer* está configurado con un *patch size* de 4, dimensión del embebido de 64, 4 cabezas y 2 bloques de *Transformers*, con un *MLP ratio* de 4. Finalmente, para la rama tabular se usó una dimensión del embebido de 64, embebidos específicos de 32, mientras que los embebidos compartidos son de 32 igualmente, un número de cabezas de 8 y 4 bloques de *Transformers* como profundidad, y la dimensión del MLP con 128.

**Clasificación autosupervisada** Para la síntesis de imágenes se empleó el optimizador Adam con tasa de aprendizaje de 0.0002 y  $\beta = (0.5, 0.999)$ , y el modo adversario LSGAN. La función objetivo combinó pérdida adversaria, consistencia cíclica  $\ell_1$  con  $\lambda_{cyc} = 10$  y identidad con  $\lambda_{id} = 5$ . Se entrenó por 300 épocas, evaluando métricas cada 10, con batch size de 8, sobre tensores de entrada/salida de  $3 \times 256 \times 256$ .

A partir del generador  $G_{CT-PET}$  previamente entrenado se extrajeron las representaciones latentes de dimensión 256 mediante una operación de *global average pooling* en la última capa convolucional, seguidos de una normalización  $\ell_2$ . Estos vectores sirvieron como descriptores de cada paciente y se emplearon como entrada a un clasificador de Regresión Logística. El cual se configuró con un esquema de regularización estándar ( $C = 1.0$ ), con un máximo de

2000 iteraciones. Previo al ajuste del modelo, cada vector de características fue normalizado mediante *Standard Scaler*, obteniendo media cero y varianza 1. El entrenamiento se realizó de manera independiente en cada *fold*.

### 5.3. VALIDACIÓN

Para la validación de los modelos propuestos se usó una validación cruzada *Stratified K-fold* con un  $K=5$  y un *random state* = 42 con el fin de asegurar la reproducibilidad y experimentar con la misma distribución de pacientes para el desarrollo de los objetivos. En el conjunto con 104 pacientes, esta configuración nos deja con una partición de 83 pacientes en el entrenamiento y 21 de prueba por *fold*. En el caso del conjunto de 56 pacientes, la partición resultante es de 44 pacientes para el entrenamiento y 12 de prueba por *fold*.

Adicionalmente, en la fase de preentrenamiento del modelo de detección se empleó el conjunto público *LIDC-IDRI*, que fue dividido en dos subconjuntos: 11.010 imágenes para entrenamiento y 1.599 imágenes para validación, con una resolución de entrada de  $512 \times 512$  píxeles. En este caso, no se aplicó validación cruzada, ya que el objetivo fue únicamente proveer un entrenamiento inicial robusto al detector antes de transferirlo al conjunto *NSCLC-Radiogenomics*.

## 6. EVALUACIÓN Y RESULTADOS

Se evaluaron las diferentes etapas afectadas en la ruta de atención, de forma independiente, *i.e.*, localización y clasificación de NP, considerando CT, PET, datos clínicos y su correspondiente mutación genómica (EGFR o KRAS). A continuación se detallan los resultados:

### 6.1. EVALUACIÓN DE MODELOS DE DETECCIÓN

En primera instancia los experimentos realizados para la detección de NP, consideraron tanto el desempeño del modelo *YOLO* y su posterior reducción de falsos positivos mediante el uso de PET. En esta sección se presentan los resultados de los experimentos unimodales realizados con diferentes arquitecturas de la familia *YOLO* (v5, v8, v9 y v10). Se evaluaron tres configuraciones de entrada: imágenes CT y PET originales, imágenes de CT y PET segmentadas para delimitar el parénquima pulmonar e imágenes PET-CT resultantes de una fusión simple, como un primer acercamiento a la integración multimodal.

Todas las métricas corresponden al promedio y desviación estándar reportados por la validación automática de *YOLO* en un esquema de 5 *folds*. Estos resultados se emplearon exclusivamente para comparar el desempeño relativo entre arquitecturas y determinar el mejor modelo de referencia para los siguientes experimentos, sin considerarse aún como la evaluación final del sistema propuesto. En la modalidad CT original (Tabla 2), el mejor mAP50 lo obtuvo YOLOv5s\_coco ( $0.602 \pm 0.044$ ), donde *coco* hace referencia a pesos preentrenados en el conjunto MSCOCO para detección generalizada, además del mayor recall ( $0.571 \pm 0.059$ ). Aunque YOLOv8s\_coco presentó la mayor precisión ( $0.773 \pm 0.069$ ).

**Tabla 2.** Resultados para la modalidad CT original (validación automática YOLO, 5 folds)

Modelo	Precisión	Recall	mAP50
YOLOv5s_coco	$0.711 \pm 0.098$	<b><math>0.571 \pm 0.059</math></b>	<b><math>0.602 \pm 0.044</math></b>
YOLOv8s_coco	<b><math>0.773 \pm 0.069</math></b>	$0.526 \pm 0.080$	$0.579 \pm 0.085$
YOLOv9s_coco	$0.755 \pm 0.126$	$0.516 \pm 0.080$	$0.571 \pm 0.058$
YOLOv10s_coco	$0.748 \pm 0.145$	$0.493 \pm 0.086$	$0.555 \pm 0.087$

En la modalidad PET original (Tabla3), el mejor mAP50 lo obtuvo YOLOv5s\_coco ( $0.583 \pm 0.143$ ), junto con el mayor recall ( $0.565 \pm 0.129$ ). Aunque YOLOv9s\_coco registró la mayor precisión ( $0.852 \pm 0.146$ ).

**Tabla 3.** Resultados para la modalidad PET original (validación automática YOLO, 5 folds).

Modelo	Precisión	Recall	mAP50
YOLOv5s_coco	$0.801 \pm 0.136$	<b><math>0.565 \pm 0.129</math></b>	<b><math>0.583 \pm 0.143</math></b>
YOLOv8s_coco	$0.655 \pm 0.159$	$0.477 \pm 0.166$	$0.501 \pm 0.196$
YOLOv9s_coco	<b><math>0.852 \pm 0.146</math></b>	$0.485 \pm 0.154$	$0.528 \pm 0.159$
YOLOv10s_coco	$0.676 \pm 0.272$	$0.463 \pm 0.145$	$0.487 \pm 0.148$

En la modalidad PET-CT fusionada (Tabla 4), el mejor mAP50 lo obtiene YOLOv5s\_coco ( $0.618 \pm 0.202$ ), que además presenta el mayor recall ( $0.582 \pm 0.207$ ). Frente a las modalidades unimodales, la fusión aporta una ganancia moderada en mAP50 respecto a **CT original** ( $0.602$ ) y **PET original** ( $0.583$ ), manteniendo un compromiso favorable entre sensibilidad y localización a  $\text{IoU} \geq 0.5$ .

**Tabla 4.** Resultados para la modalidad PET-CT fusionada con  $\alpha = 0.5$  (validación automática YOLO, 5 folds).

Modelo	Precisión	Recall	mAP50
YOLOv5s_coco	$0.808 \pm 0.153$	<b><math>0.582 \pm 0.207</math></b>	<b><math>0.618 \pm 0.202</math></b>
YOLOv8s_coco	$0.732 \pm 0.176$	$0.573 \pm 0.109$	$0.616 \pm 0.162$
YOLOv9s_coco	$0.790 \pm 0.156$	$0.556 \pm 0.111$	$0.613 \pm 0.172$
YOLOv10s_coco	<b><math>0.821 \pm 0.100</math></b>	$0.538 \pm 0.227$	$0.583 \pm 0.190$

En CT segmentada (Tabla 5), con mAP50 como métrica prioritaria, el mejor resultado lo alcanza YOLOv9s\_COCO+LIDC ( $0.697 \pm 0.074$ ), que además presenta el mayor recall

( $0.663 \pm 0.064$ ). Donde *COCO+LIDC* hace referencia al uso de pesos preentrenados en el conjunto MSCOCO para detección generalizada, posteriormente ajustados con LIDC para un preentrenamiento específico del dominio. Dentro de los modelos *solo COCO*, YOLOv8s\_coco y YOLOv9s\_coco quedan en *empate técnico* en mAP50 (0.669 vs. 0.667), lo que motivó evaluar preentrenamiento específico de dominio. Ese sesgo de dominio desplaza la frontera: en v9s la estrategia COCO+LIDC mejora mAP50 respecto a COCO (+0.030), mientras que en v8s el LIDC puro supera a COCO+LIDC (0.673 vs. 0.628), sugiriendo que la combinación con COCO no siempre es ventajosa para esa arquitectura. Comparado con CT original (mejor mAP50 = 0.602), la segmentación del parénquima aporta una ganancia en mAP50 de  $\approx 0.095$  (0.697–0.602), indicando que restringir el campo visual reduce el ruido de fondo y favorece la localización correcta a  $\text{IoU} \geq 0.5$ . En consecuencia, se selecciona YOLOv9s\_COCO+LIDC como modelo de referencia para CT segmentada.

**Tabla 5.** Resultados unificados para CT segmentada (validación automática YOLO, 5 folds).

Modelo	Precisión	Recall	mAP50
YOLOv5s_coco	$0.783 \pm 0.129$	$0.606 \pm 0.086$	$0.654 \pm 0.112$
YOLOv8s_coco	$0.782 \pm 0.087$	$0.626 \pm 0.158$	$0.669 \pm 0.154$
YOLOv8s_LIDC_puro	$0.713 \pm 0.174$	$0.623 \pm 0.134$	$0.673 \pm 0.160$
YOLOv8s_COCO+LIDC	$0.794 \pm 0.074$	$0.608 \pm 0.155$	$0.628 \pm 0.118$
YOLOv9s_coco	<b><math>0.796 \pm 0.068</math></b>	$0.588 \pm 0.141$	$0.667 \pm 0.131$
YOLOv9s_LIDC_puro	$0.691 \pm 0.162$	$0.641 \pm 0.112$	$0.634 \pm 0.156$
YOLOv9s_COCO+LIDC	$0.713 \pm 0.166$	<b><math>0.663 \pm 0.064</math></b>	<b><math>0.697 \pm 0.074</math></b>
YOLOv10s_coco	$0.729 \pm 0.165$	$0.598 \pm 0.088$	$0.649 \pm 0.117$

En la modalidad PET segmentada (Tabla 6), el mejor **mAP50** lo obtuvo YOLOv10s\_coco ( $0.573 \pm 0.167$ ), además de registrar la mayor precisión ( $0.755 \pm 0.201$ ). Aunque YOLOv8s\_coco presentó el recall más alto ( $0.569 \pm 0.148$ ), por ello se selecciona YOLOv10s\_coco para PET segmentada.

**Tabla 6.** Resultados para la modalidad PET segmentada (validación automática YOLO, 5 folds).

Modelo	Precisión	Recall	mAP50
YOLOv5s_coco	$0.651 \pm 0.195$	$0.503 \pm 0.172$	$0.509 \pm 0.183$
YOLOv8s_coco	$0.650 \pm 0.169$	<b><math>0.569 \pm 0.148</math></b>	$0.560 \pm 0.158$
YOLOv9s_coco	$0.694 \pm 0.222$	$0.520 \pm 0.125$	$0.519 \pm 0.136$
YOLOv10s_coco	<b><math>0.755 \pm 0.201</math></b>	$0.543 \pm 0.130$	<b><math>0.573 \pm 0.167</math></b>

En síntesis, los resultados confirman que la CT segmentada ofrece ventajas frente a las demás configuraciones y que YOLOv9s\_COCO+LIDC alcanza el mejor balance entre sensibilidad y desempeño global. Lo anterior puede justificarse con la riqueza textural superior, observada en el CT, con respecto a las secuencias PET. Además, como en la rutina clínica esta es la secuencia ideal para obtener una localización de los nódulos, y luego se procede con su caracterización involucrando otras modalidades como el PET. Es por ello, que tomando estos resultados como referencia, se incorporó un módulo de reducción de falsos positivos mediante un clasificador *ResNet-18* entrenado con parches PET extraídos a partir de las detecciones en CT. En validación cruzada (5 *folds*), este modelo alcanzó un AUC de  $0.894 \pm 0.032$ , con una sensibilidad promedio de  $0.920 \pm 0.075$  y una precisión de  $0.646 \pm 0.221$ , lo que se tradujo en un F1-score de  $0.729 \pm 0.179$ . Dichos valores evidencian una alta capacidad para recuperar verdaderos positivos, a la vez que aporta una reducción efectiva de los falsos positivos, cumpliendo así su rol como filtro complementario al detector inicial.

Posteriormente, se evaluó el impacto del RFP sobre la detección; las Tablas 7 y 8 sintetizan los resultados en validación cruzada (5 *folds*), reportando Precisión, Recall y mAP (media  $\pm$  desviación estándar) a IoU 0.25 y 0.50.

**Tabla 7.** Resultados **pre-reductor** en detección (5 *folds*).

IoU	Precisión	Recall	mAP
0.25	$0.417 \pm 0.236$	$0.861 \pm 0.108$	$0.660 \pm 0.072$
0.50	$0.373 \pm 0.234$	$0.735 \pm 0.132$	$0.721 \pm 0.029$

**Tabla 8.** Resultados **post**-reductor en detección (5 *fold*s).

IoU	Precisión	Recall	mAP
0.25	$0.646 \pm 0.247$	$0.788 \pm 0.093$	$0.679 \pm 0.081$
0.50	$0.568 \pm 0.256$	$0.680 \pm 0.067$	$0.733 \pm 0.043$

En conjunto, el RFP incrementó la **precisión** de  $0.417 \rightarrow 0.646$  (+0.229) a IoU 0.25 y de  $0.373 \rightarrow 0.568$  (+0.195) a IoU 0.50, con una reducción moderada del *recall* ( $-0.073$  y  $-0.055$ , respectivamente). Este intercambio se tradujo en una mejora neta del **mAP** (+0.019 a IoU 0.25; +0.012 a IoU 0.50), confirmando que el filtro reduce de forma efectiva los falsos positivos sin afectar de manera sustantiva la sensibilidad global.

## 6.2. EVALUACIÓN DE MODELOS DE CLASIFICACIÓN SUPERVISADA Y AUTOSUPERVISADA

En esta sección se presentan los resultados de clasificación, abordados como un problema de clasificación binaria (EGFR/KRAS), desarrollados bajo una estructura progresiva, desde una perspectiva supervisada unimodal para luego exponer la capacidad del modelo multimodal propuesto. En complemento con este enfoque binario, también se realizaron experimentos con un problema multiclase y con un enfoque autosupervisado teniendo en cuenta las representaciones latentes. A continuación, se detallan los resultados obtenidos.

### Evaluación de modelos de clasificación supervisada

**Evaluación de modelos de datos clínicos** Para este apartado, primero se implementaron modelos clásicos de aprendizaje de máquina como *Random Forest* y aprendizaje profundo como Redes Neuronales Densas (*DNN*, por sus siglas en inglés). Esta evaluación se realizó primero con los datos por separado para estudiar la capacidad discriminatoria que puede tener la información para clasificar entre mutaciones y después se realizaron pruebas integrando ambos datos en una sola matriz para ver si pueden complementarse y lograr mejorar los

resultados individuales. Estos análisis se realizaron en diferentes tipos de codificación: *One-hot encoding* y *Label encoding*. En Tabla 9 se reportan todos los resultados obtenidos después de hacer múltiples y exhaustivos experimentos con todas las configuraciones anteriormente mencionadas. Alternativamente, también se probó con métodos más avanzados de codificación como los vectores embebidos, estos solo aceptan números positivos para codificar, por lo que los datos faltantes fueron convertidos en la categoría siguiente en cada columna de forma dinámica. Esta estrategia fue comparada con el *Tab Transformer*, ver Tabla 9.

**Tabla 9.** Estudio de diversos modelos ante diferentes métodos de codificación de datos para modalidades clínicas, anotaciones y fusión de ambas. Se reportan las métricas de especificidad, F1-Score y AUC (media  $\pm$  desviación estándar) calculadas a partir de los diferentes *folders*.

Modalidad	Modelo	Codificación	Especificidad	F1-Score	AUC
Clínico	Random Forest	Label	<b>0.838 <math>\pm</math> 0.105</b>	0.740 $\pm$ 0.174	0.881 $\pm$ 0.081
	DNN	Label	0.738 $\pm$ 0.153	0.700 $\pm$ 0.082	0.873 $\pm$ 0.110
	Random Forest	One Hot	0.737 $\pm$ 0.163	0.729 $\pm$ 0.163	0.870 $\pm$ 0.100
	DNN	One Hot	0.798 $\pm$ 0.157	<b>0.798 <math>\pm</math> 0.158</b>	<b>0.907 <math>\pm</math> 0.084</b>
	Embeddings	Embeddings	0.776 $\pm$ 0.142	0.767 $\pm$ 0.181	0.879 $\pm$ 0.084
	Tab Transformer	Embeddings	0.742 $\pm$ 0.084	0.773 $\pm$ 0.117	0.861 $\pm$ 0.101
Anotaciones	Random Forest	Label	0.809 $\pm$ 0.115	0.776 $\pm$ 0.062	0.904 $\pm$ 0.065
	DNN	Label	0.790 $\pm$ 0.294	0.726 $\pm$ 0.058	0.905 $\pm$ 0.056
	Random Forest	One Hot	0.838 $\pm$ 0.072	<b>0.834 <math>\pm</math> 0.071</b>	0.923 $\pm$ 0.054
	DNN	One Hot	0.774 $\pm$ 0.075	0.757 $\pm$ 0.093	<b>0.931 <math>\pm</math> 0.055</b>
	Embeddings	Embeddings	0.842 $\pm$ 0.101	0.793 $\pm$ 0.152	0.920 $\pm$ 0.086
	Tab Transformer	Embeddings	<b>0.904 <math>\pm</math> 0.087</b>	0.831 $\pm$ 0.123	0.927 $\pm$ 0.070
Fusión	Random Forest	Label	0.871 $\pm$ 0.064	0.833 $\pm$ 0.064	0.897 $\pm$ 0.063
	DNN	Label	0.804 $\pm$ 0.141	0.784 $\pm$ 0.109	0.911 $\pm$ 0.057
	Random Forest	One Hot	0.781 $\pm$ 0.074	0.779 $\pm$ 0.072	0.905 $\pm$ 0.060
	DNN	One Hot	0.858 $\pm$ 0.098	<b>0.855 <math>\pm</math> 0.099</b>	0.933 $\pm$ 0.081
	Embeddings	Embeddings	0.871 $\pm$ 0.072	0.785 $\pm$ 0.088	<b>0.942 <math>\pm</math> 0.052</b>
	Tab Transformer	Embeddings	<b>0.876 <math>\pm</math> 0.123</b>	0.810 $\pm$ 0.138	0.939 $\pm$ 0.084

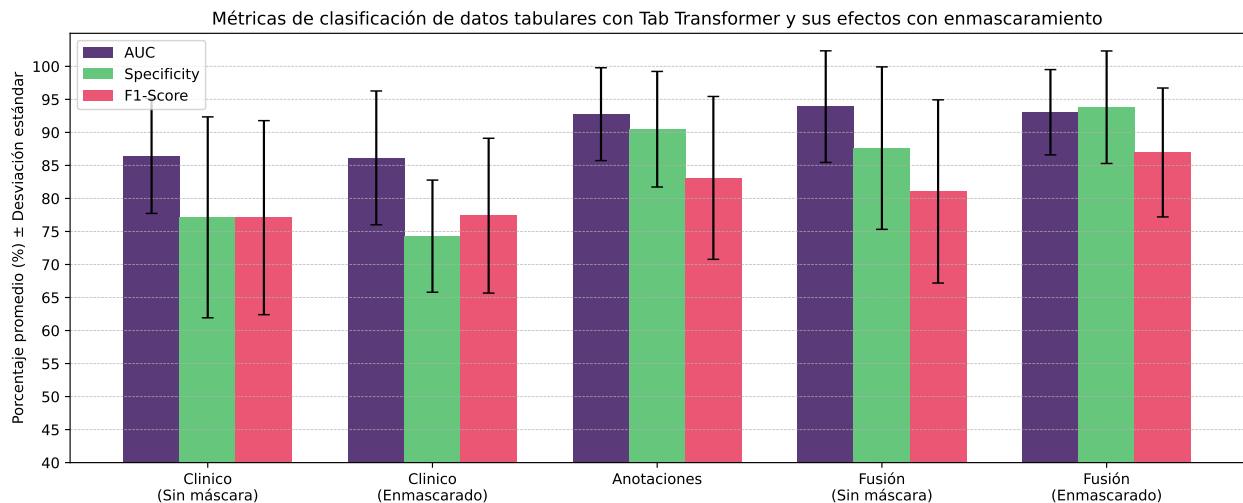
En este primer estudio, ver Tabla 9, se evidenció un desempeño limitado con los datos clínicos, probablemente debido al ruido en la clasificación generado por la imputación de información faltante en variables relacionadas con el estadio TNM y grado histopatológico. Esto contrasta con las anotaciones del nódulo, al estar completas y directamente vinculadas a fenotipos tumorales, alcanzaron un rendimiento superior. Sin embargo, al integrar ambas modalidades

se obtuvo una mejora en la discriminación de clases.

Para posteriores experimentos, se adoptó la arquitectura *Tab Transformer* como base para los datos tabulares del modelo multimodal, debido a que destacó en las modalidades de anotaciones y fusiones con la especificidad más alta, lo que indica que prioriza la distinción de la clase KRAS, tiene el segundo mejor AUC el cuál está sobre 90% lo que implica que su capacidad de discriminación entre las clases es competente y el F1-score es mayor a 80%.

Finalmente, se evaluó el *Tab Transformer* con la integración de la máscara, como se ve en la Figura 18, en la cual se puede evidenciar desde los datos clínicos y la fusión una mayor estabilidad en los modelos en la desviación estándar e incluso un aumento en las métricas de especificidad y F1-score, sobre todo en la fusión con los datos clínicos. La diferencia entre AUC entre el *Tab Transformer* con la fusión de datos sin máscara y con máscara es de 0.0085 puntos, siendo  $0,939 \pm 0,0845$  y  $0,9305 \pm 0,0646$  respectivamente.

**Figura 18.** Comparación usando el *Tab Transformer* de manera unimodal e implementando la fusión temprana entre datos clínicos y anotaciones. La figura muestra los efectos de integrar una máscara de atención para ignorar los datos faltantes que hay en el conjunto de datos clínico.

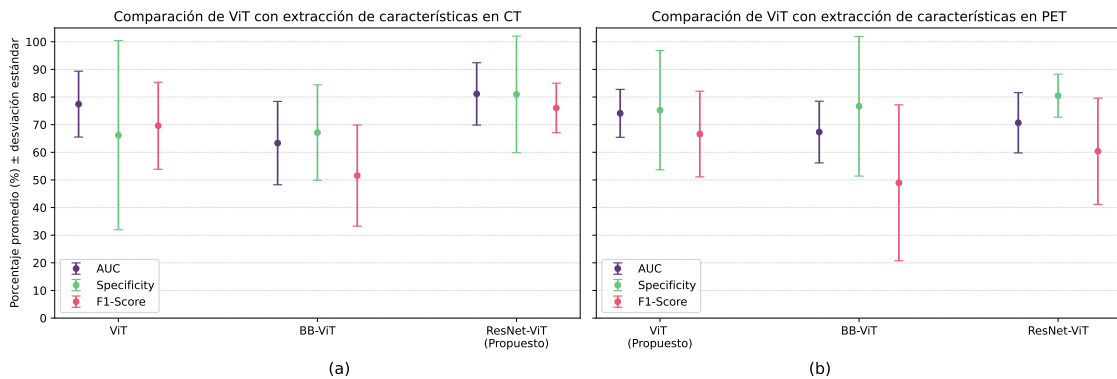


**Evaluación de modelos CT y PET.** Para el análisis del desempeño de estas ramas de procesamiento y codificación visual, se realizó una comparación con un *Vision Transformer*,

variando su extractor de características. En la Figura 19-a se observa que para la modalidad CT, se presentó variabilidad en los métodos, particularmente en la especificidad. En este caso, cabe resaltar el *Vision Transformer* con extractor de características para la rama de CT, el cual reporta los resultados mas estables y coherentes con respecto a la clasificación de las mutaciones. Este modelo fué seleccionado como la rama para la integración, en el esquema multimodal, ente otras, porque permite capturar la anatomía de los nódulos en capas iniciales, lo que facilita su clasificación.

Por otra parte, para las imágenes PET (ver Figura 19-b) se adoptó la arquitectura de *Vision Transformer* sin extractor de características dado que, por su naturaleza, no es necesario extraer características morfológicas, sino capturar los patrones funcionales, representados directamente en la imagen PET como hiperintensidades. Además, el *Vision Transformer* presenta un mejor desempeño en términos de AUC y F1-score.

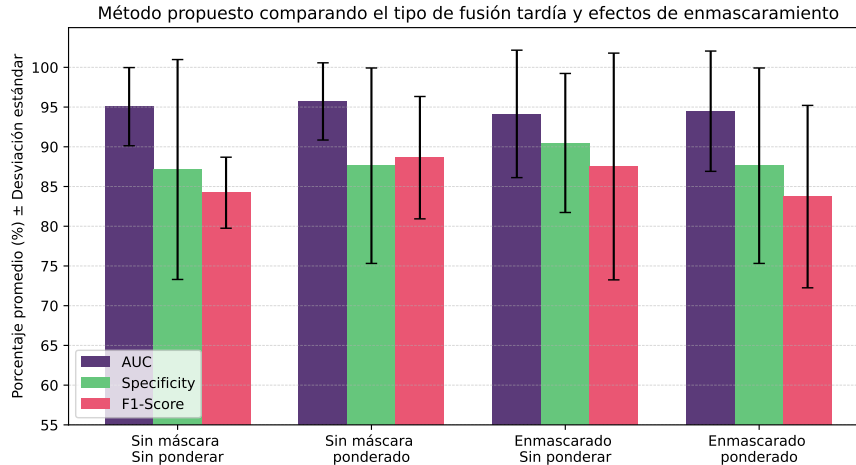
**Figura 19.** Comparación del rendimiento de AUC, Especificidad y F1-Score de un *Vision Transformer* para (a) parches de CT y (b) parches de PET, con y sin un extractor de características, calculadas a partir de los diferentes  *folds*



**Evaluación de modelos multimodales** Teniendo en cuenta el rendimiento y las limitaciones que poseen cada una de las fuentes de información de forma aislada para caracterizar mutaciones genómicas, en una tercera etapa de evaluación se buscó explotar la complementariedad que pueden ofrecer estos datos al integrar diferentes perspectivas del diagnóstico, robusteciendo el método de clasificación mediante una fusión tardía. Para evaluar este método

multimodal, se realizaron dos formas: Ponderando qué tanta información puede aportar cada rama  $\alpha \cdot Y^{CT} + \beta \cdot Y^{PET} + \gamma \cdot Y^{TAB} = 1$  y sin realizar una ponderación. De forma conjunta, se evaluó qué tanto aporta el método de enmascaramiento realizado para la rama tabular de los datos perdidos a la arquitectura multimodal, como se puede ver en la Figura 20. Cabe destacar los resultados obtenidos, en las diferentes configuraciones evaluadas, con un contundente AUC que evidencia la capacidad para separar las muestras evaluadas, según las etiquetas de mutación. Además, la ponderación entre las ramas mejora el rendimiento de la arquitectura que no aplica enmascaramiento. Por otra parte, parece que la ponderación y el uso de máscaras podrían afectar principalmente la relación entre la especificidad y el F1-score, siendo el uso del método de enmascaramiento reductor de la variabilidad en la especificidad, pero aumenta la inestabilidad en el F1-score.

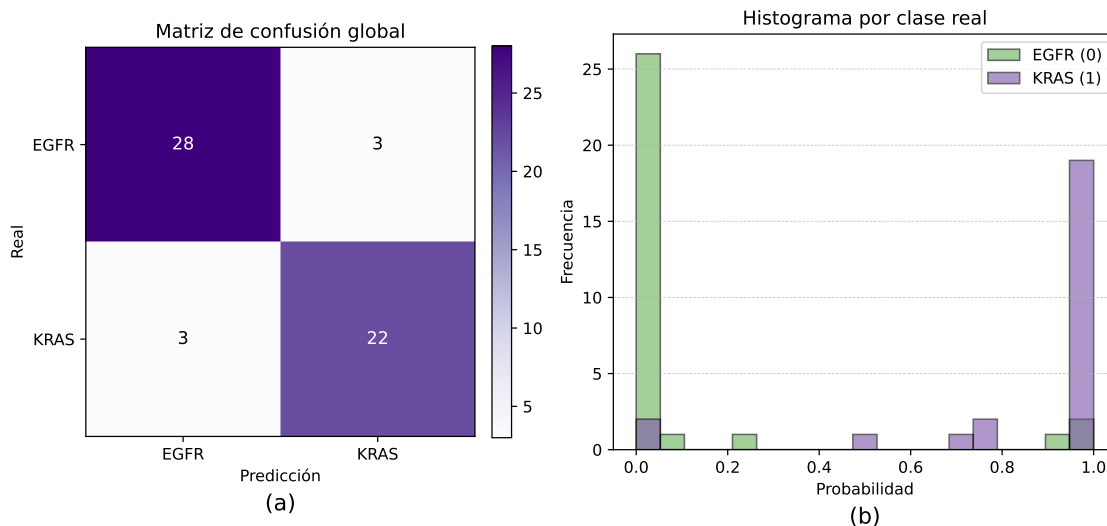
**Figura 20.** Gráfica de comparativa entre las estrategias de fusión tardía: Logits ponderados y sin ponderar para realizar la clasificación, y los efectos de la inclusión de una máscara para reducir los efectos de sesgo por imputación. Las métricas que se usaron de comparación son AUC, Especificidad y F1-Score. El método propuesto corresponde al tercer modelo de enmascaramiento sin ponderación.



En una mayor descripción de los resultados, en la Figura 21-a, se puede observar como desde la matriz de confusión el modelo es capaz de acertar en la gran mayoría de pacientes por lo que distingue muy bien entre ambas clases, indicando un buen equilibrio entre sensibilidad y especificidad. De forma complementaria, en la Figura 21b para la clase EGFR, la mayoría de

probabilidades están asignadas a la clase correcta. De manera similar, para la clase KRAS, la mayoría de las probabilidades están cerca de 1. No obstante, existen casos en los que, de pronto ajustando el umbral, podrían ser clasificados de una mejor manera como se puede ver en la Figura 21c.

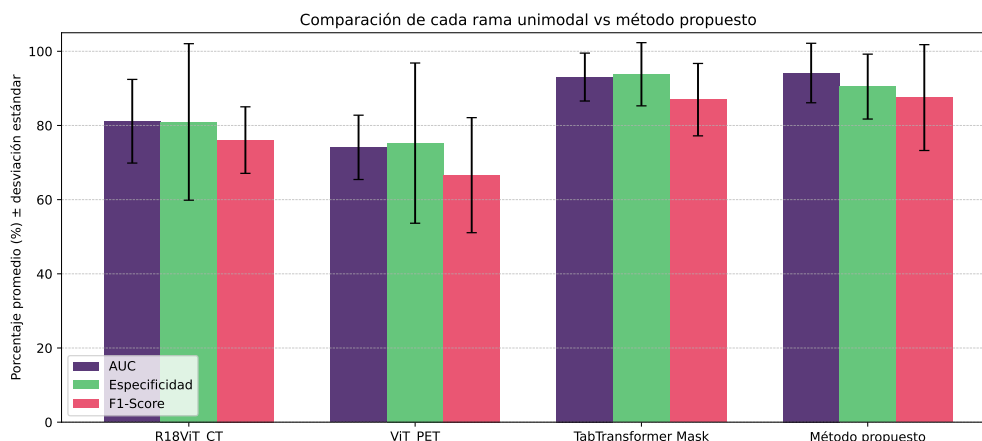
**Figura 21.** (a) La matriz de confusión global del método propuesto, (b) Histograma complementario para ver a detalle las predicciones del modelo. Donde la clase 0 corresponde a la clase EGFR y la clase 1 corresponde a la clase KRAS.



De igual forma, si comparamos el rendimiento de cada rama comparada con la arquitectura propuesta, como en la Figura 22, el método propuesto supera las ramas imagenológicas de CT y PET en todas sus métricas mostrando un método mucho más estable, mientras que en la rama tabular solo es superado en especificidad y una variación menor en F1-score y AUC. Por lo cual, se puede especificar que la integración multi-modal de diferentes dominios (imagenológicos y tabulares) permite capturar patrones discriminativos y determinantes de cada una de las mutaciones genómicas (EGFR y KRAS).

**Evaluación de método propuesto en un problema multiclase** En una experimentación extra, incluida en este trabajo, se utilizó el método propuesto en un problema de clasificación multiclase, en donde se pone a prueba la evaluación de una tercera clase además de EGFR

**Figura 22.** Comparación en métricas de AUC, Especificidad y F1-Score de los modelos usados en cada rama y en conjunto con el método propuesto.



y KRAS: Desconocido o Indeterminado. En la Tabla 10 se reportan los resultados obtenidos, comparando el método sin y con método de enmascaramiento. Como se puede ver, las métricas bajan considerablemente, lo cual se puede atribuir a varias razones: (1) la cantidad limitada de datos, o (2) la ambigüedad de las características tanto imagenológicas como tabulares en la clase "desconocido", limitando la caracterización apropiada de las mutaciones. Esto plantea la necesidad de ahondar en esta línea de trabajo, para investigaciones siguientes donde se pueda hacer un diseño experimental que permita la caracterización de múltiples clases.

**Tabla 10.** Métricas del modelo propuesto para el problema multiclase: EGFR, KRAS, Desconocido.

Modelo	Accuracy (%)	Precision (%)	Recall (%)	Especificidad (%)	F1-Score (%)	AUC (%)
Propuesto	0.586 ± 0.123	0.585 ± 0.132	0.584 ± 0.113	0.788 ± 0.058	0.573 ± 0.117	0.709 ± 0.101
Propuesto sin máscara	0.596 ± 0.107	0.602 ± 0.116	0.593 ± 0.126	0.793 ± 0.060	0.572 ± 0.112	0.737 ± 0.128

### Evaluación de modelos autosupervisados para la clasificación de mutaciones genómicas.

Como se presentó en la metodología, en este trabajo también se exploró una alternativa autosupervisada, bajo un enfoque de traducción CT-PET, para recuperar descriptores con carácter discriminatorio del tipo de mutación. En esta sección se presentan los resultados obtenidos por el enfoque en cuanto a la clasificación de mutaciones genómicas, en el cual se

aprovechó la síntesis para la extracción de características discriminativas. En los experimentos se consideraron tres versiones de imágenes: *original*, *segmentada* y *segmentada recortada*.

La Tabla 11 muestra el desempeño de los clasificadores a partir de vectores latentes extraídos en el quinto bloque de la *ResNet*, empleando la síntesis de imágenes con *CycleGAN*. La configuración con el conjunto *original*, pese a que el *Random Forest* (*RF*, por sus siglas en inglés) y la máquina de soporte vectorial (*SVM*, por sus siglas en inglés) alcanzaron valores puntuales competitivos en especificidad y F1, la regresión lineal (*LR*, por sus siglas en inglés) con la configuración de *segmentada recortada* mostró un rendimiento más estable y equilibrado en todas las métricas, a la vez que contó con un AUC más significativo ( $0.846 \pm 0.090$ ). En *segmentada*, la *LR* se destacó al superar en consistencia a los demás clasificadores, mostrando que la segmentación con *LungMask* puede aportar un factor discriminativo, alcanzando un AUC de 0.812. Finalmente, en *mascara recortada*, la *LR* mostró su ventaja al obtener los resultados globales más estables (Especificidad = 0.720, F1 = 0.688 y AUC = 0.846), siendo similares a los de *original* pero complementados con el AUC, lo que indica una mayor capacidad de discriminación entre mutaciones EGFR y KRAS cuando se concentra la atención en el parénquima pulmonar.

Apoyado en esto, los hallazgos en clasificación tuvieron coherencia con los resultados en la fase de síntesis, donde en la comparación global de las traducciones, *CycleGAN con las imágenes segmentadas* había mostrado los mejores resultados, lo que guió hacia la variante con la *segmentada recortada*, que, aunque no presentó los mejores valores al compararlos directamente, en clasificación resaltó un aspecto importante en el análisis del rendimiento, ya que demostró el mejor desempeño. Este contraste indica que, pese a que las métricas no mostraron una calidad visual elevada en síntesis, al centrar la imagen en el parénquima pulmonar se reveló que las representaciones latentes extraídas mediante la *segmentada recortada* resultan más discriminativas, aportando mayor información para la tarea de predicción de mutaciones.

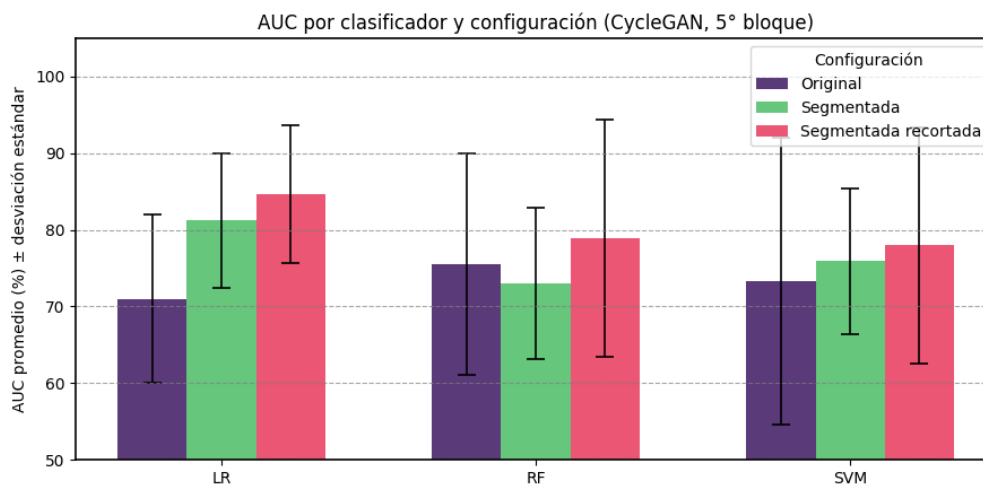
Finalmente, la Figura 23 resume los valores más representativos de AUC para cada arquitec-

**Tabla 11.** Resultados de clasificación utilizando vectores latentes extraídos en el 5° bloque de la ResNet a partir de imágenes generadas con CycleGAN. Se reportan métricas de especificidad, F1 y AUC (media  $\pm$  desviación estándar, validación cruzada en 5 folds) bajo tres configuraciones: Original, segmentada y Segmentada recortada.

Configuración	Clasificador	Especificidad	F1-score	AUC
Original	LR	0.678 $\pm$ 0.140	0.629 $\pm$ 0.149	0.710 $\pm$ 0.110
	RF	<b>0.726 <math>\pm</math> 0.131</b>	0.678 $\pm$ 0.133	0.755 $\pm$ 0.144
	SVM	0.720 $\pm$ 0.142	<b>0.692 <math>\pm</math> 0.133</b>	0.733 $\pm$ 0.187
Segmentada	LR	0.697 $\pm$ 0.077	0.660 $\pm$ 0.106	<b>0.812 <math>\pm</math> 0.087</b>
	RF	0.683 $\pm$ 0.056	0.648 $\pm$ 0.074	0.730 $\pm$ 0.099
	SVM	0.638 $\pm$ 0.112	0.599 $\pm$ 0.101	0.759 $\pm$ 0.095
Segmentada recortada	LR	<b>0.720 <math>\pm</math> 0.137</b>	<b>0.688 <math>\pm</math> 0.166</b>	<b>0.846 <math>\pm</math> 0.090</b>
	RF	0.690 $\pm$ 0.169	0.632 $\pm$ 0.202	0.789 $\pm$ 0.155
	SVM	0.650 $\pm$ 0.196	0.621 $\pm$ 0.205	0.780 $\pm$ 0.154

tura, mostrando que *CycleGAN* fue el modelo que produjo *embeddings* más discriminativos para la clasificación entre mutaciones. En contraste, pese a que la *segmentada recortada* se reflejó en todos los modelos de clasificación, la *LR* mostró el mejor desempeño en las versiones que hacen uso de la máscara extraída con *LungMask*, evidenciando que la fidelidad visual no tiene el mismo peso en tareas discriminativas.

**Figura 23.** Comparación entre métricas de AUC más representativas para cada método de clasificación propuesto en la arquitectura de síntesis.



## 7. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo de investigación, se abordó un flujo de trabajo para localizar y caracterizar mutaciones genómicas desde observaciones en imágenes CT, PET y datos tabulares. En detección, se utilizó la *YOLOv9* sobre imágenes CT segmentadas para localizar los potenciales candidatos de nódulos pulmonares y así posteriormente integrar un clasificador *ResNet18* con parches PET el cual permitió reducir falsos positivos, mostrando la utilidad de combinar información morfológica y funcional. Este esquema evidencia que la segmentación en CT aporta una base anatómica más focalizada en la región de interés para la detección, mientras que PET refuerza la confiabilidad del sistema al actuar como filtro adicional.

En clasificación supervisada, se logró implementar una arquitectura multimodal basado en *Transformers* con imágenes CT, PET y con datos clínicos. Se desarrollaron experimentos exhaustivos que permitieron evidenciar el gran potencial de integrar estos tres componentes de información para brindar un soporte a la clasificación de las mutaciones genómicas. Además, se logró constatar que en las ramas independientes se aprenden patrones con cierta correlación a estas mutaciones, lo cual es un hallazgo prometedor, como herramienta extra en la caracterización de los nódulos. También en experimentos extras se mostró cierta capacidad para hacer la clasificación en múltiples clases, pero con ciertas limitaciones, teniendo en cuenta el conjunto de datos considerados en este trabajo.

En la clasificación autosupervisada de mutaciones genómicas EGFR y KRAS, la *CycleGAN* mostró un comportamiento prometedor para la clasificación, siendo sus descriptores latentes discriminativos, alcanzando en la configuración *segmentada recortada* y con el clasificador de *Regresión Logística* los valores más estables de especificidad, F1 y AUC. En conjunto, estos hallazgos muestran que la combinación *CycleGAN + segmentada recortada + LR* constituye la estrategia más apropiada para la fase de clasificación. Considerando que durante el entrenamiento, la arquitectura no tuvo consideraciones de información genómica, los resultados son contundentes al evidenciar una codificación textural y funcional que tiene relación con esta

información. Lo anterior complementa los hallazgos supervisados y podrían abrir una línea de trabajo para aprender firmas genómicas a partir de la codificación visual de las imágenes. Finalmente, como trabajo a futuro, desde la perspectiva de detección se plantea explorar estrategias de fusión multimodal intermedias o tardías directamente dentro de la arquitectura *YOLO*, de manera que la información PET complemente de forma más eficaz a la CT en etapas avanzadas de extracción de características. Del mismo modo, desde la perspectiva de clasificación, empezando desde el aporte supervisado, se contempla la posibilidad de explorar enfoques de atención cruzada debido a que todas las ramas están basados en *Transformers* con mecanismos de atención, así como implementar otros extractores de características y abordar la clasificación desde problemas multiclase. A partir de los hallazgos obtenidos en los modelos de síntesis y la clasificación autosupervisada, es importante analizar distintas capas de los generadores para identificar qué niveles de abstracción aportan mayor información discriminativa. Además, se considera integrar los métodos propuestos en un sistema de Diagnóstico Dirigido por Computadora (CAD), con el fin de aprovechar las fortalezas complementarias de cada modelo implementado. Adicionalmente, con los hallazgos obtenidos, las perspectivas del trabajo se deberán enfocar en ampliar el conjunto de datos, para hacer exploraciones más exhaustivas que permitan dirigir los resultados a un despliegue en un escenario clínico.

## BIBLIOGRAFÍA

AL., Ferlay J et. *Global Cancer Observatory: Cancer Today (version 1.1)*. <https://gco.iarc.who.int/today>. Lyon, France: International Agency for Research on Cancer. Accessed: [26 08 2024]. 2024 (cit. on pp. 12, 15, 41).

AL., Qinghua Zhou et. “The Society for Translational Medicine: indications and methods of percutaneous transthoracic needle biopsy for diagnosis of lung cancer”. In: *Journal of Thoracic Disease* 10.9 (2018) (cit. on pp. 21, 41).

AL., Xin Huang et. “TabTransformer: Tabular Data Modeling Using Contextual Embeddings”. In: *CoRR* abs/2012.06678 (2020). arXiv: [2012.06678](https://arxiv.org/abs/2012.06678) (cit. on p. 53).

ALJOHANI ABEER, Alharbe Nawaf. “Generating Synthetic Images for Healthcare with Novel Deep Pix2Pix GAN”. In: *Electronics* 11.21 (2022), p. 3470. DOI: [10.3390/electronics11213470](https://doi.org/10.3390/electronics11213470) (cit. on p. 32).

AMERICAN CANCER SOCIETY. “Lung Cancer Risk Factors”. In: *American Cancer Society* (2025). <https://www.cancer.org/cancer/lung-cancer/causes-risks-prevention/risk-factors.html> [Accedido: 13-mar-2025] (cit. on p. 15).

BAKR S., Gevaert O. *Data for NSCLC Radiogenomics (Version 4)*. 2017. DOI: [10.7937/K9/TCIA.2017.7hs46erv](https://doi.org/10.7937/K9/TCIA.2017.7hs46erv) (cit. on pp. 19, 23).

BEN-COHEN, Avi et al. “Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection”. In: *Engineering Applications of Artificial Intelligence* 78 (2019), pp. 186–194 (cit. on p. 18).

BOCHKOVSKIY, Alexey and WANG, Chien-Yao. “YOLOv4: Optimal Speed and Accuracy of Object Detection”. In: *arXiv preprint arXiv:2004.10934* (2020) (cit. on p. 25).

BOULAHIA, Said Yacine et al. “Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition”. In: *Machine Vision and Applications* 32.6 (2021), p. 121 (cit. on p. 28).

BRADLEY STEPHEN H Kennedy Martyn PT, Neal Richard D. “Recognising lung cancer in primary care”. In: *Advances in therapy* 36.1 (2019), pp. 19–30 (cit. on p. 12).

BRONCONEUMOLOGÍA, Archivo de. “El diagnóstico y manejo de los nódulos pulmonares en adultos”. In: *Archivos de Bronconeumología* 56.7 (2020). Accedido: 2025-03-19, pp. 452–461 (cit. on p. 16).

CAICEDO CARLOS Garcia-Velloso Maria Jose, Lozano Maria Dolores. “Role of [18 F] FDG PET in prediction of KRAS and EGFR mutation status in patients with advanced non-small-cell lung cancer”. In: *European journal of nuclear medicine and molecular imaging* 41 (2014), pp. 2058–2065 (cit. on p. 22).

CARDONA DIANA, Mejía Ricardo and VIOLA ANTONIO Rojas Pedro, Serna Fernando. “Lung cancer in Colombia”. In: *Journal of Thoracic Oncology* 17.6 (2022), pp. 719–727. DOI: [10.1016/j.jtho.2022.02.015](https://doi.org/10.1016/j.jtho.2022.02.015) (cit. on p. 15).

CENTER, Roswell Park Comprehensive Cancer. *When to Worry About Lung Nodules*. Accedido: 2025-03-19. 2024 (cit. on p. 16).

CHEN, H. et al. “CenterNet-based multimodal attention for lung nodule detection”. In: *European Radiology* 32 (2022), pp. 1234–1243 (cit. on p. 35).

DEL CIELLO Franchi A, Contegiacomo P. “Missed lung cancer: when, where, and why?” In: *Diagnostic and Interventional Radiology (Ankara, Turkey)* (2017), pp. 118–126 (cit. on p. 41).

EMILIO CARRIZOSA, Dolores Romero Morales. “Supervised classification and mathematical optimization”. In: *Computers & Operations Research* 40 (2013), pp. 150–165. DOI: [10.1016/j.cor.2012.05.015](https://doi.org/10.1016/j.cor.2012.05.015) (cit. on p. 27).

FERNÁNDEZ-ARRIETA ALEXANDER, Martinez-Jaramillo Stella Isabel. “Características clinicopatológicas de nódulos pulmonares: Experiencia en Clínica Reina Sofía, Bogotá, Colombia”. In: *Revista Colombiana de Cirugía* 37 (2022), pp. 49–59. DOI: [10.30944/20117582.903](https://doi.org/10.30944/20117582.903) (cit. on p. 17).

GEVAERT OLIVIER Echegaray Sebastian, Khuong Amanda. “Predictive radiogenomics modeling of EGFR mutation status in lung cancer”. In: *Scientific reports* 7.1 (2017), p. 41674 (cit. on p. 38).

GOODFELLOW, Ian et al. “Generative adversarial networks”. In: *Communications of the ACM* 63.11 (2020), pp. 139–144 (cit. on p. 30).

GUOBIN ZHANG Shan Jiang, Zhiyong Yang. “Automatic nodule detection for lung cancer in CT images: A review”. In: *Computers in Biology and Medicine* 103 (2018). DOI: <https://doi.org/10.1016/j.combiomed.2018.10.033> (cit. on p. 41).

HAO, K. et al. “Lung nodule false positive reduction using a central attention convolutional neural network on imbalanced data”. In: *Proceedings of SPIE Medical Imaging*. Vol. 12466. 2023, p. 124661X. DOI: [10.1117/12.2654216](https://doi.org/10.1117/12.2654216) (cit. on p. 35).

HILLNER BRUCE E Siegel Barry A, Liu. “Impact of Positron Emission Tomography/Computed Tomography and Positron Emission Tomography (PET) Alone on Expected Management of Patients With Cancer: Initial Results From the National Oncologic PET Registry”.

In: *Journal of Clinical Oncology* (2008). DOI: [10.1200/JCO.2007.14.5631](https://doi.org/10.1200/JCO.2007.14.5631) (cit. on pp. 12, 41).

HUANG Shih-Cheng, Et al. “Self-supervised learning for medical image classification: a systematic review and implementation guidelines”. In: *NPJ Digital Medicine* (2023). DOI: [10.1038/s41746-023-00811-0](https://doi.org/10.1038/s41746-023-00811-0) (cit. on p. 29).

HUANG CUIQING, Liang Jianye. “Diagnostic Performance of Perfusion Computed Tomography for Differentiating Lung Cancer from Benign Lesions: A Meta-Analysis”. In: *Med Sci Monit* 25 (2019), pp. 3485–3494. DOI: [10.12659/MSM.914206](https://doi.org/10.12659/MSM.914206) (cit. on p. 17).

IYER R SUMANTH Needham Sarah R, Galdadas Ioannis. “Drug-resistant EGFR mutations promote lung cancer by stabilizing interfaces in ligand-free kinase-active EGFR oligomers”. In: *Nature Communications* 15.1 (2024), p. 2130 (cit. on p. 20).

JIN, Qiangguo et al. “Free-form tumor synthesis in computed tomography images via richer generative adversarial network”. In: *Knowledge-Based Systems* 218 (2021), p. 106753 (cit. on p. 39).

JOCHER, Glenn; STOKEN, Alex, and CHAURASIA. *YOLOv5: Implementation of YOLO for object detection*. GitHub repository. 2021 (cit. on p. 25).

JOHNSON, B. et al. “Genomic Biomarkers in Lung Cancer: Clinical Applications”. In: *Cancer Genomics Journal* 54 (2022), pp. 200–215. DOI: [10.1002/cgj.22045](https://doi.org/10.1002/cgj.22045) (cit. on p. 16).

JUEFEI-XU, Felix; BODDETI, Vishnu Naresh, and SAVVIDES, Marios. “Gang of GANs: Generative Adversarial Networks with Maximum Margin Ranking”. In: *arXiv preprint arXiv:1704.04865* (2017). Accessed: 2025-05-01 (cit. on p. 30).

KIM, Seonhwa et al. “Deep learning–radiomics integrated noninvasive detection of epidermal growth factor receptor mutations in non-small cell lung cancer patients”. In: *Scientific Reports* 14.1 (2024), p. 922 (cit. on pp. 20, 37).

KUANG YANG Liu Yuxin, Wang Zhi. “Unsupervised Multi-Discriminator Generative Adversarial Network for Lung Nodule Malignancy Classification”. In: *Medical Image Analysis* 64 (2020), p. 101713. DOI: [10.1016/j.media.2020.101713](https://doi.org/10.1016/j.media.2020.101713) (cit. on p. 39).

LAZEBNIK, Teddy and SIMON-KEREN, Liron. “Cancer-inspired genomics mapper model for the generation of synthetic DNA sequences with desired genomics signatures”. In: *Computers in Biology and Medicine* (2023). DOI: <https://doi.org/10.1016/j.compbimed.2023.107221> (cit. on p. 38).

LEVANTINI ELENA Maroni Giorgia, Del Re Marzia. “EGFR signaling pathway as therapeutic target in human cancers”. In: *Seminars in cancer biology*. Vol. 85. Elsevier. 2022, pp. 253–275 (cit. on p. 20).

LIU, Jiao; KANG, Rui, and TANG, Daolin. “The KRAS-G12C inhibitor: activity and resistance”. In: *Cancer gene therapy* 29.7 (2022), pp. 875–878 (cit. on p. 20).

LIU, Jin et al. “Detection of pulmonary nodules with YOLOv3 in chest CT”. In: *Quantitative Imaging in Medicine and Surgery* 10.10 (2020), pp. 1917–1929 (cit. on pp. 13, 33).

MAMMERI, A. et al. “YOLOv7 and VGG16-based classification of pulmonary nodules”. In: *Scientific Reports* 14 (2024), p. 5678 (cit. on p. 34).

NAIR JAY KUMAR RAGHAVAN Saeed Umar Abid, McDougall Connor C. “Radiogenomic models using machine learning techniques to predict EGFR mutations in non-small cell lung cancer”. In: *Canadian Association of Radiologists Journal* 72.1 (2021), pp. 109–119 (cit. on p. 37).

NEUBECK, Axel and VAN GOOL, Luc. “Efficient Non-Maximum Suppression”. In: *Proceedings of the International Conference on Pattern Recognition (ICPR)*. IEEE, 2006, pp. 850–855 (cit. on p. 26).

ONISHI, Yuya et al. “Automated Pulmonary Nodule Classification in Computed Tomography Images Using a Deep Convolutional Neural Network Trained by Generative Adversarial Networks”. In: *BioMed Research International* 2019 (Jan. 2019), pp. 1–9. DOI: [10.1155/2019/6051939](https://doi.org/10.1155/2019/6051939) (cit. on p. 31).

PHILLIP ISOLA Jun-Yan Zhu, Tinghui Zhou. “Image-to-Image Translation with Conditional Adversarial Networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 1125–1134. DOI: [10.1109/CVPR.2017.632](https://doi.org/10.1109/CVPR.2017.632) (cit. on p. 33).

PINHEIRO, Gil et al. “Identifying relationships between imaging phenotypes and lung cancer-related mutation status: EGFR and KRAS”. In: *Scientific reports* 10.1 (2020), p. 3625 (cit. on pp. 38, 40).

PRATAP CHANDRA SEN Mahimarnab Hajra, Mitadru Ghosh. “Supervised Classification Algorithms in Machine Learning: A Survey and Review”. In: *Emerging Technology in Modelling and Graphics*. Ed. by J. K. MANDAL and D. BHATTACHARYA. Vol. 937. Advances in Intelligent Systems and Computing. Springer Nature Singapore Pte Ltd., 2020, pp. 99–109. DOI: [10.1007/978-981-13-7403-6\\_11](https://doi.org/10.1007/978-981-13-7403-6_11) (cit. on p. 27).

RECK, M et al. “Targeting KRAS in non-small-cell lung cancer: recent progress and new approaches”. In: *Annals of Oncology* 32.9 (2021), pp. 1101–1110 (cit. on p. 20).

REDMON, Joseph and FARHADI, Ali. “YOLO9000: Better, Faster, Stronger”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 7263–7271 (cit. on p. 24).

REDMON JOSEPH Divvala Santosh, Girshick Ross. “You Only Look Once: Unified, Real-Time Object Detection”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 779–788 (cit. on p. 23).

REN, Ge et al. “A Transfer Learning Framework for Deep Learning-Based CT-to-Perfusion Mapping on Lung Cancer Patients”. In: *Frontiers in Oncology* 12 (2022). DOI: [10.3389/fonc.2022.883516](https://doi.org/10.3389/fonc.2022.883516) (cit. on p. 39).

RIZZO STEFANIA Petrella Francesco, Buscarino Valentina. “CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer”. In: *European radiology* 26 (2016), pp. 32–42 (cit. on pp. 21, 22).

RIZZO STEFANIA Raimondi Sara, de Jong Evelyn EC. “Genomics of non-small cell lung cancer (NSCLC): Association between CT-based imaging features and EGFR and K-RAS mutations in 122 patients—An external validation”. In: *European journal of radiology* 110 (2019), pp. 148–155 (cit. on p. 22).

ROBERTS, J. et al. “Advances in Lung Cancer Imaging: The Role of CT”. In: *Journal of Oncology Imaging* 45 (2023), pp. 120–135. DOI: [10.1016/j.joi.2023.04.003](https://doi.org/10.1016/j.joi.2023.04.003) (cit. on p. 16).

RUBIN, Geoffrey D. “Lung Nodule and Cancer Detection in Computed Tomography Screening”. In: *Journal of Thoracic Imaging* 30.2 (2015), pp. 130–138. DOI: [10.1097/RTI.000000000000140](https://doi.org/10.1097/RTI.000000000000140) (cit. on p. 17).

SALEHJAHROMI MORTEZA, Karpinets Tatiana V. “Synthetic PET from CT improves diagnosis and prognosis for lung cancer: Proof of concept”. In: *Cell Reports Medicine* (2024). DOI: [10.1016/j.xcrm.2024.101463](https://doi.org/10.1016/j.xcrm.2024.101463) (cit. on pp. 18, 38, 39).

SETIO, A. A. A. et al. “Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks”. In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1160–1169. DOI: [10.1109/TMI.2016.2536809](https://doi.org/10.1109/TMI.2016.2536809) (cit. on p. 35).

SHARMA PARUL Sharma Siddharth, Sobti Ranbir Chander. “Molecular Pathogenesis of Lung Cancer: Evolving Landscapes Based on Past Experience”. In: *Molecular Biomarkers for Cancer Diagnosis and Therapy*. Springer, 2024, pp. 347–380 (cit. on p. 20).

SHIRI ISAAC Amini Mehdi, Nazari Mostafa. “Impact of feature harmonization on radiogenomics analysis: Prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images”. In: *Computers in biology and medicine* 142 (2022), p. 105230 (cit. on pp. 13, 37).

SHURRAB, Saeed and DUWAIRI, Rehab. “Self-supervised learning methods and applications in medical imaging analysis: a survey”. In: *PeerJ Computer Science* (2022), e1045. DOI: [10.7717/peerj-cs.1045](https://doi.org/10.7717/peerj-cs.1045) (cit. on p. 29).

SIBILLE, L. et al. “Multimodal deep learning for lung lesion detection in PET/CT”. In: *Journal of Nuclear Medicine* 61.4 (2020), pp. 521–527 (cit. on p. 35).

SILVA FRANCISCO Pereira Tania, Morgado Joana. “EGFR assessment in lung cancer CT images: analysis of local and holistic regions of interest using deep unsupervised transfer learning”. In: *IEEE Access* 9 (2021), pp. 58667–58676 (cit. on p. 36).

SMITH, A. et al. “PET Imaging in Lung Cancer: A Review of Recent Studies”. In: *Lung Cancer Research* 38 (2021), pp. 88–102. DOI: [10.1016/j.lcr.2021.06.007](https://doi.org/10.1016/j.lcr.2021.06.007) (cit. on p. 16).

SONG, Han-Seop; MUGABI, Jophous, and JEONG, Jae-Ho. “Pix2Pix and Deep Neural Network-Based Deep Learning Technology for Predicting Vortical Flow Fields and Aerodynamic Performance of Airfoils”. In: *Applied Sciences* 13.2 (2023), p. 1019. DOI: [10.3390/app13021019](https://doi.org/10.3390/app13021019) (cit. on p. 32).

SUN, Benyuan et al. “Integrating Feature and Image Pyramid: A Lung Nodule Detector Learned in Curriculum Fashion”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer. 2018, pp. 683–691 (cit. on p. 35).

SUN WEI Yuan Xun, Tian Yijun. “Non-invasive approaches to monitor EGFR-TKI treatment in non-small-cell lung cancer”. In: *Journal of hematology & oncology* 8 (2015), pp. 1–9 (cit. on pp. 21, 41).

TAKAMOCHI KAZUYA Mogushi Kaoru, Kawaji Hideya. “Correlation of EGFR or KRAS mutation status with 18F-FDG uptake on PET-CT scan in lung adenocarcinoma”. In: *PLoS One* 12.4 (2017), e0175622 (cit. on pp. 13, 22).

TONGHE WANG Yang Lei, Yabo Fu. “A review on medical imaging synthesis using deep learning and its clinical applications”. In: *Journal of Applied Clinical Medical Physics* 22.1 (2020), pp. 11–36. DOI: [10.1002/acm2.13121](https://doi.org/10.1002/acm2.13121) (cit. on p. 38).

TRIPATHI SATVIK Moyer Ethan Jacob, Augustin Alisha Isabelle. “Radgennets: Deep learning-based radiogenomics model for gene mutation prediction in lung cancer”. In: *Informatics in medicine unlocked* 33 (2022), p. 101062 (cit. on pp. 14, 38, 40).

UNG YEE C. Maziak Donna E., Vanderveen. “18 Fluorodeoxyglucose Positron Emission Tomography in the Diagnosis and Staging of Lung Cancer: A Systematic Review”. In: *JNCI: Journal of the National Cancer Institute* (Dec. 2007). DOI: [10.1093/jnci/djm232](https://doi.org/10.1093/jnci/djm232). eprint:

<https://academic.oup.com/jnci/article-pdf/99/23/1753/7689425/djm232.pdf>

(cit. on p. 18).

URREHMAN, Zia et al. “Dual attention CNN for pulmonary nodule detection in CT”. In: *European Radiology* 33 (2023), pp. 4567–4575 (cit. on p. 36).

USUI, Keisuke et al. “A cycle generative adversarial network for improving the quality of four-dimensional cone-beam computed tomography images”. In: *Radiation Oncology* 17.1 (2022), p. 69 (cit. on p. 33).

WANG SHUO Shi Jingyun, Ye Zhaoxiang. “Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning”. In: *European Respiratory Journal* 53.3 (2019) (cit. on pp. 36, 40).

WANG TONGHE, Yang Xiaofeng. “Take CT, get PET free: AI-powered breakthrough in lung cancer diagnosis and prognosis”. In: *Cell Reports Medicine* 5 (2024). <https://doi.org/10.1016/j.xcrm.2024.101486>, accessed [11 09 2024], p. 2. DOI: [10.1016/j.xcrm.2024.101486](https://doi.org/10.1016/j.xcrm.2024.101486) (cit. on pp. 39, 41).

WANG ZHENGWEI She Qi, Ward Tomás E. “Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy”. In: *ACM Computing Surveys (CSUR)* 54.2 (2020), pp. 1–38. DOI: [10.1145/3439723](https://doi.org/10.1145/3439723). arXiv: [1906.01529](https://arxiv.org/abs/1906.01529) [cs.CV] (cit. on p. 30).

WU, X. et al. “YOLOv7 with multi-scale receptive field for small pulmonary nodule detection”. In: *Pattern Recognition* 145 (2024), p. 109876 (cit. on pp. 13, 34).

WU JINGRAN Meng Hao, Zhou Lin. “Habitat radiomics and deep learning fusion nomogram to predict EGFR mutation status in stage I non-small cell lung cancer: a multicenter study”. In: *Scientific Reports* 14.1 (2024), p. 15877 (cit. on p. 37).

WU YIHUA Li Peiwei, Zhang Honghe. “Diagnostic value of fluorine 18 fluorodeoxyglucose positron emission tomography/computed tomography for the detection of metastases in non-small-cell lung cancer patients”. In: *International Journal of Cancer* (2013). DOI: <https://doi.org/10.1002/ijc.27779> (cit. on pp. 18, 41).

YUAN MIN Huang Li-Li, Chen Jian-Hua. “The emerging treatment landscape of targeted therapy in non-small-cell lung cancer”. In: *Signal transduction and targeted therapy* 4.1 (2019), p. 61 (cit. on pp. 13, 20).

ZHANG, R. et al. “Ensemble of 3D CNNs for enhanced pulmonary nodule detection”. In: *European Radiology* 33 (2023), pp. 3456–3465 (cit. on p. 40).

ZHAO WEI Chen Weidao, Li Ge. “GMILT: a novel transformer network that can noninvasively predict EGFR mutation status”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022) (cit. on p. 36).

ZHENG, Zhaohui et al. “Distance-IoU loss: Faster and better learning for bounding box regression”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. 07. 2020, pp. 12993–13000 (cit. on p. 26).

ZHOU, Y. et al. “Improved YOLOv5 with Transformers for pulmonary nodule detection”. In: *Scientific Reports* 14 (2024), p. 1234 (cit. on p. 34).

ZHOU MU Leung Ann, Echegaray. “Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications”. In: *Radiology* 286.1 (2018), pp. 307–315 (cit. on p. 37).